APEX STANDARDS
AI基礎設施 · 市場 · 標準情報
C20004202 · 公開繁體中文版 · 2026年10月3日

峰會前簡報 · OCP全球峰會（OCP Global Summit），聖荷西，2026年10月12日至15日 · 研究資料截止日2026年10月2日

OCP聖荷西2026：AI基礎設施從機櫃邁向電網

加速器（accelerator）已不再單獨決定步調。記憶體、封裝、電力、冷卻、光學、電網併接與資本，如今共同制約交付進程，從機櫃設計到場址送電皆然。

運算 → 系統 | 48 V → 800 VDC | 氣冷 → 液冷 | GPU稀缺 → 記憶體、電力與電網稀缺 | 壓力的方向，而非完全取代

變動內容

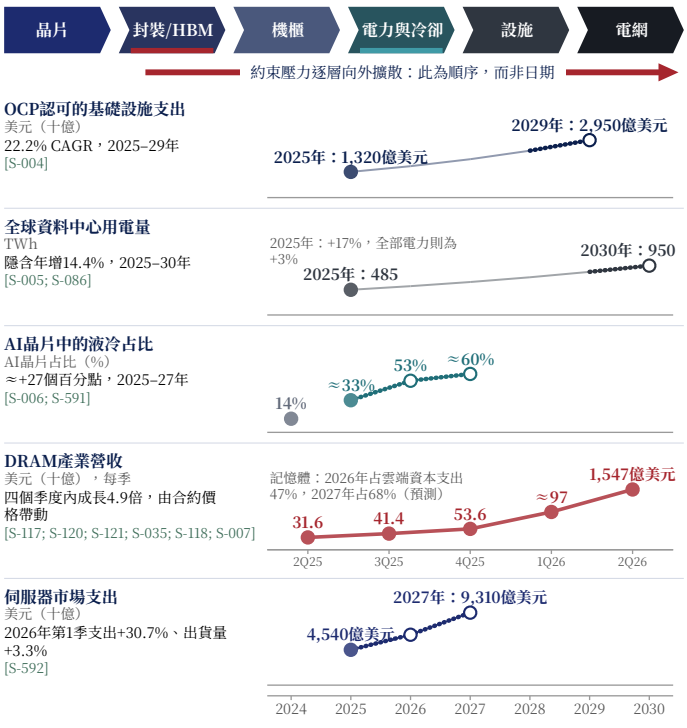
AI基礎設施競賽正延伸至加速器（accelerator）之外。 運算需求仍快速攀升，但一套可部署的AI系統如今取決於更長的鏈條：高頻寬記憶體（high-bandwidth memory, HBM）、先進封裝（advanced packaging）、機櫃供電與網路、液冷（liquid cooling），以及已送電的電網併接。Open Compute Project（OCP）的工作循此一邊界展開：由AMD、Google、Meta與NVIDIA制定的AI設施設計原則已於2026年8月27日生效[S-001]。

稀缺與議價能力（bargaining power）正隨之移動。 即使加速器（accelerator）供應無虞，記憶體、封裝產能、雷射、渦輪機產能時段與電網併接仍可能延後營收實現。NVIDIA的供應承諾在一個季度內增加逾一倍，達2,790億美元，主要用於記憶體[S-002]。買方已將此計入預算：Microsoft將其2026年4月資本支出（capex）計畫中約250億美元歸因於零組件價格，Meta則將其2026年資本支出預測增幅的大部分歸因於零組件成本，尤以記憶體為主[S-045; S-046]。緯穎（Wiwynn）為因應零組件成本上升而調整其商業模式，自2026年4月起依IFRS 15以代理人身分為特定客戶採購（procurement）記憶體，故該等金額不計入其營收與銷貨成本[S-063; S-483; S-484]。掌握稀缺環節（scarce step）者即取得籌碼（leverage）。

下一道檢驗是驗證（qualification），而非發明。 關鍵在於哪一項技術能夠從展示走向已驗證、多供應商的現場運行，且速度足以紓解下一個瓶頸。開放介面可降低轉換成本（switching cost），但無法創造產能。各層級各依自身時鐘增添產能；機會與延誤風險，均落在這些時鐘之間的落差（第1章）。

圖1。 AI基礎設施的稀缺正從晶片擴散至電網

五項指標顯示資本、稀缺與驗證壓力正往何處移動。



五項指標置於同一時間軸，DRAM列以季為單位，各自採用其自身的單位與尺度。實心標記為已結束期間的估計值，點狀標記為第三方預測；細線為推導橋接值，即已公布或隱含的複合年成長率（compound annual growth rate, CAGR）在兩個已公布端點之間所給出的數值，並非預測。色帶呈現的是約束壓力的順序，而非日期。

範圍 OCP認可支出係市場研究公司IDC為OCP所作的預測，涵蓋依OCP所認可規格（specification）製造的設備，並非AI支出總額。2024年的液冷（liquid cooling）數值為TrendForce在AI資料中心的滲透率，此一基礎與其2025–27年AI晶片占比數列所採者不同，故兩者不相連接。

成長校準：基礎各異，不可相加

全球GDP 2026年3.0%、2027年3.4%（IMF預測） | **全部電力** 2025年+3% | **資料中心用電量** 2025年+17% | **OCP認可支出** 2025–29年複合年成長率22.2% | **實體基礎設施** 2025–30年複合年成長率22% | **伺服器支出** 2026年+42.7%（IDC推估） | **DRAM營收** 2025年第2季至2026年第2季+389% [S-593; S-086; S-004; S-594; S-592; S-117; S-118]

經濟校準 CAGR · GDP貢獻度 · 資本密集度 · dot-com時代對比 → 附錄E

代表意義

財務。 資本正從運算擴及記憶體、電力與場址；凡採購（procurement）或會計基礎改變之處，解讀利潤率與成長均須謹慎。
→ 第2.1、2.2、3.2、5.3、8.4與9.3節；附錄E探討成長率、GDP與dot-com時代

產業。 部署所需時間愈來愈取決於驗證（qualification）、預訂，以及加速器之外的稀缺投入要素。
→ 第5.4–5.6、6.1與8.3節；附錄A

產品。 記憶體、電力、冷卻與網路正成為相互耦合的藍圖選擇；後期變動會擴散至整座機櫃。
→ 第3.3、4.1–4.7與8.2–8.3節；附錄A

研發。 最困難的問題落在各個介面上：電力轉換、熱控制、光學、遙測（telemetry）、測試與可靠性。
→ 第3.4、4.2、4.3、4.6、4.7、6.5與8.2節

創新。 當一項可運作的技術尚未成為可複製、已驗證的多供應商量產系統時，機會最大。
→ 第1.4、3.5與6.7節；附錄B

智慧財產權。 隨著介面開放，差異化轉向實作、材料、控制、封裝、測試、良率（yield）與服務。
→ 第6.1、6.3、6.6與6.7節

政府。 電網併接、出口管制（export control）、認證（certification）與區域集中度，可決定部署於何處、於何時發生。
→ 第7.1、7.2、7.4、7.5、8.5與9.6節

目錄、標記與來源的閱讀方式，以及依領域劃分的閱讀路徑，均見第2頁。

六個數字

1,320億美元 → 2,950億美元
OCP認可的資訊科技基礎設施支出，2025年 → 2029年（IDC為OCP所作預測）[S-004]

890億美元
NVIDIA Data Center營收，截至2026年7月26日之當季（2027財年第2季）[S-010; S-002]

11.33%對11.49%
美國智慧財產與設備投資占GDP比重：2026年第1季對照2000年高峰（聯準會札記、BEA資料）[S-595]

3.7 GW中的1.5 GW
CoreWeave於2026年6月30日的已啟用電力對比已簽約電力[S-011]

116 GW
GE Vernova於2026年6月30日的燃氣發電設備未出貨訂單（backlog）與產能時段預訂[S-012]

3,180億美元
2025年AI基礎設施支出，為2024年的兩倍以上；2026年預測為4,870億美元（IDC）[S-596]

後續觀察重點

1 **記憶體。** 伺服器DRAM與HBM合約價格、在一家以上供應商完成驗證（qualification），以及新產能的爬坡。價格若持續鬆動，將削弱記憶體的議價能力（bargaining power）。

2 **800 VDC與冷卻。** 來自無關聯供應商的量產硬體、安全認證（certification），以及OCP認可的實作：此即參考設計（reference design）與可規模化採用之間的分界。

3 **電網與電力設備。** 併網排隊佇列（interconnection queue）、渦輪機與開關設備的未出貨訂單（backlog），以及大型負載（large load）規則。最慢的實體層級，可能為其餘已可取得的運算能力設下上限。

讀者須知

本報告彙整公開的技術、市場與產業資訊，供資訊與教育用途。有來源支持的事實與分析及條件式展望分別呈現，並提供來源網址以供獨立審閱。本報告自身的前瞻性觀點均屬情境（scenario），而非預測或機率；第三方預測均以預測形式呈現，並歸屬於其發布者。讀者應自行評估底層紀錄，並形成自己的結論。

F 來自註明日期之公開來源的事實 | A 對所引事實的分析 | IF 條件式展望 | ? 未解，尚未確立 | = 推導計算 | [S-001] 來源編號，連結至附錄D中該來源的卡片

目錄

九章及其各節，以及五項附錄；每一條目皆為連結。

1 執行觀點：變動內容及其重要性何在 3
1.1 AI基礎設施競賽正延伸至加速器之外 3
1.2 四項結構性轉變 3
1.3 十一項發現 4
1.4 約束在何處形成研究課題 5
1.5 三項風險與三項訊號 5
2 需求、資本與市場規模 7
2.1 資本支出持續上升，但會計基礎各異 7
2.2 未出貨訂單、已簽約電力及其背後的融資 7
2.3 AI伺服器與NVIDIA訊號 9
2.4 OCP認可市場所衡量的對象 10
2.5 電力使AI成長成為一項能源系統課題 11
3 運算、記憶體與封裝 13
3.1 GPU與客製晶片同步成長 13
3.2 記憶體除了成為效能限制，也成為預算限制 13
3.3 為何決定節奏的是記憶體，而非算術運算 15
3.4 先進封裝決定了能夠建置多少運算能力 16
3.5 新基板、鍵合與測試顯示複雜度移向何處 17
4 機櫃架構：電力、冷卻與網路 19
4.1 機櫃成為產品本身 19
4.2 自48 V至±400 V與800 VDC 19
4.3 液體進入機櫃 20
4.4 設施與電網介面 22
4.5 機櫃外型規格相互競爭，而目前並存 23
4.6 垂直擴展交換網路維持多元，其較低層級則趨於收斂 24
4.7 光學正朝封裝靠近 25
5 供應鏈與價值擷取 28
5.1 價值落在產能持續稀缺之處 28
5.2 ODM靠規模、產品組合與已改變的營收基礎而成長 28
5.3 企業數字所涵蓋與未涵蓋者 29
5.4 電力、冷卻與光學零組件均為預售 29
5.5 渦輪機與電網設備依更長的時鐘運行 30
5.6 誰能讓一項部署停擺 31
5.7 沒有任何區域掌握所有層級 31
5.8 資本追隨瓶頸 33
6 標準、互通性與智慧財產權 35
6.1 OCP標準化了什麼，又有什麼是它無法做到的 35
6.2 聖荷西2026能夠顯示什麼，又不能顯示什麼 35
6.3 OCP所倚賴的介面由其他機構擁有 36
6.4 信任與可管理性成為採購要求 36
6.5 邊緣與電信遇上機櫃 37
6.6 專利體制：開放介面，專屬實作 37
6.7 介面開放時差異化移往何處 38
7 政策、能源與國家層級曝險 41
7.1 出口管制已出現於企業規劃之中 41
7.2 電網併接與大型負載規則決定時程 42
7.3 電網必須滿足的能源需求 43
7.4 800 VDC與液體的安全與認證 44
7.5 國家層級曝險：集中度與韌性 45
8 開放權重模型、記憶體與硬體 47
8.1 中國的開放權重推動精簡記憶體的設計 47
8.2 五項設計選擇降低每個token的記憶體用量 47
8.3 這些組態對記憶體的意涵 48
8.4 使用量的成長速度大致與每token記憶體用量的下降速度相當 48
8.5 中國的硬體與其模型相互形塑 50
8.6 至2035年記憶體需求的條件式路徑 50
9 展望：2027年、2030年、2035年 52
9.1 基準路徑與三條替代路徑 52
9.2 2027年：各項計畫迎向其產能時點 52
9.3 2030年：電力、資本與封裝決定步調 53
9.4 2035年：結構而非時程 55
9.5 至2035年的路標 55
9.6 何者會改變此一觀點 56
附錄A 技術規格狀態 58
附錄B 企業與區域圖譜 61
附錄C 方法與定義 66
附錄D 來源查核登錄表 74
附錄E 經濟與財務校準：成長、GDP、資本循環與dot-com時代 132

圖目次

圖1。 AI基礎設施的稀缺正從晶片擴散至電網 1
圖2。 AI基礎設施的各個層級依不同時鐘增添產能 4
圖3。 具約束力的限制已從晶片與封裝擴散至記憶體、雷射、渦輪機與電網 5
圖4。 每一家大型AI買方的資本支出於2026年躍升，但各依其自身的會計基礎 8
圖5。 2026年的資本計畫在年內持續上修，惟會計基礎變動者除外 9
圖6。 伺服器支出的成長遠快於伺服器出貨台數 9
圖7。 NVIDIA的Data Center營收在十個季度內增至近四倍 9
圖8。 已簽約需求的成長超越已交付容量，而各公司所用指標並不相同 10
圖9。 OCP認可基礎設施的支出預估將於2025年至2029年間成長逾一倍 11
圖10。 IDC與Gartner對AI基礎設施支出的估計相差三倍，但兩者均認為2026年將成長約一半 11
圖11。 全球資料中心用電量預估將於2025年至2030年間大致倍增 11
圖12。 客製加速器在AI伺服器中的占比上升，是與GPU並存，而非取代GPU 13
圖13。 記憶體在主要雲端業者資本支出中的占比正在擴大 14
圖14。 合約價格飆升，DRAM產業營收在四個季度內增至近五倍 14
圖15。 SK hynix的營業利益率在四個季度內自41%上升至76% 14
圖16。 記憶體相關設備支出至2028年持續上升，是對稀缺較慢的回應 15
圖17。 ±400 VDC路徑已有規格；原生800 VDC則取決於已公告的時間區間與條件 20
圖18。 液冷正成為AI晶片的預設配置 21
圖19。 從1,400 W晶片到2 MW CDU，液體散熱路徑的每一段如今都有已公布的規格範圍 22
圖20。 交換網路的各項工作在通道層與鏈路層趨於收斂，在交易層則維持多元 25
圖21。 AI叢集光學2025年成長60%，2026年預測亦為60%，並在前五大雲端業者資本支出中占比持續上升 25
圖22。 緯穎的營收與利潤率跨越2026年4月的基礎變動 28
圖23。 超大規模雲端業者持有近半數的資料中心容量，並預測至2031年將持有三分之二 31
圖24。 除加速器供應商外，另有數方能讓一項AI部署停擺 32
圖25。 各區域均依賴他處所掌握的層級 32
圖26。 已發布的標準仍在等待產品，而已出貨的光學則運行於封閉的供應商鏈上 39
圖27。 LBNL的參考情境使2030年美國資料中心用電量增至三倍以上 43
圖28。 TrendForce預估2030年資料中心電力需求與可取得電網容量之間將出現268 GW的缺口 45
圖29。 自H100以來，每具GPU的HBM已成長3.6倍 49
圖30。 2026年推論在AI雲端支出上超越訓練 49
圖31。 token量在兩年內成長數倍 50
圖32。 註明日期的事件群聚於2030年之前；其後，展望所依憑者為條件，而非時程 56
圖33。 2026年的各項預測在年內遭上調：雲端資本支出、AI基礎設施，以及自7月起的AI伺服器成長率 133
圖34。 美國設備與智慧財產投資占GDP的比重已回升至接近2000年水準，且正以1990年代後期的速度上升 134
圖35。 AI相關類別2025年對美國成長的增益高於資訊科技類別於2000年的增益；此一上升由軟體帶動 134
圖36。 1990年代的熱潮是一段持續的高實質成長期；當前數列則於2025年急遽加速 134
圖37。 美國通訊設備投資於1990年代後期成長逾一倍，其後一年內下滑近三分之一 135
圖38。 在LBNL的各項情境中，AI伺服器的使用方式對2030年美國資料中心用電量的影響，大於增加ASIC 138

表目次

表1。 每一項約束都留下一道實作難題，研究與差異化仍可在其中移動 6
表2。 四種市場規模概念回答四個不同的問題 11
表3。 HBM供應商正預售產能並擴產HBM4，但無一業者以位元陳述其HBM產能 15
表4。 台積電附日期的藍圖延伸至2029年，而其支出財測指引則在年內上調 17
表5。 ±400 VDC路徑已有規格、800 VDC硬體已公告，但認證才剛起步 21
表6。 光學正沿多個時鐘同時向封裝靠近 27
表7。 沿AI基礎設施鏈條，價值池的性質與規模均不相同 29
表8。 企業數字的涵蓋範圍各不相同 30
表9。 附日期的企業承諾將於2027年至2029年間增添美國的晶圓、封裝與記憶體產能 34
表10。 每一受限層級均正在增添產能，其時鐘自一年延伸至五年以上 34
表11。 OCP撰寫系統；其下層的介面則屬於十餘個其他機構 37
表12。 僅OCP與（通常而言）CHIPS Alliance以免權利金約束貢獻者；JEDEC、IEEE、UEC、DMTF與SNIA則容許收取權利金 39
表13。 每一列都有已量化的產能與營收訊號；標準欄所記錄的是文件，而非已完成驗證的第二供應來源 40
表14。 各份展望與申報文件對中國的處理方式各不相同 43
表15。 出口法律工具繫於產品類別、當事方與目的地，而2027年4月之前有四個日期可能改變現況 44
表16。 開放權重旗艦模型結合稀疏專家、精簡注意力與低位元格式 48
表17。 每段對話的快取用量差距逾100倍 49
表18。 三條路徑中僅有一條使記憶體需求下降 51
表19。 企業自行揭露的預測期限在2028年之後趨於稀薄；2030年之後僅存一項投資指引 57
表20。 每一項發現均有一項具名的推翻訊號，以及一次註明日期的下次檢視 57
表21。 OCP的機櫃與冷卻迴路文件已達1.0版；800 VDC電力路徑則仍依憑草案與白皮書 59
表22。 介面速率的規格制定領先於出貨系統；800 VDC設備的認證才剛起步 60
表23。 五個階段標籤區分交付與意圖 62
表24。 平台供應商與買方決定了技術堆疊其餘環節必須達到的數量與日期 62
表25。 記憶體倚賴三家HBM供應商，而其新產能排定於數年之後 63
表26。 晶圓、封裝、微影與測試各自倚賴少數依計畫增添產能的供應商 63
表27。 機櫃整合商報告的快速成長，係基於彼此不可比較的營收基礎 63
表28。 電力與散熱供應商出貨的是當前設備，而多數800 VDC角色仍停留在已宣布階段 64
表29。 渦輪機製造商的產出與產能時段已預先承諾至數年之後 64
表30。 光學正向封裝體靠近，但多數新角色仍處於試點階段、研發或已宣布階段 65
表31。 本圖譜中沒有任何區域掌握所有層級 65
表32。 每一項標記都向讀者說明該段落所作陳述的類別，以及何種情形會使其有誤 67
表33。 顏色標示某一數值所描述的領域，同一色調並貫穿所有圖表 67
表34。 八種資料類別將已公布數字與財測指引、目標、預測及推導橋接值區分開來 67
表35。 五種前瞻性陳述各有不同的分量；情境絕不被研判為財測指引 68
表36。 每一個經濟用語只回答一個問題，用於其他問題即生誤導 133
表37。 十四項指標用以研讀資本循環，自總體規模到地緣政治；其中數項尚未以可比形式公開 135
表38。 兩輪循環在投資占GDP比重上相似，在組成、獲利、融資、能源與地緣政治上則有差異 137
表39。 六種資金管道使同一批產能出現在帳上的不同位置 137
表40。 四條經濟路徑指明此輪建置獲得驗證、移轉、趨於正常化或留下過剩的各項條件 137
表41。 每一預測期限各以不同變數研讀：先看執行與價格，最後看結構 138

本報告的閱讀方式

一條主線，四個層次。 第一頁以一分鐘說明整個系統；第1章為主線敘述與十一項發現；第2至8章為證據；第9章為展望；各附錄則為規格（specification）狀態、圖譜、方法、來源與經濟校準。

標記。 證據或研判（reading）段落均以一個標記起首。F：註明日期之公開來源所述內容，包含第三方預測在內，且該預測仍標示為預測；企業的主張則保留為該企業的主張。A：本報告對所引事實的分析。IF：條件式觀點，指明應觀察的事項及使其不成立的條件。?：公開紀錄未能確立的重大事項。=：推導計算，並引註其輸入數據。未標記的段落為敘述性內容。

來源與研判。 如[S-014]這類引註為公開來源編號，連結至來源查核登錄表（附錄D）中該來源的卡片；載有原始網址、該來源以綠色斜體呈現的原文（非英文者另附譯文）、其所支持的內容、其無法據以確立的內容，以及本報告引用之處。如A-01這類研判編號，則連結至詮釋登錄表中十四項研判之一的證據、推論邏輯與適用邊界。

圖表。 線條與標記表示數值的類別：已公布值為實線、估計值較淺、企業財測指引為虛線並搭配實心菱形、管理層目標為空心菱形、容量為方形、第三方預測為點線並搭配空心圓、條件式情境為灰色點線、已公布區間為帶狀、推導橋接值為細線。顏色表示系統的組成部分，紅色僅標示稀缺或價格壓力（第C.3節）。不同單位絕不共用同一數值軸；各項評分絕不相加。

階段與展望。 角色標示為產品、試點、研發、標準或已宣布；前瞻性觀點僅於所載明的條件（conditions）下成立，且不附機率。

閱讀路徑

財務與投資。 資本支出（capex）在不同基礎上的呈現（第2.1節）；未出貨訂單（backlog）與融資（第2.2節）；記憶體作為預算項目（第3.2節）；價值與數據邊界（第5.1與5.3節）；模型與記憶體（第8.4節）；資本與瓶頸（第5.8節）；至2030年的各條路徑（第9.3節）；成長、GDP與dot-com時代（附錄E）。

產業與採購。 誰能夠終止一項部署（第5.6節）；已預售的零組件（第5.4節）；渦輪機與電網設備（第5.5節）；記憶體分層（第8.3節）；驗證（qualification）與開放介面（第6.1節）；規格（specification）狀態（附錄A）。

產品與架構。 以機櫃為產品（第4.1節）；記憶體作為決定步調的資源（第3.3節）；48 V邁向800 VDC與液冷（liquid cooling）（第4.2–4.3節）；機櫃外型規格、交換網路與光學（第4.5–4.7節）；模型端的記憶體效率與記憶體分層（第8.2–8.3節）；規格成熟度（附錄A）。

技術與研發。 48 V邁向800 VDC（第4.2節）；液冷（liquid cooling）（第4.3節）；交換網路與光學（第4.6節、第4.7節）；封裝與記憶體（第3.4節、第3.3節）；記憶體精簡型模型（第8.2節）；邊緣與電信（第6.5節）。

創新。 以各項約束為研究課題（第1.4節）；基板、鍵合與測試（第3.5節）；企業圖譜與階段標籤（附錄B）。

智慧財產權與標準。 OCP所標準化的對象（第6.1節）；專利制度（第6.6節）；差異化的移轉方向（第6.7節）；其他機構的介面（第6.3節）。

政府與政策。 企業規劃中的出口管制（export control）（第7.1節）；中國的模型與晶片（第8.5節）；電網與大型負載（large load）規則（第7.2節）；安全與認證（certification）（第7.4節）；國家層級的曝險（第7.5節）；何者將改變此一觀點（第9.6節）。

第1章

執行觀點：變動內容及其重要性何在

AI基礎設施仍在擴張，但其具約束力的限制已從加速器（accelerator）移向記憶體、封裝、光學、電力、冷卻與電網；開放介面在各交接點發揮作用，而規劃時鐘之間的落差則決定時程。

適用讀者： 所有讀者。各項發現均指向深入闡述該發現的章節；附錄D列出十四項主要研判（reading）及其證據與適用邊界。

人工智慧（AI）基礎設施仍在快速成長，但加速器（accelerator）已不再是唯一供應短缺的投入要素。2026年，企業自身的報告亦將記憶體、先進封裝（advanced packaging）、雷射與電力設備列為限制因素，而電網業者正在訂定新規則以併接大型負載（large load）。產品的範圍已自晶片擴及機櫃與設施，且各層級各依自身時鐘增添產能：晶片以季為單位，晶圓廠與封裝產線以年為單位，渦輪機與電網併接則更久。本章說明變動內容、其背後的十一項發現，以及應觀察的訊號。本章要問的是：限制現在落在何處，接下來又將移往何方？

1.1 AI基礎設施競賽正延伸至加速器之外

2026年AI基礎設施的主線可分為六個環節：需求仍在上升；交付的限制自加速器（accelerator）向外移動；產品先成為機櫃、再成為設施；開放標準在系統交會之處發揮作用；稀缺使議價能力（bargaining power）移轉；以及稀缺下一步將移往何處的問題。有一個概念將這些環節串連起來：各層級各依自身時鐘增添產能。每一環節皆以註明日期的公開來源為依據，並於後續各章深入闡述。

F 需求仍在加速。 NVIDIA公布截至2026年7月26日當季的Data Center營收達創紀錄的890億美元，較一年前成長117%[S-002]。Alphabet於7月將其2026年資本支出（capital expenditure, capex）財測指引上調至1,950億至2,050億美元，先前4月所定為1,800億至1,900億美元，而該數字本身又取代了2月的1,750億至1,850億美元[S-014; S-015; S-016]；TrendForce亦將其2026年AI伺服器出貨成長預測上調至近31%[S-013]。國際能源總署（International Energy Agency, IEA）報告指出，資料中心用電量於2025年成長17%，並預估將自2025年的485 TWh約倍增至2030年的950 TWh[S-005]（第2章）。

A 瓶頸正向外移動。 加速器（accelerator）已不再是唯一限制產能到位速度的投入要素。NVIDIA的供應承諾在一個季度內自1,190億美元增至2,790億美元，「primarily related to the procurement of memory」[S-002]。TrendForce報告指出，CoWoS先進封裝（advanced packaging）自2023年起即供不應求[S-017]；Siemens Energy表示其燃氣渦輪機產能已售罄至2028財年（其財年於9月結束）[S-018]。此一限制如今自記憶體與封裝，歷經光學、電力設備與冷卻，延伸至電網併接，而一處場址的速度，取決於其最慢的投入要素何時到位（圖3）。

A 產品正成為機櫃，繼而成為設施。 國際能源總署（IEA）報告指出，AI伺服器的功率密度在2020年至2025年間上升十一倍，並預計至2027年將再上升四倍[S-005]。在此密度之下，電力轉換、液冷（liquid cooling）、網路與遙測（telemetry）均須與晶片一同設計，而非由建築後續加裝。Open Compute Project（OCP）的工作亦隨之而來：AMD、Google、Meta與NVIDIA撰寫了適用於AI設施的Open Data Center設計原則，於2026年8月27日生效[S-001]。接受工程設計、驗證（qualification）與融資的單元，正從伺服器擴及機櫃與場址（第4章）。

A 開放標準的作用在介面上，而不在稀缺的投入要素上。 共通介面使買方得以驗證第二供應來源，並使眾多系統整合商得以依同一份文本進行建置。Google、Microsoft與NVIDIA在OCP內針對800伏特直流（DC）供電（即800 VDC）的共同工作表示，「the goal is not a single prescriptive design」[S-019]；而Ethernet for Scale-Up Networking（ESUN）倡議至其1.0版發布時，參與企業已逾175家[S-020]。但沒有任何規格（specification）能夠創造出高頻寬記憶體（high-bandwidth memory, HBM）、渦輪機產能時段或電網併接。OCP的價值在以下情形最高：於各層級之間的交接點降低整合摩擦、驗證（qualification）成本與鎖定（lock-in）（第6章）。

A 稀缺使議價能力（bargaining power）移轉至掌握稀缺環節（scarce step）者。 稀缺投入要素的供應商如今訂立多年期條件：SK hynix與約十家客戶簽有長期協議（long-term agreement, LTA）[S-003]，Infineon公布多年期產能預留協議並收取部分預付款[S-021]，GE Vernova的Gas Power設備未出貨訂單（backlog）與產能時段預訂於2026年6月底達116 GW[S-012]。機櫃組裝廠公布的全公司利潤率則更為微薄：鴻海（Hon Hai）第2季合併毛利率為6.12%[S-022]。各項數字均為該公司自身的衡量方式，並非獲利能力排名（第5.1節；A-12）。

IF 下一個問題是稀缺將移往何處。 若記憶體隨新產能開出而鬆動（SK hynix已將其龍仁第1期無塵室排定於2027年初，印第安納州HBM量產排定於2029年下半年[S-003; S-023]），具約束力的環節可能移向光學輸入／輸出、封裝良率（yield）、直流故障保護、電網併網或資金成本。此一觀點僅在需求持續上升的速度快於任何單一層級增添產能的速度時成立。應觀察記憶體合約價格、雷射供應、保護裝置認證（certification）與併網規則；若需求放緩至使所有層級同時鬆動的程度，此一觀點即不成立（第1.5節；第9章）。

A 各層級依不同時鐘擴張。 晶片與伺服器供應以季為單位變動：NVIDIA的Vera Rubin平台於2026年8月「ramping into full production」[S-010]。新的記憶體晶圓廠與新一代封裝則需以年計，SK hynix印第安納州HBM的2029年時點，以及台積電（TSMC）14倍光罩尺寸CoWoS的2028年量產時點，均可見其然[S-023; S-024]；渦輪機產出與電網併接的時程則更長[S-012; S-018]。圖2將各時鐘並列對照。商業機會與延誤風險即落在彼此之間的落差：依快時鐘下訂的產能，必須等待依較慢時鐘到位的投入要素（A-11）。

1.2 四項結構性轉變

四項轉變描述此一系統正如何改變：運算邁向系統、48 V邁向800 VDC、氣冷邁向液冷（liquid cooling），以及圖形處理器（graphics processing unit, GPU）稀缺邁向記憶體、電力與電網容量的稀缺。每一項均為壓力的方向，而非完全取代。就每一項轉變，以一段說明紀錄，另一段給出研判（reading）。

運算 → 系統

F NVIDIA的平台如今以機櫃形式交付客戶。 NVIDIA表示其Vera Rubin平台正「ramping into full production」，已在CoreWeave、Google Cloud、Microsoft Azure、Oracle Cloud Infrastructure與Nebius等合作夥伴處運行機櫃[S-010]。鴻海（Hon Hai／Foxconn）預期其AI機櫃出貨量於2026年將成長逾一倍[S-022]。台積電（TSMC）14倍光罩尺寸的CoWoS封裝排定於2028年量產，可整合約十顆大型運算晶粒與20組HBM堆疊[S-024]。

A 設計、採購（procurement）與驗證（qualification）的單元正從晶片移向機櫃。 效能如今取決於運算、記憶體、互連、電力與冷卻如何整合並一同測試[S-010; S-024]，因此系統中任一處的後期變動都代價更高。價值轉向那些能以良率（yield）交付可運作機櫃者：先進封裝（advanced packaging）、機櫃整合與測試。晶片依然重要，但愈來愈是在系統之內被購入（第4.1節；A-01）。

48 V → 800 VDC

F 在高階領域，配電正延伸至48 V匯流排（busbar）之外。 Mt Diablo規格Diablo 400（0.7.0版，2026年3月1日）自獨立的電源機櫃（power rack）配送±400 VDC[S-027]。NVIDIA於2025年5月表示，800 VDC資料中心的全面量產將與其2027年的Kyber機櫃級系統（rack-scale system）同步[S-028]；TrendForce預期NVIDIA的800 V電源機櫃將先以Vera Rubin的選配組態登場，並於2028年更廣泛部署（為一項預測）[S-029]。OCP的800 VDC工作將其定位為交流（AC）供電的補充，而非取代[S-019]。

A 更高的電壓是一項架構選擇，並受保護與認證（certification）所節制。 在功率固定的情形下，電壓愈高則電流愈低；NVIDIA指出，將機房配電自415 VAC改為800 VDC可省銅45%[S-028]。困難的問題轉移至直流故障切斷、保護協調與已認證硬體，而OCP的相關工作正就安全認證與UL Solutions、美國國家防火協會（National Fire Protection Association, NFPA）、IEEE及國際電工委員會（International Electrotechnical Commission, IEC）合作[S-019]。800 VDC要形成廣泛的多供應商量產常規，仍屬有條件（第4.2節；A-03）。

氣冷 → 液冷

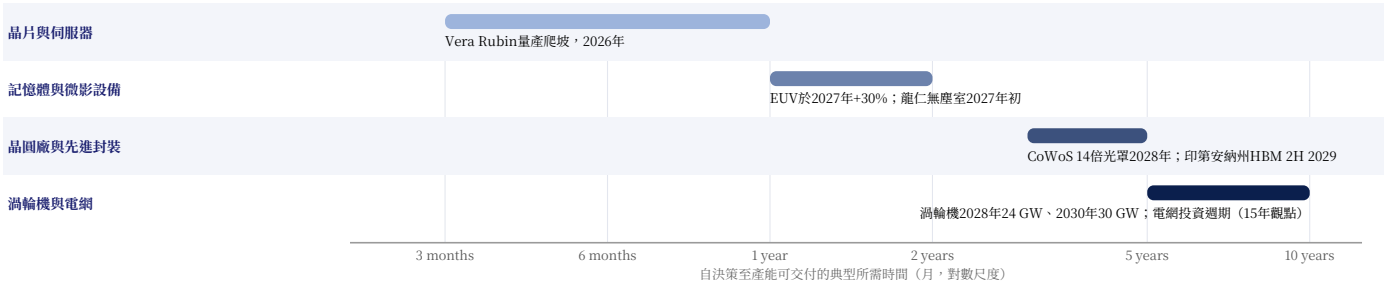
F 液冷（liquid cooling）正成為高階AI系統的標準配置。 TrendForce預估AI晶片的液冷滲透率將自2025年約33%升至2026年的53%，並於2027年趨近60%（均為預測）[S-006]；該機構描述Vera Rubin的液冷已下達中央處理器（central processing unit, CPU）、網路卡、電源板與光模組。OCP的Project Deschutes規格（specification）訂定一款適用於高功率、高密度IT pod的2 MW冷卻液分配單元（cooling distribution unit, CDU），自2026年2月26日生效[S-030]。

A 冷卻的問題已從熱傳遞轉向營運。 一旦液體進入機櫃，洩漏管理、冷卻液品質、試運轉驗收（commissioning）與維修即成為系統可靠性的一部分，而冷板（cold plate）、快速接頭（quick disconnect）與CDU亦成為交付的投入要素：TrendForce指出Vera Rubin的散熱片目前僅有單一來源供應商[S-006]。氣冷並未過時；它仍是機房層級的散熱路徑，並服務功率較低的設備。壓力的方向是在高階領域轉向液冷，而非全面轉向（第4.3節）。

GPU稀缺 → 記憶體、電力與電網稀缺

F 2026年各項揭露所指名的稀缺投入要素為記憶體、封裝、電力設備與電網併接。 SK hynix描述記憶體市場的情形為客戶需求超過供應能力[S-003]。Advantest將「advanced packaging capacity constraints」列為該產業的外部不確定因素之一[S-031]。Alphabet以「the supply constraint environment」為由，計畫於第3季擴大運用第三方產能作為過渡策略[S-016]。Siemens Energy描述燃氣渦輪機市場為「supply-constrained rather than demand-constrained」[S-026]；德州電網業者ERCOT於2026年6月追蹤的大型負載（large load）

圖2。 AI基礎設施的各個層級依不同時鐘增添產能



自投資或下單決策至產能可交付的典型所需時間，依層級列示，採以月為單位的對數尺度。各時鐘的時長為本報告的分析。企業實例均為註明日期的陳述，各為計畫、藍圖或產能里程碑，仍有待執行：NVIDIA的Vera Rubin平台「ramping into full production」（2026年8月）[S-010]；ASML計畫於2027年將其低數值孔徑（low-NA）極紫外光（extreme-ultraviolet, EUV）微影產能增加30%[S-025]；SK hynix龍仁第1期無塵室於2027年初、印第安納州HBM量產於2029年下半年[S-003; S-023]；台積電（TSMC）14倍光罩尺寸CoWoS於2028年量產[S-024]；GE Vernova燃氣渦輪機產出於2028年達24 GW、2030年達30 GW[S-012]；Siemens Energy燃氣渦輪機產能已售罄至2028財年，並就電網投資週期提出「over the next fifteen years」的看法[S-018; S-026]。
長條呈現的是指示性區間，而非專案時程；個別專案的進度可能快於或慢於其所屬層級的典型區間。

併接申請逾438,000 MW，其中近89%來自資料中心[S-032]。

A 稀缺是沿著鏈條移轉，而非消失。 某一層級的紓解會將壓力傳遞至下一層級：加速器（accelerator）增加則需要更多HBM，機櫃增加則需要更多電力與冷卻，電力增加則需要渦輪機、變壓器與併網。這些層級中的每一個，增添產能的時鐘都比前一個更慢[S-023; S-012]，因此本報告將各時鐘之間的落差，而非任何單一短缺，研判為延誤的主要來源（圖2；A-11）。

1.3 十一項發現

以下十一項發現構成本報告的核心模型。每一項均簡要陳述證據、給出研判（reading），並指向深入闡述該項的章節；詮釋登錄表（附錄D）則列出各項研判的推論邏輯、替代解釋與適用邊界。第1至10項發現對應該表中的研判A-01至A-10，第11項為A-13；A-11與A-12出自第1.1節，A-14出自第5.3節。圖3逐層顯示註明日期的來源在何處記錄到約束，期間自2024年至本十年末。

A 1. AI基礎設施的限制正從晶片擴及系統。 NVIDIA的供應承諾在一個季度內增加逾一倍，達2,790億美元，主要用於記憶體[S-002]。TrendForce報導Lumentum執行長表示，用於光鏈路雷射的磷化銦（indium phosphide, InP），其供需缺口如今已超過動態隨機存取記憶體（dynamic random-access memory, DRAM）與NAND快閃記憶體[S-037]。Siemens Energy的燃氣渦輪機產能已售罄至2028財年[S-018]。研判（reading）：交付取決於數項投入要素中最慢的一項，而非僅取決於加速器（accelerator），因此產能規劃如今同時繫於記憶體、光學、電力設備與電網併接（第5.6節；A-01）。

A 2. 記憶體正同時成為預算上的限制與效能上的限制。 TrendForce估計，DRAM與NAND快閃記憶體將占主要雲端服務供應商（CSP）2026年資本支出（capex）的47%，2027年則占68%（均為預測）[S-007]。Microsoft於2026年4月對2026日曆年資本支出的預期，包含約250億美元來自零組件價格上漲[S-045]；Meta則將其2026年基礎設施資本支出預測增幅的大部分，歸因於「higher component costs, particularly memory pricing」[S-046]。研判（reading）：記憶體如今不僅影響系統效能，更牽動資本預算與合約條件，且其價格變動會在相隔一個季度之間即傳導至買方的資本支出財測指引（第3.2節；A-02）。

A 3. 電力與冷卻如今屬於架構，而非機房層面的事後補充。 國際能源總署（IEA）報告指出，AI伺服器功率密度在2020年至2025年間上升十一倍[S-005]。TrendForce預估2026年將有53%的AI晶片採用液冷（liquid cooling）[S-006]；Google、Microsoft與NVIDIA正在OCP內訂定共通的800 VDC需求，且已有逾80家夥伴開發相容的基礎設施[S-019]。研判（reading）：機櫃供電、電力轉換與散熱如今在設計階段即與晶片一併定案，其供應商因而進入平台的驗證（qualification）流程，而非建築的裝修工程（第4章；A-03）。

A 4. 即使運算能力已可取得，電網併接與電力設備仍可能左右部署時程。 美國區域電網業者PJM已請求聯邦能源管理委員會（FERC）接受一套框架：在緊急情況下，未自帶新增容量的新建資料中心及其他大型負載（large load），其用電需求將較其他用戶優先被削減或移轉[S-042]。CoreWeave於2026年6月底的已啟用電力為1.5 GW，而已簽約電力約3.7 GW[S-011]；GE Vernova的Gas Power設備未出貨訂單（backlog）與產能時段預訂則達116 GW[S-012]。研判（reading）：決定產能何時開始創造營收者，愈來愈是送電日期而非晶片交付，且已簽約電力並不等於已送電電力（第7.2節；第5.5節；A-04）。

A 5. 客製晶片正與GPU並行成長，尚未取代GPU。 TrendForce預測，以特殊應用積體電路（application-specific integrated circuit, ASIC）為基礎的伺服器將占2026年AI伺服器出貨量的27.8%，為2023年以來最高，以GPU為基礎的伺服器則占69.7%[S-008]；至2030年ASIC占比將趨近40%[S-009]。Alphabet自2026年第2季起向客戶的資料中心交付張量處理器（tensor processing unit, TPU）系統[S-016]，同期NVIDIA的Data Center營收年增117%[S-002]。研判（reading）：兩條路徑正一同成長，而客製晶片在改變GPU的地位之前，先改變了記憶體、封裝與驗證（qualification）需求的組成（第3.1節；A-05）。

A 6. 開放介面在可藉多供應商驗證降低轉換成本的交接點上最為重要。 依OCP的智慧財產權管理政策（Intellectual Property Rights Management Policy, IPR），貢獻者除於30日期間內提出排除外，均視為已就實施最終規格（Final Specification）所需的專利請求項給予免權利金（royalty-free）授權（licence）[S-047]。平台擁有者仍保有自身的驗證關卡，例如NVIDIA針對電力與冷卻產品的DSX Ready計畫[S-048]。研判（reading）：在買方能於共通交接點驗證（qualification）第二供應來源之處，開放即有回報，並於該處降低轉換成本（switching cost）；但開放無法增添記憶體、渦輪機產能時段或電網併接（第6.1節；A-06）。

A 7. 資本結構正成為基礎設施風險的一部分。 國際能源總署（IEA）表示，資料中心投資的規模已過大，無法僅以企業自身資產負債表支應，資本市場資金對其建置將至為關鍵[S-005]。Oracle預期於其2027財年（2026年6月至2027年5月）募集約400億美元的債務與股權，其中包含第1季已完成的200億美元隨市增發[S-044; S-049]。NVIDIA已就一座將出租給OpenAI的園區提供信用支持保證，上限為1,050億美元[S-002]。研判（reading）：由誰為誰融資、以何者作為擔保，如今將形塑產能到位的速度，以及需求放緩時由誰承擔風險（第2.2節；A-07）。

A 8. 地緣政治限制已直接進入企業的規劃假設。 NVIDIA對2027財年第3季的展望假設來自中國的Data Center運算營收為零[S-010]；該假設涵蓋的是此一類別，而非NVIDIA的中國營收整體，且同一句話出現於迄今全部三次2027財年展望之中[S-050; S-051]。AMD將對中國的MI308銷售自其2025年下半年展望中排除，於2026年第1季展望中計入約1億美元，最新展望則未附任何中國相關的限定說明[S-485; S-502; S-486; S-111]。2025年10月，KLA就失去中國部分客戶對營收的影響，估計2025年12月當季與2026日曆年各約為3億至3.5億美元[S-487]。研判（reading）：美國與中國的限制可以明確假設的形式進入財測指引，惟數家企業的最新展望已不再載明（第7.1節；A-08）。

A 9. 台灣的角色正從「組裝」這一簡略代稱，轉向系統整合、先進封裝（advanced packaging）、光學與驗證。 鴻海（Hon Hai）表示Vera Rubin機櫃於2026年第3季進入量產，其共同封裝光學（co-packaged optics, CPO）全光交換器亦有望於同一季量產[S-022]。台積電（TSMC）已排定自2026年起在其COUPE平台上量產CPO，並於2028年量產14倍光罩尺寸CoWoS[S-024]。友達（AUO）正與夥伴共同開發Micro LED CPO模組與玻璃核心基板，其定位處於研發（R&D）階段[S-052]。研判（reading）：台灣的分量愈來愈落在封裝、機櫃整合與測試（第5.7節；A-09）。

A 10. 下一波研發與智慧財產權機會，圍繞在尚未解決的實作、材料、控制、測試與服務邊界。 2026年8月一篇*Nature Electronics*的回顧文章，將「thermal management, manufacturability and standardization」列為共同封裝光學（co-packaged optics, CPO）的障礙[S-053]。OCP的固態變壓器（solid-state transformer）規格（0.3.0修訂版）訂有效能要求，例如在使用率50%至100%之間、僅計入功率鏈路損耗時效率至少達98%[S-054]；SolarEdge與Infineon則正在開發適用於800 VDC的固態斷路器（solid-state circuit breaker）[S-055]。研判（reading）：隨著介面趨於共通，差異化轉向這些介面如何建置、控制、測試與維修（表1；第6.7節；A-10）。

A 11. 更精簡的模型使記憶體需求在各層級之間移動，而非減少需求。 DeepSeek表示其V4.1-Flash模型的快取僅需前一代HBM用量的四分之一[S-488]；依本報告自開發者公布組態所作的16位元計算，一段一百萬token的對話在Kimi K3上約需29 GB快取，而在2025年設計的Qwen3-235B-A22B上則接近200 GB（表17）[S-489; S-490]。在截至2026年5月的一年間，Google每月的token量約成長七倍[S-491]；記憶體製造商並將代理式AI列為需求來源之一[S-003; S-129]。研判（reading）：每token的記憶體用量正在下降，但記憶體需求總量並未下降；它正轉向頻寬、超大型模型的常駐，以及成本較低的快取層級（第8章；A-13）。

圖3。 具約束力的限制已從晶片與封裝擴散至記憶體、雷射、渦輪機與電網

	2024	2025	2026	2027–30	來源
加速器	B 受CoWoS限制	B 受CoWoS限制	B CoWoS；HBM4驗證		S-017; S-033; S-034
HBM/DRAM	—	E 2H25價格上漲	B 需求 > 供給	新晶圓廠2027–29年	S-007; S-003; S-035; S-023
CoWoS/基板	B 自2023年起短缺	B 短缺	B 缺口縮小	14倍光罩2028年	S-017; S-033; S-036; S-024
光學/雷射	—	—	B InP雷射缺口		S-037
電力電子	—	—	E 交期		S-038; S-021
冷卻	—	—	單一來源零件		S-006
燃氣渦輪機	—	E 2025年底前接單80 GW（預期）	B 已售罄至FY2028	B 已售至FY2028；FY2029正在填滿	S-039; S-012; S-018; S-026
電網併接	—	E 大型負載規則	B 排隊佇列、負載削減	2030年缺口預測	S-040; S-041; S-032; S-042; S-043
資本/融資	—	—	E 需要資本市場資金		S-005; S-044; S-002

— 所引來源未記錄約束　E 約束正在形成　B 重大/具約束力　展望不確定

依期間列示各層級供應約束的狀態，自2024年至2027–30年，以註明日期的公開來源所記錄者為準。唯有在有註明日期的來源支持時，該格才標示為正在形成（E）或具約束力（B），來源欄則列出該列的來源編號。前瞻性的格位係以企業計畫、預留與第三方預測為依據，除企業明確表示產能已售出者外，均標示為不確定。此圖記錄的是狀態，並不對各層級評分或排名。
破折號表示本報告所引註明日期的來源，均未記錄該期間存在重大約束；並不表示該約束不存在。簡短標籤說明各狀態所記錄的依據。

1.4 約束在何處形成研究課題

A **本報告中的每一項約束，同時也是一道尚未解決的工程難題。** 每一項短缺背後都有一個技術問題：如何在加速器（accelerator）旁容納更多記憶體、如何讓封裝內的雷射保持低溫、如何安全切斷直流故障、如何併接一個能夠故障穿越（ride-through）電網擾動（grid disturbance）的大型負載（large load）。當某一介面趨於共通，其文本即不再是差異的來源，差異化轉而落在它如何建置、控制、測試與維修[S-047; S-054]。表1將每一項約束與各來源留待解答的研究問題，以及實作仍保有差異化的區域相互對應；第6.7節與第6.6節則對兩者加以闡述。

1.5 三項風險與三項訊號

三項風險可能使建置放緩或使其成本移轉，三項訊號則可顯示約束正往哪個方向移動。各項風險均為對所引事實的研判（reading），而非預測；各項訊號則可於未來四個季度的企業財報、規格（specification）發布與法規申報文件中觀察到。

三項風險

IF **記憶體成本衝擊可能擠壓預算與系統整合商。** TrendForce預期伺服器DRAM合約價格在2025年下半年累計上漲64%之後，於2026年將再上漲約270%[S-007]。緯穎（Wiwynn）表示其一直在與客戶調整商業模式，「in response to changes in the AI product mix and rising component costs」；自2026年4月起，特定客戶的記憶體採購（procurement）改採採購代理模式，相關金額不再計入營收或銷貨成本[S-063; S-064]。其經核閱的期中財報並未指名記憶體，而是將該公司視為代理人，處理來自客戶指定供應商、且其並未控制的物料[S-483]。此一風險在記憶體價格漲幅超過預算增幅期間持續成立；應觀察合約價格的季變動與多供應商HBM驗證（qualification）。若價格連續數季下跌，此一風險即趨於緩和（第5.2節）。

IF **電力與電網的延誤可能為可開機的運算能力設下上限。** PJM將其2024年至2030年預測的32 GW需求成長中的30 GW歸因於資料中心[S-042]。2026年3月，Siemens Energy表示其預期2029財年的燃氣渦輪機產能將於本年度結束前大致售罄，2030財年可供產能則有限[S-018]。TrendForce預測2030年全球資料中心電力需求與資料中心可取得的電網容量之間，將出現268 GW的缺口[S-043]。此一風險在電網與設備交期（lead time）長於伺服器週期期間持續成立；應觀察大型負載（large load）規則、渦輪機未出貨訂單（backlog），以及已簽約電力與已啟用電力的對比。若彈性併接條件與場內自發電擴散，此一風險即趨於緩和（第7.2節）。

IF **融資與交易對手集中度可能使需求放緩沿鏈條傳遞。** CoreWeave 2025年營收中約67%來自Microsoft[S-065]。Nebius首筆擔保債務係以GPU基礎設施及來自一家投資等級客戶的已簽約現金流作為擔保[S-066]；Oracle預期於2027財年新增約400億美元的債務與股權[S-044]。此一建置在已簽約現金流與資本市場持續為其提供資金期間持續成立。應觀察再融資條件、預付款與保證，以及最大買方承諾的任何縮減；若放款人對以合約為擔保的融資額度重新定價，此一風險即加劇（第2.2節；A-07）。

> “Data centre investments have grown too large to be funded from company balance sheets alone, and large amounts of funding from capital markets will be critical for their buildout.” [S-005]

三項訊號

- **記憶體。** 每季的伺服器DRAM與HBM合約價格；HBM4與HBM4E在一家以上供應商完成驗證（qualification），其中Samsung預期客製HBM樣品將於2027年送達客戶[S-056]；SK hynix的M15X產能爬坡及其預定於2027年初完工的龍仁無塵室[S-003]；新發布開放權重模型（open-weight model）的每token快取記憶體用量（表17）。
- **800 VDC與冷卻。** 來自無關聯供應商、超越參考設計（reference design）的量產電力硬體；已公布的直流保護安全認證（certification）；OCP認可的實作；原生800 VDC之Mt Diablo後繼規格的註明日期草案[S-019]。首次檢視時點：OCP全球峰會（OCP Global Summit），2026年10月12日至15日[S-067]。
- **電網與電力設備。** 大型負載（large load）規則，包括PJM提出的大型負載斷接可靠度標準草案，以及聯邦能源管理委員會（Federal Energy Regulatory Commission, FERC）於2026年6月就大型負載併接所發布的命令[S-061; S-068]；渦輪機未出貨訂單（backlog），其中GE Vernova預期至2026年底，燃氣相關未出貨訂單與產能時段預訂將至少達125 GW[S-012]。

讀者視角

本章對各類讀者的意義

財務。 資本支出（capex）、未出貨訂單（backlog）與利潤率均依各公司自身的定義；記憶體價格與融資條件如今與加速器（accelerator）價格並列，一同牽動預算。

產業。 記憶體、雷射、渦輪機與電網併接各依自身時鐘運行；決定交付日期的是預留與驗證（qualification）名單，而非產品目錄。

政府。 併網規則、直流保護的認證（certification）與出口核准，對部署時程的決定力更甚於任何開放規格（specification）。

研發。 尚未解決的量產問題集中於直流故障保護、CPO的熱與良率（yield）、冷卻液營運，以及HBM之外的記憶體分層。

智慧財產權。 介面正趨於共通；差異化轉向實作、控制、材料、測試與服務，惟仍須經請求項層級的審查。

其餘各章的閱讀方式

標記與編號。 第2頁說明五項標記及圖表的閱讀方式；A-01至A-14為詮釋登錄表中的十四項主要研判（reading）。

來源。 方括號中的引註連結至來源查核登錄表（附錄D），其中載明各來源的原始網址、逐字摘錄，以及該來源所支持與無法據以確立的內容。其詮釋登錄表列出十四項主要研判（reading），即A-01至A-14。

閱讀路徑。 資本與市場：第2章、第5章。技術：第3章、第4章。標準與智慧財產權：第6章。政策與能源：第7章。模型與記憶體：第8章。展望：第9章。完整路徑見第二頁。

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-002 NVIDIA (26 Aug 2026); S-003 SK hynix (29 Jul 2026); S-010 NVIDIA (26 Aug 2026); S-011 CoreWeave (11 Aug 2026); S-012 GE Vernova (22 Jul 2026); S-014 Alphabet (4 Feb 2026); S-015 Alphabet (29 Apr 2026); S-016 Alphabet (22 Jul 2026); S-018 Siemens Energy (30 Mar 2026); S-021 Infineon Technologies (5 Aug 2026); S-022 Hon Hai (12 Aug 2026); S-023 SK hynix (28 Aug 2026); S-024 TSMC (23 Apr 2026); S-025 ASML (15 Jul 2026); S-026 Siemens Energy (29 Jun 2026); S-028 NVIDIA Technical Blog (20 May 2025); S-031 Advantest (29 Jul 2026); S-039 GE Vernova (9 Dec 2025); S-044 Oracle (10 Jun 2026); S-045 Microsoft (29 Apr 2026); S-046 Meta Platforms (29 Apr 2026); S-048 NVIDIA (21 Sep 2026); S-049 Oracle (10 Sep 2026); S-050 NVIDIA (25 Feb 2026); S-051 NVIDIA (20 May 2026); S-052 AUO (31 Aug 2026); S-055 SolarEdge and Infineon (9 Sep 2026); S-056 Samsung Electronics (12 Feb 2026); S-057 Besi (18 Jun 2026); S-063 Wiwynn (7 Aug 2026); S-064 Wiwynn (7 May 2026); S-065 CoreWeave (2 Mar 2026); S-066 Nebius (17 Jul 2026); S-111 AMD Investor Relations (4 Aug 2026); S-129 Micron Technology (24 Jun 2026); S-483 Wiwynn (7 Aug 2026); S-485 Advanced Micro Devices (5 Aug 2025); S-486 Advanced Micro Devices (3 Feb 2026); S-487 KLA (29 Oct 2025); S-488 DeepSeek (10 Sep 2026); S-489 Moonshot AI

表1。 每一項約束都留下一道實作難題，研究與差異化仍可在其中移動

未解的問題在於介面如何建置、控制、測試與維修，而不在介面文本本身。

約束	近期研究或創新問題	智慧財產權或差異化區域	來源
HBM/封裝	如何提高每顆加速器（accelerator）的可用記憶體與封裝良率（yield）：更高的堆疊、客製基礎晶粒、混合鍵合（hybrid bonding），以及HBM之外的記憶體分層	基礎晶粒邏輯與晶粒間供電；堆疊、鍵合與堆疊測試；封裝熱協同設計（通道介面由JEDEC確立）	[S-056; S-057; S-024; S-058]
光學/CPO	一旦光學元件置入封裝之內，如何控制雷射熱、良率（yield）與現場更換，以及如何測試來自多家供應商的光引擎	雷射與光引擎熱設計、光纖接合、光引擎測試與更換方法；Micro LED短距鏈路（研發階段）	[S-053; S-037; S-024; S-052]
800 VDC/電力	如何偵測、切斷並協調直流故障，以及如何在百萬瓦等級將中壓轉換為800 VDC	斷路器控制與故障協調邏輯；變壓器拓撲、磁性元件與冷卻，其中規格（specification）訂定的是效能要求；測試方法	[S-019; S-054; S-059; S-055]
液冷	如何驗證（qualification）冷卻液與兩相流體、如何偵測洩漏，以及如何在機隊規模下對液冷機櫃進行試運轉驗收（commissioning）與維修	冷板（cold plate）與蒸發器幾何、流體配方、快速接頭（quick disconnect）密封、洩漏偵測與流量控制	[S-030; S-060; S-006]
電網/設施	如何以功率平滑、儲能與負載削減（curtailment）控制，使AI負載具備彈性並能故障穿越（ride-through）電網擾動（grid disturbance）	機櫃層級儲能與功率爬升的控制、故障穿越（ride-through）與負載削減邏輯、電網遙測（telemetry）；規則由電網業者訂定	[S-042; S-061; S-062]

就每一項約束，列出所引來源留待解答的近期（near term）研究或創新問題，以及實作仍保有差異化的區域；適用於研發、創新與智慧財產權讀者。最後一欄所列來源即各項研判（reading）的依據。

範圍 智慧財產權欄標示實作仍保有差異化、並可能含有可受保護之發明的區域，惟仍須經請求項層級的新穎性與自由實施（freedom to operate）審查。並非每一項未解問題都是專利機會。

本章來源 （續）

(Jul 2026); S-490 Alibaba Qwen team (21 Jul 2025); S-491 Google (19 May 2026); S-502 Advanced Micro Devices (4 Nov 2025)

標準組織與聯盟. S-001 OCP (27 Aug 2026); S-019 OCP (11 Aug 2026); S-020 OCP (10 Mar 2026); S-027 OCP (1 Mar 2026); S-030 OCP (26 Feb 2026); S-047 OCP (14 Aug 2026); S-054 OCP (22 Jun 2026); S-058 JEDEC Solid State Technology Association (Dec 2025); S-059 OCP (2 Sep 2026); S-060 OCP (6 Aug 2026); S-062 OCP (29 Apr 2026); S-067 OCP (未註明日期)

政府與監管機關. S-005 IEA (16 Apr 2026); S-032 ERCOT (18 Jun 2026); S-040 Texas Legislature (2025); S-041 Federal Energy Regulatory Commission (18 Dec 2025); S-042 PJM Inside Lines (13 Aug 2026); S-061 PJM Inside Lines (10 Sep 2026); S-068 Federal Energy Regulatory Commission (18 Jun 2026)

研究機構. S-053 Nature Electronics (19 Aug 2026)

市場研究. S-006 TrendForce (17 Aug 2026); S-007 TrendForce (25 Aug 2026); S-008 TrendForce (20 Jan 2026); S-009 TrendForce (18 Mar 2026); S-013 TrendForce (3 Aug 2026); S-017 TrendForce (30 Apr 2026); S-029 TrendForce (25 Jun 2026); S-034 TrendForce (8 Apr 2026); S-035 TrendForce (1 Jun 2026); S-036 TrendForce (18 Sep 2026); S-038 TrendForce (15 Apr 2026); S-043 TrendForce (16 Sep 2026)

媒體與法律評論. S-033 TrendForce (15 Jun 2026); S-037 TrendForce (6 Aug 2026)

第2章

需求、資本與市場規模

最大的AI買方持續上調其2026年資本計畫，未出貨訂單（backlog）與已簽約電力均超前於已交付的產能，融資亦進入整個堆疊之中；各項指標各有其定義，因此各數字僅並列呈現，絕不相加。

適用讀者： 財務與投資、採購（procurement）、政府與政策；OCP認可市場見第2.4節，電力見第2.5節。

人工智慧（AI）基礎設施的每一個層級，自記憶體晶圓廠至燃氣渦輪機，都依同一項訊號規劃其產能：最大買方打算花多少錢。2026年，此一訊號持續上升。Alphabet、Amazon、Meta與CoreWeave在年內上調其資本計畫，雲端未出貨訂單（backlog）在兩個季度內約增加一倍，NVIDIA的Data Center營收在十個季度內增至接近四倍。然而這些主要數字如今變動的原因已不只是數量：還包括記憶體價格、租賃會計與外部融資。本章區分每一個數字所衡量的內容與其容易被誤解的意涵，並要問：需求之中有多少已簽約、已取得資金且可交付？

2.1 資本支出持續上升，但會計基礎各異

最大的雲端與AI企業的資本支出（capital expenditure），是堆疊中其餘環節據以下單的需求訊號：記憶體製造商、晶圓代工廠、機櫃製造商、冷卻設備供應商與渦輪機製造商，都圍繞它規劃產能。2026年，此一訊號同時變得更強，也更難解讀。金額再度上升，且多數在年內經過修正，但各公司計算資本支出的方式各不相同，而增幅的一部分是為同一批設備支付更高的價格。將這些數字相加，或在公司之間進行排名，都是在比較性質不同的事物。本節依各買方自身的基礎，列出其所述內容。

F Alphabet已兩度上調其2026年區間，並預期2027年將顯著更高。 Alphabet對2026年資本支出（capital expenditure）的財測指引為：2月1,750億至1,850億美元、4月1,800億至1,900億美元、7月1,950億至2,050億美元，相較之下2025年實際支出為914億美元[S-014; S-015; S-016]。4月的上調，是將與3月完成的Intersect併購相關的投資納入該區間[S-015]。7月的上調，以該公司的措辭而言，係「primarily due to an acceleration in the delivery of capacity to meet growing demand」[S-016]。就2027年，Alphabet仍預期資本支出將「increase significantly」，細節將於日後公布[S-016]。

F Amazon與Meta上調計畫的主因，是零組件、尤以記憶體為首的成本上升。 Amazon於2月公布財報時，預期2026年全公司資本支出（capital expenditure）約2,000億美元；在7月的法說會上，其將2026年現金資本支出定於約2,200億美元，並指出「the higher cost of memory」推升了該數字[S-069; S-070]。Meta的數字含融資租賃本金償付，其於4月將2026年區間自1,150億至1,350億美元上調至1,250億至1,450億美元[S-071; S-072]，主因為「higher component costs, particularly memory pricing」[S-046]，並於7月收窄至1,300億至1,450億美元[S-073]。Meta尚未就2027年資本支出提供具體展望[S-074]。

F Microsoft 7月較低的數字是會計基礎的變動，而非縮減。 4月時，Microsoft預期於2026日曆年投資約1,900億美元，其中約250億美元來自零組件價格上漲[S-045]。自其2027財年（2026年7月至2027年6月）起，該公司將資料中心與辦公大樓的估計使用年限自15年延長至25年；因此，未來更多的資料中心租賃將屬不計入資本支出（capital expenditure）的營業租賃，而非計入的融資租賃[S-075]。2026日曆年的預期因而成為約1,750億美元，其餘投資預期則維持不變；Microsoft並預期2027財年資本支出將「grow year-over-year」[S-075]。

> "Outside of this useful life impact, our calendar year 2026 CapEx investment expectations remain unchanged." [S-075]

F 建置產能以供出租的Oracle與CoreWeave，正依其合約規模調整支出。 財年於5月31日結束的Oracle，於2026財年支出557億美元，並預期2027財年資本支出為900億至950億美元，扣除客戶預付款後的淨現金資本支出則不超過700億美元[S-044; S-076]。專業AI雲端服務供應商CoreWeave將其2026年區間自2月的300億至350億美元，以零組件價格為由於5月上調至310億至350億美元，又於8月以其預期於本年度交付的產能與近期取得的合約為由上調至350億至390億美元；該公司2025年支出為149億美元[S-077; S-078; S-079]。

F 台積電（TSMC）與其客戶同步上調資本預算。 台積電於1月預期2026年資本預算為520億至560億美元，4月表示將落在該區間的「towards the high end」，7月則提高至600億至640億美元，理由為「continued strong structural demand」，其中包含「the newly emerging agentic AI market」[S-082; S-083; S-084]。該公司現表示未來三年的資本支出（capital expenditure）將較過去三年「even more significantly higher」，而1月時其所述為「significantly higher」[S-084; S-082]。第3.4節解讀該筆預算在封裝與先進製程節點上買到了什麼。

= 全年財測指引隱含下半年支出快於上半年。 公式：（全年財測指引 − 上半年支出）÷ 2 = 下半年隱含的每季平均支出。Alphabet：（1,950億至2,050億美元 − 806億美元）÷ 2 ≈ 每季570億至620億美元，相較之下第2季為449億美元[S-016; S-080]。Amazon：（約2,200億美元 − 963億美元，即432億美元與531億美元之和）÷ 2 ≈ 每季620億美元，相較之下第2季為531億美元[S-070; S-085]。因此兩份財測指引都要求支出在2026年下半年持續加速，而非在年中的水準上趨於平穩。

F 多數資金購買的是短年限機器，而非建築。 Alphabet表示其2025年技術基礎設施投資中約60%用於伺服器，40%用於資料中心與網路設備；2026年第1季與第2季所公布的比例相同，其財務長並預期2026年全年將「fairly similar」[S-014; S-015; S-016]。Microsoft表示其截至2026年3月當季的資本支出（capital expenditure）中約三分之二投入短年限資產，主要為圖形處理器（graphics processing unit, GPU）與中央處理器（central processing unit, CPU）[S-045]。

A 上升的資本支出項目如今同時反映價格與數量。 六家買方中有四家將其2026年計畫的一部分歸因於零組件或記憶體成本：Amazon、Meta與CoreWeave見於其修正說明，Microsoft則見於其歸因於零組件價格上漲的250億美元[S-070; S-046; S-078; S-045]。相對地，Alphabet 7月的上調是為了加快產能交付[S-016]。本報告的研判（reading）是：2026年增幅的一部分是為同一批伺服器支付更高價格，因此資本支出（capital expenditure）的成長快於其所購得的實體產能。TrendForce估計動態隨機存取記憶體（dynamic random-access memory, DRAM）與NAND快閃記憶體占主要雲端服務供應商2026年資本支出的47%，亦指向同一方向（圖13）[S-007]。公開資料並未將價格效應與數量效應分開。

F 第三方彙總數字是以其自身所選企業集合所作的預測，而非財測指引。 國際能源總署（International Energy Agency, IEA）報告指出，五家大型科技公司的資本支出（capital expenditure）於2025年超過4,000億美元，並於2026年「expected to jump by another 75%」[S-086; S-005]。TrendForce預測九家最大雲端服務供應商（Google、Amazon、Meta、Microsoft、Oracle、ByteDance、Tencent、Alibaba與Baidu）2026年合計資本支出將超過8,867億美元，增幅約90%，並於2027年達到約1.3兆美元[S-013]。各項彙總數字各自採用其自身的企業集合與定義；兩者皆不能取代企業的財測指引，且本報告不將各公司數字相加以進行核對。

F 各項總計數字在2026年間持續上修，而上修本身即是一項訊號。 TrendForce對最大雲端業者2026年資本支出（capital expenditure）的預測，自2月涵蓋八家企業的逾7,100億美元，上調至5月涵蓋九家企業的約8,300億美元，再至8月的逾8,867億美元[S-597; S-598; S-013]。Gartner將其2026年AI基礎設施支出預測——該類別涵蓋AI最佳化雲端容量（IaaS）、伺服器、網路交換架構、處理用半導體與裝置——自1月的約1.37兆美元上調至5月的1.43兆美元，再至9月的1.48兆美元[S-599; S-600; S-601]。第E.2節將各版次本身視為一組數列加以研讀，並標示TrendForce企業樣本的變動。

> **範圍 資本支出**
> Alphabet與Amazon公布的是不動產與設備的現金購置金額，Amazon係扣除處分價款與獎勵後的淨額[S-080; S-069]；Meta另加計融資租賃本金償付[S-071]；Microsoft的主要數字於融資租賃開始時全額計入[S-087]；Oracle的財年於5月結束[S-044]；CoreWeave在法說會公布的數字與其現金流量表列數不同[S-088; S-077]。這些數字絕不加總，亦不排名。

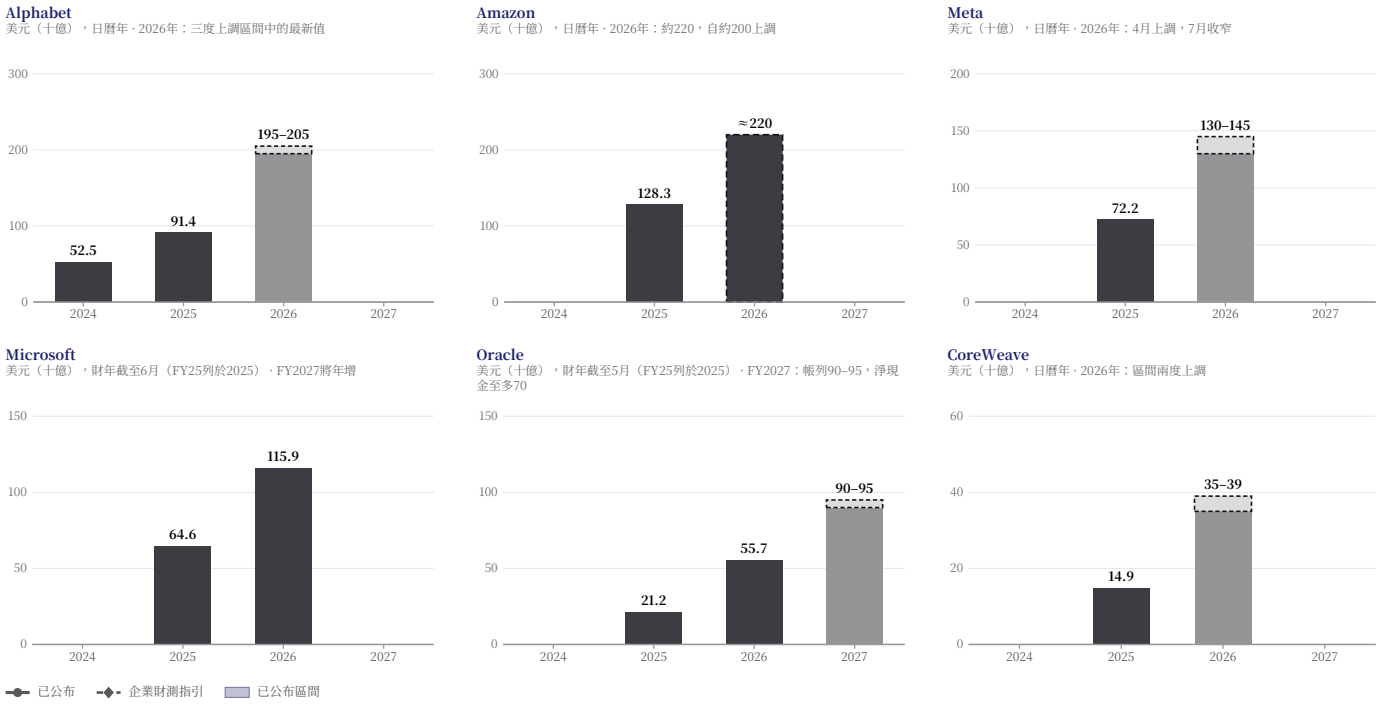
IF 支出要持續上升至2027年，前提是已簽約需求持續以快於產能的速度到來。 Alphabet預期的顯著增加、Microsoft預期的2027財年成長，以及TrendForce的1.3兆美元預測，均取決於此一條件[S-016; S-075; S-013]。應觀察2026年曆年第3季的財報、Alphabet承諾提供的2027年細節，以及Amazon與Meta首度提出的2027年說法。若有大型買方因會計以外的原因下調其計畫，或支出持續而未出貨訂單（backlog）停止成長，則此一研判（reading）不成立。記憶體價格下跌會使資本支出（capital expenditure）下降而需求並未下降，因此在將較低的數字視為需求轉弱之前，須先研讀其所載明的原因。

2.2 未出貨訂單、已簽約電力及其背後的融資

資本計畫說明買方打算支出多少；未出貨訂單（backlog）則說明其客戶已簽下多少。對雲端業者而言，未出貨訂單是未來需求中已簽約的部分、其調高支出時所引用的證據，並日益成為放款人與投資人據以融資的資產。這也是AI基礎設施各種不同時鐘最清楚顯現之處：合約可在一季之內簽訂，而供其履約的已供電容量卻需要數年。本節闡述已簽約需求、其背後的容量，以及其資金來源。

F 雲端未出貨訂單（backlog）在2026年上半年大致倍增。 Google Cloud的未出貨訂單自2025年底的2,400億美元，上升至3月底的4,620億美元與2026年6月底的5,140億美元[S-014; S-015; S-016]。Amazon Web Services（AWS）的未出貨訂單在同樣幾季中自2,440億美元上升至3,640億美元，再至4,960億美元；3月的數字未計入該季之後才宣布的逾1,000億美元Anthropic協議[S-089; S-085; S-070]。Oracle的剩餘履

圖4。 每一家大型AI買方的資本支出於2026年躍升，但各依其自身的會計基礎



依各公司自身定義的資本支出（capital expenditure），單位為十億美元：已公布年度與最新財測指引（有給定區間者列示區間）。Alphabet：不動產與設備購置[S-080; S-016]。Amazon：現金資本支出，已扣除處分收入與獎勵補助；2026年為第2季法說會所給出的預期[S-069; S-070]。Meta：含融資租賃本金償付的資本支出[S-071; S-073]。Microsoft：不動產與設備的現金增添額，財年截至6月[S-081]。Oracle：資本支出，財年截至5月，為FY2027財測指引[S-044; S-076]。CoreWeave：各次法說會所述的資本支出[S-077; S-079]。

範圍 各公司的定義與財年並不相同（現金制或應計制、融資租賃計入或不計入、日曆年或財年），因此各面板彼此不可比，且絕不相加。Microsoft對2026日曆年約1,750億美元（含融資租賃）的預期，係因使用年限延長與租賃會計而自約1,900億美元重設基礎，其基礎與其財年現金數列不同，故未予繪製[S-045; S-075]。

約義務（remaining performance obligations, RPO）於2026年8月31日達6,640億美元，Microsoft的商用RPO於6月30日為6,780億美元，一年內成長84%[S-049; S-075]。

F 各項指標在所計數的對象，以及轉化為營收的速度上並不相同。 CoreWeave的營收未出貨訂單（revenue backlog）係在其剩餘履約義務（remaining performance obligations, RPO）之外，再加計已承諾合約下預期的其他金額；該數字於2026年6月30日約為1,040億美元，高於2025年底的668億美元，且未計入第3季最初數週新增的逾250億美元淨新增承諾[S-011; S-088; S-079]。Oracle預期其剩餘履約義務約有半數將於未來36個月內轉化為營收[S-076]。Amazon表示其2027年容量絕大部分已大致被預訂，且2028年部分容量亦已被預訂[S-070]。

F 各家業者都表示需求超過其供給能力，而各以自身用語表述。 Oracle指出，客戶對其AI雲端訓練與推論服務的需求「continues to grow faster than supply」[S-049]。Microsoft談到，在「as customer demand continues to exceed supply」的情況下須設法銜接缺口[S-075]。Alphabet預期將以「in a supply-constrained way」度過2026年，並計劃於第3季更多運用第三方容量作為銜接策略[S-014; S-016]。Amazon表示，即使2026年支出達約2,200億美元，其容量仍不足以滿足2026年的全部需求，並預期2027年亦將如此[S-070]。各項說法均涉及該公司自身的雲端容量，而非整體市場。

F 各家業者現以吉瓦為單位公布容量，而各自採用不同的衡量方式。 Oracle於2026年9月10日表示，自其財年第4季結束（2026年5月31日）以來，已向客戶交付850百萬瓦（megawatt, MW）的AI容量，內含逾300,000顆GPU[S-049; S-076]。Microsoft在截至2026年6月的兩個季度中各新增約一吉瓦容量，並表示其仍按既定進度，可「roughly double our overall capacity in just two years」[S-045; S-075]。Amazon於2025年新增約4吉瓦（gigawatt, GW）電力，並預期至2027年底其電力容量將為2025年的兩倍[S-093; S-070]。已交付、已新增與已安裝容量屬不同的指標，且均非已簽約電力。

F 在CoreWeave，已簽約電力為已啟用電力的兩倍以上。 CoreWeave的已啟用電力自2025年底的逾850 MW上升至2026年6月30日的1.5 GW，其已簽約電力則自約3.1 GW上升至約3.7 GW，並於8月11日達4.2 GW[S-088; S-011; S-079]。CoreWeave預期至2026年底已啟用電力將逾1.85 GW，並以2030年前達到至少8 GW已啟用電力為目標[S-079; S-078]。2026年2月時，該公司預期2025年底已簽約的容量將幾乎全部於2027年底前上線[S-077]。

= 尚未啟用的已簽約電力，相當於CoreWeave近期新增量的四個季度以上。 計算公式：（已簽約電力 − 已啟用電力）÷ 最近一季新增的已啟用電力。2026年6月30日：（3.7 GW − 1.5 GW）÷ 近0.5 GW ≈ 4.4個季度或更長[S-011]；其後新增簽約的約0.5 GW則使此一倍數更長[S-079]。此一比率描述的是某一日期的在手訂單；它並非交付預測。CoreWeave自身對2026年底已啟用電力逾1.85 GW的財測指引，僅為下半年新增量設定了約0.35 GW的下限[S-079]。

F 部分已承諾的投資完全落在資本支出（capital expenditure）之外。 2026年6月30日，Microsoft已簽訂但尚未開始的租賃達3,291億美元，主要用於資料中心，其中部分以合約條件達成為前提；這些租賃將於2027財年至2033財年間開始，年限為一至20年[S-094]。營業租賃不計入Microsoft的資本支出，且自2027財年起，其未來資料中心租賃將有更多屬營業租賃[S-075]。Alphabet公布同一日期尚未開始的租賃未來付款額為852億美元，主要用於資料中心；Meta則為3,493.1億美元的不可取消合約承諾，多數用於第三方雲端容量、伺服器、網路基礎設施、資料中心與Reality Labs硬體[S-492; S-493]。這些承諾既非資本支出亦非電力容量，不與其中任一項相加，亦不彼此相加。

F Oracle以債務、股權與客戶的資金支應其建置。 Oracle於2026財年募得430億美元債務與50億美元股權，該財年其自由現金流為負237億美元[S-044]。2027財年，該公司預期募得約400億美元債務與股權，其中包括於第1季完成的200億美元市價股權發行[S-044; S-049]。其大型AI合約中預付款與客戶提供硬體的部分，於2026年5月31日合計達750億美元；該公司並表示，第1季入帳的逾300億美元合約未動用Oracle額外資本[S-044; S-076]。

F 供應商、客戶與外部投資人之間的財務連結已超出訂單關係。 NVIDIA於2026年1月投資20億美元於CoreWeave的A類普通股，雙方並擴大合作，以加速CoreWeave於2030年前建置逾5 GW的AI工廠（AI factory）[S-095]。CoreWeave 2025年營收約有67%來自單一客戶Microsoft[S-065]。NVIDIA公布上限為1,050億美元的保證義務，依條件達成分階段生效，與將以20年期租約租予OpenAI、用以承載NVIDIA基礎設施的場址相關[S-002]。2026年7月，Meta宣布與BlackRock合組事業，於德州艾爾帕索開發1 GW的資料中心[S-074]。

> “Data centre investments have grown too large to be funded from company balance sheets alone, and large amounts of funding from capital markets will be critical for their buildout.” [S-005]

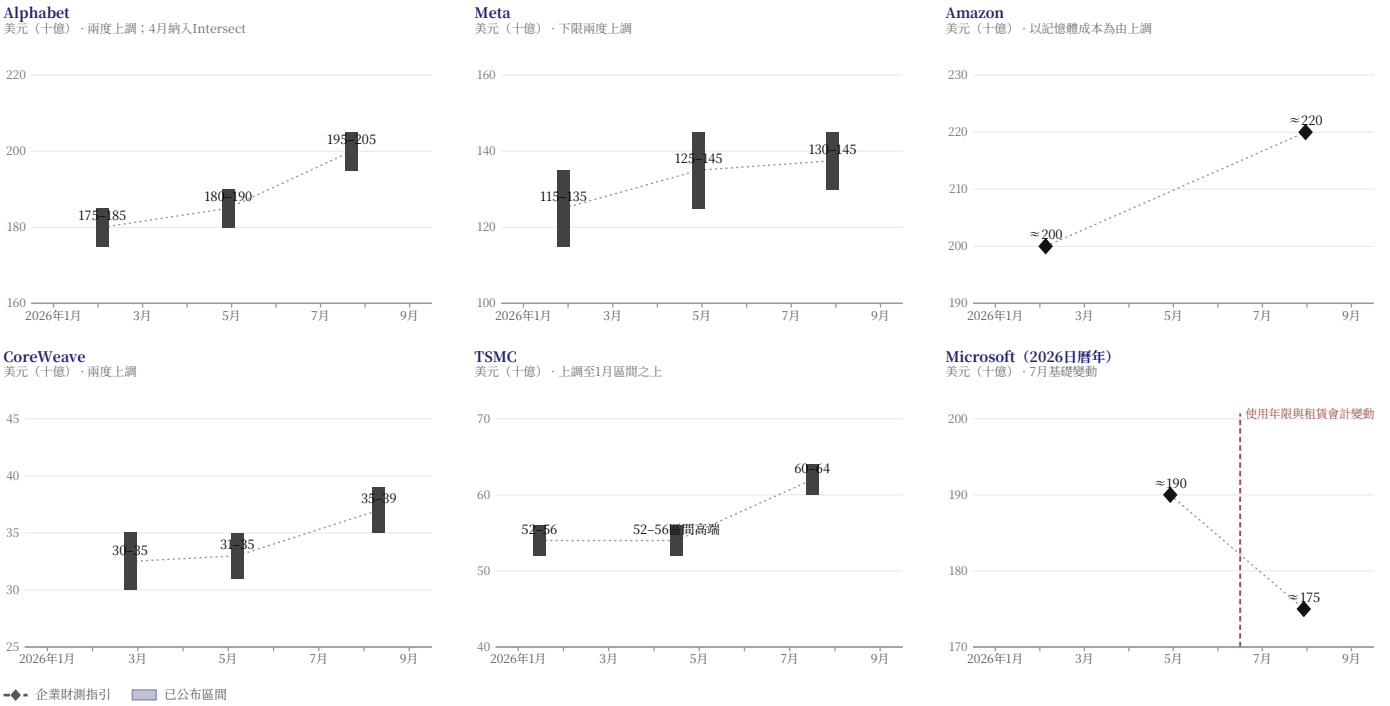
A 資本結構已成為基礎設施風險的一部分。 國際能源總署（IEA）的這項說法可在公司申報文件中得到印證。Amazon在截至2026年6月的十二個月、Oracle在2026財年，自由現金流均轉為負值[S-096; S-044]；債務與股權係以已簽約的未出貨訂單（backlog）為基礎募集；客戶預付款將部分資本成本移轉至買方；NVIDIA持有客戶的股權，並為客戶的容量提供保證[S-095; S-002]；Meta則以外部資本進行建置[S-074]。當供應商、出租人與客戶之間存在財務連結時，同一份合約可能同時呈現為供應商的訂單、雲端未出貨訂單與放款人的擔保品。因此，未出貨訂單同時也是交易對手曝險的一項指標；而在該項曝險有所揭露之處，例如CoreWeave，其曝險是集中的[S-065]。

觀察重點

Oracle，截至2026年11月30日的季度

剩餘履約義務（remaining performance obligations, RPO）、已交付容量，以及資本支出（capital expenditure）與2027財年900億至950億美元財測指引的對比，其中淨現金資本支出不超過700億

圖5。 2026年的資本計畫在年內持續上修，惟會計基礎變動者除外



各公司2026年資本支出（capital expenditure）財測指引，依其給出日期列示，單位為十億美元（台積電為資本預算）。Alphabet[S-014; S-015; S-016]；Meta[S-071; S-072; S-073]；Amazon[S-069; S-070]；CoreWeave[S-077; S-078; S-079]；台積電[S-082; S-083; S-084]；Microsoft為2026日曆年並含融資租賃[S-045; S-075]。

範圍 Microsoft 7月較低的數字，反映使用年限延長以及更多租賃被歸類為營業租賃，而營業租賃不計入資本支出（capital expenditure）；此為基礎的變動，而非縮減[S-075]。台積電4月的陳述將支出定於520億至560億美元的「towards the high end」，圖中即以該區間繪示[S-083]。

> 美元[S-076]；以及募集約400億美元債務與股權的計畫是否有任何變動[S-044]。

IF 已簽約需求要如期轉化為營收，前提是融資與交付能同步跟上。 此一研判（reading）成立的前提是：資本市場持續在營收之前為建設提供資金，且大型交易對手依約接受交付。應逐季觀察Oracle的自由現金流與剩餘履約義務（remaining performance obligations, RPO）的轉換情況、CoreWeave下一份年度報告中的客戶集中度，以及大型合約是否伴隨新的供應商保證或股權投資。若未出貨訂單（backlog）停止成長而債務持續上升、大型客戶延後或重新議定容量，或融資成本上升至足以使已簽約的建置放緩，則此一研判不成立。

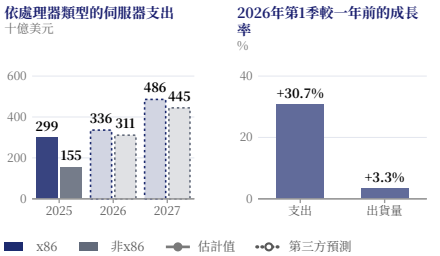
2.3 AI伺服器與NVIDIA訊號

資本計畫與未出貨訂單（backlog）是承諾；AI伺服器出貨量、伺服器支出與加速器（accelerator）營收則顯示支出已在多大程度上轉化為設備。承載此一訊號的主要有三組數列。TrendForce的出貨預測計數的是台數。IDC的伺服器追蹤報告計數的是各類伺服器的支出[S-592]。NVIDIA每季公布的Data Center營收計數的是金額，並在NVIDIA 2026年改採Data Center與Edge Computing兩個市場平台之後仍不中斷[S-097]。NVIDIA的揭露內容另顯示出貨量統計無法呈現的兩件事：地緣政治假設在何處進入供應商的展望，以及加速器需求如何在供應鏈更上游轉化為記憶體的採購承諾[S-097; S-002]。

F 隨著資本計畫上調，TrendForce調高其2026年AI伺服器成長預測。 2026年1月，TrendForce預測全球AI伺服器出貨量2026年將成長逾28%，其中以GPU為基礎的系統占出貨量的69.7%，以特殊應用積體電路（application-specific integrated circuit, ASIC）為基礎者占27.8%[S-008]。8月，該機構將成長預測調高至近31%，並將此次修正與九家最大雲端業者資本支出（capital expenditure）預期成長約90%相連結[S-013]。兩者皆為市場研究對出貨台數的預測，而非公司資料；第3.1節研讀GPU與客製加速器（accelerator）之間的組合。

F IDC的伺服器追蹤報告顯示支出的成長遠快於台數。 IDC指出，2026年第1季全球伺服器支出年增30.7%，「while unit growth was modest at 3.3% year over year」[S-592]。其表格將2025年伺服器支出列為4,535億美元，成長78.2%，並預估2026年為6,470億美元、2027年為9,306億美元；非以x86處理器為基礎的伺服器則自1,550億美元上升至2026年預估的3,111億美元[S-592]。IDC的基準假設為「a constrained memory/NAND supply environment with elevated pricing through at least the first half of 2027」，並指出供應商認為成長的主要上限在於供給而非需求[S-592]（圖6）。

圖6。 伺服器支出的成長遠快於伺服器出貨台數



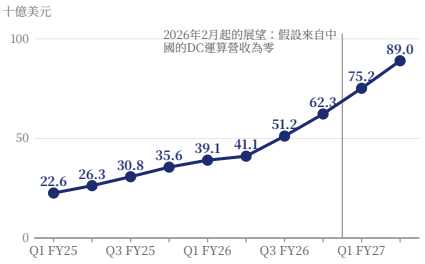
左圖：依處理器類型劃分的全球伺服器支出，以十億美元計，出自IDC《Servers Market Insights》，2026年7月28日更新；2026年與2027年數值為該日期的預估值，惟表格並未加以標示[S-592]。右圖：2026年第1季較一年前的成長率，單位為%[S-592]。

範圍 涵蓋全部伺服器，而非僅AI伺服器。x86與非x86是IDC依處理器架構所作的劃分，並非加速與非加速系統之別。數值為IDC所印的美元金額；該頁面未載明價格基準。

A 支出成長超越台數成長，所衡量的既是數量，也同樣是價格與組態。 支出成長率約為台數成長率的九倍，意味著每台出貨伺服器所承載的價值更高：以加速器（accelerator）、記憶體與機櫃級系統（rack-scale system）取代通用機器，且兩者的記憶體價格均上升[S-592; S-035]。第1季27.4個百分點的落差，是組合與價格的指標，而非平均售價——公開資料並未提供平均售價。這也是本報告將IDC的支出、TrendForce的台數與NVIDIA的營收並列研讀，且絕不將其中之一換算為另一者的原因。

F NVIDIA的Data Center營收在十個季度內增至近四倍。 NVIDIA的財年於1月下旬結束，故其2027財年延續至2027年1月。Data Center營收在2025財年第1季（結束於2024年4月28日）為226億美元，在2027財年第2季（結束於2026年7月26日）則達創紀錄的890億美元，一年內成長117%，較前一季成長18%[S-098; S-010; S-002]。NVIDIA將最近一季的表現歸因於其Blackwell Ultra基礎設施的量產爬升，並表示其下一個平台Vera Rubin正「ramping into full production」，將於第3季推出[S-002; S-010]。

圖7。 NVIDIA的Data Center營收在十個季度內增至近四倍



NVIDIA依財年季度劃分的Data Center營收，以十億美元計，採各季財報新聞稿首度公布的數字[S-098; S-099; S-100; S-101; S-102; S-103; S-104; S-050; S-051; S-010]。自2026年2月起提出的每一份展望，均假設來自中國的Data Center運算營收為零[S-050; S-051; S-010]。

範圍 此一數列具連續性：2027財年第1季的報導方式變更僅將其他平台重新歸入Edge Computing，而2027財年第2季對Hyperscale與ACIE兩個次市場的重編，並未改變Data Center總額[S-097; S-002]。NVIDIA的財年於1月下旬結束（2027財年第2季結束於2026年7月26日）。

F NVIDIA對2027財年第3季的展望，假設來自中國的Data Center運算營收為零。 NVIDIA預期該季全公司營收為1,080億美元，上下浮動2%，並表示其「is not assuming any Data Center compute revenue from China in its outlook」[S-010; S-002]。此一假設涵蓋展望期間內的Data Center運算營收；它並非表示NVIDIA預期全公司來自中國的營收為零。第

圖8。 已簽約需求的成長超越已交付容量，而各公司所用指標並不相同

各公司自身對已承諾未來營收的指標，以十億美元計，為各季末數；以及CoreWeave的已啟用與已簽約電力，以GW計。Google Cloud未出貨訂單（backlog）[S-014; S-015; S-016]；AWS未出貨訂單係Amazon執行長於法說會上所述：2月稱「our backlog」，4月為針對「what the AWS backlog looks like」一問的答覆，7月則為以「I'll start with AWS」開場的談話中所提的數字[S-089; S-085; S-070]；Oracle剩餘履約義務（remaining performance obligations, RPO）為其財年季末數（11月30日、2月28日、5月31日與8月31日），分別繪於2025年12月、2026年3月、2026年6月與2026年8月各欄[S-090; S-091; S-044; S-049]；CoreWeave營收未出貨訂單與電力[S-088; S-092; S-011; S-079]；Microsoft商用剩餘履約義務[S-075]。CoreWeave在2026年8月欄的已簽約電力係截至2026年8月11日[S-079]；其2026年3月的已啟用電力資料點為2026年5月7日所稱的「surpassed 1 GW」[S-092]。

範圍 未出貨訂單（backlog）、剩餘履約義務（remaining performance obligations, RPO）與營收未出貨訂單屬不同的會計概念，且已簽約電力的GW數並不等於已交付容量的GW數；各面板並非排名，亦不加總。

2季輸往中國的Data Center Hopper產品出貨量，低於Data Center營收的1%[S-002]。

A 中國已成為NVIDIA展望中的一項敏感因素，而非其成長的動能。 同一句話出現於2026年2月25日、5月20日與8月26日提出的三份2027財年展望中[S-050; S-051; S-010]；2025年8月針對2026財年第3季的展望則採用不同措辭，專指H20出貨[S-103]。其後公布的營收超出前兩份展望區間的上限：第1季為816億美元，展望為780億美元上下浮動2%；第2季為962億美元，展望為910億美元[S-050; S-051; S-010]。第1季輸往中國的Data Center Hopper出貨為零，一年前則為46億美元[S-097]。依NVIDIA自身的數字，2026年的成長並非建立在對中國的Data Center運算銷售之上；第7.1節研讀此一假設背後的出口管制（export control）。

F 需求正擴大至最大雲端業者之外。 NVIDIA指出，來自AI雲端、工業與企業（AI clouds, industrial and enterprise, ACIE）客戶的營收——此為其Data Center兩個次市場之一，另一為Hyperscale——在2027財年第2季年增138%、季增25%，「driven by end-demand from AI natives, enterprises, and sovereign customers」[S-002]。NVIDIA於該季重編兩個次市場之間的劃分，因此若未經重編表述，其數值在不同季度間不可比[S-002]。TrendForce預期北美主要雲端業者在NVIDIA機櫃採購中的占比，將自2025年的約70%下降至2026年的60%[S-105]。

F NVIDIA的採購承諾在一個季度內成長逾一倍，主要用於記憶體。 NVIDIA財務長指出，其承諾金額自2027財年第1季末的1,190億美元上升至第2季末的2,790億美元，「primarily related to the procurement of memory」[S-002]。隨著NVIDIA為推出Vera Rubin作準備，存貨自258億美元上升至316億美元[S-002]。TrendForce預估GB300、VR200與VR300世代NVL72機櫃系統的合計產值，2027年將超過7,100億美元[S-105]。

A 對加速器（accelerator）的需求正以記憶體承諾的形式向供應鏈上游傳遞。 當NVIDIA在一個季度內將採購承諾提高逾一倍、且主要用於記憶體時，其客戶所面對的稀缺正在更上游、於交付之前先行簽約鎖定[S-002]。這正是AI支出上升轉化為記憶體限制之處，而第3.2節循合約價格與長期協議（long-term agreement）追蹤此一限制。Data Center數列衡量的是金額而非台數：其成長有一部分反映系統價值更高，而非僅是數量更多；TrendForce的機櫃預估同樣是價值的衡量，而非台數的計數[S-105]。因此，出貨預測與營收數列係並列研讀，絕不彼此換算。

IF 加速器（accelerator）營收要持續上升，前提是新平台如期量產爬升且已承諾的記憶體順利到貨。 此一研判（reading）建立於NVIDIA對2026年10月結束之該季1,080億美元的展望，以及其已承諾採購的記憶體之上[S-010; S-002]。應觀察NVIDIA 2027財年第3季財報與該展望的對比、其採購承諾的水準，以及TrendForce對AI伺服器出貨成長的下一次修正。若承諾金額下降而營收持平，或第2.1節中的某一買方因供給以外的原因削減加速器訂單，則此一研判不成立。中國許可政策的變動會使營收落在展望假設之外，但不改變此一研判的其餘部分。

2.4 OCP認可市場所衡量的對象

Open Compute Project（OCP）自行發布一組市場規模數列，而該數列極易被誤讀為AI資料中心市場的規模。它並不是。它所衡量的是依OCP所認可規格（specification）製造的資訊科技（information technology, IT）設備支出，因此它是OCP開放硬體設計進入採購之程度的公開量尺，而對其他事項則是不佳的量尺。依其自身界定研讀，它回答了第2.1節中資本支出（capital expenditure）數字無法回答的一個問題：有多少設備採購是建立在不只一家供應商能據以製造的設計之上。

F OCP與IDC將OCP認可基礎設施的支出列為2025年1,320億美元，並於2029年上升至2,950億美元。 OCP的採用頁面載有與市場研究機構IDC共同完成、並於2025年12月1日宣布的研究結果：「OCP-recognized IT infrastructure and solutions」的總支出2025年為1,320億美元，預估2029年為2,950億美元，年複合成長率（compound annual growth rate, CAGR）為22.2%[S-004; S-106]。該研究計數的是全球依OCP認可規格（specification）採購的資料中心與資訊科技設備，並包含OCP會員之外的組織所採購者[S-004]。IDC係依市場模型，並透過訪談產品納入OCP認可標準的供應商，建構這些數字[S-107]。

F AI僅是該總額中的一部分，而散熱則屬成長最快的類別之一。 採用頁面預估AI基礎設施支出「around $155 billion by 2029」，主要由超大規模雲端業者（hyperscaler）、新興GPU雲端業者（neocloud）與其他組織的GPU加速AI伺服器所帶動[S-004]。IDC 2026年1月的網路研討會簡報顯示同一數字並標示為156，作為OCP基礎設施支出中的AI應用案例，是數個應用案例之一，因此它是2,950億美元的一部分，而非整個AI基礎設施市場[S-108]。在設備類別中，該研究指出散熱技術（年複合成長率逾32%）、電力技術（23.6%）與資訊科技機櫃基礎設施（23.5%）屬支出成長最強勁者；伺服器的預估年複合成長率為35.7%，仍為最大類別[S-004; S-108]。

A 該數字衡量的是與標準連結的設備支出，而非AI資料中心市場。 它不是OCP Foundation的營收、不是AI資料中心的總體潛在市場（total addressable market, TAM）、不是超大規模雲端業者（hyperscaler）資本支出（capital expenditure），也不是任何單一規格（specification）的採用程度；且除非方法經過對齊，否則不可與先前的市場研究比較[S-004; S-108]。公司資本支出亦用於支付資料中心建築：Alphabet將其技術基礎設施投資的約40%投入資料中心與網路設備[S-016]。因此，OCP數列與第2.1節中的資本支出數字，不僅來源不同，涵蓋範圍亦不同；以其中之一除以另一者，並不會得出OCP在AI支出中的占比。表2列出四種容易混淆的市場規模概念。

F IDC與Gartner所公布的AI基礎設施總額，其涵蓋範圍相差三倍。 IDC的追蹤報告將2025年全年AI基礎設施支出列為3,180億美元，「more than double the $153 billion recorded in 2024」，並預估2026年為4,870億美元，預測至2029年將逾1兆美元[S-596]。第4季支出899億美元中，伺服器為877億美元，占97.6%[S-596]。Gartner的AI基礎設施涵蓋

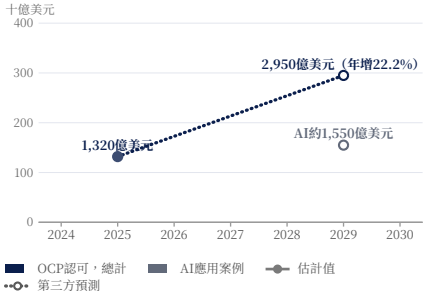
表2。 四種市場規模概念回答四個不同的問題

某一列的數字絕不可替代另一列的數字，亦不可與之相加。

概念	所衡量的對象	本章中的實例	不可用於何種用途
總體潛在市場（TAM）	在界定的市場中，界定產品的所有供應商可取得的全部營收	本報告未予估計；將以下各列的數字相加會混用不同性質的指標	在其範圍、期間與方法未載明之前，不可用於任何用途
支出預測	第三方對一組指名買方支出金額的預估	TrendForce：九家最大雲端業者2026年資本支出（capital expenditure）逾8,867億美元[S-013]	供應商的營收或市占率；企業財測指引——本項並不取代財測指引
公司營收機會	某一公司自稱在其自身領域預期可賺取或成長的幅度	TSMC：2024年至2029年AI加速器（accelerator）營收成長的年複合成長率接近「mid-to-high-fifties %」[S-082]	市場規模：公司對其自身業務具權威性，對市場則不然
認可（與標準連結）市場	依標準社群所認可規格（specification）製造的設備支出	OCP與IDC：2025年1,320億美元，2029年上升至2,950億美元[S-004]	OCP Foundation的營收、AI資料中心市場、超大規模雲端業者（hyperscaler）資本支出（capital expenditure），或單一規格（specification）的採用程度

各市場規模概念所衡量的對象、本章中的一項實例，以及其無法據以支持的內容；來源如各欄所引註。

圖9。 OCP認可基礎設施的支出預估將於2025年至2029年間成長逾一倍

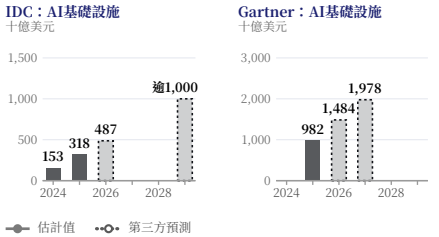


全球依OCP認可規格（specification）製造的資料中心與資訊科技設備支出，以十億美元計，出自IDC為OCP Foundation建立的市場模型，於2025年12月1日宣布[S-004; S-106]。AI應用案例是同一總額中的一部分：採用頁面載為2029年約1,550億美元，IDC 2026年1月的網路研討會簡報則標示為156[S-004; S-108]。

範圍 此為與標準連結的市場指標，並非OCP Foundation的營收、AI資料中心支出總額或超大規模雲端業者（hyperscaler）資本支出（capital expenditure），且除非方法經過對齊，否則不可與先前的市場研究比較。2025年數值於發布時為當年度估計值。

「AI-optimized IaaS, AI-optimized servers, AI network fabric, AI processing semiconductors and devices」，2025年達9,820億美元，並預測2026年為1.48兆美元、2027年為1.98兆美元，占其全部AI支出總額的半數以上[S-601]（圖10）。

圖10。 IDC與Gartner對AI基礎設施支出的估計相差三倍，但兩者均認為2026年將成長約一半



左圖：IDC的全球AI基礎設施支出，以十億美元計：2024年與2025年為全年數值，2026年與2029年為預測值（「exceed $1 trillion」，以1兆美元為下限繪製）[S-596]。右圖：Gartner的AI基礎設施（AI最佳化IaaS、伺服器、網路交換架構、處理用半導體與裝置），以十億美元計，為2026年9月的預測；2025年為剛結束的年度[S-601]。

範圍 定義不同：此二水準並非對同一市場的估計，絕不相加或平均。OCP與標準連結的數列（圖9）則是第三種涵蓋範圍。

A 各項水準不可比較，成長率則可以。 IDC與Gartner對AI基礎設施的定義不同，而OCP的數列計數的是依其規格（specification）製造的設備，不論是否與AI相關；因此三者之中無一是另一者的占比或查核依據[S-596; S-601; S-004]。三者的成長率比其水準更為一致：IDC 2026年的數字較2025年高出「approximately 53%」，Gartner高出51%，而OCP認可數列至2029年每年成長22.2%[S-596; S-601; S-004]。因此本報告將各數列分別置於自身面板，並僅比較成長率，絕不跨數列比較金額（第E.1節）。

? 未解問題——開放設計是否正在AI基礎設施支出中取得更高占比。 OCP數列預估至2029年年複合成長率為22.2%[S-004]，但並無公開預測以相同範圍與方法衡量AI基礎設施設備支出總額，因此無法計算OCP認可部分在其中的占比；IDC與Gartner的總額採用其他範圍（圖10）。這一點對採購與標準策略具重要意義：該數列的成長，既與開放介面逐步擴展相符，也與開放平台和專有平台並行成長相符（第6.7節）。

IF 此一數列要顯示開放設計成長於何處，前提是其基礎維持不變。 依其自身界定研讀，它將部分成長最強勁之處置於機櫃變動最大的地方，即散熱與電力（第4章）[S-004]。此一研判（reading）成立的前提是，每年更新均維持該研究的範圍與方法。應觀察下一次OCP與IDC的更新，以了解各類別成長率與AI部分，並觀察OCP產品認可制度的變更——依該變更，OCP Inspired標章原定於2026年7月31日之後廢止——是否改變了所計數的對象[S-109]。若後續更新變更基礎而未重編先前年度，或散熱與電力的成長率回落至總體水準，則此一研判不成立。

2.5 電力使AI成長成為一項能源系統課題

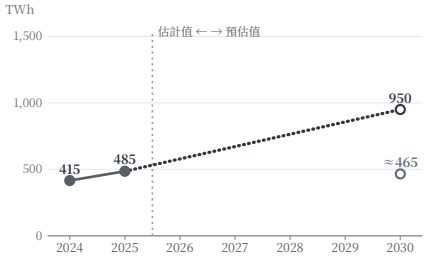
以第2.1節中的資本所購置的每一台伺服器，都會成為電網上的負載。相關金額如今已可與能源產業本身相比：國際能源總署（IEA）指出，僅五家科技企業的資本支出（capital expenditure），目前即大於全球石油與天然氣生產的投資[S-005]。這擴大了必須為AI需求進行規劃者的範圍：不僅是晶片製造商與雲端業者，還包括依較慢時鐘運作的電力公司、電網營運者與能源監管機關。國際能源總署的全球數列顯示，資料中心用電需求的成長速度為整體用電需求的數倍，而為AI建造的資料中心成長更快。

F 國際能源總署（IEA）預估全球資料中心用電量將於2025年至2030年間大致倍增。 國際能源總署估計資料中心用電量2024年約為415太瓦時（terawatt-hour, TWh）、2025年為485 TWh，並在其基準情境下預估2030年約為950 TWh，相當於屆時全球用電需求的約3%[S-005]。2025年資料中心用電需求成長17%，而全球用電需求僅成長3%[S-086]。2030年的預估與國際能源總署2025年報告所載的約945 TWh相比變動不大[S-110]。

F 以AI為核心的資料中心用電量於2025年成長一半。 國際能源總署（IEA）指出，以AI為核心的資料中心用電量2025年成長50%，而全部資料中心為17%，並預估至2030年將成長逾兩倍，在基準情境下達約465 TWh[S-005]。依國際能源總署自身的摘述，資料中心用電量「set to double by 2030, and power use from those focused on AI is poised to triple」[S-086]。兩項預估均為同一份2026年4月更新中的基準情境數字，因此可一併研讀，無須跨版本接續。

= 依國際能源總署（IEA）的數字，以AI為核心的資料中心自占資料中心用電量不到三分之一，上升至約半數。 計算公式：以AI為核心的用電量 ÷ 全部資料中心用電量。2030年：465 ÷ 950 TWh ≈ 49%[S-005]。國際能源總署未提供以AI為核心資料中心的2025年水準，但其成長逾兩倍至約465 TWh，意味該水準的上限低於約155 TWh，即低於485 TWh總量的32%[S-005]。組成與總量同等重要：負載中成長最快的部分，正是與第2.1節中AI資本計畫相連結的部分。

圖11。 全球資料中心用電量預估將於2025年至2030年間大致倍增



全球資料中心用電量，單位為TWh：2024年與2025年為國際能源總署（IEA）估計值，2030年為其基準情境預估值，均出自2026年4月的更新[S-005]。2025年用電量成長約17%；以AI為核心的資料中心成長約50%，並預估2030年約為465 TWh[S-005; S-086]。

範圍 僅涵蓋全球範圍。圖27中的美國估計值採用不同方法，不併入此一線圖。

A 負載正變得更密集，也更具爭議，這使問題超出能源供給的範疇。 國際能源總署（IEA）指出，AI伺服器的功率密度於2020年至2025年間增至11倍，並將於2027年前再增至四倍[S-005]。該署並指出，社會接受度是日益受關注的課題，「as communities push back against data centre projects」[S-005]。密集化使電力問題進入機櫃，第4.2節與第4.3節於此研讀工程細節；地方接受度則使問題進入規劃與許可核發，第7.2節於此研讀相關規則。

A 2026年投入的資本，須歷經數年才會成為電網負載，其時鐘比其所購置的伺服器更慢。 國際能源總署（IEA）將2025年資料中心用電量的增幅列為約70 TWh，而該年五家大型科技企業的支出逾4,000億美元；該署並預期其支出2026年將再成長75%[S-005; S-086]。容量合約可在數季之內簽訂，但併網與送電啟用需時更久，正如CoreWeave一家公司內部已簽約電力與已啟用電力之間的落差所顯示（圖8）[S-011]。本報告研判電力供給而非資本，才是兩項限制中較慢的一項：當電力晚於伺服器到來時，已承諾的支出便須等待送電啟用。

> **範圍 能源與容量**
>
> 太瓦時（terawatt-hour, TWh）衡量的是一年內使用的電量；吉瓦（gigawatt, GW）衡量的是電力容量。國際能源總署（IEA）2030年的950 TWh，與TrendForce對2026年全球資料中心電力需求容量161 GW的估計[S-043]，所衡量的是不同事物，若無使用率假設便無法彼此換算。美國數列與全球數列不予接續。

IF 國際能源總署（IEA）2030年的路徑要成立，前提是併網與發電能跟上資本計畫的步調。 950 TWh的預估要求資料中心容量大致依支出所隱含的時程完成併網與供電[S-005]。應觀察國際能源總署的下一次更新、主要市場的併網排隊佇列與大型負載（large load）規

則，以及雲端業者已簽約電力與已啟用電力之間的落差。若併網延遲使已併網容量低於計畫，或效率改善使每單位運算所耗能源下降的速度快於部署成長，則此一路徑不成立。第7.3節闡述以不同方法建構的美國區間（圖27），第7.2節則闡述併網規則。

本章各項指標指向同一方向，但所計數的對象並不相同。資本計畫、雲端未出貨訂單（backlog）與NVIDIA的營收在2026年間均上升，但此一上升有一部分來自價格上漲，有一部分是在買方自身資產負債表之外取得融資，還有一部分則在等待電力。後續各章追蹤資金與稀缺的交會之處：記憶體與封裝見第3章，機櫃內的電力與散熱見第4章，供應商的產能與利潤率見第5章，電網則見第7章。

讀者視角

本章對財務、產業與政府的意義

財務。 研讀資本計畫時應併同未出貨訂單（backlog）、預付款、保證與自由現金流一併檢視；各項定義不同，且融資如今已置於供應層次之內。

產業。 2026年增幅的一部分用於支付較高的記憶體價格，因此以金額表示的數字高估了台數成長；Amazon表示其2027年容量大部分已被預訂。

政府。 在國際能源總署（IEA）基準情境下，資料中心用電量至2030年大致倍增；決定已承諾容量何時成為負載的是併網，而非資本。

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-002 NVIDIA (26 Aug 2026); S-010 NVIDIA (26 Aug 2026); S-011 CoreWeave (11 Aug 2026); S-014 Alphabet (4 Feb 2026); S-015 Alphabet (29 Apr 2026); S-016 Alphabet (22 Jul 2026); S-044 Oracle (10 Jun 2026); S-045 Microsoft (29 Apr 2026); S-046 Meta Platforms (29 Apr 2026); S-049 Oracle (10 Sep 2026); S-050 NVIDIA (25 Feb 2026); S-051 NVIDIA (20 May 2026); S-065 CoreWeave (2 Mar 2026); S-069 Amazon.com (5 Feb 2026); S-071 Meta Platforms (28 Jan 2026); S-072 Meta Platforms (29 Apr 2026); S-073 Meta Platforms (29 Jul 2026); S-074 Meta Platforms (29 Jul 2026); S-075 Microsoft (29 Jul 2026); S-077 CoreWeave (26 Feb 2026); S-078 CoreWeave (7 May 2026); S-079 CoreWeave (11 Aug 2026); S-080 Alphabet (23 Jul 2026); S-081 Microsoft (29 Jul 2026); S-082 TSMC (15 Jan 2026); S-083 TSMC (16 Apr 2026); S-084 TSMC (16 Jul 2026); S-087 Microsoft (30 Jul 2025); S-088 CoreWeave (26 Feb 2026); S-090 Oracle (10 Dec 2025); S-091 Oracle (10 Mar 2026); S-092 CoreWeave (7 May 2026); S-094 Microsoft (29 Jul 2026); S-095 CoreWeave / NVIDIA (26 Jan 2026); S-096 Amazon.com (30 Jul 2026); S-097 NVIDIA (20 May 2026); S-098 NVIDIA (22 May 2024); S-099 NVIDIA (28 Aug 2024); S-100 NVIDIA (20 Nov 2024); S-101 NVIDIA (26 Feb 2025); S-102 NVIDIA (28 May 2025); S-103 NVIDIA (27 Aug 2025); S-104 NVIDIA (19 Nov 2025); S-492 Alphabet (23 Jul 2026); S-493 Meta Platforms (30 Jul 2026)

標準組織與聯盟. S-004 OCP (1 Dec 2025); S-109 OCP (11 May 2026)

政府與監管機關. S-005 IEA (16 Apr 2026); S-086 IEA (16 Apr 2026); S-110 IEA (10 Apr 2025)

市場研究. S-007 TrendForce (25 Aug 2026); S-008 TrendForce (20 Jan 2026); S-013 TrendForce (3 Aug 2026); S-035 TrendForce (1 Jun 2026); S-043 TrendForce (16 Sep 2026); S-105 TrendForce (28 Aug 2026); S-106 IDC for OCP (12 Nov 2025); S-107 IDC (Dec 2025); S-108 OCP / IDC (22 Jan 2026); S-592 IDC (28 Jul 2026); S-596 IDC, Resource Center blog (16 Apr 2026); S-597 TrendForce (25 Feb 2026); S-598 TrendForce (6 May 2026); S-599 Gartner (15 Jan 2026); S-600 Gartner (19 May 2026); S-601 Gartner (16 Sep 2026)

媒體與法律評論. S-070 StockAnalysis (30 Jul 2026); S-076 StockAnalysis (10 Sep 2026); S-085 The Motley Fool (29 Apr 2026); S-089 The Motley Fool (5 Feb 2026); S-093 Investing.com (5 Feb 2026)

第3章

運算、記憶體與封裝

加速器（accelerator）已不再是AI運算中唯一稀缺的環節：記憶體、先進封裝（advanced packaging）、鍵合與測試如今決定了可建置的規模、成本高低，以及議價能力（bargaining power）歸誰所有。

適用對象： 財務、採購、研發（R&D）與智慧財產權（IPR）讀者；市場脈絡見第3.1節與第3.2節，技術細節見第3.3節、第3.4節與第3.5節。

AI建置的第一階段受加速器（accelerator）配給所限。2026年，稀缺日益落在其周邊環節：堆疊於各處理器旁的高頻寬記憶體（high-bandwidth memory, HBM）、其後方的伺服器記憶體與快閃記憶體、將兩者結合的先進封裝，以及決定良率（yield）的鍵合與測試步驟。記憶體價格上漲之快，足以改變資本計畫，而記憶體製造商的利潤率則顯示議價能力（bargaining power）流向何處。本章自晶片向外追蹤此一限制，並探討這些層級中何者如今限制了可建置的AI運算規模，以及各層級依何種時鐘擴張。

3.1 GPU與客製晶片同步成長

AI加速器（accelerator）是執行模型運算的處理器，而購置加速器的兩種途徑如今並行成長。以NVIDIA的圖形處理器（graphics processing unit, GPU）為首的市售平台，同樣銷售給雲端業者、新興GPU雲端業者（neocloud）與企業。客製加速器則是雲端業者與Broadcom等晶片合作夥伴共同設計的特殊應用積體電路（application-specific integrated circuit, ASIC），服務單一買方自身的工作負載。對整個系統而言，問題不在於何種途徑勝出，而在於兩者都仰賴同一批晶圓代工晶圓、先進封裝（advanced packaging）、HBM與測試產能。

正因存在此一共同依賴，加速器（accelerator）組合的重要性才超出晶片設計本身。每一項客製專案都新增一款必須經過封裝、與記憶體共同驗證並完成測試的設計，因此客製占比上升會擴大對稀缺投入要素的需求，而非減輕之。以下證據顯示兩條途徑均有成長，且並未出現其中一者取代另一者的情形。

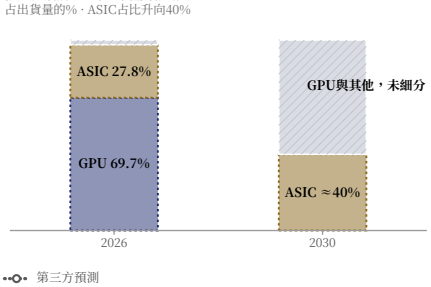
F 市售GPU平台仍逐季擴張。 NVIDIA公布截至2026年7月26日該季的Data Center營收為890億美元，較一年前成長117%[S-010]。該公司表示其Vera Rubin平台正「ramping into full production」，機櫃已在CoreWeave、Google Cloud、Microsoft Azure、Oracle Cloud Infrastructure與Nebius等合作夥伴處運行[S-010]。隨著該公司準備於其財年第3季推出Vera Rubin，其存貨上升至316億美元[S-002]。第2.3節說明該營收數列與展望的脈絡。

F AMD正在機櫃層級建立第二個市售平台。 AMD公布2026年第2季Data Center部門營收為67億美元，較一年前成長107%，係受其EPYC處理器與Instinct GPU的需求帶動；該公司並表示其Helios機櫃「is being deployed by leading AI labs and cloud providers」[S-111]。Helios係「developed jointly by AMD and Meta through the Open Compute Project」（OCP），屬於一項擴大合作的一部分，該合作將於Meta部署最多6吉瓦的AMD GPU[S-112]。第4.5節比較承載此二市售平台的機櫃設計。

F 客製晶片正在多條戰線上同時放大規模。 Broadcom 2026財年第3季的AI半導體營收為167億美元，較一年前成長221%；該公司預期下一季為217億美元，並表示對其客製AI加速器（accelerator）與網路產品的需求「continues to be very strong」[S-113]。Alphabet於2026年第2季開始向客戶的資料中心交付Tensor Processing Unit（TPU）系統[S-016]。TrendForce預測Amazon的Trainium在2026年將占Amazon自身AI伺服器出貨量的逾40%[S-114]。

F 市場研究將客製加速器（accelerator）列為2026年AI伺服器出貨量的四分之一以上，並於2030年前升向40%。 2026年1月，TrendForce預測以GPU為基礎的系統將占2026年AI伺服器出貨量的69.7%，以ASIC為基礎者占27.8%，為2023年以來ASIC的最高占比[S-008]；3月，該機構預測ASIC占比至2030年將達近40%[S-009]。8月，該機構將2026年AI伺服器出貨成長預測自28%調高至近31%，但未重編該項組合[S-013]。這些占比計數的是出貨台數，而非營收或運算能力（圖12）。

圖12。 客製加速器在AI伺服器中的占比上升，是與GPU並存，而非取代GPU



依加速器（accelerator）類型劃分的AI伺服器出貨量占比，為TrendForce的預測：2026年取自2026年1月[S-008]，2030年ASIC占比（「nearly 40%」）取自2026年3月[S-009]。TrendForce於1月將2026年AI伺服器出貨成長列為逾28%，8月則為近31%[S-008; S-013]。

範圍 為出貨台數的占比，而非營收或已安裝基數的占比。TrendForce未提供2030年的GPU占比；2030年的其餘部分以未細分方式繪製。占比預測並不足以推論出勝出者。

A 紀錄顯示的是兩條同時成長的管道，而非相互取代。 NVIDIA的Data Center營收與Broadcom的AI半導體營收在一年內各成長逾一倍，而TrendForce上升的ASIC占比所對應的出貨總量，該機構預期2026年將成長近31%[S-010; S-113; S-013]。占比預測並不指名勝出者。紀錄確實顯示的是：更多不同的加速器（accelerator）設計，正競逐同一批先進封裝（advanced packaging）產線、HBM驗證（qualification）與測試機台時間。因此，客製途徑是增加而非減輕本章所檢視各層級的壓力。

F Samsung與Broadcom為一項延續至2030年的記憶體暨晶圓代工合作，附上逾2,000億美元的共同估計金額。 該數字隨2026年7月25日宣布的一份合作意向書（memorandum of understanding, MOU）提出：「The companies expect the collaboration to be estimated at more than $200 billion across memory and foundry over the next five years through 2030」[S-115]。記憶體範圍包括供Broadcom次世代AI加速器（accelerator）使用的HBM；晶圓代工範圍涵蓋Samsung的2奈米及以下製程，並預期擴及2.3D與2.5D封裝[S-115]。該新聞稿未說明此一數字究係營收、訂單或採購承諾，亦未提供記憶體與晶圓代工之間的劃分。

A 此一估計金額屬策略情境（scenario，S類），而非財測指引或合約金額。 它所依據的是合作意向書而非確定性供貨協議，且其單位未經界定，因此不可與任一公司的財測指引相加，亦不可與未出貨訂單（backlog）類指標比較[S-115]。其意義在於結構層面：它將供客製加速器（accelerator）使用的HBM與尖端晶圓代工產能，繫於一項位於GPU供應商供應鏈之外的多年期安排。對客製途徑而言，這相當於記憶體供應商現與GPU供應商及雲端業者所簽長期協議（long-term agreement）的對應安排（第3.2節）。

F 測試設備供應商認為客製途徑正在擴大。 Advantest指出，其系統單晶片（system-on-chip, SoC）測試機在GPU、客製ASIC與中央處理器（central processing unit, CPU）各領域的需求「very strong」，並認為此一需求可延續至2026年並進入2027年[S-116]。該公司並表示，雖然目前GPU仍是最大市場，但「custom ASICs could become as large as, if not significantly larger than, the GPU market over time」[S-116]。此一說法涉及的是對Advantest測試機的需求，且屬一種可能性（「could」），而非對加速器（accelerator）營收的預測；其重要性在於每一款加速器設計都必須通過測試產能（第3.5節）。

IF 客製晶片將在2027年之前持續與GPU並存並提高占比。 此一觀點成立的前提是，雲端業者持續為其自有加速器（accelerator）專案提供資金，且這些專案能與市售平台一同取得封裝與HBM配額。應觀察TrendForce下一次的GPU與ASIC劃分、Broadcom的AI半導體營收與其財測指引的對比、Alphabet向客戶交付TPU的情況，以及Samsung與Broadcom的合作意向書是否成為確定性協議。若客製專案因封裝或記憶體不足而延後一個世代，或TrendForce下調其ASIC占比，則此一研判（reading）不成立。

3.2 記憶體除了成為效能限制，也成為預算限制

記憶體是AI系統中保存模型與對話內容的部分。HBM位於加速器（accelerator）封裝之上並為其供料；伺服器動態隨機存取記憶體（dynamic random-access memory, DRAM）與固態硬碟（solid-state drive, SSD）則位於其後方。在建置的大部分期間，記憶體是一項效能問題：每顆加速器可達到多少頻寬與容量。2026年，它同時成為一項預算問題，因為記憶體價格上漲之快，足以改變最大買方的支出計畫。

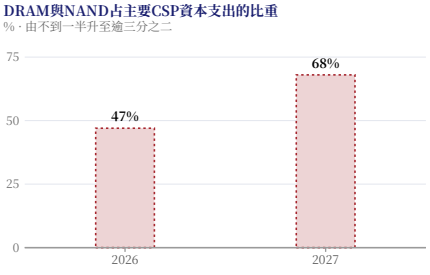
此一轉變使議價能力（bargaining power）發生移轉。當少數幾家記憶體供應商以多年期協議預售產能時，記憶體的價格與配額便開始決定買方能部署多少加速器（accelerator）、成本幾何，而利潤也移向記憶體製造商。本節依四個步驟追蹤相關證據：價格與預算、供應商的財報與協議、買方與系統整合商的說法，以及何者可望緩解此一壓力。

價格與預算

F 市場研究估計，記憶體將占主要雲端業者2026年資本支出（capital expenditure）的近半數，2027年則逾三分之二。 TrendForce估計，DRAM與NAND快閃記憶體合計將占全球主要雲端服務供應商（cloud service provider, CSP）2026年資本支出總額的47%，2027年為68%[S-007]。該機構將此一上升主要歸因於合約價格：伺服器DRAM合約價格於2025年下半年累計上漲64%，並預期2026年將再跳升約270%[S-007]。此等估計假設CSP資本支出總額2026年成長約98%、2027年成長50%；TrendForce未指明相關業者（圖13）。

= 依TrendForce自身的假設，2027年記憶體支出大致倍增，而其他所有項目的支出則下降。 主要CSP 2026年資本支出每100單位中，記憶體占47單位，其他所有項目占53單位。若支出總額成長50%，2027年支出為150單位，其中記憶體以68%計為102單位，其他所有項目為48單位[S-007]。據此，記憶體支出將增至約2.2倍（102 ÷ 47），而其他資本支出合計則下降約

圖13。 記憶體在主要雲端業者資本支出中的占比正在擴大



DRAM與NAND快閃記憶體占全球主要雲端服務供應商資本支出（capital expenditure）總額的比重，為TrendForce於2026年8月25日發布的估計，該估計假設CSP資本支出總額2026年成長約98%、2027年成長50%[S-007]。

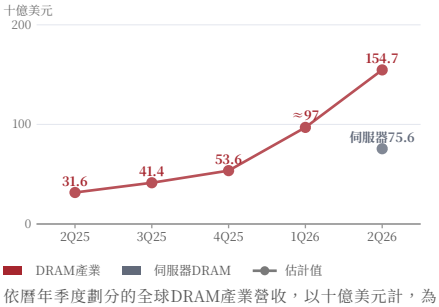
範圍 TrendForce未指明亦未計數相關業者。此為市場研究的估計，而非公司數字。

9%（48 ÷ 53 = 0.91）。所有輸入值均出自同一份TrendForce發布資料，因此業者樣本與版次一致。

A 2027年或是記憶體排擠其他採購，或是該項估計證明過高；兩者都使記憶體成為應觀察的變數。 若支出總額成長一半而非記憶體支出下降，將意味吸收預算增額的是記憶體價格，而非加速器（accelerator）數量。另一種研判（reading）則是，該68%內含的價格漲幅會在2027年預算執行之前被新增供給所侵蝕。無論如何，此一轉變是由價格而非數量所主導[S-007]，而由價格主導的占比，其反轉速度可能快於由數量主導者。這些數字是針對一群未指名業者的估計值，不應與任何公司的財測指引相乘。

F 受合約價格帶動，DRAM產業營收在四個季度內增至近五倍。 TrendForce估計全球DRAM產業營收2025年第2季為316.3億美元，2026年第2季則達近1,547.3億美元，其間第1季季增81%、第2季季增59.5%[S-117; S-035; S-118]。該機構將第1季的跳升歸因於傳統DRAM合約價格在一個季度內上漲約93%至98%[S-035]。僅伺服器DRAM一項，2026年第2季即帶來755.8億美元營收，較第1季成長53%，約占總額的半數[S-119]（圖14）。

圖14。 合約價格飆升，DRAM產業營收在四個季度內增至近五倍



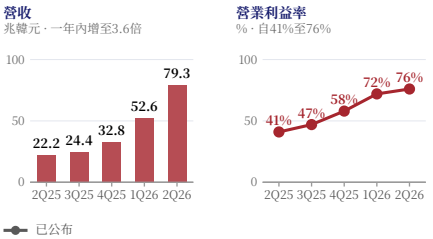
依曆年季度劃分的全球DRAM產業營收，以十億美元計，為TrendForce的估計[S-117; S-120; S-121; S-035; S-118]；伺服器DRAM營收為2026年第2季[S-119]。2026年第1季的數值在來源中係四捨五入至十億美元整數。

範圍 為市場研究對已結束季度的估計，含高頻寬記憶體（high-bandwidth memory, HBM）；伺服器DRAM是總額的一部分。

供應商的利潤率與協議

F SK hynix的財報顯示價格漲幅中有多少到達供應商手中。 SK hynix公布2026年第2季營收為79.3兆韓元，營業利益為60.5兆韓元，營業利益率為76%，營收較一年前成長257%，營業利益成長557%[S-003]。一年前，其營業利益率為41%[S-122]。其淨利93.9兆韓元之所以超過營收，係因業外項目所致，這也是圖15僅繪製營業利益率的原因[S-003]。

圖15。 SK hynix的營業利益率在四個季度內自41%上升至76%



依曆年季度劃分的SK hynix合併營收（兆韓元）與營業利益率（%）[S-122; S-123; S-124; S-125; S-003]。

範圍 2026年第2季118%的淨利率反映業外項目，未予繪製。

> "In a market environment where customer demand exceeds supply capabilities, the ability to deliver requested volumes in a timely manner has emerged as a core business competitiveness." [S-003]

F SK hynix描述的是一個客戶需求超出供應能力的市場，並正以長期協議（long-term agreement）將客戶綁定。 上文引述的說法出自其第2季業績發布，所指為記憶體整體，而非僅HBM[S-003]。SK hynix已與約10家客戶完成長期協議（LTA）簽訂，其中包括主要策略夥伴，並正擴大多年期合約的洽談；該公司於第2季開始量產出貨最新一代HBM產品HBM4，並於上半年完成HBM4E樣品出貨[S-003]。6月，該公司宣布與NVIDIA建立多年期技術夥伴關係，稱其"supports supply"，惟未載明數量、期間或產品[S-126]。

F Samsung預期記憶體市場供不應求的狀況將持續至2026年下半年。 Samsung Electronics於7月表示，預期下半年以伺服器為中心的記憶體需求將維持強勁，伺服器DRAM、企業級SSD與HBM的需求成長"expected to accelerate"，而這"projected to keep the market undersupplied"[S-127]。其含記憶體業務在內的Device Solutions部門，第2季以127.5兆韓元的營收取得89.2兆韓元的營業利益[S-127]。2月，Samsung表示預期其HBM銷售額在2026年將較2025年"more than triple"；至7月，該公司已向主要客戶出貨其自稱為業界首批的HBM4E樣品[S-056; S-127]。

F Micron與Kioxia描述了同樣轉向多年期合約的動向。 Micron表示，其多年期策略客戶協議（Strategic Customer Agreements）對其財務表現將"significantly enhance the durability and predictability"[S-128]，並預估現金保證金及相關財務承諾達220億美元[S-129]。快閃記憶體製造商Kioxia正自一年期轉向多年期長期協議（long-term agreement，LTA），其財務長表示，計劃在2028財年（即截至2029年3月的年度）之前，將資料中心產品提升至營收的逾60%[S-130; S-131]。Kioxia的這項數字屬管理層目標（類別T），而非財測指引。

A 記憶體現已按合約配置，而利潤率顯示議價能力（bargaining power）移向何方。 SK hynix與Samsung描述需求高於供給；SK hynix與Micron報告與客戶簽有長期協議（long-term agreement），Kioxia則表示正自一年期轉向多年期長期協議[S-003; S-127; S-128; S-130]（表3）。SK hynix的營業利益率在四個季度內自41%上升至76%，同期營收成長3.6倍，此一型態顯示的是定價能力，而非僅有量的成長[S-122; S-003]。這些合約的作用是雙向的：Micron收取保證金[S-129]，而TrendForce報告指出，數家美國雲端服務供應商（CSP）握有限制供應商調漲價格的多年期長期協議，因此漲幅更多落在此類協議之外的買方身上[S-132]。

買方與系統整合商感受到成本壓力

F 兩家最大的買方已表示，記憶體或零組件價格推升了其2026年的支出。 Amazon於7月表示，目前預期2026年的現金資本支出（capital expenditure）約為2,200億美元，並稱"the higher cost of memory pushing this number up from our prior estimate of about $200 billion"[S-070]。Microsoft於4月表示，其2026日曆年約1,900億美元的預期金額中，包含約250億美元"from the impact of higher component pricing"[S-045]。7月，該公司將該預期金額改述為約1,750億美元，原因是使用年限延長使部分租賃移出資本支出，其投資計畫則維持不變[S-075]（第2.1節）。

F NVIDIA的承諾金額在一個季度內增加逾一倍，主要用於記憶體。 依其財務長的說明，NVIDIA的承諾金額在2027財年第2季（結束於2026年7月26日）自1,190億美元增至2,790億美元，"primarily related to the procurement of memory"[S-002]。NVIDIA的財政年度於1月底結束，因此2027財年自2026年初延續至2027年初。該數字是購買的承諾，而非短缺的陳述；第2.3節將其與NVIDIA的營收及展望並列。

F 緯穎（Wiwynn）的營收基礎於2026年4月發生變動。 就部分客戶而言，記憶體採購已改採代理模式，因此該部分記憶體金額不再計入營收或銷貨成本[S-064; S-063]。緯穎表示，其持續與客戶合作調整商業模式，"in response to changes in the AI product mix and rising component costs"[S-063]。跨越2026年4月的營收與毛利率比較需附基礎說明。來源所指為記憶體而非HBM，且該項變動涵蓋部分客戶，而非全部（第5.2節；圖22）。

A 記憶體現已成為沿供應鏈流動、且各方皆須據以規劃的一項預算項目。 雲端服務供應商因記憶體成本而上調資本計畫、GPU供應商在需求之前即承諾採購記憶體、伺服器製造商改變記憶體在其帳上的計入方式，這三者是同一項轉變的三種視角：記憶體已成為一項與其所搭載的伺服器分開議價、融資與配置的成本[S-070; S-002; S-063]。這些說法沒有任何一項本身足以證明短缺的具體規模。但合起來，它們顯示記憶體價格正同時形塑資本預算、供應商承諾與系統整合商的營收。

何者可望緩解此一壓力

? 未解問題——HBM缺口的規模並未公開。 供應商以文字陳述需求高於供給，而已公布的數字衡量的是價格、營收、支出占比與封裝產能；就同一期間而言，無一業者以位元或金額陳述HBM的需求與供給[S-003; S-007; S-033]。缺少此一衡量，買方與投資人即無從判斷已公告的產能是在2027年或要到2029年才能彌平缺口。

F 新增記憶體產能的到位時程以一至三年為單位。 SK hynix正加速其M15X廠的量產時程，並計劃在龍仁第1期無塵室於2027年初啟用後迅速擴充產能[S-003]。其位於印第安納州的廠區為該公司在美國的首座HBM生產基地，預定於2028年10月前完成無塵室，並於2029年下半年開始量產次世代HBM[S-023]。Samsung的長期計畫訂出2026年至2040年約2,450兆韓元的投資規模，其中約2,100兆韓元用於半導體，並說明該計畫是依現行市場狀況所訂的指導方針[S-134]。

F 記憶體相關設備支出正在上升，這是較慢的紓解途徑。 SEMI預測DRAM設備銷售額2026年成長39.0%至388億美元，其後2027年成長27.4%、2028年成長15.0%，達569億美元；NAND設備則於2026年增至139億美元、2028年增至208億美元[S-602]。SEMI於6月發布的300 mm晶圓廠展望，將記憶體晶圓廠設備投資列為2026年520億美元、2027年570億美元，300 mm記憶體產能則列為2026年每月410萬片晶圓、2027年420萬片[S-603]；其第3季更新則將記憶體設備成長率列為2026年32%、2027年15%[S-604]（圖16）。

A 設備是可能侵蝕記憶體議價能力（bargaining power）的延遲機制，且它並非一對一地轉化為HBM位元。 2026年購入的設備，須待晶圓廠興建與驗證（qualification）完成後才會形成產出，且其新增的晶圓由HBM、伺服器DRAM與一般產品共同分攤；HBM較高的堆疊也增加了鍵合與測試步驟（第3.5節）。TrendForce預期HBM位元出貨量2027年將成長50%至60%，但仍不足以滿足需求[S-135]。因此，記憶體資本支出上升是紓解的領先指標，而非紓解本身，其效果將依前述一至三年的時程到來。

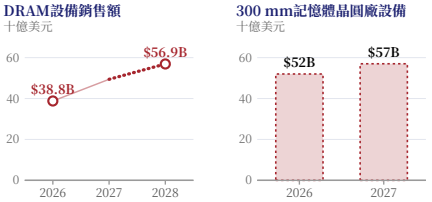
表3。 HBM供應商正預售產能並擴產HBM4，但無一業者以位元陳述其HBM產能

在HBM4擴產、HBM4E送樣的同時，記憶體透過多年期協議進行配置，因此驗證（qualification）與合約條件決定了2026年至2027年間誰能取得供應。

供應商	需求與供給（依其說法）	協議	HBM4與HBM4E狀態	產能日期	來源
SK hynix	客戶需求"exceeds supply capabilities"（記憶體整體，2026年7月）	與約10家客戶簽有長期協議（long-term agreement，LTA），其中包括主要策略夥伴；與NVIDIA的多年期夥伴關係稱可"supports supply"，惟未載明數量	HBM4自2026年第2季起量產出貨，並於2026年下半年（2H）擴產；HBM4E樣品於2026年上半年（1H）完成	M15X時程提前；龍仁第1期無塵室於2027年初啟用；印第安納無塵室於2028年10月前完成，次世代HBM於2029年下半年（2H）量產	[S-003; S-126; S-023]
Samsung Electronics	市場"projected to"在2026年下半年維持供不應求；伺服器DRAM、企業級SSD與HBM的需求成長"expected to accelerate"；HBM銷售額預期在2026年"more than triple"（2月）	與Broadcom簽署涵蓋記憶體（含HBM）的備忘錄（MOU），供Broadcom次世代AI加速器（accelerator）使用；連同晶圓代工合計，至2030年的共同估計值逾2,000億美元	HBM4自2026年2月起量產並商業出貨；HBM4E樣品於2026年第2季前交付主要客戶；客製HBM樣品於2027年	HBM4產能正在擴充（2026年2月）；至2040年約2,450兆韓元的長期投資指導方針	[S-127; S-056; S-115; S-134]
Micron	—	多年期策略客戶協議（Strategic Customer Agreements）；預估現金保證金及相關財務承諾達220億美元	HBM4（1-beta）已為其主要客戶的平台大量出貨；HBM4E（1-gamma）預計於2027日曆年量產	—	[S-128; S-129]

各HBM供應商就需求、協議、產品狀態與產能所公布的內容，依其自身用語呈現，資料截至2026年9月27日；來源列於最末欄。
範圍 除指名HBM者外，需求說法與協議均指記憶體整體。Samsung與Broadcom的數字係備忘錄（MOU）下的共同估計值（類別S），並非訂單或營收財測指引。
破折號表示所引述的發布文件未有此類說法。據報導（Investing.com引述Bloomberg），三家供應商均於2026年6月通過NVIDIA Vera Rubin所需的HBM4認證（certification）[S-133]。

圖16。 記憶體相關設備支出至2028年持續上升，是對稀缺較慢的回應



左圖：設備製造商的DRAM設備銷售額，單位為十億美元，SEMI於2026年7月14日發布的年中預測；2027年數值係依SEMI所給27.4%的成長率推導，並以推導橋接值繪製[S-602]。右圖：記憶體領域的300 mm晶圓廠設備投資，單位為十億美元，取自SEMI《300mm Fab Outlook》2026年第2季版（2026年6月4日發布），並依SEMI於2026年6月29日的發布內容轉述[S-603; S-604]。
範圍 兩項基礎不同的SEMI衡量指標——設備製造商的DRAM設備銷售額，以及300 mm記憶體晶圓廠的設備支出——並列呈現，不予加總。設備支出並不等於位元產能。

F 另一家市場研究機構則以其自有且更廣的尺度衡量記憶體營收。 TechInsights預估記憶體IC銷售額將"grow a record-breaking 248% in 2026, reaching $706 billion"，以HBM與企業級QLC固態硬碟為核心，其中DRAM與NAND快閃記憶體占市場營收的99%[S-605]。該機構並表示，"while AI components have low unit volume, their soaring prices account for the majority of total revenue"[S-605]。其定義與TrendForce的季度DRAM營收不同（圖14），因此兩者被視為價格主導市場的獨立訊號，不予平均或合併。

IF 壓力唯有在位元供給追上需求時才會緩解。 只要合約價格持續上漲、新廠房仍在一至三年之後，此一預算研判（reading）即成立；TrendForce預期DRAM供給至2027年仍將緊俏，並報告NVIDIA已開始評估Rubin Ultra採用較低HBM配置的方案，最終規格（specification）尚未確定[S-135]。應逐季觀察伺服器DRAM合約價格、各新平台在多家供應商處的HBM驗證（qualification）進度、M15X與龍仁的擴產，以及TrendForce下一次更新中記憶體占CSP資本支出的比重。若合約價格連續數季下跌，同時多家供應商已在主要平台上完成驗證，或供應商報告庫存上升、協議期間縮短，則此一研判不成立。

> **觀察重點**
> **下次檢視時點。** SK hynix與Samsung的2026年第3季業績（2026年10月下旬）；TrendForce對第3季DRAM營收與第4季合約價格的估計；SK hynix龍仁第1期無塵室預定於2027年初啟用[S-003]。

3.3 為何決定節奏的是記憶體，而非算術運算

記憶體所承受的商業壓力有其技術根源。大型語言模型每次僅生成一個token，而每個token都須從記憶體讀取模型權重與該段對話已儲存的上下文。算術運算單元的成長速度已超過為其供料的記憶體，因此在服務推論時，決定一具加速器（accelerator）能服務多少使用者、以及速度多快的，是記憶體的速度與容量，而非算術運算的次數。

本節以淺顯的方式說明該項算術關係，說明模型與軟體的進展改變了什麼，並提出記憶體展望所繫的問題：效率上的增益究竟能紓解記憶體需求，抑或會被更長的上下文與更多的使用者所吸收。第8章將同一套算術關係套用於2025年與2026年發布的開放權重模型（open-weight model）。

解碼階段讀取記憶體

F 算術運算的成長已超越為其供料的記憶體。 一項由加州大學柏克萊分校主導的研究指出，過去二十年間伺服器浮點運算吞吐量的尖峰值每兩年成長3.0倍，而DRAM頻寬僅成長1.6倍、互連頻寬僅成長1.4倍，使記憶體而非運算成為AI應用的主要瓶頸，在服務推論時尤為明顯[S-136]。來自華盛頓大學與Microsoft的Splitwise，將推論劃分為運算密集的提示階段與記憶體密集的token生成階段[S-137]。FlashAttention重新組織了精確注意力的計算方式，以減少HBM與晶片本身上較快記憶體之間的資料傳輸量[S-138]。

A 在服務推論時，記憶體頻寬決定速度，記憶體容量決定規模。 每一個新的token都在解碼（decode）步驟中產生，須讀取模型每一個作用中的權重，以及該段對話已儲存的上下文，即鍵值（key-value，KV）快取，而每讀取一個權重僅執行約兩次算術運算。將多位使用者批次合併，可重複利用權重的讀取，但無法重複利用快取的讀取，因為每一段對話僅關注自身的快取。因此，頻寬決定每秒token數，容量則決定有多少段對話、多長的對話能與處理器並存[S-137; S-136]。一具更快的算術運算單元若未減少每個token所搬移的位元組數，解碼速度便不會改變。

F HBM4加寬了記憶體介面，而供應商已以超出該標準基準速度的規格運行。 記憶體標準組織JEDEC於2025年4月發布HBM4標準（JESD270-4），採2,048位元介面，每接腳最高8 Gb/s，每堆疊最高2 TB/s[S-139]。Samsung所列其HBM4E在單一16層堆疊中最高達64 GB與4 TB/s，每接腳最高16 Gb/s，並表示客戶驗證（qualification）時程尚未公開[S-140]。2026年7月，JEDEC發布SPHBM4，其採用與HBM4相同的DRAM晶粒，將HBM4級的頻寬帶到標準有機基板上[S-141]。

= 堆疊頻寬等於介面寬度乘以接腳速率，因此決定可用頻寬的是驗證（qualification），而非標準本身。 以GB/s計的頻寬等於介面位元數×以Gb/s計的接腳速率÷8。在JEDEC的8 Gb/s下，2,048×8÷8 = 2,048 GB/s，即該標準所稱的"up to 2 TB/s"；在Samsung同樣寬度下的16 Gb/s，2,048×16÷8 = 4,096 GB/s，即其HBM4E的4 TB/s[S-139; S-140]。因此，同一介面可依誰的零件在何種速度下通過驗證，而提供兩倍的頻寬。一具加速器（accelerator）的記憶體頻寬在供應商驗證時即已固定，這也是多供應商驗證既緩慢又有價值的原因之一。

KV快取

= 一段長對話即可占滿一個HBM堆疊的大部分容量。 每個token的KV快取大小為2×層數×KV頭數×頭維度×每元素位元組數，其中的2對應鍵與值。Meta的Llama 3 70B有80層與8個維度為128的KV頭，因此在16位元（2位元組）下，每個token儲存2×80×8×128×2 = 327,680位元組，而一段131,072個token（128K）的序列約占42.9 GB[S-142]。這相當於Samsung所列64 GB HBM4E堆疊的三分之二[S-140]，此外還須加上700億個參數在16位元下約140 GB的權重。若無分組查詢注意力（grouped-query attention，GQA）讓64個查詢頭共用8個KV頭，該快取將大上八倍。

A 需求正沿三個軸向同時延長該快取。 上下文長度使每段對話的快取倍增：MiniMax-M1支援100萬個token的原生上下文[S-143]。並行數使每具加速器（accelerator）上的快取倍增，每位作用中的使用者各占一份完整快取。代理式工作流程則延長每份快取必須存續的時間：InferCept研究發現，在工具呼叫之後重新計算上下文，占用了模型前向傳遞時間的37%至40%，而在工具執行期間保留快取即可避免此一成本[S-144]。推理模式，例如Alibaba Qwen3中的思考預算，則增加每次請求的token數[S-145]。每一軸向都使記憶體需求線性上升，而各軸向之間則是相乘的關係。

效率改變了什麼

F 開放模型與服務推論研究同時降低了每個token的運算量與快取量。 DeepSeek-V2以多頭潛在注意力（multi-head latent attention，MLA）將鍵與值壓縮為潛在向量，並報告其KV快取較稠密型DeepSeek 67B縮小93.3%[S-146]。DeepSeek-V3透過混合專家（mixture of experts，MoE）架構，為每個token啟用其6,710億個參數中的370億個；Moonshot AI的Kimi K2則在1兆個參數中啟用320億個[S-147; S-148]。服務推論研究透過分頁配置（vLLM）、2位元快取量化（quantisation）（KIVI）與汰除較少使用的token（H2O），進一步縮小該快取，各項結果均在其作者所選的模型與基準上衡量[S-149; S-150; S-151]。

F 較低精度的數值格式減少了每個權重的位元組數。 OCP Microscaling（MX）格式背後的多家企業共同論文，作者來自Microsoft、AMD、Intel、Meta、NVIDIA與Qualcomm，報告以低於8位元的權重、啟動值與梯度訓練生成式語言模型，而準確度損失極小[S-152]。NVIDIA報告其4位元NVFP4格式在DeepSeek-R1-0528的基準測試中，相較於8位元的FP8格式呈現"1% or less accuracy degradation versus FP8"[S-153]。因此，數值格式既是算術運算上的槓桿，也同樣是記憶體上的槓桿：每個權重的位元數愈少，承載模型所需的容量愈小，讀取模型所需的頻寬也愈小。

A 效率改變的是記憶體算術關係中的常數，而非其形態。 混合專家降低了每個token的算術運算量，卻未免除將每一位專家的權重都保持在可及範圍內的需要；

MLA、GQA與低位元快取降低每個token的位元組數，4位元格式降低每個權重的位元組數，然而只要多保留一個token、多服務一位使用者，快取就會繼續成長[S-147; S-146; S-153]。每項結果都是在其作者所選的模型與基準上衡量，因此這些增益屬已展示（demonstrated），而非已標準化（standardised），且不能單純相加。對規劃者而言，快取或格式上的縮減是一項與加購HBM相互競爭的容量槓桿。

HBM以下的記憶體層級

F 服務推論系統現已將上下文分散到成本較低的記憶體層級。 DistServe將受運算限制的提示階段與受記憶體限制的解碼（decode）階段置於不同硬體上執行，報告在滿足其中逾90%請求的延遲限制下，所服務的請求量最高可達7.4倍[S-154]。Moonshot AI旗下Kimi背後的服務架構Mooncake，將GPU叢集中未充分利用的CPU、DRAM與SSD容量用作解耦的KV快取，報告可多處理75%的請求[S-155]。NVIDIA的開源Dynamo將推論資料卸載至成本較低的記憶體與儲存裝置[S-156]，SK hynix則與SanDisk合作，正將一項快閃記憶體產品以High Bandwidth Flash（HBF）的名稱推向市場，並共同推動標準化[S-157]。

F Compute Express Link（CXL）以共享記憶體池提供了第二條路徑。 CXL Consortium於2025年11月發布CXL 4.0，將連結速率加倍至128 GT/s[S-158]。Microsoft的Pond研究發現，在8至16個插槽之間共享DRAM即可取得多數效益，可使DRAM成本降低7%，而效能與本機配置相差在1%至5%之內[S-159]。OCP的Composable Memory Systems（CMS）子專案將Meta與Google的「Hyperscale CXL Tiered Memory Expander」規格（specification）列為審查中的草案[S-160]；而一份CMS白皮書（white paper）則指出，AI垂直擴展工作負載"do not require the hardware-based cache-coherence or the low-latency features that CXL specification offers"[S-161]。

A 較低的層級增加的是承載上下文的空間，而非解碼（decode）所需的頻寬。 快取在其對話處於作用中時，每一個解碼步驟都會被讀取，而在輪次之間或工具呼叫之間則閒置，因此閒置的快取可下移至主機DRAM、共享記憶體池或快閃記憶體，並在使用前返回。上述系統無一將HBM自解碼路徑中移除[S-155; S-154; S-156]。因此，分層化把AI記憶體需求擴散到伺服器DRAM與企業級快閃記憶體，也就是Samsung預期需求成長將加速、而Kioxia正重新聚焦的市場，同時每位使用者的每秒token數仍由HBM決定[S-127; S-130]。

反彈效應的問題

A 效率與需求從相反兩端作用於同一組算術關係。 潛在注意力、低位元快取與4位元格式降低每個token的位元組數；混合專家（mixture of experts）降低每個token的運算次數；分頁、解耦與分層則提高每位元組HBM所能服務的token數。每一項增益都降低單一token的成本，使更長的上下文、更大的批次、推理模式與多步驟代理變得可負擔，而上述每一項又都提高記憶體中所保留的token數。2026年的紀錄符合反彈效應的樣貌：供應商陳述需求高於供給，TrendForce預期DRAM至2027年仍將緊俏，而這遠在主要效率技術發表之後[S-003; S-135; S-146]。

IF 效率唯有在超越上下文與使用者的成長速度時，才能紓解記憶體需求。 只要HBM在HBM4E擴產期間仍處於配額狀態，即使已發表的模型持續降低每個token的記憶體用量，反彈效應的研判（reading）即成立。應觀察新平台中每具加速器（accelerator）的HBM容量、TrendForce關於HBM配置的報告、潛在注意力快取與4位元格式在廣泛提供服務的模型中的採用情況，以及KV快取分層介面是否形成開放規格（specification）。若供應商報告庫存過剩或HBM價格下跌，而上下文長度仍持續上升，則此一研判不成立，這將意味效率已超越需求。

3.4 先進封裝決定了能夠建置多少運算能力

現代的AI加速器（accelerator）並非單一晶片，而是一個封裝體：數個邏輯晶粒與一圈HBM堆疊並列安裝在矽中介層或橋接結構上，由其承載彼此之間數千條連線。其中領先的形式是Taiwan Semiconductor Manufacturing Company（TSMC）的Chip-on-Wafer-on-Substrate（CoWoS）；依市場研究與媒體報導，該技術在2026年仍供不應求，因此能夠組裝的封裝體數量，限制了無論由誰設計的加速器出貨量。

封裝的時鐘也比晶片銷售來得慢。台積電（TSMC）的封裝藍圖已延伸至2029年，而每一次封裝體放大，都會使其所消耗的記憶體、基板與測試量倍增。本節說明台積電已陳述的內容、媒體報導與市場研究對產能的說法，以及開放小晶片（chiplet）介面能夠與不能夠改變什麼。

CoWoS產能

F TrendForce轉述的媒體報導預期CoWoS產能將持續上升，而缺口在2026年將縮小而非消除。 TrendForce於2026年6月報導，台積電（TSMC）每月CoWoS產能在2026年可達12萬至14萬片晶圓，供需缺口預期將由約20%縮小至年底的約10%，同時委外半導體封裝測試（outsourced semiconductor assembly and test，OSAT）業者每月再增加5萬至6萬片晶圓[S-033]。該機構9月報導指出，CoWoS產能可自2026年底每月約13萬片晶圓，於2028年底倍增至26萬片，重心在台積電位於台灣的AP7廠區與其亞利桑那園區[S-162]。兩者均為TrendForce News轉述的媒體報導，而非台積電的揭露或TrendForce自有的研究。

> **範圍 CoWoS產能**
> 產能的報價單位是每月晶圓片數，而非加速器（accelerator）數量。隨著封裝體朝14個光罩的尺寸發展，每片晶圓所能容納的封裝體數量減少，因此晶圓產能倍增並不代表加速器產出倍增[S-162; S-024]。

F TrendForce表示NVIDIA已提早鎖定供應，而CoWoS客戶正在尋求替代方案。 該機構描述AI競爭已轉為供應鏈之爭，NVIDIA為最早察覺產能趨緊者之一，先行鎖定大量3奈米與4奈米晶圓產能、CoWoS封裝、基板、T-glass、印刷電路板、HBM與SSD[S-017]。該機構並指出，CoWoS自2023年以來一直供應短缺，促使客戶尋求替代的產能來源[S-017]。該機構預期CoWoS-L至2028年仍將是AI晶片的主流封裝方式，並報導台積電封裝產能的持續受限，已提高業界對Intel嵌入式多晶粒互連橋接（Embedded Multi-die Interconnect Bridge，EMIB）作為替代方案的關注[S-036]。

台積電的藍圖與支出

F 台積電的藍圖將封裝轉化為一項多年期產能計畫。 2026年4月，台積電表示一種14個光罩尺寸的CoWoS，可整合約10個大型運算晶粒與20個HBM堆疊，預定於2028年投入生產，其後將於2029年擴展至14個光罩以上，另有一種40個光罩尺寸的System-on-Wafer（SoW-X）同樣預計於2029年推出[S-024]。其共同封裝光學（co-packaged optics，CPO）平台COUPE原定於2026年在基板上開始生產，而台積電於7月表示生產已經開始[S-024; S-084]。表4將這些日期與台積電的邏輯製程藍圖及其支出財測指引並列。

A 封裝體尺寸每提高一級，其內部的記憶體、基板與測試量都隨之倍增。 光罩是微影設備單次曝光所能圖案化的最大面積，因此一個14個光罩尺寸的封裝體，是在同一基板上接合相當於許多顆晶片的矽面積。在每個封裝體約20個HBM堆疊的情況下，封裝產能與HBM供給成為同一項限制：唯有在其所有堆疊都通過驗證（qualification）、完成鍵合與測試後，該封裝體才會出貨[S-024]。較大的封裝體提高了每個封裝體的產出，但也提高任一部分失效時所損失的價值，這正是封裝體愈大、基板、鍵合與測試的重要性愈高的原因（第3.5節）。

F 台積電在2026年內同時上調了其AI成長看法與資本支出。 1月，台積電將其2024年至2029年AI加速器（accelerator）營收成長預測上調至"approach a mid-to-high-fifties % CAGR"，即年複合成長率[S-082]；7月則未提出新數字，但表示AI成長率"stronger than what we said before"[S-084]。該公司將2026年資本預算自1月的520億至560億美元上調至7月的600億至640億美元，並表示目前預期未來三年的資本支出相較於前三年的1,010億美元將"even more significantly higher"，而1月時的說法為"significantly higher"[S-082; S-084]。

F 部分新增產能正設置於美國。 台積電採用3奈米技術的第二座亞利桑那晶圓廠，在該公司因客戶需求而將時程提前後，預定於2027年下半年開始量產[S-082; S-083]。1月，台積電表示正為第四座亞利桑那晶圓廠及其在當地的首座先進封裝（advanced packaging）廠申請許可[S-082]。7月，該公司宣布在亞利桑那州追加1,000億美元投資，美國商務部則指其投資總額達2,650億美元[S-084; S-163]。第7.5節討論此事對台灣集中度的影響。

A 台積電的計畫使封裝落後於季度需求訊號兩至三年。 資本預算可在一年之內上調，但決定每具加速器（accelerator）搭載多少HBM堆疊的封裝體尺寸，其日期落在2028年與2029年，而亞利桑那第二座晶圓廠也要到2027年下半年才達到量產[S-084; S-024; S-082]。加速器需求的季度時鐘與封裝產能的多年期時鐘之間的落差，正是配置權、以及隨之而來的定價權所在之處。台積電表示AI成長強於其1月所給的區間、卻未提出新數字，其指向亦相同。

小晶片與開放介面

F 開放小晶片（chiplet）規格（specification）現已描述一個封裝體如何在供應商之間分工。 Arm於2025年10月將其Foundation Chiplet System Architecture（FCSA）貢獻給OCP，1.0.0版的日期為2026年2月12日[S-164; S-165]。FCSA將小晶片的角色對應到既有介面上，其中包括Universal Chiplet Interconnect Express（UCIe）、Arm的AMBA CHI晶片對晶片協定、CXL與PCI Express，並將任何標準均未涵蓋的連接標示為"TO BE DEFINED"[S-166; S-165]。2025年8月發布的UCIe 3.0支援最高64 GT/s，而UALink Consortium的小晶片規格則被描述為與其完全相容[S-167; S-168]。

A HBM堆疊的底層正成為客戶可以形塑的邏輯晶片。 Samsung在4奈米邏輯基礎晶粒上建構其HBM4，並表示客製HBM樣品將於2027年"according to their respective specifications"交付客戶[S-140; S-056]。JEDEC的標準所固定的是記憶體與處理器之間的通道介面，而非堆疊下方的邏輯[S-058]。因此，設計自由度移入基礎晶粒，那裡是記憶體製造商、晶圓代工廠與加速器（accelerator）設計者交會之處，也是差異化能在開放介面之下存續之處（第6.7節）。

A 開放介面降低了取得第二家晶粒供應商的成本；它們並不創造封裝產能或HBM。 一份共通的對應關係圖（FCSA）與一條具公開速率階梯的晶粒間連結（UCIe），使在同一封裝體內驗證來自不同供應商的晶粒更為低廉，而JEDEC的SPHBM4則擴大了HBM可用基板的選擇[S-141]。但這些都不會增加封裝產線、鍵合設備或HBM晶圓。能否大量建置小晶片（chiplet）系統，決定於封裝產能與記憶體供給的持有者，而非規格（specification）組織，這也是OCP在此處的價值在於交接點而非產能的原因。

? 未解問題——CoWoS產能與未獲滿足的封裝需求，僅以媒體報導的數字形式公開。 上述產能與缺口數字來自TrendForce News轉述的媒體報導，而非出自台積電或TrendForce自有的研究[S-033; S-162]。加速器（accelerator）供應何時趨於寬鬆，以及隨之而來的定價權與配置權何時轉移，取決於沒有任何企業揭露的一項平衡關係。

IF 封裝唯有在產能的成長快於其所須建置之封裝體尺寸時，才會不再構成限制。 認為CoWoS限制加速器

表4。 台積電附日期的藍圖延伸至2029年，而其支出財測指引則在年內上調

支出依季度時鐘變動，但決定每具加速器（accelerator）HBM含量的封裝體尺寸，其日期落在2028年至2029年，因此封裝產能落後於需求的單位是年，而非季。

項目	日期	類別	台積電的用語	來源
基板上的COUPE共同封裝光學（co-packaged optics）	2026	R	"beginning production in 2026"；至7月生產已"started"	[S-024; S-084]
亞利桑那第2座晶圓廠（3 nm）量產	2027年下半年（2H）	C	建廠完成；"pulling forward the production schedule"	[S-082; S-083]
A14風險試產	2027	R	稱"Risk production will start in 2027"	[S-084]
A14量產；N2U投入生產	2028	R	A14稱"scheduled for 2028"；N2U稱"scheduled for production in 2028"	[S-084; S-024]
14個光罩尺寸的CoWoS，約10個運算晶粒與20個HBM堆疊	2028	R	稱"slated for production in 2028"	[S-024]
14個光罩以上的CoWoS；40個光罩的SoW-X	2029	R	擴展稱"in 2029"；SoW-X稱"also expected in 2029"	[S-024]
A13與A12量產；A14對A14的SoIC	2029	R	稱"scheduled for volume production in 2029"；SoIC稱"set to be available for production in 2029"	[S-084; S-024]
2026年資本預算	2026	G	520億至560億美元（1月）；"towards the high end"（4月）；600億至640億美元（7月）	[S-082; S-083; S-084]
2026年至2028年資本支出，對比2023年至2025年的1,010億美元	2026–2028	G	"significantly higher"（1月）；"even more significantly higher"（7月）	[S-082; S-084]
AI加速器（accelerator）營收成長	2024–2029	G	"approach a mid-to-high-fifties % CAGR"（1月）；"stronger than what we said before"，未提出新數字（7月）	[S-082; S-084]
亞利桑那州追加投資	2026年7月16日公告	C	規劃1,000億美元；依美國商務部所述，投資總額達2,650億美元	[S-084; S-163]

台積電與AI加速器（accelerator）相關的附日期藍圖項目、產能推進步驟與財測指引，依台積電的用語呈現，取自其2026年4月技術論壇的發布文件，以及2026年1月、4月與7月的法人說明會。類別：R為藍圖日期；C為產能承諾，並載明階段；G為財測指引或管理層預測。

範圍 藍圖日期為台積電的時程，仍須視執行情況而定。台積電的資本預算為其自有的衡量指標，與雲端服務供應商的資本支出（capital expenditure）不可比。

除數字歸屬其他來源者外，文字均為台積電自身的用語。

（accelerator）產出的看法，在封裝體尺寸與HBM堆疊數的上升速度與晶圓產能同步時成立。應觀察CoWoS供需缺口的下一個報導數字、台積電的資本預算及其為AP7與亞利桑那所載明的任何封裝產能、Intel EMIB與其他替代方案的採用情況，以及首批進入加速器封裝體的SPHBM4零件。若報導數字顯示缺口已彌平而2028年的封裝體仍如期推進，或替代方案在AI加速器封裝中取得實質份額，則此一研判（reading）不成立。

3.5 新基板、鍵合與測試顯示複雜度移向何處

隨著封裝體愈來愈大、記憶體堆疊愈來愈高，困難度移入多數讀者從未見過的環節：承載封裝體的基板、接合晶粒的鍵合，以及在把每一顆晶粒與每一個堆疊投入昂貴封裝體之前、證明其功能正常的測試。這些環節決定良率（yield），而良率決定有多少已規劃的產能能轉化為已出貨的運算能力。

這些環節的供應商與晶片製造商相比規模甚小，但其訂單簿與產能計畫是封裝複雜度走向的早期訊號。它們也是新進工業參與者——包括來自傳統半導體供應鏈之外的台灣企業——投注研發的所在。

F 友達光電（AUO）已自鄰接的顯示器業務，轉入直接載明的AI基礎設施技術定位。 2026年8月31日，友達宣布將與Corning在SEMICON Taiwan（2026年9月2日至4日）展出玻璃核心基板（glass-core substrate，GCS）封裝技術，結合Corning的半導體級玻璃與友達的重佈線層（redistribution layer，RDL）製程，以因應更大的AI伺服器處理器晶粒、高密度互連與低損耗訊號傳輸[S-052; S-169]。友達並表示，正與生態系夥伴共同開發一款系統級Micro LED共同封裝光學（co-packaged optics）模組，用於AI資料中心內最長10公尺的短距連結[S-052]（第4.7節）。其階段為研發：相關發布文件描述的是開發與展出，而非量產、客戶或營收。

A 問題如今在於可製造性與驗證（qualification），而非友達是否應被列入版圖。 玻璃核心是友達與Corning對更大封裝體所產生之需求的回應：更大的處理器晶粒、更密的互連與更低的訊號損耗[S-052]。友達帶來逾三十年來自顯示器領域的玻璃基板製程經驗，這是台灣工業基礎由組裝走向封裝與光學的一個例證（第5.7節）[S-052]。應觀察客戶驗證、量產規模、連結效能、每位元功耗、良率（yield）、封裝生態系的採用情況，以及是否進入具名的伺服器或網路平台。

F 混合鍵合（hybrid bonding）的客戶與應用案例正在增加，其中包括記憶體與光學。 荷蘭封裝設備製造商Besi報告，其混合鍵合客戶數自2025年底的15家增至2026年6月底的21家，應用案例在邏輯、記憶體、共同封裝光學（co-packaged optics）與消費性應用上均有增加[S-170]。該公司估計，AI應用在2026年上半年約占其系統訂單的60%，相較於一年前約為50%，並於第2季再自兩家回購客戶與一家新的超大規模雲端業者（hyperscaler）客戶取得混合鍵合訂單[S-170]。

F Besi以AI封裝需求為由上調其長期目標。 在2026年6月的投資人日上，Besi將其長期營收目標自15億至19億歐元上調至17億至22億歐元，營業利益率目標自40%至55%上調至45%至55%，但未載明目標年度[S-057]。該公司所給理由為2.5D AI資料中心與光子應用需求增加，以及混合鍵合（hybrid bonding）在邏輯、記憶體與共同封裝光學（co-packaged optics）上的新用途[S-057]。這些屬管理層目標（類別T），而非財測指引。

F 更高的HBM堆疊提高了每個封裝體中的鍵合數與測試數。 Tokyo Electron（TEL）表示，HBM中堆疊的DRAM晶粒數將自12顆增至16顆，而在先進封裝（advanced packaging）領域，其探針台銷售呈現"outstanding growth"，截至2027年3月的年度先進封裝全年銷售額"likely to go significantly beyond 100 billion yen"[S-171]。該公司在其商機項目中列出鍵合相關設備至2030年的累計銷售額為5,000億日圓，並表示供AI資料中心使用的晶圓廠設備正成長至超過總量的50%[S-172]。

F 測試產能正依封裝產能進行規劃。 Advantest表示，其加速器（accelerator）測試機的產能規劃與先進封裝（advanced packaging）產能的擴充緊密相連，並且還須吸收每一新世代晶片不斷上升的測試量[S-116]。該公司正提前其目標，即在2027年3月前達到5,000台SoC測試系統、2028年底或2029年初達到10,000台，並表示下一步"could be at a scale of 15,000 systems or more"；此係一項情境（scenario），屬類別S，而非計畫[S-116]。該公司將"advanced packaging capacity constraints and memory supply dynamics"列為業界所面對的外部不確定因素之一[S-031]。

F 微影產能亦在上升，2027年的計畫較2028年更為確定。 ASML計劃在2027年將其2026年的低數值孔徑極紫外光（extreme ultraviolet，EUV）產能增加30%，該產能約為65台；並將其約130台的深紫外光（deep ultraviolet，DUV）浸潤式產能增加30%；2028年再增加30%則仍在評估中，尚未納入計畫[S-025]。該公司表示，持續進行的AI相關投資正帶動先進邏輯與記憶體晶片的需求，目前預期2026年總淨銷售額為430億至450億歐元[S-025]。

A 設備製造商的計畫顯示複雜度累積於何處：層數、封裝體尺寸與良率（yield）。 更高的堆疊意味每個HBM堆疊中有更多鍵合；更大的封裝體意味在組裝之前須證明良好的晶粒與堆疊更多；兩者都提高每一次良率失效的成本。Besi的客戶數、TEL的探針台與鍵合展望，以及Advantest的測試機計畫，全都是對這些壓力的回應，而Advantest更明確地將其產能規模對應到封裝產能[S-170; S-171; S-116]。因此，它們的訂單簿是封裝與HBM擴產的早期指標。本節所討論的供應商分屬荷蘭與日本，提醒我們沒有任何單一區域掌握每一個層級（第5.7節）。

IF 在堆疊與封裝體持續變大的同時，複雜度將繼續移入基板、鍵合與測試。 此一看法在HBM堆疊自12顆晶粒增至16顆、且封裝體依台積電的時程朝14個光罩發展的情況下成立[S-171; S-024]。應觀察Besi報告中來自記憶體製造商的混合鍵合（hybrid bonding）訂單、TEL預定隨其截至2026年9月季度業績一併發布的全年預測[S-173]、Advantest的產能推進步驟，以及首個通過驗證並進入具名AI封裝體的玻璃核心基板（glass-core substrate）。若封裝藍圖延後，或HBM製造商在未新增鍵合與測試產能的情況下達到16層堆疊，則此一研判（reading）不成立。

讀者視角

本章對研發、智慧財產權、財務與產業的意義

研發。 尚不穩定的介面是HBM基礎晶粒（HBM base die）、16層堆疊、玻璃核心與KV快取層級；採用與否由良率（yield）與測試決定。

智慧財產權。 HBM4、UCIe與FCSA使介面趨於共通；差異化移入基礎晶粒邏輯、鍵合、測試方法與快取管理。

財務。 記憶體占雲端資本支出的比重持續上升；應將Samsung與Broadcom的案例視為情境（scenario），並將緯穎跨越2026年4月的營收視為已重設基礎。

產業。 HBM、CoWoS與測試機產能在數年前即已配置；能否轉換，取決於是否已與一家以上的記憶體供應商完成驗證（qualification）。

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-002 NVIDIA (26 Aug 2026); S-003 SK hynix (29 Jul 2026); S-010 NVIDIA (26 Aug 2026); S-016 Alphabet (22 Jul 2026); S-023 SK hynix (28 Aug 2026); S-024 TSMC (23 Apr 2026); S-025 ASML (15 Jul 2026); S-031 Advantest (29 Jul 2026); S-045 Microsoft (29 Apr 2026); S-052 AUO (31 Aug 2026); S-056 Samsung Electronics (12 Feb 2026); S-057 Besi (18 Jun 2026); S-063 Wiwynn (7 Aug 2026); S-064 Wiwynn (7 May 2026); S-075 Microsoft (29 Jul 2026); S-082 TSMC (15 Jan 2026); S-083 TSMC (16 Apr 2026); S-084 TSMC (16 Jul 2026); S-111 AMD Investor Relations (4 Aug 2026); S-112 AMD (24 Feb 2026); S-113 Broadcom (2 Sep 2026); S-115 Samsung and Broadcom (25 Jul 2026); S-116 Advantest (29 Jul 2026); S-122 SK hynix (24 Jul 2025); S-123 SK hynix (29 Oct 2025); S-124 SK hynix (28 Jan 2026); S-125 SK hynix (22

本章來源 （續）

Apr 2026); S-126 SK hynix (8 Jun 2026); S-127 Samsung Electronics (30 Jul 2026); S-128 Micron Technology (24 Jun 2026); S-129 Micron Technology (24 Jun 2026); S-130 KIOXIA (2 Jun 2026); S-131 KIOXIA (2 Jun 2026); S-134 Samsung Electronics (29 Jun 2026); S-140 Samsung Semiconductor (2026); S-153 NVIDIA (24 Jun 2025); S-156 NVIDIA (18 Mar 2025); S-157 SK hynix Newsroom (20 May 2026); S-164 Arm Newsroom (14 Oct 2025); S-169 AUO (31 Aug 2026); S-170 Besi (23 Jul 2026); S-171 Tokyo Electron (30 Jul 2026); S-172 Tokyo Electron (30 Jul 2026); S-173 Tokyo Electron (30 Jul 2026)

標準組織與聯盟. S-058 JEDEC Solid State Technology Association (Dec 2025); S-139 JEDEC (16 Apr 2025); S-141 JEDEC (13 Jul 2026); S-158 CXL Consortium (18 Nov 2025); S-160 OCP (22 Aug 2026); S-161 OCP (19 Dec 2025); S-165 OCP (12 Feb 2026); S-166 OCP (8 Apr 2026); S-167 UCIe Consortium (5 Aug 2025); S-168 Ultra Accelerator Link Consortium (7 Apr 2026)

政府與監管機關. S-163 U.S. Department of Commerce (16 Jul 2026)

研究機構. S-136 Gholami et al. (21 Mar 2024); S-137 Patel et al. (30 Nov 2023); S-138 arXiv (27 May 2022); S-142 Meta AI (31 Jul 2024); S-143 MiniMax (16 Jun 2025); S-144 arXiv (2 Feb 2024); S-145 arXiv (14 May 2025); S-146 DeepSeek-AI (7 May 2024); S-147 arXiv (27 Dec 2024); S-148 Moonshot AI (Kimi) (28 Jul 2025); S-149 arXiv (12 Sep 2023); S-150 arXiv (5 Feb 2024); S-151 arXiv (24 Jun 2023); S-152 Bita Darvish Rouhani et al., arXiv (Oct 2023); S-154 Zhong et al. (18 Jan 2024); S-155 Qin et al. (24 Jun 2024); S-159 Li et al. (1 Mar 2022)

市場研究. S-007 TrendForce (25 Aug 2026); S-008 TrendForce (20 Jan 2026); S-009 TrendForce (18 Mar 2026); S-013 TrendForce (3 Aug 2026); S-017 TrendForce (30 Apr 2026); S-035 TrendForce (1 Jun 2026); S-036 TrendForce (18 Sep 2026); S-114 TrendForce (20 May 2026); S-117 TrendForce (2 Sep 2025); S-118 TrendForce (7 Sep 2026); S-119 TrendForce (9 Sep 2026); S-120 TrendForce (26 Nov 2025); S-121 TrendForce (26 Feb 2026); S-132 TrendForce (9 Jul 2026); S-135 TrendForce (4 Aug 2026); S-602 SEMI (14 Jul 2026); S-603 SEMI (29 Jun 2026); S-604 SEMI (2026); S-605 TechInsights (24 Apr 2026)

媒體與法律評論. S-033 TrendForce (15 Jun 2026); S-070 StockAnalysis (30 Jul 2026); S-133 Investing.com via Yahoo Finance (5 Jun 2026); S-162 TrendForce (14 Sep 2026)

第4章

機櫃架構：電力、冷卻與網路

功率密度已使機櫃成為產品本身：電力轉換、液冷（liquid cooling）與網路現已一併設計，而尚未解決的介面則向外移至保護、認證（certification）、冷卻液處理，以及設施的電網併接。

適用對象： 產業與採購、研發、智慧財產權與政府部門；電力見第4.2節、冷卻見第4.3節、設施與電網見第4.4節；機櫃、網路架構與光學見第4.5至4.7節

加速器（accelerator）已不再單獨決定一套AI系統的上限。旗艦級機櫃如今的耗電達數百千瓦，而其電力路徑、冷卻液迴路與內部網路，已作為同一項產品一併設計。這使困難的問題向外推移：800 V如何配送與保護、液體迴路可以倚賴哪些流體與接頭，以及一整廳的百萬瓦級機櫃作為電網上的負載會有何表現。本章追蹤由機櫃向設施的轉移，並探問這些介面中哪些已經形成文字，哪些仍倚賴單一供應商的設計。

4.1 機櫃成為產品本身

過去，營運業者購買伺服器，電力與冷卻則交由建築物處理。這樣的分工對AI已不再成立。一具加速器（accelerator）如今最有用的形態，是作為機櫃級系統（rack-scale system）的一部分，其中數十顆晶片共用同一個垂直擴展網路（scale-up network），因此這些晶片必須彼此靠近；靠近使功率集中，而集中的功率又使熱量集中。一旦機櫃耗電達數百千瓦，其供電、冷卻與監控方式便在機櫃設計之時即已固定，而非在安裝之時。

這項轉移觸及本章的每一位讀者。買方現在連同機櫃一併購買電力轉換與冷卻容量；工程師承接的是冷卻液接頭與電源連接器等介面，而這些介面決定哪些供應商可以混搭；規劃者則發現建築物的電力引入、冷卻液廠務與電網併接，都成為產品規格（specification）的一部分。接下來三節依序討論電力、熱與設施；第4.5節至第4.7節則涵蓋機櫃、網路架構與光學。

F 加速器（accelerator）與機櫃功率在兩個產品世代之內即已躍升。 NVIDIA為其Blackwell圖形處理器（graphics processing unit，GPU）給出最高1,200 W的功率，Blackwell Ultra則為最高1,400 W[S-174]。在機櫃層級，NVIDIA於2026年8月19日的DSX設施設計指南，將機櫃熱設計功率（thermal design power，TDP）自MGX Gen 1.1的198 kW提高至Vera Rubin NVL72的330 kW[S-175]，相較於Meta的Open Rack V3（ORv3）High Power Rack在2024年10月所支援的140 kW[S-176]。Google於2025年4月表示，在2030年之前，機器學習每個資訊科技（information technology，IT）機櫃將需要超過500 kW[S-177]。

F 旗艦平台是以整座機櫃為單位設計，並採用液冷機箱。 NVIDIA於2026年8月26日報告，其Vera Rubin平台正"ramping into full production"，機櫃已在CoreWeave、Google Cloud、Microsoft Azure、Oracle Cloud Infrastructure與Nebius運行[S-010]；該公司將其計劃貢獻給Open Compute Project（OCP）的Vera Rubin NVL72運算機箱描述為100%液冷[S-178]。AMD的Helios參考設計（reference design）在一座依Open Rack Wide標準建置的液冷機櫃中配置72顆MI455X GPU，每顆搭載432 GB的第四代高頻寬記憶體（high-bandwidth memory，HBM4），AMD並預期2026年下半年進入量產部署[S-179]。

A 記憶體與網路把電力與熱綁在機櫃上，這正是三者如今一併設計的原因。 如Helios這樣的機櫃搭載31 TB的HBM4[S-179]，與加速器（accelerator）封裝在一起（第3.4節），因此其熱量也經由同一組冷板（cold plate）帶出。垂直擴展網路架構偏好短距電氣連結，因此共用同一網路架構的加速器會被盡可能壓縮在最少的機櫃內（第4.6節）。耦合愈密，機櫃功率愈高；功率愈高，則必須提高配送電壓並採用液冷（liquid cooling）；而兩者又共同決定了機櫃列與建築物必須提供什麼。

A OCP的工作範圍隨著限制條件而擴大，由機櫃延伸至設施及其電網併接。 ORv3於2022年固定了一款機櫃及其48 V電源架（power shelf）[S-180]。2025年10月啟動的Open Data Center for AI倡議，一開始即納入機櫃以外的貢獻文件（contribution）——Diablo 400電源機櫃（power rack）與Google的Project Deschutes冷卻液分配單元（coolant distribution unit，CDU）——並與美國暖氣、製冷與空調工程師學會（American Society of Heating, Refrigerating and Air-Conditioning Engineers，ASHRAE）結成聯盟[S-181; S-182]。第4.4節說明標準化單元（unit of standardisation）為何移轉。

> **重要性何在**
> 選擇一個加速器（accelerator）平台，如今就使買方承諾一套電力與冷卻架構：一座為48 V氣冷機櫃建造的機房，在能夠容納下一世代之前，可能需要新的配電系統與冷卻液迴路[S-183; S-175]。

4.2 自48 V至±400 V與800 VDC

配電是AI機櫃中最不顯眼的部分，也是日後最難改變的部分之一。匯流排（busbar）電壓決定機櫃需要多少銅、電力轉換要從運算空間占去多少位置，以及適用哪些保護裝置、連接器與安全法規（safety code），而這些選擇在硬體到場數年之前，就已內建於機房與供應合約之中。

目前有三條路徑並存。48 V等級的ORv3機櫃是已安裝的基準。±400伏特直流（volts direct current，VDC）路徑將電力轉換移入IT機櫃旁的電源機櫃（power rack），並已載入一份OCP規格（specification）[S-027]。原生800 VDC路徑由NVIDIA於2025年5月提出[S-028]，目前與Google及Microsoft共同推進，將800 V直接送入機櫃本身；逾80家夥伴正開發相容的基礎設施，但OCP尚未為其訂定規格[S-019]。

F ORv3基準是由機櫃電源架（power shelf）供電的48 V等級匯流排（busbar）。 ORv3基礎規格（specification）1.0版（2022年9月22日，Meta與Google作者）訂定標稱51 VDC的電源架輸出，要求IT設備在46 V至52 V之間運作，並為每個IT設備輸入連接器的電力與接地路徑訂出"100A continuous"的最低額定值[S-180]；其備用電池架容納六個鋰離子單元，採5+1冗餘配置[S-184]。NVIDIA表示，為千瓦級機櫃設計的傳統54 V機櫃內配電"isn't designed to support"百萬瓦級機櫃[S-028]。

= 在330 kW的機櫃上，800 V配電所承載的電流僅為48 V所需的6%。 其規則為I = P/V。以NVIDIA設計指南中330 kW的Vera Rubin NVL72機櫃為例[S-175]，在48 V下電流為6,875 A，在54 V下為6,111 A[S-028]，在800 V極對極下則為412.5 A，後者同時涵蓋0–800 V與±400 V兩種供電方式。電流下降94%（1 − 48/800）；在導體相同的情況下，電阻損耗（I^2R）降至其48 V數值的$(48/800)^2 = 0.36\%$。在1 MW、即Diablo 400電源機櫃（power rack）的規模下[S-027]，該規則得出48 V下為20,833 A，800 V下為1,250 A。

A 已公布的增益是系統層級的數字，遠低於單一導體的算術結果。 NVIDIA表示，較低的電流可減少"copper requirements by 45%"，相同導體線徑所承載的功率比在415伏特交流（volts alternating current，VAC）下多85%，而端到端效率相較於54 V最多可提升5%[S-028]。此一落差是可預期的：匯流排（busbar）與連接器的尺寸還須考量機械強度、短路耐受能力與絕緣，415 VAC的起點本就遠高於48 V，而大部分轉換損耗仍留在未被取代的各級電路中。決定性的增益在於機櫃內外的銅用量與空間。

F ±400 VDC路徑是OCP已訂定規格（specification）的路徑。 Microsoft與Meta於2024年10月15日宣布Mt Diablo[S-185]，Google則於2025年4月加入，提出每座機櫃最高1 MW的±400 VDC供電[S-177]。由這三家公司共同撰寫、2026年3月1日生效的「Diablo 400 Project: Rack and Power」基礎規格0.7.0版，描述一款可"from 800kW to 1MW+"擴展的側置電源機櫃（sidecar）[S-027]。Google、Microsoft與NVIDIA於2026年8月11日發表於OCP部落格的聯合文章補充指出，"the published Mt Diablo +/-400V sidecar spec enables current solutions, while a Mt Diablo 2.0 will be coming later to support native 800V"[S-019]。

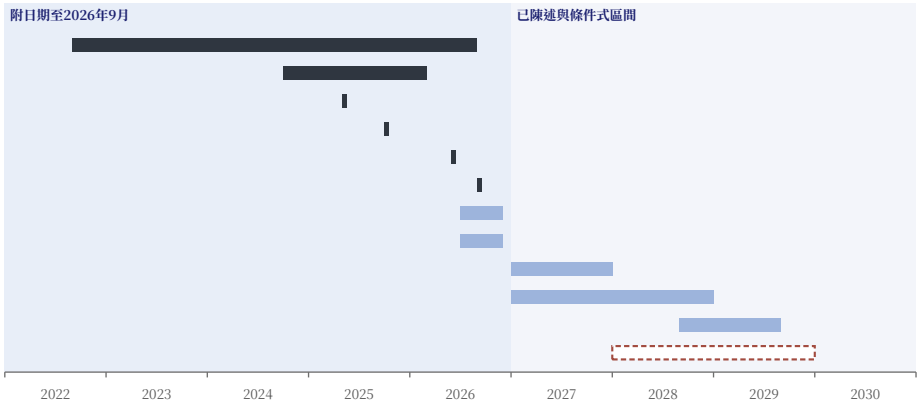
F NVIDIA的800 VDC時程表最為具體；獨立的研判（reading）則認為廣泛採用會更晚。 NVIDIA於2025年5月表示，"full-scale production of 800 VDC data centers will coincide with NVIDIA Kyber rack-scale systems in 2027"[S-028]。市場研究機構TrendForce預測，NVIDIA的800 V電源機櫃（power rack）將自2026年第3季起成為Vera Rubin的選配組態，並於2028年普及[S-029]。Tom's Hardware引述SemiAnalysis，於2026年7月報導Kyber機櫃已延後至2028年；NVIDIA則回應其藍圖並未改變[S-186]。

F 2026年8月，Google、Microsoft與NVIDIA將800 VDC重新定位為一組共通介面。 三家公司的OCP聯合文章報告了一項中壓交流（medium-voltage alternating current，AC）至800 VDC的工作流（workstream），公開了固態變壓器（solid-state transformer，SST）規格（specification）v0.3，並表示三家公司正同步與UL Solutions、美國國家防火協會（National Fire Protection Association，NFPA）、電機電子工程師學會（Institute of Electrical and Electronics Engineers，IEEE）及國際電工委員會（International Electrotechnical Commission，IEC）就安全認證（certification）展開合作[S-019]。該文章允許以側置電源機櫃（power rack）將既有的480 VAC轉換為±400或0–800 VDC，長期則由整流器或SST整裝機組將中壓直接轉換為800 VDC主幹，並載明"800 VDC is not a replacement for existing AC infrastructure"[S-019]。

> "The goal is not a single prescriptive design. The goal is a common set of interfaces and system requirements that lets operators deploy safely, suppliers build interoperable products and the industry move faster together." [S-019]

A 這篇聯合文章把競爭移到共通介面之後。 由於側置機櫃處允許兩種電壓配置，機房層級則可採整流器或SST，互通性必須落在機櫃電力輸入、連接器、保護介面、電力品質限值與儲能耦合上，而供應商則在這些介面之後的部分相互競爭[S-019]。依共通介面設計的供應商可以賣給多個買方，而買方也可以驗證多家供應商，這降低了轉換成本（switching cost）。若OCP

圖17。 ±400 VDC路徑已有規格；原生800 VDC則取決於已公告的時間區間與條件



機櫃配電的規格（specification）、認證（certification）與產品里程碑，2022年至2030年，來源有載明月份者即標至月。深色橫條為附日期的紀錄；淺色橫條為企業所陳述或TrendForce所預測的時間區間；虛線橫條為本報告的條件式區間。來源：[S-180; S-185; S-027; S-028; S-187; S-054; S-188; S-029; S-189; S-190; S-186; S-191]。

範圍 企業所給的時間區間是藍圖陳述，而非出貨紀錄，而TrendForce的橫條是市場研究預測。針對原生800 V的Mt Diablo 2.0已公告但未附日期，故未繪製[S-019]。

該條件式區間是將NVIDIA關於2027年的陳述、TrendForce關於2028年的預測、ABB兩至三年的預期，以及經報導的Kyber延後一併研讀的結果；它僅在本節末尾IF段落所列的條件（conditions）之下成立。

的後續修訂版固定了某一種電壓配置或轉換拓撲，則此一研判（reading）將趨於薄弱（A-03）。

供應商生態系

F 功率半導體製造商正在量產之前即為800 VDC進行開發。 NVIDIA為該架構列出十家矽晶片夥伴[S-028]，並表示逾80家企業正依其透過OCP與Google及Microsoft共同制定的800 VDC規格（specification）開發產品[S-190]。Infineon於2026年5月加入NVIDIA的MGX生態系：其產品將把800 V轉換為50 V、12 V或6 V，並在原生800 V伺服器板卡上，以其碳化矽（silicon carbide，SiC）接面場效電晶體（junction field-effect transistor，JFET）提供保護與熱插拔功能，未給出出貨時程[S-192]。

F Infineon已上調其AI電力業務營收展望，並正接受多年期產能預訂。 該公司預期2026財年（即截至2026年9月30日的年度）專用於AI電力的營收為"more than 1.6 billion Euros"，高於原先規劃的15億歐元，並將於11月上調其2027財年AI資料中心營收"2.5 billion Euros plus"的預估，且預期該次上調為"material"[S-193]。該公司表示，AI資料中心生態系中數家主要客戶已簽署或正在洽談多年期產能預訂協議，累計金額達高個位數十億歐元，並附帶部分預付款[S-021; S-193]。

F 系統供應商已公告800 VDC產品，多數未附交貨日期。 Eaton發表一套搭配超級電容進行快速循環備援的800 VDC參考架構[S-194]；Schneider Electric正與NVIDIA共同開發適用於最高1.2 MW機櫃的800 VDC側置電源機櫃（sidecar），未附上市日期[S-195]；Vertiv則將其800 VDC產品組合規劃於2026年下半年推出，並提及位於台灣高雄的一處驗證場址[S-189; S-196]。鴻海（Foxconn）位於高雄、40 MW的Kaohsiung-1場址，於2025年10月被描述為依800 VDC建置[S-178]；所引述的來源並未記載截至2026年9月有任何800 VDC機房投入運轉。

保護、法規與認證

保護是直流最困難之處。交流每一週期兩次通過零點，機械式斷路器的電弧即在該零點熄滅。直流（direct current，DC）沒有零點，因此直流斷路器必須自行迫使電流下降，手段是高於系統電壓的電弧電壓、反向電流或半導體切換。在較高電壓下，故障電流上升得更快，匯流排電容也會對故障點放電，因此保護每多花一微秒，所釋放的能量就愈大。而協調配合，也就是決定自整流器到機箱之間哪一個裝置先行動作，其重要性不亞於任何單一斷路器。

F 固態斷路器正以產品與開發計畫兩種形式問世。 ABB將其SACE Infinitus固態斷路器（solid-state circuit breaker，SSCB）描述為"the world's first IEC-certified solid-state circuit breaker"，斷路時間為25 µs[S-197; S-198]；2026年9月21日，該公司推出Infinitus自電源端到機櫃的800 VDC產品組合，並表示預期首批完整的DC原生設施將在兩至三年內出現[S-191]。2026年9月9日，SolarEdge與Infineon將其合作延伸至適用於800 VDC、且"in line with leading industry framework"的SSCB，由SolarEdge主導設計，Infineon供應SiC JFET[S-055]。

A SolarEdge與Infineon的斷路器是一項開發合作（研發），而非產品。 該發布文件表示SSCB"are designed to interrupt faults within a few microseconds"，未給出上市、量產或展示日期，也未將該項工作與任何具名的加速器（accelerator）平台相連[S-055]。ABB的25 µs與OCP聯合文章所稱的"solid-state breakers that clear faults in milliseconds"衡量的是不同事物，前者是裝置的斷路時間，後者是系統的故障清除時間；一份採購規格（specification）需要兩者兼備[S-198; S-019]。

F 法規與認證正分段逐步補上。 UL 489I為固態斷路器（solid-state circuit breaker）標準的第一版，於2025年10月22日發布，涵蓋額定至1,000 VAC與1,500 VDC、依《國家電氣法規》（National Electrical Code, NFPA 70）提供保護的固態與混合式斷路器及開關[S-187]。UL Solutions於2026年9月24日表示，已依其中壓電源研究大綱UL 2877，為可將中壓交流直接轉換為800 VDC的固態機組推出認證（certification）計畫[S-188]。在OCP的低壓直流（low-voltage direct current, LVDC）工作中，UL Solutions主導法規與標準工作流（workstream），ABB則主導保護與安全[S-199]。

? 未解問題——一座800 VDC機房可依循哪一套已認證的保護與協調配合方案來建置。 至1,500 VDC的固態斷路器（solid-state circuit breaker）已有標準，中壓轉換器的認證（certification）計畫也剛啟動，但所引來源並未記錄任何已認證的端到端800 VDC機櫃系統，亦無已公布的協調配合方案[S-187; S-188; S-059]。在此之前，每一個案場都須與其具管轄權之主管機關及承保公司逐案議定保護方案，這會拉長時程，並有利於已持有認證的供應商（第7.4節）。

A 靠近負載側的儲能正成為電力路徑的一部分。 AI訓練使加速器（accelerator）的取用功率出現波動；依ServeTheHome的報導，NVIDIA的GB300 NVL72以一項工作負載技術與更高的電源電容來抑制這類波動[S-200]。OCP的Energy子專案將AI工作負載暫態響應、緩衝與電力品質列入其儲能工作的考量項目[S-201]；Eaton與Schneider Electric的800 VDC設計則加入超級電容或模組化儲能，用於備援與負載平滑（load smoothing）[S-194; S-195]。

IF 800 VDC唯有在硬體、保護與認證（certification）同步推進時，才會擴展至參考設計（reference design）之外。 這需要一份針對機櫃原生800 V的OCP文件、來自一家以上供應商且附已公布協調配合數據的固態斷路器（solid-state circuit breaker），以及已認證的中壓轉換器。應觀察聖荷西峰會（2026年10月12日至15日）是否出現Mt Diablo 2.0草案或0.3.0版之後的SST修訂版、首批依UL 2877取得認證的產品、Infineon於2026年11月的財測指引，以及Kyber級系統的時程[S-019; S-188; S-193; S-186]。若2027至2028年的機房仍維持交流饋電，並在側置電源機櫃（sidecar）之後採用48 V或54 V機櫃，則此一觀點不成立。

4.3 液體進入機櫃

冷卻過去是機房的屬性，其規模由建築物的機械設備決定。在第4.1節所述的功率水準下，熱量必須由液體自每一顆封裝帶走，並以封閉迴路攜出機櫃，因此迴路的零件、流體與維修作業也成為買方規格（specification）所涵蓋的內容。問題已從液體是否可行——機隊數據如今已給出答案——轉向在機櫃的整個使用年限中，由誰負責加注、認證（certification）與維護該流體。

其物理原理很簡單。功率加倍而接觸面積不變的封裝，需要熱阻更低的冷卻液路徑，而空氣最先觸及其極限。自矽晶片至設施水的每一段，都會耗用固定溫差預算的一部分：冷板（cold plate）上的溫升、機櫃內的溫升，以及冷卻液分配單元（coolant distribution unit, CDU）的趨近溫度（approach temperature），亦即其所供應冷卻液與其所接收設施水之間的溫差。趨近溫度愈小，在相同晶片溫度下即可使用溫度較高的設施水。

F 晶片已跨過OCP自身寫下的門檻。 2023年9月的OCP Accelerator Module（OAM）規格（specification）r2.0版建議，僅熱設計功率（thermal design power, TDP）在600 W以下的模組採用氣冷，超過者則採用液冷（liquid cooling）[S-183]；Blackwell Ultra最高1,400 W約為該水準的2.3倍[S-174]。實際運轉走在規格之前：Google表示其自2018年起已在逾2,000座張量處理單元（Tensor Processing Unit, TPU）Pod上以吉瓦規模採用液冷，正常運行率約99.999%，而其Deschutes冷卻液分配單元（coolant distribution unit, CDU）自2020年起亦達到相近的機隊整體可用率[S-177]。

F TrendForce預期自2026年起多數AI晶片將採用液冷（liquid cooling）。 TrendForce在2026年8月17日的一份新聞稿中預估，AI晶片的液冷滲透率將自2025年約33%升至2026年的53%，並於2027年趨近60%[S-006]（圖18）；其分母為AI晶片，而非伺服器或資料中心。該機構將此一轉變歸因於晶片熱設計功率（thermal design power, TDP）超過1 kW，以及NVIDIA機櫃級系統（rack-scale system）的出貨量（其預期該出貨量在2026年將成長一倍），並表示Google的AI伺服器已有逾80%採用液冷[S-006]。

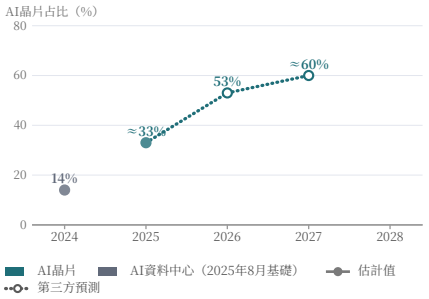
表5。 ±400 VDC路徑已有規格、800 VDC硬體已公告，但認證才剛起步

唯有±400 VDC側置電源機櫃（sidecar）可被引為已發布的OCP規格（specification）；機櫃原生800 V、中壓轉換器與直流保護，目前仍僅倚賴公告、一份初版草案，以及新開放的認證（certification）途徑。

環節	文件或規格	產品	認證	階段	來源
48 V機櫃匯流排（busbar）（ORv3）	基礎規格（specification）1.0版，2022年9月22日；1.1版列於2023年12月	Meta High Power Rack，最高140 kW（2024年10月）	尚無記錄	標準	[S-180; S-202; S-176]
±400 VDC側置電源機櫃（sidecar）（Diablo 400）	基礎規格（specification）0.7.0版，2026年3月1日生效	尚無記錄	尚無記錄	標準	[S-027]
800 VDC側置電源機櫃（power rack）	OCP聯合文章，2026年8月：480 VAC至±400或0–800 VDC	Vertiv產品組合計畫於2026年下半年（2H）推出；Schneider 1.2 MW側置電源機櫃（sidecar），未註明日期；TrendForce：自2026年第3季起為選配（預測）	尚無記錄	已宣布	[S-019; S-189; S-195; S-029]
機櫃原生800 V	Mt Diablo 2.0「coming later」，未註明日期	NVIDIA：2027年隨Kyber進入量產；有報導指出延後至2028年	尚無記錄	已宣布	[S-019; S-028; S-186]
機櫃列電力中心	無記錄	每列最高2 MW；預期2027年可供應（NVIDIA）	尚無記錄	已宣布	[S-190]
中壓至800 VDC（SST）	SST規格（specification）0.3.0版，2026年6月22日生效，為初版草案	Amperesand：計畫於2026年交付30 MW商用系統；Heron Power：計畫於2026年夏季推出原型機	UL 2877計畫於2026年9月24日啟動	試點階段	[S-054; S-203; S-204; S-188]
固態斷路器（solid-state circuit breaker）（ABB）	UL 489I涵蓋至1,500 VDC的SSCB	SACE Infinitus：2,500 A、1,250 VDC、25 μs；800 VDC產品組合，2026年9月	依ABB所述已取得IEC認證（certification）	產品	[S-197; S-198; S-191; S-187]
固態斷路器（solid-state circuit breaker）（SolarEdge–Infineon）	就800 VDC而言「In line with leading industry framework」	開發合作；未載明可供應日期	未載明	研發	[S-055]
800 V板上保護與轉換（Infineon）	支援NVIDIA的MGX 800 VDC架構	SiC JFET熱插拔；800 V降至50、12或6 V；未載明時程	尚無記錄	已宣布	[S-192]
電池儲能（BESS）	OCP儲能白皮書1.0版，2026年1月；NVIDIA BESS指引v1.0	來自Hitachi Energy、LG Energy Solution與Tesla的DSX Ready BESS	NFPA 855（2026年版）大規模火災測試	產品	[S-205; S-206; S-048]

截至2026年9月27日機櫃電力路徑各環節的狀態，取自規格（specification）、企業新聞稿與認證（certification）機構；階段標籤之定義見第B.1節。
「尚無記錄」意指所引來源未就該環節載有任何產品或認證（certification）。階段標籤給出的是該環節在公開紀錄上最進階的狀態，而非其出貨規模。

圖18。 液冷正成為AI晶片的預設配置



AI晶片的液冷（liquid cooling）滲透率，以%計，TrendForce：2025年為估計值，2026年與2027年為預測值（「approach 60%」），發布於2026年8月17日[S-006]。2024年的數值14%，為TrendForce在其2025年8月21日新聞稿中AI資料中心的滲透率，單獨繪製[S-591]。
範圍 2025–27年數列的分母為TrendForce所界定的AI晶片，而非全部伺服器或全部資料中心；2024年的數值則以資料中心為基礎，故兩者不相連接，亦不跨兩者計算任何變動率。

A 液冷（liquid cooling）如今已是旗艦系統的預設配置，而混合並存的機隊仍屬常態。 OAM門檻、Google的機隊紀錄，以及NVIDIA全液冷的Vera Rubin運算機箱，都指向同一方向：新的旗艦機櫃自設計之初即為濕式，冷卻能力隨機櫃一併購入[S-183; S-177; S-178]。即便以TrendForce預測的53%占比計算，2026年仍有近半數AI晶片以氣冷散熱[S-006]。對研發而言，自晶粒至冷卻液的熱阻已成為一項產品參數；對設施而言，液體迴路如今已是新建機房基本機電服務的一部分。

F 設備製造商的營收從供給端說出同樣的故事。 Dell'Oro Group預測，至2030年散熱管理仍將是資料中心實體基礎設施中成長最快的區隔，「with liquid cooling the fastest-growing technology as rack densification moves the technology from an option to a precondition」[S-594]。營運業者方面的公開證據較為薄弱：Uptime Institute 2026年針對1,034位受訪者所作的冷卻調查，聚焦於直接液冷（direct liquid cooling）的採用，但其公開頁面並未給出結果占比，故本報告不採用任何營運業者自行回報的占比[S-606]。

CDU成為買方的要求項目

F Project Deschutes使冷卻液分配單元（coolant distribution unit, CDU）成為一項已公布的需求。 Google的「Project Deschutes: Data Center Facilities」規格（specification）1.0.0版於2026年2月26日生效，「defines the requirements for the 2MW Next-Generation Cooling Distribution Unit (CDU)」，目標為在80至90磅／平方英寸（psi）的可用壓差下達到每分鐘500美制加侖（GPM），並以去離子水或體積濃度約25%的丙二醇（PG25）作為冷卻液[S-030; S-207]。該規格允許液體溫度為18至55 °C，並就IT側訂定25至44 μm的過濾精度，另設500 μm一次濾網與0.2 μm旁流過濾[S-030]。

= 在Deschutes的目標值下，水在通過負載後升溫約15 K；若改以空氣承載，所需體積約為其3,500倍。 移除的熱量等於體積流量×體積熱容×溫升。若由完整的500 GPM承載2 MW的目標值[S-030; S-207]，流量為0.0315 m³/s，而水的體積熱容約為每m³每K 4,180 kJ，則溫升為2,000 ÷（0.0315 × 4,180）≈ 15 K。以空氣在相同溫升下承載同樣的2 MW，其體積熱容約為每m³每K 1.2 kJ，需要約110 m³/s，約為水流量的3,500倍。使百萬瓦級負載以液體形式離開機櫃的原因是這個比值，而非偏好。

F CDU的供給正同時擴大與整併。 Eaton列有一款CDU，稱其「Inspired by OCP's fifth generation Project Deschutes CDU specification」，並標示在3 °C的趨近溫度（approach temperature）下達到2 MW，但未取得OCP Accepted標章[S-208]。NVIDIA的DSX Ready計畫於2026年9月21日啟動，啟動時即完成LG Electronics、LiquidStack與Vertiv的CDU驗證（qualification）[S-048]；LG表示其2.6 MW CDU符合DSX參考設計（reference design）的要求[S-209]。市場研究機構Dell'Oro於2025年9月統計約有40家企業的產品組合中納入CDU，並預期至本十年結束時將由不到10家企業掌握大部分市場[S-210]。

A 已公布的CDU需求使買方得以驗證（qualification）多家供應商，但如今有兩條驗證途徑並行。 Deschutes為營運業者提供了一個共同的採購目標值，在供應商基礎整併之際降低了轉換成本（switching cost）——例如Eaton於2026年3月完成收購Boyd的散熱事業，該事業即提供一款以Deschutes為藍本的CDU[S-211; S-208]。該規格（specification）並未就符合性作出規定，因此買方須自行查證效能主張[S-208]。NVIDIA的DSX Ready在開放規格之外，另加上一道平台擁有者的驗證[S-048]；若其中一條途徑成為供應的前提條件，則對CDU製造商的議價能力（bargaining power）將轉移至執行該驗證的一方。

接頭、流體與ASHRAE的分界

F OCP對迴路接合點的標準化程度，高於對迴路本身。 Cold Plate子專案涵蓋技術冷卻系統（technology cooling system, TCS），即自冷板（cold plate）通往CDU的封閉迴路[S-214]。其接頭年份較早但相當穩定：Universal Quick Disconnect（UQD）規格（specification）1.0修訂版（2020年9月4日，列為1.0.2版）訂定UQD02至UQD08各尺寸，最大運作壓力為100 psi，溫度範圍為17至65 °C；盲插版本（UQDB）則另加1.0 mm的徑向自動對位公差[S-212; S-215]。涵蓋兩者的UQD第2版為由CPC與Etched主導的beta草案[S-060]。

F 多數產品列表所引用的二醇類流體僅倚賴一份指引；唯有水基冷卻液有正式規格（specification）。 OCP的丙二醇文件是一份最終版但未註明日期的指引，作者來自Dow、Third Coast Chemicals、Intel與Meta：該文件將PG25定義為體積濃度24.5%至29.5%的丙二醇，載明其冰點為−13至−9 °C，並建議就10%的流量施以優於5 μm的旁流過濾[S-216]。由ChemTreat主導、2026年8月6日生效的Base Specification for Single Phase Water-Based Cold Plate Coolants 1.3版，則訂定pH值8.0至10.5、導電度低於2,000 μS/cm[S-217]。

F ASHRAE訂定這一分界的設施側。 其設施水等級於2021年改名為W17、W27、W32、W40、W45與W+，各名稱中的數字即為該等級的水溫上限（以°C計）[S-213]。OCP與ASHRAE於2025年10月13日宣布結成聯盟，透過ASHRAE第9.9技術委員會（Technical Committee, TC）推動，「focused on data center liquid cooling technology standards and best practices」，但未訂有註明日期的交付項目[S-182]。該委員會頁面在其新出版品中列有一份技術公報，內容關於TCS冷卻液完整性與系統就緒最佳實務[S-218]。

A OCP把接合點寫成文字——多供應商零件最先在此處無法對接——並讓迴路設計保持競爭。 Deschutes為IT迴路提供80至90 psi的壓差，並訂定0至130 psig的運作壓力範圍，而UQD接頭的最大運作壓力額定值為100 psi，因此接頭處的運作壓力必須維持在其額定

圖19。 從1,400 W晶片到2 MW CDU，液體散熱路徑的每一段如今都有已公布的規格範圍

液冷（liquid cooling）AI機櫃的散熱路徑，並列出各段的已公布數值：加速器（accelerator）最大功率[S-174]；NVIDIA DSX設計指南中的機櫃熱設計功率（thermal design power, TDP）[S-175]；接頭額定值[S-212]；CDU容量與流量[S-030; S-207]；ASHRAE設施水等級[S-213]。第二列列出規範各段的文件。
標示強調的方塊為適用OCP文件的接合點。在Deschutes的目標值下，水在通過負載後升溫約15 K，此為本報告於本節所示的計算結果。

值之內，此為整合時須查證的一項限制[S-030; S-212]。CDU正位於ASHRAE水等級與OCP接頭之間的分界上，因此該聯盟在此處最為重要[S-213; S-182]。冷板（cold plate）設計則仍保留創新空間：Microsoft表示其與Corintis共同開發的晶片內微流體冷卻（microfluidic cooling），在實驗室規模的測試中散熱效果最高達冷板的三倍[S-219]。

兩相、浸沒式與氣冷

F 在OCP內，兩相冷卻（two-phase cooling）仍處於指引與展示階段，浸沒式冷卻（immersion cooling）則僅倚賴一份2023年的需求文件，而工業集團已相繼入股。 Cold Plate專案持有一份關於兩相冷板（cold plate）介電流體的指引（2024年7月）[S-060]；ZutaCore則列於2026年峰會議程，將在一台運行中的Blackwell伺服器上展示晶粒直接流動沸騰[S-220]。OCP的核心浸沒式文件為2023年8月18日的Immersion Requirements 2.10修訂版[S-221]。Carrier於2026年4月對ZutaCore進行後續投資[S-222]；依Data Center Frontier的報導，Trane Technologies於2026年2月同意收購LiquidStack[S-223]。第7.4節討論流體安全與核准。

F 氣冷仍存在於現有產品線與既有機房中。 AMD的MI350系列涵蓋「standard air cooled UBB-based servers to ultra dense Direct Liquid Cooled (DLC) platforms」[S-224]。Google的Brazos（2026年6月16日）是一套供既有氣冷機房使用的封閉迴路液對氣冷卻（liquid-to-air cooling）系統：三組模組化冷卻機組，每組為11個Open Rack單元的ORv3相容機殼，合計可將每座機櫃標稱60 kW的熱量排至熱通道[S-225]。相對地，旗艦機櫃級系統（rack-scale system）則透過其機箱進行液冷：NVIDIA將其Vera Rubin運算機箱描述為100%液冷，AMD的Helios則將冷卻液分配至其運算機箱與交換機箱[S-178; S-179]。

營運層面的系統

IF 液冷機櫃自出廠起即為濕式，因此未解的問題是每一次交接時由誰負責該流體。 若充液出貨，PG25迴路在−13至−9 °C有結凍風險[S-216]；若排空出貨，則有腐蝕與微生物滋生風險。NVIDIA在2026年峰會的議程涵蓋部署前調理所用的沖洗加注推車，以及液冷機櫃的運送、儲放與試運轉驗收（commissioning）[S-226]。在這些作業仍因供應商而異的期間，「如今決定步調者是維修與責任歸屬、而非硬體」這一研判（reading）成立。應觀察OCP或ASHRAE是否推出附驗收限值的作業程序，例如水基冷卻液規格（specification）已訂定者[S-217]；若系統整合商在標準保固中承擔加注、結凍與污染風險，則此一研判不成立。

4.4 設施與電網介面

百萬瓦級機櫃是一個建築問題。在該密度下，機房的配電電壓、冷卻液設備、儲能與電網併接均依IT負載選定，而其使用年限會跨越數個世代的IT設備。建築物的融資期間長達數十年，而具有不同功率、冷卻與重量規格範圍的加速器（accelerator），卻每一至兩年就推出一代。正是這種落差，使OCP的關注焦點自伺服器轉向設施，也使2026年最具影響力的文件所描述的是機房、轉換器與電網行為，而非電路板。

這些文件的意義不止於工程。主機代管或新興GPU雲端業者（neocloud）需要它們，才能就尚未問世的硬體向承租客戶承諾電力與冷卻；買方需要它們，才能以共同的條件比較各設施；電網業者則需要知道，當一條輸電線跳脫時，一整座AI機櫃機房作為負載會有何表現。本節研讀OCP的設施文件，接著研讀與之銜接的儲能與電網規則。

F 機櫃周邊設備的預測成長速度，大致與OCP認可市場相當。 Dell'Oro Group預估資料中心實體基礎設施的製造商營收——涵蓋不斷電系統、散熱管理、機櫃與機架配電、匯流排槽、機櫃與氣流封閉設施，以及其軟體與服務——將自2025年至2030年以每年22%的速度成長，達1,200億美元，並預估至2030年資料中心容量將新增近200 GW[S-594]。該機構對風險的看法是：「Demand is no longer the open question—delivery is.」它並指出設備交期（lead time）、施工人力、電網併接與社區同意等因素[S-594]。

F OCP的Open Data Center for AI倡議列出五項優先事項。 其白皮書（white paper）（0.5版，2026年3月18日，封面標示2026年4月）的作者來自OCP Foundation、Google、Oracle、NVIDIA、Meta、AMD與Microsoft，列出：開放資料中心參考設計（reference design）；能源與電網解決方案；「Open and Standard Service Level Agreements with Colocation, Neocloud and other 3rd party facility operators」；整合一致的IT與設施遙測（telemetry）；以及用於減少閒置電力容量（stranded power）的功率估算[S-227]。該白皮書並將設施可替換性，即同一座機房得以容納相繼各代硬體的能力，列為其目標之一[S-227]。

A 其中三項優先事項涉及在各方之間分配風險，因此比一張機櫃圖面更難定案。 閒置電力容量（stranded power）指已配置並已付費、但從未被取用的容量，原因是負載係按銘牌額定值規劃；在電網併接須耗時數年之處，更精準的估算可使同一併接容量後方承載更多運算資源。可替換性則處理使用年限落差的問題：依某一代加速器（accelerator）的規格範圍所設計的機房，在下一代到來時會使資本閒置。服務水準這一項優先事項則決定由誰承擔這些風險——營運業者或其承租客戶[S-227]。OCP 2026年4月的已核准貢獻文件（contribution）清單中，並無任何關於服務水準或功率估算的項目[S-062]。

F 已核准的文件涵蓋設計原則、電力路徑與儲能。 由AMD、Google、Meta與NVIDIA撰寫的Open Data Center設計原則已達0.7.0修訂版，於2026年8月27日生效，其修訂紀錄將該版描述為「Collaborative 2027–2028 guidance unified across AMD, Google, Meta, and Nvidia, generalizing the scope for industry AI/ML systems」[S-001]。低壓直流（low-voltage direct current, LVDC）白皮書（white paper）1.1.0版（2026年9月2日）指出，產品標準、法規與安裝規章將低壓界定為「at or below 1000 VAC or 1500 VDC」的運作狀態，並增列關於IT機櫃、控制與運轉，以及電磁相容性的章節[S-059]。OCP 2026年4月的更新曾列出兩者的較早版本[S-062]。

F 自中壓邁向800 VDC這一段已有初版草案規格（specification）與早期供應商。 由Google、Microsoft與NVIDIA撰寫、2026年6月22日生效的SST規格0.3.0修訂版自稱為「first draft of SST product for OCP community」：該規格界定自13.8 kV或34.5 kV轉換為800 V單極輸出的轉換器，額定5 MW（在34.5 kV下最高10 MW），並要求在使用率50%至100%之間效率至少達98%，且僅計入功率鏈路損耗[S-054]。Amperesand於2025年11月表示，計畫於2026年交付30 MW的商用SST系統[S-203]；Heron Power則計畫於2026年夏季推出其首批工程原型機[S-204]。

A 機櫃與冷卻液文件已達1.0版；電力路徑與設計原則尚未。 Open Rack Wide（ORW）與Deschutes均已以1.0.0版發布[S-228; S-030]，而Diablo 400與設計原則停留在0.7.0版、SST規格（specification）停留在0.3.0版，電力相關內容大部分由白皮書（white paper）承載[S-027; S-001; S-054; S-059]。白皮書記錄的是共同實務，不含可用以檢驗產品的符合性條文，因此2026年的標案可以引用ORW或Deschutes作為已發布的規格，而電力路徑文件僅能作為設計參考。

儲能、平滑化與電網邊緣

F 儲能正透過兩條途徑取得驗證（qualification）：OCP的需求文件與NVIDIA的計畫。 OCP的白皮書（white paper）「Requirements for Energy Storage Systems Used in Data Centers」（1.0修訂版，2026年1月）記載，NFPA 855（2026年版）現已要求進行大規模火災測試[S-205]。NVIDIA 2026年5月28日的DSX BESS Self-Qualification Guidelines v1.0，界定由合作夥伴自行執行的電池儲能系統（battery energy storage system, BESS）驗證，適用於AI負載緩衝、需量反應與低壓故障穿越（ride-through）[S-206]；DSX Ready則於2026年9月21日啟動，首批通過驗證的BESS來自Hitachi Energy、LG Energy Solution與Tesla[S-048]。NVIDIA的設計指南將BESS置於34.5 kV開關設備匯流排上[S-175]。

F 電網業者正在訂定AI負載在擾動中應有行為的規則。 2026年7月22日，北維吉尼亞一條230 kV線路跳脫，使PJM Interconnection（PJM）系統中約3,800 MW的資料中心負載脫離，這些負載未經預警即轉為備用發電，為PJM歷來規模最大的此類事件[S-229]。PJM於2026年9月10日提案要求大型負載（large load）在正常清除的輸電擾動中維持故障穿越（ride-through），並擬於2026年11月向聯邦能源管理委員會（Federal Energy Regulatory Commission, FERC）提交[S-061]。北美電力可靠度公司（North American Electric Reliability Corporation, NERC）正在研擬三項運算負載標準，其中一項關於保護協調與擾動監測，首批標準預定於2026年底完成[S-230]。

A 故障穿越（ride-through）使設施的電力架構成為電網資產，或成為電網風險。 如7月事件所示，為保護自身設備而脫離的負載，會把擾動轉嫁給電網[S-229]。置於可控整流器或SST之後直流匯流排（busbar）上的儲能，可在前端穿越故障期間維持IT負載，而800 VDC設計、儲能需求與電網規則正在此處交會[S-019; S-205; S-061]。對營運業者而言，這把一項電氣設計選擇與併接條件綁在一起；第7.2節則討論這些規則本身。

? 未解問題——OCP的電網介面文件是否會界定電網業者所接受的故障穿越（ride-through）行為。 在此處所引的OCP文本中，LVDC白皮書（white paper）就電壓與頻率故障穿越提供指引[S-059]，Open Data Center for AI白皮書則把低壓故障穿越列為一項挑戰[S-227]；PJM已提案、NERC正在研擬故障穿越要求，兩者各依自身時程進行[S-061; S-230]。在尚無任一文件能同時滿足兩者之前，一座機房的保護與儲能設計取決於由哪一套規則管轄，而設計者必須按兩者中較嚴格者規劃。

F 在電網併接緩慢之處，營運業者正在表後（behind the meter）增設發電。 Bloom Energy與Oracle於2026年4月將雙方合作擴大為一份主服務協議，涵蓋最高2.8 GW的固態氧化物燃料電池容量，其中1.2 GW已簽約；Bloom表示其在55天內為Oracle交付了一套可完全運轉的系統[S-231]。Vertiv於2026年9月同意收購UtilityInnovation Group，該公司為資料中心設計微電網與表後電力，Vertiv並將此交易定位為在公用電力併接緩慢的情況下加快送電時程的途徑[S-232]。渦輪機與電網設備則依較慢的時鐘運行（第5.5節）。

A 標準化單元（unit of standardisation）已隨約束條件自伺服器移向設施。 在每座機櫃1 MW、每列2 MW的水準下[S-027; S-190]，其約束力的限制位於伺服器之外，落在機房的配電、冷卻液設備、儲能與電網併接。OCP的文件如今橫跨這一整條鏈條，自冷卻液接頭到中壓轉換與整座機房的設計原則，其聯盟亦延伸至訂定設施與電網實務的機構：ASHRAE、電力研究院（Electric Power Research Institute, EPRI）與Current/OS[S-182; S-062]。每一步都把原本位於某一家供應商設備內部的功能，轉為各方之間的介面，並使差異化轉入實作（第6.7節）。

IF Open Data Center系列文件唯有在作者群之外的營運業者依其建置時，才會成為第三方設施的共同基礎。 這需要主機代管與新興GPU雲端業者（neocloud）在設計、服務方案或合約中按修訂版引用該設計原則，並需要首份服務水準或功率估算文件。應觀察設計原則是否推出1.0版、0.3.0版之後是否出現SST修訂版，以及2026年10月峰會或其後是否出現任何服務水準草案[S-001; S-054; S-227]。若後續修訂版仍維持同樣的四位作者而無第三方引用，或如DSX Ready這類平台計畫成為設施供應商實際的關卡[S-048]，則此一研判（reading）不成立。

> **觀察重點**
> 聖荷西，2026年10月12日至15日：一份Mt Diablo 2.0草案、0.3.0版之後的SST修訂版，以及NVIDIA關於800 VDC保護與CDU的議程[S-019; S-226]。
> 至2026年底：PJM向FERC提交的故障穿越（ride-through）申報，以及NERC首批運算負載標準[S-061; S-230]。

4.5 機櫃外型規格相互競爭，而目前並存

機櫃是一套系統各項決定性介面交會之處：匯流排（busbar）決定一個機箱可取用多少電力，歧管與接頭決定熱量如何離開機箱，垂直擴展佈線則決定有多少加速器（accelerator）能作為一體運作。因此，撰寫機櫃外型規格——即每一個機架與機箱都必須遵循的尺寸與對接點組合——的一方，便決定了原始設計製造商（original design manufacturer, ODM）必須建造什麼、一家機架供應商被另一家取代的難易程度，以及哪些機箱可裝入哪些機房。

這使機櫃介面成為商業控制權的所在。僅指定伺服器規格（specification）的買方，等於把決定性介面交給供應商；指定機櫃標準的買方則把它們收回。在Open Compute Project（OCP）內，目前有兩條機櫃脈絡並立，一條由營運業者與機架製造商共同撰寫，另一條由NVIDIA貢獻；客製化加速器（accelerator）機櫃則走第三條路線，而機櫃之下的模組標準則決定什麼能裝入一個插槽。

Open Rack：ORv3、High Power Rack與Open Rack Wide

F Open Rack V3固定的是介面，而非機架。 Open Rack V3（ORv3）基礎規格（specification）由Meta與Google的作者於2022年9月22日發布；OCP的規格維基目前列出2023年12月的1.1修訂版[S-180; S-202]。該規格界定資訊科技（information technology, IT）設備所對接的對象，即以標稱51 VDC供電的48 V等級匯流排（busbar），而未就機架的外部深度、高度或寬度訂定任何要求[S-180]。其High Power Rack（HPR）隨Meta液冷的Catalina系統於2024年10月15日問世，功率最高達140 kW[S-176]，但該維基至今仍將HPR列為審查中的Meta規格資料夾，而非附版本編號的文件[S-202]。

F Open Rack Wide使占地面積加倍，並把基礎規格與設計分開。 由Rittal與Meta作者撰寫的Open Rack Wide（ORW）基礎規格（specification）V1.0.0版於2026年4月28日發布，適用於需要「a larger footprint than a traditional ORv3 based IT gear rack」的機櫃；該規格載明「There are no prescribed requirements in the base specification for external frame dimensions such as depth, height, base geometry, or maximum payload capability」[S-228]。OCP於2026年4月核准其為貢獻文件（contribution）[S-062]，而Rittal在OCP的產品列表已標示基礎規格v1.0.1版[S-233]。Meta於2026年5月5日生效的設計規格V1.0.1版，訂定標稱寬度1,200 mm、深度1,219 mm，以及「SHALL be 4700 kg (~10,360 lbs) minimum」的機架承載額定值[S-234]。

A Open Rack使機架供應商成為可替換者，並使規格範圍成為合約變數。 由於ORv3與ORW基礎規格（specification）僅固定IT設備所對接的對象，依某一基礎規格製造的機箱理應可裝入任何符合規範的機架，而第二家實作者也可在不修改Meta文件的情況下發布自己的設計[S-180; S-228]；高度、承載與底座幾何則移入採購合約。以HPR功能組合為基礎的Rittal雙倍寬度ORW機架即展現此一結果：液冷匯流排（busbar）達700 kW以上，「with an adjacent power rack」，並同時可容納Open Rack單元（OU）與機櫃單元（RU）[S-233]。若日後的ORW設計需要超出基礎介面的機箱變更，則此一研判（reading）將趨於薄弱。

Helios、MGX與Kyber：建置於Open Rack之上與其旁的供應商系統

F AMD的Helios是一款建置於Open Rack Wide之上的參考設計（reference design）。 AMD稱Helios為「a reference design, not a product for sale」：配置72顆Instinct MI455X圖形處理器（graphics processing unit, GPU）、EPYC處理器與Pensando網路卡，並透過兩顆「512 lane 200G UALoE switch ASICs」以「using UALink」方式組網[S-179]。AMD於2025年10月首次以靜態展示方式呈現，建置於ORW規格（specification）之上，並表示Helios是其透過OCP與Meta共同開發[S-235; S-112]。Meta於2026年2月同意部署最高6 GW的AMD GPU，自Helios機櫃起步，首個吉瓦的出貨落在2026年下半年；7月AMD表示Helios「now in production」，且OpenAI預期自第4季起將其上線[S-112; S-236]。

A Helios顯示一家晶片供應商正按買方的機櫃來設計其旗艦系統。 一家營運業者與一家機架製造商共同撰寫了機櫃格式，並在OCP發布；AMD隨後依該格式設計其系統[S-228; S-112]。對採購而言，這一劃分很重要：機架及其介面是開放的，而該系統仍屬AMD的參考設計（reference design），而非OCP規格（specification），其符合ORW之說亦出自供應商自身的陳述[S-179]。若AMD將Helios設計本身貢獻給OCP，則此一研判（reading）即須改變。

F NVIDIA的MGX脈絡是以貢獻文件（contribution）形式進入OCP。 2024年10月15日，NVIDIA貢獻其GB200 NVL72設計的若干要素，包括「the rack architecture, compute and switch tray mechanicals, liquid-cooling and thermal environment specifications」[S-237]。OCP記載了一款支援19吋美國電子工業協會（Electronic Industries Alliance, EIA）規格、配備1,400 A匯流排（busbar）的強化型機櫃，並表示該貢獻有助於「from multiple vendors」的機櫃與機箱在「in power, cooling and mechanical interfaces」上互通[S-238]。2025年10月，NVIDIA表示其「plans to contribute the upgraded rack as well as the compute tray innovations」，所指為其Vera Rubin世代[S-178]；該貢獻是否受理，尚未宣布。

F Kyber把這條脈絡帶到800 VDC，但時程有所爭議。 NVIDIA將其下一代機櫃Kyber描述為「by 2027」可容納576顆Rubin Ultra GPU[S-178]，並曾表示800 VDC資料中心的全面量產「will coincide with NVIDIA Kyber rack-scale systems in 2027」[S-028]。2026年7月6日，Tom's Hardware引述SemiAnalysis報導，指出因機櫃的中介板電路板而延後至2028年；NVIDIA則回應「Our roadmap is intact」[S-186]。當前世代正在出貨：NVIDIA表示Vera Rubin正「ramping into full production」，機櫃已在指名的雲端合作夥伴處運行，而鴻海（Foxconn）表示Vera Rubin機櫃於2026年第3季進入量產[S-010; S-022]。

A OCP如今對同一級系統持有兩條機櫃脈絡。 MGX以19吋EIA機架為起點，搭配大電流匯流排（busbar）與NVIDIA自有的垂直擴展佈線；Open Rack則以ORv3匯流排為起點，最重型的系統另以ORW基礎規格（specification）為起點[S-238; S-180; S-228]。兩者都正把電力轉換移出IT機櫃，移入Rittal的相鄰電源機櫃（power rack）或NVIDIA為Kyber所設計的800 VDC架構（第4.2節），因此下一場競爭既在機架，也同樣在機櫃之間的電力與冷卻介面[S-233; S-028]。市場已開始以機櫃計數：TrendForce預估GB300、VR200與VR300世代NVL72機櫃系統的產值2027年將超過7,100億美元[S-105]。

機櫃之下：模組與建構區塊

F 模組標準的推進速度慢於其上方的機櫃。 OCP的Modular Hardware System（MHS）工作致力於「by providing consistent interfaces and form factors among modular building blocks」以達成互通[S-239]。Data Center Modular Hardware System（DC-MHS）維基載明，其「workstreams are still working the MHS R1 v1.x specifications」，且「The Final Spec Agreement (FSA) for the latest specs has not yet been signed」[S-240]。2023年9月14日的OCP Accelerator Module（OAM）規格（specification）r2.0版「supports up to 1000W」，而NVIDIA為其Blackwell Ultra GPU訂出最高1,400 W[S-183; S-174]。網路卡外型規格OCP NIC 3.0目前為1.6.0版，另有液冷實體模型列於2026年峰會議程[S-241; S-220]。

A 模組的規格範圍為標準插槽可容納的對象設下上限，而旗艦產品已超出開放規格的範圍。 OAM為1,000 W、旗艦GPU為1,400 W，因此最高功率的加速器（accelerator）係由其供應商在機箱與機櫃層級訂定規格，而非作為可互換模組[S-183; S-174]；通用基板路線則仍為氣冷伺服器保留空間，AMD至今仍提供此類產品[S-224]。DC-MHS的效益落在別處：共通的模組介面使ODM得以跨處理器世代沿用機殼、電源與纜線，而另行訂定的Datacenter-ready Secure Control Module（DC-SCM，2.2修訂版）則讓各供應商的主機板共用同一個管理平面[S-242]。其代價是步調；最新的協議至今仍未簽署[S-240]。

IF 沒有任何單一機櫃外型規格已確立取代其他規格。 紀錄所支持的是Open Rack系列、MGX系列與客製化加速器（accelerator）機櫃並存，後者例如OpenAI與Broadcom計畫「entirely with Ethernet and other connectivity solutions from Broadcom」加以擴展者[S-228; S-238; S-243]。此一研判（reading）僅在各系列持續貢獻新世代並持續出貨系統的期間成立。應觀察Vera Rubin機櫃是否被記載為已受理的貢獻文件（contribution）、是否出現第二家晶片供應商的ORW系統、HPR是否以附版本編號的規格（specification）發布，以及Kyber的800 VDC設計以何種形式貢獻。若其中一個系列停止貢獻，或買方將新建機房標準化於單一規格範圍，則此一研判不成立。

觀察重點

OCP全球峰會（OCP Global Summit），聖荷西，2026年10月12日至15日。 Vera Rubin機櫃與運算機箱是否被記載為已受理的貢獻文件（contribution），以及初步版的ORW Canister IT Gear Design Guide（0.1版，2026年4月，Meta與AMD）是否推進[S-178; S-202]。

4.6 垂直擴展交換網路維持多元，其較低層級則趨於收斂

一座AI叢集運行兩個網路。垂直擴展網路（scale-up network）把同一座機櫃或同一個pod內的加速器（accelerator）緊密連接，使彼此都能讀寫對方的記憶體；水平擴展網路（scale-out network）則把各個pod連成數千顆規模的叢集。當一個模型被切分到多顆加速器上時，每一步都要等待同儕的部分結果，而這些交換既頻繁又短小，因此每則訊息的固定成本（標頭、為復原丟失封包所需的一次往返、排隊）會在每一步重複發生。因此，垂直擴展設計偏好記憶體語意（memory semantics）、逐鏈路復原與無損流量控制；水平擴展設計則以延遲換取傳輸距離與路徑多樣性。

A 垂直擴展網域正是鎖定（lock-in）如今的所在。 它把加速器（accelerator）、交換器與轉接卡綁在一起，其規模受訊號在銅線上可傳輸距離的限制，這也是它已等同於機櫃的原因（第4.7節）。三種途徑在此競爭：依NVIDIA自身條件向合作夥伴開放的NVIDIA NVLink；由聯盟撰寫的Ultra Accelerator Link（UALink）；以及乙太網路，OCP就此持有一份交換器端規格（specification）與一份端點傳輸文件[S-244; S-245; S-246; S-247]。水平擴展則有其自身圍繞Ultra Ethernet規格的收斂。圖20依層級標示各項工作的位置。

F AI後端網路的交換器銷售額已超越前端網路（front-end network）。 Dell'Oro Group報告指出，2026年第2季連接加速伺服器的AI後端網路交換器銷售額「surpassed those in front-end networks for the first time」[S-607]。該機構預測此類交換器的累計支出「to approach $1 trillion over the next five years (2026-2030)」，且至2030年垂直擴展將占市場半數以上；自1月以來該機構已「substantially raised」其預測，當時其預期支出至2030年將突破1,000億美元[S-608; S-609]。該機構並預期在其預測期限（horizon）內，市場將「remain supply-constrained rather than demand-constrained」[S-608]。

A 網路支出隨加速器（accelerator）而動，其成長最快的部分正是鎖定（lock-in）所在的垂直擴展網域。 Dell'Oro預期乙太網路「to emerge as the long-term winner across both scale-up and scale-out architectures」，而目前垂直擴展則由NVLink主導[S-609; S-608]。Dell'Oro 1月的預測預期，AI後端網路中目前以800 Gb/s為主的多數交換器埠，至2027年將為1,600 Gb/s，至2030年將為3,200 Gb/s[S-609]，而每一步都代表新一代的交換晶片、光學元件與驗證（qualification）。因此，垂直擴展的成長提高了本節各項交換網路選擇的商業重要性。

垂直擴展用乙太網路：ESUN與SUE

F ESUN撰寫的是乙太網路垂直擴展的交換器端。 OCP於2025年10月13日啟動Ethernet for Scale-Up Networking（ESUN），創始成員12家，其中包括AMD、Broadcom、Meta、Microsoft、NVIDIA與OpenAI；該倡議「focuses solely on open, standards-based Ethernet switching and framing for scale-up networking」[S-248]。其網路營運業者需求的基礎規格（specification）於2026年2月9日生效，並於3月10日發布，當時參與企業已逾175家[S-246; S-020]。該規格就壅塞訂定優先權流量控制（priority flow control, PFC）與基於信用額度的流量控制（credit-based flow control, CBFC），就可靠性訂定鏈路層重傳（Link Level Retry, LLR），並載明「L2 encapsulation to carry all the SU traffic with additional 4B and removal of IP headers (saving 20-40B)」[S-020; S-246]。

= 當訊息很小時，少數幾個標頭位元組就有影響。 取ESUN規格（specification）所指的位元組數，即其移除的20至40位元組IP標頭與其增加的4位元組，並對照單一訊息的酬載，且不計兩者共有的乙太網路訊框開銷[S-246]：開銷＝標頭位元組數÷酬載位元組數。就64位元組的訊息而言，約為31%至63%，相對於6.25%；就1,024位元組的訊息而言，則為2.0%至3.9%，相對於0.4%。這項節省對於短小而頻繁的垂直擴展交換具決定性，對於大量傳輸則微不足道，這也是兩個網路規格不同的原因。

A ESUN縮小交換器之間的差異，並讓端點保持開放。 它是一份由營運業者撰寫的需求清單，而非新的協定，由乙太網路訊框、精簡標頭，以及「defined by UEC」（即Ultra Ethernet Consortium）的鏈路機制組合而成[S-246]。交換晶片供應商因而獲得一套統一的需求，營運業者則獲得一份採購參考，而加速器（accelerator）供應商仍在端點保有其差異化，該部分並不在ESUN的範圍之內[S-248]。若後續修訂版將範圍擴及主機端軟體堆疊，則此一研判（reading）將趨於薄弱。

? 未解問題——乙太網路垂直擴展的端點這一半，尚無多方規格（specification）。 唯一已公布的端點傳輸文件是Broadcom的Scale Up Ethernet（SUE）1.0.0，於2025年9月5日以單一企業貢獻文件（contribution）的形式獲核准，適用於加速器（accelerator）之間的load/store記憶體操作[S-247]。OCP其後將該項工作更名並釐清為SUE-T，這是一個參與者須簽署回饋協議（Feedback Agreement）的工作流（workstream），至今尚未發布規格[S-249; S-248]。今日所設計的端點，必須引用SUE 1.0.0或供應商自有的傳輸方式。

UALink：走在晶片之前的聯盟堆疊

F UALink在1.0版晶片出貨之前，規格已達2.0版。 UALink Consortium於2025年4月8日發布其200G 1.0規格（specification），以每通道200 Gb/s提供記憶體語意（memory semantics），即讀取、寫入與原子操作，適用於「up to 1,024 accelerators within an AI computing pod」[S-245; S-250]。2026年4月7日，該聯盟又批准四份文件：新增網路內運算的Common Specification 2.0、資料鏈路層與實體層2.0、採用Switch Abstraction Interface（SAI）的Manageability 1.0，以及對應Universal Chiplet Interconnect Express（UCIe）3.0晶粒間介面（die-to-die interface）標準的Chiplet 1.0[S-251; S-168]。The Register報導，1.0版晶片將於2026年下半年進入實驗室，產品則於2027年較晚時推出[S-252]。

A UALink是一套獨立的堆疊，但與乙太網路共用電氣通道。 它有自己的鏈路層與實體層、一個具記憶體語意（memory semantics）的交易層，以及自有的可管理性與小晶片（chiplet）規範檔[S-251]。然而它與乙太網路共用224G電氣等級：Credo的重定時器（retimer）自2026年1月起送樣、量產數量預定於第3季供應，可支援UALink、ESUN與乙太網路，「which converge at layer 1 on the 224G 802.3 standard」[S-253]。AMD的Helios則混用各層，以乙太網路交換晶片承載UALink[S-179]。由於規格（specification）走在產品之前，2027年之前的UALink採購僅能倚賴文件與已公告的晶片；若1.0版晶片未能趕上2027年已公告的產品，則此一研判（reading）將趨於薄弱。

NVLink：依NVIDIA自身條件開放

F NVIDIA透過自有計畫開放NVLink，同時也在建構乙太網路產品。 NVIDIA於2025年5月18日宣布NVLink Fusion，使其他企業得以與其夥伴生態系共同建構半客製化AI基礎設施[S-244]。其Rubin平台將NVLink 6交換器與ConnectX-9轉接卡及Spectrum-6乙太網路交換器搭配；NVIDIA於2026年1月表示Rubin「is in full production」，並於8月表示Vera Rubin機櫃已在指名的雲端合作夥伴處運行[S-254; S-010]。NVIDIA同時是ESUN的創始成員，並為Meta的Minipack3N交換器供應Spectrum乙太網路產品[S-248; S-255]。

A NVIDIA同時站在交換網路競爭的兩邊。 NVLink的開放是透過NVIDIA自有計畫，而非聯盟規格（specification），因此Fusion擴大了可連接NVLink的對象，卻未改變由誰來界定它[S-244]。同時，NVIDIA共同創立了ESUN，並把乙太網路晶片賣進一家OCP營運業者的交換器[S-248; S-255]。對買方而言，其風險在於議價能力（bargaining power）：一套NVLink系統把加速器（accelerator）、交換器與轉接卡綁在單一供應商的藍圖上[S-254]。若NVIDIA將NVLink規格貢獻給某一開放機構，或Fusion夥伴出貨同時實作ESUN或UALink的晶片，則此一研判（reading）即不成立。

水平擴展：Ultra Ethernet、MRC與開放叢集設計

F 水平擴展乙太網路正收斂於同一套機制。 Ultra Ethernet Consortium於2025年6月11日發布規格（specification）1.0；2026年7月16日的1.0.3版為現行版本，並新增每通道200 Gb/s的訊號傳輸[S-256; S-257]。其傳輸層會把一個連線的封包噴灑（packet spraying）至多條路徑，並接受這些封包「may arrive in a different order than how they were sent」[S-258]。2026年5月5日，OpenAI宣布Multipath Reliable Connection（MRC）規格，該規格與AMD、Broadcom、Intel、Microsoft及NVIDIA共同開發，並以OCP貢獻文件（contribution）形式發布；MRC「takes the packets from a single transfer and sprays them across hundreds of paths」[S-259; S-260]。

A MRC把同一套機制帶到既有設備基礎，而OCP則把設計打包給買方。 MRC是延伸而非取代融合乙太網路遠端直接記憶體存取（remote direct memory access over converged Ethernet, RoCE）的可靠連線語意，OpenAI表示其已在自身最大的GB200叢集上運行[S-260; S-259]。OCP的AI Inference Fabric Reference Architecture 1.2版（2026年4月26日），在開放網路作業系統SONiC之上，為64至1,024顆加速器（accelerator）打包了一套乙太網路設計，具備無損後端與「two isolated redundant fabrics」[S-261]。該文件由兩家企業撰寫，因此它是一份範本，而非多供應商共識，且發布並不等於採用。

A 依層級研讀，這些交換網路在底層趨於收斂，在上層則維持多元。 乙太網路路線共用200 Gb/s通道等級，以及ESUN所採納、SUE所承載的Ultra Ethernet鏈路機制，而管理層面正朝SAI這類共通介面移動[S-246; S-247; S-168]。決定一顆加速器（accelerator）如何讀取另一顆記憶體的交易層，則維持多元：垂直擴展有UALink、SUE與NVLink；水平擴展有Ultra Ethernet傳輸與MRC。專利條件亦有差異：ESUN與SUE係依Modified Open Web Foundation Agreement 0.9發布，而Ultra Ethernet文本則讓實作者「solely responsible for obtaining any necessary licenses」，且專利池的徵集仍在進行（第6.6節）[S-246; S-247; S-267; S-268]。

IF 在每一條垂直擴展路線都有大型部署支撐的期間，並存的局面成立。 目前每一條都有一個：乙太網路落在OpenAI與Broadcom計畫自2026年下半年至2029年底部署的機櫃[S-243]，NVLink落在NVIDIA的Vera Rubin系統[S-010]，UALink則落在據報導將於2027年推出的產品[S-252]。應觀察UALink 1.0版晶片是否在2027年內進入已公告的產品、SUE-T是否發布草案，以及是否有供應商將已公告的系統自一種垂直擴展交換網路轉換至另一種。若出現已公布的對應方式，使某一交易層得以在另一方的訊框之上運行且買方加以採用，或其中一條路線失去其支撐性部署，則此一研判（reading）不成立。

圖20。 交換網路的各項工作在通道層與鏈路層趨於收斂，在交易層則維持多元

層級	OCP的工作	其他機構與供應商
叢集營運	Open Cluster Designs參考交換網路	UALink Manageability 1.0（SAI）；SONiC
水平擴展傳輸	MRC 1.0	Ultra Ethernet 1.0.3
垂直擴展端點	SUE 1.0.0；SUE-T工作流	UALink Common Specification 2.0；NVLink Fusion
垂直擴展交換	ESUN第2層與第3層訊框	UALink資料鏈路層與實體層2.0；NVLink 6交換器
鏈路層	ESUN：LLR、PFC、CBFC	Ultra Ethernet鏈路機制；IEEE P802.1Qdt
通道與光學	OCS子專案白皮書	IEEE P802.3dj；OIF CEI-224G與CEI-448G

截至2026年9月27日，AI叢集交換網路各層級由哪一個機構或企業撰寫。OCP欄列出OCP規格（specification）、貢獻文件（contribution）與工作流（workstream）；另一欄則列出各聯盟、IEEE與OIF的專案，以及企業計畫。各項的位置依其文件所載範圍而定。來源：[S-248; S-246; S-247; S-249; S-260; S-261; S-262; S-251; S-244; S-254; S-257; S-263; S-264; S-265; S-266]。
由最底層開始。同一層級中的項目可在一套系統中並存；本圖顯示的是各層級由誰撰寫，而非市場占有率。

重要性何在

共通通道降低的是零組件的轉換成本（switching cost），而非系統的轉換成本。 將交易層、通道速率與管理介面分別訂定規格（specification），即使垂直擴展協定來自單一供應商，仍可使重定時器（retimer）、纜線與光學元件維持多來源供應[S-253; S-168]。

4.7 光學正朝封裝靠近

每通道資料速率每加倍一次，銅質通道在相同長度上流失的訊號就更多，因此銅的傳輸距離縮短，而用以復原訊號的電子元件耗電也更高。這限制了可共用同一個銅質垂直擴展網域的加速器（accelerator）數量，也是該網域最終定在機櫃層級的原因之一。光學可在損耗更低的情況下把訊號傳得更遠，但可插拔模組是在交換器前面板把光轉為電，仍留下一段通往晶片的長電氣路徑。解法是把轉換移到更靠近晶片之處，而這會把雷射、光纖接合及其熱量帶入機櫃中最昂貴的封裝。

A 銅與光學都受到限制，而光學的推進分為三步。 依The Register的報導，NVIDIA網路部門負責人形容銅具成本效益且不耗電，但超過數英尺後即明顯衰減，這迫使加速器（accelerator）必須集中於單一機櫃；NVIDIA執行長則表示，產業需要同時擴增銅、光學與共同封裝光學的產能[S-269]。光學依三種時鐘推進：速率更高的可插拔模組，包括不含數位訊號處理器（digital signal processor, DSP）的線性設計；把光引擎置入交換器或處理器封裝內的共同封裝光學（co-packaged optics, CPO）；以及改變拓撲本身的光路交換（optical circuit switching, OCS）。表6列出各項技術的階段。

電氣通道與可插拔模組

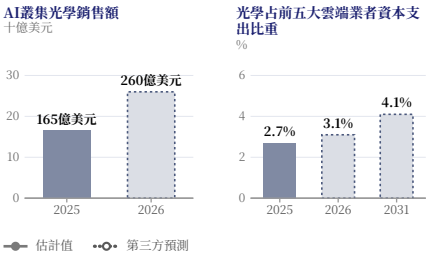
F 200 Gb/s以上的電氣通道標準仍在制定中。 負責200 Gb/s至1.6 Tb/s乙太網路的IEEE P802.3dj任務小組於2026年8月6日仍在運作，而IEEE一個關於每通道400 Gb/s訊號傳輸的研究小組，涵蓋電氣連結與最長500公尺的單模光纖傳輸距離[S-263; S-270]。Optical Internetworking Forum（OIF）將CEI-224G與CEI-448G電氣介面專案列為現行工作[S-264]。Converge Digest報導，OIF將於2026年9月的ECOC會議展示多供應商每通道448 Gb/s的示範，但CEI-448G實作協議並無完成日期[S-271]。

F 可插拔模組持續提升速率，而線性設計則捨去訊號處理器。 線性可插拔光學（linear pluggable optics, LPO）多源協議主張其「lower power, lower cost, and lower latency than competing solutions」，但未公布相關數據[S-272]。線性接收光學（linear receive optics, LRO）僅在接收端移除DSP；由Celestial AI與Cisco作者撰寫的IEEE Electronics Packaging Society綜覽文章傾向在每通道200G採用LRO，並在該條件下將初期1.6 Tb/s模組的功耗列為30 W以上[S-273]。Coherent在OFC 2025展示了每通道200G的多模1.6T模組，「even as the standards are still being written」，並宣布將於2026年3月的OFC推出每通道400G的3.2T光收發模組[S-274; S-275]。

F 出貨量與製造產能集中於少數幾家模組製造商。 市場研究機構Cignal AI預測，1.6 Tb/s乙太網路模組的出貨量2026年將超過500萬台，並報告Innolight在800 Gb/s出貨量上領先，但「the gap is narrowing with Eoptolink, Coherent, and Nvidia」[S-276]。TrendForce估計，中國製造商2026年掌握全球光模組代工產能約56%[S-277]。該機構並估計CPO在2026年僅占AI資料中心所用光收發模組約0.5%，另依不同的衡量方式預測，矽光子（silicon photonics）CPO至2030年前後可能達到AI資料中心約35%的滲透率[S-278]。這些是市場估計值，而非企業揭露資訊。

F AI叢集用光學在雲端預算中的占比正快速上升。 LightCounting估計，供AI叢集使用的乙太網路光收發模組與共同封裝光學（co-packaged optics, CPO）銷售額2025年達165億美元，2026年將達260億美元，兩年各成長60%；並假設售予前五大雲端企業的光學產品，占其資本支出（capex）的比重將自2025年的2.7%升至2026年的3.1%，並於2031年達4.1%[S-610]。規模更大的叢集在水平擴展網路中每顆GPU最多需要六個光收發模組，而垂直擴展所需頻寬則接近其十倍[S-611]。LightCounting認為至2030年AI叢集光學年銷售額達1,000億美元有「a reasonable chance」，其基準預測則被該機構稱為「Soft Landing」[S-611]（圖21）。

圖21。 AI叢集光學2025年成長60%，2026年預測亦為60%，並在前五大雲端業者資本支出中占比持續上升



左圖：供AI叢集使用的乙太網路光收發模組與共同封裝光學（co-packaged optics, CPO）銷售額，以十億美元計，LightCounting，2026年1月：2025年為估計值，2026年為預測值[S-610]。右圖：售予前五大雲端企業的光學產品占其資本支出（capex）的比重，以%計，為LightCounting對2025年、2026年與2031年的假設[S-610]。
範圍 採LightCounting自有的市場定義。其3月所載2025年成長70%的數字涵蓋全部乙太網路光收發模組，並非同一數列[S-611]；至2030年達1,000億美元為該機構所指出的一種可能性，而非其基準預測，故未予繪製。

A 產品走在通道標準之前，交換網路亦然。 Ultra Ethernet 1.0.3已新增每通道200 Gb/s的訊號傳輸，224G重定時器（retimer）也已在送樣，而P802.3dj仍未完成[S-257; S-253; S-263]。可插拔模組仍是量產主力與便於維修的預設選項，線性設計則可降低功耗，但其效益至今僅以定性方式陳述[S-272]。可插拔模組無法消除的，是自前面板至交換晶片的那段電氣路徑，而這正是共同封裝光學（co-packaged optics, CPO）所針對的路徑。

共同封裝光學：逐一供應商的早期量產

F NVIDIA的共同封裝交換器已自公告進入首批交付。 NVIDIA於2025年3月18日宣布其Spectrum-X Photonics（乙太網路）與Quantum-X Photonics（InfiniBand）CPO交換器，稱其所需雷射數量為傳統設計的四分之一[S-279]；其產品頁面目前將Spectrum-X Photonics列為「Available in the second half of 2026」，並主張「5x better power efficiency over pluggable transceivers」[S-280]。Lambda表示其於2026年6月收到NVIDIA的工程樣品，並明確說明是在量產之前[S-281]。2026年8月26日，NVIDIA表示其「supporting both pluggable and co-packaged optics」的Spectrum-6交換器系統正「arriving across the world's gigascale AI factories」[S-010]。

F 據報導Broadcom的CPO交換器已進入量產，鴻海（Foxconn）與台積電（TSMC）則把自身產品定在2026年。 TrendForce報告指出，Broadcom 51.2 Tb/s的Bailly CPO交換器已與Delta Electronics及Micas Networks共同進入量產[S-282]；ServeTheHome則於2025年10月報導，Broadcom 102.4 Tb/s的Tomahawk 6「Davisson」CPO交換器已在出貨[S-283]。鴻海表示其CPO全光交換器「are on track for mass production and shipment in the third quarter」，所指為2026年[S-022]。台積電排定其基板上COUPE光引擎於2026年開始量產，並於7月表示：「We started the production right now, and it will be ramped up」[S-024; S-084]。

A CPO已跨入早期量產，但每一家供應商各自驗證（qualification）自身的技術堆疊。 每一款交換器各自搭配自有的光引擎、雷射配置與組裝夥伴，而TrendForce將更大規模的量產爬坡定在2027至2028年，瓶頸則在光引擎良率（yield）、矽光子（silicon photonics）與先進封裝（advanced packaging）[S-282]。供應商的論據是效率：台積電（TSMC）主張COUPE的電源效率為板載可插拔模組的兩倍[S-024]。可維修性則是實務上的檢驗；依ServeTheHome的報導，Broadcom之所以把Davisson的雷射設計為現場可更換，是因為「light sources are usually the part with the highest failure rate」[S-283]。2026年一篇*Nature Electronics*的回顧文章，將「thermal management, manufacturability and standardization」列為仍待克服的障礙[S-053]。

? 未解問題——目前100 Tb/s級的CPO交換器並無共通介面可循。 OIF於2023年3月的3.2 Tb/s共同封裝模組協議，是作為51.2 Tb/s交換器的建構區塊而撰寫[S-284]。Tech Times於2026年8月報導，OCP內由20家企業組成的聯盟預期於2026年第4季提交其首批CPO規格（specification），該日期指的是提交而非發布，且僅有該篇報導為據[S-285]。在共通介面出現之前，光引擎的第二供應來源（second source）必須逐一交換器進行驗證（qualification）。

IF 唯有共通的測試、介面與可維護性標準隨之到位，CPO才能達到廣泛互通。 早期市場正圍繞單一供應商驗證（qualification）形成：Tech Times報導日月光（ASE）與台積電（TSMC）講者於OCP 2026年亞太峰會上的發言，指出缺少共用的測試標準，正推使買方走向單一來源（single-source）採購（procurement）[S-286]。應觀察OCP或OIF是否公布適用於光學引擎與外接雷射的介面、是否有第二供應來源（second source）的引擎在同一台交換器上通

過驗證，以及雷射更換的現場數據。若CPO僅憑單一供應商驗證即達到量產規模，而無任何開放介面草案，則此一研判（reading）不成立；屆時OCP的角色將局限於機櫃、佈線與管理層級。

> **觀察重點**
> **OCP全球峰會（OCP Global Summit）的OIF研討會，2026年10月13日。** 議程涵蓋224與448 Gb/s電氣通道、次世代光通道，以及CPO與近封裝光學（near-packaged optics）架構：即多供應商CPO市場所需的各項介面[S-287]。

雷射：光學稀缺已移往何處

F NVIDIA已簽約鎖定未來的雷射產能。 2026年3月2日，NVIDIA宣布對Lumentum與Coherent各投資20億美元；兩項非專屬協議均包含數十億美元的採購承諾與未來產能權利，對Lumentum為「advanced laser components」，對Coherent為「advanced laser and optical networking products」[S-288; S-289]。Tech Times於2026年5月報導，主要電吸收調變雷射（electro-absorption modulated laser, EML）供應商對NVIDIA以外買方的交期（lead time）已延長至2027年之後[S-290]。該交期數字為該媒體的報導，而非供應商的揭露。

F Lumentum描述雷射需求超前於供給。 據Tom's Hardware於2026年7月31日報導，Lumentum執行長表示，即使擁有五座磷化銦（indium phosphide, InP）晶圓廠，該公司EML的出貨量仍較客戶需求低逾30%[S-291]。TrendForce報導其看法，認為InP的供需缺口「has now surpassed that of DRAM and NAND」，並補充產業分析師預期此一失衡將持續至2027年之後，且中國約占全球精煉銦產量的70%[S-037]。上述為一家供應商的說法與所報導的分析師預期，並非經查核的產能數據。

A 光學稀缺已向上游移動，進入雷射製造與光子學產能環節。 CPO減少了每一鏈路所需的雷射數量，卻使依賴集中於少數InP雷射晶圓廠與光學引擎的封裝環節[S-279; S-037; S-282]。產能正在需求之前即被簽約鎖定：NVIDIA的採購承諾，以及Tower Semiconductor已簽訂、計入2027年營收的13億美元矽光子（silicon photonics）合約，另有2.9億美元的客戶預付款用於產能預留[S-288; S-292]。因此議價能力（bargaining power）落在雷射與光子學晶圓廠，以及有能力向其預付款的買方手中。若包含NVIDIA出資興建者在內的新增InP產能於2027年間填平此一缺口，則此一研判（reading）即被削弱。

光路交換與新進者

F 光路交換（optical circuit switching）在單一機群內已進入量產，作為開放堆疊則仍處於白皮書（white paper）階段。 Google的TPU v4超級電腦以光路交換器重組其互連；其作者指出「Deployed since 2020, TPU v4 outperforms TPU v3 by 2.1x and improves performance/Watt by 2.7x」，此為整個系統層級的增益[S-293]。OCP於2025年7月21日成立的OCS子專案，於2026年4月發表一份文件，自稱為「a technical white paper rather than a hardware or software specification」，並將迄今的採用歸於Google[S-294; S-262]。Lumentum銷售一款300×300的市售交換器，該公司與Marvell原計畫於2026年3月的OFC上展示[S-295; S-296]。

F 友達（AUO）已自相鄰的顯示器業務，轉入明確表態的AI基礎設施技術定位。 友達於2026年8月31日宣布，將於SEMICON Taiwan（9月2日至4日）展出一款與生態系夥伴共同開發的系統級Micro LED共同封裝光學（co-packaged optics）模組，目標為「short-reach interconnects of up to 10 meters in AI data centers」[S-052]。集團旗下企業Ennostar供應Micro LED發射器，Tyntek供應光偵測器接收端；Daxin Materials的介電材料「expected to be incorporated」於封裝之中，Corning則供應光纖[S-052; S-169]。友達與Corning亦計畫展出適用於較大型AI伺服器處理器晶粒的玻璃核心基板（glass-core substrate）（第3.5節）[S-052]。其階段為研發：友達稱正在「accelerating the commercialization」，並未公布量產、客戶或鏈路效能[S-052]。

A 如今的問題在於可製造性與驗證（qualification），而非友達（AUO）是否應列入版圖。 友達將顯示器領域的玻璃加工、巨量轉移與重佈線層封裝，帶入短距光鏈路，並由集團旗下企業供應光源與材料[S-052]。其所列Micro LED發射器的數據——熱穩定性達125°C、操作壽命逾30,000小時——為該公司自有數據，未經獨立查證[S-052]。應觀察客戶驗證、量產規模、鏈路效能、每位元功耗、良率（yield）、封裝生態系的採用，以及是否進入指名的伺服器或網路平台。在此之前，友達為研發階段的新進者，而非AI伺服器供應商。

> **讀者視角**
> **本章對產業、研發、智慧財產權與創新的意義**
> **產業。** 誰撰寫機櫃、電源機櫃（power rack）與交換器封裝的介面，誰就掌握驗證（qualification）；雷射、光子學產能與電力設備如今均於數年前即簽約鎖定。
> **研發。** 於量產規模下仍未解決者：800 VDC保護與故障協調、液冷迴路可靠性、光學引擎良率（yield）、光纖接合，以及現場可更換雷射。
> **智慧財產權。** 隨著機櫃底座、ESUN與200G通道趨於共通，差異化將移向光學引擎、雷射、冷板（cold plate）、保護裝置與控制韌體（第6.7節）。
> **創新。** 機會存在於尚無共用標準之處：CPO測試與燒機、開放的光交換器控制、機櫃層級的功率平滑，以及液冷（liquid cooling）遙測（telemetry）。

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-010 NVIDIA (26 Aug 2026); S-021 Infineon Technologies (5 Aug 2026); S-022 Hon Hai (12 Aug 2026); S-024 TSMC (23 Apr 2026); S-028 NVIDIA Technical Blog (20 May 2025); S-048 NVIDIA (21 Sep 2026); S-052 AUO (31 Aug 2026); S-055 SolarEdge and Infineon (9 Sep 2026); S-084 TSMC (16 Jul 2026); S-112 AMD (24 Feb 2026); S-169 AUO (31 Aug 2026); S-174 NVIDIA Technical Blog (22 Aug 2025); S-175 NVIDIA (19 Aug 2026); S-176 Engineering at Meta (15 Oct 2024); S-177 Google Cloud Blog (29 Apr 2025); S-178 NVIDIA Blog (13 Oct 2025); S-179 AMD (未註明日期); S-185 Microsoft Azure Blog (15 Oct 2024); S-188 UL Solutions (24 Sep 2026); S-189 Vertiv (13 Oct 2025); S-190 NVIDIA (11 Aug 2026); S-191 ABB (21 Sep 2026); S-192 Infineon Technologies (28 May 2026); S-193 Infineon Technologies (5 Aug 2026); S-194 Eaton (13 Oct 2025); S-195 Schneider Electric (13 Oct 2025); S-196 Vertiv (29 Jul 2026); S-197 ABB (13 Oct 2025); S-198 ABB (未註明日期); S-203 Amperesand (18 Nov 2025); S-204 Heron Power (18 Feb 2026); S-206 NVIDIA (28 May 2026); S-209 LG Electronics (22 Sep 2026); S-211 Boyd (12 Mar 2026); S-219 Microsoft Source (24 Sep 2025); S-222 Carrier (29 Apr 2026); S-224 AMD (未註明日期); S-225 Google Cloud Blog (16 Jun 2026); S-226 NVIDIA (未註明日期); S-231 Bloom Energy (13 Apr 2026); S-232 Vertiv (2 Sep 2026); S-235 AMD Newsroom (14 Oct 2025); S-236 AMD (23 Jul 2026); S-237 NVIDIA Newsroom (15 Oct 2024); S-243 Broadcom Investor Relations (13 Oct 2025); S-244 NVIDIA (18 May 2025); S-253 Credo Technology Group (29 Jan 2026); S-254 NVIDIA (5 Jan 2026); S-255 NVIDIA (13 Oct 2025); S-259 OpenAI (5 May 2026); S-268 Sisvel (未註明日期); S-274 Coherent (1 Apr 2025); S-275 Coherent (17 Mar 2026); S-279 NVIDIA Newsroom (18 Mar 2025); S-280 NVIDIA (18 Mar 2025); S-281 Lambda (1 Jun 2026); S-288 NVIDIA (2 Mar 2026); S-289 NVIDIA (2 Mar 2026); S-292 Tower Semiconductor (13 May 2026); S-295 Lumentum (未註明日期); S-296 Marvell Technology (16 Mar 2026); S-297 Ayar Labs (Mar 2026); S-298 Marvell Technology (2 Feb 2026)

標準組織與聯盟. S-001 OCP (27 Aug 2026); S-019 OCP (11 Aug 2026); S-020 OCP (10 Mar 2026); S-027 OCP (1 Mar 2026); S-030 OCP (26 Feb 2026); S-054 OCP (22 Jun 2026); S-059 OCP (2 Sep 2026); S-060 OCP (6 Aug 2026); S-062 OCP (29 Apr 2026); S-168 Ultra Accelerator Link Consortium (7 Apr 2026); S-180 OCP (22 Sep 2022); S-181 OCP (13 Oct 2025); S-182 OCP / ASHRAE (13 Oct 2025); S-183 OCP (14 Sep 2023); S-184 OCP (12 Sep 2023); S-187 ANSI / UL (22 Oct 2025); S-199 OCP (未註明日期); S-201 OCP (未註明日期); S-202 OCP (未註明日期); S-205 OCP (未註明日期); S-207 OCP (13 Oct 2025); S-208 OCP (未註明日期); S-212 OCP (4 Sep 2020); S-213 ASHRAE (未註明日期); S-214 OCP (未註明日期); S-215 OCP (4 Sep 2020); S-216 OCP (未註明日期); S-217 OCP (6 Aug 2026); S-218 ASHRAE TC 9.9 (未註明日期); S-220 OCP (22 Sep 2026); S-221 OCP (18 Aug 2023); S-227 OCP (Apr 2026); S-228 OCP (28 Apr 2026); S-233 OCP (未註明日期); S-234 OCP (5 May 2026); S-238 OCP (15 Oct 2024); S-239 OCP (未註明日期); S-240 OCP (24 Apr 2026); S-241 OCP (6 Aug 2026); S-242 OCP (27 Jul 2022); S-245 UALink Consortium (8 Apr 2025); S-246 OCP (9 Feb 2026); S-247 OCP (5 Sep 2025); S-248 OCP (13 Oct 2025); S-249 OCP (13 Jun 2026); S-250 UALink Consortium (8 Apr 2025); S-251 UALink Consortium (7 Apr 2026); S-256 Ultra Ethernet Consortium (11 Jun 2025); S-257 Ultra Ethernet Consortium (16 Jul 2026); S-258 Ultra Ethernet Consortium (16 Jul 2026); S-260 OpenAI / OCP (5 May 2026); S-261 OCP (26 Apr 2026); S-262 OCP (Apr 2026); S-263 IEEE 802.3 Working Group (6 Aug 2026); S-264 OIF (未註明日期); S-265 IEEE 802.1 Working Group (未註明日期); S-266 SONiC Foundation / Linux Foundation (未註明日期); S-267 Ultra Ethernet Consortium (28 Jan 2026); S-270 IEEE 802.3 Ethernet Working Group (3 Aug 2026); S-272 Linear Pluggable Optics MSA (未註明日期); S-284 OIF (29 Mar 2023); S-287 OCP (未註明日期); S-294 OCP (21 Jul 2025)

政府與監管機關. S-061 PJM Inside Lines (10 Sep 2026); S-229 PJM Inside Lines (11 Aug 2026); S-230 NERC (Aug 2026)

研究機構. S-053 Nature Electronics (19 Aug 2026); S-273 IEEE Electronics Packaging Society (未註明日期); S-293 Jouppi et al. (Apr 2023)

市場研究. S-006 TrendForce (17 Aug 2026); S-029 TrendForce (25 Jun 2026); S-105 TrendForce (28 Aug 2026); S-210 Dell'Oro Group (30 Sep 2025); S-276 Cignal AI (7 Jan 2026); S-277 TrendForce (5 Aug 2026); S-278 TrendForce (11 Mar 2026); S-282 TrendForce (27 Jul 2026); S-591 TrendForce (21 Aug 2025); S-594 Dell'Oro Group (19 Aug 2026); S-606 Uptime Institute (12 Aug 2026); S-607 Dell'Oro Group (3 Sep 2026); S-608 Dell'Oro Group (28 Jul 2026); S-609 Dell'Oro Group (4 Feb 2026); S-610 LightCounting (Jan 2026); S-611 LightCounting (Mar 2026)

媒體與法律評論. S-037 TrendForce (6 Aug 2026); S-186 Tom's Hardware (6 Jul 2026); S-200 ServeTheHome (28 Jul 2025); S-223 Data Center Frontier (11 Feb 2026); S-252 The Register (7 Apr 2026); S-269 The Register (5 Apr 2026); S-271 Converge Digest (26 Aug 2026); S-283 ServeTheHome (12 Oct 2025); S-285 Tech Times (15 Aug 2026); S-286 Tech Times (12 Aug 2026); S-290 Tech Times (27 May 2026); S-291 Tom's Hardware (31 Jul 2026)

表6。 光學正沿多個時鐘同時向封裝靠近

可插拔模組已量產出貨，CPO交換器則處於初期、特定供應商的量產階段；而能使CPO走向多供應商的開放介面，以及最新的進入者，仍停留在標準或研發階段。

技術	階段	對象與內容	日期	來源
1.6T可插拔模組	產品	模組製造商；預測2026年1.6T模組逾500萬套（Cignal AI）	2026	[S-276]
線性可插拔與接收光學	標準	LPO MSA規格，附有定性的功耗聲明；每通道200G時傾向採用LRO（IEEE EPS綜述）	2026	[S-272; S-273]
3.2T可插拔模組，每通道400G	已宣布	Coherent，為OFC 2026展出	2026年3月	[S-275]
CPO模組介面	標準	OIF 3.2 Tb/s共同封裝模組，適用於51.2T交換器	2023年3月	[S-284]
CPO交換器：Broadcom	產品	Bailly 51.2T已量產（TrendForce）；Tomahawk 6 Davisson 102.4T已出貨（ServeTheHome）	2025年10月至2026年7月	[S-282; S-283]
CPO交換器：NVIDIA	產品	工程樣品，隨後為首批合作夥伴出貨（TrendForce）；Spectrum-6系統支援可插拔與CPO	2026年6月至8月	[S-281; S-282; S-010]
CPO交換器：鴻海（Foxconn）	已宣布	全光CPO交換器的量產與出貨「on track」	2026年第3季	[S-022]
CPO光學引擎：台積電COUPE	產品	COUPE置於基板上；已開始量產，將逐步擴產	2026	[S-024; S-084]
光學I/O小晶片（optical I/O chiplet）	已宣布	Ayar Labs取得量產資金；Marvell預期Celestial AI營收將於2028財年爬坡	2026年2月至3月	[S-297; S-298]
Micro LED CPO模組	研發	友達（AUO）與Ennostar、Tyntek、Daxin及Corning合作；展出開發中的模組，目標10 m	2026年9月	[S-052; S-169]
光路交換，單一營運商	產品	Google TPU v4與資料中心網路架構	自2020年起	[S-293; S-262]
光路交換，開放與市售	研發	OCP白皮書；Lumentum R300（300×300）交換器，計畫於OFC 2026搭配Marvell光學元件展示	2026年3月至4月	[S-262; S-295; S-296]
448G電氣通道	研發	OIF CEI-448G專案；IEEE每通道400 Gb/s研究群組	2026	[S-264; S-270]

第4.7節所討論各項光互連技術的階段，資料截至2026年9月27日，並列出相關企業或機構及註明日期的證據。階段標示：產品，指出貨中或已明確進入量產；試點；研發；標準，指已公布的規格（specification）或協議；以及已宣布，指已陳述的計畫或展示，但無量產證據。市場研究與產業媒體來源於該列中指名。

階段描述的是指名的產品或文件，而非市場。CPO交換器標示為產品，意指出貨已開始，而非大量部署已然確立。

第5章

供應鏈與價值擷取

價值與籌碼（leverage）正移向掌握持續稀缺事物者——記憶體、封裝、測試、驗證（qualification）、保護硬體、渦輪機產能時段、電網併接與資本——而機櫃組裝則在微薄的利潤率下成長，其中一家組裝廠的營收基礎更已改變。

適用讀者：財務、採購、產業策略與政府；企業證據見第5.2節、第5.4節與第5.5節，企業數字所涵蓋與未涵蓋者見第5.3節，籌碼（leverage）見第5.6節，區域曝險見第5.7節。

人工智慧（AI）基礎設施的需求仍在上升，但收益並不會單純歸於出貨機箱最多者。記憶體製造商、為加速器（accelerator）進行封裝的晶圓代工廠、測試與微影設備供應商、電力設備製造商與渦輪機製造商，均依其各自的時鐘增添產能，其中數家如今更透過長期協議（long-term agreement）、預留與預付款提前出售該等產能。系統組裝廠以個位數利潤率出貨創紀錄的數量，其中一家更已改變其營收的計算範圍。本章沿著鏈條並跨越各區域追蹤價值與籌碼（leverage），並提出一個直白的問題：哪一方能夠讓一項部署停擺，原因何在？

5.1 價值落在產能持續稀缺之處

一項AI部署是一連串的採購：加速器（accelerator）與堆疊於其旁的記憶體、將兩者接合的晶圓代工與封裝產能、由原始設計製造商（original design manufacturer, ODM）組裝的機櫃、其周邊的電力與冷卻設備，以及建築背後的發電與電網併接。每一環節的獲利條件各不相同。部分供應商出售價格隨需求上漲的稀缺產能；其他供應商則就整合零組件收取費用，而該等零組件大多由客戶自行選定，有時並由客戶直接支付。因此，解讀價值擷取須先將各公司自身公布的經濟數據並列，再追問其間差異的原因。本節並未就任一公司與另一公司進行排名。

F 加速器（accelerator）、晶圓、記憶體與設備的供應商公布的利潤率偏高，且各依其自身的衡量方式。 NVIDIA對其2027財年第3季（結束於2026年10月的該季）的展望為毛利率74.0%，上下各50個基點[S-010]。台積電（TSMC）對2026年第3季的財測指引為毛利率65%至67%、營業利益率56%至58%[S-084]。SK hynix公布第2季營業利益率為76%，營收為79.3兆韓元[S-003]；Besi同一季的毛利率為65.7%[S-170]。

F 組裝廠與基礎設施供應商則依其他基礎、在其他水準上公布數據。 鴻海（Hon Hai／Foxconn）公布2026年第2季毛利率與營業利益率分別為6.12%與3.75%，營收為新台幣2.53兆元[S-022]。緯穎（Wiwynn）同一季公布的數字為9.3%與7.3%，係在特定客戶記憶體的入帳方式變更之後（第5.2節）[S-063]。Vertiv對2026年的財測指引為調整後營業利益率23.3%至24.3%[S-196]；Infineon公布截至2026年6月之三個月的分部業績利潤率（Segment Result Margin）為19.1%[S-021]。

A 利潤率反映的是對稀缺產能的掌握，而非在鏈條中的位置。 公布高利潤率的企業所出售的產能，是買方必須提前確保的：SK hynix已與約10家客戶敲定長期協議（long-term agreement, LTA）[S-003]，Micron預期客戶現金保證金與相關承諾達220億美元[S-129]。組裝廠整合的是買方指定的零組件；自2026年4月起，緯穎（Wiwynn）就特定客戶的記憶體改以代理人而非主要義務人身分採購，使該等金額不計入營收與銷貨成本[S-063; S-483]。在營收基礎上升的同時維持微薄利潤率，是另一種商業模式，而非較弱的商業模式：表7顯示的是定價能力所在，而非哪一家公司經營得更好。

IF 在記憶體、封裝與電力設備持續採配額供應期間，此一價值分配將維持不變。 若液冷（liquid cooling）機櫃整合變得稀缺到足以反映於價格，此一分配將移向系統整合商，其表徵為ODM毛利率在可比基礎上連續數季上升。若記憶體合約價格連續數季下跌，且供應商取消目前部分協議所附的保證金與最低價格，此一分配將移向買方[S-129]。應觀察SK hynix、Micron與台積電（TSMC）下一季的利潤率，並與鴻海（Hon Hai）及緯穎（Wiwynn）的利潤率並列，各自對照其自身歷史解讀，同時觀察新的供應協議是否仍附帶預付款。

5.2 ODM靠規模、產品組合與已改變的營收基礎而成長

原始設計製造商（original design manufacturer, ODM）將加速器（accelerator）、記憶體、網路與冷卻，轉化為可供出貨、安裝與維修的機櫃。其營收取決於機櫃數量與經其帳上流過的零組件價值，因此不僅隨出貨量變動，亦隨零組件價格變動。這使ODM的財報成為AI機櫃抵達資料中心速度的良好訊號，卻是組裝廠留存多少價值的拙劣指引。兩家台灣ODM於2026年揭露的資訊足以顯示上述兩項效應，其中一家並已改變其營收的涵蓋範圍。

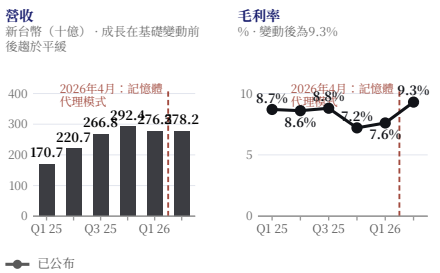
F 鴻海（Foxconn）預期其AI機櫃出貨量於2026年將成長逾一倍。 鴻海公布第2季營收為新台幣2.53兆元，年增41%，並表示隨著Vera Rubin機櫃於第3季進入量產，預期AI伺服器機櫃營收將較上季上升，出貨量將以高雙位數的速度成長[S-022]。該公司預期800G及以上交換器的全年營收將成長一倍，其共同封裝光學（co-packaged optics, CPO）全光交換器並如期於第3季量產與出貨[S-022]。輪值執行長Michael Chiang表示，2027年客戶對AI產能的需求「will still be very strong」[S-022]。

F 緯穎（Wiwynn）的營收基礎於2026年4月改變。 就特定客戶而言，記憶體採購改採代理模式，故該等記憶體金額不再計入營收或銷貨成本[S-063; S-064]。緯穎於2026年5月7日隨第1季財報首次揭露此一變動，公布4月營收為新台幣827.31億元，較3月減少16.14%，並表示預期商業模式的部分變動不會影響全年總獲利[S-064]。跨越2026年4月的營收與毛利率比較，須附上基礎說明。

F 2026年第2季是首個完全依新基礎公布的季度。 緯穎（Wiwynn）公布營收為新台幣2,781.53億元，年增26.0%，較第1季高出0.6%，毛利率為9.3%，營業利益率為7.3%[S-063; S-064]。該公司將毛利率上升歸因於代理模式的變動，「combined with higher NRE income from a new product introduction」，即一次性工程（non-recurring engineering, NRE）費用[S-063]。該公司表示，其持續與客戶合作調整商業模式，「in response to changes in the AI product mix and rising component costs」[S-063]。

F 緯穎（Wiwynn）經核閱的期中財務報表以會計用語說明此一變動。 其截至2026年6月30日止六個月的合併財務報表，經KPMG核閱並載明日期為2026年8月7日，其中指出就特定客戶交易，該集團協助自客戶指定的供應商處採購原材料，並載明「As the Group does not obtain control of such materials before transfer to the customers, the Group is considered an agent in these transactions」[S-483]。依IFRS 15，代理人僅就其預期有權取得的「in the amount of any fee or commission」認列營收，而非商品價格[S-484]。該財務報表既未指名記憶體，亦未就金額予以量化；財報發布則指名記憶體與2026年4月[S-063]。

圖22。 緯穎的營收與利潤率跨越2026年4月的基礎變動



依季度列示的緯穎（Wiwynn）合併營收（新台幣十億元）與毛利率（%）[S-302; S-303; S-304; S-305; S-064; S-063]。2025年第4季的數字於發布時未經查核。

範圍 自2026年4月起，特定客戶的記憶體採購改採代理模式，相關金額不計入營收與銷貨成本[S-063]；該分界線前後的營收與利潤率並非同基礎可比。緯穎（Wiwynn）將第2季的利潤率歸因於此一變動，加上新產品導入帶來較高的NRE收入[S-063]。

> 範圍 **緯穎**
> 2026年4月前後的營收與毛利率並非同基礎可比：特定客戶的記憶體改採代理模式，不計入營收與銷貨成本[S-063]；而經核閱的期中財務報表就客戶指定供應商所提供的原材料，依IFRS 15的用語將緯穎（Wiwynn）視為代理人[S-483; S-484]。來源所述為「memory」與「materials」，而非高頻寬記憶體（high-bandwidth memory, HBM），且為「certain customers」而非全部客戶；兩者均未就排除的金額予以量化。

A 代理模式使記憶體的價格波動脫離緯穎（Wiwynn）的營收；此一安排本身並非短缺的證據。 當高價值零組件以極低加價流經營收時，其價格上漲會使營收膨脹並稀釋毛利率百分比；在代理模式下，該零組件同時脫離營收與銷貨成本，故相同的獲利會呈現為較高的百分比，此與緯穎預期總獲利不受影響一致[S-064]。本報告稱此為營收基礎的變動；緯穎則描述為採購代理模式，並將零組件成本上升列為其理由之一[S-063]。此一安排並未顯示所有客戶的記憶體均以此方式採購，亦未顯示出貨趨緩：緯穎預期出貨量將呈雙位數成長[S-064]。

? 未解問題——緯穎（Wiwynn）並未就脫離其營收的記憶體金額予以量化。 其財報發布與經核閱的期中財務報表均未提供該金額[S-063; S-483]。缺少該數字，2026年4月之後的成長率與毛利率均無法依先前基礎重編，而任何納入緯穎2026年數字的ODM營收比較，都混合了兩種基礎。

A ODM的籌碼（leverage）在於整合產能與地理位置，而非稀缺的投入要素。 Amazon與緯穎（Wiwynn）在緯穎位於德州索科羅（Socorro）廠房逾16億美元的擴建，使為Amazon基礎設施生產客製硬體的作業移入美國[S-306]；鴻海（Foxconn）表示其資本支出在台灣、美國、墨西哥與越南將持續上升[S-022]。圍繞單一買方建置的廠房，會將產能綁定於該買方，而共通的機櫃介面則使競標更易於比較（第4.5節）；即使數量成長，兩者皆限制了定價能力。ODM所出售的，是在接近客戶之處大量整合、測試與出貨液冷（liquid cooling）機櫃的能力。

IF ODM在數量上升之際維持微薄利潤率，而高價值零組件僅流經其帳上。 鴻海（Foxconn）與緯穎（Wiwynn）的陳述描述的是其自身的客戶與產品，並非ODM市場的代表；其他組裝廠則依其自身基礎公布數據[S-022; S-063]。若ODM毛利率在可比基礎上連續數季擴大，或若某家ODM表示限制其出貨者為組裝

表7。 沿AI基礎設施鏈條，價值池的性質與規模均不相同

配額式產能的供應商公布高利潤率，並提前出售該等產能；組裝廠則在龐大營收上賺取微薄利潤率，而該等營收有一部分僅是過帳。

層級	範例企業	各公司自行公布的經濟數據	商業模式說明	來源
加速器	NVIDIA	2027財年第3季（結束於2026年10月的該季）毛利率展望74.0%±50個基點	供應承諾於一個季度內自1,190億美元增至2,790億美元，主要用於記憶體	[S-010; S-002]
晶圓代工與先進封裝（advanced packaging）	台積電（TSMC）	2026年第3季財測指引：毛利率65%至67%、營業利益率56%至58%	產能於需求之前即完成配置；2026年資本預算上調至600億至640億美元	[S-084]
記憶體	SK hynix；Micron	SK hynix 2026年第2季營業利益率76%；Micron 2026財年第3季毛利率84.9%（2026年6月24日公布）	與約10家客戶簽有長期協議（long-term agreement）（SK hynix）；預期客戶保證金與承諾達220億美元（Micron）	[S-003; S-129]
設備與測試	ASML；Besi；Advantest	ASML 2026年第2季毛利率54.0%；Besi 2026年第2季毛利率65.7%；Advantest 2026財年預測：銷售額17,140億日圓、營業利益8,460億日圓	訂單隨客戶的產能計畫而動；Advantest將測試機產能與封裝產能相連結	[S-025; S-170; S-031; S-116]
機櫃組裝（ODM）	鴻海（Hon Hai／Foxconn）；緯穎（Wiwynn）	鴻海2026年第2季毛利率6.12%、營業利益率3.75%；緯穎2026年第2季毛利率9.3%、營業利益率7.3%	高價值零組件部分僅為過帳；緯穎自2026年4月起就特定客戶的記憶體以代理人基礎入帳；NRE收入會影響利潤率	[S-022; S-063]
功率半導體	Infineon	截至2026年6月當季分部業績利潤率（Segment Result Margin）19.1%；預期2026財年約為20%	多年期產能預留協議，並收取部分預付款	[S-021; S-193]
電力與冷卻設備	Vertiv；Eaton	Vertiv對2026年調整後營業利益率的財測指引為23.3%至24.3%；Eaton電氣事業未出貨訂單（backlog）於2026年6月年增43%	以未出貨訂單為主導；併購延伸至散熱與現地供電	[S-196; S-299; S-211; S-232]
渦輪機與電網設備	GE Vernova；Siemens Energy	GE Vernova於2026年6月總未出貨訂單（backlog）1,760億美元；Siemens Energy截至2026年6月當季訂單未出貨訂單1,620億歐元、訂單出貨比1.57	產能時段預留協議先於確定訂單	[S-012; S-300; S-301]

各層級範例企業最新公布或財測指引的經濟數據，依各公司自身的用語與期間呈現，資料截至2026年9月27日；來源列於最末欄。

範圍 本表僅示意價值池所在，並非跨公司的獲利能力排名。毛利率、營業利益率、分部利潤率與調整後利潤率的定義各不相同；所涉期間混有曆年季度與財年季度、實際數與財測指引；公司層級的利潤率涵蓋整家公司，而非其在單一機櫃價值中所占的份額。

產能而非零組件，則此一研判（reading）不成立。應觀察鴻海下一季財報中Vera Rubin機櫃與CPO交換器的數量，以及其他ODM是否就記憶體採用代理條件。

5.3 企業數字所涵蓋與未涵蓋者

一家公司的營收、展望與未出貨訂單（backlog）會被整條鏈條上下游解讀為需求訊號，但2026年有更多此類數字附帶了明示的界線：所涵蓋者看似需求卻並非需求，或未涵蓋者其實正是需求。一家組裝廠不再入帳的記憶體、另一家按成本收取的附加費、第三家得以回收的關稅、第四家已承諾但未入帳的數量：每一項都改變了公布的數字，而未改變零組件的實體流動。本節列出各公司在其自身申報文件與發布文件中明示的界線，以及各項界線對解讀供需的意義（表8）。

F 其他公司則將記憶體成本轉嫁收取，或將其納入財測指引。 寬頻系統製造商Calix表示，其記憶體成本的上升反映「the demand from the building of AI data centers」，其附加費僅為「cost recovery only」，且隨時間推移「the incremental memory costs will be neutral to gross profit」，惟仍對毛利率構成壓力[S-494]。KLA對截至2026年3月當季的毛利率財測指引，已納入其系統隨機出貨的影像處理電腦中「the rapidly escalating cost of DRAM chips」所帶來的增額影響，該公司將2026曆年的此一影響估為約75至100個基點，同時在其較長期預測中假設記憶體價格回歸正常[S-495]；4月時該公司仍認為「a roughly 100bps negative impact on gross margins over the next several quarters」[S-586]。

F 在機櫃系統整合商處，銷貨成本多為外購材料，其中GPU的比重日益提高。 Super Micro Computer截至2026年6月的財年淨銷售額自220億美元增至391億美元，毛利率則自11.1%降至10.8%[S-496]。其年報載明，GPU「have represented an increasing portion of our cost of sales」，銷貨成本包含關稅，而供應商折讓則使其減少[S-497]。Celestica 2026年1月的展望假設關稅或貿易限制並無重大變動，並載明「Substantially all tariffs paid by Celestica are expected to continue to be recovered from our customers」[S-498]；至7月，該假設則列於其前瞻性陳述之中[S-499]。

F 供應合約如今將未來數年的數量、保證金與價格區間一併固定。 Micron已簽訂16項策略客戶協議，通常自2026曆年起至2030年底，涵蓋該期間其約20%的DRAM數量與三分之一的NAND數量[S-129]。該等協議「structured as take-or-pay agreements」；其中最大者對既有產品一般訂有以2026曆年第2季市價為準的上限，並訂有下限；16項中有14項在其剩餘期間內以最低合約價格計算約有1,000億美元的營收[S-129]。TrendForce估計，HBM合約價格「could still rise by 70–140% in 2027」[S-007]。

F 部分未出貨訂單（backlog）並未納入已經承諾的需求。 Generac表示，其服務資料中心市場產品的未出貨訂單於2026年7月29日達約16億美元，「which does not include any committed volumes from the second hyperscale customer」，該公司已於6月與該客戶簽訂全球供應協議[S-500]。CoreWeave於6月30日約1,040億美元的營收未出貨訂單，並未納入其後數週新增的逾250億美元承諾（第2.2節）[S-011]。Oracle表示，其大型AI合約中預付與客戶提供硬體的部分「now total $75 billion」：前者由客戶就Oracle所購買的GPU向其預付；後者由客戶自行購買GPU並提供，而Oracle未就兩者的占比予以說明[S-501]。

A 公布的數字如今在被解讀為需求之前，須先經界線查核。 同一機櫃中的記憶體，可能以全額計入某一系統整合商的營收，而在另一家僅以手續費呈現；記憶體價格上漲可能推升某一供應商的營收，而在另一家僅推升成本；已承諾的數量亦可能落在未出貨訂單（backlog）之外[S-483; S-494; S-500]。因此本報告並列解讀五項指標：單位、位元與百萬瓦的實體流動；公布的營收；已簽約需求，例如未出貨訂單與取或付數量；資金來源，無論為資本支出、租賃或客戶提供的設備；以及出口與進口規則允許被滿足的需求（第7.1節）。五者相互補充，絕不相加。

A 一家公司將界線劃在何處，顯示由誰承擔記憶體的價格風險。 在緯穎（Wiwynn）的代理條件下，記憶體價格不再流經其營收與銷貨成本，使該風險留在客戶一方；在Calix的附加費下，客戶按成本支付漲幅中相當大的一部分；在Micron的取或付（take-or-pay）協議下，買方承諾數量與保證金，下限保護供應商、上限保護買方[S-063; S-494; S-129]。KLA預期其記憶體成本將回歸正常，Celestica預期能回收幾乎全部關稅；NVIDIA則吸收H200在美國境內受檢、獲許可出口至中國前所適用的25%美國進口關稅[S-495; S-498; S-499; S-503]。此一型態與A-12相符：掌握稀缺投入要素或大額訂單者，會將價格風險移轉給他人。

IF 唯有上述界線不再變動，公布的營收才會重新貼合實體需求。 此一研判（reading）須以代理、附加費與取或付等安排維持其所揭露的狀態為前提[S-483; S-494; S-129]。應觀察其他ODM是否就記憶體採用代理條件、Micron是否履行其所計畫的後續協議、Calix的附加費是否隨記憶體價格鬆動而下降，以及Generac第二家超大規模雲端業者的數量是否於其下一次財報中計入未出貨訂單[S-500]。若企業將先前期間依單一基礎重編，或就所排除的金額予以量化，使公布的數字得以再次直接比較，則此一研判不成立。

5.4 電力、冷卻與光學零組件均為預售

每一台出貨的機櫃都需要電力轉換、電路保護、液冷（liquid cooling）硬體，以及日益不可或缺的光鏈路。這些零組件的成本低於其所服務的加速器（accelerator），但缺少其中任一項，對機櫃的阻滯與缺少一顆晶片一樣確實。2026年，其中數家供應商開始以記憶體製造商的方式出售產能，透過預留、預付款與多年期協議進行；另有供應商則以併購切入相鄰產品。以下的訂單簿均為各公司自身的衡量方式；其定義與日期各不相同，既不相加，亦不排名。

F Vertiv於2025年底的未出貨訂單（backlog）為150億美元，其後並上調其2026年財測指引。 其未出貨訂單一年內成長109%，2025年第4季的訂單出貨比約為2.9[S-307]。2026年7月，該公司將全年財測指引的所有關鍵指標上調，淨銷售額上調至138億至142億美元，調整後營業利益率上調至23.3%至24.3%，理由為「growing pipelines」與持續的產能擴充，但未提供新的未出貨訂單或訂單數字[S-196]。其第2季簡報顯示在台灣高雄一處場址推進800伏特直流（800 VDC）驗證的路徑[S-196]。

F Eaton與Schneider Electric公布資料中心業務強勁，但未提供資料中心的訂單數字。 截至2026年6月底，Eaton電氣部門的未出貨訂單（backlog）年增43%，其電氣業務以滾動十二個月計算的訂單出貨比為1.2；Electrical Global的未出貨訂單成長103%，可能包含3月所收購的Boyd散熱事業，故不能解讀為有機成長[S-299; S-211]。Schneider Electric表示其Data Center & Networks終端市場的需求於第2季「up triple-digit」，但未說明所指為訂單或營收，並預期該業務將維持「strong momentum, supported by a robust backlog」[S-308]。

F 供應商正透過併購擴大其在散熱與電氣產品的範圍。 Eaton於2026年3月12日完成以95億美元收購Boyd的散熱事業[S-211]。Vertiv收購晶片直接液冷（direct-to-chip cooling）冷板（cold plate）專業廠商Strategic Thermal Labs[S-309]，並同意以14.5億美元現金外加或有對價（earn-out），收購為資料中心設計微電網與表後（behind the meter）供電的UtilityInnovation Group[S-232]。Vertiv就第二筆交易的說明，著眼於客戶能多快走完「from site selection to first token」的歷程[S-232]。

表8。 企業數字的涵蓋範圍各不相同

在將企業數字解讀為需求之前，須先查核其界線：目前已有數家公司的數字未納入真實需求，或納入了僅為過帳的金額。

公司（層級）	數字	公司所明示的界線	對解讀需求的影響	來源
緯穎（機櫃ODM）	自2026年4月起的營收與銷貨成本	特定客戶的記憶體採採購代理模式，相關金額不計入營收與銷貨成本；財務報表載明：就客戶指定供應商所提供的原材料為代理人；未提供金額	依IFRS 15之代理人處理：營收低估已出貨機櫃中的記憶體；毛利率可能在獲利不變的情況下上升	[S-063; S-483; S-484]
NVIDIA（加速器）	自2026年2月起的營收展望	假設不含來自中國的Data Center運算營收	實際發生的獲許可出貨不在展望之內：於2027財年第2季占Data Center營收不到1%	[S-010; S-002; S-503]
AMD（加速器）	2025年第3季至2026年第3季的展望	排除對中國的MI308銷售（2025年第3與第4季），其後納入約1億美元（2026年第1季）；此後未再附中國相關的限定說明	單一產品的中國銷售退出財測指引後又重新納入；重新納入未載明依據	[S-485; S-502; S-486; S-111]
KLA（設備）	2025年10月29日針對2025年12月當季與2026曆年的展望	納入因美國管制擴大而失去中國部分客戶市場通路的營收影響，約3億至3.5億美元；於所閱讀的後續股東信中未予重編	此一觀點已扣除一項出口規則影響的估計值	[S-487; S-587]
KLA（設備）	2026年3月當季的毛利率財測指引	納入其影像處理電腦中DRAM成本攀升的影響，2026年為75至100個基點；2026年4月則為未來數季約100個基點	記憶體價格已影響設備製造商的利潤率	[S-495; S-586]
Calix（網路系統）	營收與毛利率	記憶體附加費僅為回收成本	營收上升而毛利並未增加	[S-494]
Celestica（製造服務）	2026年展望	假設關稅並無重大變動；幾乎全部關稅均自客戶回收	利潤率取決於能否將關稅轉嫁	[S-498; S-499]
Super Micro（機櫃系統整合商）	2026財年淨銷售額	GPU占銷貨成本的比重上升；銷貨成本含關稅；供應商折讓使其減少	營收成長混合了過帳零組件的價值與自身的附加價值	[S-497; S-496]
Micron（記憶體）	策略客戶協議	取或付（take-or-pay），2026年至2030年；訂有下限與上限；14項協議以最低價格計算約1,000億美元	需求於數年前即已簽約，價格風險由雙方分擔	[S-129]
Generac（電力設備）	資料中心未出貨訂單（backlog），約16億美元	不含來自第二家超大規模雲端業者客戶的已承諾數量	未出貨訂單低估了已承諾的需求	[S-500]
Oracle（雲端）	大型AI合約	客戶預付或客戶提供的GPU，750億美元；未說明兩者占比	部分已簽約產能由客戶出資，而非由Oracle自有資金支應	[S-501]
Microsoft（雲端）	自2027財年起的資本支出（capital expenditure）	更多資料中心租賃被歸類為營業租賃，不計入資本支出	公布的資本支出下降，而投資並未減少	[S-075]

各公司在其自身申報文件與發布文件中明示的界線，依其用語呈現，資料截至2026年9月27日；來源列於最末欄。

範圍 本表為範例，並非所有公司的普查。影響欄為本報告的研判（reading）（A）。

美元數字均依各公司所述；所涉期間與定義各不相同，各列彼此不可比。

F Infineon正以多年期預留方式出售功率半導體產能。 AI資料中心生態系中數家主要客戶已與Infineon簽訂或正在洽談多年期產能預留協議，累計銷售量達「a high single-digit billion euro amount」，並附帶部分預付款[S-021; S-193]。Infineon預期其2026財年（截至2026年9月30日的年度）專用於AI電力的營收將逾16億歐元，高於其原先計畫的15億歐元[S-193; S-310]。其電源管理產品將支援NVIDIA的MGX與800 VDC架構，並以原生800 V伺服器板上的保護與熱插拔功能為目標[S-192]。

> **觀察重點**
>
> **2026年11月**
>
> Infineon將於其2026年11月的年度財測指引中，向上修正其對2027財年AI資料中心營收「2.5 billion Euros plus」的預估，並預期該次更新具重大性[S-193]。

F 雷射供應是光學領域中有紀錄可循的約束。 據Tom's Hardware於2026年7月31日報導，Lumentum執行長在該公司2026年5月的法人說明會上向分析師表示，電吸收調變雷射（electro-absorption modulated laser, EML）的供需失衡已擴大至「somewhere greater than 30%」[S-291]。TrendForce報導同一位高階主管表示，磷化銦（indium phosphide, InP）的供需缺口如今已超過動態隨機存取記憶體（dynamic random-access memory, DRAM）與NAND快閃記憶體；該機構並報導NVIDIA於2026年3月對兩家未指名的供應商各承諾20億美元，且將額外壓力歸因於中國自2025年起加強的InP與銦出口管制（export control），中國約占精煉銦產量的70%[S-037]。TrendForce將光學引擎良率（yield）列為共同封裝光學（co-packaged optics）交換器的瓶頸之一[S-282]。

A 零組件供應商如今如同記憶體製造商一般，將稀缺寫入合約。 Infineon附預付款的產能預留、Vertiv與Eaton的大額未出貨訂單（backlog），以及買方對雷射供應商的直接承諾，將供應不足風險的一部分自買方的時程轉移至供應商的訂單簿，同時使供應商取得投資所需的能見度[S-193; S-307; S-037]。將冷板（cold plate）、冷卻液分配單元（coolant distribution unit, CDU）與現地供電整合在一起的併購，使買方面對的是數量更少、規模更大的供應商[S-211; S-208; S-309; S-232]。對買方而言，實際的後果是電力、冷卻與光學零組件必須與記憶體依同一份前瞻日程下訂，而不能臨時自現貨採買。

IF 唯有新增產能超越機櫃的爬坡速度，零組件的緊俏才會緩解。 「電力、冷卻與光學零組件在2027年之前持續採配額供應」此一研判（reading），取決於AI機櫃出貨量於2027年維持強勁——鴻海（Foxconn）預期屆時客戶對AI產能的需求仍將「very strong」[S-022]——亦取決於800 VDC設計是否使每一機櫃的電力設備含量增加（第4.2節）。若Vertiv或Eaton公布訂單出貨比下滑、若新的產能協議不再附帶預付款，或若雷射製造商表示出貨已能滿足需求，則此一研判即被削弱。應觀察Infineon 11月的財測指引、雷射製造商下一季的陳述，以及Vertiv是否恢復揭露其未出貨訂單。

5.5 渦輪機與電網設備依更長的時鐘運行

零組件一旦到位，一台機櫃可在數週內安裝完成；而為其送電的發電廠與電網併接則需以年計。燃氣渦輪機、變壓器與開關設備出自產出增加緩慢的生產線，其製造商如今更在客戶下確定訂單之前，即出售未來的生產時段。其揭露中出現四項指標，且必須彼此區分：計畫中的工廠產出、透過時段協議預留的產能、以確定未出貨訂單（backlog）簽約的產能，以及已出貨的設備。上述任一項均未衡量已交付至資料中心的發電量。

F GE Vernova正逐步上調其計畫中的燃氣渦輪機產出，目標至2030年。 該公司表示，其仍如期於2026年第3季達到年產20吉瓦（gigawatt, GW）燃氣渦輪機的產出，2028年達24 GW，並正「implementing actions to produce 30 GW in 2030」[S-012]。其燃氣設備未出貨訂單（backlog）與產能時段預留協議於第2季自100 GW成長至116 GW，並預期至2026年底簽約量至少達125 GW[S-012]。該116 GW係將已簽約的確定未出貨訂單與已預留時段合計而成。

F GE Vernova就其Electrification事業的資料中心訂單提出量化數字。 其Electrification事業的資料中心訂單於2026年上半年逾50億美元，較其2025年全年總額高出一倍以上[S-012]。總未出貨訂單（backlog）於2026年6月底達1,760億美元[S-012]；該公司於2025年12月表示，預期至2028年底約達2,000億美元，其中Electrification的未出貨訂單將自300億美元倍增至600億美元[S-039]。Utility Dive報導，GE Vernova正接受2031年交貨的預留[S-311]。

F Siemens Energy描述其燃氣渦輪機產能已售罄至2028財年。 2026年3月，其投資人關係負責人表示燃氣渦輪機產能已售罄至2028財年（Siemens Energy的財年於9月30日結束），2029財年的時段正「filling up very quickly」，部分2030財年的時段亦已預訂[S-018]。6月，他描述該公司已「booked out onto '28」，並認為未來數年市場平均約為每年110至120 GW，此為公司的市場看法而非財測指引[S-026]。於2026年6月30日，Siemens Energy持有69 GW的確定未出貨訂單（backlog）與26 GW的產能時段預留協議（slot reservation agreement, SRA）[S-301]。

> “From a structural perspective, the market remains supply-constrained rather than demand-constrained, with strong backlog visibility and a growing number of commitments including reservation agreements.” [S-026]

= 以相同基礎計算，兩家製造商於2026年6月30日的燃氣渦輪機合計承諾量分別為116 GW與95 GW。 GE Vernova將確定未出貨訂單（backlog）與產能時段預留合併為一個數字，即116 GW[S-012]；Siemens Energy則分別列示，故其可比數字為已簽約69 GW加上已預留26 GW，合計95 GW[S-301]。公式：已簽約＋已預留＝合計承諾量；兩家公司僅在此一基礎上並列對照，絕不以確定未出貨訂單單項相互比較。就產出而言，116 GW除以每年20 GW，相當於GE Vernova依其2026年第3季產速計算的5.8年產量[S-012]，此為上限：產出計畫於2028年提高至24 GW，且並非每一項預留都會轉為訂單。

F 與資料中心相關的電網設備訂單正在上升。 Siemens Energy於2026財年上半年在其電網業務取得約20億歐元與資料中心相關的訂單，幾乎相當於2025財年全年的規模[S-026]，並表示其變壓器業務（包含資料中心專案）的成長帶動了第3季的電網訂單[S-300]。其集團訂單未出貨訂單（backlog）於2026年6月30日達1,620億歐元，訂單出貨比為1.57[S-300]。該公司認為電網領域將出現「a once-in-a-generation investment cycle over the next fifteen years」[S-018]。

A 發電與電網設備如今決定AI建置部分環節的步調。 機櫃以季為單位規劃：鴻海（Foxconn）於2026年第3季內將Vera Rubin機櫃投入量產[S-022]。渦輪機時段正為2029年至2031年的交貨而預訂[S-018; S-311]，電網投資則以十五年為框架[S-018]（圖2）。加速器（accelerator）與機櫃均如期到位的營運商，仍可能等待一台渦輪機、一台變壓器或一項併接。在Siemens Energy所描述的市場——「supply-constrained rather than demand-constrained」——配置供給者為預留及其轉為訂單的情形，而非定價表[S-026]。

IF 唯有新增產出超越新增預留，渦輪機與電網的交期（lead time）才會縮短。 「發電設備在2028年之前仍是具約束力的時鐘」此一研判（reading），須以合計未出貨訂單（backlog）在產出爬坡之際持續上升為前提。若GE Vernova的燃氣合計未出貨訂單在產出達到每年20 GW後即停止成長、若Siemens Energy公布2029或2030財年仍有未售出的時段，或若營運商轉向繞過渦輪機的現地電源，則此一研判不成立。應觀察GE Vernova的年末數字對照其至少125 GW的預期，以及2029財年是否如Siemens Energy管理層所預期，於年底前達到「largely sold-out」[S-012; S-018]。

5.6 誰能讓一項部署停擺

這條鏈條中的議價能力（bargaining power），透過一個問題最容易看清：誰的拒絕會讓一項部署停擺？當一方掌握他人在既有時間內無法繞道的環節時，該方即取得籌碼（leverage）。其中部分環節屬實體，例如記憶體晶圓、封裝產線與渦輪機時段；部分屬行政，例如驗證（qualification）清單、安全證書或電網併接研究；另有一項屬財務。圖24自下單延伸至已送電的產能；其後各段指名各環節由誰掌握，以及何者會削弱該項掌握。

A 記憶體製造商決定誰的加速器（accelerator）能夠出貨。 SK hynix描述的記憶體市場是「customer demand exceeds supply capabilities」，且依請求數量及時交貨已成為核心競爭因素，該公司並透過與約10家客戶簽訂的長期協議（long-term agreement, LTA）配置供給[S-003]。NVIDIA的供應承諾在一個季度內自1,190億美元增至2,790億美元，「primarily related to the procurement of memory」[S-002]。TrendForce報導，數家美國雲端業者持有多年期DRAM協議，限制供應商調漲價格[S-132]，故價格上漲落在未簽有此類協議的買方身上。不在配置之內的買方只能等待、付出更高價格，或重新設計（第3.2節）。

A 封裝與測試產能決定能有多少加速器（accelerator）完工。 TrendForce於2026年6月報導，Chip-on-Wafer-on-Substrate（CoWoS）封裝的供需缺口預期將自約20%縮小至2026年底的約10%，但需求仍高於供給[S-033]。台積電（TSMC）較大尺寸的規格依藍圖推出：2028年為14倍光罩尺寸CoWoS，2029年逾14倍光罩尺寸[S-024]。Advantest將其加速器測試機的產能規劃與晶圓代工廠及委外半導體封裝測試（outsourced semiconductor assembly and test, OSAT）業者的封裝產能擴充相連結，並將封裝約束列為其外部不確定因素之一[S-116; S-031]。誰配置封裝，誰就配置加速器。

A 客製晶片為加速器（accelerator）環節開啟第二條路徑，卻無法繞過記憶體與封裝。 TrendForce預測，特殊應用積體電路（application-specific integrated circuit, ASIC）將占2026年AI伺服器出貨量的27.8%，圖形處理器（graphics processing unit, GPU）則占69.7%（圖12）[S-008]。客製設計仍需高頻寬記憶體（high-bandwidth memory, HBM）與先進封裝（advanced packaging）：Samsung與Broadcom之間的備忘錄涵蓋為Broadcom次世代AI加速器計畫供應記憶體（含HBM）[S-115]；Advantest並認為GPU、客製ASIC與中央處理器（central processing unit, CPU）的測試機需求強勁[S-116]。自行設計加速器的買方，對單一市售供應商的配置依賴較低；但其在記憶體、封裝、測試與電力上仍排在同一條佇列中。

A 驗證（qualification）清單如今左右哪些電力與冷卻產品能進入買方的設計。 NVIDIA於2026年9月21日推出的DSX Ready計畫，為AI工廠（AI factory）驗證電力與冷卻產品；該計畫先納入Hitachi Energy、LG Energy Solution與Tesla的電池儲能系統（energy storage system），以及LG Electronics、LiquidStack與Vertiv的CDU[S-048]。LG Electronics表示其2.6 MW CDU符合NVIDIA DSX參考設計（reference design）的要求[S-209]。Open Compute Project的OCP Accepted認可則是開放途徑，任何依已公布規格（specification）建置的供應商均可取得（第6.1節）。未列入買方所用清單的產品並不稀缺，但在通過驗證之前無法進入該買方的設計，這使清單的擁有者對供應商的上市時程握有籌碼（leverage）。

A 經認證（certification）的保護硬體是邁向更高電壓供電的關卡。 Google、Microsoft與NVIDIA的800 VDC更新文件將安全認證與法規框架列為「prerequisites for global deployment」[S-019]。ABB自述其擁有全球首款依國際電工委員會（International Electrotechnical Commission, IEC）標準完整認證的固態斷路器（solid-state circuit breaker, SSCB）[S-191]；而SolarEdge以Infineon碳化矽（silicon carbide）元件設計的SSCB則屬開發合作，未載明供貨日期[S-055]。在多家供應商的保護裝置均取得800 VDC認證之前，少數已認證的產品與為其認證的機構將決定採用的步調（第4.2節；第7.4節）。

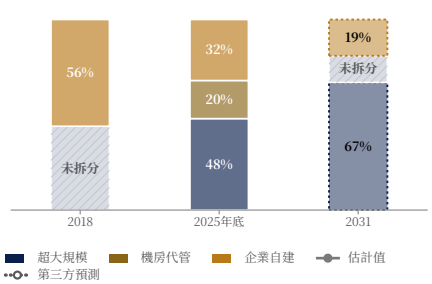
A 電網業者握有任何供應商都無法推翻的拒絕權。 德州電力可靠度委員會（Electric Reliability Council of Texas, ERCOT）於2026年6月追蹤逾438,000 MW的大型負載（large load）併接申請，其中近89%來自資料中心，並依其監管機關於2026年6月18日核准的程序，將75 MW以上的申請併同研究[S-032]。PJM於2026年8月13日提交的文件，將自帶供給的大型負載納入單一框架併接，其餘則置於一項過渡性服務之下，於緊急情況時其用電可較其他用戶優先被削減[S-042]。以現地發電避開電網佇列的開發商，則轉而加入渦輪機的佇列（第5.5節）。並無任何OCP規格（specification）能拘束電網業者（第7.2節）。

A 資本是最後一個停擺點，且其高度集中。 CoreWeave 2025年營收中約67%來自Microsoft[S-065]；Nebius的首筆擔保債務則以GPU基礎設施及「contracted cash flows from an agreement with an investment-grade customer」作為擔保[S-066]。因此放款人係以少數買方的合約為依據為產能提供融資，使較小型雲端營運商能否建置，取決於買方的信用，而非僅取決於硬體（第2.2節）。國際能源總署（International Energy Agency, IEA）表示，資料中心投資「have grown too large to be funded from company balance sheets alone」[S-005]。

F 買方與供應商同樣正在集中。 Synergy Research Group報告指出，超大規模雲端業者（hyperscaler）於2025年底擁有1,360座大型資料中心，持有全球資料中心容量的48%，非超大規模的機房代管業者占20%，企業自有設施占32%，而後者在2018年曾占56%；該機構預測超大規模雲端業者於2031年將占容量的67%，企業自建設施則占19%[S-612]。Amazon、Microsoft與Google合計占全部超大規模容量的58%，而每季的超大規模資本支出（capital expenditure）三年內成長近180%，於2025年第3季達1,420億美元[S-613]（圖23）。

A 雙方同時集中，使議價呈現雙邊性質。 少數記憶體、封裝與電力供應商掌握稀缺產能，而少數買方則持有需要該產能的容量中龐大且持續上升的份額：供應商進行配置，而買方則能以他人無法企及的規模預付款、共同投資並驗證（qualification）第二供應來源（second source）[S-612; S-003]。同一集中現象亦說明了第3.2節的長期協議（long-term agreement）與第2.2節的預付款。這也意味著少數幾家買方的修正即會牽動整條鏈條，因此首份載有具體數字的2027年資本支出（capital expenditure）財測指引即為一項路標（signpost）（第9.1節）。

圖23。 超大規模雲端業者持有近半數的資料中心容量，並預測至2031年將持有三分之二



依營運商類型劃分的全球資料中心容量占比（%），Synergy Research Group，2026年4月7日；2018年與2025年底為已公布數字，2031年為預測[S-612]。Synergy未提供2018年的超大規模占比與2031年的機房代管占比；該等部分以未拆分方式繪示。

範圍 係實體容量的占比，而非支出的占比。超大規模容量包含承租設施，於2025年底約占其40%。

A 價值正遷移至持續稀缺的環節。 能讓一項部署停擺的各方，掌握的是已配置的記憶體、封裝或測試產能、一份驗證（qualification）清單、一項經認證的保護裝置、一個渦輪機時段、一項電網併接，或一紙具信用的合約；而這些也正是如今收取保證金、預留或預付款的各方[S-129; S-193; S-012]。機櫃組裝——鴻海（Foxconn）預期其AI機櫃出貨量將於一年內成長逾一倍[S-022]——在2026年的紀錄中並不在其列。約束已自加速器（accelerator）擴及其周邊的整個系統（圖3），價值亦隨之移動；A-12列出此一研判（reading）、其替代解釋，以及何者會削弱該研判。

> **重要性何在**
> 一份只確保了加速器（accelerator），而未確保記憶體、封裝時段、已驗證的電力與冷卻零組件、渦輪機時段或電網時程的採購計畫，所確保的其實是最不稀缺的那個環節。籌碼（leverage）屬於掌握買方時程上最長佇列者。

IF 若稀缺發生遷移，停擺點亦會移動。 在記憶體與封裝持續採配額供應、電網併接持續排隊期間，此一圖譜成立。若伺服器DRAM合約價格連續數季下跌，且HBM在多家供應商、一個以上平台完成驗證（qualification），使記憶體環節鬆動；若多家供應商推出經認證的800 VDC保護裝置；或若大型負載（large load）規則給予資料中心確定的併接日期，此一圖譜即會改變。應觀察記憶體協議與保證金、CoWoS產能的陳述、DSX Ready與OCP Accepted的列名，以及PJM計畫於2026年11月提交的大型負載故障穿越（ride-through）標準[S-061]。

5.7 沒有任何區域掌握所有層級

沒有任何國家或區域掌握AI基礎設施鏈條的所有層級。加速器（accelerator）設計、資本與大部分需求位於美國；最先進晶圓、先進封裝（advanced packaging）與機櫃整合位於台灣；HBM位於韓國；測試設備、晶圓製程設備與快閃儲存位於日本；微影、鍵合設備、功率半導體與電網設備位於歐洲；而中國則有一套平行的國內堆疊。波斯灣國家與印度則透過主權層級的建置增添需求。圖25依企業與政府的揭露解讀各區域；其描述的是依賴關係，而非排名或國家實力。

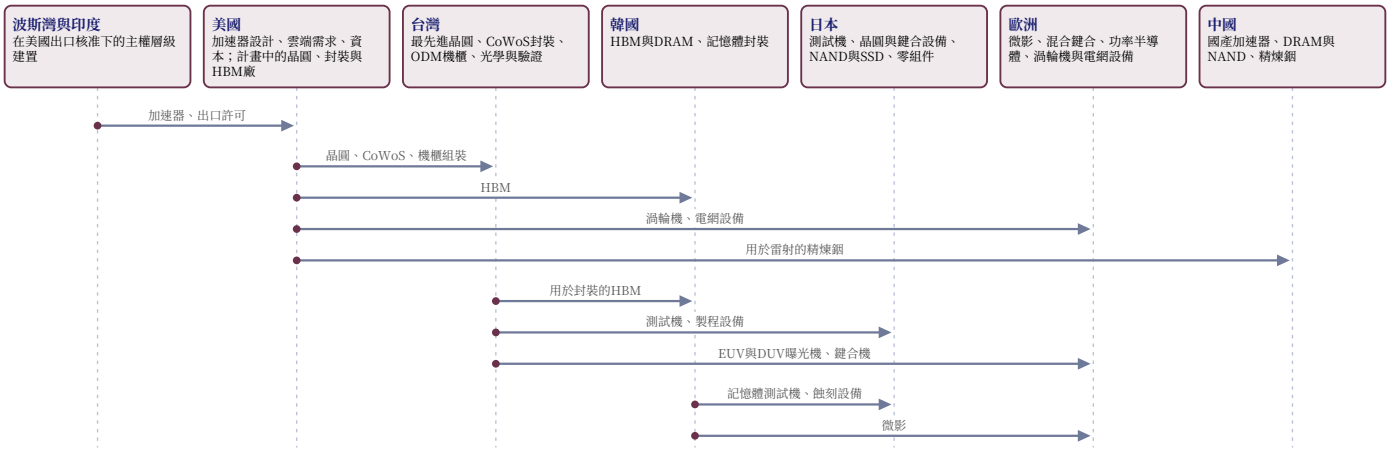
A 美國掌握加速器（accelerator）設計、雲端需求與資本，並正將實體鏈條的部分環節移回國內。 TrendForce估計，北美前五大雲端業者占2026年NVIDIA GB系列與VR系列伺服器需求的逾60%[S-114]。台積電（TSMC）於2026年7月16日宣布在亞利桑那州追加投資1,000億美元，美國商務部則認定其投資總額

圖24。 除加速器供應商外，另有數方能讓一項AI部署停擺

一項AI部署自採購訂單至已送電產能所經過的各環節，分為三條平行通道並於送電時匯合，並列出掌握各環節的一方。標示強調的環節，係有註明日期的2026年來源顯示掌握方正在配置、預留、驗證或排隊處理供給者：記憶體[S-003; S-002]；驗證（qualification）[S-048]；保護裝置認證（certification）[S-019]；資本[S-065; S-066]；渦輪機時段[S-012; S-026]；電網併接[S-032; S-042]。封裝與測試之所以標示強調，係因據報導CoWoS需求高於供給，且Advantest正提前其測試機產能目標[S-033; S-116]。

範圍 其排序顯示的是依賴關係，而非時間、規模或排名。未標示強調的環節並無此類紀錄，但這並非產能有餘裕的證據。

圖25。 各區域均依賴他處所掌握的層級



AI基礎設施鏈條中主要的跨區域依賴關係，資料截至2026年9月。方框列出該區域企業所掌握的層級。每一列為一項依賴關係：箭頭自依賴方區域（圓點）指向其所依賴的區域（箭頭端），並在其上方標明所依賴的對象。來源：美國[S-114; S-163; S-023]；台灣[S-024; S-022]；韓國[S-003; S-056]；日本[S-116; S-031; S-172; S-130]；歐洲[S-025; S-170; S-021; S-300]；中國[S-312; S-313; S-037]；波斯灣與印度[S-314; S-315; S-316]。

範圍 箭頭顯示的是依賴關係，而非規模、份額或排名，亦未涵蓋所有流向；一個區域可以掌握某一層級，同時仍進口該層級的產品。地點係依企業製造或計畫製造該項目之處認定。

為2,650億美元[S-084; S-163]。SK hynix計畫於2029年下半年在印第安納州量產次世代HBM，並於2028年10月前完成無塵室[S-023]；Amkor則已將其計畫中的亞利桑那州封裝測試園區規模提高至120億美元[S-317]。

A 台灣的角色正從「組裝」這一簡略代稱，轉向系統整合、先進封裝（advanced packaging）、光學與驗證。 台積電（TSMC）的封裝藍圖決定了加速器（accelerator）的尺寸上限：2028年為14倍光罩尺寸的CoWoS，2029年為40倍光罩尺寸的System-on-Wafer形式，而其基板上COUPE共同封裝光學（co-packaged optics, CPO）已進入量產[S-024; S-084]。鴻海（Foxconn）將於2026年第3季使Vera Rubin機櫃與CPO全光交換器進入量產[S-022]；Vertiv則正於高雄一處場址驗證800 VDC路徑[S-196]。聯華電子（UMC）正與數家新客戶合作開發用於光收發模組的光子積體電路，風險試產排定於2026年與2027年[S-318]。

F 友達光電（AUO）已自鄰接的顯示器業務，轉入直接載明的AI基礎設施技術定位。 友達於2026年8月31日宣布，將在SEMICON Taiwan（2026年9月2日至4日）展出一款正與生態系夥伴共同開發的系統級Micro LED共同封裝光學（co-packaged optics）模組：發射器來自集團企業Ennostar、接收器來自Tyntek、光纖來自Corning，而來自Daxin的介電材料則「expected to be incorporated」[S-052; S-169]。該模組的目標應用為AI資料中心內最長10公尺的短距互連；友達並表示將與Corning共同展示用於更大AI伺服器處理器晶粒、高密度互連與低損耗訊號傳輸的玻璃核心基板（glass-core substrate）封裝[S-052]。其階段為研發（R&D）。

A 問題如今在於可製造性與驗證（qualification），而非友達是否應被列入版圖。 友達將巨量轉移、重佈線層（redistribution layer, RDL）封裝、光耦合與系統架構設計列為其所帶來的能力，並自述正在「accelerating the commercialization」該技術[S-052]；該發布文件並未提及量產、客戶、設計導入、營收或OCP採用情形。應觀察客戶驗證、量產規模、連結效能、每位元功耗、良率（yield）、封裝生態系的採用情況，以及是否進入具名的伺服器或網路平台。

A 韓國供應著決定加速器（accelerator）產出步調的記憶體。 SK hynix於2026年第2季開始大量出貨HBM4，並於上半年完成HBM4E樣品出貨，且在2027年初龍仁第1期無塵室啟用之前，正加速其M15X晶圓廠的量產時程[S-003]。Samsung於2026年2月表示，預期其HBM銷售額於2026年將增至逾三倍；至其第2季財報時，已向主要客戶出貨HBM4E樣品[S-056; S-127]。韓國的記憶體產出又取決於他處製造的設備：ASML將其系統的需求與AI帶動的先進邏輯與記憶體晶片需求相連結[S-025]，而Advantest則因高效能DRAM需求轉強，上調其記憶體測試機市場預測[S-031]。

A 日本掌握上層各層級所倚賴的測試、製程設備、快閃儲存與零組件。 Advantest認為其測試機在GPU、客製ASIC與中央處理器（central processing unit, CPU）各領域的需求均屬強勁，並正提前其產能目標：2026財年結束前（2027年3月）達5,000台系統單晶片（system-on-chip, SoC）測試機，2028年底或2029年初達10,000台[S-116]。Tokyo Electron預期其2027財年（2026年4月至2027年3月）的先進封裝（advanced packaging）銷售額將顯著超過1,000億日圓[S-172]。Kioxia計畫至2028財年使資料中心產品占其營收逾60%，並正將客戶轉向多年期長期協議（long-term agreement, LTA）[S-130]。TrendForce報告指出，日本與韓國的電容器製造商正縮減消費性產品供應，以聚焦於AI[S-319]。

A 歐洲供應微影設備、鍵合設備、功率半導體、發電設備與光子學研究。 ASML計畫在其2026年約65台低數值孔徑（low-NA）極紫外光（extreme ultraviolet, EUV）與約130台深紫外光（deep ultraviolet, DUV）浸潤式系統的產能之上，於2027年增加30%，並正研議於2028年再增加30%[S-025]。Besi的混合鍵合（hybrid bonding）客戶數在2026年上半年自15家增至21家[S-170]。Siemens Energy將美國需求（包含大型資料中心訂單）列為其燃氣渦輪機訂單的推動因素之一[S-300]；研究機構imec則已將其iSiPP300矽光子（silicon photonics）製程授權（licence）予台灣的聯華電子（UMC）[S-320; S-318]。

A 中國正在出口管制（export control）的限制範圍內建立一套平行堆疊。 Huawei的藍圖將其Ascend 950DT加速器（accelerator）定於2026年第4季，並已將其UnifiedBus 2.0垂直擴展互連的規格（specification）以開放協定形式發布[S-312]。CXMT於2026年7月在上海科創板掛牌；據Tom's Hardware報導，其公開說明書並未載有專門的HBM計畫，所列具名支出大多投向DRAM晶圓產線與製程[S-313; S-321]。依Counterpoint Research的資料、經Tom's Hardware報導，YMTC於2026年第2季出貨了全球14%的NAND快閃記憶體，此係出貨量份額而非營收份額[S-322]。中國在精煉銅的份額，使其與雷射限制相連（第5.4節）[S-037]；出口規則見第7.1節。

A 波斯灣國家、印度與歐盟增添了主權需求；在波斯灣，此一需求取決於美國的出口核准。 美國商務部於2025年11月授權（authorisation）阿拉伯聯合大公國（United Arab Emirates, UAE）的G42與沙烏地阿拉伯的HUMAIN購買相當於最多35,000顆NVIDIA GB300晶片[S-314]；2026年7月的一項規則放寬了對阿聯的許可作業，而G42與Core42的授權若無進一步通知，將於2027年4月6日到期[S-315]。印度政府於2026年2月表示，將在其共用運算容量既有的38,000顆GPU之外再增加20,000顆[S-316]。歐洲聯盟（European Union, EU）計畫動員200億歐元，用於最多五座AI Gigafactories[S-323]。

A 沒有任何區域能夠獨力打造一整座完整的AI機櫃。 一款交換器即可說明此點：業界媒體報導，NVIDIA的Spectrum-X Photonics供應鏈經由台積電（TSMC）取得矽光子（silicon photonics）、由SPIL負責封裝與測試、由Lumentum提供雷射晶片、由TFC Communication提供雷射模組，並由鴻海（Foxconn）進行組裝[S-324]。美國的本土化生產改變了部分環節在2027年至2029年間的發生地點[S-083; S-023]，但在此之前並未改變對韓國記憶體、歐洲微影設備或日本測試的倚賴。因此，韌性問題在於買方能以合約分散哪些依賴（記憶體靠協議、光學靠第二供

應來源、組裝靠第二家原始設計製造商（original design manufacturer, ODM）），以及哪些無法分散，例如微影設備、電網併接與出口許可（第7.5節）。

IF **區域版圖唯有在某一區域以量產規模補上一個層級時才會改變。** 若台積電（TSMC）第二座亞利桑那晶圓廠如期於2027年下半年達到量產、亞利桑那封裝產能爬坡，且SK hynix印第安納州HBM於2029年投產，則美國加速器（accelerator）設計業者對台灣與韓國的依賴將會收窄[S-083; S-023]。反之，若出口規則擴及設備或材料，或中國在精煉鎵的份額成為雷射產出的約束性限制，則該依賴將更形緊縮[S-037]。應觀察表9中附日期的各項里程碑——每一項均為企業或政府自身的承諾——以及它們是否成立。

5.8 資本追隨瓶頸

資本追隨稀缺，但各依不同時鐘。記憶體製造商可以加快一座已在興建中的晶圓廠；晶圓代工廠的新封裝形式則依藍圖在兩至三年後到來；渦輪機製造商每次增添的產出以數吉瓦（gigawatt, GW）計。表10逐層列出各企業所稱將於一年內、三年內及其後增添的內容，並各依其陳述的類別加以標示。橫向對讀，即可看出投資在何處一致、在何處分歧，以及其到來的先後順序。

A **買方、晶圓代工廠、記憶體製造商與電力設備供應商均在提高投資。** 台積電（TSMC）將其2026年資本預算上調至600億至640億美元，並表示現在預期未來三年的支出將比過去三年「even more significantly higher」[S-084]。SK hynix正加速M15X，並將在龍仁於2027年初啟用後擴充產能[S-003]；ASML計畫於2027年將低數值孔徑（low-NA）EUV與DUV浸潤式產能再增30%[S-025]；GE Vernova正將規劃中的渦輪機產出提升至2030年的30 GW[S-012]。各家企業均將該項投資與其所描述為強勁的需求相連結，而在記憶體與燃氣渦輪機方面則描述為需求超過供給[S-003; S-026]。

A **產能並非均勻地追隨需求。** 部分擴產屬確定計畫，部分則屬選擇權：SK hynix將「in phases based on customer demand and investment efficiency」執行其中長期專案[S-003]，ASML的2028年推進步驟係在研議中而非已規劃[S-025]，而Besi的長期目標則未載明年度[S-057]。渦輪機產出每年增加數吉瓦，預訂量則已排至2031年[S-012; S-311]；某一主要雷射製造商的出貨量低於需求逾30%[S-291]。依分期並依預訂量增添產能的供應商，會透過保證金、預付款與最低價格，將一部分需求風險轉移給買方[S-129; S-193]。

A **這波投資依序到來，自資本預算到電網併接。** 紀錄中可見的順序，自買方的資本預算（第2.1節）起，到記憶體協議與保證金[S-003; S-129]，再到封裝產能及與之相連的測試機[S-033; S-116]，再到原始設計製造商（original design manufacturer, ODM）的機櫃產能爬坡[S-022]，再到800 VDC供電與液冷（liquid cooling）硬體[S-196; S-193]，最後到渦輪機與電網的未出貨訂單（backlog）[S-012; S-026]。每一步驟均在等待前一步驟傳來的確定需求，而最後各步驟依最長的時鐘運行（圖2）。因此，上游紓解的瓶頸會將需求與稀缺一併往序列的更下游推送。

A **各步驟之間的時間落差，正是延誤風險與機會所在。** 企業為自身的步驟載明日期，卻未提出共同的落差：SK hynix的印第安納無塵室預定於2028年10月前完成，約在2026年8月動土後兩年，量產則在2029年下半年[S-023]；台積電（TSMC）14倍光罩尺寸的CoWoS排定於2028年[S-024]；Siemens Energy致力於平均在6至12個月內將產能時段預訂轉為訂單[S-026]；GE Vernova的預訂量則已排至2031年[S-311]。因此，2026年規劃的一項部署，所面對的是2025年至2026年間決定的記憶體與封裝產能，以及數年前即已預訂的發電容量。愈早預訂者，即掌握各時鐘之間的落差。

IF **在買方預算與供應商訂單簿同步上升的情況下，此一投資循環即成立。** 此一研判（reading）取決於2027年的資本支出計畫維持較高水準（第2.1節），以及供應商持續收取保證金與預訂。若有大型買方因會計以外的原因下調其2027年計畫、記憶體或渦輪機的預訂量未獲轉換或遭取消，或ASML的30%、Advantest提前的測試機目標等2027年產能計畫遭到延後，則此一研判將趨於薄弱。應觀察ASML於2027年6月10日的資本市場日[S-025]、Infineon 2026年11月的財測指引[S-193]，以及GE Vernova年底的未出貨訂單（backlog）[S-012]。

讀者視角

本章對財務、產業與政府的意義

財務。 利潤率僅可在同一公司內部比較；緯穎（Wiwynn）於2026年4月變更營收基礎，使其自身的數列出現斷點。保證金與預訂顯示需求風險由誰承擔。

產業。 應將記憶體、封裝時段、已驗證的電力與冷卻零組件及渦輪機產能，納入同一份前瞻行程表下訂；組裝很少是阻斷交付的環節。

政府。 電網併接、保護裝置認證（certification）與出口許可，是任何標準都無法提供的韌性槓桿；附日期的本土化生產承諾將美國製HBM定在2029年。

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-002 NVIDIA (26 Aug 2026); S-003 SK hynix (29 Jul 2026); S-010 NVIDIA (26 Aug 2026); S-011 CoreWeave (11 Aug 2026); S-012 GE Vernova (22 Jul 2026); S-018 Siemens Energy (30 Mar 2026); S-021 Infineon Technologies (5 Aug 2026); S-022 Hon Hai (12 Aug 2026); S-023 SK hynix (28 Aug 2026); S-024 TSMC (23 Apr 2026); S-025 ASML (15 Jul 2026); S-026 Siemens Energy (29 Jun 2026); S-031 Advantest (29 Jul 2026); S-039 GE Vernova (9 Dec 2025); S-048 NVIDIA (21 Sep 2026); S-052 AUO (31 Aug 2026); S-055 SolarEdge and Infineon (9 Sep 2026); S-056 Samsung Electronics (12 Feb 2026); S-057 Besi (18 Jun 2026); S-063 Wiwynn (7 Aug 2026); S-064 Wiwynn (7 May 2026); S-065 CoreWeave (2 Mar 2026); S-066 Nebius (17 Jul 2026); S-075 Microsoft (29 Jul 2026); S-083 TSMC (16 Apr 2026); S-084 TSMC (16 Jul 2026); S-111 AMD Investor Relations (4 Aug 2026); S-115 Samsung and Broadcom (25 Jul 2026); S-116 Advantest (29 Jul 2026); S-127 Samsung Electronics (30 Jul 2026); S-129 Micron Technology (24 Jun 2026); S-130 KIOXIA (2 Jun 2026); S-169 AUO (31 Aug 2026); S-170 Besi (23 Jul 2026); S-172 Tokyo Electron (30 Jul 2026); S-191 ABB (21 Sep 2026); S-192 Infineon Technologies (28 May 2026); S-193 Infineon Technologies (5 Aug 2026); S-196 Vertiv (29 Jul 2026); S-209 LG Electronics (22 Sep 2026); S-211 Boyd (12 Mar 2026); S-232 Vertiv (2 Sep 2026); S-299 Eaton (31 Jul 2026); S-300 Siemens Energy (5 Aug 2026); S-301 Siemens Energy (5 Aug 2026); S-302 Wiwynn (8 May 2025); S-303 Wiwynn (8 Aug 2025); S-304 Wiwynn (7 Nov 2025); S-305 Wiwynn (8 Jan 2026); S-306 Wiwynn (10 Sep 2026); S-307 Vertiv (11 Feb 2026); S-308 Schneider Electric (30 Jul 2026); S-309 BGL (20 Jul 2026); S-310 Infineon Technologies (6 May 2026); S-312 Huawei (18 Sep 2025); S-317 Amkor Technology (8 Sep 2026); S-318 UMC (8 Dec 2025); S-325 Rapidus (5 Jun 2026); S-326 KIOXIA (2 Jun 2026); S-328 ASML (14 Nov 2024); S-483 Wiwynn (7 Aug 2026); S-485 Advanced Micro Devices (5 Aug 2025); S-486 Advanced Micro Devices (3 Feb 2026); S-487 KLA (29 Oct 2025); S-494 Calix (20 Jul 2026); S-495 KLA (29 Jan 2026); S-496 Super Micro Computer (11 Aug 2026); S-497 Super Micro Computer (31 Aug 2026); S-498 Celestica (28 Jan 2026); S-499 Celestica (27 Jul 2026); S-500 Generac (29 Jul 2026); S-501 Oracle (10 Jun 2026); S-502 Advanced Micro Devices (4 Nov 2025); S-503 NVIDIA (26 Aug 2026); S-586 KLA (29 Apr 2026); S-587 KLA (28 Jul 2026)

標準組織與聯盟. S-019 OCP (11 Aug 2026); S-208 OCP (未註明日期)

政府與監管機關. S-005 IEA (16 Apr 2026); S-032 ERCOT (18 Jun 2026); S-042 PJM Inside Lines (13 Aug 2026); S-061 PJM Inside Lines (10 Sep 2026); S-163 U.S. Department of Commerce (16 Jul 2026); S-313 Shanghai Stock Exchange (24 Jul 2026); S-314 U.S. Department of Commerce (19 Nov 2025); S-315 Federal Register (BIS) (14 Jul 2026); S-316 PIB, Government of India (17 Feb 2026); S-323 European Commission (9 Apr 2025); S-484 European Commission (9 Nov 2017)

研究機構. S-320 imec (8 Dec 2025)

市場研究. S-007 TrendForce (25 Aug 2026); S-008 TrendForce (20 Jan 2026); S-114 TrendForce (20 May 2026); S-132 TrendForce (9 Jul 2026); S-282 TrendForce (27 Jul 2026); S-319 TrendForce (24 Sep 2026); S-612 Synergy Research Group (7 Apr 2026); S-613 Synergy Research Group (19 Dec 2025)

媒體與法律評論. S-033 TrendForce (15 Jun 2026); S-037 TrendForce (6 Aug 2026); S-291 Tom's Hardware (31 Jul 2026); S-311 Utility Dive (23 Jul 2026); S-321 Tom's Hardware (27 Jul 2026); S-322 Tom's Hardware (12 Aug 2026); S-324 StorageReview (15 Aug 2026); S-327 The National (5 Dec 2025)

表9。 附日期的企業承諾將於2027年至2029年間增添美國的晶圓、封裝與記憶體產能

本土化生產於本十年後期增添美國廠房，而決定近期（near term）產出的封裝藍圖、記憶體產能爬坡、測試機與設備，則仍掌握在台灣、韓國、日本與歐洲企業手中。

區域	晶片與記憶體	封裝、測試與設備	機櫃與設施	電力設備與發電
美國	台積電（TSMC）亞利桑那第2座晶圓廠於2027年下半年（2H）量產（R）；追加1,000億美元用於晶圓廠與先進封裝（advanced packaging），2026年7月16日宣布（C）[S-083; S-084]	SK hynix印第安納州HBM基地：無塵室於2028年10月前完成（C）；2029年下半年（2H）量產（R）[S-023]；Amkor亞利桑那園區投資額提高至120億美元，第2期工程自2027年底起（C）[S-317]	緯穎（Wiwynn）德州Socorro：逾16億美元，至2027年底提供近1,000個職位（C）[S-306]；鴻海（Foxconn）於德州、威斯康辛州、俄亥俄州與加州擴建（C）[S-022]	GE Vernova燃氣渦輪機產出於2028年達24 GW、2030年達30 GW（C）[S-012]
台灣	台積電（TSMC）A14於2028年量產；A13與A12於2029年（R）[S-084; S-024]	台積電（TSMC）基板上COUPE自2026年起量產；14倍光罩尺寸CoWoS於2028年；SoW-X於2029年（R）[S-024]；聯華電子（UMC）光子積體電路，風險試產2026–27年（R）[S-318]	鴻海（Foxconn）Vera Rubin機櫃與CPO交換器，2026年第3季量產（R）[S-022]；Vertiv 800 VDC驗證場址，高雄（R）[S-196]	—
韓國	SK hynix M15X時程提前；龍仁第1期無塵室於2027年初（C）[S-003]；Samsung客製HBM樣品於2027年（R）[S-056]	SK hynix P&T7先進封裝（advanced packaging）廠，依需求分期推進（C）[S-003]	—	—
日本	Rapidus 2 nm製造於2027年前（T）[S-325]；Kioxia資本支出（capex）約每年4,700億日圓，為期三年（G）[S-326]	Advantest SoC測試機：2027年3月前達5,000台，2028年底或2029年初達10,000台，兩項目標均已提前（T）[S-116]；Tokyo Electron鍵合相關設備至2030年累計銷售額5,000億日圓，係作為商業機會提出（S）[S-172]	—	—
歐洲	—	ASML於2027年將低數值孔徑（low-NA）EUV與DUV浸潤式產能增加30%（C）；2028年再增30%仍在研議中[S-025]；Besi長期目標，未載明年度（T）[S-057]	—	Infineon第四座德勒斯登廠區模組（C）[S-021]；Siemens Energy於2027財年每年約50台大型燃氣渦輪機，自約35台提升（C）[S-026]
中國	Huawei Ascend 950DT，2026年第4季（R）[S-312]	—	—	—
波斯灣與印度	—	—	Stargate UAE首期200 MW於2026年第3季，據報導引述Mubadala執行長所述（R）[S-327]；印度政府：在38,000顆GPU之外再增加20,000顆，2026年2月宣布[S-316]	—

企業自行揭露的前瞻性承諾，依產能所在區域（未指名場址者，列於該公司母國所在區域）與層級列示，資料截至2026年9月27日；波斯灣與印度一列的出資方為一家主權投資機構與一國政府。類別：C為產能承諾（已規劃或在建）；R為藍圖；T為管理層目標；G為財測指引；S為情境（scenario）或機會。破折號表示本報告未引述任何由企業揭露的承諾。來源列於各欄位內。

表10。 每一受限層級均正在增添產能，其時鐘自一年延伸至五年以上

每一受限層級都有附日期的擴產計畫，但一年期的推進步驟多集中於晶片、測試與零組件，而發電與電網的增添則需三至五年以上才會到來。

層級	1年訊號	3年訊號	5年以上訊號	來源
HBM與DRAM	SK hynix M15X量產時程提前；龍仁第1期無塵室於2027年初（C）；Samsung客製HBM樣品於2027年（R）	SK hynix印第安納無塵室於2028年10月前完成（C）；次世代HBM於2029年下半年（2H）量產（R）	—	[S-003; S-056; S-023]
先進封裝	台積電（TSMC）基板上COUPE自2026年起量產（R）	14倍光罩尺寸CoWoS於2028年；14個光罩以上與40個光罩的SoW-X於2029年（R）	—	[S-024; S-084]
尖端邏輯	台積電（TSMC）2026年資本預算600億至640億美元（G）；亞利桑那第2座晶圓廠於2027年下半年（2H）量產（R）	A14於2028年量產；A13與A12於2029年（R）；三年期資本支出（capex）「even more significantly higher」（G）	—	[S-084; S-083; S-024]
測試	Advantest於2026財年結束前（2027年3月）達5,000台SoC測試系統，且正持續提前（T）	2028年底或2029年初達10,000台系統，且正持續提前（T）	下一步「could」達到15,000台或更多（S）	[S-116]
微影	ASML於2027年在約65台低數值孔徑（low-NA）EUV與約130台DUV浸潤式系統的基礎上增加30%（C）	2028年再增30%，仍在研議中（S）	2024年11月提出的2030年情境（scenario），將於2027年6月10日更新（S）	[S-025; S-328]
混合鍵合（Hybrid bonding）	Besi的混合鍵合（hybrid bonding）客戶數於2026年上半年自15家增至21家；AI約占上半年系統訂單的60%（actual）	—	長期目標為營收17億至22億歐元、營業利益率45%至55%，未載明年度（T）	[S-170; S-057]
功率半導體	Infineon於2026財年AI電力相關營收逾16億歐元（G）；附預付款的多年期產能預訂（C）	2027財年預估25億歐元以上，將於2026年11月上調（G）；第四座德勒斯登廠區模組（C）	—	[S-021; S-193]
電力與冷卻設備	Vertiv上調2026年財測指引（G）；高雄的800 VDC驗證場址（R）；Eaton Electrical部門未出貨訂單（backlog）+43%（actual）	—	—	[S-196; S-299]
燃氣渦輪機發電	GE Vernova自2026年第3季起每年20 GW（C），且至2026年底未出貨訂單（backlog）加預訂量至少達125 GW（G）；Siemens Energy於2027財年每年約50台大型機組（C）	GE Vernova於2028年達24 GW（C）；Siemens Energy預期2029財年大致售罄（C，已預訂）	GE Vernova於2030年達30 GW，相關行動正在進行（C）；Siemens Energy認為市場至少將維持高位至2035年（S）	[S-012; S-018; S-026]
電網設備	GE Vernova於2026年上半年的資料中心Electrification訂單逾50億美元；Siemens Energy於2026財年上半年約20億歐元的資料中心電網訂單（actual）	GE Vernova至2028年底Electrification未出貨訂單（backlog）達600億美元，總未出貨訂單約2,000億美元（G）	Siemens Energy：一輪「over the next fifteen years」的電網投資循環（S）	[S-012; S-026; S-039; S-018]

各企業所稱將增添的內容，依層級與自2026年9月起算的預測期限（horizon）列示：1年（至2027年底）、3年（2028年至2029年）、5年以上（2030年及以後）。類別：C為產能承諾（已規劃或在建）；R為藍圖；T為管理層目標；G為財測指引；S為情境（scenario）或期望；actual表示已公布數值。來源列於最末欄。

範圍 僅為企業自身的陳述，各依其自身的效力呈現；各步驟之間的時間落差並未推導，且各列在規模上並不可比。

第6章

標準、互通性與智慧財產權

開放介面在可藉多供應商驗證（qualification）降低轉換成本（switching cost）的交接點上最為重要；它們無法創造稀缺的產能，而一旦某個介面開放，差異化便移往實作、材料、控制、測試與服務。

適用對象： 智慧財產權與標準策略人員、研發（R&D）與創新團隊、採購（procurement）與政府部門；專利體制與技術空白區（white space）見第6.6至6.7節。

開放標準是人工智慧（AI）基礎設施中唯一不為任何單一買方或供應商所擁有的槓桿。它們無法增添高頻寬記憶體（high-bandwidth memory, HBM）產能、渦輪機產能時段或電網併接。然而，在機櫃、電力、冷卻、網路與管理之間的交接點上，它們能夠降低驗證（qualification）第二供應來源的成本。本章將Open Compute Project（OCP）視為常規的形成者，而非正式的標準組織，並梳理擁有其下層各介面的其他機構，以及規格（specification）發布時隨之成立的專利承諾（patent commitment）。本章提出一個問題：當一個介面開放時，差異化與可受保護的發明將移往何處？

6.1 OCP標準化了什麼，又有什麼是它無法做到的

在記憶體、封裝產能與電力均短缺的市場中，共通介面具有商業價值。當來自不同供應商的機櫃、電源架（power shelf）、冷卻液分配單元（coolant distribution unit, CDU）與管理控制器，能依文件化的條件彼此相容時，買方便可驗證（qualification）第二供應來源而無須重新設計整個系統，供應商也能將同一套設計賣進數個機隊。這正是OCP所提供的。這同時也是OCP所提供的極限：規格（specification）改變的是誰可以製造某一零組件，而非可以製造多少零組件，而且它對未選擇依其建置或採購者並不具拘束力。

F **OCP將其社群描述為常規的形成者，而非標準組織。** OCP於2011年由Facebook（今Meta）發起，並由一個依美國《國內稅收法典》第501(c)(6)條設立為德拉瓦州非營利無股份公司的基金會營運[S-329; S-330]。其九人董事會中有七席為企業席次，其中三席屬超大規模雲端業者（hyperscaler）（Google、擔任主席的Meta、Microsoft），四席屬晶片企業（AMD、Arm、Intel、NVIDIA），另有一席個人董事與執行長一席[S-331]。電力、冷卻與設施企業則透過另設的顧問委員會（Advisory Board）進入，該委員會於2026年1月15日新增六家企業，其中包括台達電子（Delta Electronics）、Schneider Electric與Vantage Data Centers[S-332]。

> “Though not a standards body, the OCP Community forges new technology norms” [S-329]

A **當撰寫規格（specification）的企業同時也依該規格採購時，常規便形成。** 電機電子工程師學會（Institute of Electrical and Electronics Engineers, IEEE）等正式機構係依書面共識規則進行投票，其標準並可為法規所引用；OCP的核准則透過其自身的指導委員會（Steering Committee）進行[S-047]。賦予OCP規格效力的，是其作者中各營運業者的採購分量（其中數家並擁有董事席次）[S-331]，以及發布時隨之成立的免權利金（royalty-free）專利承諾（patent commitment）（第6.6節）。由此衍生兩項限制。治理權掌握在買方與晶片供應商手中，而最直接受從晶片到電網（chip to grid）議程影響的電力、冷卻與設施供應商，則位於一個其權限未見於公開紀錄的顧問委員會之中[S-332]。此外，OCP並不撰寫安全法規（safety code）：800伏特直流（volts direct current, VDC）供電與液冷（liquid cooling）仍須在他處取得認證（certification）（第7.4節）。

F **規格（specification）係透過載明於紀錄的行為而定案，產品則正統一至單一標記。** 僅會員得提交規格，且每一份均受貢獻者授權協議（Contributor License Agreement, CLA）所涵蓋，而最終規格（Final Specification）則指經指導委員會（Steering Committee）核准者[S-047]。2026年5月11日，OCP宣布將設置30日回饋期間，於2026年7月31日後終止OCP Inspired標記，並以OCP Accepted作為其單一產品標示；該標示仍須經審查，並對照已核准的貢獻文件（contribution）進行驗證[S-109]。即使未提交任何設計檔案，產品只要符合一份已核准的規格，即可獲認可為OCP Accepted，且申請者無須揭露設計檔案[S-333; S-334]。

A **該標記所證明的是經審查的符合性，而非開放的設計，且其本身並非授權（licence）。** 買方可將OCP Accepted的列示，解讀為該產品曾對照一份具名規格（specification）接受審查的證據[S-333]。該列示並不等於取得設計的管道，既非專利授權，亦非數量的證據。OCP自身的簡介手冊列有逾500家會員、逾150家Marketplace解決方案供應商與逾800項產品列示[S-335]；這些數字衡量的是參與的廣度，而非部署或貢獻。對一份規格的影響力，應從每一份已發布文件上所印的貢獻者名單研讀而得[S-047]，而非從會員數研讀。

F **工作如今形成迅速，且往往在半私有群組中、以機櫃或設施規模進行。** 半私有工作流（semi-private workstream）係由已簽署貢獻者授權協議（Contributor License Agreement, CLA）並共同開發一份規格（specification）的貢獻組織所組成；其他會員則透過簽署回饋協議（Feedback Agreement）加入[S-336]。Ethernet for Scale-Up Networking（ESUN）工作流於2025年10月13日以12家企業起步，並於2026年3月10日發布ESUN 1.0，至當時參與企業已逾175家[S-248; S-020]。Open Data Center for AI倡議於2026年4月29日列出八份新核准的貢獻文件（contribution），其中包括Open Rack Wide（ORW）規格[S-062]。

A **開放性更多體現於產出，而非起草過程。** 在半私有群組之外，供應商與專利團隊可能要到規格（specification）張貼或發表時才初次得見，這縮短了他們的交期（lead time）[S-336]。協作加速基金（Collaboration Acceleration Fund）於2026年4月7日以「a non-equity, pre-competitive grant program」的形式啟動，按專案資源的25%至50%提供相對資金，條件是其成果須依OCP核准的授權（licence）條款回饋貢獻[S-337; S-338]，因此受資助的成果亦進入同一套專利體制。標準化單元（unit of standardisation）亦已自伺服器移至設施：針對次世代機器學習基礎設施的Open Data Center設計原則，於2026年8月27日達到0.7.0修訂版，並載有AMD、Google、Meta與NVIDIA對2027年至2028年的指引[S-001]（第4.4節）。

A **開放介面在可藉多供應商驗證（qualification）降低轉換成本（switching cost）的交接點上最具回報。** 此為本報告的第六項發現。共通的機櫃外廓、匯流排（busbar）、冷卻液接頭、冷卻液分配單元（coolant distribution unit, CDU）介面或管理協定，使買方得以在該邊界上以一家供應商替換另一家，而無須重新設計系統，且在投入要素稀缺時，此一選擇權更具價值。但在限制屬實體性質之處，其作用即告終止。SK hynix描述記憶體市場的情形為客戶需求超過供應能力[S-003]；GE Vernova的Gas Power設備未出貨訂單（backlog）與產能時段預訂量，於2026年6月底達116 GW[S-012]。沒有任何規格（specification）能增添記憶體晶圓或渦輪機產能時段；它只能使既有的產能更易於組合。

IF **常規形成者這一研判（reading），僅在制定規格（specification）的企業確實部署其所制定的內容時才成立。** 其成立的前提是：新規格須以載明版本的方式，出現在數家互不相關供應商的OCP Accepted產品之中，且各營運業者持續依所貢獻的文本採購。應觀察指名ESUN 1.0、ORW或冷卻液規格的產品列示，以及每一次發布所印的貢獻者名單[S-047]。若大型營運業者部署的介面與其自身的貢獻文件（contribution）出現分歧，或規格發布後並無產品隨之出現，則此一研判將趨於薄弱。標準發布並不等於採用。

重要性何在

交接點上的第二供應來源

當記憶體、光學與電力設備採配額分配時，買方的籌碼（leverage）便落在其所掌控邊界上的轉換能力。共通的機櫃、電力或冷卻液介面可使該項轉換更為便宜；但它並不會縮短稀缺投入要素本身的排隊佇列。

6.2 聖荷西2026能夠顯示什麼，又不能顯示什麼

全球峰會是一年之中OCP公告最為密集的一週，其中許多內容都會被報導為進展。對於要判斷究竟有何改變的讀者而言，應守的紀律是將關注度與狀態區分開來。主題演講、攤位、示範與夥伴數量，顯示的是企業希望被看見之處；真正改變一份規格（specification）內容的，只有少數幾項載明於紀錄的行為，而其中多數都發生在活動之後。

F **2026年峰會於10月12日至15日在聖荷西舉行，主題為「Scaling Innovation for the AI Era」。** 其主題頁呼籲以「from Chip to Grid」重新思考資料中心技術，所有場次均錄影並於會後分享[S-067; S-339]。10月13日於聖荷西會議中心，DMTF（前稱Distributed Management Task Force）舉辦一場關於Redfish、Security Protocol and Data Model（SPDM）與Platform Level Data Model（PLDM）的可管理性研討會；Optical Internetworking Forum（OIF）則舉辦一場關於224 Gbps與448 Gbps電氣通道以及共同封裝光學（co-packaged optics）與近封裝光學（near-packaged optics）的研討會[S-287]。OCP 2025年全球峰會的頁面在其同期場次中並未列有OIF條目，因此撰寫電氣通道與共同封裝協議的機構，如今在峰會上與OCP的網路架構工作相遇[S-340]。

F **控制、電力與電網議題與原型展示共同構成議程。** Lattice Semiconductor宣布將有一場主題演講，探討「the growing need for a common management and control model for AI and compute infrastructure」[S-341]。NVIDIA的活動頁面列出由Ian Buck主講的主題演講與約30場場次，其中包括800 VDC架構、大型負載（large load）的安全與電網併接，並表示已有逾200家生態系夥伴採用其MGX設計[S-226]。OCP將其創新村（Innovation Village）描述為原型示範[S-342]；展品包括一台標示為「sized to」Project Deschutes冷卻液分配單元（coolant distribution unit, CDU）架構的1 MW熱交換器、在一台運作中的Blackwell伺服器上運行的兩相冷卻（two-phase cooling），以及一個為符合電網法規而設的機櫃級層[S-220]。

A 關注度不等於採用，示範也不等於標準。 依OCP的政策，一份規格（specification）的狀態唯有透過載明於紀錄的行為才會改變：張貼附編號的草案、將定稿文件上傳至Contribution Hub（此舉即開啟30日的退出期間（opt-out period）），以及發布最終規格（Final Specification）並附上其授權（licence）通知、貢獻者名單與任何排除事項[S-047; S-343]。對產品而言，該行為則是經審查後獲得OCP Accepted認可[S-109]。主題演講、攤位、標示為「sized to」某一規格的展品或夥伴數量，均不改變上述任何一項。它們可以先於狀態改變而出現，但本身並非狀態改變，而展出者所主張的相容性亦非符合性認定。

F 2025年峰會顯示了產出如何成熟：在活動之後，見於附日期的文件之中。 AMD於2025年10月14日以靜態展示方式展出其依Meta的ORW規格（specification）打造的Helios機櫃[S-235]。2026年2月24日，AMD與Meta宣布達成協議，將部署最多6吉瓦（gigawatt, GW）的AMD Instinct圖形處理器（graphics processing unit, GPU），其中包括經OCP共同開發的Helios機櫃[S-112]。OCP於2026年4月29日將ORW規格列為新核准，即該次展出後六個半月[S-062]。商業上的步驟與標準上的步驟各自在稍後、以各自的文件到來。

IF 有三項研判（reading）可能在聖荷西出現變動，但每一項都須有附日期的文件為憑。 若固態變壓器（solid-state transformer, SST）規格（specification）v0.3之後出現附編號的修訂版，或張貼了原生800 V的Mt Diablo 2.0草案，則800 VDC這項研判即生變動[S-019]；若ESUN或Scale-Up Ethernet Transport（SUE-T）端點工作流（workstream）張貼新的附編號規格，則垂直擴展這項研判即生變動[S-249]；若設計原則推進至0.7.0修訂版之後，則開放資料中心這項研判即生變動[S-001]。若峰會在某一領域僅產出演講與展示，則該領域的研判即不成立。自2027年起，峰會移至舊金山的Moscone Center舉行，自2027年10月4日至7日開始[S-344]；場地改變，核准流程則未改變。

> **觀察重點**
> **聖荷西，2026年10月12日至15日**
> 關鍵在上傳與發布，而非台上時間：v0.3之後附編號的SST修訂版、Mt Diablo 2.0草案，或新的ESUN或SUE-T文件[S-019; S-249]。上傳即開啟30日的退出期間（opt-out period）[S-047]。區域性活動同樣會發布文件：800 VDC的更新即是在台北亞太峰會的首日張貼[S-019; S-345]。

6.3 OCP所倚賴的介面由其他機構擁有

OCP撰寫的是系統規格（specification）：機櫃、電源架（power shelf）、冷卻液介面、管理需求，以及日益涵蓋整座設施。其下層的通道、鏈路、記憶體堆疊、小晶片（chiplet）介面、管理協定與散熱指引，則屬於其他機構，各有其自身的會員規則、發布時鐘與專利政策。一份指名200 Gb/s通道或HBM4堆疊的機櫃設計，即承繼其下層文件的成熟度與授權（licence）條件。表11列出哪一介面由誰擁有，以及OCP在紀錄中的角色為何。

F 除OCP之外，如今另有兩個機構發布垂直擴展與水平擴展網路架構規格（specification）。 Ultra Accelerator Link（UALink）Consortium於2025年4月8日發布UALink 200G 1.0，適用於單一pod內最多1,024具加速器（accelerator），並於2026年4月7日再發布四份規格[S-245; S-346]。Ultra Ethernet Consortium（UEC）於2025年6月11日發布其1.0版規格；2026年7月16日的1.0.3版為現行版本[S-256; S-257]。OCP的ESUN 1.0採用UEC所定義的鏈路層重傳（Link Level Retry, LLR）與基於信用額度的流量控制（credit-based flow control, CBFC），而非自行定義[S-246]。

F 位於這些網路架構之下的電氣與光學通道，正處於週期中段。 IEEE 802.3工作小組的P802.3dj任務組涵蓋200 Gb/s至1.6 Tb/s的乙太網路，目前仍在運作[S-347]；而一個400 Gb/s/Lane訊號研究小組已於2026年3月13日獲准成立，處理最長500公尺的電氣與單模光纖鏈路[S-270]。OIF將其CEI-224G與CEI-448G專案列為現行工作[S-264]，並於2023年3月發布一份3.2 Tb/s共同封裝模組實作協議[S-284]。

F 記憶體、小晶片（chiplet）與主機輸入／輸出介面則由另外四個機構訂定。 JEDEC於2025年4月16日發布HBM4標準（JESD270-4），每一堆疊最高達2 TB/s，並於2025年12月發布其修訂版JESD270-4A[S-139; S-058]。Compute Express Link（CXL）Consortium於2025年11月18日發布128 GT/s的CXL 4.0；Universal Chiplet Interconnect Express（UCIe）Consortium則於2025年8月5日發布最高64 GT/s的UCIe 3.0[S-348; S-167]。PCI-SIG於2025年6月11日向會員發布128.0 GT/s的PCI Express（PCIe）7.0，並計畫於2028年前向會員發布256.0 GT/s的PCIe 8.0[S-349; S-350]。

A OCP的角色依層級而異：作者、設定檔制定者、管道或使用者。 在機櫃、電力、冷卻與管理各層級，OCP是作者。在網路架構層，它為其他機構的成果制定設定檔：ESUN建立於乙太網路之上並採用UEC所定義的功能[S-246]，而Foundation Chiplet System Architecture（FCSA）1.0.0則在既有標準存在之處，將小晶片（chiplet）傳輸對應至UCIe、CXL與PCIe[S-165]。在通道與記憶體堆疊層，它使用自己並未撰寫的文件。它所宣布的與UALink、Storage Networking Industry Association（SNIA）及ASHRAE的協作，所記錄的是管道，而非共同文件[S-351; S-352; S-182]。以兩個機構名義共同發布的附日期文件，才會是共同標準的證據。

A 紀錄顯示網路架構層呈現並存局面，而非單一贏家。 Credo於2026年1月推出一款224G重定時器（retimer），可在單一零組件中支援UALink、ESUN與乙太網路[S-253]；UALink並將其Chiplet 1.01規格（specification）描述為完全相容於UCIe 3.0[S-168]。零組件供應商正同時為多種協定進行開發，而單一產品也可以實作多個機構的文件。此一層疊關係具有雙重意義。就互通性而言，一套系統的穩定程度，僅等同於其所倚賴之最不成熟的文件。就授權（licence）而言，每一份文件各自帶有其專利體制，因此OCP設定檔的免權利金（royalty-free）條件並不延伸至其下層的標準（第6.6節）。

IF 開放網路架構唯有在多供應商產品與符合性測試依循這些文件時，才能降低轉換成本（switching cost）。 此一研判（reading）成立的前提是：出現來自互不相關供應商、載明符合UALink 2026年各項規格（specification）、ESUN 1.0或Ultra Ethernet 1.0.x的交換器與端點，並有測試計畫可供買方查核該項主張。應觀察首批指名上述版本的產品列示，以及IEEE 802.3的400 Gb/s專案在脫離研究小組階段後的進展[S-270]。若垂直擴展部署仍停留在單一供應商的網路架構上，而開放規格則仍在等待產品，則此一研判將趨於薄弱；第4.6節追蹤此一競逐。

6.4 信任與可管理性成為採購要求

一個AI機隊在每一具加速器（accelerator）、網路卡、硬碟與管理控制器上運行來自數十家供應商的韌體，而一份遭入侵的映像檔即可擴及數千台機器。因此，安全與可管理性正從供應商的功能，轉為書面的採購（procurement）條件：根植於晶片的身分、經簽章的韌體、共享的安全審查，以及買方可在驗收前測試的標準遙測（telemetry）。OCP撰寫裝置需求與一套開放的信任根（root of trust），DMTF撰寫線路協定，各國政府則訂定後量子的期限。拒絕一項裝置的權力屬於買方與監管機關，而非OCP。

F Caliptra為資料中心晶片提供一套開放的信任根（root of trust），其後量子支援則分兩步到來。 Caliptra為系統單晶片（system-on-chip, SoC）設計提供「IP and firmware for an integrated Root of Trust block」，其程式碼依Apache License 2.0在CHIPS Alliance中維護[S-362; S-363]。2025年7月31日由AMD、Google、Microsoft與NVIDIA提出的2.0版規格（specification），新增一種子系統模式，並在其憑證上採用雙簽章，其一為橢圓曲線（secp384r1），其一為後量子密碼學（post-quantum cryptography, PQC）簽章演算法ML-DSA-87[S-364]。2025年10月15日的2.1版暫存器傳輸層級（register-transfer-level，硬體設計）版本，新增ML-KEM金鑰封裝演算法，並支援OCP L.O.C.K.，後者將硬碟加密金鑰置於硬碟韌體之外[S-365; S-366]。

A 晶片所整合的Caliptra版本，決定了其在整個服役期間的後量子起點。 Caliptra韌體係透過一條起始於Caliptra唯讀記憶體的信任鏈進行驗證[S-364]，因此一項零組件在首次開機時所執行的檢查，在其製造之時即已固定。因此，採購（procurement）上真正有用的問題，是該零組件整合了哪一版Caliptra，而非它是否「supports Caliptra」。對於在2026年採購並運行至2030年代的加速器（accelerator）而言，此一答案決定了它能否滿足落在其使用年限之內的簽章期限。若能證明僅憑韌體更新即可為較早的晶片加入首次開機時的後量子驗證，則此一研判（reading）即會改變。

F 共享審查與可更換的管理模組，使安全得以查核。 OCP S.A.F.E.「standardizes firmware security reviews across the data center supply chain」，並列有11家已核准的安全審查服務供應商[S-367]。Datacenter-ready Secure Control Module（DC-SCM）將基板管理控制器（baseboard management controller, BMC）置於一個可更換的模組上；2025年10月16日提交的Rev 2.2 Ver 1.0為最新的公開修訂版，而3.0的工作流（workstream）正在進行[S-242; S-368]。Lattice Semiconductor於2026年7月27日完成對安全與可管理性韌體供應商AMI的收購[S-369]。

F DMTF的協定承載身分與控制，而混合式後量子支援仍處於回饋階段。 SPDM 1.4.0是2025年5月25日的一項DMTF標準，為首個納入美國國家標準與技術研究院（National Institute of Standards and Technology, NIST）後量子演算法的版本，並載明「No hybrid support in SPDM 1.4」[S-370]。1.4.1版於2026年7月2日隨之發布，DMTF並就SPDM 1.5的傳統與後量子混合式支援，徵求至2026年8月31日止的回饋[S-356; S-357]。OCP正推動一項經可信證明（attestation）的憑證請求功能納入SPDM，該功能在SPDM中屬「anticipated for addition in SPDM 1.5」[S-359]。

F 買方如今將可靠性與診斷寫入需求之中。 由Microsoft、Google、NVIDIA、Meta與AMD提出的圖形處理器（graphics processing unit, GPU）與加速器（accelerator）可靠性、可用性與可維護性（reliability, availability and serviceability, RAS）需求1.7版，要求具備錯誤注入、以頻內與頻外方式回報的標準錯誤紀錄，以及在致命錯誤時自動蒐集的當機傾印[S-371]。2025年12月14日的RAS應用程式介面（application programming interface, API）規格（specification）0.9版，是在方向變更之後的過渡版本[S-372]。Compliance Tool for Accelerator Management（CTAM）「provides acceptance testing for Accelerator Management in cloud data centers」[S-358]。

F 後量子期限為信任鏈定下了時程表。 NIST於2024年8月13日發布其前三項後量子標準FIPS 203、204與205[S-373]。就美國國家安全系統而言，CNSA 2.0演算法套組要求韌體簽章（firmware signing）「support and prefer CNSA 2.0 by 2025 where available, exclusively use CNSA 2.0 by 2030」[S-374]。2026年6月22日簽署的第14412號行政命令，就高價值聯邦系統訂定後量子金鑰建立應於2030年12月31日前完成、簽章應於2031年12月31日前完成，並指示為承包商訂定一項採購（procurement）規則

表11。 OCP撰寫系統；其下層的介面則屬於十餘個其他機構

依OCP規格（specification）打造的機櫃，承繼了其他機構文件的成熟度與授權（licence）條件，而這些文件多數是OCP所使用而非共同撰寫的。

介面	制定機構	現行文件與狀態	OCP的角色	來源
垂直擴展加速器（accelerator）鏈路	UALink Consortium	UALink 200G 1.0（2025年4月8日）；2026年4月7日的四份規格（specification），其中包括Chiplet 1.0（現列為1.01）	協作於2025年4月29日宣布	[S-245; S-346; S-168; S-351]
用於AI與HPC的乙太網路傳輸	UEC	Ultra Ethernet Specification 1.0（2025年6月11日）；1.0.3為現行版本（2026年7月16日）	自2023年10月17日起協作；ESUN 1.0採用UEC所定義的重傳與流量控制	[S-256; S-257; S-353; S-246]
電氣通道；共同封裝模組	OIF	CEI-224G與CEI-448G專案為現行工作；3.2T共同封裝模組協議（2023年3月）	同期研討會（co-located workshop），2026年10月13日	[S-264; S-284; S-287]
乙太網路實體層	IEEE 802.3	P802.3dj（200 Gb/s至1.6 Tb/s）任務組運作中；400 Gb/s/lane研究小組於2026年3月13日獲准成立	ESUN於啟動時即承諾與IEEE 802.3接觸	[S-347; S-270; S-248]
HBM與記憶體元件	JEDEC	HBM4 JESD270-4（2025年4月16日）；JESD270-4A（2025年12月）；SPHBM4（2026年7月13日）	使用者；JEDEC列為聯盟夥伴	[S-139; S-058; S-141; S-354]
主機輸入／輸出	PCI-SIG	128.0 GT/s的PCIe 7.0已向會員發布（2025年6月11日）；PCIe 8.0計畫於2028年前推出	使用者；FCSA將小晶片（chiplet）傳輸對應至PCIe	[S-349; S-350; S-165]
記憶體擴充與資源池化	CXL Consortium	128 GT/s的CXL 4.0（2025年11月18日）	使用者；FCSA對應至CXL	[S-348; S-165]
晶粒間小晶片（chiplet）鏈路	UCIe Consortium	48與64 GT/s的UCIe 3.0（2025年8月5日）	使用者；FCSA對應至UCIe	[S-167; S-165]
管理、可信證明（attestation）、遙測（telemetry）	DMTF	Redfish 2026.2（規格（specification）1.25.0）；SPDM 1.4.1（2026年7月2日）；SPDM 1.5的混合式後量子支援處於回饋階段	CTAM透過Redfish測試加速器（accelerator）管理；OCP推動憑證配發納入SPDM 1.5	[S-355; S-356; S-357; S-358; S-359]
用於AI的儲存與記憶體	SNIA	紀錄中無共同文件	協作於2025年10月13日宣布	[S-352]
設施散熱邊界	ASHRAE TC 9.9	液冷（liquid cooling）白皮書（white paper），界定W17至W+的設施供水等級	自2025年10月13日起結盟；無附日期的共同交付成果	[S-213; S-182]
資料中心網路訊號傳遞	IETF	快速網路通知（fast network notification, FANN）：問題陳述草案引述AI訓練；BoF申請已獲核准（頁面最後更新於2026年6月15日），FANN工作小組正在籌組成立	紀錄中無	[S-360; S-361]

介面、擁有該介面的機構、現行文件與狀態，以及OCP在紀錄中的角色，資料截至2026年9月27日；供梳理依賴關係的標準策略人員與產品團隊使用。來源列於最末欄。
協作、對齊聲明或同期研討會（co-located workshop）屬於管道，而非共同規格（specification）。GT/s為每秒十億次傳輸；HPC為高效能運算；CTAM為Compliance Tool for Accelerator Management；BoF為Birds of a Feather會議；IETF為Internet Engineering Task Force。

[S-375]。2026年6月24日的M-26-15號備忘錄表示，各機關對於其所使用、且屬網路安全暨基礎設施安全局（Cybersecurity and Infrastructure Security Agency）列為採用後量子標準類別的產品，應要求整合後量子密碼學（post-quantum cryptography, PQC）[S-376]。

A **有三類主體能夠拒絕一項裝置，而OCP並不在其中。** OCP規格（specification）對任何人均不生義務；信任鏈唯有在具備採購數量或法律權限的一方將其寫入合約或規則時，才會成為進入市場的條件。大型買方確實如此行事，而他們與作者群相互重疊：Caliptra 2.0的貢獻者與加速器（accelerator）RAS的作者，大致是同一批雲端與晶片企業[S-364; S-371]。各國政府則透過採購（procurement）規則行事[S-375; S-376]。韌體供應商占據第三個位置，因為可信證明（attestation）、更新與遙測（telemetry）邏輯都在管理韌體中運行，而其中一家供應商如今已隸屬於一家可程式邏輯供應商[S-369]，另有一條路徑則為開源[S-377]。

IF **若採購規則擴及韌體，可信證明（attestation）便會成為伺服器晶片的銷售條件。** 此一研判（reading）成立的前提是：第14412號行政命令所指示的承包商規則須涵蓋韌體與管理控制器，且SPDM 1.5須發布傳統與後量子的混合式支援[S-375; S-357]。應觀察該規則草案、SPDM 1.5的發布、自2026年6月3日起處於候選發布版RC4的OCP L.O.C.K. 1.1定稿版，以及0.9版之後的RAS介面修訂版[S-366; S-372]。若採購（procurement）接受供應商自我宣告而不要求裝置可信證明，則此一研判不成立。

> 未解問題
> **採用哪一種後量子簽章**
> CHIPS Alliance將Caliptra 2.0的ML-DSA-87簽章呈現為符合CNSA 2.0[S-378]。美國國家安全局（NSA）的CNSA 2.0常見問答建議韌體簽章（firmware signing）採用以雜湊為基礎的LMS或XMSS方案，但表示ML-DSA已核准用於所有簽章用途，且一旦完成標準化與驗證，對部分軟體與韌體簽章而言可能屬合理選擇[S-374]。仍未確定的是，某一特定零組件的ML-DSA-87實作是否已取得該項驗證；此一答案有助於判斷哪些世代的晶片符合美國國家安全系統的要求。

6.5 邊緣與電信邊上機櫃

隨著推論作業向使用者端靠近，加速伺服器正離開超大規模園區，進入區域代管機房、電信業者網路據點（telco point of presence），並隨著無線接取網路人工智慧（AI in the radio access network, AI-RAN）進入基地台台址。這正是OCP硬體規格（specification）可能與電信標準世界交會之處。該世界依3rd Generation Partnership Project（3GPP）的發布時鐘與須付權利金的專利承諾（patent commitment）運作，而非依OCP的貢獻規則，因此此一交會點須謹慎研讀。

F **OCP與IOWN Global Forum於2026年2月提出AI Computing Continuum的框架；其下層的硬體產品線則處於候選發布版階段。** 2026年2月10日，IOWN Global Forum與OCP Foundation宣布一項合作框架，其中由IOWN定義以光子學為基礎的通訊架構，OCP則定義開放硬體規格（specification）[S-379]。OCP的Edge-MHS規格頁面所記錄的候選發布版，最新為2025年12月8日的v0.9 RC1[S-380]。由早先Evenstar倡議重構而來的AI-Native Wireless Access子專案，其章程與工作流（workstream）仍標示為「TBA」；其頁面上重製的Evenstar Open RU章程，則列有「Collaborate and partner with Open Air Interface (OAI), 3GPP, TIP, and Open RAN (O-RAN)」的目標[S-381]。

F **無線接取網路人工智慧（AI in the radio access network, AI-RAN）正處於電信業者試驗階段，且在資料中心級伺服器上運行。** NVIDIA與Nokia於2025年10月28日宣布針對Nokia無線產品組合的AI-RAN產品，NVIDIA並將對Nokia投資10億美元[S-382]。2026年9月16日，Nokia表示有八家電信業者（其中包括中華電信（Chunghwa Telecom）與TPG Telecom）正在NVIDIA平台上執行AI-RAN概念驗證與實地試驗，且該平台已帶來逾20%的頻譜效率提升[S-383]。上述均為企業關於試驗的陳述，而非部署數量。

F **電信標準機構依延伸至2028年的發布時鐘運作。** 3GPP Release 20為最後一個「5G-Advanced only」版本，其Stage 3目標定於2027年3月，並將AI與機器學習列為重點領域之一[S-384]；Release 21為首個6G版本，其Stage 3凍結時點定於2028年12月[S-385]。O-RAN ALLIANCE議定其R006版本涵蓋與Release 20對齊的6G研究，R007則涵蓋與Release 21對齊的規範性工作[S-386]。在Internet Engineering Task Force（IETF）中，一份快速網路通知的問題陳述，將AI訓練列為需要此項功能的應用之一[S-360]。

A **交會點在伺服器，而彼此的聯繫是具名的夥伴，而非聯絡關係（liaison）。** 公開紀錄顯示，3GPP、O-RAN或IETF的交付文件中並未引用任何OCP規格（specification）；而OCP無線子專案的頁面，透過其所載的Evenstar章程，將3GPP與O-RAN列為預定夥伴，卻無任何載明於紀錄的協議[S-381]。由於無線接取網路人工智慧（AI in the radio access network, AI-RAN）將無線軟體置於加速伺服器之上[S-382]，OCP的機箱、機櫃、電力與管理規格可以降低加速器（accelerator）周邊的轉換成本（switching cost），而非加速器本身的轉換成本。專利條件亦有差異：在電信領域依公平、合理且無歧視（FRAND）條件授權（licence）標準必要專利（standard-essential patent）的無線設備供應商，若對某一OCP規格有所貢獻，所面對的則是免權利金（royalty-free）承諾（第6.6節）。

IF **OCP的邊緣規格（specification）唯有在三項落差彌合之後，才會成為電信的設計基礎。** 此一研判（reading）成立的前提是：AI Computing Continuum各子專案已公布章程、一份Edge-MHS規格獲接受並超越候選發布版階段，以及OCP與3GPP或O-RAN ALLIANCE之間出現聯絡關係（liaison）聲明或相互引用[S-381; S-380]。應觀察3GPP Release 20為5G-Advanced最後一批協定規格所定的2027年3月Stage 3目標，其重點領域包含AI與機器學習[S-384]。若各電信業者在這些落差仍然存在的情況下，即以專屬伺服器設計擴大AI-RAN與邊緣站點的規模，則此一研判不成立。

6.6 專利體制：開放介面，專屬實作

每一個標準機構同時也是一套授權（licence）安排。對專利權人而言，真正有用的問題不在於某機構是否自稱開放，而在於參與者承諾了什麼、該承諾於何時成立、涵蓋哪些請求項，以及如何退出。OCP的答案與記憶體及電信標準中常見的須付權利金體制不同，而此一差異決定了介面共享之後，價值仍可在何處獲取。本節依各治理文本所載內容說明相關規則；本節並非法律意見。

F OCP的政策（2026年8月14日版）使免權利金（royalty-free）承諾於發布時成立，而非於簽署時成立。 僅會員得提交貢獻，且每一份規格（specification）貢獻文件（contribution）均受Open Web Foundation（OWF）貢獻者授權協議（Contributor License Agreement, CLA）0.9所涵蓋，該協議授予著作權授權（licence），並載明「No express or implied patent, trademark, or other property rights are granted」[S-047; S-387]。規格授權（Specification License）則為Open Web Foundation Agreement 0.9（OWFa）。對一份提案規格的每一位貢獻者，在最終規格（Final Specification）發布時即視為已同意該授權，但有效排除通知（Exclusion Notice）中所列的內容除外[S-047]。

> “a 30-day period starting on the date the final document is uploaded to the OCP Contribution Hub to initiate the Steering Committee vote” [S-047]

F 退出窗口自上傳時開啟，且是唯一的專利例外安排。 上文所引的退出期間（opt-out period）可經全體貢獻者書面同意而提前結束，其後即不再接受任何排除通知（Exclusion Notice）[S-047]。排除通知須載明該規格（specification）、所排除的專利或申請案號，以及所涉及的條節；就尚未公開的申請案，則提供申請案文本或規格中受影響的部分[S-047]。未列入的專利仍受規格授權（Specification License）拘束，而最終規格（Final Specification）發布時會附上授權（licence）通知、貢獻者名單與任何有效的排除通知[S-047; S-343]。

F OWFa 0.9涵蓋實施規格（specification）必備部分所必要的請求項，包括其後取得的請求項。 簽署人給予不可撤回的承諾，不就該等請求項提出主張；或者作為替代方案，給予「a no charge, royalty free license to my Necessary Claims on reasonable and non-discriminatory terms solely for your Implementation」[S-388]。必要請求項（Necessary Claims）包括簽署後取得、且為實施必備部分所必要的請求項，其中亦含選擇性部分中的必備要素[S-388]。若有一方就該等請求項對簽署人提起訴訟，上述授予即對該方終止[S-388]。

A 對貢獻者而言，對OCP規格（specification）的必要性是一項承諾，而非一個營收地位。 該承諾僅及於一份已發布規格所要求的內容。文本所未涵蓋的部分，例如必備部分以外的實作、材料、製造流程、控制軟體與韌體、測試方法與服務，均落在其範圍之外[S-388]。從未對某一規格有所貢獻的權利人，在該規格之下不負任何承諾[S-047]。對貢獻者而言，真正關鍵的日期是上傳日，而非發布日或某場主題演講；窗口期間內保持沉默即視為未提出排除；而依文本，窗口結束時尚無申請案在案的發明並無可列之物，其後取得的必要請求項（Necessary Claims）則仍受涵蓋[S-047; S-388]。

F 軟體、硬體設計與周邊程式碼適用各自不同的條款。 軟體貢獻文件（contribution）須與一份提案規格（specification）相連結，並依經Open Source Initiative核准的授權（licence）條款發布；其他貢獻則歸入公共領域，或依Creative Commons或著作權授權條款發布[S-047]。OCP較早的硬體授權條款OCPHL-R與OCPHL-P已列為封存，「for use only with the OCP CLA」[S-389]。數個周邊開源專案適用Apache License 2.0，該條款帶有明示的貢獻者專利授權：Caliptra、SONiC網路作業系統的建置映像檔，以及OpenBMC的Redfish伺服器均為其例[S-363; S-390; S-377]。

F JEDEC、IEEE與UEC以容許收取權利金的授權（licence）條件約束參與者。 JEDEC參與者同意，以參與為條件，依合理且無歧視（RAND）條件就必要專利請求項提供授權，並揭露已知的可能必要專利[S-391]。IEEE標準協會的保證函（Letter of Assurance）承諾不收取報酬或依合理費率提供授權，該費率並排除納入標準所生的價值，同時禁止對願意本於善意協商的實施者尋求禁止令[S-392]。UEC的章程要求依FRAND條件就必要請求項（Necessary Claims）提供授權，而Ultra Ethernet Specification本身則不授予任何專利權[S-393; S-267]。DMTF與SNIA的政策同樣以RAND為基礎[S-394; S-395]；表12將各體制並列對照。

? 未解問題——UALink、CXL與UCIe三個聯盟並未在其公開頁面載明其專利條件。 其會員頁面僅提及會員協議中的智慧財產權政策，而CXL的評估版並未授予任何實施授權（licence）[S-396; S-397; S-398; S-399]。此點之所以重要，是因為垂直擴展與小晶片（chiplet）產品會同時實作其中數份文件，因此自由實施（freedom to operate）審查不能假定OCP設定檔的免權利金（royalty-free）條件可延伸至其下層的介面。

A 同一項功能可能同時落在兩套體制之下。 ESUN 1.0採用UEC所定義的鏈路層重傳（Link Level Retry, LLR）與基於信用額度的流量控制（credit-based flow control, CBFC）[S-246]；UEC參與者依RAND條件承諾[S-393]，OCP貢獻者則依免權利金（royalty-free）條件承諾[S-047]。專利池管理機構Sisvel已就一項擬設立的專利池，向專利權人徵求對應Ultra Ethernet Specification v1.0.2的請求項對照表[S-268]。同理，OCP正推動裝置身分配發納入DMTF的SPDM[S-359]；DMTF 2025年的智慧財產權政策聲明，旨在確保其標準可依RAND條件實施[S-394; S-400]。紀錄顯示的是各體制觸及同一項功能；它並未顯示任何請求項如何涵蓋任何產品，本處亦未指明任何一方受哪一套體制規範。

F 紀錄中的爭訟涉及AI系統與記憶體模組，而非OCP規格（specification）。 2026年3月11日，統一專利法院慕尼黑地方分庭就ParTec針對NVIDIA、以歐洲專利3 743 812為基礎提起的侵權訴訟，於一審予以駁回，ParTec並宣布上訴[S-401]。2026年9月24日，美國國際貿易委員會就Netlist關於DDR5記憶體模組的訴狀，表決對Micron等公司展開調查[S-402]；另於2026年8月5日，Netlist與Samsung宣布一項交互授權（licence），依該協議雙方「settle and mutually release all pending legal actions」[S-403]。上述各案依其公開內容，均未指名任何OCP規格。

IF 免權利金（royalty-free）介面唯有在貢獻者不提出排除、監管機關不提出異議的情況下，才會維持免權利金。 此一研判（reading）成立的前提是：各次發布持續僅列少數或完全未列排除通知（Exclusion Notice），且近期規格（specification）（例如ESUN 1.0與ORW基礎規格）上的「Modified OWFa 0.9」通知並未縮小該項承諾[S-246; S-228]。應觀察每一次發布所印的排除通知，以及美國的反壟斷表態：2025年9月，一名美國司法部官員批評「private consortia that impose mandatory, royalty-free cross-licensing obligations on their members」，但未指名任何一方[S-404]。若有主管機關就某一免權利金規格體制啟動案件，或主要貢獻者開始就其自行撰寫的介面排除請求項，則此一研判不成立。

6.7 介面開放時差異化移往何處

介面開放並不會終結競爭，而是使競爭移位。最終規格（Final Specification）一旦固定了接頭、協定或效能範圍，供應商便轉而在介面背後的產品如何製造、控制、散熱、測試與維護上競爭，專利策略亦隨之移往該處。此為本報告的第十項發現：下一波研究與智慧財產權機會，落在尚未定案的實作、材料、控制、測試與維護邊界周邊。開放並未消除智慧財產權機會，而是將其移轉至共用文本所留待開放的各層級。

A OCP規格（specification）固定的是要求與介面，而非實作。 2026年6月22日生效、由Google、Microsoft與NVIDIA撰寫的SST Specification 0.3.0，描述一款自中壓交流（medium-voltage alternating current）轉換至800 VDC的百萬瓦級轉換器，並訂出諸如在50%至100%使用率之間、僅計入功率鏈路損耗時效率至少達98%之類的要求，而非轉換器拓撲[S-054]。FCSA 1.0.0僅在已有產業標準之處載明傳輸方式，其餘則將介面標示為「TO BE DEFINED」[S-165]。ORW基礎規格未規定任何外部框架尺寸[S-228]。由於專利承諾（patent commitment）僅涵蓋必備部分[S-388]，拓撲、磁性元件、控制韌體與材料仍屬各供應商自有。

A 在OCP尚無草案之處，介面文本根本還未定案。 OCP的光路交換（optical circuit switching, OCS）子專案仍在研擬「guiding principles, open standards, and a fully open network technology stack」[S-406]，雖然Google自2020年起便已在生產環境中運行自有的光路交換器[S-293]。2026年9月2日的低壓直流（low-voltage direct current, LVDC）文件是一份白皮書（white paper），其列出保護裝置，卻未訂定保護動作時序[S-059]；兩相冷卻（two-phase cooling）則以流體指引為依據，流體規格與測試仍列為工作項目[S-060]。在這些領域中，研究成果仍可能形塑未來最終規格（Final Specification）的要求內容，以及其貢獻者的承諾何時生效。

A 六項技術邊界顯示介面開放之後，實作在何處仍保有差異化。 在每一項邊界上，共用文件固定的是必須連接或必須達成的要求，而困難的工程則落在其背後：在電路、材料、控制邏輯、測試方法與現場維護之中。以下清單列出每一項邊界，以及該邊界仍屬開放的證據。它標示的是可受保護之發明可能存在之處，而非確實存在之處；且此為本報告對所引文件的研判（reading），而非對任何企業產品的技術評估。

- **800 VDC供電。** SST草案固定了電壓、效率，以及轉換器為協調配合所設的故障電流行為（150%持續150 ms，其後得以跳脫）；系統層級的保護協調配合、斷路器控制與故障偵測仍屬開放，SolarEdge與Infineon亦仍在開發固態斷路器（solid-state circuit breaker）[S-054; S-055]。
- **液冷（liquid cooling）。** OCP的冷卻液與CDU規格（specification）訂出流體化學組成與機組要求[S-217; S-030]；冷板（cold plate）結構、兩相流體、洩漏處理與試運轉驗收（commissioning）則落在其背後[S-060; S-220]。
- **共同封裝光學（co-packaged optics）。** 2026年一篇回顧文章將熱管理、可製造性與標準化列為尚待解決的障礙[S-053]；雷射模組、光纖接合與光引擎測試則是實作區域。
- **記憶體與封裝。** JEDEC固定了HBM4的通道介面[S-058]；基礎晶粒已是一項設計變數，因為SPHBM4沿用HBM4的記憶體晶粒，並搭配適用於有機基板的新基礎晶粒[S-141]。
- **垂直擴展網路架構。** ESUN 1.0固定了交換器側的框架格式與鏈路行為[S-246]；端點傳輸則屬另一個工作流（workstream）SUE-T，至今仍未有已發布的規格（specification）[S-249]。
- **管理與安全。** 各項協定固定了訊息與演算法[S-370]；當機傾印處理、機隊遙測（telemetry）分析，以及回應可信證明（attestation）的韌體，則仍留在供應商手中[S-371]。

A 置於兩條軸線上，本報告所涵蓋的技術分為四組。 圖26依註明日期的紀錄所示，將標準成熟度（standards maturity）對照商業部署。部分介面已發布且已量產：JEDEC已將HBM4標準化，SK hynix、Samsung與Micron亦已出貨[S-058; S-003; S-056; S-128]。另有一些已發布但尚未達到多供應商量產，例如UALink 2026年各項規格（specification）、ESUN 1.0與CXL 4.0[S-346; S-020; S-348]。共同封裝光學（co-packaged optics）已在特定供應商的交換器中出貨：NVIDIA表示其Spectrum-6交換器系統支援共同封裝光學，鴻海（Foxconn）則表示其CPO全光

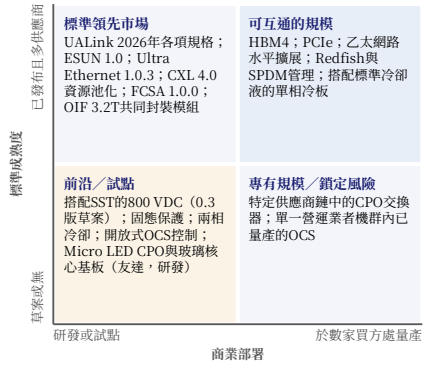
表12。 僅OCP與（通常而言）CHIPS Alliance以免權利金約束貢獻者；JEDEC、IEEE、UEC、DMTF與SNIA則容許收取權利金
單一產品可以同時實作數個機構的文件，因此其專利地位結合了自免權利金（royalty-free）不主張承諾至須付權利金的RAND授權（licence）等各種承諾。

制定機構	專利體制	貢獻者承諾的內容	來源
OCP	免權利金（royalty-free）：最終規格（Final Specification）發布時即視為已接受OWFa 0.9；排除通知（Exclusion Notice）須於上傳後30日內提出	就必備部分所必要的請求項（含其後取得的請求項）承諾不主張，或提供免費授權（licence）；貢獻者授權協議（Contributor License Agreement, CLA）僅授予著作權	[S-047; S-388; S-387]
CHIPS Alliance（Caliptra）	規格（specification）採免權利金（royalty-free），依OWFa 0.9；CHIPS Alliance常見問答稱此為其各專案的通例；Caliptra程式碼依Apache 2.0授權	規格（specification）承諾與OCP之下相同；所貢獻的程式碼依Apache 2.0提供明示的專利授權（licence）	[S-405; S-363]
JEDEC	以參與為條件採RAND，不必然為免權利金（royalty-free）	依RAND條件就必要專利請求項提供授權（licence）；揭露已知的可能必要專利，但無搜尋義務	[S-391]
IEEE SA	保證函（Letter of Assurance）：不行使權利的聲明，或不收取報酬或依合理費率提供授權（licence）	費率排除納入標準所生的價值；不得對願意本於善意協商的實施者尋求禁止令	[S-392]
UEC	依其Joint Development Foundation章程採RAND	依FRAND條件就必要請求項（Necessary Claims）提供授權（licence）；規格（specification）文本依CC BY-ND 4.0授權，且不授予專利	[S-393; S-267]
DMTF	RAND（2025年聲明）；2001年政策：免權利金（royalty-free）或公平合理的費率	依RAND條件提供授權（licence）；預期揭露已知專利	[S-394; S-400]
SNIA	RAND，收取或不收取報酬均可	依RAND條件就必要請求項提供非專屬、全球性授權（licence）	[S-395]
UALink、CXL、UCIe	由公開頁面未載明的會員智慧財產權政策訂定	UALink與CXL會員依其協議中所述的智慧財產權保護進行實作；CXL的評估版不授予任何實施授權（licence）	[S-396; S-397; S-399]

各機構的專利體制與貢獻者承諾，依其治理文本於2026年9月27日前所載內容呈現；供須將單一產品對應至多份規格（specification）的法務人員與標準策略人員使用。各項條件均依所載呈現，未經評估。
RAND為合理且無歧視；FRAND為公平、合理且無歧視；IEEE SA為IEEE標準協會；CC BY-ND為Creative Commons姓名標示—禁止改作。此處未載明PCI-SIG與OIF的專利條件。本表所載內容均非法律意見。

交換器有望於2026年第3季量產[S-010; S-022]。800 VDC供電仍停留在草案與參考設計（reference design）階段[S-054; S-192]。

圖26。 已發布的標準仍在等待產品，而已出貨的光學則運行於封閉的供應商鏈上



依本報告對標準成熟度（standards maturity）（無文件、草案、已發布標準）與商業部署（研發、試點、於單一營運業者或單一供應商鏈中的產品、於數家買方處量產）的研判（reading），定位各項技術，並依2026年9月27日註明日期的公開紀錄呈現。來源：HBM4 [S-058; S-003]；PCIe [S-349]；乙太網路與管理[S-347; S-355]；冷板（cold plate）[S-006; S-217]；網路架構與小晶片（chiplet）[S-346; S-020; S-257; S-348; S-165; S-284]；800 VDC [S-019; S-054; S-055]；兩相冷卻（two-phase cooling）[S-060]；OCS [S-293; S-406]；CPO交換器[S-010; S-022]；友達（AUO）[S-052]。
此為定性定位，而非評分。已發布的標準不等於採用，示範也不等於量產。

A 介面工作在對角線之外的兩個角落回報最大，原因卻恰好相反。 在左上角，已發布的文本仍在等待多供應商產品；符合性測試與首個完成驗證（qualification）的第二供應來源（second source）可使其轉化，而這正是新進者得以競爭之處。在右下角，商業成功則透過封閉的供應鏈達成：依Tech Times報導，日月光（ASE）與台積電（TSMC）的高階主管於2026年亞太峰會上表示，CPO交換器在共通測試標準出現之前即已出貨，且晶圓級測試、光纖陣列整合與光學封裝組裝仍未解決[S-286]。在該處，開放介面將降低買方的轉換成本（switching cost）。左下角擁有最多的研究與專利空間，右上角則在成本、良率（yield）、服務與供應上競爭。

IF CPO唯有在共通的測試、介面與可維護性文件隨之到位時，才會邁向可互通的規模。 此一研判（reading）成立的前提是：出現一份涵蓋光引擎機械、熱、雷射與管理介面的開放文件，且光引擎與雷射均有一家以上完成驗證（qualification）的來源。依Tech Times報導，應觀察由20家企業組成的OCP聯盟預期於2026年第4季提交的首批CPO規格（specification）[S-285]；台積電COUPE的量產爬升，該公司自稱已自2026年中起進入量產[S-084]；以及鴻海第3季的CPO交換器出貨[S-022]。若CPO交換器在沒有介面草案的情況下，透過單一供應商驗證即達到量產，則此一研判不成立，而OCP將僅餘機櫃、佈線與管理層級。

A 產能訊號註明日期且已量化；標準訊號所記錄的是文件，而非已完成驗證的第二供應來源。 表13就六項技術，並列其最新技術里程碑、企業產能或資本支出（capital expenditure）訊號、營收或利潤率訊號，以及標準或驗證（qualification）訊號。每一列的前三項均具體：HBM4已進入量產出貨、14倍光罩尺寸的CoWoS封裝排定於2028年量產、燃氣渦輪機產出規劃於2028年達24 GW[S-003; S-024; S-012]。標準欄所記錄的則是文件：在介面已成熟之處為已發布的規格（specification），例如JEDEC的HBM4與OCP的冷卻液與CDU規格；800 VDC仍為草案；共同封裝光學與電網則落在OCP之外。其中任何一項均不足以單獨顯示已有第二家供應商完成驗證。產能預留不等於標準，標準也不等於客戶。

A 技術落差尚不等於專利地位。 本節所列邊界標示的是實作仍保有差異化、並可能含有可受保護之發明的區域。它們並未確立任何發明具有新穎性：相關技術有相當部分早已公開，自2022年發表的記憶體資源池化與分層研究[S-159; S-407]，至OIF 2023年的共同封裝協議[S-284]皆然。在作出任何專利結論之前，均須經請求項層級的新穎性與自由實施（freedom to operate）審查，且並非每一項未解的工程問題都是專利機會。本處並未陳述任何一方擁有、侵害或必須就任何專利取得授權（licence）。

IF 隨著各項最終規格（Final Specification）陸續發布，技術空白區（white space）正逐一領域收攏。 貢獻者的必要請求項（Necessary Claims）於最終規格發布時即受承諾拘束[S-047]，因此今日尚未定案的草案，正是研究成果與請求項撰寫與該項承諾交會之處：0.3修訂版之後的SST、支援原生800 V的Mt Diablo 2.0、SUE-T端點傳輸、Universal Quick Disconnect v2修訂版，以及下一版FCSA[S-019; S-249; S-060; S-408]。應觀察上傳至Contribution Hub的文件，每一次上傳均開啟30日期間。若前沿規格（specification）日後開始規定控制演算法、流體配方或其他實作細節，則此一研判（reading）不成立，因為那會將今日的差異化區域拉入該項承諾之內。

讀者視角

本章對智慧財產權、研發、創新、政府與產業團隊的意義

智慧財產權。 應追蹤上傳文件，而非主題演講；30日的排除期間自上傳時開啟，且唯有在該期間屆滿前已申請在案的請求項才得列入。

研發。 在OCP僅有草案或尚無文件之處，投入仍能形塑文本內容：800 VDC保護、兩相流體、CPO測試、垂直擴展端點。

創新。 UALink、ESUN與CXL資源池化等已發布但尚未被採用的介面，會回報於符合性測試工具，以及首個完成驗證（qualification）的第二供應來源（second source）。

政府。 決定可信證明（attestation）是否成為強制要求的，是採購（procurement）規則與後量子期限，而非OCP規格（specification）；800 VDC與冷卻液安全則取決於認證（certification）機構。

產業。 OCP Accepted的列名顯示的是經審查的符合性，而非開放設計、授權（licence）或量產規模；應對照指名的規格（specification）版本驗證供應商。

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-003 SK hynix (29 Jul 2026); S-010 NVIDIA (26 Aug 2026); S-012 GE Vernova (22 Jul 2026); S-022 Hon Hai (12 Aug 2026); S-023 SK hynix (28 Aug 2026); S-024 TSMC (23 Apr 2026); S-052 AUO (31 Aug 2026); S-055 SolarEdge and Infineon (9 Sep 2026); S-056 Samsung Electronics (12 Feb 2026); S-084 TSMC (16 Jul 2026); S-112 AMD (24 Feb 2026); S-128 Micron Technology (24 Jun 2026); S-192 Infineon Technologies (28 May 2026); S-193 Infineon Technologies (5 Aug 2026); S-196 Vertiv (29 Jul 2026); S-226 NVIDIA (未註明日期); S-235 AMD Newsroom (14 Oct 2025); S-253 Credo Technology Group (29 Jan 2026); S-268 Sisvel (未註明日期); S-301 Siemens Energy (5 Aug 2026); S-307 Vertiv (11 Feb 2026); S-341 Lattice Semiconductor (12 Aug 2026); S-369 Lattice Semiconductor (27 Jul 2026); S-382 NVIDIA (28 Oct 2025); S-383 Nokia (16 Sep 2026); S-401 ParTec AG via EQS News (11 Mar 2026); S-402 Netlist (24 Sep 2026); S-403 Netlist (5 Aug 2026)

標準組織與聯盟. S-001 OCP (27 Aug 2026); S-019 OCP (11 Aug 2026); S-020 OCP (10 Mar 2026); S-030 OCP (26 Feb 2026); S-047 OCP (14 Aug 2026); S-054 OCP (22 Jun 2026); S-058 JEDEC Solid State Technology Association (Dec 2025); S-059 OCP (2 Sep 2026); S-060 OCP (6 Aug 2026); S-062 OCP (29 Apr 2026); S-067 OCP (未註明日期); S-109 OCP (11 May 2026); S-139 JEDEC (16 Apr 2025); S-141 JEDEC (13 Jul 2026); S-165 OCP (12 Feb 2026); S-167 UCIe Consortium (5 Aug 2025); S-168 Ultra Accelerator Link Consortium (7 Apr 2026); S-182 OCP / ASHRAE (13 Oct 2025); S-187 ANSI / UL (22 Oct 2025); S-213 ASHRAE (未註明日期); S-216 OCP (未註明日期); S-217 OCP (6 Aug 2026); S-220 OCP (22 Sep 2026); S-228 OCP (28 Apr 2026); S-242 OCP (27 Jul 2022); S-245 UALink Consortium (8 Apr 2025); S-246 OCP (9 Feb 2026); S-248 OCP (13 Oct 2025); S-249 OCP (13 Jun 2026); S-256 Ultra Ethernet Consortium (11 Jun 2025); S-257 Ultra Ethernet Consortium (16 Jul 2026); S-264 OIF (未註明日期); S-267

表13。 每一列都有已量化的產能與營收訊號；標準欄所記錄的是文件，而非已完成驗證的第二供應來源

六項技術的產能與營收里程碑，企業均揭露了註明日期且已量化的數字，而標準欄所收錄的已發布、草案或外部文件，本身並不足以顯示多供應商驗證（qualification），因此產能承諾絕不應被解讀為可互通性。

技術	技術里程碑	產能或資本支出訊號	營收或利潤率訊號	標準或驗證訊號
HBM4／HBM4E	SK hynix自2026年第2季起量產出貨HBM4，並於2026年上半年（1H）出貨HBM4E樣品[S-003]；Micron預期HBM4E於2027曆年進入量產[S-128]	SK hynix龍仁第1期無塵室於2027年初啟用；印第安納州HBM量產排定於2029年下半年（2H）[S-003; S-023]	SK hynix 2026年第2季營業利益率76%[S-003]；Samsung於2026年2月預期2026年HBM銷售額將增至三倍以上[S-056]	JEDEC JESD270-4A，2025年12月[S-058]；SK hynix與約10家客戶簽有長期協議（long-term agreement, LTA），係就記憶體整體而言[S-003]
CoWoS／先進封裝（advanced packaging）	14倍光罩尺寸的CoWoS（約10顆運算晶粒、20組HBM堆疊）排定於2028年量產；超過14倍光罩尺寸者與SoW-X則於2029年[S-024]	台積電2026年資本預算上調至600億至640億美元[S-084]	台積電2026年第3季毛利率財測指引65%至67%（全公司）[S-084]	UCIe 3.0，2025年8月[S-167]；OCP FCSA 1.0.0，2026年2月[S-165]；JEDEC適用於有機基板的SPHBM4，2026年7月[S-141]
CPO	台積電基板上的COUPE已於2026年進入量產[S-024; S-084]；NVIDIA Spectrum-6交換器系統支援可插拔光學與共同封裝光學（co-packaged optics）[S-010]	鴻海（Foxconn）的CPO全光交換器有望於2026年第3季量產並出貨[S-022]	無CPO專屬數字；鴻海預期800G以上交換器的2026年營收將成長一倍[S-022]	OIF 3.2T共同封裝模組協議，2023年3月[S-284]；2026年一篇回顧文章將標準化列為尚待解決的障礙之一[S-053]；友達（AUO）的Micro LED CPO模組處於研發階段[S-052]
800 VDC	Mt Diablo ±400 V側置電源機櫃（sidecar）規格（specification）已發布，支援原生800 V的Mt Diablo 2.0將隨後推出；SST Specification v0.3[S-019; S-054]	Infineon已簽署或正在協商的多年期產能預留協議，累計達數十億歐元的高個位數水準[S-193]	Infineon於2026財年的AI專用電源營收逾16億歐元；2027財年為「2.5 billion Euros plus」，並將向上修正[S-193]	與UL Solutions、NFPA、IEEE及IEC進行的安全認證（certification）工作[S-019]；適用於固態斷路器（solid-state circuit breaker）的UL 489I，2025年10月[S-187]
液冷	Project Deschutes 2 MW CDU規格（specification）1.0.0，2026年2月[S-030]；TrendForce估計2025年AI晶片中的滲透率約為33%，並預測2026年為53%[S-006]	Vertiv於2025年底未出貨訂單（backlog）150億美元，訂單出貨比約2.9（集團）[S-307]	Vertiv 2026財年調整後營業利益率財測指引23.3%至24.3%（集團）[S-196]	OCP水基冷卻液規格（specification）1.3版，2026年8月，以及丙二醇指引[S-217; S-216]；與ASHRAE的聯盟，無註明日期的交付項目[S-182]
電網／發電	機櫃層級的電網法規協調配合已列為2026年10月峰會的示範項目[S-220]；中壓至800 VDC的轉換仍為草案[S-054]	GE Vernova的燃氣渦輪機產出自2026年第3季起為每年20 GW，2028年為24 GW[S-012]；Siemens Energy於2026年6月30日的燃氣渦輪機確定未出貨訂單（backlog）為69 GW，另有26 GW的產能時段預訂[S-301]	GE Vernova未出貨訂單（backlog）1,760億美元（集團）；電氣化事業的資料中心訂單年初至今逾50億美元[S-012]	OCP與EPRI就資料中心作為彈性電網資源所締結的聯盟，2026年4月[S-062]；併接條件由電網業者與監管機關訂定（第7.2節）

六項技術截至2026年9月27日所揭露的最新註明日期技術里程碑、企業產能或資本支出（capital expenditure）訊號、營收或利潤率訊號，以及標準或驗證（qualification）訊號。企業的陳述均保留其自身效力（計畫、預期、財測指引、目標）。來源見各欄位。

範圍 企業數字涵蓋所指名的整家公司或整個部門，而非僅限該列所述技術，並依各公司自身的用語與期間呈現；各列之間不可比較，亦不相加或排名。

1H／2H為上半年／下半年；CDU為冷卻液分配單元；SST為固態變壓器；COUPE為台積電的共同封裝光引擎技術；EPRI為電力研究院；NFPA為美國國家防火協會；IEC為國際電工委員會。

本章來源 （續）

Ultra Ethernet Consortium (28 Jan 2026); S-270 IEEE 802.3 Ethernet Working Group (3 Aug 2026); S-284 OIF (29 Mar 2023); S-287 OCP (未註明日期); S-329 OCP (未註明日期); S-330 OCP (14 Jun 2024); S-331 OCP (未註明日期); S-332 OCP (15 Jan 2026); S-333 OCP (未註明日期); S-334 OCP (未註明日期); S-335 OCP / IDC (未註明日期); S-336 OCP (未註明日期); S-337 OCP (7 Apr 2026); S-338 OCP (未註明日期); S-339 OCP (未註明日期); S-340 OCP (未註明日期); S-342 OCP (未註明日期); S-343 OCP (14 Aug 2026); S-344 OCP (1 Sep 2026); S-345 OCP (27 Aug 2026); S-346 UALink Consortium (7 Apr 2026); S-347 IEEE 802.3 Ethernet Working Group (6 Aug 2026); S-348 CXL Consortium (18 Nov 2025); S-349 PCI-SIG (11 Jun 2025); S-350 PCI-SIG (5 Aug 2025); S-351 OCP (29 Apr 2025); S-352 OCP (13 Oct 2025); S-353 Ultra Ethernet Consortium (17 Oct 2023); S-354 OCP (未註明日期); S-355 DMTF (14 Sep 2026); S-356 DMTF (2 Jul 2026); S-357 DMTF (24 Sep 2026); S-358 OCP (未註明日期); S-359 CHIPS Alliance (27 Aug 2026); S-360 IETF (7 May 2026); S-361 IETF Datatracker (15 Jun 2026); S-362 OCP (18 Apr 2024); S-363 CHIPS Alliance (未註明日期); S-364 OCP (31 Jul 2025); S-365 CHIPS Alliance (15 Oct 2025); S-366 OCP (3 Jun 2026); S-367 OCP (未註明日期); S-368 OCP (15 Jul 2026); S-370 DMTF (25 May 2025); S-371 OCP (未註明日期); S-372 OCP (14 Dec 2025); S-377 OpenBMC project (未註明日期); S-378 CHIPS Alliance (2025); S-379 IOWN Global Forum (10 Feb 2026); S-380 OCP (未註明日期); S-381 OCP (未註明日期); S-384 3GPP (未註明日期); S-385 3GPP (10 Jun 2026); S-386 O-RAN ALLIANCE (8 Dec 2025); S-387 Open Web Foundation (未註明日期); S-388 Open Web Foundation (未註明日期); S-389 OCP (14 Aug 2026); S-390 SONiC project (未註明日期); S-391 JEDEC Solid State Technology Association (未註明日期); S-392 IEEE SA (3 Mar 2025); S-393 Ultra Ethernet Consortium (21 Oct 2024); S-394 DMTF (18 Sep 2025); S-395 SNIA (18 Sep 2015); S-396 Ultra Accelerator Link Consortium (未註明日期); S-397 Compute Express Link Consortium (未註明日期); S-398 Universal Chiplet Interconnect Express (未註明日期); S-399 CXL Consortium (12 Feb 2026); S-400 DMTF (17 Oct 2001); S-405 CHIPS Alliance (未註明日期); S-406 OCP (未註明日期); S-408 OCP (21 Sep 2026)

政府與監管機關. S-373 NIST (13 Aug 2024); S-374 U.S. National Security Agency (Apr 2024); S-375 Federal Register (25 Jun 2026); S-376 Office of Management and Budget (24 Jun 2026); S-404 U.S. Department of Justice (19 Sep 2025)

研究機構. S-053 Nature Electronics (19 Aug 2026); S-159 Li et al. (1 Mar 2022); S-293 Jouppi et al. (Apr 2023); S-407 Maruf et al. (6 Jun 2022)

市場研究. S-006 TrendForce (17 Aug 2026)

媒體與法律評論. S-285 Tech Times (15 Aug 2026); S-286 Tech Times (12 Aug 2026)

第7章

政策、能源與國家層級曝險

出口規則、電網併接條件與安全認證（certification），如今決定了AI容量可於何時、何地交付；其時鐘慢於晶片，所涉曝險亦非任何單一司法管轄區所能掌控。

適用讀者： 政府與政策人員、產業與採購（procurement）、財務；出口規則及企業展望對中國的假設見第7.1節，電網規則與能源需求見第7.2節與第7.3節，認證（certification）見第7.4節，國家層級曝險見第7.5節。

晶片以季為單位下訂；環繞其周邊的各項許可則依較慢的時鐘運行。出口許可決定加速器（accelerator）可運往何處，電網併接決定場址何時能取用電力，安全列名則決定一套新的800伏特直流系統是否得以安裝。上述各道程序均非由Open Compute Project（OCP）所撰寫，而每一道都可能使已建置完成的容量無法投用。本章先將各項規則視為法律工具來研讀，記明其發布機關、日期與狀態，其後再看這些規則必須容納的用電需求，以及其背後的國家集中度。本章要問的是：如今決定人工智慧（AI）容量何時到位的，是哪些許可，而非哪些零組件？

7.1 出口管制已出現於企業規劃之中

出口管制（export control）看似是銷售完成之後才處理的合規事項。但對AI基礎設施而言，它如今已置於營收展望、產能計畫與場址選擇之中，因為效能最強的加速器（accelerator）屬須經許可的貨品。自2025年1月起，美國已將一套全球性許可框架改以範圍較窄的法律工具取代，分別繫於產品類別、目的地或指名當事方；中國亦已宣布對電子業所倚賴的礦產實施管制。對買方或供應商而言，其結果並非只有一條規則可讀，而是一份法律工具的行事曆，每一項各有其發布機關、日期與失效時點（表15）。

F 2025年1月的全球性許可框架，其撤銷於四個月內即獲宣布。 美國商務部產業安全局（Bureau of Industry and Security, BIS）於2025年1月15日發布臨時最終規則「Framework for Artificial Intelligence Diffusion」[S-409]。該規則就出口管制分類編號（Export Control Classification Numbers, ECCN）3A090.a與4A090.a之下的先進晶片，於全球範圍內要求取得許可；對包括日本、大韓民國與台灣在內的18個盟友目的地給予許可例外（licence exception）；並對以逾10^{26}次運算訓練而成的封閉模型權重實施管制[S-409]。2025年5月13日，該部宣布撤銷該規則，並發布指引，其中包含對特定Huawei Ascend處理器的風險警示[S-410]。一份後續規則草案於2026年3月初的媒體報導中有所描述，但在貿易法評論者於2026年3月13日就其內容作出摘述時，該草案尚未公開[S-411]。

F 2026年1月，一類運往中國的晶片轉為逐案審查。 產業安全局2026年1月15日發布並同日生效的規則「Revision to License Review Policy for Advanced Computing Commodities」，就出口至中國與澳門、總處理效能（total processing performance, TPP）低於21,000且動態隨機存取記憶體（dynamic random-access memory, DRAM）總頻寬低於6,500 GB/s的晶片設定逐案審查，此一類別即NVIDIA的H200與AMD的MI325X所屬[S-412]。申請人須就若干條件提出證明，其中包括美國本地供應充足，且晶圓代工產能未自美國客戶處移轉他用[S-412]。同日生效的第11002號公告，依第232條對特定先進運算晶片課徵25%關稅，並訂有豁免，其中包括供美國資料中心使用的進口品[S-413]。

F 2025年11月議定的美國暫停措施，將於OCP全球峰會（OCP Global Summit）結束後數週內到期。 產業安全局的關係企業規則（Affiliates rule）為2025年9月30日的臨時最終規則，將實體清單（Entity List）限制自動擴及由已列名當事方持股50%以上的實體；產業安全局自2025年11月10日起至2026年11月9日止暫停該規則[S-414]。中國於2025年10月9日宣布對稀土實施範圍廣泛的出口管制（export control）；依2025年11月1日白宮事實說明所載的安排，中國承諾暫停其全球實施，並就稀土、鎵、鍺、銻與石墨核發一般許可，美國則暫停關係企業規則一年[S-415]。該安排之下的美國暫停措施施行至2026年11月10日；該事實說明並未載明中國方面的終止日期[S-415]。

> “NVIDIA is not assuming any Data Center compute revenue from China in its outlook.” [S-010]

F NVIDIA對2027財年第3季的展望，假設來自中國的Data Center運算營收為零。 該假設涵蓋的是展望期間內的Data Center運算營收；並非表示NVIDIA預期全公司在中國均無營收[S-010]。第2季運往中國的Data Center Hopper產品出貨，占Data Center營收不到1%[S-002]。同一句話出現於迄今就2027財年（NVIDIA截至2027年1月下旬的年度）所提出的全部三次展望之中，分別於2026年2月25日、5月20日與8月26日提出[S-050; S-051; S-010]；NVIDIA 2025年8月針對2026財年第3季的展望，則採用不同的、僅就H20而言的措辭[S-103]。

F H20的紀錄顯示許可決定傳導至帳列營收的速度有多快。 NVIDIA公布其H20出口至中國自2025年4月起須取得許可，並於2026財年第1季就H20認列45億美元的費用[S-102]。該公司2025年8月提出的2026財年第3季展望，假設不會有H20運往中國，且該年度第2季未向位於中國的客戶銷售任何H20[S-103]。2027財年第1季並無Data Center Hopper產品運往中國，而一年前為46億美元[S-097]。

A 規劃變數已從一項產品轉為一個市場。 在NVIDIA 2025年8月與2026年2月的展望之間，其措辭自單一產品（H20出貨）轉為一個營收類別（來自中國的Data Center運算營收）[S-103; S-050]。因此，2026年2月以來所提出的每一次展望（圖7）均假設該類營收為零，而第2季售往中國的Hopper產品等出貨（占Data Center營收不到1%）[S-002]，係落在該假設之外，而非之內。這是本報告各項來源中，將中國專屬假設寫入企業財測指引的最直接案例（第8項發現）；該展望未載明理由，而NVIDIA的10-Q則指向美國的許可條件與中國政府的限制[S-503]。此事並未顯示NVIDIA對中國的較長期看法，亦未顯示其Edge Computing營收的情況。

F 中國正在建立國內加速器（accelerator）供應，且依媒體報導，亦限制部分進口。 Huawei 2025年9月18日的藍圖載明，Ascend 950DT將於2026年第4季供應，供可連結最多8,192顆該晶片的Atlas 950 SuperPoD使用；Huawei並發布其UnifiedBus 2.0垂直擴展互連的規格（specification），表示希望產業夥伴採用該協定[S-312]。依South China Morning Post於2026年8月的報導，北京在推動國產替代方案之際，限制了NVIDIA H200的採購[S-416]。NVIDIA財務長的說明提及Data Center Hopper產品，並指其對中國的出貨占Data Center營收不到1%[S-002]；其10-Q則表示，已獲許可的H200對位於中國客戶的銷售受到中國政府限制[S-503]。

F 美國現已將AI出口推廣為附帶含量規定的整套方案。 依第14320號行政命令，美國商務部國際貿易署於2026年4月10日徵求由預先組成的聯合體提出方案，以供應全堆疊AI出口套案，其硬體與基礎設施層包括晶片、伺服器、加速器（accelerator）、資料中心儲存、雲端服務與網路；方案提交期限為2026年6月30日[S-417]。依Simpson Thacher的摘述，該徵求公告禁止中華人民共和國（People's Republic of China, PRC）企業在任何聯合體中擔任角色，排除由中國企業開發的開放權重模型（open-weight model），並要求硬體與基礎設施層依價值計算至少含51%的美國成分[S-418]。2025年7月的AI Action Plan，在待出口的「full AI technology stack」中，於硬體、模型、軟體與應用之外並列入標準[S-419]。

F 晶片出口立法仍在審議中。 倡議團體The AI Policy Network於2026年8月31日的一封信函，列出針對2027財年國防授權法（National Defense Authorization Act, NDAA）所提出的三項參議院修正案：關於最高效能晶片許可要求的AI OVERWATCH Act；關於半導體製造設備多邊管制的MATCH Act；以及關於位置驗證的Chip Security Act[S-420]。美國國會研究處（Congressional Research Service）記載，眾議院於2026年7月22日以216票對212票通過其版本法案，參議院於2026年7月14日就程序動議的終結辯論表決以50票對46票未獲通過，且至2026年9月1日國防授權法尚未完成立法[S-421]。

其他企業在展望與申報文件中載明中國及出口規則的影響

F NVIDIA的季度申報文件，為其展望中的該句話提供了脈絡。 NVIDIA在其截至2026年7月26日當季的報告中表示，美國政府「granted licenses that would allow us to ship small amounts of H200 products to specific China-based customers, but such sales were restricted by the PRC government」[S-503]。該公司於2027財年上半年就過剩存貨與採購義務認列4億美元的H200相關費用，其後已獲許可的出貨在最近一季占Data Center營收不到1%[S-503]。已獲許可的H200於出貨前須在美國接受查驗，因而在輸入美國時須負擔25%關稅，而NVIDIA迄今未能將其中任何部分轉嫁給客戶[S-503]。

F AMD將對中國的MI308銷售自其展望之外移入其中，其後便不再提及中國。 該公司2025年第3季的展望「not include any revenue from AMD Instinct MI308 shipments to China as our license applications are currently under review by the U.S. Government」[S-485]，其第4季展望則再次將其排除，且未說明理由[S-502]。就2026年第1季，AMD預期營收約98億美元，「including approximately $100 million of AMD Instinct MI308 sales to China」，此前2025年第4季的該類營收約為3.9億美元，惟未載明許可依據[S-486]。該公司2026年第2季與第3季的展望均未附中國相關的限定說明；就第3季，AMD「expects revenue to be approximately $13 billion, plus or minus $300 million」[S-588; S-111]。

F 設備製造商於2025年公布了出口規則影響的估計值。 Applied Materials在2025年10月的一份即時報告中估計，產業安全局的關係企業規則（Affiliates rule）將使其2025財年第4季淨營收減少約1.1億美元，2026財年減少約6億美元；該等估計係在該規則自2025年11月10日起暫停之前所作[S-504; S-505]。KLA於同月表示，其展望「is inclusive of the revenue impact related to additional market access loss related to certain customers in China」，並就2025年12月當季與2026曆年各估為約3億至3.5億美元[S-487]。

KLA 2026年7月的信函將中國列為季度營收的26%，但未提供此類估計[S-587]。

F ASML將其中國業務下滑歸因於客戶需求，並將其預測區間與出口管制（export control）的協商相連結。 ASML原預期中國客戶需求、因而其2026年中國淨銷售額，將較2024年與2025年「decline significantly」[S-506]。2026年4月，該公司表示其2026年區間「accommodates potential outcomes of ongoing discussions around export controls」，但未指名任何規則或國家[S-507]。該公司7月的發布將2026年區間上調至「between €43 billion and €45 billion」，惟未再重複該句話[S-025]。

F 設計軟體曾短暫受此類規則規範，記憶體晶圓廠至今仍受其規範，中國亦有自己的一套。 Synopsys在產業安全局2025年5月29日的信函告知其涉及中國的新出口限制後，暫停其2025財年的財測指引；Cadence則公布，當交易一方位於中國時，電子設計自動化軟體與技術的出口須取得新的許可[S-508; S-509]；產業安全局於2025年7月2日撤銷上述兩項[S-589; S-590]。產業安全局自2025年12月31日起，將Samsung China Semiconductor、SK hynix Semiconductor（China）與Intel Semiconductor（Dalian）自經認可最終使用者清單中移除，並預期每年將增加約1,000件許可申請[S-510]。在相反方向上，中國的網路安全監管機關禁止關鍵資訊基礎設施營運者採購Micron產品[S-511]，Cambricon則將其被列入實體清單（Entity List）列為供應鏈風險[S-512]。

F 依地區劃分的營收，衡量的是產品運往何處或客戶位於何處，而非其使用地點。 Marvell公布截至2026年8月1日當季營收的42%為運往中國的出貨，並表示「a substantial majority of the shipments made to China relate to sales to non-China based customers that have factories or contract manufacturing operations located within China」[S-515]。Lam Research依客戶設施所在地歸屬營收，公布2026財年中國占34%[S-516]；KLA依客戶所在地歸屬營收，公布2026財年中國占29.8%，2024財年則為42.8%[S-514]。ASML 2025年對中國的淨銷售額為95億歐元，占327億歐元中的29.1%，2024年則為36.1%[S-517]。

A 展望可以將基本需求與規則允許被滿足的需求區分開來。 儘管已有獲許可的出貨發生，NVIDIA仍未在其展望中計入任何來自中國的Data Center運算營收；AMD曾排除對中國的MI308銷售，其後計入，之後則不再提及中國；Applied Materials與KLA則於2025年公布了某項規則影響的估計值[S-010; S-503; S-486; S-111; S-504; S-487]。客戶願意購買者，與合法得以出貨、進口與定價者之間的落差，形成一道監管楔子：它會改變展望、存貨與費用，一如NVIDIA的H20與H200費用（NVIDIA將其與該等產品需求減少相連結），卻未必改變運算需求總量[S-102; S-503]。至於關稅由誰負擔——已獲許可的H200由NVIDIA自行承擔，Celestica則由客戶承擔——所改變的是利潤率，而非數量[S-503; S-498]。

A 區域營收占比並非區域需求的完善指引。 在申報文件依出貨目的地或客戶設施歸屬營收之處，中國的占比即如Marvell所述，混合了中國終端需求與全球客戶在中國的製造[S-515]。KLA與ASML自2024年以來中國占比的下滑，主要來自其他地區成長較快：KLA的中國營收維持在約40億至42億美元，而其總額成長38%；ASML的中國銷售額下滑7%，而其總額成長16%[S-514; S-517]。各份申報文件並未將此一變化歸因於出口規則或晶圓廠的設置地點。就第5.7節與第7.5節中的區域研判（reading）而言，製造地點與最終使用地點須分別對應，而一家企業的中國占比並非中國AI需求的衡量指標。

IF 若2026年11月的暫停措施到期而未獲延長，企業展望將在全球需求顯現之前先吸收其影響。 此一研判（reading）成立的前提是：關係企業規則（Affiliates rule）的暫停於2026年11月9日到期，且無後續安排[S-505; S-415]；Applied Materials 2025年就關係企業規則所作的估計，以及KLA就管制擴大所作的估計，顯示影響最先會出現在何處[S-504; S-487]。應觀察NVIDIA下一次展望中的該句話、AMD是否重新納入中國相關的限定說明，以及設備製造商在11月9日之後的首次陳述。若延長措施獲得公布，或美國許可範圍擴大且中國政府的限制放寬至足以使NVIDIA再次將中國納入其展望，則此一研判不成立[S-503]。

出口管制對開放社群的意義

F 美國規則將公開發表資訊（published information）與依其製成的物項分別處理。 依出口管理條例（Export Administration Regulations, EAR）第734.3(b)(3)條，已公開發表，或為第734.10條所述標準相關活動而釋出的資訊與軟體，不受EAR管轄[S-422]。第734.7(a)條將技術或軟體在未限制其後續散布的情形下向公眾提供者，視為已公開發表；而第734.7(b)條則使歸類於ECCN 5D002之下已公開發表的加密軟體，除符合所載條件外，仍受EAR管轄[S-423]。第734.10條僅適用於所載分類的物項，以及已公開發表或預定公開發表的標準[S-424]。

A OCP的規格（specification）層資訊；依其製成的機櫃與加速器（accelerator）則屬物項。 表15中的美國法律工具繫於交易——一項出口、一項進口、一個指名當事方或一個目的地——而無一項繫於已公開發表規格的文本[S-422; S-412; S-315]。有三類素材須由各參與者自行進行合規審查。發表前在會員之間共享的工作，以及公開會議之外的技術交流，僅在屬所載分類的物項、且係為已公開發表或預定公開發表的標準而釋出時，才為第734.10條所排除；而歸類於ECCN 5D002之下已公開發表的加密軟體，涉及韌體與安全工作（第6.4節），除符合所載條件外仍受EAR管轄[S-424; S-423]。以上所述為各項規則的結構，而非任何文件、參與者或出貨的狀態，亦非法律意見。

IF 若2026年11月的各項暫停措施失效，曝險將移入供應鏈，而非移入規格（specification）。 若關係企業規則（Affiliates rule）的暫停於2026年11月9日到期而未獲延長，實體清單（Entity List）限制即自動擴及由已列名當事方持股50%以上的實體[S-414]；產業安全局已於2023年3月將浪潮集團（Inspur Group）列名，並於2025年3月將包含Inspur Taiwan在內的六家浪潮子公司列名[S-425; S-426]。依該等條文的文義，此將涉及與該類關係企業之間的硬體交易，而非已公開發表的OCP文件。應觀察聯邦公報（Federal Register），以及2025年11月的安排在2026年11月9日至10日之前是否獲得續行[S-415]。若延長措施或後續安排先行公布，或若有某項法律工具直接針對參與開放規格工作本身加以規範，則此一研判（reading）不成立。

> **觀察重點**
> **2026年11月9日至10日。** 產業安全局是否延長關係企業規則（Affiliates rule）的暫停，以及2025年11月的安排是否續行；2026年9月25日的白宮事實說明既未提及半導體，亦未提及延長[S-414; S-415; S-428]。
> **2026年12月4日。** 第11052號公告之下的多晶矽關稅生效[S-427]。

7.2 電網併接與大型負載規則決定時程

一座裝滿加速器（accelerator）的機櫃可在數月內建成；為其送電所需的變電所、輸電線路與併接研究則可能耗時數年。此一落差使大型負載（large load）併接規則——誰先獲得研究、誰負擔升級費用、誰在緊急情況下必須削減用電、誰得以自帶電力供應——對交付日期的決定性，與任何零組件相當。2025年與2026年間，美國、歐洲與亞洲的監管機關與電網業者，已著眼於資料中心而開始改寫這些規則，或其周邊條件。方向是一致的：對能夠彈性調節、緩衝或自行供電的負載給予較佳條件，對無法做到者則課予更嚴格的義務。

F 2026年，聯邦層級對大型負載（large load）的行動已自公告進入命令階段。 2025年7月23日的第14318號行政命令，加速了對新增逾100百萬瓦（megawatt, MW）AI負載或承諾投入至少5億美元的資料中心專案的聯邦許可程序[S-429]。2025年10月23日，美國能源部長指示聯邦能源管理委員會（Federal Energy Regulatory Commission, FERC）就20 MW以上負載的併接，於RM26-4案號下考慮發布擬議規則制定預告；至2026年7月22日，該案仍停留在該階段[S-430]。2026年6月18日，FERC命令其所監管的六家區域電網業者，其中包括PJM Interconnection（PJM），須於60日內就大型負載研究、成本移轉、共址與彈性負載（flexible load）服務，為其費率章程提出正當理由或提出改革方案[S-068]。

F FERC亦已命令為運算類負載訂定可靠度標準。 2026年7月16日，FERC於RD26-7-000案號下指示北美電力可靠度公司（North American Electric Reliability Corporation, NERC），就將運算類負載併入大電力系統所生風險，研擬新訂或修訂的可靠度標準，該等標準及NERC程序規則的相關修訂須於2026年12月31日前完成[S-431]。NERC將該問題描述為在故障與正常情況下管理用電量的快速波動，而此類波動可能使電網失穩[S-230]。2025年12月，FERC要求PJM就服務共址負載的發電設施訂定費率條件，並為該類負載提供四項輸電服務，其中兩項為非確定性服務[S-041]。

F 2026年7月的單一事件，顯示了故障穿越（ride-through）規則的用意。 依PJM與Dominion所述，2026年7月22日北維吉尼亞一條230 kV線路跳脫，使約3,800 MW的大型負載（large load）未經預警即脫離並轉為備用供電[S-229]。2026年9月10日，PJM提出針對大型負載斷接事件的可靠度標準，並將於2026年11月向FERC提交，同時指出NERC正在研擬大型負載指引與故障穿越建議[S-061]。PJM 2026年8月13日的提案框架，將在緊急情況下使未自帶新增電力供應的新建大型負載，較其他用戶優先被削減或移轉，並將其2024年至2030年預測需求成長32 GW中的30 GW歸因於資料中心[S-042]。

F 德州已將大型負載（large load）條件寫入法律。 2025年德州參議院第6號法案要求德州公用事業委員會（Public Utility Commission of Texas, PUCT）就75 MW以上的負載訂定併接標準，研究費至少10萬美元，並須揭露重複申請與場內備用發電，且賦予德州電力可靠度委員會（Electric Reliability Council of Texas, ERCOT）在能源緊急情況下指示大型負載使用備用發電或削減用電的權力；PUCT關於批售輸電成本分攤的修訂規則須於2026年12月31日前完成[S-040]。2026年6月18日，PUCT核准ERCOT用於併接大型用戶的「Batch Zero」流程，當時ERCOT正追蹤逾438,000 MW的大型負載併接申請，其中近89%來自資料中心[S-032]。

A 併接條件如今取決於一個負載能做什麼，而不僅取決於其在排隊佇列中的位置。 德州的法律與PJM的提案方向趨同，FERC 2026年6月要求說明理由的命令亦指向同一方向：能夠削減用電、故障穿越（ride-through）、緩衝其功率波動或自帶發電的大型負載（large load）可獲較佳條件，而無法做到者在緊急情況下會最先被削減[S-040; S-042; S-068]。兩項研究說明了彈性為何具有價值：杜克大學發現，76 GW的美國新增負載可在年平均負載削減（curtailment）0.25%的條件下併入[S-432]；2026年一篇預印本則發現，2023年的PJM系統尚有空間容納最多32 GW的彈性負載（flexible load），而削減量為該負載年用電量的0.1%[S-433]。因此，機櫃儲能、可控的功率爬升與電網遙測（telemetry）便取得了併接上的價值（第4.4節）。

IF 若最終規則以技術性能界定彈性，機櫃層級的緩衝即成為一項併接資產。 此一研判（reading）成立的前提是：FERC就RM26-4案號所採行動與六家業者的回應、NERC須於2026年12月31日前完成的標準、PJM 2026年11月的提交，以及PUCT的成本規則，均以可衡量的方式載明彈性負載（flexible load）服務，例如功率爬升率、故障穿越（ride-through）與負載削減（curtailment）回應[S-430; S-431; S-061; S-040]。

表14。 各份展望與申報文件對中國的處理方式各不相同

NVIDIA仍排除一個中國營收類別；AMD先排除、後計入一項產品，如今則不再提及；設備製造商則於2025年載明了估計值。

企業	文件與期間	形式	所述內容	來源
NVIDIA	2026財年第2季展望，2025年5月28日提出	影響估計	該展望「reflects a loss in H20 revenue of approximately $8.0 billion due to the recent export control limitations」	[S-102]
NVIDIA	2026財年第3季展望，2025年8月27日	排除一項產品	假設無H20運往中國	[S-103]
NVIDIA	2026財年第4季展望，2025年11月19日	未載明	無中國相關文句；第3季H20銷售「insignificant」	[S-513]
NVIDIA	2027財年第1、2與3季展望，2026年2月25日、5月20日與8月26日	排除一個營收類別	假設來自中國的Data Center運算營收為零；2027財年第2季運往中國的Hopper出貨占Data Center營收不到1%	[S-050; S-051; S-010; S-002]
NVIDIA	截至2026年7月26日當季的10-Q	影響與風險	就少量H200核發許可；銷售受中國政府限制；4億美元費用；25%關稅未予轉嫁	[S-503]
AMD	2025年第3季與第4季展望	排除	無來自中國的MI308營收；8月時許可申請仍在美國審查中	[S-485; S-502]
AMD	2026年第1季展望	計入	對中國的MI308銷售約1億美元；2025年淨費用約4.4億美元	[S-486]
AMD	2026年第2季與第3季展望，2026年5月5日與8月4日	未載明	未附中國或MI308相關的限定說明	[S-588; S-111]
Applied Materials	8-K，2025年10月2日	影響估計	關係企業規則（Affiliates rule）將使2025財年第4季淨營收減少約1.1億美元，2026財年減少約6億美元	[S-504]
KLA	展望，2025年10月29日；信函，2026年7月28日；2026財年10-K	計入；風險	因失去中國部分客戶而產生的營收影響，2025年12月當季與2026曆年各約3億至3.5億美元；2026年7月的信函未提供估計；中國占2026財年營收29.8%；未能取得出口許可已損及未出貨訂單（backlog）	[S-487; S-587; S-514]
ASML	展望，2025年10月15日、2026年4月15日與7月15日	已載明的假設	2026年中國銷售將「decline significantly」；4月的區間將出口管制的可能結果納入考量；7月未再重複	[S-506; S-507; S-025]
Synopsys；Cadence	8-K，2025年5月與7月	暫停；風險	Synopsys於產業安全局信函之後暫停財測指引；Cadence公布涉及中國當事方的EDA出口須取得許可；產業安全局於2025年7月2日撤銷上述兩項	[S-508; S-509; S-589; S-590]
Samsung、SK hynix與Intel Semiconductor（Dalian）在中國的實體	產業安全局規則，2025年12月31日生效	規則	經認可最終使用者資格遭移除；預期每年將增加約1,000件許可申請	[S-510]
Micron	2025財年10-K	來自中國措施的風險	中國的網路安全監管機關禁止關鍵資訊基礎設施營運者採購Micron產品	[S-511]
Cambricon	2025年年度報告	風險	列入實體清單（Entity List）為供應鏈穩定的風險	[S-512]
Marvell	截至2026年8月1日當季的10-Q	地區界定	依出貨目的地計算，中國占淨營收42%，其中多數與非中國籍客戶在中國的工廠有關	[S-515]

各份文件就中國與出口規則所述內容，依各公司自身用語、於所示日期呈現；較晚的文件可能未再提及。文件閱讀至2026年9月27日；來源見最後一欄。
範圍 各季別均為各公司的財務季度。各項估計值均為各公司自行所作，並作成於所示日期；其中數項早於關係企業規則（Affiliates rule）於2025年11月10日暫停之前[S-505]。
引號內為各公司自身的用語；其他欄位則為摘述。

應觀察上述四項決定。若彈性僅被界定為一項契約上的削減權利，從而獎勵契約而非設備，或若大型負載（large load）一律被視為確定性負載，則此一研判不成立。

F 歐洲的規則在愛爾蘭透過併接條件發揮作用，在歐洲聯盟全境則透過資訊揭露。 愛爾蘭公用事業管理委員會（Commission for Regulation of Utilities, CRU）於2025年12月12日決議，併入電網的新建資料中心仍將被要求在場內或鄰近地點提供與其所申請最大輸入容量相當的發電或儲能容量，並須以愛爾蘭境內額外的再生能源電力專案滿足其年度需求的至少80%[S-434]。依歐洲聯盟（European Union, EU）修訂後的能源效率指令，已裝置資訊科技（information technology, IT）功率至少500 kW的資料中心，須向歐洲資料庫報送能源與水資源指標[S-435; S-436]。2026年9月21日，歐盟執委會通過一套共通評級制度；其預期首批標示於2027年出現，並預期至2030年資料中心將超過歐盟用電需求的3%[S-437]。

F 在亞太地區，效率條件與容量得於何處增設相互綁定。 新加坡第二輪資料中心申請徵求由資訊通信媒體發展局於2025年12月1日開放，規模至少200 MW，要求在100% IT負載下電力使用效率（power usage effectiveness, PUE）達1.25或更佳[S-438]。中國國家發展和改革委員會於2024年7月23日發布的綠色資料中心專項行動計畫，引導新建大型資料中心優先設於國家樞紐節點叢集，並將樞紐節點專案的PUE上限定為1.2[S-439]。日本要求自2029會計年度起新建的資料中心須達到PUE 1.3或以下[S-440]；韓國的人工智慧基本法（Framework Act on AI）預定於2026年1月22日施行，允許國家對建設AI資料中心提供支援[S-441]。

A 在美國以外，操作槓桿是資訊揭露、選址與自行供電，而非排隊制度改革。 歐盟的制度將所報送的數據轉換為公開等級，並須先經歐洲議會與理事會兩個月的審查才會生效[S-437]；新加坡以效率作為其容量配置的條件，中國對樞紐節點專案設定PUE上限，日本則對自2029會計年度起興建的資料中心訂定PUE限值[S-438; S-439; S-440]；愛爾蘭則使併接本身以自帶電力供應為條件[S-434]。由於各項PUE門檻係在不同條件下載明，單一設計值並不足以顯示在各司法管轄區均屬合規。對設備供應商而言，所報送PUE與用水量背後的衡量方法，正是OCP的效率與遙測（telemetry）工作得以發揮作用之處。

? 未解問題——一個大型負載（large load）如今要等多久才能取得電力。 上述各項法律工具訂定了研究程序、成本規則與削減義務，但公開紀錄並未提供自大型負載申請至送電啟用的典型時間；ERCOT的數字計算的是申請件數，而非併接件數[S-032]。缺少該項數字，送電時程就必須逐一場址確立，而它仍是整條鏈條中最慢的時鐘。

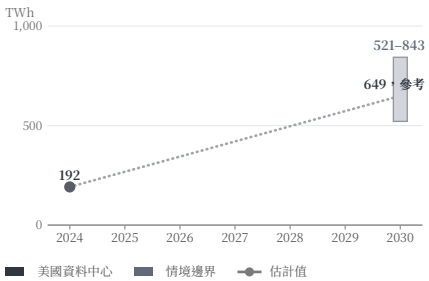
> **觀察重點**
> **2026年11月。** PJM計畫向FERC提交其所提出的大型負載（large load）故障穿越（ride-through）標準[S-061]。
> **2026年12月31日。** NERC針對運算類負載的可靠度標準[S-431]，以及PUCT關於批售輸電成本分攤的修訂規則[S-040]，均須於此日前完成。

7.3 電網必須滿足的能源需求

有兩類數字描述資料中心對電網的要求，而兩者極易混淆。以太瓦時（terawatt-hour, TWh）計的電量，是一個機群一年內的消耗量；它決定燃料用量、排放與帳單。以吉瓦計的容量，則是必須在同一時刻可供使用的功率；它決定變電所、輸電與發電廠。在缺少機群運行強度假設的情況下，以其中一種單位所作的預測，不能當作另一種單位的預測來閱讀；而美國數列與全球數列亦採不同方法建立。本節將上述四者分別處理。

F 勞倫斯柏克萊國家實驗室（Lawrence Berkeley National Laboratory, LBNL）的參考情境，使2030年美國資料中心用電量增至三倍以上。 LBNL在其2026年6月發布的報告中，將2024年美國資料中心用電量定為192 TWh，相當於美國用電量的4.7%[S-442]。其2030年參考情境為649 TWh，相當於NERC 2025年長期可靠度評估所載美國預測用電量的11.8%，並落在521至843 TWh（即9.5%至15.3%）的綜合不確定性區間內[S-442]。該報告摘要表示，至2030年資料中心「could account for」美國電力的11.8%；該區間涵蓋的是情境（scenario）邊界，並非統計上的區間[S-442]。

圖27。 LBNL的參考情境使2030年美國資料中心用電量增至三倍以上



美國資料中心用電量，單位為TWh：LBNL對2024年的估計（占美國電力的4.7%）、其2030年參考情境（占美國預測用電量的11.8%），以及其綜合不確定性情境（scenario）的上下界（9.5%至15.3%），取自2026年6月發布的報告[S-442]。
範圍 僅限美國範圍，與圖11中國際能源總署（IEA）的全球數列不可比。該區間涵蓋LBNL的情境（scenario）邊界，並非統計上的區間。

A 在LBNL的各項敏感度分析中，營運假設對2030年估計值的影響大於出貨假設。 在該報告所測試的四項單一參數變動中，AI伺服器閒置功耗與使用率提高對2030年估計值的影響最大，使其自649 TWh的參考值升至782 TWh；而專用繪圖晶片出貨與安裝量增加，僅使其升至664 TWh[S-442]。資訊科技設備安裝量減少（578 TWh）與AI晶片壽命縮短（590 TWh）則使其下降[S-442]。就電網規劃而言，機群如何運行是至少與出貨晶片數量同等重要的需求變數，而這正是機櫃電力管理與遙測（telemetry）進入能源議題之處（第4.2節）。

F 全球估計值建立於不同基礎之上而上升，美國的各項預測亦指向同一方向。 國際能源總署（International Energy Agency, IEA）將2025年全球資料中心用電量定為485 TWh，成長17%，並預估

表15。 出口法律工具繫於產品類別、當事方與目的地，而2027年4月之前有四個日期可能改變現況

各項法律工具作用於產品、當事方、目的地與關稅，而非作用於已公開發表的規格（specification），且可能改變這些工具的日期中有三個落在2026年11月9日至12月4日之間。

法律工具	發布機關	日期	狀態（2026年9月27日）	涵蓋內容	來源
Framework for Artificial Intelligence Diffusion（臨時最終規則）	BIS	2025年1月15日發布	2025年5月13日宣布撤銷；後續規則草案經報導，但至2026年3月13日仍未公開	ECCN 3A090.a與4A090.a晶片於全球範圍須取得許可；18個盟友目的地適用例外；逾10^{26}次運算的封閉模型權重	[S-409; S-410; S-411]
關係企業規則（Affiliates rule，臨時最終規則）	BIS	2025年9月30日	自2025年11月10日起至2026年11月9日止暫停	實體清單（Entity List）限制擴及由已列名當事方持股50%以上的實體	[S-414]
稀土出口管制（rare-earth export controls）	中國	2025年10月9日宣布	已承諾暫停全球實施；美國事實說明未載明終止日期（美國方面的條款施行至2026年11月10日）	稀土；就稀土、鎵、鍺、銻與石墨核發一般許可	[S-415]
對G42（阿拉伯聯合大公國）與HUMAIN（沙烏地阿拉伯）的核准	美國商務部	2025年11月19日	已核准	相當於最多35,000顆GB300晶片，並附安全與報告要求	[S-314]
Revision to License Review Policy for Advanced Computing Commodities	BIS	2026年1月15日發布並生效	現行有效	對中國與澳門、TPP低於21,000且DRAM頻寬低於6,500 GB/s者實施逐案審查	[S-412]
第11002號公告（第232條）	美國總統	2026年1月15日生效	現行有效	對特定先進運算晶片課徵25%關稅；豁免範圍包括供美國資料中心使用的進口品	[S-413]
American AI Exports Program：徵求預先組成的聯合體（第14320號行政命令）	國際貿易署	2026年4月10日	方案提交期限為2026年6月30日	全堆疊AI出口套案；排除中國企業；硬體須含51%美國成分	[S-417; S-418]
Enhanced Favorable Treatment for the United Arab Emirates	BIS	2026年7月10日生效	現行有效；G42與Core42的授權（authorisation）若無通知，將於2027年4月6日到期	阿拉伯聯合大公國移至國家組別A:5；指名當事方得免許可接收	[S-315]
第11052號公告（第232條，多晶矽）	美國總統	2026年8月11日發布	關稅自2026年12月4日起課徵	對多晶矽與晶圓課徵15%關稅並設最低進口價格；日本、韓國、台灣與歐洲聯盟等適用15%的合計上限	[S-427]
出口管理條例第734.3(b)(3)、734.7與734.10條	BIS	現行條文	現行有效	公開發表資訊（published information）與標準相關釋出不受EAR管轄；已公開發表的5D002加密軟體除符合所載條件外仍受管轄	[S-422; S-423; S-424]

涉及AI硬體流通的各項法律工具，依發布機關與日期排列，並附2026年9月27日的狀態；狀態可能變動。BIS：美國商務部產業安全局。來源見最後一欄。

各日期為各來源所載的發布日或生效日；狀態欄位記錄的是2026年9月27日的情形。

2030年約為950 TWh，相當於全球用電需求的約3%[S-005]；第2.5節詳述該數列，該數列不與LBNL的數列接續。就美國而言，美國能源資訊署（Energy Information Administration）預測商業部門售電量2026年成長3.3%、2027年成長2.7%，而用電量在資料中心開發的帶動下達到歷史高位[S-443]。顧問公司Grid Strategies指出，各公用事業對美國尖峰需求成長的五年預測為166 GW，其中資料中心約占需求成長的55%[S-444; S-445]。

F TrendForce預期2026年全球資料中心電力需求容量將成長約31%。 在2026年9月16日發布的一項預測中，TrendForce將2025年全球資料中心電力需求容量定為122.9 GW、2026年為161 GW，其中AI伺服器估計占2026年需求容量的33.4%[S-043]。此一指標涵蓋一般伺服器、AI伺服器、其他資訊科技設備，以及散熱與電力供應等非資訊科技設備[S-043]。就2030年，TrendForce預估需求容量為490.7 GW，而資料中心可取得的電網容量為222.6 GW（假設所有其他用電需求均獲完全滿足）——缺口為268 GW——且美國的缺口將超過170 GW[S-043]。

> **範圍 GW與TWh**
> 吉瓦（gigawatt, GW）數字指的是必須在某一時刻可供使用的功率；太瓦時（terawatt-hour, TWh）數字指的是一年內所使用的電量。TrendForce的容量數列與LBNL及國際能源總署（IEA）的電量數列所衡量的是不同事物，絕不換算、相加或繪為同一條線。

F 部分買方正自行簽約取得發電，而非等待電網。 Oracle與Bloom Energy於2026年4月擴大雙方合作，為AI基礎設施部署最高2.8 GW的Bloom燃料電池[S-231]。Meta於2026年1月宣布多項核能協議，支持至2035年最高6.6 GW的新增與既有潔淨能源：其中逾2.1 GW來自兩座營運中的Vistra電廠及其機組增容，另有新建反應爐，最早可於2030年與2032年供電[S-446]。日本對美國戰略投資的第二批專案於2026年3月19日宣布，其中包括售電承購將供應共址資料中心的專案[S-447]。

A 場內供電移動了限制，而非消除限制。 一組燃料電池機群或一座專用電廠可繞過部分併網排隊佇列（interconnection queue），但仍須仰賴訂單已滿的發電設備：GE Vernova公布其燃氣設備未出貨訂單（backlog）與產能時段預訂達116 GW，並預期至2026年底至少125 GW已簽約[S-012]；Siemens Energy則表示其燃氣渦輪機已售罄至2028年[S-026]。這些協議中的新建反應爐最早自2030年起才會到位[S-446]；TrendForce並預期至2028年止，美國資料中心需求中有相當大的部分仍將透過電網供電，而表後（behind the meter）電力正成為一項額外解方[S-043]。電網時鐘因此被換成了設備與許可時鐘（第5.5節）。

IF 若電網供電於2028年之後趨緩，為部署設下上限的將是可取得的電力，而非加速器（accelerator）。 TrendForce預期電網供電延誤於2028年之後將更為頻繁[S-043]；此一研判（reading）成立的前提是，該缺口在各公用事業的規劃中持續存在，且場內供電仍僅為部分途徑。應觀察各公用事業的尖峰需求預測與併接核准、第7.2節中FERC與PJM的各項決定、渦輪機的交付時段，以及LBNL與國際能源總署（IEA）的下一次更新。若彈性負載（flexible load）服務使既有電網得以吸納大部分成長——一如杜克大學與PJM的研究所指出，在削減量很小的情況下即屬可行[S-432; S-433]——或若需求容量的成長遠低於TrendForce的預測，則此一研判不成立。

7.4 800 VDC與液體的安全與認證

規格（specification）告訴供應商設備應如何連接；法規條文或產品列名則告訴檢查人員設備是否得以安裝。對於以±400或800伏特直流（volts direct current, VDC）供電的機櫃而言，規格的推進速度快於法規。至於液冷（liquid cooling），法規所繫者較少在於迴路設計，而更在於流體的化學組成，以及場址所消耗的水量。這兩者都是部署須等待OCP之外機構的環節。

F 800 VDC路徑已有規格（specification）與認證（certification）工作，但尚無法規條文。 2026年8月11日，來自Google、Microsoft與NVIDIA的工程師於OCP部落格撰文表示，三方正同步與UL Solutions、美國國家防火協會（National Fire Protection Association, NFPA）、IEEE及國際電工委員會（International Electrotechnical Commission, IEC）就安全認證與法規架構進行協作，並表示已有逾80家合作夥伴正在開發相容800 VDC的基礎設施[S-019]。此項工作背後的文件為：適用於±400 VDC側置電源機櫃（sidecar）的Diablo 400機櫃與電力規格，0.7.0版，2026年3月1日[S-027]；固態變壓器（solid-state transformer）規格0.3.0修訂版，2026年6月22日[S-054]；以及貼文所稱將於日後推出的原生800 V之Mt Diablo 2.0[S-019]。

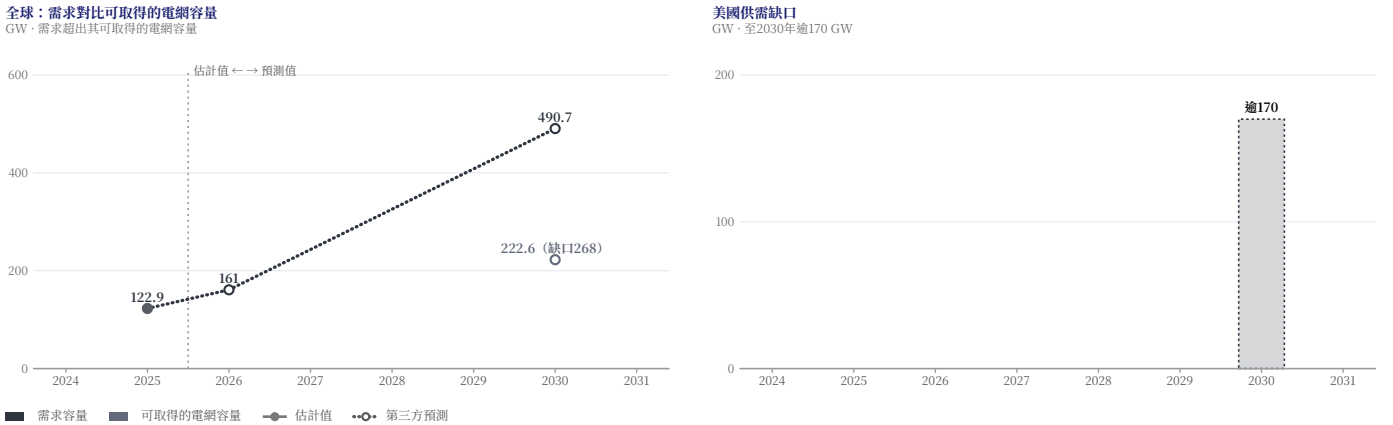
F 保護硬體與儲能各自負有其認證（certification）要求。 ABB將其SACE Infinitus描述為全球首款取得國際電工委員會（International Electrotechnical Commission, IEC）認證的固態斷路器（solid-state circuit breaker），斷路時間為25微秒[S-197; S-198]。在電池儲能方面，UL Solutions依UL 9540對系統進行認證，並依UL 9540A測試火勢蔓延[S-448]；OCP的儲能白皮書（white paper）記載，NFPA 855（2026年版）現已要求進行大規模火災測試[S-205]。OCP的低壓直流（low-voltage direct current, LVDC）白皮書1.1.0版，2026年9月2日，指出產品標準、法規與安裝規章將低壓界定為在1,000伏特交流電（VAC）或1,500 VDC及以下運作[S-059]。

A 落差在於已發布文件與可安裝產品之間。 由於800 VDC落在各項法規所界定的低壓範圍之內[S-059]，未解問題與其說是一個新的電壓等級，不如說是斷路器、匯流排槽、接頭與側置電源機櫃（sidecar）在Diablo 400規格（specification）所處理的百萬瓦等級規模下的列名認證[S-027]。一份規格可以獲得受理、一座機櫃可以完成展示，而列名認證仍在審理之中；介於兩者之間的步驟是認證（certification），而非規格（第4.2節）。因此，持有已列名保護裝置的供應商，以及撰寫測試方法的認證機構，掌握著一道OCP文件無法移除的關卡[S-019; S-197]。

? 未解問題——±400與800 VDC機櫃將受哪些法規條文管轄。 公開紀錄顯示已與UL Solutions、NFPA、IEEE及IEC展開協作[S-019]，但並無專門針對±400或800 VDC資料中心機櫃的已公布法規條文或產品列名。在尚無任一者之前，每一處安裝均取決於當地核准，這會拖慢案場之間的複製速度，並使規劃日期取決於認證（certification）日期，而非規格（specification）日期。

F 在液冷（liquid cooling）方面，法規所規範的對象是流體與水，而非迴路本身。 OCP關於丙二醇冷卻液的指引載明，PG25的冰點為−13至−9 °C，封閉迴路的典型運作溫度低於49 °C[S-216]。含氟氣體法規Regulation (EU) 2024/573規定，氫氟碳化物至2050年前自歐盟淘汰[S-449]；而針對全氟及多氟烷基物質（per- and polyfluoroalkyl substances, PFAS）的全面限制，仍處於歐洲化學總署（European Chemicals Agency）的限制程序之中[S-450]。歐盟的

圖28。 TrendForce預估2030年資料中心電力需求與可取得電網容量之間將出現268 GW的缺口



資料中心電力需求容量與資料中心可取得的電網容量，單位為GW：TrendForce對2025年的估計，以及對2026年與2030年的預測，於2026年9月16日發布[S-043]。可取得的電網容量假設所有其他用電需求均獲完全滿足；全球缺口為TrendForce的268 GW（490.7減222.6）。美國面板呈現TrendForce對2030年美國缺口所給的下限，該公司稱其「will surpass 170 GW」[S-043]。

範圍 此為容量（必須同時可供使用的功率，單位GW），而非年用電量（TWh），因此與圖27或圖11不可比。全球數值與美國數值分置於不同面板，不相加亦不相抵。此為市場研究預測，而非公用事業或政府的數字。

評級機制就能源與水分別給予每一申報場址A至G的等級；依Freshfields的歸納，A級在能源方面意指電力使用效率（power usage effectiveness, PUE）為1.15或以下，在水方面意指水資源使用效率（water usage effectiveness, WUE）為0.1或以下，且兩者不合併為單一分數[S-451]。

A 法規曝險繫於流體的選擇。 如PG25之類的水—二醇迴路不含氟，而氫氟碳化物冷水機冷媒則落在含氟氣體淘汰的範圍，含氟的兩相冷卻（two-phase cooling）與浸沒式冷卻（immersion cooling）流體則面臨PFAS限制程序的曝險[S-216; S-449; S-450]。因此，在單相冷板（cold plate）、兩相與浸沒式之間的選擇（第4.3節），既是散熱決策，也是法規與供應決策。又因歐盟機制就能源與水分別評級[S-451]，一項以蒸發水分來節省能源的排熱設計，可能在取得一個等級的同時付出另一個等級的代價。

IF 若PFAS限制在未就封閉傳熱迴路設有排除條款的情況下獲得採納，兩相冷卻（two-phase cooling）與浸沒式冷卻（immersion cooling）設計將面臨流體供應的問題。 此一研判（reading）在該限制仍未定案、且含氟流體仍在其適用範圍內的期間成立[S-450]。應觀察歐洲化學總署（European Chemicals Agency）各委員會的意見、歐盟執行委員會的任何提案，以及流體供應商的驗證（qualification）公告。若獲採納的條文豁免封閉迴路傳熱用途，或不含氟的兩相流體達到已驗證產品的階段，使選擇重新回歸單純的散熱效能，則此一研判不成立。

7.5 國家層級曝險：集中度與韌性

AI硬體鏈中的每一個法律管轄區都倚賴其他管轄區。美國設計的加速器（accelerator）倚賴集中於台灣的晶圓代工與封裝產能，以及多由韓國企業供應的高頻寬記憶體（high-bandwidth memory, HBM），而中國則掌握電子業所用礦物的管制權[S-162; S-452; S-415]。各國政府正透過投資計畫與貿易條件採取行動。對政策而言，有用的區分在於韌性問題與標準問題之別：前者關乎產能實際坐落於何處，後者關乎共通介面能否讓買方驗證（qualification）第二供應來源（second source）。

F 台灣的晶圓代工與封裝集中度正向亞利桑那州延伸，而非被取代。 2026年7月16日，台灣積體電路製造股份有限公司（Taiwan Semiconductor Manufacturing Company, TSMC）宣布在亞利桑那州追加投資1,000億美元[S-084]，美國商務部則認定其投資總額為2,650億美元[S-163]；台積電的魏哲家（C.C. Wei）表示，「probably」會再建造四座或更多晶圓廠[S-084]。台積電預期其第二座亞利桑那晶圓廠將於2027年下半年開始量產[S-083]。TrendForce報導指出，台積電的Chip-on-Wafer-on-Substrate（CoWoS）封裝產能可自2026年底每月約13萬片晶圓，倍增至2028年底的26萬片，擴產以其位於台灣的AP7廠區與亞利桑那園區為重心[S-162]。

F 台灣與美國的貿易條件將送交立法院審議。 行政院經貿談判辦公室（Office of Trade Negotiations）表示，2026年1月15日簽署的投資備忘錄（memorandum of understanding, MOU）與2026年2月12日簽署的對等貿易協定（Agreement on Reciprocal Trade），訂定15%的對等關稅且不疊加於最惠國稅率之上；該備忘錄並未載明特定公司的投資金額，兩份文件將一併送交立法院[S-453]。美國商務部的一份摘要列有2,500億美元用於半導體晶圓廠，以及2,500億美元用於更廣泛半導體供應鏈的保證[S-454]。在國內，行政院計畫至2033年前投入新台幣3,000億元（約95億美元）於半導體[S-455]。

A 移轉在台灣以外增添了產能，但尚未降低台灣的集中度。 亞利桑那州的承諾屬台積電（TSMC）全公司層級的計畫，所報導的CoWoS產能倍增亦為全公司層級的目標，而擴產預期將以台灣與亞利桑那州為共同重心[S-162; S-084]。在美國各廠區公布量產產出之前，供AI加速器（accelerator）所用的先進製程晶圓代工與先進封裝（advanced packaging）仍集中於單一法律管轄區，而此一集中度屬韌性問題，非任何開放規格（specification）所能移轉。台灣的角色也正從「組裝」這一簡略代稱，轉向系統整合、先進封裝、光學與驗證，第5.7節即追溯此一轉變（發現9）。

F 依一項市場估計，兩家韓國企業供應大部分的高頻寬記憶體（high-bandwidth memory, HBM）。 據Seoul Economic Daily報導，Counterpoint Research將2026年第2季HBM營收占比列為SK hynix 50%、Samsung 33%、Micron 18%，以上為其公布之數字（三者合計為101%）[S-452]。SK hynix於第2季開始HBM4的大量出貨，並與約10家客戶簽有長期協議（long-term agreement）[S-003]；Samsung於2026年2月開始HBM4的量產與出貨，並預期其HBM銷售額在2026年將成長逾兩倍[S-056]。Micron表示HBM4已為其主要客戶的平台進入大量出貨[S-128]。

F 日本與歐洲資助國內產能，並訂定主權相關條件。 在日本，資訊處理推進機構（Information-technology Promotion Agency）於2026年6月5日完成對Rapidus 1,500億日圓的增資，Rapidus並以2027年前全面量產2 nm製程為目標[S-325]。歐盟執行委員會（European Commission）2025年4月9日的AI Continent Action Plan，擬動員200億歐元，用於最多五座AI Gigafactories[S-323]。2026年6月，執行委員會提出Cloud and AI Development Act，其最高主權等級要求服務供應商須由歐盟境內持有並控制[S-456]；另提出Chips Act 2.0以降低戰略依賴[S-457]。

F 中國、波斯灣與印度將需求與國內或主權計畫相互搭配。 中國掌握礦物方面的籌碼（leverage）（第7.1節），並正增添國內加速器（accelerator），例如Huawei公告於2026年第4季推出的Ascend 950DT[S-415; S-312]。2025年11月，美國商務部核准阿拉伯聯合大公國（United Arab Emirates, UAE）的G42與沙烏地阿拉伯的HUMAIN購買相當於最多35,000顆NVIDIA GB300晶片[S-314]；2026年7月10日生效的一項工業及安全局（Bureau of Industry and Security, BIS）規則，將阿聯移入Country Group A:5，而G42與Core42的授權（authorisation）若無進一步通知，將於2027年4月6日到期；沙烏地阿拉伯則不在適用範圍內[S-315]。印度的IndiaAI運算計畫列有來自2024年獲選入列的10家投標者的逾18,000顆圖形處理器（graphics processing unit, GPU）[S-458]；2026年2月，印度政府表示將在既有的38,000顆GPU之外再增加20,000顆[S-316]。

A 多數國家層級的曝險屬韌性問題，非標準所能解決。 產能實際坐落之處——先進製程晶圓代工與CoWoS在台灣、HBM在韓國、稀土受中國管制、電網併接受各國主管機關管轄——共通規格（specification）無法改變；能改變的是投資、貿易條件與許可[S-162; S-452; S-415]。標準發揮作用的範圍較窄：開放的機櫃、電力、冷卻或管理介面，可讓買方更快驗證（qualification）第二家供應商或組裝廠區（第6.1節）。中國則從另一端顯示了此一界限：其開放數據中心委員會（Open Data Center Committee）在OCP當地區域社群之外，另推動GPU超節點與冷板（cold plate）液冷（liquid cooling）的專案[S-459; S-460]，可見共通介面能流通至硬體無法到達之處。

IF 若設於美國的晶圓代工與封裝達到量產規模，部分集中度將自2027年起緩解。 此一研判（reading）須以台積電（TSMC）第二座亞利桑那晶圓廠如預期於2027年下半年達到量產[S-083]、設於美國的先進封裝（advanced packaging）隨後到位，以及HBM驗證（qualification）在各供應商之間擴大為前提[S-128; S-003]。應觀察台積電就亞利桑那州所作的季度說明、TrendForce的CoWoS產能報告，以及台美備忘錄與貿易協定的立法審議[S-453]。若美國各廠區維持低產量，或2026年11月的到期使實體管制恢復，或中國重新實施其已承諾暫停的稀土出口管制（rare-earth export controls），因而同時衝擊相同的供應鏈，則此一研判不成立[S-414; S-415]。

> **重要性何在**
> **集中度是交付日期的風險。** 任一法律管轄區的晶圓代工、封裝或HBM供應若發生中斷，無論買方採用何種機櫃標準，部署都會延後；於開放介面上進行第二供應來源

（second source）的驗證（qualification）可縮短恢復時間，卻無法創造產能。

讀者視角

本章對政府、產業、財務與研發的意義

政府。 註明日期的決策包括2026年11月的出口管制到期、NERC預定於2026年12月31日完成的運算負載標準，以及800 VDC的列名認證；集中度所需的是投資，而非標準。

產業。 決定許多交付日期的是取得電力的時間，而非加速器（accelerator）供應；負載削減（curtailment）、儲能與故障穿越（ride-through）能力，如今可換得更好的併接條件。

財務。 NVIDIA 2027財年的展望假設來自中國的Data Center運算營收為零；電網缺口與待決的列名認證，可能使已採購設備的營收遞延。

研發。 直流故障保護、故障穿越（ride-through）控制與不含氟的兩相流體，是認證（certification）與化學品關卡背後的工程問題。

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-002 NVIDIA (26 Aug 2026); S-003 SK hynix (29 Jul 2026); S-010 NVIDIA (26 Aug 2026); S-012 GE Vernova (22 Jul 2026); S-025 ASML (15 Jul 2026); S-026 Siemens Energy (29 Jun 2026); S-050 NVIDIA (25 Feb 2026); S-051 NVIDIA (20 May 2026); S-056 Samsung Electronics (12 Feb 2026); S-083 TSMC (16 Apr 2026); S-084 TSMC (16 Jul 2026); S-097 NVIDIA (20 May 2026); S-102 NVIDIA (28 May 2025); S-103 NVIDIA (27 Aug 2025); S-111 AMD Investor Relations (4 Aug 2026); S-128 Micron Technology (24 Jun 2026); S-197 ABB (13 Oct 2025); S-198 ABB (未註明日期); S-231 Bloom Energy (13 Apr 2026); S-312 Huawei (18 Sep 2025); S-325 Rapidus (5 Jun 2026); S-446 Meta Platforms (9 Jan 2026); S-485 Advanced Micro Devices (5 Aug 2025); S-486 Advanced Micro Devices (3 Feb 2026); S-487 KLA (29 Oct 2025); S-498 Celestica (28 Jan 2026); S-502 Advanced Micro Devices (4 Nov 2025); S-503 NVIDIA (26 Aug 2026); S-504 Applied Materials (2 Oct 2025); S-506 ASML (15 Oct 2025); S-507 ASML (15 Apr 2026); S-508 Synopsys (29 May 2025); S-509 Cadence Design Systems (29 May 2025); S-511 Micron Technology (3 Oct 2025); S-512 (13 Mar 2026); S-513 NVIDIA (19 Nov 2025); S-514 KLA (6 Aug 2026); S-515 Marvell Technology (28 Aug 2026); S-516 Lam Research (7 Aug 2026); S-517 ASML (25 Feb 2026); S-587 KLA (28 Jul 2026); S-588 Advanced Micro Devices (5 May 2026); S-589 Synopsys (2 Jul 2025); S-590 Cadence Design Systems (2 Jul 2025)

標準組織與聯盟. S-019 OCP (11 Aug 2026); S-027 OCP (1 Mar 2026); S-054 OCP (22 Jun 2026); S-059 OCP (2 Sep 2026); S-205 OCP (未註明日期); S-216 OCP (未註明日期); S-448 UL Solutions (未註明日期); S-459 ODCC (未註明日期); S-460 OCP (未註明日期)

政府與監管機關. S-005 IEA (16 Apr 2026); S-032 ERCOT (18 Jun 2026); S-040 Texas Legislature (2025); S-041 Federal Energy Regulatory Commission (18 Dec 2025); S-042 PJM Inside Lines (13 Aug 2026); S-061 PJM Inside Lines (10 Sep 2026); S-068 Federal Energy Regulatory Commission (18 Jun 2026); S-163 U.S. Department of Commerce (16 Jul 2026); S-229 PJM Inside Lines (11 Aug 2026); S-230 NERC (Aug 2026); S-314 U.S. Department of Commerce (19 Nov 2025); S-315 Federal Register (BIS) (14 Jul 2026); S-316 PIB, Government of India (17 Feb 2026); S-323 European Commission (9 Apr 2025); S-409 Federal Register (BIS) (15 Jan 2025); S-410 Bureau of Industry and Security (13 May 2025); S-412 Federal Register (BIS) (15 Jan 2026); S-413 Federal Register (20 Jan 2026); S-414 Federal Register (BIS) (12 Nov 2025); S-415 The White House (1 Nov 2025); S-417 Federal Register (ITA) (10 Apr 2026); S-419 The White House (Jul 2025); S-421 Congressional Research Service (1 Sep 2026); S-422 eCFR (22 Sep 2026); S-423 eCFR (22 Sep 2026); S-424 eCFR (18 Sep 2026); S-425 Federal Register (BIS) (6 Mar 2023); S-426 Federal Register (BIS) (28 Mar 2025); S-427 Federal Register (11 Aug 2026); S-428 The White House (25 Sep 2026); S-429 The White House (23 Jul 2025); S-430 Federal Energy Regulatory Commission (22 Jul 2026); S-431 FERC (16 Jul 2026); S-434 CRU (Ireland) (12 Dec 2025); S-435 European Commission (未註明日期); S-437 European Commission (21 Sep 2026); S-438 IMDA (1 Dec 2025); S-439 NDRC (23 Jul 2024); S-440 METI (20 May 2026); S-443 EIA (9 Sep 2026); S-447 U.S. Department of Commerce (19 Mar 2026); S-449 European Commission (未註明日期); S-450 European Chemicals Agency (未註明日期); S-453 Executive Yuan (Taiwan) (15 Feb 2026); S-454 U.S. Department of Commerce (20 Jan 2026); S-455 Executive Yuan (Taiwan) (10 Sep 2026); S-456 European Commission (3 Jun 2026); S-457 European Commission (11 Sep 2026); S-458 IndiaAI (MeitY) (未註明日期); S-505 U.S. Department of Commerce (12 Nov 2025); S-510 U.S. Department of Commerce (2 Sep 2025)

研究機構. S-432 Nicholas Institute, Duke (11 Feb 2025); S-433 SSRN preprint (未註明日期); S-442 LBNL (Jun 2026); S-444 Grid Strategies (未註明日期); S-445 Grid Strategies (未註明日期)

市場研究. S-043 TrendForce (16 Sep 2026)

媒體與法律評論. S-162 TrendForce (14 Sep 2026); S-411 Global Trade & Sanctions Law (13 Mar 2026); S-416 South China Morning Post (27 Aug 2026); S-418 Simpson Thacher & Bartlett (16 Apr 2026); S-420 The AI Policy Network (31 Aug 2026); S-436 White & Case (20 Oct 2025); S-441 Future of Privacy Forum (18 Apr 2025); S-451 Freshfields (24 Sep 2026); S-452 Seoul Economic Daily (3 Sep 2026)

第8章

開放權重模型、記憶體與硬體

中國的開放權重模型（open-weight model）開發者如今公布了精簡記憶體設計工作的大部分成果，使每個token所需的快取在各次發布之間以大幅倍數下降；同時，代理、推理與百萬token上下文則使所保留的token數倍增。依2026年的紀錄，使用量的成長速度大致與每token記憶體用量的下降速度相當，因此需求是在各記憶體層級之間移動，而非減少。

適用讀者： 財務與投資、記憶體與系統供應商、研發與產品團隊、政府與政策；模型見第8.1節，記憶體算術關係見第8.3節，需求紀錄見第8.4節，中國的硬體路徑見第8.5節，至2035年的條件式路徑見第8.6節。

每一具AI加速器（accelerator）出貨時旁側皆配有記憶體，而在其上運行的模型則決定每位使用者占用其中多少。自2024年12月的DeepSeek-V3以來，中國的開發者已發布多款開放權重模型（open-weight model），其訓練所得參數任何人皆可下載並用於提供服務；其中數款重新設計了注意力機制，使每一個上下文token所需的記憶體僅為一年前的一小部分。同一段時間也帶來了代理、推理模式與百萬token上下文，使一個模型所保留的token數倍增。本章自開發者自身的模型卡與組態、自使用量與供應商資料，以及自中國的硬體路徑來研讀這兩股動向，並提出一個問題：更精簡的設計是減少了記憶體與AI硬體的需求，抑或只是使需求移動？

8.1 中國的開放權重推動精簡記憶體的設計

開放權重模型（open-weight model）係指其訓練所得參數公開供下載者，通常並附有模型卡、組態檔與一份授權（licence）條款；訓練資料與程式碼可能仍不公開，因此開放權重與開放原始碼並不相同。就基礎設施而言，重點在於任何人都能在自選的硬體上提供此類模型的服務，這使其記憶體設計成為整個市場的屬性，而非某一家服務供應商雲端的屬性。中國的五家開發者——DeepSeek、Alibaba的Qwen團隊、Moonshot AI、Z.ai（Zhipu）與MiniMax——發布了本章所研讀的大部分開放權重模型（表16）。

F 各家開發者如今皆發布一款參數量達數千億至數兆的旗艦模型，多數具備約一百萬token的上下文。 DeepSeek於2026年4月的V4預覽版，列出V4-Pro為1.6兆個參數、每個token啟用490億個，V4-Flash則為2,840億與130億個，兩者的上下文長度均為一百萬token[S-518]。Alibaba於2026年8月發布的Qwen3.8-2.4T-A95B具有2.4兆個參數，啟用950億個[S-519]；Moonshot AI於7月發布的Kimi K3則為2.8兆與1,040億個[S-489]。Z.ai的GLM-5.3具有7,440億個參數，作用中為400億個[S-520]；MiniMax-M3約為4,280億個，作用中約230億個[S-521]。各款模型就任一token所啟用的參數，均不足其總量的6%。

F 在數個系列的頂層，授權（licence）已從寬鬆條款轉為客製條款。 Alibaba於2026年2月以Apache 2.0發布Qwen3.5-397B-A17B，而Qwen3.8-2.4T-A95B與Qwen3.8-Flash-Next則採客製的Qwen授權[S-522; S-519; S-523]。Moonshot AI自Kimi K2所採的修訂版MIT授權，轉為其自有的Kimi K3 License；MiniMax則自MiniMax-M1的Apache 2.0，轉為MiniMax-M3的社群授權[S-524; S-489; S-525; S-521]。DeepSeek-V3.1的權重與Z.ai的GLM-5.3-Flash採MIT授權[S-526; S-527]。本報告僅指名這些授權，未審閱其條款；任何計畫以商業方式提供這些模型服務者，均須自行研讀。

F 在開發者平台上，中國的開放權重模型（open-weight model）占有龐大且持續上升的token份額。 一項針對經OpenRouter平台路由的逾100兆個token所作的研究發現，來自中國開發者（包含台灣與香港者）的開放原始碼模型，其每週份額自2024年底最低1.2%，上升至2025年若干週接近30%，全年平均約13.0%[S-536]。在2026年9月14日起的那一週，DeepSeek占OpenRouter文字請求的25.4%，為各開發者中最高者；Z.ai的GLM 5.3 Flash在過去三十日間處理的token數最多[S-537]。在Vercel的AI Gateway上，各來源的開放權重模型於2026年8月運行了56%的token，而2025年12月時尚不及十分之一[S-538]。

F 企業調查與下載數呈現出不同的面貌。 Menlo Ventures於2025年11月針對495位美國企業決策者所作的調查發現，開放原始碼在企業市場的份額一年內自19%降至11%，而中國的開放原始碼模型則約占大型語言模型（LLM）API總使用量的1%[S-539]。Hugging Face統計建構於Qwen之上的衍生模型達151,448款，為Meta足跡的2.6倍，同時指出下載數並未涵蓋API使用量或私有部署[S-540]。

= 在Vercel的閘道上，以定價計算，2026年8月一個開放權重token的成本約為封閉權重token的八分之一。 計算公式：開放權重模型的每token價格 ÷ 封閉權重模型的每token價格 =（開放權重的支出份額 ÷ 開放權重的token份額）÷（封閉權重的支出份額 ÷ 封閉權重的token份額）。開放權重模型占token的56%、占支出的14%，因此封閉權重模型分別為44%與86%：（14 ÷ 56）÷（86 ÷ 44）= 0.25 ÷ 1.95 ≈ 0.13[S-538]。Vercel係依定價估算支出，因此此一比值比較的是單一閘道上的定價，而非使用者實付金額或提供服務的成本。它說明了token量為何比支出增長得更快。

A 這些份額衡量的是不同的對象，合併起來僅支持一項範圍狹窄的研判（reading）。 彙整平台與閘道的份額所計算的，是開發者在眾多模型之間路由的token；調查所計算的是企業自行申報的內容；下載數所計算的是檔案。開放權重在彙整平台、閘道與下載等指標上均有所成長，但在Menlo的調查中份額下降；在最近一週的紀錄中，由中國開發者DeepSeek領先OpenRouter的請求量，而在美國企業之間，中國模型仍約占API使用量的1%[S-536; S-537; S-538; S-539]。就基礎設施而言，重點不在於哪一項份額才正確：這些模型由眾多業者提供服務，包括Amazon Bedrock、Microsoft Foundry與Google的Vertex AI[S-542; S-543; S-544]，因此其記憶體設計所形塑的需求，遠超出開發者自身的雲端。

8.2 五項設計選擇降低每個token的記憶體用量

一個模型在為使用者提供服務時所需的記憶體分為兩部分：所有使用者共用的權重，以及保有每段對話上下文、並隨每一個token而增長的鍵值（key-value, KV）快取（第3.3節）。表16中的開發者對兩者同時下手，所採用的五項選擇在各系列之間反覆出現。其中沒有任何一項為中國獨有；突出之處在於中國開發者結合這些選擇的速度，以及將成果以開放權重形式發布，並附上足以讓任何人驗算的組態。

F 混合專家（mixture of experts, MoE）設計使大部分參數在任一token上保持閒置。 DeepSeek-V3為每一個token啟用其6,710億個參數中的370億個[S-528]；其後的各系列採用同一原則，直至DeepSeek-V4-Pro在1.6兆中啟用490億個、Kimi K3在2.8兆中啟用1,040億個[S-518; S-489]。MiniMax將MiniMax-M2的100億個啟用參數定位為服務代理的一種途徑，並宣稱可帶來「Simpler capacity planning with smaller per-request memory and steadier tail latency」[S-533]。Alibaba的Qwen3.8-Flash-Next自一個1,250億參數的主模型中，為每個token啟用60億個參數[S-530]。

F 壓縮式注意力快取縮減了每個token所儲存的上下文。 DeepSeek-V3在每一層儲存512維的潛在向量加上64個位置維度，而非128個頭的完整鍵與值；Kimi K2與GLM-5.3的組態採用相同的維度[S-528; S-524; S-520]。DeepSeek的V4系列另加壓縮稀疏注意力（sparse attention）：在一百萬token的上下文下，V4-Pro所需的單token推論運算量為DeepSeek-V3.2的27%，KV快取為其10%[S-518]。2026年9月10日發布的DeepSeek-V4.1-Flash，以FP4快取方式為每個token儲存890位元組的全域KV快取，約為V4-Flash的四分之一；DeepSeek並表示其快取所需的HBM為前一代的四分之一，SSD儲存為前一代的八分之一[S-529; S-488]。

F 混合式設計僅在其四分之一或更少的層中保有持續增長的快取。 Qwen3-Next與Qwen3.5以三層狀態不隨上下文增長的Gated DeltaNet，與一層全注意力層交替排列；Qwen3.8-2.4T-A95B則在92層的範圍內維持此一配置[S-545; S-522; S-519]。Kimi Linear以三比一的比例將Kimi Delta Attention與潛在注意力配對，並報告KV快取最多可減少75%；Kimi K3則有69層Kimi Delta Attention與24層閘控潛在注意力層[S-546; S-489]。Z.ai表示GLM-5.3-Flash結合稀疏注意力（sparse attention）與線性注意力，使注意力運算量與KV快取相較GLM-5.3分別減少3.01倍與4.44倍[S-532]。MiniMax早在2025年6月即發布MiniMax-M1，這是一款建構於lightning attention之上的開放權重混合注意力模型[S-525]。

F 4位元權重正成為預設做法，且已有一家開發者以4位元進行快取。 DeepSeek-V3.1以UE8M0 FP8比例格式進行訓練，「to ensure compatibility with microscaling data formats」；DeepSeek-V4則以FP4儲存其專家參數[S-526; S-518]。Kimi K2 Thinking透過量化（quantisation）感知訓練，以原生INT4形式出貨，Moonshot AI表示此舉可帶來約兩倍的生成速度；Kimi K3則採MXFP4權重搭配MXFP8激活值[S-547; S-489]。OpenAI的gpt-oss-120b對其專家權重採用MXFP4，使其得以在單一具80 GB的GPU上運行[S-534]。DeepSeek-V4.1-Flash以FP4進行快取[S-529]。在4位元下，一個參數占半個位元組，為16位元下大小的四分之一，且尚未計入比例因子。

F 快取，以及部分模型資料表，正日益被重複利用或移出加速器（accelerator）。 Moonshot AI表示，其Mooncake解耦式服務使官方Kimi API在程式撰寫工作負載下的快取命中率高於90%[S-548]。Alibaba的Qwen3.8-Flash-Next另加510億個參數的N-gram嵌入，並表示「the embedding table can be offloaded to host memory and overlapped with model computation through asynchronous prefetching」[S-530]。DeepSeek-V4.1-Flash不將滑動視窗的快取狀態持久化至SSD，使其持久化快取縮減至V4-Flash的約八分之一[S-529]。TrendForce報導，在中國「AI architectures led by DeepSeek are promoting the use of high-capacity QLC enterprise SSDs for KV cache offloading」[S-549]。

A 這五項選擇降低的是每個token所觸及的記憶體，而非一個模型所占用的記憶體。 混合專家（mixture of experts, MoE）減少的是每個token所讀取的權重，

表16。 開放權重旗艦模型結合稀疏專家、精簡注意力與低位元格式

各系列均已轉向具備長上下文（long context）的超大型稀疏模型，而各自以不同途徑降低單一token所需的記憶體。

模型（開發者）	已發布	參數，總量／作用中	上下文	注意力與快取設計	格式	授權	來源
DeepSeek-V3（DeepSeek）	2024年12月	6,710億／370億	128K	多頭潛在注意力（multi-head latent attention, MLA），61層	FP8權重	DeepSeek Model License	[S-528]
DeepSeek-V4-Pro與V4-Flash（DeepSeek）	2026年4月	1.6兆／490億，以及2,840億／130億	1M	壓縮稀疏注意力（sparse attention）；在100萬token下，V4-Pro所需的KV快取為V3.2的10%	FP4專家，其餘為FP8	—	[S-518]
DeepSeek-V4.1-Flash（DeepSeek）	2026年9月	主幹5,520億；預填充（prefill）時作用中80億，解碼（decode）時160億	1M	因果編碼器－解碼器；每個token的全域KV快取為890位元組	FP4的KV快取與專家	—	[S-529; S-488]
Qwen3-235B-A22B-Instruct-2507（Alibaba Qwen）	2025年7月	2,350億／220億	原生262K，擴充1M	分組查詢注意力（grouped-query attention, GQA），4個KV頭，94層	BF16	Apache 2.0	[S-490]
Qwen3.5-397B-A17B（Alibaba Qwen）	2026年2月	3,970億／170億	原生262K，擴充1M	每三層Gated DeltaNet搭配一層閘控注意力，共60層	BF16	Apache 2.0	[S-522]
Qwen3.8-2.4T-A95B（Alibaba Qwen）	2026年8月	2.4兆／950億	原生262K，擴充1M	同樣的3:1混合配置，共92層	BF16與FP8版本	客製（qwen3.8-max）	[S-519]
Qwen3.8-Flash-Next（Alibaba Qwen）	2026年8月	1,250億，另加510億N-gram嵌入／60億	原生262K，擴充1M	Gated DeltaNet搭配Qwen Sparse Attention；嵌入表可卸載至主機記憶體	—	客製（qwen-community-1.0）	[S-530; S-523]
Kimi K2-Instruct-0905（Moonshot AI）	2025年9月	1兆／320億	256K	MLA，61層	區塊式FP8	修訂版MIT	[S-524]
Kimi K3（Moonshot AI）	2026年7月	2.8兆／1,040億	1M	69層Kimi Delta Attention與24層閘控MLA	MXFP4權重，MXFP8激活值	Kimi K3 License	[S-489]
GLM-4.5（Z.ai）	2025年7月	3,550億／320億	128K	GQA，8個KV頭，92層	BF16與FP8版本	MIT	[S-531]
GLM-5.3與GLM-5.3-Flash（Z.ai）	2026年8月	7,440億／400億，以及3,200億／180億	1M	潛在注意力搭配DeepSeek Sparse Attention；Flash另加線性注意力，KV快取較GLM-5.3少4.44倍	FP8與BF16版本	Flash：MIT	[S-520; S-532; S-527]
MiniMax-M2（MiniMax）	2025年10月	2,300億／100億	196K個位置	全注意力，GQA，8個KV頭，62層	FP8	—	[S-533]
MiniMax-M3（MiniMax）	2026年6月	約4,280億／230億	1M	在GQA之上的MiniMax Sparse Attention	—	客製（minimax-community）	[S-521]
OpenAI的gpt-oss-120b	2025年8月	1,170億／51億	131K個位置	128-token滑動視窗與全注意力交替	MXFP4專家	Apache 2.0	[S-534]
Llama 3.3 70B（Meta）	2024年12月	700億，稠密型	128K	GQA，8個KV頭，80層	—	Llama 3.3 Community License	[S-535]

依開發者與發布月份選列的開放權重模型（open-weight model），內容為各開發者的模型卡、組態或發布說明所載，資料截至2026年9月27日，並納入OpenAI與Meta作為對照。參數分為總量與每個token的啟用量。

範圍 數字均為開發者自有，本報告未予實測。上下文為其所載的原生或擴充長度。授權（licence）僅予指名，未經審閱。

破折號標示所引段落未提供的細節。發布月份為各開發者所提供者。

而非所儲存的權重：每一位專家都必須保持在為該模型提供服務的加速器（accelerator）可及範圍之內。壓縮式、稀疏與混合式注意力降低每個token的快取量，低位元格式則同時降低兩者。其結果是記憶體約束所在之處的移轉：每段對話所需較少，而在一個垂直擴展範圍內保有超大型模型所需者較多，保存可重複利用快取的主機記憶體與儲存所需者亦較多[S-518; S-489; S-529]。Moonshot AI建議將Kimi K3部署於64具或更多加速器構成的超節點上，此與第二項效應一致，惟Moonshot係以通訊頻寬作為其依據[S-548]。

8.3 這些組態對記憶體的意涵

一個模型的組態檔決定了每一個上下文token所占用的記憶體量。每個token的KV快取，等於保有持續增長快取的層數，乘以每一該類層針對一個token所儲存的數值個數，再乘以每個數值的位元組數（第3.3節）。將此一算術關係套用於已公布的組態，可顯示第8.2節所述的設計將這些數字推移了多遠，以及哪些部分未被改變：權重、讀取權重所需的頻寬，以及在各輪對話之間保存快取的記憶體。

= 在128K token下，一段對話在2024年的稠密型設計上約需43 GB快取，而在Kimi K3上則不足4 GB。 分組查詢注意力（grouped-query attention, GQA）的計算公式：2 × 層數 × KV頭數 × 頭維度 × 位元組數；潛在注意力則為：層數 ×（512 + 64）× 位元組數；均以16位元與131,072個token計。Llama 3.3 70B（頭維度8,192 ÷ 64 = 128）：2 × 80 × 8 × 128 × 2 = 每個token 327,680位元組，即42.9 GB[S-535]。Qwen3-235B-A22B：2 × 94 × 4 × 128 × 2 = 192,512位元組，即25.2 GB[S-490]。DeepSeek-V3：61 × 576 × 2 = 70,272位元組，即9.2 GB[S-528]。Kimi K3僅在其24層潛在注意力層中進行快取：24 × 576 × 2 = 27,648位元組，即3.6 GB[S-489]。DeepSeek載明V4.1-Flash在FP4下為890位元組，即0.12 GB[S-529]。

= 在約一百萬token處，每段對話的差距自接近200 GB至不足1 GB。 同一組公式得出：Qwen3-235B-A22B在其1,010,000個token的擴充上下文下為194.4 GB，Qwen3.8-2.4T-A95B在同一長度下為95.2 GB，Kimi K3在1,048,576個token下為29.0 GB，而依DeepSeek就V4.1-Flash所提的FP4數字則為0.93 GB[S-490; S-519; S-489; S-529]。Qwen自身的指引與第一項相符：在Qwen3-235B-A22B上處理一百萬token的上下文需要「approximately 1000 GB of total GPU memory」，其中涵蓋權重、快取與激活值[S-490]，而僅其16位元權重即約占470 GB（2,350億 × 2位元組）。在Qwen3-235B-A22B上，一段此類對話的快取將占滿一具Rubin GPU 288 GB HBM4的約三分之二（194.4 ÷ 288）；在Kimi K3上則約為10%（29.0 ÷ 288）[S-550]。

= 如今決定最大型開放模型最小占用量的是權重，而非快取。 在4位元下，一個參數在計入比例因子前占半個位元組，因此Kimi K3的2.8兆個參數至少占1.4 TB（2.8兆 × 0.5位元組），僅權重一項即至少需要五具288 GB的加速器（accelerator）（1.4 TB ÷ 288 GB = 4.9）[S-489]。Qwen3.8-2.4T-A95B的8位元版本約占2.4 TB，至少需要九具此類加速器（2.4 TB ÷ 288 GB = 8.3）[S-519]。在加入任何使用者的快取之前，僅權重即至少需要五至九具最大型的加速器；Moonshot AI建議為Kimi K3採用64具或更多加速器構成的超節點，理由是更大的高頻寬通訊範圍可提升推論效率[S-548]。一座Vera Rubin NVL72機櫃在其各GPU上合計搭載20.7 TB的HBM4[S-551]。

A 更精簡的快取，使具約束力的限制自容量轉向頻寬與模型駐留。 每一個解碼（decode）步驟就整個批次讀取作用中的權重一次，並讀取批次中每一段對話的快取；因此當快取縮小時，決定一具加速器（accelerator）能服務多少使用者的便是頻寬，而非容量（第3.3節）。NVIDIA將解碼階段描述為「fundamentally memory subsystem bound」，並在其2026年7月的技術部落格中列出每具Rubin GPU最高22 TB/s的HBM4頻寬，而其9月的產品頁面則列為19.2 TB/s[S-550; S-551]。一項使快取縮減十倍的設計，可讓最多十倍數量的對話共用扣除權重後所剩的記憶體，但每一個步驟仍須讀取作用中的權重與所有駐留的快取：每位使用者的容量下降，而每具加速器的頻寬需求並未下降。

F 自H100至Blackwell Ultra，每具加速器（accelerator）的記憶體隨每一代平台上升，並在Rubin上維持288 GB；其下一步則不確定。 NVIDIA每具GPU的HBM自H100的80 GB，提高至Blackwell的192 GB與Blackwell Ultra的288 GB，NVIDIA將後者描述為「3.6x more on-package memory than H100 and 50% more than Blackwell」[S-174]；Rubin維持288 GB，但改為HBM4，而一座Vera Rubin NVL72機櫃另加最高54 TB的LPDDR5X主機記憶體[S-551]。TrendForce預期NVIDIA的Rubin Ultra平台將於2027年「to further increase HBM capacity per GPU to 384GB」[S-552]，但同時報導NVIDIA自2026年第3季起一直在評估Rubin Ultra較低容量的HBM組態，並因LPDDR5X的供應而將其下一代Vera Rubin模組的SOCAMM容量減半[S-135]。

? 未解問題——已有多少實際提供服務的流量是在最精簡的設計上運行。 開發者公布其模型的需求，平台公布依開發者劃分的份額，但本報告所研讀的來源中，並無任何一項提供實際提供服務的token在注意力設計、快取格式或上下文長度上的組成[S-537; S-538]。缺少此一資料，表17中每個token的節省量便無法換算為記憶體總需求的變動。

8.4 使用量的成長速度大致與每token記憶體用量的下降速度相當

每token記憶體用量降低，會降低單一token的成本，而成本降低又會招致更多的token。記憶體總需求究竟隨之下降或上升，取決於何者移動得更快：每個token

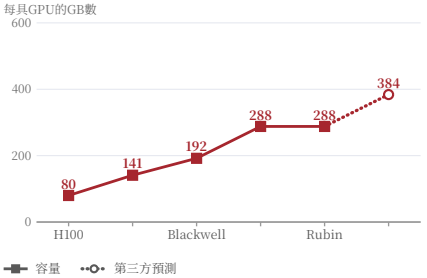
表17。 每段對話的快取用量差距逾100倍

在相同的上下文長度下，各開放權重設計之間每段對話的快取記憶體用量差距超過兩個數量級。

模型	保有持續增長快取的層數	每一該類層針對每個token所儲存的內容	每token位元組數	128K token下的GB數	1M token下的GB數	來源
Llama 3.3 70B（Meta，2024年）	80層中的80層（GQA）	鍵與值，8個頭 × 128	327,680	42.9	—	[S-535]
GLM-4.5（Z.ai，2025年）	92層中的92層（GQA）	鍵與值，8個頭 × 128	376,832（在FP8下為188,416，即Z.ai所載的設定）	49.4	—	[S-531]
MiniMax-M2（2025年）	62層中的62層（GQA）	鍵與值，8個頭 × 128	253,952	33.3	—	[S-533]
Qwen3-235B-A22B（Alibaba，2025年）	94層中的94層（GQA）	鍵與值，4個頭 × 128	192,512	25.2	194.4	[S-490]
DeepSeek-V3；Kimi K2（2024–25年）	61層中的61層（MLA）	潛在向量，512 + 64	70,272	9.2	—	[S-528; S-524]
gpt-oss-120b（OpenAI，2025年）	36層中的18層（其餘層保有128個token）	鍵與值，8個頭 × 64	36,864	4.8	—	[S-534]
Qwen3.8-2.4T-A95B（Alibaba，2026年）	92層中的23層（混合式）	鍵與值，4個頭 × 256	94,208	12.3	95.2	[S-519]
Qwen3.5-397B-A17B（Alibaba，2026年）	60層中的15層（混合式）	鍵與值，2個頭 × 256	30,720	4.0	31.0	[S-522]
Kimi K3（Moonshot AI，2026年）	93層中的24層（混合式MLA）	潛在向量，512 + 64	27,648	3.6	29.0	[S-489]
DeepSeek-V4.1-Flash（2026年）	開發者數字	以FP4壓縮的全域快取	890	0.12	0.93	[S-529]

每個token與每段對話的KV快取記憶體用量，由本報告依各開發者的組態計算，除另有載明者外，均以每個儲存數值16位元計；DeepSeek-V4.1-Flash為DeepSeek自有數字，採FP4。
範圍 本表為計算值，而非任何部署的實測值。服務系統可能以8位元或4位元儲存快取、對其進行分頁，或將其移至主機記憶體。權重、激活值、滑動視窗緩衝區，以及線性注意力層的固定大小狀態均不計入。GB為10^9位元組；128K = 131,072個token，1M = 1,048,576個token，Qwen系列模型則為其所載的最大值1,010,000個token；破折號表示所載的上下文長度更短。

圖29。 自H100以來，每具GPU的HBM已成長3.6倍



依NVIDIA世代劃分的每具GPU HBM容量，單位GB（自Rubin起為HBM4）：來自產品規格（specification）[S-174; S-551]，以及TrendForce對Rubin Ultra於2027年的預期[S-552]。
範圍 本圖為每具GPU的容量，而非出貨量或頻寬。NVIDIA尚未公布Rubin Ultra的記憶體組態，TrendForce則報導有較低容量的組態正在評估之中[S-135]。

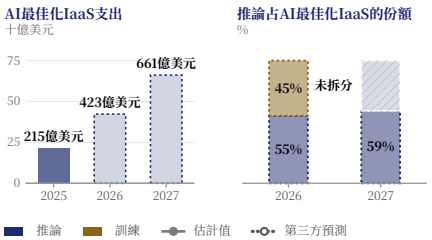
的節省量，抑或token數的成長。截至2026年9月的紀錄使兩者皆可被研讀，惟兩者單位不同；該紀錄並顯示記憶體製造商——其陳述是記憶體本身需求的唯一直接證據——如何描述其所見。

F 所報告的token量在兩年內成長數倍。 Google於2026年5月表示，其在各項服務介面上每月處理逾3,200兆個token，為一年前約480兆的七倍，而2024年5月則為9.7兆[S-491]；其模型API在2026年第2季每分鐘處理約220億個token，高於一季前的160億[S-553]。中國國家數據局（National Data Administration）負責人報告，全國每日平均token調用量自2024年初的1,000億，上升至2025年底的100兆，並於2026年3月逾140兆[S-554]。據Cailian Press報導，ByteDance的Volcano Engine表示其Doubao模型於2026年6月達到每日180兆個token，此為該公司依不同於全國數字之基礎所作的計數[S-555]。

F 在推理與代理的帶動下，每一次請求也變得更長。 OpenRouter與Andreessen Horowitz報告指出，推理最佳化模型於2025年占OpenRouter token量的半數以上，而每次請求的平均提示token數自約1,500上升至逾6,000，完成輸出則自約150上升至400[S-557]。Anthropic報告指出，其代理所用的token約為聊天互動的四倍，多代理系統則約為十五倍[S-558]。DeepSeek-R1-0528就每一道AIME題目平均使用23,000個token，而前一版本為12,000個[S-559]。NVIDIA的Vera Rubin產品頁面載明，代理式系統消耗的token量達「up to 15x more tokens than traditional AI applications」[S-551]。

F AI雲端容量的支出正向推論移轉。 在AI最佳化基礎設施即服務（IaaS）之內——Gartner預測其將自2025年的215億美元，成長至2026年的423億美元與2027年的661億美元——2026年「global spending on inference ($23.3 billion) will surpass that of training ($19 billion)」；推論於2026年占總額的55%，2027年占59%[S-614]。國際能源總署（IEA）報告指出，每項AI任務的能源用量一直在下降，「by at least an order of magnitude annually in recent years」，而影片生成、推理與代理式任務則「can consume hundreds or thousands of times more energy per query than simple text generation」[S-005]（圖30）。

圖30。 2026年推論在AI雲端支出上超越訓練



左：全球AI最佳化基礎設施即服務（IaaS）支出，單位十億美元，Gartner，2026年8月：2025年為估計值，2026年與2027年為預測值[S-614]。右：推論占該支出的份額，單位%，2026年與2027年為預測值；2026年其餘部分為訓練（190億美元），Gartner未就2027年的其餘部分進行拆分[S-614]。
範圍 本圖為Gartner的單一類別，即針對AI最佳化的雲端容量；並非全部AI支出，亦不含企業為其自有資料中心所採購的AI伺服器。

A 推論使記憶體從一項建置需求轉為一項服務需求。 訓練隨每一代模型出現高峰；推論則持續運行，權重常駐，且每一段作用中的對話都有一份快取，因此推論占比上升，會使記憶體需求繫於使用者、上下文與延遲，而非繫於模型發布[S-614]。此點將第8.3節的快取算術關係與第3.2節的預算證據相連：每一個百分點的份額移向推論，便增添一份為使用者所保留的記憶體，而更精簡的快取僅能部分抵銷之。

F 另有一些發布則使每項任務的token數朝相反方向移動。 OpenAI表示，具思考模式的GPT-5在各項能力上以「50-80% less output tokens」的輸出表現優於o3[S-560]；DeepSeek表示V3.1的思考模式在輸出token數減少20%至50%的情況下，平均表現與DeepSeek-R1-0528相當[S-561]。Alibaba表示其Agent Context可將token使用量最多減少67%[S-562]。2025年1月，Microsoft執行長表示推論的性價比已改善「more than 10X for every model generation due to software optimizations」，並就採購提出告誡，不宜「too much of anything at one time」[S-563]。

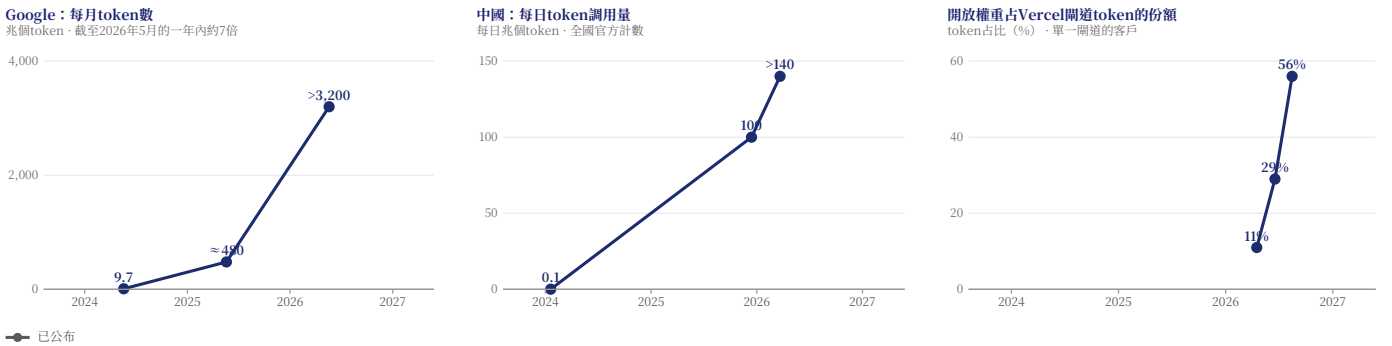
= 依2025–26年的步調，token成長與快取節省的量級相近。 計算公式：為上下文所保留的快取，其變動幅度 = token數的倍數 × 每個token快取量的倍數。截至2026年5月的一年內，Google每月的token數上升約七倍[S-491]；中國的每日調用量在2025年底至2026年3月之間上升逾40%，換算為年率至少為3.8倍（1.4^4）[S-554]。DeepSeek的各次發布使每個token的快取量降至十分之一（2026年4月，V4-Pro對比V3.2，於1M token下），其後再降至四分之一（9月，V4.1-Flash對比V4-Flash）[S-518; S-529]。token數上升七倍，對比每個token下降四至十倍，得出快取記憶體為7 ÷ 4 = 1.75倍至7 ÷ 10 = 0.7倍；且此結果僅在精簡設計承載全部流量時才成立，而目前尚非如此。

F 各服務供應商對重複利用的上下文定價，僅為新上下文的一小部分。 DeepSeek就DeepSeek-V4.1-Flash列出的價格為：在尖峰時段，當上下文對其快取而言屬全新時，每百萬個輸入token為0.30美元，而命中快取時為0.006美元，輸出則為1.20美元[S-564]。DeepSeek表示，快取命中的費用「often account for a large share of agent costs」，壓縮快取可降低這些費用，並就「2,000 GPUs + a storage cluster」規模的部署徵詢洽談[S-488]。Amazon將Kimi K3描述為Bedrock上首款支援明確提示快取的開放權重模型（open-weight model），可在跨模型調用重複利用上下文時降低延遲與輸入成本[S-542]。

= 依DeepSeek的定價，快取命中的費用為快取未命中的2%。 計算公式：快取命中價格 ÷ 快取未命中價格。以DeepSeek-V4.1-Flash為例，尖峰時段為0.006美元 ÷ 0.30美元 = 2%，離峰時段為0.003美元 ÷ 0.15美元 = 2%[S-564]。價格並非成本，但此一落差給了使用者保有並重複利用長上下文（long context）的理由，也給了服務供應商儲存快取而非重新計算的理由。因此，每個token在HBM上所節省的部分，會以對保存可重複利用快取的主機記憶體與快閃記憶體的需求形式回流；而SK hynix即於2026年8月在池化的CXL記憶體與SSD上展示了此類快取[S-565]。

F 記憶體製造商如今明確將代理與長上下文（long context）列為需求來源。 SK hynix表示，隨著AI演進為代理式形態，「the underlying demand base for memory is broadening」[S-003]。Samsung預期2026年下半年的需求將以伺服器為重心，「stemming from continued AI infrastructure capex and broader adoption of agentic AI」

圖31。 token量在兩年內成長數倍



Google：2024年、2025年與2026年5月在其各項服務介面上每月處理的token數，依其I/O 2026主題演講所述[S-491]。中國：全國每日平均token調用量，國家數據局（National Data Administration）[S-554]。Vercel AI Gateway：開放權重模型（open-weight model）占其客戶所路由token的份額[S-556; S-538]。

範圍 三項指標各依其自身的座標軸與定義：Google自有產品與API、一項全國官方計數，以及單一閘道的客戶。三者彼此不可比，絕不相加，亦不衡量開放權重模型或中國模型的市場份額。

[S-127]。Micron表示代理式AI使資料中心擴及CPU機櫃與「storage racks for rapidly expanding context store」，而其當季資料中心SSD營收逾50億美元[S-129]。Kioxia預期來自推論AI應用的NAND需求，至2028年前將以估計每年86%的速度成長[S-566]。

F 市場研究預期HBM供應在2027年仍將不足以滿足需求。 TrendForce預測HBM位元出貨量於2027年成長50–60%，「which will still be insufficient to keep pace with demand growth」，並預期供應商將在2027年全年保有定價能力[S-135]。該機構估計HBM合約價格「could still rise by 70–140% in 2027」[S-007]；2026年9月，該機構表示在供應緊俏與HBM4組合比重提高之下，已大幅上調其2027年HBM價格展望[S-567]。Micron預測HBM市場將自2025年的約350億美元，成長至2028年的約1,000億美元[S-568]。

A 依目前的紀錄，效率使記憶體需求移動，而非減少需求。 在2025–26年降低每個token快取量的每一項設計，所投入的市場都是token量倍增數倍、請求長度拉長、記憶體製造商陳述需求高於供給的市場[S-518; S-491; S-003]。這些節省改變的是記憶體的需求所在：每段對話所需的HBM較少，每具加速器（accelerator）所需的HBM頻寬較多，兆級參數模型所占用的HBM較多，而可重複利用快取所需的主機記憶體與快閃記憶體亦較多[S-529; S-129; S-548]。另一種研判（reading）——即一旦最精簡的設計承載大部分流量，效率便將超越使用量——並未被排除；惟本報告所引的記憶體製造商與市場研究機構，迄今均未如此描述。

8.5 中國的硬體與其模型相互形塑

部分中國開發者表示，其設計時已將國內硬體納入考量。自2024年12月起，美國的出口管制（export control）已涵蓋高頻寬記憶體（high-bandwidth memory, HBM）本身，而記憶體頻寬亦成為決定哪些加速器（accelerator）可出口至中國的規則中的一項參數。中國正在建立國內的加速器與記憶體，其開發者並發布對兩者需求均較少的模型；就紀錄而言，無法區分這是對管制的回應，抑或對成本的回應。本節自各項規則、自各公司的藍圖與財報，以及自各模型自身的文件來研讀此一循環。

F 美國的規則如今管制HBM，並將記憶體頻寬寫入許可門檻。 2024年12月2日生效的一項工業及安全局（Bureau of Industry and Security, BIS）規則，管制記憶體頻寬密度高於每平方毫米每秒2 GB的HBM，並指出「All HBM stacks currently in production exceed this threshold」[S-569]。一項許可例外（licence exception）規定，每平方毫米低於3.3 GB/s的HBM，僅在由共同封裝晶片的設計者直接採購並運往封裝廠區時，方得輸往中國[S-570]。2026年1月關於逐案許可審查的規則，設定DRAM總頻寬6,500 GB/s的上限，該項指標的定義載明「Copackaged DRAM ICs include, for example, high bandwidth memory (HBM)」[S-412]；同月的第232條關稅，則適用於DRAM頻寬落在特定級距內的晶片[S-571]。

F Huawei正自行開發HBM，並已將其加速器（accelerator）藍圖提前。 2025年9月，Huawei為其Ascend 950晶片公布兩款自有HBM類型：HiBL 1.0，即一款「low-cost HBM that is more cost-effective than HBM3E and HBM4E」；以及HiZQ 2.0，「delivering 144 GB of memory and 4 TB/s memory access bandwidth」[S-312]。在2026年7月的世界人工智慧大會（World Artificial Intelligence Conference）上，該公司展示了一套由1,024張卡構成、具256 TB統一記憶體定址的實體Atlas 950 SuperPoD，而2025年所規劃者為最多8,192顆晶片[S-572; S-312]。2026年9月，該公司將Ascend 960DT提前至2027年第1季，並描述了一套最多4,096顆NPU、具最高1 PB HBM的Atlas 960E SuperPoD[S-573]。

F 中國的記憶體製造商與雲端服務供應商，陳述了與市場其餘部分相同的價格壓力與高額支出。 CXMT公布2026年上半年營收為人民幣1,503億元，成長873.64%，並將其歸因於全球DRAM供應短缺且價格急升，同時表示來自AI相關領域的營收占比仍相對較低[S-574]。Alibaba 6月當季的資本支出（capital expenditure）成長75%至人民幣677億元，部分原因為「higher pricing of a broad range of chip components」[S-575]；Tencent則成長176%至人民幣528億元[S-576]。Alibaba執行長表示，AI資料中心供應鏈各環節的全球性短缺，限制了該公司擴充運算規模的速度[S-577]。

F DeepSeek的數值格式與部署說明指向國內硬體。 DeepSeek-V3.1以UE8M0 FP8比例格式進行訓練[S-526]；據National Business Daily報導，DeepSeek官方帳號的一則置頂留言表示，該格式係為即將發布的次世代國產晶片而設計[S-578]。據Cailian Press報導，DeepSeek-V4於2026年4月推出時，Huawei表示其Ascend SuperPoD全系列產品均支援這些模型；DeepSeek則表示V4-Pro的服務吞吐量受高階運算所限，並預期一旦Ascend 950 SuperPoD於下半年大量出貨，即會有大幅降價[S-579]。

A 精簡記憶體的設計同時服務於成本與供應兩項限制，而就紀錄而言無法區分兩者。 一個僅需十分之一快取、並以4位元運行的模型，能最充分地運用稀缺的HBM，無論該稀缺源自出口規則或源自價格[S-518; S-569]。中國的方向也指向其得以在國內供應的層級：來自CXMT的DRAM、TrendForce所稱由DeepSeek主導的架構正推動用於快取卸載的QLC SSD，以及Huawei用於預填充（prefill）的較低成本HiBL HBM[S-574; S-549; S-312]。相同的技術也出現在美國的開放模型中，例如gpt-oss-120b的滑動視窗注意力與MXFP4權重[S-534]，因此僅憑設計選擇並不足以顯示這是對管制的回應。

F 各國政府將中國的開放權重模型（open-weight model）視為安全與產業政策問題。 CAISI於2025年9月認定DeepSeek的模型落後於美國模型，且較易受到代理挾持與越獄攻擊[S-580]；2026年5月則認定DeepSeek V4落後於前沿約八個月，但較同等能力的模型更具成本效率[S-541]。美國2026財年《國防授權法》（National Defense Authorization Act）要求國防部自其系統中移除DeepSeek的AI，並禁止承包商於履約過程中使用[S-581]。中國國務院指示支持AI開放原始碼社群，並設定至2027年新一代智慧終端、代理及類似應用的滲透率逾70%的目標[S-582]。

IF 若美國的管制持續將記憶體頻寬作為管制參數，中國的模型與記憶體硬體將與其他地區進一步分化。 此一研判（reading）須以2026年1月之後的下一項美國規則持續將頻寬與HBM作為管制參數，以及Huawei與CXMT為Ascend 950與960系統大規模供應HBM與DRAM為前提[S-412; S-573; S-574]。應觀察關係企業規則（Affiliates rule）暫停措施於2026年11月9日屆期、Huawei於2027年第1季交付Ascend 960DT，以及DeepSeek的下一款旗艦模型是否率先在Ascend上提供服務[S-505; S-573]。若輸往中國的HBM級產品許可範圍擴大，或中國的旗艦模型回歸高快取用量的設計，則此一研判不成立。

8.6 至2035年記憶體需求的條件式路徑

證據支持的是條件式觀點，而非預測。決定這些路徑的有三項變數：token量的成長速度、最精簡的設計在實際提供服務的流量中擴散的速度，以及其餘快取有多少自HBM移出至較便宜的層級。表18就每一預測期限（horizon）列出三種組合。各組合均不附機率；每一組合均指名可顯示其正在發生的訊號，並與第9章的各項展望相連。

= 各路徑的差異在於使用量與每個token快取量如何結合。 計算公式：為上下文所保留的快取記憶體，相對於2026年，等於token數的倍數 × 每個token快取量的倍數。若token數連續兩年每年上升七倍（49倍），而每個token的快取量每年降至四分之一（1/16），則所保留的快取約上升三倍（49 ÷ 16 ≈ 3.1）；若token數每年上升三倍（9倍），而快取每年降至十分之一（1/100），則降至約十分之一（9 ÷ 100）。上述為情境（scenario）假設（assumptions），框限Google、中國國家數據局（National Data Administration）與DeepSeek在2025–26年的紀錄[S-491; S-554; S-529; S-518]。權重與頻寬不在此一算術關係之內。

IF 至2027年，唯有token量持續倍增且HBM維持短缺，更精簡的模型才會使記憶體需求在各層級之間移轉。 此一研判（reading）須以token量持續每年成長數倍、HBM供應維持短缺為前提，一如TrendForce所預期的2027年HBM位元出貨量成長50–60%卻仍屬不足[S-135; S-491]。應觀察HBM與伺服器DRAM的合約價格、中國與北美用於快取卸載的企業級SSD訂單，以及DeepSeek、Qwen與Kimi下一批發布的快取設計

表18。 三條路徑中僅有一條使記憶體需求下降

層級移轉路徑為本報告的基準研判（reading）：HBM需求隨頻寬與模型駐留持續上升，而快取需求則移向主機記憶體與快閃記憶體。

路徑	假設	至2027年	至2030年	至2035年	路標
使用量主導	token量持續每年成長數倍，一如Google在2025–26年的情形[S-491]；代理與推理持續拉長請求長度；精簡設計擴散緩慢	HBM持續採配額供應且合約價格上升，依TrendForce的估計於2027年上升70–140%[S-007]；用於快取的主機記憶體與SSD亦隨之成長	新增的HBM、DRAM與封裝產能一到位即被吸收；記憶體在AI資本支出中維持高占比	記憶體層級體系成為AI系統最大的成本區塊，而供應決定部署的步調	Google與中國國家數據局（National Data Administration）的token計數；HBM合約價格；記憶體占雲端資本支出的比重
層級移轉（基準研判，base reading）	token成長速度快於每個token快取量的下降速度；快取移向主機記憶體、池化記憶體與快閃記憶體[S-529; S-565]	HBM需求隨頻寬與模型駐留成長，而非隨每位使用者的快取成長；企業級SSD與主機記憶體需求隨快取分層而成長[S-549; S-551]	快取分層成為服務軟體與介面的標準做法；每具加速器（accelerator）的HBM成長速度慢於頻寬	價值集中於頻寬、封裝與記憶體控制器，更甚於集中在HBM容量	新平台中每具GPU的HBM；每座機櫃的SSD與主機記憶體；快取命中的定價；CXL與上下文儲存產品
效率主導	精簡設計在一兩年內承載大部分流量；隨代理採用趨於平穩，token成長趨緩	緩解最先出現在較舊HBM與伺服器DRAM的合約價格上	記憶體占AI資本支出的比重下降；2027–29年新增的產能足以滿足需求	在最先進HBM世代之外，記憶體回歸大宗商品循環	價格連續數季下跌；供應商庫存；長期協議（long-term agreement）縮短或重新議定

AI記憶體需求的三條條件式路徑，依預測期限（horizon）列出其所由產生的假設（assumptions），以及可顯示各路徑正在發生的訊號。為APEX情境（scenario），並以各儲存格所引來源為依據。
範圍 本表為情境（scenario），而非預測；不附任何機率。市場研究數字為其發布機構的預測。
依第8.4節的證據，本報告的基準研判（reading）為層級移轉路徑。

[S-549]。若HBM或伺服器DRAM的合約價格連續兩季下跌而token量仍在成長，意味效率超越使用量的時點早於此一研判所容許者，則此一研判不成立（展望7）。

IF **至2030年，其平衡取決於代理是否持續使每項任務的token數倍增。** 在層級移轉路徑上，HBM需求隨頻寬與模型駐留上升，而快取則移向較便宜的層級；記憶體製造商於2027–29年新增的產能，所面對的是一個基礎更廣、但每位使用者HBM用量較少的需求[S-003; S-129]。應觀察各平台所報告的代理使用量、廣泛提供服務之模型中每個token的快取位元組數，以及如Micron策略客戶協議之類的多年期協議是否以類似條件續約[S-129]。若代理使用量趨於平穩而效率持續累積，使需求落入效率主導路徑，則此一研判（reading）不成立（展望15）。

IF **至2035年，記憶體層級體系可能一如加速器（accelerator）已然如此，在兩個生態系之間出現差異。** 若中國的模型持續與國內HBM、DRAM與快閃記憶體共同演進，而美國規則持續將頻寬作為管制參數，則兩個生態系在記憶體設計上亦將如在晶片上一般出現差異[S-412; S-573]。國際能源總署（IEA）針對資料中心用電量的2035年情境，自AI採用較慢時的約700 TWh，至採用較快時的逾1,700 TWh，其高效率情境則約為970 TWh，可見在該預測期限（horizon）上，採用與效率的範圍之廣[S-583]。應觀察中國開發者是否持續公布完整組態，以及共用的記憶體標準是否使兩個生態系維持互通。若HBM級記憶體的貿易重新開放，則此一研判（reading）不成立（展望19）。

本章的證據目前指向同一個方向。中國的開放權重模型（open-weight model）已在一年之內將每一token所需的記憶體大幅降低數倍，且是在公開場域中達成，任何營運業者均可採用其設計。然而使用量成長同樣快速，記憶體製造商仍表示需求高於供給，而加速器（accelerator）記憶體自H100以來已成長3.6倍，惟自Blackwell Ultra至Rubin則維持持平。這些節省使需求轉向頻寬、極大的模型佔用量與更廉價的快取層級；它們尚未消除需求。

讀者視角

本章對金融、產業、研發與政府的意義

金融。 更精簡的模型降低的是每token成本，而非目前紀錄所顯示的記憶體需求；應將token用量、合約條件與記憶體價格合併研讀。

產業。 需求正從每位使用者的HBM容量轉向頻寬、模型駐留，以及主機DRAM、LPDDR與SSD中的快取層級；應使各層級均完成驗證（qualification），而非僅限HBM。

研發。 壓縮注意力、稀疏注意力（sparse attention）與線性注意力、四位元格式以及快取卸載，如今共同塑造記憶體需求；在各層級之間搬移快取的介面尚未標準化。

政府。 記憶體頻寬如今已是一項出口管制（export control）參數；中國的模型與記憶體供給共同演進，而美國國防相關規則限制使用DeepSeek的模型。

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-003 SK hynix (29 Jul 2026); S-127 Samsung Electronics (30 Jul 2026); S-129 Micron Technology (24 Jun 2026); S-174 NVIDIA Technical Blog (22 Aug 2025); S-312 Huawei (18 Sep 2025); S-488 DeepSeek (10 Sep 2026); S-489 Moonshot AI (Jul 2026); S-490 Alibaba Qwen team (21 Jul 2025); S-491 Google (19 May 2026); S-518 DeepSeek-AI (24 Apr 2026); S-519 Alibaba Qwen team (12 Aug 2026); S-520 Z.ai (18 Aug 2026); S-521 MiniMax (1 Jun 2026); S-522 Alibaba Qwen team (16 Feb 2026); S-523 Alibaba Qwen team (26 Aug 2026); S-524 Moonshot AI (Sep 2025); S-525 MiniMax (16 Jun 2025); S-526 DeepSeek-AI (21 Aug 2025); S-527 Z.ai (26 Aug 2026); S-528 DeepSeek-AI (26 Dec 2024); S-529 DeepSeek-AI (10 Sep 2026); S-530 Alibaba Qwen team (26 Aug 2026); S-531 Z.ai (28 Jul 2025); S-532 Z.ai (26 Aug 2026); S-533 MiniMax (27 Oct 2025); S-534 OpenAI (5 Aug 2025); S-535 Meta (6 Dec 2024); S-537 OpenRouter (27 Sep 2026); S-538 Vercel (17 Sep 2026); S-540 Hugging Face (14 Aug 2026); S-542 Amazon Web Services (18 Sep 2026); S-543 Microsoft (31 May 2026); S-544 Google Cloud (10 Feb 2026); S-545 Alibaba Qwen team (11 Sep 2025); S-546 Moonshot AI (30 Oct 2025); S-547 Moonshot AI (Nov 2025); S-548 Moonshot AI (Jul 2026); S-550 NVIDIA (21 Jul 2026); S-551 NVIDIA (21 Sep 2026); S-553 Google (22 Jul 2026); S-556 Vercel (13 Jul 2026); S-558 Anthropic (13 Jun 2025); S-559 DeepSeek-AI (28 May 2025); S-560 OpenAI (7 Aug 2025); S-561 DeepSeek (21 Aug 2025); S-562 Alibaba Group (24 Sep 2026); S-563 Microsoft (29 Jan 2025); S-564 DeepSeek (10 Sep 2026); S-565 SK hynix (7 Aug 2026); S-566 KIOXIA (2 Jun 2026); S-568 Micron Technology (17 Dec 2025); S-572 Huawei Technologies (17 Jul 2026); S-573 Huawei Technologies (17 Sep 2026); S-574 CXMT Corporation (29 Aug 2026); S-575 Alibaba Group (20 Aug 2026); S-576 Tencent (12 Aug 2026); S-577 Alibaba Group (22 Sep 2026)

政府與監管機關. S-005 IEA (16 Apr 2026); S-412 Federal Register (BIS) (15 Jan 2026); S-505 U.S. Department of Commerce (12 Nov 2025); S-541 National Institute of Standards and Technology (1 May 2026); S-554 National Data Administration of the People's Republic of China (23 Mar 2026); S-569 U.S. Department of Commerce (5 Dec 2024); S-570 U.S. Government (18 Sep 2026); S-571 The White House (14 Jan 2026); S-580 National Institute of Standards and Technology (30 Sep 2025); S-581 U.S. Congress (18 Dec 2025); S-582 State Council of the PRC (26 Aug 2025)

研究機構. S-536 Malika Aubakirova et al., arXiv (15 Jan 2026); S-557 OpenRouter (Dec 2025); S-583 IEA (10 Apr 2025)

市場研究. S-007 TrendForce (25 Aug 2026); S-135 TrendForce (4 Aug 2026); S-539 Menlo Ventures (9 Dec 2025); S-549 TrendForce (21 Sep 2026); S-552 TrendForce (2 Jun 2026); S-567 TrendForce (22 Sep 2026); S-614 Gartner (10 Aug 2026)

媒體與法律評論. S-555 via 中国金融信息网 (23 Jun 2026); S-578 National Business Daily (21 Aug 2025); S-579 Cailian Press (24 Apr 2026)

第9章

展望：2027年、2030年、2035年

依目前的財測指引，需求持續上升，但晶片、機櫃與電網各依不同時鐘擴張；本展望是一組條件式路徑，其轉折點均註明日期、可供觀察，且多數落在任何規格（specification）之外。

適用讀者： 金融、政府、研發與策略讀者；監測總覽表為表20，註明日期的路標則見圖32。

AI基礎設施的每一層級都仍在擴張，但並非依同一時鐘。加速器（accelerator）產能爬坡與記憶體價格以季為單位變動；機櫃、封裝產線與晶圓廠以年為單位；渦輪機、輸電與電網併接則歷時近十年。本章將前述各章的證據轉化為2027年、2030年與2035年的條件式展望。其中沒有任何一項是預測。每一項均載明其成立條件、應觀察的註明日期訊號，以及使其不成立的觀察結果。本章所回答的問題相當務實：哪些訊號、依何種順序，能顯示稀缺接下來將移往何處？

9.1 基準路徑與三條替代路徑

企業財測指引、產能計畫與能源預測在2027年之前均指向同一方向，但它們並未確定其後的情況。本節提出一條基準路徑，即與2026年9月27日紀錄最為一致的研判（reading），以及同一紀錄所支持的三條替代路徑，另有第四項架構面的變體。每一項均以條件式觀點撰寫：載明其成立條件、應觀察的事項，以及使其不成立的條件。任何一項均不附機率。

F 買方自身的財測指引指向2027年支出增加，而非減少。 Alphabet於2026年7月表示，其仍預期資本支出（capital expenditure, capex）將「to increase significantly in 2027」[S-016]。Microsoft預期其2027財年（自2026年7月至2027年6月）的資本支出將較前一年成長[S-075]。Amazon執行長表示，Amazon於2027年新增的產能已「largely reserved」，且Amazon仍有望於2027年底前使其2025年的電力容量倍增[S-070]。三家公司均尚未提出2027年的資本支出數字。

A 基準路徑是在限制範圍持續擴大之下維持成長。 依此一研判（reading），需求在2027年之前持續上升，而具約束力的限制則持續自加速器（accelerator）向外移動。它移向記憶體，SK hynix於此表示客戶需求超出其供給能力[S-003]；移向先進封裝（advanced packaging）與測試，其產能計畫延伸至2028年與2029年[S-024; S-116]；並移向發電設備與電網併接，GE Vernova於此計畫將燃氣渦輪機年產出自2026年的每年20 GW提高至2028年的24 GW與2030年的30 GW[S-012]。經由Open Compute Project（OCP）發展的開放介面，在眾多供應商相互交接之處擴散，而各平台則保有自有的垂直擴展網路。各層級依不同時鐘擴張，延誤風險即落在其間的落差之中（A-11）。

IF 在買方確認提高預算、供應商達成其自訂產能時點的情況下，基準路徑即成立。 它要求大型買方將其2027年的方向轉化為更高的數字化財測指引，並要求記憶體、封裝與電力供應商如其所言增添產能：SK hynix龍仁第1期無塵室於2027年初[S-003]、台積電（TSMC）14倍光罩尺寸的Chip-on-Wafer-on-Substrate（CoWoS）封裝於2028年[S-024]，以及GE Vernova於2028年達成24 GW的燃氣渦輪機年產出[S-012]。應觀察Alphabet、Microsoft、Amazon與Meta首次提出的2027年資本支出（capex）數字化財測指引。若有數家大型買方下調其2027年計畫，或上述產能時點之一延後一年以上，則此一路徑不成立。

IF 替代路徑一：記憶體於2028年之前鬆動。 若新增的動態隨機存取記憶體（dynamic random-access memory, DRAM）與高頻寬記憶體（high-bandwidth memory, HBM）產能到位的速度快於需求，則記憶體占雲端預算的比重將低於TrendForce的預測：DRAM與NAND快閃記憶體占主要雲端業者2026年資本支出（capex）的47%，2027年則占68%[S-007]。系統整合商的成本亦將隨之緩解。應觀察伺服器DRAM合約價格逐季下跌、某家HBM供應商表示供給已能滿足需求，以及同一加速器（accelerator）平台有數家供應商完成驗證（qualification）。目前的紀錄指向相反方向：TrendForce預期DRAM供給至2027年仍維持緊俏[S-135]。若合約價格於2027年上半年仍在上漲，則此一替代路徑不成立。

IF 替代路徑二：電力與電網延誤限制部署。 在此一路徑中，加速器（accelerator）與機櫃均可取得，但已送電的場址則否。TrendForce預測，電網送電延誤將於2028年後更為頻繁，且至2030年，全球資料中心490.7 GW的電力需求容量將面對222.6 GW的可用電網容量，落差達268 GW；上述均為容量數字，而非年度能源量[S-043]。發電設備已遭提前預訂：Siemens Energy於2026年6月表示，其燃氣渦輪機已「booked out onto '28」[S-026]。應觀察聯邦能源管理委員會（Federal Energy Regulatory Commission, FERC）就其2026年6月向六家電網業者發出之要求說明理由命令所採取的行動[S-068]。若併接時間縮短，或渦輪機未出貨訂單（backlog）連續兩季下降，則此一路徑將趨於薄弱。

IF 替代路徑三：資金趨緊或交易對手出現壓力。 此輪建置已倚賴外部融資。國際能源總署（International Energy Agency, IEA）表示，資料中心投資「have grown too large to be funded from company balance sheets alone」[S-005]。Oracle預期於其2027財年（2026年6月至2027年5月）募集約400億美元的債務與股權，其中包括第1季已完成的200億美元隨市場發股權[S-044; S-049]。若融資成本上升，或某一大型承租人的承諾轉弱，則倚賴已簽約需求的產能將遭延後。應觀察買方申報文件中的債務與股權發行、客戶預付款與租賃條件。若最大買方持續以營運現金流支應資本支出（capex）的成長，則此一路徑將趨於薄弱。

IF 第四項變體即架構轉變，它會改變稀缺的對象，而非消除稀缺。 TrendForce預期客製加速器（accelerator）將在商用平台之外持續成長[S-009]，光學正移入封裝之中[S-024]，而更高電壓的配電——目前為±400伏直流（VDC），其後為800 VDC——正被納入規格，以供可自800 kW擴展至1 MW以上的電源機櫃（power rack）使用[S-027; S-019]。上述任一項若加速轉變，都會使需求在供應商之間移轉，從可插拔模組移向雷射與封裝，或從48 V轉換移向更高電壓的電力電子，但不會降低總需求。應觀察指名新架構的量產設計導入案。若此一轉變停滯於展示階段，而既有設計吸納了下一世代平台，則此一變體將趨於薄弱。

A 這些替代路徑宜視為一組須持續監測的序列，而非相互競爭的預測。 它們共有一項機制：某一層級的鬆動會使限制移向下一層級，而非消除限制。其訊號亦依不同時鐘到來。DRAM合約價格逐季變動[S-035]；融資計畫隨各財年訂定[S-044]；渦輪機時段與電網併接則歷時數年才會變動[S-012; S-026]。僅追蹤記憶體價格的讀者，會看到第一次移轉，卻看不到限制最終落於何處。表20依可供查核的順序，列出每一項發現所對應的訊號。

觀察重點

首次提出的2027年資本支出（capex）數字化財測指引：Alphabet表示將「provide more details at a later date」[S-016]，而Meta尚未提出具體的2027年展望[S-074]。

9.2 2027年：各項計畫迎向其產能時點

2027年是2025年至2026年短缺期間所訂多項計畫首度到達交付時點的一年。新的記憶體無塵室、晶圓廠產能爬坡、機櫃平台與新一代交換器均排定於此時到位，大型買方亦將訂定其2027年預算。因此，2027年的各項展望是本章中最可供檢驗者：其中多數可在該年結束之前，由一份註明日期的揭露、一次季度財報或一項法規行動予以確認或推翻。

F 多項產能進程排定於2027年。 SK hynix正進行投資，以在其龍仁第1期無塵室於2027年初啟用後迅速擴充產能[S-003]，Micron則預期HBM4E將於2027日曆年量產[S-128]。台積電（TSMC）位於亞利桑那州的第二座晶圓廠排定於2027年下半年開始量產[S-083]。ASML計畫為2027年將其低數值孔徑（low-NA）極紫外光（extreme-ultraviolet, EUV）與深紫外光（deep-ultraviolet, DUV）浸潤式微影系統的2026年產能增加30%，並正評估於2028年再增加30%[S-025]。

F 平台與機櫃的產能爬坡亦已註明日期。 NVIDIA表示其Vera Rubin平台正進入全面量產，機櫃已在CoreWeave、Google Cloud、Microsoft Azure與Oracle Cloud Infrastructure等合作夥伴處運轉[S-010]。鴻海（Foxconn）預期其AI機櫃出貨量於2026年成長逾一倍，Vera Rubin機櫃則於第3季進入量產[S-022]。AMD表示其Helios機櫃已進入量產，OpenAI預期自2026年第4季起將其上線，MI500世代則於2027年接續推出[S-236]。2025年5月，NVIDIA表示800 VDC資料中心的全面量產將與其2027年的Kyber機櫃級系統（rack-scale system）同步[S-028]。

F 政策與市場相關時點亦落在2027年。 美國給予阿拉伯聯合大公國G42與Core42無須取得許可即可接收先進運算物項的授權（authorisation），除經工業及安全局（Bureau of Industry and Security, BIS）另行通知，將於2027年4月6日自動失效[S-315]。歐盟執行委員會（European Commission）預期，依其新評級制度對個別資料中心核發的首批永續標示將於2027年展示[S-437]。ASML將於2027年6月10日的資本市場日更新其較長期的觀點[S-025]。

A 2027年的檢驗在於這些進程能否在需求再度超前之前到位。 每一項進程僅為某一層級增添供給，而非為所有層級。新的記憶體無塵室不會增添封裝產能，而機櫃產能爬坡加快也不會增添渦輪機時段——GE Vernova的產出計畫要到2028年才將其提高至每年24 GW[S-012]。本報告將2027年研讀為限制版圖由交付而非由宣布加以檢驗的一年。若記憶體、封裝與機櫃的產能爬坡如期到位，而發電與電網併接未能如期，則限制將進一步移向電力（第7.2節）；若記憶體產能爬坡延後，則限制仍停留在零組件（第3.2節）。

? 未解問題——原生800 VDC的OCP規格（specification）何時問世。 Google、Microsoft與NVIDIA於2026年8月的更新表示，「Mt Diablo 2.0 will be coming later to support native 800V」，但未給出日期[S-019]。在註明日期的草案出現之前，800 VDC的2027年時程所依據的是企業計畫，而非OCP文件；此點對於在±400 VDC與800 VDC機房之間作選擇的買方具有重要意義。

A 聖荷西峰會與首屆Moscone峰會是檢查點，而非檢驗。 主題演講、場次或展示顯示的是關注度；它並不顯示採用情形，也不會改變規格（specification）的狀態。2027年的各項展望隨文件事件而變動：新版本、具名的貢獻者、目標日期，或買方明示的需求。對其影響最大的OCP文件為原生800 V的Mt Diablo 2.0[S-019]、固態變壓器（solid-state transformer, SST）規格v0.3.0之後的修訂版[S-054]、Open Data Center設計原則v0.7.0之後的版本[S-001]，以及首批共同封裝光學（co-packaged optics, CPO）規格——據業界媒體報導，OCP內部一個由20家公司組成的聯盟預期於2026年第4季提交該等規格[S-285]。兩屆峰會分別於2026年10月12日至15日與2027年10月4日至7日舉行[S-067; S-344]。

> **觀察重點**
> 2026年11月9日至10日：美國關係企業規則（Affiliates rule）為期一年的暫緩，除經延長，將於11月9日屆期，而2025年11月美中安排的關鍵條款則有效至11月10日[S-414; S-415]。

展望1 · 至2027年 · memory

IF 至2027年，記憶體仍為配置取得，而非可自由採購。

成立條件	AI產能需求持續上升，而新增的DRAM與HBM產能——例如SK hynix於2027年初啟用的龍仁第1期無塵室[S-003]，以及Micron於2027日曆年的HBM4E量產[S-128]——雖增添數量，卻未能追上需求。
觀察重點	長期協議（long-term agreement），SK hynix已就記憶體整體與約10家客戶完成簽訂[S-003]；伺服器DRAM合約價格，TrendForce預期2026年將上漲約270%[S-007]；同一平台在一家以上供應商完成HBM驗證（qualification）；TrendForce認為DRAM供給至2027年仍維持緊俏的看法[S-135]。
不成立條件	伺服器DRAM合約價格連續兩季下跌，或某家HBM供應商表示其供給已能滿足或超過需求。
下次檢視	SK hynix、Samsung與Micron的下一次季度財報。

展望2 · 至2027年 · power

IF 800 VDC於2027年進入量產，作為交流（alternating current, AC）與48 V配電之外的一項選項，而非取代兩者。

成立條件	NVIDIA維持其於2027年與Kyber機櫃級系統（rack-scale system）同步全面量產800 VDC的計畫[S-028]；固態保護與安全認證（certification）跨出開發階段，而SolarEdge與Infineon的斷路器工作目前仍處於該階段[S-055]；OCP的±400 VDC Mt Diablo電源機櫃（power rack）出現原生800 V的後繼版本，其貢獻者曾將800 VDC描述為「an additional option that can coexist with existing AC systems」[S-027; S-019]。
觀察重點	NVIDIA 800 V電源機櫃（power rack）的出貨，TrendForce預期其自2026年第3季起作為Vera Rubin的選配組態[S-029]；Kyber的時程，業界媒體報導其延後至2028年，而NVIDIA則表示其「roadmap is intact」[S-186]；一份註明日期的Mt Diablo 2.0草案[S-019]。
不成立條件	至2027年底仍未有800 VDC或±400 VDC機櫃的量產出貨獲得公告，或直流（direct current, DC）保護的認證（certification）延後至2027年之後。
下次檢視	OCP全球峰會（OCP Global Summit），聖荷西，2026年10月12日至15日[S-067]；OCP全球峰會，舊金山，2027年10月4日至7日[S-344]。

展望3 · 至2027年 · cooling

IF 2027年，液冷（liquid cooling）為旗艦AI系統的預設方案，氣冷則保留給密度較低的設備。

成立條件	加速器（accelerator）平台持續以液冷（liquid cooling）形式出貨，而冷卻液分配單元（coolant distribution unit, CDU）依已公布的百萬瓦級介面建置，例如2 MW的Project Deschutes規格（specification）[S-030]。
觀察重點	TrendForce所統計AI晶片的液冷（liquid cooling）滲透率，2025年約33%、2026年53%、2027年接近60%[S-006]；依加速器（accelerator）參考設計（reference design）完成驗證（qualification）的CDU，例如LG的2.6 MW機組，LG表示其符合NVIDIA DSX參考設計的要求[S-209]；列為OCP Accepted產品的CDU。
不成立條件	某一已公布的數列顯示，液冷（liquid cooling）在2027年出貨的AI晶片中占比低於一半，或旗艦平台大規模出貨氣冷版本。
下次檢視	TrendForce的下一次液冷（liquid cooling）更新；OCP全球峰會（OCP Global Summit），2026年10月12日至15日[S-067]。

展望4 · 至2027年 · optics

IF 共同封裝光學（co-packaged optics, CPO）交換器於2027年量產出貨，但僅限於單一供應商的供應鏈之內。

成立條件	交換器CPO在已啟動之處持續爬坡：鴻海（Foxconn）的CPO全光交換器原有望於2026年第3季量產並出貨[S-022]，NVIDIA的Spectrum-6交換器系統同時支援可插拔與共同封裝光學（co-packaged optics）[S-010]，台積電（TSMC）亦已啟動COUPE量產[S-084]；共用的測試與驗證（qualification）方法則持續落後。
觀察重點	首批CPO規格（specification）在OCP內部的提交，據業界媒體報導，一個由20家公司組成的聯盟預期於2026年第4季提交[S-285]；已公布的晶圓級測試方法，據業界媒體報導，日月光（ASE）與台積電（TSMC）高階主管曾於2026年8月指出其尚不存在[S-286]；磷化銦（indium phosphide）雷射供應，TrendForce稱其為一處瓶頸[S-037]。
不成立條件	既有供應鏈之外的某家封裝廠，以已公布的測試方法為某家超大規模雲端業者（hyperscaler）驗證（qualification）一條競爭性的CPO交換器供應鏈，或某家超大規模雲端業者自兩個來源採購同一類CPO交換器。
下次檢視	OCP CPO基礎規格（specification）的首次發布。

展望5 · 至2027年 · racks

IF 2027年不會有任何一種機櫃外型規格取代其他：以OCP為基礎的機櫃與NVIDIA的機櫃產品線並行出貨。

成立條件	AMD的Helios——AMD表示該產品係與Meta透過OCP共同開發——依公告內容完成部署[S-112; S-236]；Vera Rubin機櫃持續經由合作夥伴爬坡[S-010; S-022]；Open Rack Wide（ORW）v1.0.0仍是一組介面與公差，而非單一規定的框架[S-228]。
觀察重點	OpenAI自2026年第4季起將Helios上線[S-236]；Kyber的時程[S-186]；列為OCP Accepted的ORW產品。
不成立條件	某家營運業者公開在其整個機隊中以其中一條機櫃系譜取代另一條，或其中一條系譜退出量產。
下次檢視	OCP全球峰會（OCP Global Summit），2026年10月12日至15日與2027年10月4日至7日[S-067; S-344]。

展望6 · 至2027年 · geopolitics

IF 至2027年，美國對輸往中國之先進加速器（accelerator）的管制維持現行以許可為基礎的形式，關係企業規則（Affiliates rule）則處於暫緩或遭替代的狀態，而出口政策在部分企業的財測指引中仍為明示的假設。

成立條件	2026年11月的屆期窗口以延長或替代而非失效的方式結束[S-414; S-415]；2026年1月針對H200級產品的逐案許可審查維持有效[S-412]。
觀察重點	2026年11月9日，除經延長，BIS關係企業規則（Affiliates rule）的暫緩於該日結束[S-414]；2026年11月10日，2025年11月美中安排的關鍵條款於該日屆期，而白宮2026年9月25日的事實說明文件並未提及延長一事[S-415; S-428]；AI OVERWATCH Act，即參議院就2027財年國防授權法（National Defense Authorization Act）所提的條款[S-420]；NVIDIA下一次展望中關於中國的句子，因其2027財年第3季（即截至2026年10月之當季）的展望假設來自中國的Data Center運算營收為零[S-010]；最新展望對中國未置一詞的AMD與KLA，是否重新截明其對中國的假設[S-111; S-587]。
不成立條件	關係企業規則（Affiliates rule）未經修改即生效，或針對效能最高之晶片的法定禁令獲得制定；或某項持久的協議以一般許可重新開放加速器（accelerator）貿易。
下次檢視	2026年11月9日至10日；NVIDIA 2027財年第3季財報。

展望7 · 至2027年 · memory

IF 至2027年，更精簡的開放權重模型（open-weight model）會使記憶體需求在各層級之間移轉，而非減少需求。

成立條件	token用量持續每年成長數倍，一如Google在截至2026年5月的一年間所呈現者[S-491]；HBM供給維持短缺，依TrendForce的估計，2027年HBM位元出貨量成長50%至60%，仍不足以因應需求[S-135]；新設計持續降低每token快取用量，一如DeepSeek-V4.1-Flash所為[S-529]。
觀察重點	HBM與伺服器DRAM合約價格；用於KV快取（key-value cache）卸載的企業級SSD訂單，TrendForce表示由DeepSeek帶動的架構正在中國推升此類需求[S-549]；NVIDIA Rubin Ultra的每GPU HBM配置，TrendForce報導其正評估較低的組態[S-135]；下一批DeepSeek、Qwen與Kimi發布版本的每token快取位元組數。
不成立條件	在已公布的token用量仍在成長之際，HBM或伺服器DRAM合約價格連續兩季下跌。
下次檢視	TrendForce的季度記憶體價格更新；記憶體製造商2026年第3季財報。

9.3 2030年：電力、資本與封裝決定步調

至2030年，註明日期的紀錄趨於稀薄。少有企業公布該年的數字，而公布者所載明的是目標、計畫或情境（scenario），而非財測指引。2030年確實具備的，是一批關於渦輪機、封裝、記憶體晶圓廠與雲端電力的產能時點，以及這些產能所欲因應的電力需求預測。因此，2030年的各項展望所檢驗的是：較慢的層級是否能追上較快的層級，以及若未能追上，誰會取得籌碼（leverage）。

F 延伸至2030年的企業預測期限（horizon）層目標與情境（scenario），而非財測指引。 CoreWeave表示其有望於2030年前達到至少8 GW，該目標定義為已啟用電力[S-079; S-078]，相對於2026年6月底的約1.5 GW[S-011]。GE Vernova表示其正為2030年的燃氣渦輪機產出「implementing actions to produce 30 GW」[S-012]。ASML約440億至600億歐元的2030年營收區間，係取自其2024年11月投資人日的一組「potential 2030 scenarios」，預定於2027年6月更新[S-328; S-025]。Samsung與Broadcom依一份合作意向書（memorandum of understanding, MOU），

共同估計其記憶體暨晶圓代工合作至2030年將逾2,000億美元；該數字並未載明為營收或訂單[S-115]。

F 這些產能所因應的需求，預測將大致倍增。 國際能源總署（IEA）預估全球資料中心用電量將自2025年的485 TWh上升至2030年的950 TWh，約占全球電力需求的3%[S-005]。就美國而言，勞倫斯柏克萊國家實驗室（Lawrence Berkeley National Laboratory, LBNL）給出2030年參考情境為649 TWh，相當於美國預測用電量的11.8%，並落在521至843 TWh的情境（scenario）區間之內[S-442]。TrendForce以吉瓦而非能源量衡量電力需求容量，預測2030年全球為490.7 GW[S-043]。此三組數列在單位、範圍與方法上各異；它們僅並列研讀，絕不合併。

A 本報告將2030年研讀為對較慢時鐘的檢驗。 2026年公告的記憶體與封裝產能排定於兩至三年內投產：SK hynix於2026年8月在印第安納州動工，以供2029年下半年的HBM量產[S-023]，而台積電（TSMC）14倍光罩尺寸的CoWoS則排定於2028年[S-024]。發電與電網容量所需時間更長，TrendForce亦預期電網送電延誤將於2028年後更為頻繁[S-043]。若較慢的層級持續落後，則一處已送電、已併接的場址相對於其中的硬體將更具價值，而議價能力（bargaining power）則移向掌握場址、設備產能時段與資金者（第5.6節）。

> “From a structural perspective, the market remains supply-constrained rather than demand-constrained, with strong backlog visibility and a growing number of commitments including reservation agreements.” [S-026]

F 封裝、測試與微影的計畫延伸至2028年與2029年。 台積電（TSMC）已將一種14倍光罩尺寸的CoWoS排定於2028年投產，該封裝可整合約10個大型運算晶粒與20個HBM堆疊，其後於2029年擴展至14個光罩以上，並推出40倍光罩尺寸的System-on-Wafer（SoW-X）[S-024]。Advantest以2028年底或2029年初達成10,000台系統單晶片（system-on-chip, SoC）測試機為目標，並表示正提前該目標[S-116]。ASML為2028年再增加30%微影產能一事仍在評估之中，尚非計畫[S-025]。

A 若記憶體緊動，稀缺將移轉而非消失。 紀錄顯示稀缺接下來會移往何處：移向封裝尺寸與良率（yield），因CoWoS正邁向14個光罩以上[S-024]；移向測試產能，Advantest將其與封裝擴充及每一世代晶片測試內容的增加相連結[S-116]；移向光學，TrendForce稱磷化銦（indium phosphide）雷射供應為一處瓶頸[S-037]；並移向電力設備與電網併接[S-043]。上述每一項各有其領先指標，且無一能僅憑規格（specification）即獲解決。此為本報告對下一處限制所在的研判（reading），而非對其時點的預測（A-12）。

F 預訂契約正從記憶體擴散至電力零組件與渦輪機。 SK hynix已與約10家客戶完成簽訂長期協議（long-term agreement, LTA）[S-003]，Kioxia則表示其正自一年期協議轉向多年期協議[S-130]。Infineon表示，數家主要的AI資料中心客戶已簽署或正在洽談多年期產能預訂協議，累計金額達高個位數十億歐元，並附帶部分預付款[S-021]。Siemens Energy的目標是平均於六至十二個月內將燃氣渦輪機產能時段預訂轉為訂單[S-026]。

A 預訂正是稀缺轉化為議價能力（bargaining power）的途徑。 當買方提前數年預訂產能並預付部分款項時，稀缺投入要素的供應商即取得定價籌碼（leverage）與營收可見度，而未取得預訂的買方則須等待。上述協議以多年期條件，將成長最快的買方與記憶體、電力半導體及渦輪機製造商綁在一起[S-003; S-021; S-301]。至2030年，對系統整合商與後進者而言，問題不在於硬體是否存在，而在於其在各個排隊佇列中是否占有一席。應觀察預訂條件是否擴散至冷卻、光學與電網設備，以及在需求出現停頓時是否仍然維持（第5.8節）。

範圍 OCP認可的支出

OCP／IDC數列衡量的是全球依OCP認可規格（specification）所建置之資訊科技基礎設施與解決方案的支出，並包含非會員的支出[S-004]。它並非OCP Foundation的營收、AI資料中心市場或超大規模雲端業者（hyperscaler）的資本支出（capex），而其AI數字則是該總額之中的AI應用案例部分[S-108]。

展望8 · 至2030年 · grid

IF 朝向2030年，決定新增AI產能何時送電者為電網併接與發電設備，而非加速器（accelerator）。

成立條件	電力需求在已公布的預估範圍內成長，即國際能源總署（IEA）所估2030年全球950 TWh[S-005]，以及LBNL所估美國521至843 TWh[S-442]，同時發電設備分階段增添：GE Vernova計畫2028年燃氣渦輪機年產出達24 GW，並朝2030年的30 GW推進[S-012]。
觀察重點	TrendForce預測資料中心電力需求容量與可用電網容量之間的落差，至2030年全球為268 GW，美國則逾170 GW[S-043]；各製造商所公布的渦輪機承諾，GE Vernova的未出貨訂單（backlog）與產能時段預訂合計116 GW，Siemens Energy則為確定未出貨訂單69 GW，另有產能時段預訂26 GW，兩者均為2026年6月30日之數[S-012; S-301]；聯邦能源管理委員會（FERC）就大型負載（large load）併接所採取的行動[S-068; S-430]。
不成立條件	確定的電網併接能以場址興建的步調取得，或渦輪機未出貨訂單（backlog）連續兩季下降。
下次檢視	GE Vernova與Siemens Energy的季度財報；聯邦能源管理委員會（FERC）就各電網業者回覆所發布的命令[S-068]。

展望9 · 至2030年 · finance

IF 至2030年，AI產能背後的資本結構，將與其中的硬體同樣具決定性。

成立條件	投資持續超越內部現金創造能力——國際能源總署（IEA）表示此事已然發生[S-005]——致使債務、股權、租賃與客戶預付款支應其中愈來愈高的比重。
觀察重點	Oracle計畫於2027財年募集約400億美元的債務與股權[S-044]；與Oracle資本支出（capex）相抵的客戶預付款，於2027財年第1季為114億美元[S-049]；使投資落在帳列資本支出之外的租賃重分類，一如Microsoft自2027財年起的變更[S-075]；CoreWeave自2026年6月約1.5 GW已啟用電力朝其8 GW目標推進的進程[S-011; S-079]。
不成立條件	最大買方以營運現金流支應資本支出（capex）的成長，且外部融資、租賃與預付款占投資的比重不再上升。
下次檢視	各買方首次提出的2027年資本支出（capex）數字化財測指引及其年度報告。

展望10 · 至2030年 · compute

IF 至2030年，客製加速器（accelerator）在AI伺服器中的占比上升，但並未取代圖形處理器（graphics processing unit, GPU）。

成立條件	雲端業者的自研與合作計畫持續進入量產，例如OpenAI與Broadcom合作的10 GW客製加速器（accelerator），將自2026年下半年部署至2029年底[S-243]，同時市售GPU平台維持其年度節奏[S-010]。

展望10 · 至2030年 · compute

觀察重點	TrendForce預測以特殊應用積體電路（application-specific integrated circuit, ASIC）為基礎的系統，將自2026年占AI伺服器出貨量的27.8%上升至2030年的近40%[S-009]；Advantest認為就其測試機而言，客製ASIC「could become as large as, if not significantly larger than, the GPU market over time」[S-116]；Samsung與Broadcom估計其記憶體與晶圓代工合作至2030年將逾2,000億美元，該數字屬合作意向書（MOU）階段，並非營收[S-115]。
不成立條件	已公布的數列顯示ASIC占比回落，或大型客製計畫遭取消或轉向市售GPU。
下次檢視	TrendForce的下一次AI伺服器預測；Broadcom的季度財報。

展望11 · 至2030年 · market

IF OCP認可基礎設施的支出沿OCP／IDC所示路徑，於2029年達約2,950億美元。

成立條件	超大規模雲端業者（hyperscaler）、新興GPU雲端業者（neocloud）與其他買方在GPU加速AI伺服器上的支出持續成長，此即該研究所指出的驅動因素；而該研究列為支出成長最強勁的類別，即冷卻（每年逾32%）、電力（23.6%）與資訊科技機櫃（23.5%），亦維持其步調[S-004]。
觀察重點	下一次OCP／IDC年度更新，對照2025年的1,320億美元與2029年的2,950億美元，年複合成長率為22.2%[S-004]；至2029年約1,550億美元的AI基礎設施支出，IDC 2026年1月的網路研討會簡報將其列為該總額之中的AI應用案例部分[S-004; S-108]；任何破壞可比性的方法變更。
不成立條件	後續研究下調2029年的數值，或變更其範圍，致使該數列不再可供比較。
下次檢視	下一份OCP／IDC市場研究。

展望12 · 至2030年 · regions

IF 至2030年，台灣的角色所倚仗者為系統整合、先進封裝（advanced packaging）、光學與驗證，而非組裝，同時部分新增產能建於美國。

成立條件	台積電（TSMC）的封裝藍圖按其所載日期到位，即2028年的14倍光罩尺寸CoWoS，以及2029年擴展至14個光罩以上與40倍光罩尺寸的SoW-X[S-024]；台灣系統製造商持續在機櫃與交換器的產能爬坡上領先，一如鴻海（Foxconn）在Vera Rubin機櫃與CPO交換器上的表現[S-022]。
觀察重點	台積電（TSMC）額外1,000億美元的亞利桑那州投資，美國商務部表示此使其在美承諾達2,650億美元[S-084; S-163]；台積電亞利桑那州第二座晶圓廠於2027年下半年量產[S-083]；美國的機櫃產能，例如緯穎（Wiwynn）為Amazon在德州Socorro逾16億美元的擴產[S-306]；友達（AUO）的Micro LED CPO模組，目前處於研發階段[S-052]。
不成立條件	台積電（TSMC）2028年至2029年的封裝里程碑延後一年以上，或台灣企業的封裝、光學與驗證計畫仍停留於研發階段。
下次檢視	台積電（TSMC）的季度財報與技術論壇；鴻海（Foxconn）的季度財報。

展望13 · 至2030年 · power

IF 至2030年，對於機櫃屬800 kW至1 MW級的新建AI機房，800 VDC是一項常規且經認證（certification）的選項，而交流與48 V路徑則留存於既有機房。

成立條件	2027年的觀點成立（展望2）；經認證（certification）的保護硬體由一家以上供應商提供；OCP發布適用於原生800 V的Mt Diablo 2.0[S-019]。

展望13 · 至2030年 · power

觀察重點	Diablo 400電源機櫃（power rack），其規格自800 kW至1 MW以上[S-027]；OCP固態變壓器（solid-state transformer, SST）規格（specification）v0.3.0的修訂版，該規格將13.8 kV或34.5 kV輸入轉換為800 VDC[S-054]；TrendForce預期NVIDIA的800 V電源機櫃將於2028年廣泛部署[S-029]；Infineon對其2027財年（2026年10月至2027年9月）AI資料中心營收「2.5 billion Euros plus」的預估，該公司預期於2026年11月將其大幅上調[S-193]。
不成立條件	此一電壓等級下直流配電的安全認證（certification）延後至2030年之後，或800 VDC仍屬個案專屬的設計。
下次檢視	Infineon 2026年11月的年度財測指引[S-193]；2027年至2030年於Moscone舉行的OCP全球峰會（OCP Global Summit）[S-344]。

展望14 · 至2030年 · optics

IF **至2030年，CPO在AI光鏈路中承載相當比重，且來自一家以上供應商，而在傳輸距離、維修或成本較為有利之處，可插拔光學與銅纜仍維持其鏈路。**

成立條件	2027年的觀點成立（展望4），且共用的測試、介面與可維護性方法使買方得以驗證（qualification）第二條供應鏈；雷射、光纖接合與封裝的良率（yield）在一家以上供應商達到量產水準。
觀察重點	TrendForce估計CPO於2026年約占AI資料中心光收發模組的0.5%，並預測矽光子（silicon photonics）CPO至2030年前後可能達到約35%的AI資料中心滲透率[S-278]；可現場更換的雷射設計，例如據業界媒體報導Broadcom Davisson交換器中的雷射單元[S-283]；Optical Internetworking Forum（OIF）的共同封裝實作協議[S-284]。
不成立條件	至2030年，CPO仍處於驗證與有限部署階段，或僅有一家供應商量產出貨。
下次檢視	2027年至2030年於Moscone舉行的OCP全球峰會（OCP Global Summit）[S-344]。

展望15 · 至2030年 · memory

IF **至2030年，隨著代理程式使每項任務的token數倍增，記憶體需求持續上升，但主機記憶體與快閃記憶體的成長快於HBM容量。**

成立條件	推理持續拉長請求長度，一如OpenRouter的資料所示，而代理程式持續使用數倍於聊天的token數，一如Anthropic所報告者[S-557; S-558]；將快取分層至主機記憶體、共享記憶體池與SSD，成為推論服務軟體的標準做法[S-565; S-584]；每個加速器（accelerator）的HBM頻寬持續以快於其容量的速度上升[S-550; S-551]。
觀察重點	多年期記憶體協議的續約條件，例如Micron延續至2030年的策略客戶協議[S-129]；記憶體製造商關於上下文儲存與SSD需求的說法[S-566]；一份已公布的、用於在各層級之間搬移KV快取（key-value cache）的介面，例如透過以CXL為基礎的記憶體池化[S-585]。
不成立條件	代理程式的採用趨於平緩，而每token快取用量持續下降，致使記憶體占AI資本支出的比重連續數年下降。
下次檢視	記憶體製造商2027年至2029年的年度財報與展望。

9.4 2035年：結構而非時程

2030年之後，幾乎沒有企業公布數字。紀錄中僅有少數長程陳述與一小批公開期限，其中多數屬於密碼學領域。因此，2035年的各項展望屬結構性：每一項均載明紀錄的方向所支持的格局，以及何者會使其反轉。它們是本報告中最難檢驗的觀點，宜作為監測的框架，而非作為預期。

F **為數不多的2035年定錨點是長程計畫、市場觀點與期限。** Samsung的中長期計畫訂定自2026年至2040年投資約2,450兆韓元，其中約2,100兆韓元用於半導體，並稱該金額係依當前市場條件所訂的指引[S-134]。2026年3月，Siemens Energy表示其預期「an elevated gas-turbine market at least until 2035」，惟該公司在2026年6月的法人說明會中並未重述此一觀點[S-018; S-026]。註明日期的期限位於後量子密碼學（post-quantum cryptography, PQC）領域：美國國家安全系統至2033年，於雲端服務上僅得使用Commercial National Security Algorithm Suite 2.0（CNSA 2.0）[S-374]；美國國家標準與技術研究院（National Institute of Standards and Technology, NIST）並於一份草案中提議，於2035年後不再容許易受量子攻擊的公開金鑰演算法[S-461]。

F **國家層級的產能承諾延長了供應鏈的預測期限（horizon）。** 台積電（TSMC）於2026年7月宣布在亞利桑那州額外投資1,000億美元，美國商務部表示此使台積電在美承諾達2,650億美元[S-084; S-163]。SK hynix的印第安納州廠區是其在美國的首座HBM生產基地，與韓國的生產高度整合營運，下一世代HBM量產排定於2029年下半年[S-023]。Huawei已公布延伸至2028年第4季的Ascend加速器（accelerator）藍圖，並搭配其自有的垂直擴展協定UnifiedBus 2.0[S-312]。

A **在此一尺度上，問題在於結構，而非時程。** 至2035年，決定性變數在於哪些介面已成為共通介面、發電與電網容量落在何處，以及開放AI硬體是否仍為同一個社群。紀錄給出了方向，卻未給出日期。OCP的設施文件已自中壓輸入延伸至機櫃[S-054; S-059]，OCP並與電力研究院（Electric Power Research Institute, EPRI）就資料中心作為彈性電網資源一事展開合作[S-062]。供應鏈正在各區域之間複製，而非自各區域遷出[S-163; S-023]，而中國的國內藍圖則運行於其自有的介面之上[S-312]。以下三項展望研讀該一方向，並指明何者會使其反轉。

展望16 · 至2035年 · standards

IF **至2035年，開放介面涵蓋設施邊界，即電力入口、冷卻液接頭、機櫃外廓、管理與遙測（telemetry），而差異化與鎖定（lock-in）則留存於垂直擴展網路、封裝與控制軟體。**

成立條件	設施規格（specification）持續由數家大型貢獻者經OCP提出，一如AMD、Google、Meta與NVIDIA所提的Open Data Center設計原則v0.7.0[S-001]，而安全與標準機構則將其轉化為法規。
觀察重點	依平台並存的垂直擴展網路：參與企業逾175家的Ethernet for Scale-Up Networking（ESUN）[S-020]、Ultra Accelerator Link（UALink）Common Specification 2.0[S-346]，以及專有鏈路；1.0.3版之後的Ultra Ethernet發布版本[S-257]；引用OCP低壓直流工作成果的安全法規[S-059]。
不成立條件	某一專有的垂直擴展網路大規模橫跨數家供應商的平台，或主要的設施規格（specification）在OCP之外撰寫且未參照OCP。
下次檢視	2027年至2030年於Moscone舉行的每一屆OCP全球峰會（OCP Global Summit）[S-344]。

展望17 · 至2035年 · grid

IF **至2035年，大型AI園區係以電網參與者的身分規劃，配備儲能、現地發電或已簽約的彈性，而電力則經固態變壓器（solid-state transformer）自中壓流向800 VDC。**

成立條件	2030年的電網觀點成立（展望8）；大型負載（large load）規則持續增訂負載削減（curtailment）或備援義務，一如德州對75 MW以上負載所為[S-040]；固態變壓器（solid-state transformer）發展至經認證（certification）的產品[S-054]。

展望17 · 至2035年 · grid

觀察重點	聯邦能源管理委員會（FERC）於案號RM26-4中就大型負載（large load）併接所採取的行動，該案仍處於預告階段[S-430]；OCP與EPRI就資料中心作為彈性電網資源之合作的產出，以及OCP對儲能系統（energy storage system）的要求[S-062]；固態變壓器（SST）規格（specification）的修訂版[S-054]。
不成立條件	建置消化了併網排隊佇列（interconnection queue），且不附彈性義務的確定併接成為常態，或固態變壓器（SST）產品未能通過認證（certification），致使營運業者維持以變壓器搭配機櫃端轉換的做法。
下次檢視	聯邦能源管理委員會（FERC）於案號RM26-4中的下一道命令[S-430]。

展望18 · 至2035年 · geopolitics

IF **至2035年，開放AI硬體以兩個相互重疊的生態系運行，一個環繞OCP，另一個以中國為中心，兩者共享已公布的規格（specification），但在晶片、垂直擴展網路與認證（certification）上分道而行。**

成立條件	美國的管制與中國的國內藍圖並行推進，Huawei將其Ascend 960定於2027年第4季、970定於2028年第4季，並採用其自有的垂直擴展協定[S-312]；OCP的規格（specification）維持公開發布，其中國大陸社群亦持續公開運作[S-460]。
觀察重點	2026年11月各項安排的延長、替代或失效[S-414; S-415]；先進運算物項許可政策的變更[S-412]；2026年9月設立的美中超級智慧對話，其下一次交流預定於2026年11月前進行[S-428]；UnifiedBus在Huawei之外的採用情形[S-312]。
不成立條件	某項持久的協議以一般許可重新開放加速器（accelerator）貿易，或相關規則禁止參與開放標準工作，致使OCP失去其中國大陸社群。
下次檢視	2026年11月9日至10日[S-414; S-415]。

展望19 · 至2035年 · memory

IF **至2035年，以中國為中心的生態系與其餘各方之間，記憶體階層架構將有所不同，一如加速器（accelerator）目前的情形。**

成立條件	美國的規則持續將記憶體頻寬與HBM列為管制參數[S-412; S-569]；中國的開發者持續以國產加速器（accelerator）與記憶體為考量來設計模型，DeepSeek的數值格式與Huawei的HBM計畫即顯示此點[S-578; S-312; S-573]；CXMT與其他國內製造商大規模供應DRAM[S-574]。
觀察重點	Huawei的HBM與SuperPoD記憶體藍圖[S-573]；中國的旗艦模型是否持續公布其組態，並優先於國產硬體上提供服務[S-579]；關係企業規則（Affiliates rule）與HBM級產品的許可政策[S-505]。
不成立條件	與中國之間HBM級記憶體的貿易在一般許可下重新開放，或中國的旗艦模型回歸需要進口HBM的設計。
下次檢視	每年的HUAWEI CONNECT與OCP全球峰會（OCP Global Summit）。

9.5 至2035年的路標

這些路標（signpost）區分兩類日期。註明日期的外部事件係由他人所定：某家企業排定的量產起始日、某一監管機關的期限、某項OCP活動。每一項均附來源。本報告的情境（scenario）窗口則完全不是日期；它們標示上述條件式展望最早可能獲得確認或被推翻之處。圖32將兩者分置於不同的列，而表19僅列出企業自行揭露的內容，不含分析師估計值。

A **企業自行揭露的預測期限（horizon）在2028年之後趨於稀薄，公開期限則取而代之。** 在表19中，多數目標與里程碑落在2026年至2028年；僅少數延伸至2029年與2030年，其中包括台積電（TSMC）的A13、

圖32。 註明日期的事件群聚於2030年之前；其後，展望所依憑者為條件，而非時程

取自企業、OCP與公共政策來源的註明日期外部事件（第1至3列），以及本報告的情境（scenario）窗口（第4列），期間為2026年至2035年，依月份定位。僅註明年份者置於年中，註明季別者置於該季中間月份，註明「年初」者置於2月，註明下半年者置於10月，而「2035年後」則置於2035年底。來源：[S-067; S-344; S-285; S-022; S-003; S-025; S-083; S-084; S-024; S-012; S-023; S-079; S-414; S-415; S-315; S-437; S-375; S-374; S-461]。

第1至3列為附來源的外部日期（深藍圓點）；CPO的日期是據業界媒體報導預期提交首批規格（specification）的時點，而非發布時點。第4列為APEX分析（灰色空心菱形）：每一窗口標示本章某一項展望最早可能獲得確認或被推翻之處，而非預測日期。

A12與SoW-X[S-024]、SK hynix的印第安納州HBM產線[S-023]、GE Vernova的30 GW產出計畫[S-012]，以及CoreWeave的8 GW目標[S-079]。2030年之後，該表僅存Samsung的投資指引[S-134]，而圖32中註明日期的事件則為密碼學期限，其中NIST的期限僅見於草案提議[S-374; S-461]。因此，長程展望所依憑者為條件（conditions）與結構，而來自任一層級的新多年期目標，本身即是一項路標（signpost）。

9.6 何者會改變此一觀點

本報告的十一項發現是論點，而非結論（第1.3節）。表20就每一項列出其目前所依憑的證據、會強化該項的訊號、會使該項不成立的訊號，以及下一次註明日期的檢視。該表係以監測總覽表的形式撰寫：政府、研發與投資讀者可逐季追蹤同一批列，而否證欄一旦出現變動，即構成重新檢視該項發現所出自章節的理由。

A 最早的幾次檢視落在2027年中之前。 其中三項論點可能在數月內出現變動。出口政策面對的是2026年11月9日至10日的屆期日[S-414; S-415]。記憶體面對的是每一季的合約價格與供應商財報，TrendForce並預測記憶體占雲端資本支出（capex）的比重於2027年將進一步上升[S-007]。資本支出與融資兩項論點面對的是各買方首次提出的2027年數字化財測指引[S-016]。電網、台灣與介面三項論點變動較慢，依循渦輪機未出貨訂單（backlog）、封裝藍圖與規格（specification）發布的時鐘[S-012; S-024; S-019]。僅研讀變動較快的各列，可看出限制正在移動，卻看不出它最終落於何處。

讀者視角

本展望對各類讀者的意義

金融。 在數字出現之前，應將2027年資本支出（capex）的方向視為尚未確認；對融資條件、預付款與租賃會計的追蹤，應與對資本支出本身同樣密切。

政府。 電網併接、大型負載（large load）規則與安全認證（certification）依循最慢的時鐘，且掌握在公部門手中；多數2030年的窗口由其決定。

產業。 渦輪機產能時段、電網併接與多年期記憶體協議，如今均須在硬體下單數年之前先行確保；後進者只能排在已預訂產能之後。

研發。 2027年至2030年的未解問題為直流保護與認證（certification）、CPO測試與可維護性，以及記憶體分層：均屬各項規格（specification）留待解決的層級。

智慧財產權。 隨著設施介面趨於共通，差異化將移向保護裝置、封裝、測試與控制軟體；應就該等領域評估相關請求項。

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-003 SK hynix (29 Jul 2026); S-010 NVIDIA (26 Aug 2026); S-011 CoreWeave (11 Aug 2026); S-012 GE Vernova (22 Jul 2026); S-016 Alphabet (22 Jul 2026); S-018 Siemens Energy (30 Mar 2026); S-021 Infineon Technologies (5 Aug 2026); S-022 Hon Hai (12 Aug 2026); S-023 SK hynix (28 Aug 2026); S-024 TSMC (23 Apr 2026); S-025 ASML (15 Jul 2026); S-026 Siemens Energy (29 Jun 2026); S-028 NVIDIA Technical Blog (20 May 2025); S-039 GE Vernova (9 Dec 2025); S-044 Oracle (10 Jun 2026); S-049 Oracle (10 Sep 2026); S-052 AUO (31 Aug 2026); S-055 SolarEdge and Infineon (9 Sep 2026); S-056 Samsung Electronics (12 Feb 2026); S-074 Meta Platforms (29 Jul 2026); S-075 Microsoft (29 Jul 2026); S-077 CoreWeave (26 Feb 2026); S-078 CoreWeave (7 May 2026); S-079 CoreWeave (11 Aug 2026); S-083 TSMC (16 Apr 2026); S-084 TSMC (16 Jul 2026); S-111 AMD Investor Relations (4 Aug 2026); S-112 AMD (24 Feb 2026); S-115 Samsung and Broadcom (25 Jul 2026); S-116 Advantest (29 Jul 2026); S-128 Micron Technology (24 Jun 2026); S-129 Micron Technology (24 Jun 2026); S-130 KIOXIA (2 Jun 2026); S-131 KIOXIA (2 Jun 2026); S-134 Samsung Electronics (29 Jun 2026); S-193 Infineon Technologies (5 Aug 2026); S-196 Vertiv (29 Jul 2026); S-209 LG Electronics (22 Sep 2026); S-236 AMD (23 Jul 2026); S-243 Broadcom Investor Relations (13 Oct 2025); S-301 Siemens Energy (5 Aug 2026); S-306 Wiwynn (10 Sep 2026); S-312 Huawei (18 Sep 2025); S-328 ASML (14 Nov 2024); S-462 Microsoft (29 Oct 2025); S-486 Advanced Micro Devices (3 Feb 2026); S-487 KLA (29 Oct 2025); S-491 Google (19 May 2026); S-529 DeepSeek-AI (10 Sep 2026); S-550 NVIDIA (21 Jul 2026); S-551 NVIDIA (21 Sep 2026); S-558 Anthropic (13 Jun 2025); S-565 SK hynix (7 Aug 2026); S-566 KIOXIA (2 Jun 2026); S-573 Huawei Technologies (17 Sep 2026); S-574 CXMT Corporation (29 Aug 2026); S-584 NVIDIA (18 Sep 2025); S-587 KLA (28 Jul 2026)

標準組織與聯盟. S-001 OCP (27 Aug 2026); S-004 OCP (1 Dec 2025); S-019 OCP (11 Aug 2026); S-020 OCP (10 Mar 2026); S-027 OCP (1 Mar 2026); S-030 OCP (26 Feb 2026); S-054 OCP (22 Jun 2026); S-059 OCP (2 Sep 2026); S-062 OCP (29 Apr 2026); S-067 OCP (未註明日期); S-228 OCP (28 Apr 2026); S-257 Ultra Ethernet Consortium (16 Jul 2026); S-284 OIF (29 Mar 2023); S-344 OCP (1 Sep 2026); S-346 UALink Consortium (7 Apr 2026); S-460 OCP (未註明日期); S-585 CXL Consortium (21 Aug 2026)

政府與監管機關. S-005 IEA (16 Apr 2026); S-040 Texas Legislature (2025); S-068 Federal Energy Regulatory Commission (18 Jun 2026); S-163 U.S. Department of Commerce (16 Jul 2026); S-315 Federal Register (BIS) (14 Jul 2026); S-374 U.S. National Security Agency (Apr 2024); S-375 Federal Register (25 Jun 2026); S-412 Federal Register (BIS) (15 Jan 2026); S-414 Federal Register (BIS) (12 Nov 2025); S-415 The White House (1 Nov 2025); S-428 The White House (25 Sep 2026); S-430 Federal Energy Regulatory Commission (22 Jul 2026); S-437 European Commission (21 Sep 2026); S-461 NIST (12 Nov 2024); S-505 U.S. Department of Commerce (12 Nov 2025); S-569 U.S. Department of Commerce (5 Dec 2024)

研究機構. S-442 LBNL (Jun 2026); S-557 OpenRouter (Dec 2025)

市場研究. S-006 TrendForce (17 Aug 2026); S-007 TrendForce (25 Aug 2026); S-009 TrendForce (18 Mar 2026); S-029 TrendForce (25 Jun 2026); S-035 TrendForce (1 Jun 2026); S-043 TrendForce (16 Sep 2026); S-108 OCP / IDC (22 Jan 2026); S-135 TrendForce (4 Aug 2026); S-278 TrendForce (11 Mar 2026); S-549 TrendForce (21 Sep 2026)

媒體與法律評論. S-037 TrendForce (6 Aug 2026); S-070 StockAnalysis (30 Jul 2026); S-186 Tom's Hardware (6 Jul 2026); S-283 ServeTheHome (12 Oct 2025); S-285 Tech Times (15 Aug 2026); S-286 Tech Times (12 Aug 2026); S-420 The AI Policy Network (31 Aug 2026); S-578 National Business Daily (21 Aug 2025); S-579 Cailian Press (24 Apr 2026)

表19。 企業自行揭露的預測期限在2028年之後趨於稀薄；2030年之後僅存一項投資指引

多數已公布的目標與里程碑至2029年至2030年即告中止，因此長程展望所依憑者為條件（conditions）與結構，而非企業承諾。

層級	2026	2027	2028	2029	2030	2031–35
超大規模雲端業者（hyperscaler）與新興GPU雲端業者（neocloud）	CoreWeave：至年底已啟用電力逾1.85 GW（C）[S-079]	Amazon：至年底電力容量達2025年的兩倍（T）[S-070]；CoreWeave：2025年底已簽約的3.1 GW幾近全數上線（C）[S-077]；Microsoft：資料中心總規模於2025年10月起兩年內大致倍增，2026年7月表示仍如期推進（T）[S-462; S-075]	—	—	CoreWeave：已啟用電力至少8 GW（T）[S-079; S-078]	—
運算	NVIDIA Vera Rubin進入全面量產（R）[S-010]	AMD MI500（R）[S-236]；Huawei Ascend 960，第4季（R）[S-312]	Huawei Ascend 970，第4季（R）[S-312]	OpenAI與Broadcom：至年底部署10 GW（C）[S-243]	—	—
記憶體	SK hynix HBM4於下半年（2H）產能爬坡（R）[S-003]；Samsung HBM銷售額逾2025年的3倍（G）[S-056]	SK hynix龍仁第1期無塵室，2027年初（C）[S-003]；Micron HBM4E量產（R）[S-128]；Samsung客製HBM樣品（R）[S-056]	SK hynix印第安納州無塵室於10月前完成（C）[S-023]；Kioxia至2028財年資料中心營收占比逾60%（T）[S-131]	SK hynix印第安納州HBM於下半年（2H）量產（R）[S-023]	—	Samsung：2026年至2040年投資約2,450兆韓元，係參考指引，而非承諾[S-134]
晶圓代工、封裝、測試	台積電（TSMC）資本支出（capex）600億至640億美元（G）[S-084]	台積電（TSMC）A14風險試產；亞利桑那州第2座晶圓廠量產，下半年（R）[S-084; S-083]；ASML低數值孔徑（low-NA）EUV與DUV浸潤式產能增加30%（C）[S-025]；Advantest至2027年3月達5,000台SoC測試機（T）[S-116]	台積電（TSMC）A14與N2U量產；14倍光罩尺寸CoWoS（R）[S-084; S-024]；Advantest至2028年底或2029年初達10,000台測試機（T）[S-116]	台積電（TSMC）A13、A12；CoWoS擴展至14個光罩以上；SoW-X（R）[S-024]	—	—
原始設計製造商（original design manufacturer, ODM）	鴻海（Foxconn）：Vera Rubin機櫃進入量產，第3季；AI機櫃出貨量成長逾一倍（R；G）[S-022]	緯穎（Wiwynn）德州Socorro擴產，逾16億美元，至2027年底提供近1,000個職位（C）[S-306]	—	—	—	—
光學與網路	鴻海（Foxconn）CPO交換器進入量產，第3季（R）[S-022]；台積電（TSMC）基板上COUPE進入量產（R）[S-024]	—	—	—	—	—
電力半導體	Infineon AI電力營收逾16億歐元，2026財年（G）[S-193]	Infineon AI資料中心營收25億歐元以上，2027財年（G）[S-193]；NVIDIA 800 VDC與Kyber同步全面量產，2025年5月所述（R）[S-028]	—	—	—	—
散熱	Vertiv淨銷售額138億至142億美元（G）[S-196]	—	—	—	—	—
渦輪機與電網設備	GE Vernova自第3季起燃氣渦輪機年產出20 GW；至年底已簽約至少125 GW（C；G）[S-012]	Siemens Energy每年約50部大型燃氣渦輪機，2027財年（C）[S-026]	GE Vernova年產出24 GW；至年底未出貨訂單（backlog）總額約2,000億美元（C；G）[S-012; S-039]	—	GE Vernova年產出30 GW，「implementing actions」（C）[S-012]	—

由企業自行揭露的數字化目標、量產里程碑與產能里程碑，依年份列示，資料截至2026年9月27日。字母表示陳述的類別：G為企業財測指引、T為管理層目標、R為藍圖日期、C為產能計畫或里程碑。不含分析師估計值。來源如各格所引註。

範圍 各項陳述在效力與定義上互有差異（企業財測指引、管理層目標、排定日期、產能計畫、投資指引），其財年慣例亦不相同；它們僅並列呈現，絕不相加或排名。

表20。 每一項發現均有一項具名的推翻訊號，以及一次註明日期的下次檢視

最可能提早修正的觀點為出口政策、記憶體配置與2027年資本支出（capex），三者的檢視時點均在2027年中之前。

論點	目前證據	確認訊號	否證訊號	下次檢視
1. 限制正從晶片擴及系統	記憶體與渦輪機供應商報告需求高於供給；Oracle表示其AI雲端訓練與推論服務的需求成長快於供給[S-003; S-026; S-049]	數個層級公布配置情形或產能售罄至2028年	記憶體、封裝與電力設備的交期（lead time）與未出貨訂單（backlog）同步下降	各供應商的下一次季度財報
2. 記憶體已成為預算上的限制	DRAM與NAND占主要雲端業者2026年資本支出（capex）的47%、2027年的68%，為TrendForce預測[S-007]；SK hynix與約10家客戶簽有長期協議（long-term agreement）[S-003]	記憶體占資本支出的比重上升；附預付款的多年期協議增加	伺服器DRAM合約價格連續兩季下跌	記憶體製造商的下一次季度財報
3. 電力與冷卻已屬架構的一部分	2026年53%的AI晶片採用液冷（liquid cooling），為TrendForce預測[S-006]；OCP的±400 VDC與固態變壓器（SST）規格（specification）[S-027; S-054]	數家供應商推出量產的800 VDC機櫃；取得OCP Accepted的CDU	800 VDC仍停留於參考設計（reference design）；認證（certification）延後	OCP全球峰會（OCP Global Summit），2026年10月12日至15日
4. 電網與電力設備決定時程	Siemens Energy的燃氣渦輪機已「booked out onto '28」[S-026]；TrendForce預測2028年後出現電網落差[S-043]	產能時段佇列變長；大型負載（large load）規則附帶負載削減（curtailment）義務	確定併接能以興建步調取得；渦輪機未出貨訂單（backlog）連續兩季下降	聯邦能源管理委員會（FERC）就大型負載（large load）採取的行動；渦輪機製造商財報
5. 客製晶片正與GPU並行成長	ASIC於2026年占AI伺服器的27.8%、至2030年近40%，為TrendForce預測[S-009]；Vera Rubin產能爬坡中[S-010]	客製計畫達到規模化量產	ASIC占比回落；相關計畫遭取消	TrendForce的下一次AI伺服器預測
6. 開放介面的作用在交接點上	ESUN有逾175家公司參與[S-020]；由四家公司提出的Open Data Center v0.7.0[S-001]	電力、冷卻與機櫃介面出現多供應商的OCP Accepted產品	各平台均未採用任何OCP設施介面	Moscone峰會，2027年10月4日至7日
7. 資本結構即基礎設施風險	國際能源總署（IEA）：投資規模過大，無法僅以資產負債表支應[S-005]；Oracle 2027財年約400億美元的融資[S-044]	債務、股權與預付款融資上升	資本支出成長以營運現金支應	各買方首次提出的2027年資本支出財測指引
8. 地緣政治已進入企業財測指引	NVIDIA 2027財年第3季的展望假設來自中國的Data Center運算營收為零[S-010]；AMD（2026年2月）與KLA（2025年10月）曾於較早的展望中載明中國的影響，惟其最新展望對此未置一詞[S-486; S-487; S-111; S-587]	管制趨嚴，或2026年11月的條款失效	以一般許可達成持久的貿易開放	2026年11月9日至10日
9. 台灣正轉向整合、封裝與光學	台積電（TSMC）14倍光罩尺寸CoWoS於2028年、COUPE於2026年[S-024]；鴻海（Foxconn）的CPO交換器[S-022]	封裝與光學里程碑如期達成	里程碑延後，或各項計畫仍停留於研發階段	台積電（TSMC）與鴻海（Foxconn）的季度財報
10. 研發與智慧財產權落在未解的邊界上	800 VDC固態斷路器仍在開發中[S-055]；友達（AUO）的Micro LED CPO處於研發階段[S-052]；據報導CPO測試方法尚不存在[S-286]	各項規格（specification）將保護、測試與控制留給實作者	各項規格（specification）將這些層級收攏完備，或這些層級在無新方法的情況下獲得解決	首批CPO規格（specification）預定於2026年第4季在OCP內部提交（據報導）[S-285]
11. 更精簡的模型使記憶體需求在各層級之間移轉	DeepSeek-V4.1-Flash將每token快取用量降至四分之一[S-529]；Google的月度token數一年內成長約七倍[S-491]；記憶體製造商提及代理式需求[S-003]	token成長超越快取節省幅度；用於快取的SSD與主機記憶體成長	在token用量仍成長之際，HBM或伺服器DRAM合約價格連續兩季下跌	TrendForce的季度記憶體價格更新

第1.3節的每一項發現各占一列。目前證據的資料截至2026年9月27日，來源列於該列之內；確認訊號與否證訊號為本報告的分析。

附錄A

技術規格狀態

規格（specification）的版本記錄的是其作者所達成的共識，而非買方所部署的內容；本附錄為本報告所依憑的每一份文件註明日期，並將其狀態與該技術所達到的階段並列呈現。

適用讀者： 撰寫需求的工程、採購（procurement）與標準人員；追蹤哪些介面正趨於共通的研發與智慧財產權讀者。

本報告引用數十份規格（specification），自機櫃圖紙至記憶體標準皆有，而每一份都可能遭兩種誤讀：將草案當作已完成的標準，或將已發布的標準當作已部署的標準。本附錄列出截至2026年9月27日各技術章節所依據的文件，並載明其文件本身頁面所印的版本與日期，另於每一份旁側列出所引紀錄就其所述內容顯示的最遠階段。本附錄所回答的問題相當務實：2026年的一份需求可以引用哪些文件，又有哪些文件所述的硬體是買方可以下單採購的？

F OCP為其文件設有數種承諾層級。 依其2026年8月14日修訂的智慧財產權管理政策（Intellectual Property Rights Management Policy），最終規格（Final Specification）係指經指導委員會核准、並連同其授權（licence）聲明、貢獻者及任何專利排除通知（Exclusion Notice）一併發布的提案規格[S-047]。仍由孵育委員會審議中的提交案列為審查中，並「available for public comment」[S-202]。產品認可則屬另一道程序：產品若符合一份經核准的OCP規格或其他貢獻文件（contribution），即可獲認可為OCP Accepted[S-333]；OCP並於2026年5月11日公告，第二種標章OCP Inspired將於2026年7月31日之後退場[S-109]。

A 發布並非採用。 版本編號記錄的是一份文件的作者在某一日期所達成的共識；它並不顯示產品是否符合該文件、供應商之間是否互通，或買方是否已部署其成果。兩者可能朝任一方向出現落差。Open Rack Wide（ORW）於2026年4月達到1.0.0版，依該規格（specification）建置的機櫃已列入名錄，而建置於其上的AMD Helios設計亦正在部署[S-233; S-179; S-111]。NVIDIA的Vera Rubin平台正「ramping into full production」[S-010]，但NVIDIA表示計畫貢獻給OCP的機櫃設計並無接受紀錄[S-178]。因此，各表將狀態與階段分置於不同欄位。

> **狀態欄與階段欄的閱讀方式**
>
> **草案。** 版本低於1.0、屬發布候選版（RC）或處於審查中：可作為設計參考加以引用，但不可作為已發布的需求。
>
> **已發布、已批准或已公布。** 經該機構核准；部分聯盟僅向會員發布（參見第A.2節的範圍註記）。
>
> **白皮書（white paper）或指引。** 屬實務上的指導，而非供產品據以測試的規格（specification）[S-262]。
>
> **設計貢獻文件（contribution）。** 係企業提供給OCP的設計，而非附版本編號的規格（specification）[S-237]。
>
> **階段。** 產品：已有實作列入名錄、正在出貨或已進入量產。試點階段、研發：僅屬試行或開發。標準：文件已存在，但所引來源並未顯示有任何實作（並非認定不存在實作）。已宣布：已規劃，但尚未發布。

A.1 OCP的機櫃與冷卻迴路文件已發布；其電力路徑仍處於草案階段

表21列出本報告所依憑的OCP文件，依自電力至設施的七個領域排列。版本與日期均依各文件自身所印方式呈現，因此編號樣式互有差異（「Rev 1.0」、「V1.0.0」、「r2.0 v1.0」）；破折號表示該文件未印載此項。

A 機櫃與冷卻迴路已有發布的文件；電力路徑則沒有。 ORW、Deschutes CDU與水基冷卻液規格（specification）均於2026年2月至8月間達到發布狀態[S-228; S-030; S-217]。800 VDC路徑所依憑的文件均未達1.0版——Diablo 400為0.7.0版、SST規格為0.3.0首份草案、Open Data Center設計原則為0.7.0版——另有一份低壓直流（low-voltage direct current, LVDC）白皮書（white paper）[S-027; S-054; S-001; S-059]。2026年的標案可將第一組引為已發布的需求，而電力路徑在其文件之一達到1.0版之前，僅能作為設計參考加以引用（第4.2節）。

A 在堆疊的最上層，產品的進展領先於文件。 Meta於2024年10月隨其Catalina系統推出ORv3 High Power Rack[S-176]，但HPR規格（specification）的文件迄仍列為審查中[S-202]。NVIDIA的GB200 NVL72機櫃於2024年10月以設計貢獻文件（contribution）的形式進入OCP，而非一份附版本編號的規格[S-237]。開放文件也可能記錄已在運作中的設計：OpenAI表示，MRC——其OCP規格所載的發布日期為2026年3月21日[S-260]——運行於其所有最大型的NVIDIA GB200超級電腦上[S-259]。在這些情形中，開放文件是隨產品而來：它協助周邊的供應商在電力、冷卻與機械介面上實現互通[S-238]，而非先定義一項產品，再由多家廠商據以建造。

? 未解問題——共同封裝光學（co-packaged optics, CPO）尚無共通的測試與封裝基礎。 依OCP APAC Summit 2026的報導，日月光（ASE）一位高階主管表示，產業尚未建立共通的測試標準；台積電（TSMC）一位高階主管則將晶圓級測試、光纖陣列整合與光學封裝組裝列為尚未集體解決的難題；據報導，一個由20家企業組成的OCP聯盟預期於2026年第4季提交其首批CPO規格（specification）[S-286; S-285]。在此類文件出現之前，CPO只能逐一廠商進行驗證（qualification），使採購維持單一來源（single-source）（第4.7節）。

IF 唯有在電力相關文件達到發布狀態時，800 VDC才能寫入需求，而非僅存於廠商參考設計（reference design）之中。 應觀察Mt Diablo 2.0的發布——其作者表示，支援原生800 V的版本「will be coming later」[S-019]；0.3.0草案之後的SST規格（specification）[S-054]；以及2026年9月24日啟動的UL 2877計畫下首批取得認證（certification）的機組[S-188]。若這些文件至2027年仍停留在1.0版以下，而部署卻依單一廠商設計持續進行，則此一觀點不成立：屆時廠商參考架構將成為實際適用的需求。

> **觀察重點**
>
> **下次檢視時點**
>
> OCP全球峰會（OCP Global Summit），聖荷西，2026年10月12日至15日[S-067]：Mt Diablo 2.0的發布、首批CPO規格（specification）的提交，或Vera Rubin機櫃貢獻文件（contribution）獲接受，均會改變表21中的若干列。

A.2 OCP所倚賴的介面由其他機構擁有

A OCP的文件倚賴由其他機構掌握的介面。 ESUN採納Ultra Ethernet Consortium（UEC）所定義的鏈路層重傳（Link Level Retry）與基於信用額度的流量控制（credit-based flow control）[S-246]，UALink的Manageability 1.0採用DMTF的Redfish[S-251]，FCSA則將小晶片（chiplet）介面對應至UCIe、CXL與PCIe等標準[S-166]。因此，一項OCP機櫃或網路架構需求的確定程度，僅等同於其所引用的外部標準。表22依各文件所定義的介面列示，並採用相同欄位（第6.3節）。

A AMD的Helios展示了外部介面如何在同一座機櫃內結合。 其參考設計（reference design）以UALink over Ethernet運行垂直擴展網路（scale-up network），並以UEC-ready的Pensando網路介面卡運行水平擴展網路（scale-out network）[S-179]；AMD表示，Helios正由AI實驗室與雲端服務供應商部署[S-111]。「UEC-ready」是AMD自身的描述，而非符合性紀錄，因此UEC仍停留在標準階段。在所閱讀的來源中，UALink自有的200G實體層與鏈路層尚未出現於任何出貨系統，惟一款支援UALink、ESUN與乙太網路的重定時器（retimer）自2026年1月起已開始送樣[S-253]。若出現可互通的獨立UALink交換器與端點，即顯示垂直擴展之爭正從文件轉向產品（第4.6節）。

? 未解問題——HBM4E的JEDEC標準。 JEDEC於2025年4月發布HBM4，比Samsung宣布HBM4商業出貨早十個月[S-139; S-056]，但所閱讀的來源並未記錄任何JEDEC HBM4E標準，而Micron則預期HBM4E於2027年曆年進入量產[S-128]。這對2027年的加速器（accelerator）設計具重要性：在JEDEC標準發布之前，各供應商的HBM4E只能依該供應商自有的規格（specification）進行驗證（qualification），使增添第二供應來源（second source）更為困難（第3.2節）。

? 未解問題——800 VDC的已認證保護與轉換設備。 UL於2025年10月發布適用於固態斷路器（solid-state circuit breaker）的UL 489I[S-187]，並於2026年9月24日依UL 2877為中壓轉800 VDC電源機組開設認證（certification）計畫[S-188]；所閱讀的來源並未記錄任何依該計畫取得認證的機組。這之所以重要，是因為800 VDC唯有在具管轄權的主管機關與保險人能夠接受已認證設備時，才會擴展至參考設計（reference design）之外（第7.4節）。

> **範圍 聯盟規格**
>
> 發布日期並不等於公開可取得。PCI-SIG將其規格（specification）發布給會員[S-349]，而CXL的公開版本僅授權（licence）用於評估，不含實作授權[S-399]。實作權利取決於會員身分與各機構的專利政策（第6.6節）。

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-010 NVIDIA (26 Aug 2026); S-056 Samsung Electronics (12 Feb 2026); S-111 AMD Investor Relations (4 Aug 2026); S-128 Micron Technology (24 Jun 2026); S-176 Engineering at Meta (15 Oct 2024); S-178 NVIDIA Blog (13 Oct 2025); S-179 AMD (未註明日期); S-188 UL Solutions (24 Sep 2026); S-224 AMD (未註明日期); S-237 NVIDIA Newsroom (15 Oct 2024); S-253 Credo Technology Group (29 Jan 2026); S-259 OpenAI (5 May 2026)

標準組織與聯盟. S-001 OCP (27 Aug 2026); S-019 OCP (11 Aug 2026); S-020 OCP (10 Mar 2026); S-027 OCP (1 Mar 2026); S-030 OCP (26 Feb 2026); S-047 OCP (14 Aug 2026); S-054 OCP (22 Jun 2026); S-058 JEDEC Solid State Technology Association (Dec 2025); S-059 OCP (2 Sep 2026); S-060 OCP (6 Aug 2026); S-062 OCP (29 Apr 2026); S-067 OCP (未註明日期); S-109 OCP (11 May 2026); S-139 JEDEC (16 Apr 2025); S-141 JEDEC (13 Jul 2026); S-158 CXL Consortium (18 Nov 2025); S-160 OCP (22 Aug 2026); S-161 OCP (19 Dec 2025); S-165 OCP (12 Feb 2026); S-166 OCP (8 Apr 2026); S-167 UCIe Consortium (5 Aug 2025); S-168 Ultra Accelerator Link Consortium (7 Apr 2026); S-180 OCP (22 Sep 2022); S-183 OCP (14 Sep 2023); S-187 ANSI / UL (22 Oct 2025); S-202 OCP (未註明日期); S-205 OCP (未註明日期); S-207 OCP (13 Oct 2025); S-208 OCP (未註明日期); S-212 OCP (4 Sep 2020); S-215 OCP (4 Sep 2020); S-216 OCP (未註明日期); S-217 OCP (6 Aug

表21。 OCP的機櫃與冷卻迴路文件已達1.0版；800 VDC電力路徑則仍依憑草案與白皮書

2026年的一份需求可將ORW、Deschutes CDU與冷卻液規格（specification）引為已發布的文件，但800 VDC電力路徑僅能作為設計參考加以引用。

領域	文件	版本與日期	狀態	階段	來源
電力	Diablo 400: Rack and Power（Mt Diablo，±400 VDC側置電源機櫃）	0.7.0，2026年3月1日生效	基礎規格（specification）尚未達1.0；側置電源機櫃（sidecar）自800 kW至1 MW以上	標準	[S-027]
電力	Mt Diablo 2.0（原生800 V）	已規劃；無日期	未發布	已宣布	[S-019]
電力	Solid State Transformer (SST) Specification：中壓至800 VDC	0.3.0修訂版，2026年6月22日生效	首份草案	標準	[S-054]
電力	Data Center Facility – Low Voltage Direct Current Power Distribution	白皮書v1.1.0，2026年9月2日	白皮書；低壓定義為最高1,500 VDC	標準	[S-059]
電力	Requirements for Energy Storage Systems Used in Data Centers	1.0修訂版、1.0版，2026年1月	白皮書	標準	[S-205]
散熱	Project Deschutes: Data Center Facilities（冷卻液分配單元，CDU）	1.0.0，2026年2月26日生效	已發布；約2 MW、500 GPM，壓力80至90 psi	產品	[S-030; S-207; S-208]
散熱	Universal Quick Disconnect，手動對接（UQD）與盲插對接（UQDB）	1.0修訂版，2020年9月4日（UQD檔案與wiki：1.0.2）	已發布；最大操作壓力100 psi	標準	[S-212; S-215]
散熱	UQD v2，手動對接與盲插對接	v2.X.4 beta版草案	草案	標準	[S-060; S-212]
散熱	Base Specification for Single Phase Water-Based Cold Plate Coolants	1.3版，2026年8月6日生效	已發布；pH 8.0至10.5，導電度低於2,000 µS/cm	標準	[S-217]
散熱	單相冷板機櫃用丙二醇基流體（PG25）指引	定稿；未註明日期	指引	標準	[S-216]
散熱	兩相冷板機櫃用介電流體指引	2024年7月	指引；兩相冷卻（two-phase cooling）展示排定於2026年峰會	標準	[S-060; S-220]
散熱	Immersion Requirements（浸沒式冷卻需求）	Rev 2.10，2023年8月18日	已發布需求	標準	[S-221]
機櫃	Open Rack V3 (ORv3) Base Specification	1.1，2023年12月（1.0版：2022年9月22日）	已發布；電源架標稱輸出51 VDC，資訊科技設備為46至52 V	產品	[S-202; S-180; S-176]
機櫃	ORv3 High Power Rack (HPR)	規格資料夾，2024年10月與2025年4月	審查中	產品	[S-202; S-176]
機櫃	Open Rack Wide (ORW) Base Specification	V1.0.0，2026年4月28日	已發布	產品	[S-228; S-233]
機櫃	ORW Meta Design Specification（設計規格）	V1.0.1，2026年5月5日生效	已發布	產品	[S-234; S-179; S-111]
機櫃	NVIDIA GB200 NVL72機櫃與機箱（MGX）	於2024年10月15日貢獻	設計貢獻文件（contribution），而非附版本編號的規格（specification）	產品	[S-237; S-176]
機櫃	NVIDIA Vera Rubin機櫃與運算機箱	2025年10月13日表明貢獻計畫	未有接受紀錄	產品	[S-178; S-010]
系統	DC-MHS模組化硬體系列	R1 v1.x；M-CRPS 1.06，2026年4月	R1 v1.0於2022年11月核准；最新版本的最終協議尚未簽署	標準	[S-240]
系統	DC-SCM（Datacenter-ready Secure Control Module）	Rev 2.2 Ver 1.0，於2025年10月16日提交	最新公開修訂版；3.0工作流（workstream）正在進行	標準	[S-242; S-368]
系統	OAI-OAM Base Specification（加速器模組）	r2.0 v1.0，2023年9月14日	已發布；在44至59.5 V下最高1,000 W	產品	[S-183; S-224]
系統	Foundation Chiplet System Architecture (FCSA)	1.0.0，2026年2月12日發布	已發布；2026年4月核准之貢獻文件（contribution）	標準	[S-165; S-062]
系統	CMS: Memory Fabric Orchestration Architecture	白皮書v1.0，2025年12月19日	白皮書	標準	[S-161]
系統	Hyperscale CXL Tiered Memory Expander Base Specification	–	審查中草案	標準	[S-160]
網路	ESUN Network Operator Requirements Base Specification	Rev 1.0，2026年2月9日生效	2026年3月10日公告發布；參與企業逾175家	標準	[S-246; S-020]
網路	Scale Up Ethernet (SUE)	1.0.0，2025年9月5日核准	已發布；相關工作以SUE-T延續	標準	[S-247; S-249]
網路	Multipath Reliable Connection (MRC) Specification	1.0修訂版，2026年3月21日	已發布之貢獻文件（contribution）	產品	[S-260; S-259]
網路	Optical Circuit Switching for AI and Machine Learning Infrastructure	2026年4月	白皮書，而非規格（specification）；Google自有的OCS自2020年起已部署	研發	[S-262; S-293]
網路	共同封裝光學（co-packaged optics, CPO）規格（specification）	據報導，首批規格預定於2026年第4季提交	未發布	已宣布	[S-285]
安全	Caliptra Root of Trust 2.0 Modular Base Specification	v2.0，2025年7月31日；2.1硬體（RTL）發布，2025年10月15日	已發布	標準	[S-364; S-365]
安全	OCP L.O.C.K.（自加密硬碟的金鑰管理）	1.1 RC4，2026年6月3日生效	發布候選版	標準	[S-366]
可管理性	OBMF Interface Consolidation Protocol over USB（經USB的介面整合協定）	1.0.0，2026年4月10日	已發布	標準	[S-463]
可管理性	RAS API Specification	0.9版第1修訂版，2025年12月14日	中途發布版	標準	[S-372]
可管理性	GPU and Accelerator RAS Requirements	V1.7	已發布需求	標準	[S-371]

本章來源 （續）

2026); S-220 OCP (22 Sep 2026); S-221 OCP (18 Aug 2023); S-227 OCP (Apr 2026); S-228 OCP (28 Apr 2026); S-233 OCP (未註明日期); S-234 OCP (5 May 2026); S-238 OCP (15 Oct 2024); S-240 OCP (24 Apr 2026); S-242 OCP (27 Jul 2022); S-245 UALink Consortium (8 Apr 2025); S-246 OCP (9 Feb 2026); S-247 OCP (5 Sep 2025); S-249 OCP (13 Jun 2026); S-251 UALink Consortium (7 Apr 2026); S-256 Ultra Ethernet Consortium (11 Jun 2025); S-257 Ultra Ethernet Consortium (16 Jul 2026); S-260 OpenAI / OCP (5 May 2026); S-262 OCP (Apr 2026); S-263 IEEE 802.3 Working Group (6 Aug 2026); S-270 IEEE 802.3 Ethernet Working Group (3 Aug 2026); S-284 OIF (29 Mar 2023); S-333 OCP (未註明日期); S-349 PCI-SIG (11 Jun 2025); S-350 PCI-SIG (5 Aug 2025); S-356 DMTF (2 Jul 2026); S-357 DMTF (24 Sep 2026); S-364 OCP (31 Jul 2025); S-365 CHIPS Alliance (15 Oct 2025); S-366 OCP (3 Jun 2026); S-368 OCP (15 Jul 2026); S-370 DMTF (25 May 2025); S-371 OCP (未註明日期); S-372 OCP (14 Dec 2025); S-399 CXL Consortium (12 Feb 2026); S-463 OCP (10 Apr 2026); S-464 Optical Internetworking Forum (未註明日期); S-465 DMTF Redfish Forum (20 May 2026); S-466 OpenBMC (未註明日期)

研究機構. S-293 Jouppi et al. (Apr 2023)

媒體與法律評論. S-271 Converge Digest (26 Aug 2026); S-285 Tech Times (15 Aug 2026); S-286 Tech Times (12 Aug 2026)

領域	文件	版本與日期	狀態	階段	來源
設施	Open Data Center: Next-Generation ML Infrastructure Design Principles	0.7.0修訂版，2026年8月27日生效	設計規格（specification）尚未達1.0	標準	[S-001]
設施	Open Data Center for AI（倡議白皮書）	0.5版，2026年3月18日（封面標示2026年4月）	白皮書	標準	[S-227]
設施	Third Party Integration, Telemetry & APIs	2026年4月核准	已核准之貢獻文件（contribution）	標準	[S-062]

與AI基礎設施相關的OCP規格（specification）、白皮書（white paper）與設計貢獻文件（contribution），依領域列示，並載明各文件所印的版本與日期及其截至2026年9月27日的狀態。階段標籤的定義見上方註記。來源見最後一欄。

「標準」係指所引來源並未顯示該文件有任何實作；並非表示不存在實作。企業關於部署的主張，均為該等企業自身的主張。

表22。 介面速率的規格制定領先於出貨系統；800 VDC設備的認證才剛起步

對多數外部介面而言，下一份文件已經寫成，因此買方具約束力的環節是平台驗證（qualification）與安全認證（certification），而非標準化。

領域	文件	版本與日期	狀態	階段	來源
垂直擴展	UALink 200G 1.0 Specification（UALink Consortium）	2025年4月8日	已發布；一個pod內最多1,024顆加速器（accelerator）	標準	[S-245]
垂直擴展	UALink Common Specification 2.0；200G Data Link and Physical Layers 2.0；Manageability 1.0；Chiplet 1.0	2026年4月7日批准；規格（specification）頁面列為Chiplet 1.01	已批准	標準	[S-251; S-168]
水平擴展	Ultra Ethernet Specification（UEC）	1.0版，2025年6月11日；現行1.0.3版，2026年7月16日	已發布；期間另有兩次維護版發布；AMD稱其Helios網路介面卡為UEC-ready	標準	[S-256; S-257; S-179]
水平擴展	IEEE P802.3dj（200 Gb/s、400 Gb/s、800 Gb/s與1.6 Tb/s乙太網路）	任務小組；工作小組頁面於2026年8月6日更新	草案；尚非已核准的標準	標準	[S-263]
水平擴展	IEEE 802.3每通道400 Gb/s訊號研究小組	2026年3月13日獲准成立；草案PAR與CSD於2026年7月17日核准	研究小組階段，尚未成立任務小組	研發	[S-270]
光學	OIF關於3.2 Tb/s共同封裝（CPO）模組的實作協議	2023年3月29日	實作協議	標準	[S-284]
電氣I/O	OIF CEI-448G（每通道448 Gb/s）	自2024年起進行框架工作；據報導，多廠商示範排定於2026年9月21日至23日的ECOC	探索階段；尚無實作協議	研發	[S-464; S-271]
記憶體	JEDEC HBM4（JESD270-4）	2025年4月16日；JESD270-4A（1.1版），2025年12月	已發布標準；每個堆疊最高2 TB/s	產品	[S-139; S-058; S-056]
記憶體	JEDEC SPHBM4（JESD330-4）：有機基板上的HBM4晶粒	2026年7月13日	已發布標準	標準	[S-141]
主機I/O	PCIe 7.0（PCI-SIG）	2025年6月11日	已向會員發布；128.0 GT/s	標準	[S-349]
主機I/O	PCIe 8.0（PCI-SIG）	2025年8月5日宣布；計畫於2028年前發布	開發中；256.0 GT/s	已宣布	[S-350]
記憶體擴充	CXL 4.0（CXL Consortium）	2025年11月18日	已發布；128 GT/s，綁定式連接埠	標準	[S-158]
晶粒間	UCIe 3.0（UCIe Consortium）	2025年8月5日	已發布；48與64 GT/s	標準	[S-167]
可信證明（attestation）	SPDM，DSP0274（DMTF）	1.4.0，2025年5月25日；1.4.1，2026年7月2日	DMTF標準；後量子演算法，無混合模式	標準	[S-370; S-356]
可管理性	Redfish（DMTF）	2026.1版，規格（specification）v1.24.0；2026.2版於2026年9月15日宣布	已發布	產品	[S-465; S-357; S-466]
安全	UL 489I《固態斷路器》第1版	2025年10月22日	已發布；額定至1,000 VAC與1,500 VDC	標準	[S-187]
安全	UL 2877認證（certification）計畫：中壓轉800 VDC電源機組	2026年9月24日啟動	計畫已開放；無已取得認證的機組紀錄	標準	[S-188]

本報告所倚賴的其他標準組織與聯盟的文件，並載明各文件所印的版本與日期及其截至2026年9月27日的狀態。階段標籤同表21。來源見最後一欄。

AMD Helios設計所採用的UALink over Ethernet（UALoE），係在乙太網路鏈路上承載UALink；此處並未將其列為UALink自有200G實體層與鏈路層的實作。

附錄B

企業與區域圖譜

沒有任何企業或區域掌握AI基礎設施的每一個層級。本圖譜將對本報告論證具重要性的企業——自加速器（accelerator）與記憶體，至電力、冷卻、光學與渦輪機——置於其公開紀錄所能支持的階段上。

適用讀者： 將企業定位於技術堆疊上的產業、採購（procurement）、財務、政府與智慧財產權讀者；階段標籤見第B.1節，依層級與區域劃分的表格見第B.2節。

本報告各章論證，AI基礎設施的約束正自加速器（accelerator）向外移往記憶體、封裝、電力、冷卻、光學與電網。本附錄說明這些層級各由誰掌握，以及各企業的角色在公開紀錄上推進到何種程度：已出貨、處於試點階段、開發中、已寫入規格（specification），或僅止於宣布。這是一份經篩選的地圖，而非名錄。企業唯有在有註明日期的公開來源記載其角色時才會出現，且任何標籤的強度均不超過該來源。本附錄所回答的問題相當實際：在每一層級上，今日誰在交付，而誰僅僅表明了計畫？

B.1 納入對象，以及階段標籤的意義

單憑一份企業名稱清單，所能說明的極為有限。讀者需要知道某家企業為何對整個技術堆疊具重要性，以及其角色推進到何種程度：一家量產出貨的記憶體製造商，與一家開發光學模組的顯示器製造商，兩者都具相關性，但相關的方式並不相同。因此本圖譜對每一列均適用同一項納入規則與同一套五階段尺度，並在每一標籤旁印出證據及其日期，使該標籤得以查核，而非僅能信賴。當有註明日期的公開來源記載下列五種角色之一時，一家企業或研究機構即納入本圖譜：

- 其向技術堆疊供應重要的技術或產品：處理器、記憶體、封裝體、機櫃、電力或冷卻系統、光學元件或渦輪機。
- 其具有已記載的設計採用、生產角色、驗證（qualification）、標準化角色或重大已宣布計畫。
- 其掌握重要的瓶頸或致能投入要素，例如高頻寬記憶體（high-bandwidth memory, HBM）、先進封裝（advanced packaging）或測試產能、微影、渦輪機產能時段或資本。
- 其具有與特定層級綁定且可信的研發（research and development, R&D）路徑。
- 其為說明某一區域對供應鏈其餘環節的依賴所必需。

納入是一份地圖，而非排名。出現於本圖譜之中，並不表示該企業為Open Compute Project（OCP）會員、OCP認可的供應商、峰會講者，或其所在表格中其他企業的直接競爭者，亦不代表任何投資觀點。本圖譜刻意採取選擇性原則：保留其角色與本報告關於記憶體、封裝、電力、冷卻、光學、系統整合與發電之論證相關的企業，並排除其角色無法與註明日期的公開來源相連結的企業。因此未列入並不代表該企業不重要；僅表示公開紀錄未在這些表格所需的層次上記載其角色。

A 標籤依循企業自身所用的動詞。 「Expects」、「plans」與「scheduled」會使一項角色維持在「已宣布」，直至該企業報告生產或出貨為止：SK hynix的HBM4計為「產品」，因為該公司報告了量產出貨；而其印第安納州晶圓廠在生產開始之前仍維持「已宣布」[S-003; S-023]。展示或展出無論看來多先進，均維持在「研發」，這也是友達（AUO）的模組雖以具名的AI資料中心用途為目標，仍列為研發的原因[S-052]。一家企業可在不同層級處於不同階段：Vertiv目前出貨電力與熱管理設備，但將其800 VDC產品組合描述為計畫於2026年下半年推出[S-307; S-189]。各列所記錄的證據截至2026年9月27日。

範圍 圖譜各列

國別為總部所在地。各項數字均為各企業自有的衡量方式與期間：營收、未出貨訂單（backlog）、訂單、產出與產能屬不同概念，各列之間不可比較。前瞻性陳述保留各企業自身的用語，無論其為預期、計畫、目標或排定。

B.2 稀缺的層級集中於少數企業與少數地點

A 本圖譜依部署建置的順序閱讀。 其先自決定數量與日期的平台供應商與買方開始（表24），再依序呈現其所依賴的投入要素：記憶體（表25）；晶圓代工、封裝、微影與測試（表26）；機櫃整合（表27）；電力與冷卻（表28）；發電（表29）；以及光學（表30）。其中有兩種模式反覆出現。本報告視為具約束力的各層級僅由少數供應商掌握，而這些供應商均表示供應緊俏，或產能時段已預訂至數年之後[S-003; S-018]。而最新出現的各項角色，自共同封裝光學（co-packaged optics）至800 VDC保護，多數仍處於試點階段、研發或已宣布[S-055; S-052]。

依區域研讀本圖譜

A 美國掌握平台、買方與大部分已表明的需求。 表24十一列中有十列為美國企業，其計畫決定了其他區域必須供應的數量：Alphabet已將其2026年資本支出（capex）區間上調至1,950億至2,050億美元，Amazon則預期現金資本支出約2,200億美元[S-016; S-070]。在堆疊更下游處，位於美國的最稀缺層級產能屬排定中而非已在運行：台積電（TSMC）的亞利桑那第2座晶圓廠於2027年下半年開始量產，SK hynix的印第安納州HBM晶圓廠於2029年下半年開始量產，Amkor的亞利桑那第2期則預定於2029年底前完工[S-083; S-023; S-317]。在這些場址投入運行之前，美國買方在尖端晶圓、CoWoS封裝，以及表25與表26中三家HBM來源的其中兩家上，仍須倚賴總部位於台灣與韓國的供應商。

A 台灣的各列從晶圓延伸到機櫃，而其最新的角色仍處於早期階段。 台灣在台積電（TSMC）握有尖端晶圓代工與CoWoS封裝，在鴻海（Foxconn）、緯穎與緯創握有機櫃整合，在台達電子（Delta）與光寶（LITE-ON）握有機櫃電力轉換。本報告所描述的轉變，即自「組裝」這一簡略代稱轉向整合、封裝、光學與驗證（qualification），確已可見，但各項所處階段並不一致：鴻海的共同封裝光學（co-packaged optics）交換器量產「on track」，聯電（UMC）的矽光子（silicon photonics）風險性試產預定於2026年與2027年，友達（AUO）的Micro LED模組則仍在開發中[S-022; S-318; S-052]。鴻海40 MW的高雄一號資料中心於2025年10月被描述為依800 VDC建置[S-178]。台灣的組裝作業也正移往美國場址，緯創的沃斯堡廠房已啟用，緯穎則與Amazon在德州擴產[S-475; S-306]。

A 韓國聚焦於記憶體；日本則供應設備、測試、儲存與渦輪機。 本圖譜中三家HBM供應商有兩家為韓國企業，且兩者均表示供應緊俏：SK hynix表示客戶需求超出供應能力，Samsung則預估記憶體市場在2026年下半年仍將供不應求[S-003; S-127]。韓國的另一列為散熱，其中LG Electronics 2.6 MW的冷卻液分配單元（coolant distribution unit, CDU）已通過NVIDIA DSX Ready計畫的驗證（qualification）[S-209]。日本的各列位於其他區域產能的上游：Advantest表示推論AI用ASIC、CPU與DRAM的測試機需求成長快於其預期，Tokyo Electron表示供AI資料中心使用的晶圓製造設備占比正成長至超過總量的一半，Mitsubishi Heavy Industries則訂下目標，要使2028財年燃氣渦輪機複循環機組出貨量高於現有產能30%以上[S-031; S-172; S-479]。

A 歐洲的籌碼（leverage）位於上游，且依最長的時鐘運行。 歐洲各列涵蓋ASML的微影、Besi的混合鍵合（hybrid bonding）、Infineon與STMicroelectronics的功率半導體、Schneider Electric與ABB的電氣設備，以及Siemens Energy的燃氣渦輪機；無一位於加速器（accelerator）、HBM或機櫃整合層級。其中數項角色依整個堆疊中最慢的時鐘運行：ASML計劃為2027年將其低數值孔徑（low-NA）EUV產能增加30%，Siemens Energy則表示其燃氣渦輪機產能已售罄至2028財年[S-025; S-018]。Infineon的客戶正透過多年期協議預訂功率半導體產能[S-021]。因此，歐洲對AI需求的曝險雖大，卻屬間接；應觀察的訊號是訂單簿與產能預訂，而非產品推出。

A 中國的各列形成一套處於不同階段的平行堆疊。 Huawei的Ascend 950DT排定於2026年第4季，SMIC則提供國內晶圓代工產能[S-312; S-473]。在記憶體方面，依Tom's Hardware報導，CXMT生產DRAM，其公開說明書未載有專門的HBM計畫；YMTC則生產3D NAND[S-321; S-470]。中國在本圖譜中最強的位置是光學製造：TrendForce估計，中國製造商2026年將握有全球光學模組委外製造產能的約56%[S-277]。出口管制（export control）決定這些列中哪些能夠服務哪些市場，相關內容見第7章；本圖譜記錄的是角色，而非市場進入。以色列則透過兩列出現：正與Infineon共同開發800 VDC固態斷路器（solid-state circuit breaker）的SolarEdge[S-055]，以及Tower Semiconductor的矽光子（silicon photonics）晶圓代工[S-292]。

觀察重點

應優先重新檢視的各列

最接近自「已宣布」轉為「產品」的，是鴻海（Foxconn）的共同封裝光學（co-packaged optics）交換器，其2026年第3季量產「on track」；以及Vertiv的800 VDC產品組合，計畫於2026年下半年（2H）推出[S-022; S-189]。

友達（AUO）唯有在取得客戶驗證（qualification）、達到量產規模，或進入具名平台時，才會跨出研發階段[S-052]。

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-002 NVIDIA (26 Aug 2026); S-003 SK hynix (29 Jul 2026); S-010 NVIDIA (26 Aug 2026); S-011 CoreWeave (11 Aug 2026); S-012 GE Vernova (22 Jul 2026); S-016 Alphabet (22 Jul 2026); S-018 Siemens Energy (30 Mar 2026); S-021 Infineon Technologies (5 Aug 2026); S-022 Hon Hai (12 Aug 2026); S-023 SK hynix (28 Aug 2026); S-024 TSMC (23 Apr 2026); S-025 ASML (15 Jul 2026); S-028 NVIDIA Technical Blog (20 May 2025); S-031 Advantest (29 Jul 2026); S-048 NVIDIA (21 Sep 2026); S-049 Oracle (10 Sep 2026); S-052 AUO (31 Aug 2026); S-055 SolarEdge and Infineon (9 Sep 2026); S-056 Samsung Electronics (12 Feb 2026); S-057 Besi (18 Jun 2026); S-063 Wiwynn (7 Aug 2026); S-064 Wiwynn (7 May 2026); S-073 Meta Platforms (29 Jul 2026); S-075 Microsoft (29 Jul 2026); S-082 TSMC (15 Jan 2026); S-083 TSMC (16 Apr 2026); S-111 AMD Investor Relations (4 Aug 2026); S-112 AMD (24 Feb 2026); S-113 Broadcom (2 Sep 2026); S-115 Samsung and Broadcom (25 Jul 2026); S-126 SK hynix (8 Jun 2026); S-127 Samsung Electronics (30 Jul 2026); S-128 Micron Technology (24 Jun 2026); S-129 Micron

表23。 五個階段標籤區分交付與意圖

階段標籤記錄的是已有證據的角色推進到何種程度；計畫、樣品或展品絕不標示為生產。

標籤	在本圖譜中的意義	不代表的意義	本圖譜中的實例
產品	該企業就所指角色載明生產、出貨、客戶或營收；若為買方或營運業者，則為已投入服務的容量	並非關於市場占有率、在所有客戶處均完成驗證（qualification）或技術領先地位的主張	SK hynix於2026年第2季開始HBM4量產出貨[S-003]
試點階段	樣品、工程樣機、驗證（qualification）、風險性試產、首批訂單或首次有限部署	並非量產或經常性營收	Samsung表示已向主要客戶出貨HBM4E樣品（2026年7月）[S-127]
研發	開發中、已展示（demonstrated）或已展出，但未載明向客戶出貨	並非供應關係、設計採用、生產或營收	友達（AUO）的Micro LED共同封裝光學（co-packaged optics）模組，開發中，並宣布於2026年9月展出[S-052]
標準	已有證據的角色為一份已發布的規格（specification），或對開放硬體或標準機構的貢獻文件（contribution）	發布不等於採用；貢獻文件（contribution）不等於產品	固態變壓器（solid-state transformer）規格（specification）v0.3，由Microsoft、Google、NVIDIA等共同制定[S-019]
已宣布	已載明但尚未交付的計畫、投資、產能承諾、協議或藍圖日期	並非結果；計畫可能延後、縮減或取消	SK hynix的印第安納州HBM晶圓廠，量產排定於2029年下半年（2H）[S-023]

適用於本圖譜每一列的定義，並就每一標籤附一項實例及其來源。標籤附著於該列所指的角色，而非整家企業。
若一家企業所持角色處於不同階段，階段欄會逐一列明，例如「產品；800 VDC：已宣布」（800 VDC即800伏特直流配電）。

表24。 平台供應商與買方決定了技術堆疊其餘環節必須達到的數量與日期

少數平台供應商與買方決定了步調；其前瞻數字屬財測指引、計畫與藍圖日期，而非已交付的產能。

企業	國別	層級	在AI基礎設施中的角色	階段	關鍵證據與日期	來源
NVIDIA	美國	加速器（accelerator）、機櫃、網路	GPU平台、機櫃級系統（rack-scale system）與交換；800 VDC參考架構	產品；800 VDC：已宣布	2027財年第2季（至2026年7月26日）Data Center營收890億美元；Vera Rubin「ramping into full production」（2026年8月）；800 VDC全面量產將與2027年的Kyber機櫃同步（藍圖，2025年5月）	[S-002; S-010; S-028]
AMD	美國	加速器（accelerator）、機櫃	Instinct GPU與EPYC CPU；Helios機櫃級系統（rack-scale system），與Meta透過OCP共同開發	產品	2026年第2季Data Center營收67億美元，成長107%；表示Helios正由AI實驗室與雲端服務供應商「being deployed」（2026年8月），並表示該系統「now in production」，OpenAI預期自2026年第4季起上線（2026年7月）	[S-111; S-236; S-112]
Broadcom	美國	客製化加速器（accelerator）、網路	客製化AI加速器（accelerator）與乙太網路交換晶片，其中包括一款共同封裝光學（co-packaged optics）交換器	產品	2026財年第3季AI半導體營收167億美元，並預期第4季達217億美元（2026年9月）；表示其「now shipping」Tomahawk 6 – Davisson，即一款具共同封裝光學（co-packaged optics）的102.4 Tb/s交換器（2025年10月）	[S-113; S-467]
Huawei	中國	加速器（accelerator）、互連	供中國內建置使用的Ascend加速器（accelerator）與SuperPoD互連	已宣布	Ascend 950DT將於2026年第4季供應；互連協定規格（specification）已發布供合作夥伴採用（2025年9月）	[S-312]
Alphabet（Google）	美國	買方；客製化加速器（accelerator）	雲端買方；開發並部署TPU（TPU v4自2020年起投入服務）；OCP SST規格（specification）的共同開發者	產品；SST：標準	2026年資本支出（capex）財測指引上調至1,950億至2,050億美元；預期2027年資本支出「increase significantly」（2026年7月）	[S-016; S-293; S-019]
Amazon（AWS）	美國	買方	雲端買方；與緯穎（Wiwynn）在德州共同建造供其基礎設施使用的客製化硬體	產品	預期2026年現金資本支出（capex）約2,200億美元，高於原先約2,000億美元，原因為記憶體成本上升（2026年7月）	[S-070; S-306]
Meta	美國	買方；機櫃	買方與營運業者；透過OCP共同開發AMD的Helios機櫃	產品	2026年資本支出（capex）含融資租賃本金，財測指引為1,300億至1,450億美元（2026年7月）；6 GW的AMD GPU協議（2026年2月）	[S-073; S-112]
Microsoft	美國	買方；電力規格（specification）	雲端買方與營運業者；OCP SST規格（specification）的共同開發者	產品；SST：標準	於2026年6月當季新增「another gigawatt of capacity」，並稱其容量在兩年內「on track to roughly double」	[S-075; S-019]
Oracle	美國	買方	具已簽約AI需求的雲端營運業者	產品	在超過300億美元的新AI雲端合約之後，剩餘履約義務（remaining performance obligations, RPO）達6,640億美元；需求「continues to grow faster than supply」（2026年9月）	[S-049]
CoreWeave	美國	買方（AI雲端）	採購機櫃級系統（rack-scale system）並簽約取得電力的AI雲端營運業者	產品	已啟用電力1.5 GW；已簽約電力約3.7 GW；營收未出貨訂單（backlog）約1,040億美元（2026年第2季）	[S-011]
OpenAI	美國	買方；模型開發者	具吉瓦級運算協議的模型開發者；已向OCP提出一份網路規格（specification），即MRC	已宣布；MRC：標準	與NVIDIA的10 GW，首個吉瓦於2026年下半年（2H）採用Vera Rubin（2025年9月）；與Broadcom的10 GW客製化加速器（accelerator）（2025年10月）；MRC於2026年5月提出	[S-468; S-243; S-259]

加速器（accelerator）平台供應商、客製化晶片供應商與AI基礎設施最大買方，並附各角色最新的註明日期證據。資本支出（capital expenditure, capex）數字採用各企業自身的定義與財務年度曆。
GPU為圖形處理器；CPU為中央處理器；TPU為張量處理器；SST為固態變壓器；RPO為剩餘履約義務。各財務季度依各企業自身的曆年。

本章來源（續）

Technology (24 Jun 2026); S-130 KIOXIA (2 Jun 2026); S-169 AUO (31 Aug 2026); S-170 Besi (23 Jul 2026); S-171 Tokyo Electron (30 Jul 2026); S-172 Tokyo Electron (30 Jul 2026); S-178 NVIDIA Blog (13 Oct 2025); S-189 Vertiv (13 Oct 2025); S-191 ABB (21 Sep 2026); S-192 Infineon Technologies (28 May 2026); S-193 Infineon Technologies (5 Aug 2026); S-194 Eaton (13 Oct 2025); S-195 Schneider Electric (13 Oct 2025); S-197 ABB (13 Oct 2025); S-209 LG Electronics (22 Sep 2026); S-211 Boyd (12 Mar 2026); S-231 Bloom Energy (13 Apr 2026); S-236 AMD (23 Jul 2026); S-243 Broadcom Investor Relations (13 Oct 2025); S-259 OpenAI (5 May 2026); S-275 Coherent (17 Mar 2026); S-292 Tower Semiconductor (13 May 2026); S-296 Marvell Technology (16 Mar 2026); S-297 Ayar Labs (Mar 2026); S-298 Marvell Technology (2 Feb 2026); S-299 Eaton (31 Jul 2026); S-300 Siemens Energy (5 Aug 2026); S-301 Siemens Energy (5 Aug 2026); S-306 Wiwynn (10 Sep 2026); S-307 Vertiv (11 Feb 2026); S-308 Schneider Electric (30 Jul 2026); S-312 Huawei (18 Sep 2025); S-317 Amkor Technology (8 Sep 2026); S-318 UMC (8 Dec 2025); S-325 Rapidus (5 Jun 2026); S-467 Broadcom (8 Oct 2025); S-468 NVIDIA Newsroom (22 Sep 2025); S-469 Astera Labs (15 Sep 2026); S-470 YMTC (未註明日期); S-472 Tokyo Electron (3 Aug 2026); S-473 SMIC (7 Aug 2025); S-474 NVIDIA (16 Mar 2026); S-475 Wistron (21 Jul 2026); S-476 Delta Electronics (14 Oct 2025); S-477 Wolfspeed (6 Aug 2026); S-478 STMicroelectronics (17 Mar 2026); S-479 Mitsubishi Heavy Industries (27 May 2026); S-480 Mitsubishi Heavy Industries (16 Jul 2026); S-481 Lumentum (11 Aug 2026)

標準組織與聯盟. S-019 OCP (11 Aug 2026); S-208 OCP (未註明日期)

政府與監管機關. S-313 Shanghai Stock Exchange (24 Jul 2026)

研究機構. S-293 Jouppi et al. (Apr 2023); S-320 imec (8 Dec 2025); S-482 imec (未註明日期)

市場研究. S-276 Cignal AI (7 Jan 2026); S-277 TrendForce (5 Aug 2026)

媒體與法律評論. S-070 StockAnalysis (30 Jul 2026); S-321 Tom's Hardware (27 Jul 2026); S-322 Tom's Hardware (12 Aug 2026); S-471 TrendForce (12 Aug 2026)

表25。 記憶體倚賴三家HBM供應商，而其新產能排定於數年之後

本圖譜中的HBM來自三家供應商，其中兩家表示需求超前於供給；SK hynix所排定的新產能將於2027年至2029年間到位。

企業	國別	層級	在AI基礎設施中的角色	階段	關鍵證據與日期	來源
SK hynix	韓國	HBM、DRAM、NAND	HBM供應商；與NVIDIA的多年期記憶體合作關係（2026年6月）	產品；印第安納州晶圓廠：已宣布	HBM4自2026年第2季起量產出貨；與約10家客戶簽有長期協議（long-term agreement, LTA）；表示客戶需求超出供應能力（2026年7月）；印第安納州HBM量產排定於2029年下半年（2H）	[S-003; S-126; S-023]
Samsung Electronics	韓國	HBM、DRAM、晶圓代工	HBM與晶圓代工供應商；與Broadcom簽署涵蓋HBM與晶圓代工的備忘錄（MOU）	產品；HBM4E：試點階段	HBM4自2026年2月起商業出貨；預期2026年HBM銷售額「more than triple」；表示已出貨首批HBM4E樣品（2026年7月）	[S-056; S-127; S-115]
Micron Technology	美國	HBM、DRAM、SSD	收取客戶保證金的HBM與DRAM供應商	產品；HBM4E：已宣布	HBM4已為其主要客戶的平台大量出貨；HBM4E量產預期於2027年曆年；預估現金保證金及相關承諾達220億美元（2026年6月）	[S-128; S-129]
Kioxia	日本	NAND、SSD	快閃記憶體與資料中心SSD，即HBM之下的儲存層級	產品	計劃於2028財年前將資料中心產品占營收比重提升至60%以上（管理層目標）；正自一年期轉向多年期長期協議（long-term agreement, LTA）（2026年6月）	[S-130]
Astera Labs	美國	記憶體互連	用於記憶體擴充的Compute Express Link（CXL）記憶體控制器	試點階段	Leo控制器「sampling today with hyperscaler customers」（2026年9月）	[S-469]
CXMT	中國	DRAM	國內DRAM製造商	產品（DRAM）	於上海科創板掛牌（2026年7月）；依Tom's Hardware報導，其公開說明書未載有專門的HBM計畫	[S-313; S-321]
YMTC	中國	NAND	3D NAND快閃記憶體製造商（整合元件製造商）	產品	依Tom's Hardware報導，2026年第2季占全球NAND出貨量14%，居第三大（2026年8月）	[S-470; S-322]

記憶體與儲存製造商，以及一家記憶體互連供應商，並附各角色最新的註明日期證據。長期協議（long-term agreement, LTA）與保證金均為各企業自有的合約衡量方式。
HBM為高頻寬記憶體；DRAM為動態隨機存取記憶體；SSD為固態硬碟；MOU為合作備忘錄。Samsung該列中的「first」係該公司自身的主張。

表26。 晶圓、封裝、微影與測試各自倚賴少數依計畫增添產能的供應商

自晶片設計到完成的加速器（accelerator）封裝體，其間每一環節的供應商都很少，而其所載明的產能增添均落在2027年至2029年之間。

企業	國別	層級	在AI基礎設施中的角色	階段	關鍵證據與日期	來源
台積電（TSMC）	台灣	晶圓代工、封裝	供加速器（accelerator）使用的尖端晶圓與CoWoS封裝	產品；藍圖：已宣布	預測2024年至2029年AI加速器（accelerator）營收成長的年複合成長率接近「mid-to-high-fifties」%（2026年1月）；14倍光罩CoWoS預定於2028年投入生產；亞利桑那第2座晶圓廠自2027年下半年（2H）起量產	[S-082; S-024; S-083]
日月光（ASE Technology，SPIL）	台灣	封裝、測試	委外封裝測試；透過其子公司矽品（SPIL）提供CoWoS產能	已宣布（斗六廠）	依TrendForce報導，矽品（SPIL）已於斗六動土興建CoWoS廠，第一期預定於2028年營運（2026年8月）	[S-471]
Amkor Technology	美國	封裝、測試	先進封裝（advanced packaging）與測試；正興建美國園區	已宣布	亞利桑那園區投資額提高至120億美元；第2期預期於2029年底前完工（2026年9月）	[S-317]
ASML	荷蘭	微影	供尖端邏輯與記憶體使用的EUV與DUV微影系統	產品；產能：已宣布	計劃於2027年將其2026年低數值孔徑（low-NA）EUV產能（約65台）再增加30%，並正評估2028年再增加30%（2026年7月）	[S-025]
Besi	荷蘭	混合鍵合（hybrid bonding）、晶粒接合	供2.5D AI封裝體、HBM與光子應用使用的鍵合與封裝設備	產品	2026年第2季訂單2.929億歐元，較2026年第1季成長8.6%，係因資料中心、光子與AI電源管理需求（2026年7月）；長期營收目標上調至17億至22億歐元，未載明目標年度	[S-170; S-057]
Advantest	日本	測試	供加速器（accelerator）、CPU與DRAM使用的測試系統	產品	推論用ASIC、CPU與DRAM的測試機需求「growing faster than we previously expected」；2026年曆年SoC測試機市場看法上調至105億至115億美元（2026年7月）	[S-031]
Tokyo Electron	日本	晶圓廠與封裝設備	晶圓製造設備；供先進封裝（advanced packaging）使用的鍵合機、解鍵合機與探針機	產品	表示供AI資料中心使用的晶圓製造設備占比正成長至超過總量的50%，且HBM堆疊將自12層晶粒增至16層（2026年7月）；預期鍵合機與探針機銷售將成長（2026年8月）	[S-172; S-171; S-472]
SMIC	中國	晶圓代工	供中國國內供應鏈使用的晶圓代工產能	產品	月產能為991,250片8吋約當晶圓，使用率92.5%（2025年第2季）	[S-473]
Rapidus	日本	晶圓代工	政府出資的2 nm邏輯晶圓代工計畫	研發	自日本政府取得1,500億日圓，以於2027年前自研發轉入2 nm全面量產（2026年6月）	[S-325]

晶圓代工廠、封裝廠，以及設備與測試供應商，並附各角色最新的註明日期證據。
CoWoS為Chip-on-Wafer-on-Substrate；SPIL為矽品精密工業（Siliconware Precision Industries）；CAGR為年複合成長率；EUV、DUV分別為極紫外光與深紫外光；NA為數值孔徑；WFE為晶圓製造設備；SoC為系統單晶片；ASIC為特殊應用積體電路。

表27。 機櫃整合商報告的快速成長，係基於彼此不可比較的營收基礎

原始設計製造商（original design manufacturer, ODM）的成長數字顯示的是方向，而非占有率：各企業均依其自身基礎報告，而緯穎的基礎已於2026年4月變更。

企業	國別	層級	在AI基礎設施中的角色	階段	關鍵證據與日期	來源
鴻海（Hon Hai／Foxconn）	台灣	機櫃、交換器	AI機櫃整合與高效能網路產品	產品；CPO交換器：已宣布	Vera Rubin機櫃於2026年第3季進入量產；2026年AI機櫃出貨「expected to more than double」；CPO全光交換器於第3季量產「on track」（2026年8月）	[S-022]
緯穎	台灣	伺服器、機櫃	超大規模伺服器與機櫃原始設計製造商（original design manufacturer, ODM）	產品	自2026年4月起，特定客戶的記憶體改以代理模式採購，並不計入營收與銷貨成本；2026年第2季毛利率9.3%	[S-064; S-063]
緯創（Wistron）	台灣	伺服器、系統	AI伺服器與系統製造，包括在美國生產	產品；Vera Rubin：已宣布	於德州沃斯堡啟用其D1 AI廠房，並表示NVIDIA的GB300超級晶片於該處建造並量產（2026年7月）；經NVIDIA列為預期交付Vera Rubin伺服器的製造商之一（2026年3月）	[S-475; S-474]
Quanta Cloud Technology (QCT)	台灣	伺服器、機櫃	雲端伺服器與機櫃	已宣布（Vera Rubin）	經NVIDIA列為預期交付以Vera Rubin為基礎的伺服器的製造商之一（2026年3月）	[S-474]

整合AI伺服器與機櫃的原始設計製造商（original design manufacturer, ODM）與電子製造商，並附各角色最新的註明日期證據。
CPO為共同封裝光學。NVIDIA另將Cisco、Dell Technologies、HPE、Lenovo、Supermicro、Aivres、華碩（ASUS）、技嘉（GIGABYTE）、英業達（Inventec）與和碩（Pegatron）列入預期交付以Vera Rubin為基礎的伺服器的製造商之列[S-474]。各成長率均為各企業自身的衡量方式。

表28。 電力與散熱供應商出貨的是當前設備，而多數800 VDC角色仍停留在已宣布階段

不斷電系統、開關設備、功率半導體與冷卻液分配單元目前均已出貨；多數800 VDC角色仍屬已宣布或開發中，推出時間落在2026年下半年（2H）至2027年之間。

企業	國別	層級	在AI基礎設施中的角色	階段	關鍵證據與日期	來源
Vertiv	美國	電力、散熱	不斷電系統、熱管理與冷卻液分配單元；與NVIDIA合作的800 VDC產品組合	產品；800 VDC：已宣布	未出貨訂單（backlog）150億美元，訂單出貨比約2.9倍（2025年第4季）；800 VDC產品組合計畫於2026年下半年（2H）推出；冷卻液分配單元已通過NVIDIA DSX Ready驗證（2026年9月）	[S-307; S-189; S-048]
Eaton	愛爾蘭	電力、散熱	配電與不斷電系統；透過原Boyd熱管理事業提供液冷（liquid cooling）	產品；800 VDC：已宣布	完成收購Boyd的熱管理事業（2026年3月）；電氣部門未出貨訂單（backlog）成長43%，並上調2026年有機成長財測指引（2026年7月）；800 VDC參考架構（2025年10月）	[S-211; S-208; S-299; S-194]
Schneider Electric	法國	電力、散熱	配電與資料中心基礎設施；與NVIDIA合作的800 VDC側置電源機櫃（sidecar）	產品；800 VDC：研發	表示2026年Data Center & Networks的需求「up triple-digit in Q2」；正為最高1.2 MW的機櫃開發800 VDC側置電源機櫃（sidecar）（2025年10月）	[S-308; S-195]
ABB	瑞士	電氣化、直流保護	中壓與低壓設備；固態斷路器（solid-state circuit breaker, SSCB）	產品；直流產品組合：已宣布	推出供AI資料中心使用的直流產品組合（2026年9月21日）；表示其SACE Infinitus固態斷路器已取得IEC認證	[S-191; S-197]
台達電子（Delta Electronics）	台灣	電力轉換	機櫃電力轉換；800 VDC電源解決方案	已宣布（800 VDC）	於OCP全球峰會（OCP Global Summit）展出800 VDC解決方案，其中包括效率最高98.5%的800 V轉12 V轉換（2025年10月）	[S-476]
光寶科技（LITE-ON Technology）	台灣	電力轉換	機櫃電源供應器；800 VDC側置電源機櫃（sidecar）	已宣布（800 VDC）	Wolfspeed碳化矽（silicon carbide）已通過光寶800 VDC側置電源機櫃（sidecar）與運算機櫃電源供應平台的驗證（qualification）；未載明供應日期（2026年8月）	[S-477]
Infineon Technologies	德國	功率半導體	用於NVIDIA MGX架構中800 V轉換與保護的功率元件	產品	預期2026財年AI電力相關營收逾16億歐元；多年期產能預訂協議累計金額達高個位數十億歐元（2026年8月）	[S-193; S-021; S-192]
STMicroelectronics	瑞士	功率半導體	與NVIDIA共同開發的800 VDC轉12 V與6 V轉換	已宣布	以12 V與6 V架構擴充其800 VDC電力轉換產品組合（2026年3月）	[S-478]
SolarEdge	以色列	直流保護	以Infineon碳化矽（silicon carbide）JFET為基礎的800 VDC固態斷路器（solid-state circuit breaker）	研發	延長與Infineon的合作，以開發供800 VDC使用的固態斷路器（solid-state circuit breaker）；未載明供應日期（2026年9月）	[S-055]
LG Electronics	韓國	散熱	冷卻液分配單元（coolant distribution unit, CDU）	產品	2.6 MW冷卻液分配單元（coolant distribution unit）已通過NVIDIA DSX Ready驗證（qualification），符合其參考設計（reference design）要求（2026年9月）	[S-209; S-048]

電力設備、功率半導體與散熱供應商，並附各角色最新的註明日期證據。
UPS為不斷電系統；CDU為冷卻液分配單元；JFET為接面場效電晶體；IEC為國際電工委員會（International Electrotechnical Commission）；DSX Ready為NVIDIA針對電力與散熱產品的驗證（qualification）計畫。

表29。 渦輪機製造商的產出與產能時段已預先承諾至數年之後

發電設備依本圖譜中最慢的時鐘運行：渦輪機產出、未出貨訂單（backlog）與產能時段預訂均已承諾至2028年至2030年，且三者屬不同指標，不可相加。

企業	國別	層級	在AI基礎設施中的角色	階段	關鍵證據與日期	來源
GE Vernova	美國	燃氣渦輪機、電網設備	燃氣渦輪機與電氣化設備，包括供資料中心使用者	產品；產出計畫：已宣布	有望自2026年第3季起達到每年20 GW的燃氣渦輪機產出，2028年達24 GW，並正採取行動以於2030年達30 GW；燃氣未出貨訂單（backlog）與產能時段預訂合計116 GW；年初至今資料中心電氣化訂單逾50億美元（2026年7月）	[S-012]
Siemens Energy	德國	燃氣渦輪機、電網設備	燃氣渦輪機與變壓器，包括供資料中心專案使用者	產品	燃氣渦輪機產能「sold-out until fiscal year 2028」（2026年3月）；確定未出貨訂單（backlog）69 GW，另有26 GW產能時段預訂，集團訂單未出貨金額1,620億歐元，二者均以2026年6月30日為基準（2026年8月報告）	[S-018; S-301; S-300]
Mitsubishi Heavy Industries	日本	燃氣渦輪機、散熱	燃氣渦輪機複循環（GTCC）電廠；供AI資料中心使用的冷水機	產品（GTCC）；冷水機：試點階段；產能：已宣布	目標使2028財年GTCC機組出貨量高於現有產能30%以上，其後則考慮達到2024財年水準的約兩倍（2026年5月）；已將一台10 MW級離心式冷水機測試機組運往美國，預定於7月前後抵達（2026年7月）	[S-479; S-480]
Bloom Energy	美國	現地發電	於資料中心場址的燃料電池發電	產品；Oracle計畫：已宣布	擴大合作關係，為Oracle部署最高2.8 GW（2026年4月）	[S-231]

本報告所指的燃氣渦輪機、電網設備與現地發電製造商，並附各角色最新的註明日期證據。
此處的GW數字係設備產出或訂單指標，而非發電量或資料中心負載。

表30。 光學正向封裝體靠近，但多數新角色仍處於試點階段、研發或已宣布階段

可插拔光收發模組（pluggable transceiver）目前已大量出貨；共同封裝光學（co-packaged optics）、光學I/O小晶片（optical I/O chiplet）與新的矽光子（silicon photonics）路線多數仍處於試點階段、研發或已宣布。

企業	國別	層級	在AI基礎設施中的角色	階段	關鍵證據與日期	來源
Marvell Technology	美國	光學DSP、客製化晶片	光學互連晶片與客製化晶片；來自Celestial AI的光子交換網路	產品（光學DSP）	依Cignal AI資料，其Acacia產品線在400ZR等級出貨量居首（2026年1月）；完成收購Celestial AI（2026年2月2日）；宣布將於OFC 2026展示其光學產品與Lumentum光路交換器（optical circuit switch）的互通（2026年3月）	[S-276; S-298; S-296]
Lumentum	美國	雷射、光交換	供光收發模組與共同封裝光學（co-packaged optics）使用的雷射；光路交換（optical circuit switching, OCS）設備	產品；CPO雷射、OCS：試點階段	指出高功率CPO雷射需求上升，並取得外部雷射光源模組的首批訂單；OCS「beginning to layer in」（2026年8月）	[S-481]
Coherent	美國	光收發模組、雷射	光收發模組與雷射	產品；3.2T：研發	依Cignal AI資料，其在800GbE模組出貨量上與Innolight的差距正在縮小（2026年1月）；宣布將於OFC 2026展出每通道400G的3.2T光收發模組（2026年3月）	[S-276; S-275]
友達（AUO）	台灣	光學、封裝	Micro LED共同封裝光學（co-packaged optics）模組與玻璃核心基板（glass-core substrate, GCS）封裝，兩者均在開發中	研發	於其2026年8月31日新聞稿中宣布將在SEMICON Taiwan（2026年9月2日至4日）展出：一款支援最長10 m鏈路的Micro LED CPO模組，搭配Ennostar（Tx）、Tyntek（Rx）與Corning光纖，以及與Corning合作的GCS封裝	[S-052; S-169]
Ayar Labs	美國	光學I/O	供共同封裝光學（co-packaged optics）使用的光學輸入／輸出小晶片（chiplet）	已宣布	完成5億美元E輪募資，以加速CPO小晶片（chiplet）量產（2026年3月）	[S-297]
聯電（UMC）	台灣	矽光子（Silicon photonics）	在imec所授權的iSiPP300平台上製作光子積體電路	已宣布	正與數家新客戶合作，風險性試產「slated for 2026 and 2027」（2025年12月）	[S-318]
Tower Semiconductor	以色列	矽光子（Silicon photonics）	供光收發模組使用的矽光子（silicon photonics）晶圓代工	產品	已簽客戶合約對應2027年矽光子（silicon photonics）營收13億美元；活躍客戶逾50家（2026年5月）	[S-292]
imec	比利時	研究	矽光子（silicon photonics）、CPO與封裝研究	研發	已將其iSiPP300平台授權（licence）給聯電（UMC）（2025年12月）；並將對位公差與光學I/O密度列為CPO衝接的限制之一	[S-320; S-482]
Zhongji Innolight	中國	光收發模組	高速光收發模組	產品	依Cignal AI資料，800GbE模組出貨量居首（2026年1月）	[S-276]
Eoptolink	中國	光收發模組	高速光收發模組	產品	依Cignal AI資料，其在800GbE模組出貨量上與Innolight的差距正在縮小（2026年1月）	[S-276]

光學元件、矽光子（silicon photonics）與光學互連企業，以及一家研究機構，並附各角色最新的註明日期證據。NVIDIA與Broadcom的交換器角色見表24；鴻海（Foxconn）的CPO交換器見表27。

DSP為數位訊號處理器；400ZR與800GbE為光學介面等級；I/O為輸入／輸出；Tx、Rx分別為發射器與接收器。

表31。 本圖譜中沒有任何區域掌握所有層級

各區域的列均集中於少數層級，因此每一區域的AI建置都須倚賴總部設於他處的企業。

區域	平台與買方	記憶體	晶圓代工、封裝、設備、測試	機櫃與伺服器	電力、散熱、發電	光學
美國	NVIDIA、AMD、Broadcom、Alphabet、Amazon、Meta、Microsoft、Oracle、CoreWeave、OpenAI	Micron、Astera Labs	Amkor	—	Vertiv、GE Vernova、Bloom Energy	Marvell、Lumentum、Coherent、Ayar Labs
台灣	—	—	台積電（TSMC）、日月光（ASE，SPIL）	鴻海（Foxconn）、緯穎、緯創、QCT	台達電子（Delta Electronics）、光寶（LITE-ON）	友達（AUO）、聯電（UMC）
韓國	—	SK hynix、Samsung	Samsung（晶圓代工）	—	LG Electronics	—
日本	—	Kioxia	Advantest、Tokyo Electron、Rapidus	—	Mitsubishi Heavy Industries	—
歐洲	—	—	ASML、Besi	—	Schneider Electric、ABB、Infineon、STMicroelectronics、Siemens Energy、Eaton	imec
中國	Huawei	CXMT、YMTC	SMIC	—	—	Zhongji Innolight、Eoptolink
以色列	—	—	—	—	SolarEdge	Tower Semiconductor

本附錄各企業依總部所在區域與層級列示，由前述各表彙整而成，各條目的階段、證據與來源均見於該等表格。破折號表示本圖譜中無該類企業，並非表示該區域沒有此類企業。

Eaton的總部位於愛爾蘭，STMicroelectronics的總部位於瑞士。

附錄C

方法與定義

本報告將來源所述內容與其自身結論加以區分，依類別與強度分類前瞻性陳述，並絕不相加定義不同的指標；本附錄載明這些規則、其定義及其適用限制。

適用讀者： 分析師、審閱者，以及任何重複使用本報告數字或前瞻性陳述者。

本報告的設計是可快速閱讀、慢慢查核。其表層是敘述；在其之下，每一項重大事實均可對應至來源查核登錄表（附錄D）中一份註明日期的公開來源，而每一段落均說明其係報告來源、研讀來源，或在條件（conditions）下作前瞻。本附錄載明使此成為可能的各項規則：五項標記、各類來源及其優先順位、各圖表如何將不同類別的資料分開呈現、市場規模與企業展望如何維持可比性，以及本報告不作何種主張。本附錄所回答的問題是：就任一句話而言，紀錄止於何處，研判（reading）自何處開始？

C.1 五項標記將紀錄與研判區分開來

每一段承載主張的段落均以一個標記起首。 F報告註明日期之公開來源所述內容，企業的主張則保留為該企業的主張。A為本報告對所引事實的研判（reading），並引註該等事實。IF為前瞻性觀點，僅在所載明的條件（conditions）成立期間成立；其指明應觀察的事項，以及使其不成立的條件。OPEN印為?，標示公開紀錄未能確立的重大事項。CALC印為=，標示對論證具重要性的推導計算，並引註其輸入數據、載明其計算公式。未帶標記的段落為敘述性內容：開場、轉折，或如本段這樣的方法規則。每一帶標記的段落均以一句粗體句子起首，說明該段要點，因此僅讀各段起首句即可承載整個論證。

本報告不設可信度等級（confidence tier）。 本報告不將任何陳述評為高、中或低可信度，因為此類評級讀來即如機率，而本報告並未對任何機率加以建模。每一項研判（reading）改而呈現其依據：A段落引註其所依憑的事實，IF段落則指明其條件（conditions）、應觀察的訊號，以及會使其不成立的觀察結果。讀者得以直接權衡該研判，而非信賴一個標籤。基於同一理由，任何前瞻性觀點均不附機率；除其來源本身承諾日期者外，亦不附任何承諾日期。

只有在會改變決策時，才將某一事項列為未解問題。 公開紀錄留下許多細小問題未獲解答，而其中多數並不影響讀者的選擇。本報告僅在答案會改變決策時，才將某一事項標示為OPEN：設備是否確已依某項安全標準取得認證（certification）、某份規格（specification）是否已發布、某項供給缺口是否已量化、某項產品是否已大量出貨，或某一市場規模的定義是否公開。OPEN段落說明的是何事尚未確立及其重要性何在，而非如何進行查找。表格中記為「尚無記錄」的欄位，意指所閱讀的來源中並無該項，而非已證明其在現實中不存在。

C.2 公開來源，閱讀至2026年10月2日，並依固定的優先順位

本報告依憑閱讀至2026年10月2日（其研究資料截止日）的公開來源。 每一項狀態、版本與展望均以該日為基準陳述；為第一頁校準、市場比較與附錄E所新增的經濟與市場研究來源，則閱讀於2026年9月28日。本報告首先取材於原始來源：企業申報文件、業績發布、法人說明會與投資人簡報；規格（specification）與標準組織文件；以及政府、監管機關與研究機構的出版物，例如國際能源總署（International Energy Agency, IEA）[S-005]。在缺乏原始資料之處，則採用TrendForce與IDC等機構的市場研究，並標示為估計值或預測，且載明出處。產業媒體僅在無任何原始來源時採用，並載明出處。每一項來源在來源查核登錄表中均有一張卡片，載明其原始網址、以該來源原文語言呈現的逐字摘錄、必要時的英文譯文，以及該來源所支持與無法據以確立的內容。

市場研究取自數家機構，各自對應其所衡量的鏈條環節。 沒有任何單一機構涵蓋整條鏈條，且各家定義互有差異。本報告採用IDC衡量OCP認可、AI基礎設施與伺服器支出；Gartner衡量依市場劃分的AI支出及其修正；TrendForce衡量零組件、伺服器與雲端資本支出；SEMI衡量半導體設備；TechInsights衡量記憶體營收；Dell'Oro Group衡量實體基礎設施與AI網路；Synergy Research Group衡量超大規模雲端產能；LightCounting衡量光學。各家數字絕不相加、平均，或置於同一數值軸上；若兩家衡量相近事物，則一併呈現並載明其定義（圖10）。市場研究數字僅自其發布機構自身的頁面取用，調查結果亦僅依其公開頁面所印內容採用。

原始來源優先於摘要，較晚版本優先於較早版本。 原始規格（specification）、申報文件、官方新聞稿與政府紀錄，其效力優先於對該等文件所作的摘要；而註明日期較晚的版本則取代較早者。當效力相當的來源相互衝突時，本報告呈現該項衝突，而非以假定加以解決。例如，台積電（TSMC）2026年4月自身的論壇新聞稿，對其藍圖日期的效力優先於媒體對同一活動的報導：14倍光罩CoWoS「slated for production in 2028」，A13「scheduled to enter production in 2029」[S-024]。

企業申報文件就其管理層所述內容具權威性。 一家企業的申報文件或新聞稿，是其自身業績、財測指引、產能、已揭露合約與藍圖，以及其管理層預期的最佳來源。但其本身並不足以證明市場總規模、競爭者表現、全產業短缺、「世界首創」、技術優越性或未來採用情形。因此，本報告保留企業陳述的出處歸屬。SK hynix描述「the environment where customer demand exceeds supply capacity」[S-125]：這是SK hynix對其自身市場的說法，而關於記憶體整體的研判（reading），尚須其他供應商與市場來源一併支撐（第3.2節）。

可信來源之間的衝突均予呈現，並以四種方式之一處理。 當可信來源互有歧異時，登錄表會印出兩份摘錄，以一句話載明其差異，並記錄其處理方式。若某一較弱或相互衝突的來源會改變結論，則絕不隱而不報。

- **不作結論。** 證據無法解決該項衝突；凡研判（reading）取決於該項差異者，本報告即不依憑其中任一來源。
- **原始來源優先。** 具效力的申報文件或規格（specification）取代對其所作的摘要，一如台積電（TSMC）的新聞稿對其藍圖日期的效力[S-024]。
- **較晚來源優先。** 註明日期較晚的正式來源取代較早者，一如JEDEC於2025年12月的JESD270-4A修訂版取代2025年4月的HBM4標準[S-058; S-139]。
- **範圍不同。** 兩者在不同定義下均可成立：OCP與IDC所稱至2029年AI基礎設施支出約1,550億美元，係OCP認可支出之中屬AI使用情境的部分，而非整個AI基礎設施市場[S-004; S-108]。

已被取代的來源不入本文，除非該項變動本身即為重點。 當一份正式來源被取代時，本文引用現行版本。較舊的來源僅在該項變動本身承載資訊時保留，且其登錄表卡片會標示為已被取代，並以一行註記說明變動內容。財測指引經修訂是常見情形。Alphabet 2026年資本支出（capital expenditure）區間自2月的1,750億至1,850億美元，於4月調整為1,800億至1,900億美元以納入一項收購相關投資，並於7月調整為1,950億至2,050億美元；該區間共上調兩次，其中一次係因範圍變動[S-014; S-015; S-016]。這屬於市場情報，而非來源沿革（第2.1節）。

模型卡、組態檔與申報文件均依其自身效力閱讀。 開發者的模型卡、組態檔或發布說明屬企業來源：就該開發者對其自有模型所載明的內容具權威性，例如參數、上下文長度、注意力配置與數值格式；但就該模型在他人部署環境中的表現則不具權威性。表17中的快取數字係本報告依該等組態計算所得，除開發者另有載明者外，均以每個儲存數值16位元計。財務報表的會計附註與企業展望亦以同樣方式閱讀：它們確立的是該企業自身的界限，例如依IFRS 15計算的代理人淨額營收，或對中國的某項假設，而非其所未納入的金額。

C.3 各圖表顯示每一數值點屬於何種類別的數字

每一項入圖數值均載有其資料類別與來源。 已公布的某一季數字、企業的財測指引與分析師的預測，屬不同類別的數字，各圖表亦以不同方式繪製（表34）。標題載明的是發現，而非主題；圖說則載明所繪內容、涵蓋期間、所採單位與所對比的分母，並載明來源。凡評分會隱含超出紀錄所能支持的精確度之處，即以階段標籤、狀態圖與四象限取代評分。

顏色表示某一數值描述系統的哪一部分；線條與標記則表示其屬何種類別的數字。 兩套編碼彼此獨立，因此記憶體支出預測與電力用量預測共用點線，而以色調相區別（表33）。四種色調承載大部分意義：靛藍代表運算、牛血紅代表記憶體、藍綠代表散熱、青銅代表標準；電力為石墨灰、資本為碳黑、政策為梅紫。紅色僅標示稀缺或價格壓力，絕不用於預測。歷史比較以石板灰代表較早期間，以深靛藍代表當前期間。所有區別在灰階列印下亦可辨識，因為線條樣式、標記與直接標籤同樣承載這些區別。

不同單位絕不共用同一軸。 當兩項指標同向變動但單位不同時，例如營收與利潤率，或資本支出（capital expenditure）與電力，本報告採用對齊的小型多圖：各自獨立的面板、每一面板一種單位、共用同一時間方向。本報告不使用雙軸，因雙軸會使尺度的選擇得以製造出相關性；亦不使用以100為基期的裝飾性指數化。已公布區間以區間繪製，而財測指引區間則置於其給出之日，因此修訂讀來即如該帶狀區間的移動（圖5）。

斷點與邊界均標示於圖表本身。 當一家企業改變其所公布數字的基礎時，圖表即在變動之日標示斷點，跨越該斷點的比較須附基礎說明。緯穎為主要案例：自2026年4月起，特定客戶的記憶體採購（procurement）改為代理模式，因此該等記憶體金額不再計入營收或銷貨成本[S-064; S-063]（圖22）。邊界線標示已公布資料結束、推估開始之處；若一項圖表混合了不同範圍，則會附一行範圍說明，載明其差異。

任何序數資料均不相加，階段標籤亦絕不上調。 瓶頸圖（圖3）僅在有註明日期的來源支持時才記錄各項狀態（無重大約束、正在形成、具約束力、不確定），且不附任何合計、平均或排名：某一列中具約束力的約束，並不會被另一列中的閒置產能所抵銷。階段標籤（產品、試點階段、研發、標準、已宣布）均依循公開紀錄；研發不會被稱為供應，試點階段也不會被稱為商業化。展示不等於部署，而一場已排定的會議場次代表的是關注度，而非採用。

表32。 每一項標記都向讀者說明該段落所作陳述的類別，以及何種情形會使其有誤

讀者僅憑標記即可判別一句話是報告來源、研讀來源、在條件（conditions）下作前瞻、指出缺口，或推導出一項數字。

標記	印製符號	所主張的內容	何種情形會使其有誤	本報告中的實例
F	F	註明日期的公開來源如此表示	該來源被誤述，或已被取代	NVIDIA對2027財年第3季的展望假設來自中國的Data Center運算營收為零；該假設涵蓋展望期間的Data Center運算營收，而非整家公司[S-010]
A	A	本報告對所引事實的研判（reading）	另一種研判（reading）更能契合同一批事實	由於ORW與Deschutes CDU已發布，而800 VDC電力文件仍未達1.0版，2026年的標案僅能將電力路徑作為設計參考加以引用[S-228; S-030; S-054]
IF	IF	在具名條件（conditions）下的前瞻性觀點，並附應觀察的訊號及使其不成立的條件	條件（conditions）成立，結果卻仍未隨之發生	800 VDC唯有在多供應商硬體、保護與認證（certification）隨之到位時，才會擴展至參考設計（reference design）之外；應觀察依UL 2877取得認證的機組[S-188]
OPEN	?	公開紀錄未能確立的重大事項	有公開來源予以確立	截至資料截止日，並無任何機組依中壓轉800 VDC電源機組的UL 2877計畫取得認證之紀錄[S-188]
CALC	=	推導而得的數字，並附其輸入數據與計算公式	某一輸入數據或計算公式有誤	電流I = P / V：一座1 MW機櫃在800 VDC下取用1,250 A，而在ORv3電源架（power shelf）51 VDC標稱輸出下約取用19,600 A[S-027; S-180]

五項面向讀者的標記、各項所主張的內容、何種情形會使其有誤，以及本報告中的一項實例及其來源。

表33。 顏色標示某一數值所描述的領域，同一色調並貫穿所有圖表

讀者可從顏色判別某一圖表描述系統的哪一層級，並從線條判別其呈現何種類別的數字。

色調	領域	用途	範例
墨藍	總覽	第一頁的標題與數字；OCP認可支出；路標（signpost）	圖1
靛藍	運算	加速器（accelerator）、伺服器、AI雲端產能、模型使用量	圖6
石板靛藍	封裝、機櫃、互連	網路架構、光學、機櫃外型規格	圖21
牛血紅	記憶體、儲存、價格壓力	DRAM與HBM營收、價格與預算占比	圖14
藍綠	散熱	液冷（liquid cooling）普及率、熱路徑	圖18
石墨灰	電力、設施、電網	電力用量、電力路徑、電網容量	圖11
碳黑	資本與融資	資本支出（capital expenditure）、財測指引、市場規模	圖4
青銅	標準、驗證（qualification）	規格（specification）成熟度	圖26
梅紫	政策、地緣政治	區域依賴關係	圖25
石板灰	較早的期間	1990年代後期的數列	圖34
深靛藍	當前期間	2020年代的數列	圖34

本報告的各領域色調、各色調的用途，以及一項範例圖表。來源編號與引述則保留其專用的綠色。

表34。 八種資料類別將已公布數字與財測指引、目標、預測及推導橋接值區分開來

僅憑線條與標記，讀者即可判別某一數值是已公布、屬財測指引、屬目標、屬計畫，或係他人的預測。

類別	意義	線條	標記	來源實例
已公布	由該企業或機構公布	實線	實心圓	季度業績發布[S-010]
估計值	該來源對已在進行中或剛結束期間所作的估計	實線，較淺	實心，較淺	國際能源總署（IEA）對2025年的估計[S-005]
企業財測指引	企業對所載明期間的財測指引	虛線	實心菱形	資本支出（capital expenditure）區間[S-016]
管理層目標	管理層目標，而非預測	虛線	空心菱形	長期營收目標[S-057]
容量	已安裝、已規劃、已預訂或已簽約的容量；標籤載明其為何者	實線	實心方形	燃氣渦輪機產出計畫[S-012]
第三方預測	第三方的預測	點線	空心圓	市場研究預測[S-007]
條件式情境（scenario）	本報告的條件式情境（scenario），或企業所稱的「could」	點線，灰色	空心灰色圓	「could」達15,000台的測試機產能[S-116]
推導橋接值	在兩個已公布端點之間，由已公布或隱含的成長率所推得的數值；絕非某一來源的預測	細線，較淺，採該領域的色調	無	依IDC每年22.2%推算2025年至2029年間的OCP認可支出[S-004]

本報告圖表所採用的資料類別、各類別的意義、其繪製方式，以及提供該類別資料的一項來源實例。

C.4 四種市場規模概念與四種規劃時鐘

市場規模僅在其自身定義範圍內可比。 本報告區分四種概念。總體潛在市場（total addressable market, TAM）為某一產品理論上可服務的全體。支出預測係估計在界定範圍內的買方支出。企業營收機會則為某一企業表示其可觸及的部分，常以情境（scenario）形式陳述。認可市場，或稱與標準連結的市場，則為依某一標準組織的方法所計入的支出。這四者絕不相加，亦不置於同一排名尺度上；若方法互有差異，本報告則採用各自獨立的面板（表2）。

F OCP與IDC的數列衡量的是對OCP認可之資訊科技基礎設施與解決方案的支出。 OCP的採用情形頁面依其與IDC共同進行的研究指出，此類支出將自2025年的1,320億美元成長至2029年的2,950億美元，年複合成長率為22.2%[S-004; S-106]。該頁面並載明至2029年AI基礎設施支出約1,550億美元[S-004]，而IDC的網路研討會簡報則將其呈現為OCP基礎設施支出之中屬AI使用情境的部分[S-108]。以上述用語而言，該數列屬認可市場（第2.4節）。

> **範圍 OCP與IDC的市場數字**
> 並非OCP Foundation的營收、並非AI資料中心的總體潛在市場（TAM）、並非超大規模雲端業者（hyperscaler）資本支出，也並非任何單一規格（specification）的採用程度。除非方法經過對齊，否則不可與其他市場研究比較。

技術堆疊的每一層各依自身的時鐘運行。 本報告透過四個規劃時鐘研判時程，每一個時鐘代表自投資或下單決策至產能可交付的典型所需時間（圖2）。晶片與伺服器依季度時鐘運行，為三至十二個月：加速器（accelerator）、伺服器與記憶體的出貨、價格與利潤率逐季變動。記憶體與微影設備產能、機櫃平台與冷卻驗證（qualification）依一至二年的時鐘運行。新建晶圓廠與先進封裝（advanced packaging）產能依三至五年的時鐘運行。渦輪機、電網併網與其他電力基礎設施依五至十年的時鐘運行，標準趨同、核能與其他大型發電，以及區域結構性變化亦然。各時鐘的時長為本報告的分析，並以註明日期的企業陳述為依據。

A 同一波熱潮在運算上顯得迅速，在電力上卻顯得緩慢。 企業藍圖呈現了這一落差。台積電（TSMC）14倍光罩尺寸的CoWoS排定於2028年量產[S-024]；GE Vernova計畫燃氣渦輪機產出於2028年達24 GW，並正「implementing actions」以於2030年達到30 GW[S-012]；Siemens Energy則描述燃氣渦輪機市場「remains supply-constrained rather than demand-constrained」[S-026]。商業機會與延誤風險，均落在這些時鐘之間的落差：依季度時鐘決定的產能，必須等待依多年期時鐘決定的投入要素（第1.1節）。

C.5 前瞻性陳述依類型分類，並依其自身的力度研判

並非每一句談及未來的話都是財測指引。 本報告將每一項企業前瞻性陳述歸入五種類型之一（表35），且絕不以較弱的類型充作較強的類型呈現：策略情境（scenario）不是財測指引，而目標除企業自稱其為預測外，並非預測。產能陳述一律載明該產能屬已安裝、已規劃、已預留、已簽約或在建，因為此一差異決定買方能夠倚賴其中多少。

表35。 五種前瞻性陳述各有不同的分量；情境絕不被研判為財測指引
在比較兩項企業對未來的數字之前，應先確認兩者屬於同一類型的陳述。

類型	定義	紀錄中的實例	本報告的處理方式
G — 財務指引	特定期間的營收、資本支出（capital expenditure）、利潤率或獲利	NVIDIA：2027財年第3季營收「is expected to be $108.0 billion, plus or minus 2%」[S-010]	最強的前瞻類別；呈現時一併載明其區間與提出日期
T — 管理層目標	長期財務模型、利潤率區間或產能目標	Besi於2026年6月的投資人日將其長期營收目標上調至17億至22億歐元，但未載明目標年度[S-057]	並非預測；僅在最新申報文件再次提及時，方描述為已重申
R — 藍圖	產品、製程節點或平台的日期	台積電（TSMC）：A13「scheduled to enter production in 2029」；14倍光罩尺寸CoWoS「slated for production in 2028」[S-024]	標示為藍圖或已規劃量產；日期可能延後
C — 產能承諾	產出、生產檔期或產能：已安裝、已規劃、已預留、已簽約或在建	GE Vernova：燃氣渦輪機年產出於2026年第3季達20 GW、2028年達24 GW，進度符合預期，屬已規劃的量產產能[S-012]	一律載明屬何種產能；絕不與不同的產能衡量方式相加
S — 情境（scenario）或期望	「Could」、「opportunity」等用語，以及市場情境（scenario）	Samsung與Broadcom依一項合作備忘錄，估計雙方合作規模於2030年之前逾2,000億美元[S-115]	絕不以財測指引或營收的形式呈現

本報告所使用的五種企業前瞻性陳述類型，並各附一則取自紀錄的實例，以及本報告的處理方式。

企業用語一律保留其自身的力度。 「Expects」是展望；「plans」是計畫，而非結果；「targets」是管理層目標，而非財測指引；「aims」是期望；「could」是情境（scenario）；「scheduled」則是仍有待執行的藍圖日期。稱某項產品「sold out」或稱「demand exceeds supply」的說法，均連同其產品、期間與範圍（HBM、伺服器DRAM、記憶體整體、某一製程節點或雲端產能）一併報導，且絕不加以擴大。因此，Advantest稱其下一階段測試機產能「could be at a scale of 15,000 systems or more」的陳述屬情境，與其5,000台及10,000台的目標並列呈現[S-116]。

單位相同的數字並不因此即可比較。 各企業各以自身方式界定資本支出（capital expenditure），包括融資租賃或營業租賃、現金制或應計制、使用年限等，因此本報告將各家資本支出並列呈現，絕不加總或排序。Microsoft即為一例：自2027財年起，該公司將其資料中心與辦公建物的使用年限自15年延長至25年，使更多未來租賃成為不計入資本支出的營業租賃，其2026日曆年的預期金額因而自約1,900億美元調整為約1,750億美元，投資計畫並未改變，此為認定基礎的變動而非削減[S-045; S-075]。未出貨訂單（backlog）與剩餘履約義務（remaining performance obligations），以及不同的GW與占比衡量方式，亦以同樣方式分開處理。

採用最新的財測指引，較早的數字僅用於呈現修訂情形。 各企業的數字均取自截止日前最新的申報文件或發布資料。於最新申報文件中再次提及的長期目標，描述為已重申；未再提及者，則不假定其仍然成立。企業已排定更新時程者，現行觀點均註明日期：ASML關於2030年年營收約440億至600億歐元的情境（scenario），出自其2024年11月的投資人日[S-328]；ASML並表示將於2027年6月10日的資本市場日更新其較長期的觀點[S-025]。

C.6 本報告未主張之事項，以及其限制所在

本報告屬情報分析，而非建議。 本報告不提供投資建議：企業數字均連同其日期與認定基礎一併報導，任何陳述均非對某項證券的看法。本報告不提供法律建議：專利政策、授權（licence）與出口管制（export control）工具均依其文本所載內容描述，絕不構成任何一方擁有、侵害、必須取得授權或得以出口任何事物的認定，且任何專利結論均須經請求項層級的新穎性與自由實施（freedom to operate）審查。本報告不提供會計、安全認證（certification）或工程建議：諸如特定電壓下電流值之類的算術，僅用於說明機制，並非為系統進行容量設計。

前瞻性觀點均屬條件式情境（scenario），而非預測。 任何前瞻性觀點均未附機率，因為並未就機率進行建模；除來源本身作出承諾者外，任何觀點均不帶有承諾日期。各項觀點均載明其所需條件，以及足以改變該觀點的訊號。展示不等於量產，標準發布不保證獲得採用，而列入OCP全球峰會（OCP Global Summit）議程的一場會議屬關注度，而非部署。本報告不對企業或政府歸因動機，亦不將向標準制定組織（standards development organisation）提交貢獻文件（contribution）的數量，研判為影響力、技術優勢或市場力量。

主要限制傾向於低估紀錄所能支持的內容。 第一，本研究涵蓋截止日前公開網頁所提供的內容；需註冊方可取得的資料、由用戶端產生的議程，以及部分貢獻文件（contribution）資料庫未能充分閱讀，因此誤差方向偏向將真實的論點列為未解，而非將錯誤的論點列為已確立（established）。第二，部分數字在無一手來源的情況下，倚賴市場研究或產業媒體；這些數字均已標註出處，但可能延續該來源的誤差。第三，企業財測指引、產能與藍圖均為管理層自身的陳述，僅標註出處，並未經獨立查證。第四，各項狀態很快過時，首先即是2026年10月12日至15日的OCP全球峰會（OCP Global Summit）[S-067]。

自本報告引用的數字，應連同其適用範圍一併使用。 自圖表中擷取的數字，若未同時帶上其類別、認定基礎與日期，即會失去這些資訊。重複使用某一數字時，應保留其資料類別、該企業自身的定義（資本支出認定基礎、未出貨訂單概念、產能衡量方式）、陳述日期及其來源編號，並應優先採用一手數字，而非未附來源而流傳的較大數字。例如，NVIDIA公布其800 VDC架構可減少45%的銅用量[S-028]，而已發布的OAM規格（specification）支援功率最高達1,000 W的模組[S-183]。

C.7 用語界定

本節界定讀者閱讀本報告所需的技術與市場用語。各條目說明該用語的含義，及其對AI基礎設施的重要性，並指出本報告中加以展開的部分；數字與狀態均以2026年9月27日為基準。縮寫的全稱見第C.8節。

800 VDC與±400 VDC 以800伏特向AI機櫃進行直流供電，可為兩導體之間（0–800 V），亦可為+400 V與−400 V兩極的形式（±400 VDC）。電壓較高者以較小的電流傳送相同功率，可節省銅材與機櫃空間，但直流保護與認證（certification）仍在確立之中：OCP已制定±400 VDC側置電源機櫃（sidecar）規格，並宣布推出原生800 V的後繼方案，惟未載明日期[S-027; S-019]。見第4.2節。

加速器（Accelerator） 專為執行AI模型運算而打造的處理器，可能是銷售給眾多買方的圖形處理器（graphics processing unit, GPU），也可能是為單一買方的工作負載所設計的客製晶片。AI產能以加速器（accelerator）數量計算，但依本報告的研判，記憶體、封裝、電力與電網接取現已限制可部署的數量。見第3.1節。

先進封裝 將數顆晶片——例如邏輯晶粒與高頻寬記憶體（high-bandwidth memory, HBM）堆疊——接合於同一封裝體內的組裝工序，可為並列於中介層或橋接結構之上（2.5D），亦可為垂直堆疊（3D）。對AI加速器（accelerator）而言，這本身即構成一項產能約束；依TrendForce報導，買方已預先取得CoWoS封裝產能[S-017]。見第3.4節。

關係企業規則 美國商務部產業安全局（Bureau of Industry and Security）2025年9月30日的一項規則，將實體清單（Entity List）限制——依該限制，與清單所列外國當事方往來須取得許可（licence）——擴及由清單所列當事方持股50%以上的實體；該規則自2025年11月10日起至2026年11月9日止暫停適用[S-414]。該規則若失效，將影響含有此類關係企業的硬體供應鏈，而非已發布的規格（specification）。見第7.1節。

代理人模式（本人與代理人，Agency model） 依IFRS 15，若一企業於特定商品移轉給客戶之前即控制該商品，則為本人；若其係安排他方提供該商品，則為代理人；代理人僅就其服務費或佣金認列收入[S-484]。緯穎（Wiwynn）截至2026年6月30日之六個月經核閱期中財務報表，就客戶指定供應商所供應的物料將其視為代理人，而其營運結果發布資料自2026年4月起載明記憶體，因此該等金額不再計入其營收與銷貨成本[S-483; S-063]。跨越此一變動的營收與利潤率須附認定基礎說明；該變動本身並不構成供應短缺的證據。見第5.2節。

特定應用積體電路（Application-specific integrated circuit, ASIC） 為單一用途所設計的晶片；在AI領域中，指雲端業者與晶片夥伴共同為自身工作負載設計的客製加速器（accelerator），例如Google的Tensor Processing Unit（TPU）。TrendForce預測以ASIC為基礎的系統將占2026年AI伺服器出貨量的27.8%，至2030年接近40%，其成長係與GPU系統並行，而非取而代之[S-008; S-009]。見第3.1節。

未出貨訂單（Backlog） 企業尚未交付或認列的訂單或已簽約收入，依各企業自身的衡量方式申報，例如未出貨訂單（backlog）、剩餘履約義務（remaining performance obligations, RPO）或營收未出貨訂單。各項衡量方式均顯示已承諾的需求，但在計入範圍與轉為營收的速度上各有不同，因此並列呈現，絕不相加或排序。見第2.2節。

基礎晶粒（Base die） 位於高頻寬記憶體（high-bandwidth memory, HBM）堆疊最底層的邏輯晶片，負責將堆疊的記憶體晶粒與處理器相連；Samsung的HBM4係建構於4奈米邏輯基礎晶粒之上[S-140]。JEDEC的標準所確定的是通往處理器的通道介面，而非堆疊下方的邏輯，因此設計自由度與差異化空間轉移至基礎晶粒[S-058]。見第3.4節。

電池儲能系統（Battery energy storage system, BESS） 電池及其電力轉換與控制裝置，用於為場址提供備援，並緩衝AI負載的快速變動。此類系統現已透過OCP的儲能要求與NVIDIA的DSX Ready計畫進行驗證（qualification），而NFPA 855（2026年版）則要求進行大規模火災測試[S-205; S-048]。見第4.4節。

表後發電（Behind-the-meter generation） 於場址之上或緊鄰場址處，以燃料電池、渦輪機或微電網所產生，並在當地使用而不先經過公用電網的電力。在併網緩慢之處，營運業者會增設此類電源，一如Oracle依其與Bloom Energy的協議採購最高2.8 GW的燃料電池；但此舉只是把等待移轉至訂單已滿的發電設備[S-231; S-012]。見第4.4節。

訂單出貨比（Book-to-bill ratio） 一期間內所接獲新訂單的金額，除以同一期間已開立帳單的營收；比率高於1表示在手訂單正在成長。電力、冷卻與電網設備供應商申報的比率高於1，例如Vertiv於2025年第4季約為2.9[S-307]。見第5.4節。

匯流排（Busbar） 沿機櫃輸送電力至其所連接各機箱的剛性金屬導體。其電壓決定配電所占用的銅材與空間：ORv3機櫃採用以標稱51 VDC供電的48 V級匯流排（busbar），而針對百萬瓦級機櫃，目前正制定更高電壓的規格[S-180; S-019]。見第4.2節。

Caliptra 一項開放的矽信任根（root of trust），即自晶片第一道指令起即檢查其韌體的模組；其規格於OCP制定，程式碼則由CHIPS Alliance維護[S-362; S-363]。晶片所整合的版本，決定了該晶片在其服務年限內的後量子密碼學（post-quantum cryptography）支援能力：2.0版新增ML-DSA-87簽章，2.1版硬體則新增ML-KEM[S-364; S-365]。見第6.4節。

容量（GW）與電量（TWh） 電力的兩種不同衡量方式：容量以吉瓦為單位，指某一時點必須可供使用的功率；電量以太瓦時為單位，指一年之內的消耗量。變電所與發電廠的規模依容量設計，而燃料用量與電費則隨電量而定，因此本報告絕不將兩者相互換算、相加或繪製為同一序列。見第7.3節。

資本支出（Capital expenditure, capex） 用於伺服器、網路設備與建物等長期性資產的支出。最大AI買方的資本支出（capital expenditure），是供應鏈其餘環節據以規劃的需求訊號，但各企業的定義各異，例如是否計入融資租賃，因此各項數字並列比較，絕不加總。見第2.1節。

小晶片（Chiplet） 為與其他晶粒組合於同一封裝體內而設計的晶粒，使處理器得以由不同製程或不同供應商所製造的部件

組裝而成。開放規格——例如用於晶粒間連結的UCIe，以及將小晶片（chiplet）角色對應至既有介面的OCP之Foundation Chiplet System Architecture（FCSA）——使第二家晶粒供應來源的驗證（qualification）成本降低，但並未增添封裝產能[S-167; S-165]。見第3.4節。

Chip-on-Wafer-on-Substrate (CoWoS) 台積電（TSMC）的先進封裝（advanced packaging）技術，將邏輯晶粒與HBM堆疊並列置於矽中介層或橋接結構之上，再置於封裝基板上；TrendForce預期其CoWoS-L形式將持續為AI晶片的主流封裝[S-036]。其產能限制了加速器（accelerator）的出貨數量；台積電已將可容納約10顆運算晶粒與20座HBM堆疊的14倍光罩尺寸版本，排定於2028年量產[S-024]。見第3.4節。

冷板（Cold plate） 安裝於處理器之上、內具流道的金屬板，液態冷卻液流經其中，直接將熱自晶片帶走。OCP所標準化的，主要是迴路的接頭與流體，例如快速接頭（quick disconnect）與水基冷卻液，而非冷板本身——冷板的設計仍開放給發明空間[S-212; S-217]。見第4.3節。

Compute Express Link (CXL) 一項開放的互連標準，用於在處理器自身記憶體通道之外附掛與匯集記憶體；2025年11月發布的CXL 4.0將鏈路速率加倍至128 GT/s[S-158]。在AI服務中，其為保存上下文提供較廉價的層級，惟一份OCP白皮書（white paper）指出，AI垂直擴展工作負載並不需要其硬體快取一致性（cache coherence）[S-161]。見第3.3節。

已簽約電力與已啟用電力（Contracted power and active power） 已簽約電力，指資料中心營運業者為未來場址依合約取得的電力容量；已啟用電力，則指已通電並正服務客戶的容量。兩者的落差顯示有多少已承諾的需求仍在等待電力：截至2026年6月30日，CoreWeave已啟用1.5 GW，而已簽約者約為3.7 GW[S-011]。見第2.2節。

合約價格（Contract price） 記憶體製造商與大型買方就一段期間的供應所議定的價格，與公開市場上的現貨價格有別。伺服器DRAM合約價格是本報告主要的記憶體訊號：TrendForce報導2025年下半年累計上漲64%，並預期2026年將再上漲約270%[S-007]。見第3.2節。

冷卻液分配單元（Coolant distribution unit (CDU)） 由泵與熱交換器構成的單元，在使兩個迴路彼此隔離的同時，將熱自機櫃液冷迴路移轉至設施水系統（facility water system）。OCP的Project Deschutes規格（1.0.0版，2026年2月）界定了一款約每分鐘500美制加侖的2 MW單元，此一公開的目標值使買方得以依同一項要求驗證（qualification）數家供應商[S-030; S-207]。見第4.3節。

共同封裝光學（CPO） 將光引擎置於與交換器或處理器晶片同一封裝體之內，使訊號以光的形式離開封裝，而非經銅線傳輸至前面板的可插拔模組。供應商稱其電力效率優於可插拔光收發模組（pluggable transceiver）[S-024]，但雷射、光纖接合、測試與現場維修均移入封裝之內；一篇2026年的回顧文獻則將熱管理、可製造性與標準化列為障礙[S-053]。見第4.7節。

負載削減（Curtailment） 由電網營運機構指示或依合約允許的客戶用電削減，通常發生於緊急狀況期間。新的大型負載（large load）規則，將未自備電源的資料中心列為負載削減（curtailment）的首要對象，因而使具備彈性調節、緩衝或自行供電能力的負載，取得較佳的併接條件[S-042; S-040]。見第7.2節。

直流故障保護（DC fault protection） 偵測直流匯流排（busbar）上的短路，並在損害擴散之前將其切斷的裝置與設定。此事比在交流系統中更為困難，因為直流電不具可使電弧自然熄滅的零交越點；而決定何一裝置先行動作——即所謂的協調配合——其重要性不亞於任何單一斷路器。見第4.2節。

解碼（Decode） AI推論中逐一產生輸出token的階段，每產生一個新token，均須自記憶體讀取模型的作用中權重與該對話已儲存的上下文。由於其每讀取一位元組所執行的運算量甚少，決定其運行速度的是記憶體頻寬而非處理器速度，這也是HBM之所以成為AI服務節奏決定因素的原因[S-137; S-136]。見第3.3節。

動態隨機存取記憶體（Dynamic random-access memory, DRAM） 電腦的主要工作記憶體，以矽晶圓上的晶片形式製造；HBM堆疊即由DRAM晶粒構成，而伺服器DRAM則位於主板上處理器的旁側。它已同時成為預算項目與效能項目：TrendForce估計DRAM與NAND快閃記憶體將占主要雲端業者2026年資本支出（capital expenditure）的47%，2027年則占68%[S-007]。見第3.2節。

Ethernet for Scale-Up Networking (ESUN) 一個OCP工作流（workstream），其基礎規格（specification）自2026年2月9日生效，訂定網路營運者對垂直擴展網路（scale-up network）中乙太網路交換與訊框處理的要求；至該規格發布時，參與企業已逾175家[S-246; S-020]。該規格縮小了各款交換器之間的差異，但將加速器（accelerator）端點留給另一個工作流SUE-T處理，而後者尚未發布規格[S-249]。見第4.6節。

極紫外光（EUV）微影（Extreme ultraviolet lithography） 用於製作先進邏輯與記憶體晶片最精細圖形的晶片圖案化技術；深紫外光（deep ultraviolet, DUV）設備用於曝印較粗的層，數值孔徑（numerical aperture, NA）則用以描述其光學特性。微影產能在封裝之前即限制了晶圓產出：ASML計畫於2027年將其低數值孔徑EUV與DUV浸潤式產能增加30%[S-025]。見第3.5節。

最終規格（Final Specification） 在OCP的流程中，指經指導委員會（Steering Committee）核准的規格（specification）提案，發布時並附其授權（licence）聲明、貢獻者名單及任何有效的排除通知（Exclusion Notice）[S-047]。發布之時即為貢獻者免權利金（royalty-free）專利承諾（patent commitment）成立之時；其並不表示產品已符合該規格，或買方已予部署。見第6.6節。

融資租賃與營業租賃（Finance lease and operating lease） 租賃的兩種會計分類：融資租賃的處理方式近似以賒購方式購入資產，部分企業將其計入資本支出（capital expenditure）；營業租賃則於租期內列為費用，不計入資本支出。其分類部分取決於資產的使用年限，這也是Microsoft自2027財年起延長資料中心使用年限，使其2026日曆年資本支出預期自約1,900億美元調整為約1,750億美元，而投資計畫並未改變的原因[S-045; S-075]。見第2.1節。

自由實施（Freedom to operate） 一項法律審查，用以判斷某項產品能否在不侵害他人專利的情況下製造與銷售，與某項發明是否具新穎性的問題有別。本報告標示出實作仍具差異化、並可能含有可受保護發明之處，但任何專利結論均須經請求項層級的新穎性與自由實施（freedom to operate）審查。見第6.7節。

玻璃核心基板（Glass-core substrate） 以玻璃而非有機疊層材料作為核心的封裝基板，目標在於實現佈線更密、訊號損耗更低的大型處理器封裝。友達（AUO）於2026年8月31日宣布將與Corning於SEMICON Taiwan共同展出玻璃核心基板封裝技術，此一定位處於研發階段，未見量產或客戶的相關報導[S-052]。見第3.5節。

圖形處理器（Graphics processing unit, GPU） 具備眾多並行運算單元的處理器，最初為圖形處理而打造，如今則是公開市場上由NVIDIA、AMD等業者銷售的主要AI加速器（accelerator）。TrendForce預測以GPU為基礎的系統將占2026年AI伺服器出貨量的69.7%，而GPU與客製ASIC動用的是同一批晶圓代工、封裝、HBM與測試產能[S-008]。見第3.1節。

電網併接（Grid interconnection） 將新的負載或發電機併入電網的程序，包括研究、電網升級與議定條件；申請案在接受研究期間須排隊等候。對AI資料中心而言，這已成為最慢的規劃時鐘：僅ERCOT一地，2026年6月即在追蹤逾438,000 MW的大型負載（large load）申請，而公開紀錄並未給出自申請至送電的典型所需時間[S-032]。見第7.2節。

毛利率與營業利益率（Gross margin and operating margin） 毛利率為營收減去銷貨成本後占營收的比率；營業利益率則另行扣除研究、銷售與管理等營業費用。各企業申報其自有的變體（毛利、營業、分部或調整後），因此本報告係就同一企業在不同時間的利潤率進行研判，絕不跨企業排序。見第5.1節。

指引 企業對特定期間營收、資本支出（capital expenditure）、利潤率或其他財務數字所作的自身預測（即本報告表格中的G類），為最強的一種前瞻性陳述。管理層目標（T）、藍圖日期（R）、產能承諾（C）與情境（S）均屬較弱的類型，絕不被研判為財測指引。見第C.5節。

HBM4與HBM4E 第四代高頻寬記憶體（high-bandwidth memory）及其擴充版後繼產品：JEDEC於2025年4月發布的HBM4標準，訂定2,048位元介面、每座堆疊最高2 TB/s，而Samsung所列的HBM4E則最高達4 TB/s[S-139; S-140]。SK hynix於2026年第2季開始大量出貨HBM4，Micron則預期HBM4E將於2027日曆年進入量產[S-003; S-128]。見第3.2節。

Helios AMD為其Instinct GPU所設計的機櫃級系統（rack-scale system），係透過OCP與Meta共同開發，並以OCP的Open Rack Wide格式為基礎[S-112; S-235]；AMD稱其為「a reference design, not a product for sale」[S-179]。此例顯示一家晶片供應商依營運商所撰寫的機櫃標準，設計其旗艦系統。見第4.5節。

高頻寬記憶體（High-bandwidth memory, HBM） 由DRAM晶粒垂直堆疊而成、置於處理器封裝體內其旁側的記憶體，具備極寬的介面，可提供遠高於一般伺服器記憶體的頻寬。它決定了加速器（accelerator）產出的節奏，因為每一個封裝體唯有在其所有堆疊均完成驗證（qualification）、鍵合與測試之後才能出貨；台積電（TSMC）排定於2028年的14倍光罩尺寸CoWoS，可整合約20座HBM堆疊[S-024]。見第3.2節。

Hopper NVIDIA在Blackwell之前的加速器（accelerator）世代，其中包括自2025年4月起輸往中國須取得許可（licence）的H20，以及H200——後者所屬的產品類別，美國目前就輸往中國者逐案審查[S-102; S-412]。NVIDIA報告2027財年第2季輸往中國的Data Center Hopper產品出貨，占Data Center營收不到1%[S-002]。見第7.1節。

混合鍵合（Hybrid bonding） 一種以直接的銅與氧化物鍵合取代錫球凸塊、將晶粒面對面接合的方式，可實現更密的連接與更薄的堆疊。其應用正擴散至邏輯、記憶體與共同封裝光學（co-packaged optics）：Besi的混合鍵合（hybrid bonding）客戶數自2025年底的15家增至2026年6月底的21家[S-170]。見第3.5節。

混合線性注意力（Hybrid linear attention） 一種模型架構配置，僅在部分層保留快取會隨每個token增長的完整注意力，其餘各層則採用狀態大小固定的線性注意力層。Qwen3.5與Qwen3.8採三層Gated DeltaNet對一層注意力層的比例，Kimi K3則具69層Kimi Delta Attention與24層潛在注意力層，因此其快取僅在約四分之一的層中增長[S-522; S-519; S-489]。見第8.2節。

超大規模雲端業者（Hyperscaler） 以最大規模建設並營運資料中心的少數雲端與網際網路企業之一，例如Alphabet（Google）、Amazon、Meta與Microsoft。其資本計畫訂定了AI供應鏈其餘環節所依循的數量與日期，而Google、Meta與Microsoft在OCP董事會中擁有企業席次[S-331]。見第2.1節。

磷化銦（InP） 一種化合物半導體，光鏈路所用的雷射——包括電吸收調變雷射（electro-absorption modulated laser, EML）——即製作於其上。雷射供應是光學領域中已載明的約束：TrendForce報導Lumentum執行長表示，磷化銦的供需缺口現已超過DRAM與NAND，並指出中國約掌握70%的精煉銦產量[S-037]。見第5.4節。

鍵值快取（KV cache） 一段對話中所有先前token已儲存的鍵與值，語言模型在每一個解碼（decode）步驟均須讀取。其規模隨上下文長度及每一位並行使用者而增長，因此決定長上下文服務所需記憶體量的，並非僅是模型權重，而是它；分組查詢注意力與多頭潛在注意力（multi-head latent attention）即為縮小其規模的設計[S-142; S-146]。依本報告依據開發者組態所作的計算，一段約一百萬token的對話，視設計而定，所需快取自接近200 GB至不足1 GB[S-490; S-529]。見第3.3節與第8.3節。

Kyber NVIDIA的下一代機櫃，據述至2027年將容納576顆Rubin Ultra GPU，並與NVIDIA於2027年全面量產800 VDC的計畫相連結[S-178; S-028]。產業媒體曾報導其延後至2028年，NVIDIA對此有異議，因此Kyber的時程是800 VDC展望的一項訊號[S-186]。見第4.5節。

通道（Lane） 一條鏈路之內的單一電氣或光學訊號路徑；鏈路速率為單一通道速率乘以通道數，因此八條200 Gb/s通道可構成一個1.6 Tb/s模組。通道速率每提高一次，訊號在銅線上可傳輸的距離即縮短，從而限制了可共用同一個銅質垂直擴展網域的加速器（accelerator）數量。見第4.7節。

大型負載（Large load） 規模大到足以使監管機關為其訂定個別併接條件的電網客戶；FERC關於此類負載併接的規則制定，針對的是20 MW以上的負載，而德州則為75 MW以上的負載訂定標準[S-430; S-040]。AI資料中心即屬此一類別，因此為大型負載（large load）所訂的研究程序、成本規則與負載削減（curtailment）義務，決定了它們何時能夠併接。見第7.2節。

液冷 以液體迴路而非空氣將熱自晶片移除，透過冷板（cold plate），或在部分設計中透過兩相流體或浸沒方式進行。此法正成為高階AI系統的標準做法：TrendForce預測2026年將有53%的AI晶片採用液冷，高於2025年的約33%[S-006]。見第4.3節。

長期協議（Long-term agreement (LTA)） 記憶體製造商與客戶之間的多年期供應合約，可能附帶保證金、預付款或價格條件。當記憶體短缺時，長期協議（long-term agreement）決定誰優先獲得供應：SK hynix已與約10家客戶敲定此類協議，而TrendForce報導部分美國雲端業者的協議限制供應商不得調高價格[S-003; S-132]。見第3.2節。

低壓直流（Low-voltage direct current (LVDC)） 在產品標準與裝置法規所界定的低壓範圍內進行的直流配電，上限為1,500 VDC（交流則為1,000 VAC），一如OCP的低壓直流（low-voltage direct current）白皮書（white paper）所述[S-059]。由於800 VDC落在此一範圍之內，AI機房仍需要的是百萬瓦級的已認證產品與保護裝置，而非一個新的電壓等級。見第7.4節。

中壓（Medium voltage） 介於建物低壓供電與高壓輸電之間的配電電壓；OCP的固態變壓器（solid-state transformer）規格（specification）採13.8 kV或34.5 kV輸入[S-054]。將中壓直接轉換為800 VDC，可省去電網併接點與機櫃之間的轉換級數。見第4.4節。

記憶體分層（Memory tiering） 依資料被需用的時間先後，將其配置於數種不同記憶體之上，自處理器旁的HBM到主機DRAM、匯集的Compute Express Link（CXL）記憶體與快閃儲存。在AI服務中，此法將閒置的上下文快取移至較廉價的層級，並於使用前移回，因而增添了保存上下文的空間，卻未增添解碼（decode）所需的頻寬[S-155; S-156]。見第3.3節。

MGX NVIDIA的模組化伺服器與機櫃架構；NVIDIA於2024年10月15日將其GB200 NVL72機櫃設計的部分要素貢獻給OCP，形式為設計貢獻文件（contribution），而非具版本編號的規格（specification）[S-237]。它與Open Rack並列，構成OCP兩條機櫃譜系之一；NVIDIA表示已有逾200家生態系夥伴採用其MGX設計[S-226]。見第4.5節。

Micro LED 為顯示器而開發的微型發光二極體，亦可作為極短距光鏈路的光源。友達（AUO）正在開發一款Micro LED共同封裝光學（co-packaged optics）模組，適用於AI資料中心內最長10公尺的鏈路，此一定位處於研發階段，未見量產或客戶的相關報導[S-052]。見第4.7節。

微縮放（MX）格式（Microscaling formats） 一類數值格式，其中一組數值共用同一個縮放因子，使每一個數值得以4至8位元儲存而幾無精度損失；此類格式出自Microsoft、AMD、Intel、Meta、NVIDIA與Qualcomm作者共同撰寫的一篇論文[S-152]。Kimi K3以32個為一組提供MXFP4權重，並搭配MXFP8活化值；DeepSeek-V3.1則以一種FP8縮放格式訓練，以求與微縮放格式相容[S-489; S-526]。見第8.2節。

混合專家（MoE） 一種模型設計，其中每一個MoE層都包含許多專家子網路，並由路由器將每個token送往其中少數幾個，因此任一token僅動用一小部分參數。DeepSeek-V3每個token在其6,710億個參數中啟用370億個，Kimi K3則在2.8兆個參數中啟用1,040億個[S-528; S-489]；然而每一個專家仍必須儲存於服務該模型的加速器（accelerator）可及範圍之內。見第8.2節。

Mt Diablo（Diablo 400） 由Microsoft、Meta與Google撰寫的OCP電源機櫃規格（specification），將交流至直流的轉換移入側置電源機櫃（sidecar），以±400 VDC向IT機櫃供電，可擴充範圍為「from 800kW to 1MW+」；基礎規格0.7.0版自2026年3月1日生效[S-027]。這是OCP已制定規格的較高電壓路徑，而針對原生800 V的Mt Diablo 2.0則僅經宣布，未載明日期[S-019]。見第4.2節。

多頭潛在注意力（MLA） 隨DeepSeek-V2一併提出的一種注意力設計，就每一個token與每一層儲存一個經壓縮的潛在向量，而非為每一個頭儲存完整的鍵與值；DeepSeek報告其KV快取較其稠密型67B模型縮小93.3%[S-146]。DeepSeek-V3、Kimi K2與GLM-5.3每層儲存512個潛在維度與64個位置維度[S-528; S-524; S-520]。見第8.2節。

NAND快閃記憶體（NAND flash） 在無電力情況下仍能保存資料的記憶體，用於固態硬碟（solid-state drive, SSD）；其速度慢於DRAM，但每位元成本遠為低廉。它構成HBM與伺服器DRAM之下的儲存層級；TrendForce在計算記憶體占雲端業者資本支出（capital expenditure）比重上升時，將其與DRAM併計[S-007]。見第3.2節。

必要請求項（Necessary Claims） 實施規格（specification）必要部分所需的專利請求項，包括選用部分中的必要要素，以及簽署後方取得的請求項；OCP貢獻者依OWFa 0.9所作的承諾僅涵蓋此類請求項[S-388]。規格未予規定之處，例如材料、控制韌體或測試方法，均不在該承諾範圍之內。見第6.6節。

新興GPU雲端業者（Neocloud） 在最大型雲端平台之外，專門出租AI用GPU產能的雲端業者，例如CoreWeave或Nebius。新興GPU雲端業者（neocloud）依少數大型客戶的合約進行建置，並部分以該等合約擔保的債務為產能融資，因而使交易對手風險趨於集中[S-065; S-066]。見第2.2節。

NVLink NVIDIA專有的垂直擴展互連技術，將其GPU相連，使彼此都能讀寫對方的記憶體。NVIDIA透過其NVLink Fusion計畫，依NVIDIA自訂的條件向其他公司的晶片開放；此舉擴大了可連接者的範圍，卻未改變由誰定義該技術[S-244]。見第4.6節。

OCP Accepted 自OCP於2026年終止其OCP Inspired標章後，OCP唯一的產品認可標章；它表示某項產品已依一份經核准的OCP規格（specification）或其他貢獻文件（contribution）受過審查，即使並未貢獻任何設計檔案亦然[S-109; S-333]。它並不代表可取用該設計、取得專利授權（licence），亦不構成出貨量的證據。見第6.1節。

OCP認可支出（OCP-recognized spending） 出自OCP與IDC共同研究的市場衡量指標：依OCP認可規格（specification）所建置的IT基礎設施與解決方案支出，2025年為1,320億美元，至2029年上升至2,950億美元，其中AI應用情境於2029年約為1,550億美元[S-004; S-108]。這是一個與標準相連結的市場，而非OCP Foundation的營收、AI資料中心市場或超大規模雲端業者（hyperscaler）的資本支出（capital expenditure）。見第2.4節。

Open Compute Project（OCP） 由OCP Foundation營運的一個社群，資料中心營運業者與供應商在其中發布開放硬體規格（specification），範圍自伺服器與機櫃，至電力、冷卻與整座設施。它自我描述為常規的形成者，而非標準組織，其規格僅對選擇依其建置或採購者具約束力[S-329]。見第6.1節。

Open Data Center for AI OCP針對AI設施的倡議，於2025年10月啟動，其白皮書（white paper）列出五項優先事項，自參考設計（reference design）與電網解決方案，至服務水準協議、遙測（telemetry）與電力估算[S-181; S-227]。由AMD、Google、Meta與NVIDIA撰寫的設計原則於2026年8月27日達到0.7.0修訂版，仍低於已發布的1.0版[S-001]。見第4.4節。

Open Rack V3（ORv3） OCP的機櫃規格（specification），由Meta與Google的作者於2022年發布，具一條48 V級匯流排（busbar），由電源架以標稱51 VDC供電；它所確定的是IT設備的對接對象，而非機架的外部尺寸[S-180]。它是OCP機櫃已安裝的基準線，其High Power Rack（HPR）變體隨Meta的Catalina系統於2024年支援最高140 kW[S-176]。見第4.5節。

Open Rack Wide（ORW） OCP針對最重型AI系統的雙倍寬機櫃格式；其基礎規格V1.0.0版由Rittal與Meta的作者於2026年4月28日發布，訂定介面與公差，但未規定外部機架尺寸[S-228]。由於依該基礎規格製造的機箱可裝入任何符合規格的機架，機架供應商因而成為可替代者，而以其為基礎建構的系統——例如AMD的Helios——則仍屬各供應商自身的設計[S-179]。見第4.5節。

Open Web Foundation Agreement 0.9（OWFa 0.9） OCP規格（specification）背後的專利授權（licence），依其規定，各貢獻者承諾不主張其實施最終規格（Final Specification）必要部分所需的請求項，或者選擇就該等請求項給予免權利金（royalty-free）授權[S-388; S-047]。它使OCP介面於發布時即為免權利金，而超出必要部分的實作則不在該承諾範圍之內。見第6.6節。

開放權重模型（Open-weight model） 其訓練後參數公開供任何人下載並運行的模型，通常並附模型說明卡、組態檔與一份授權（licence）。訓練資料與程式碼可能仍不公開，因此開放權重與開放原始碼並不相同，而其授權條款自MIT與Apache 2.0以至各種自訂條款[S-526; S-522; S-519]。由於任何營運業者均可提供此類模型的服務，其記憶體設計會形塑整個市場的需求。見第8.1節。

光路交換（Optical circuit switching, OCS） 直接在光纖之間導引光訊號的交換方式，可在不將訊號轉換為電訊號的情況下，改變哪些加速器（accelerator）或機櫃彼此連接。Google自2020年起即在其TPU超級電腦中以生產方式運行光路交換機，而OCP的相關工作仍處於白皮書（white paper）階段[S-293; S-262]。見第4.7節。

退出期間（Opt-Out Period） 自最終文件上傳至OCP Contribution Hub以供指導委員會（Steering Committee）表決時起算的30日期間，貢獻者可在此期間內提出排除通知（Exclusion Notice），將指名的專利自其免權利金（royalty-free）承諾中排除[S-047]。這是唯一的除外機制，且其自上傳之時起算，而非自發布或峰會主題演講之時起算。見第6.6節。

原始設計製造商（Original design manufacturer, ODM） 依客戶要求設計並製造伺服器與機櫃的企業，例如鴻海（Hon Hai／Foxconn）或緯穎（Wiwynn）。原始設計製造商（original design manufacturer）以微薄的利潤率，將加速器（accelerator）、記憶體與冷卻整合為可出貨的機櫃，其營收除隨出貨量變動外，亦隨零組件價格變動[S-022; S-063]。見第5.2節。

委外半導體封裝測試（OSAT） 為他人進行晶片封裝與測試的企業，例如日月光（ASE）與Amkor。委外半導體封裝測試（outsourced semiconductor assembly and test）業者，是先進封裝（advanced packaging）產能得以在單一晶圓代工廠之外成長的途徑之一；TrendForce報導其於2026年每月增添50,000至60,000片晶圓的產能[S-033]。見第3.4節。

PG25 一種冷卻液，為丙二醇於水中以體積比約25%（OCP指引中為24.5%至29.5%）所構成，在Project Deschutes的CDU規格（specification）中與去離子水併列獲准使用[S-216; S-030]。它所依據的是OCP指引而非規格，且已加注的迴路在運送或儲存期間可能於−13至−9 °C結冰[S-216]。見第4.3節。

規劃時鐘（Planning clock） 本報告用以指稱AI基礎設施各層級增添產能所需典型時間的用語：晶片與伺服器以季計，晶圓廠、封裝線與記憶體廠以年計，渦輪機與電網併接則更久。各項時長為本報告的分析，並以註明日期的企業陳述為依據；商業機會與延誤風險，均落在這些時鐘之間的落差。見圖2。

可插拔光模組（Pluggable optical module） 插入交換器前面板的收發模組，可將電訊號轉換為光訊號並反向轉換；它是光鏈路的量產產品，也是可現場維修的預設選項。線性設計移除數位訊號處理器以節省電力；Cignal AI預測2026年1.6 Tb/s乙太網路模組出貨量將超過五百萬個[S-272; S-276]。見第4.7節。

後量子密碼學（Post-quantum cryptography, PQC） 設計用以抵禦未來量子電腦攻擊的密碼演算法，例如NIST於2024年8月發布之首三項後量子標準中所列者[S-373]。美國所訂的具體期限——包括國家安全系統至2030年韌體簽章（firmware signing）須專用CNSA 2.0——使晶片內建的後量子支援能力，成為將運行至2030年代之硬體的採購課題[S-374]。見第6.4節。

電力使用效率（Power usage effectiveness（PUE）） 資料中心的總能源使用量除以其IT設備的能源使用量；1.0即表示冷卻與電力轉換完全不產生額外耗能。監管機關現已為其設定上限，例如新加坡最近一次新增產能招標要求在IT滿載下達1.25或更佳，日本則對2029財年起興建的資料中心要求1.3，惟各項數值僅在相同量測條件下方可相互比較[S-438; S-440]。見第7.2節。

製程節點（Process node） 晶片製造技術的一個世代，以3奈米或2奈米之類的奈米級別標示，在台積電（TSMC）則以其註明日期藍圖上的N2U、A14、A13與A12等名稱標示[S-024; S-084]。新製程節點決定加速器（accelerator）的密度與電力效率，而運行這些製程的晶圓廠，其興建與產能爬坡均需數年。見第3.4節。

提示快取（Prompt caching） 保存一則提示已處理完成的上下文（即其KV快取），使日後重複使用相同上下文的請求無須重新計算。供應商對快取命中的定價遠低於新輸入：DeepSeek就V4.1-Flash在尖峰時段的每百萬輸入token，命中列為0.006美元，未命中則為0.30美元[S-564]；Amazon Bedrock則對Kimi K3支援明示的提示快取[S-542]。已儲存的快取使需求移向主機記憶體與快閃儲存。見第8.4節。

驗證階段 買方就供應商零組件在指名平台上所進行的測試與認可，涵蓋效能、熱表現、功率、韌體與可靠性。此一過程緩慢且成本高昂，因此決定了誰能供應HBM之類的稀缺投入要素；而驗證（qualification）清單的掌握者——例如NVIDIA的DSX Ready計畫——則取得對供應商上市時程的籌碼（leverage）[S-048]。見第5.6節。

機櫃級系統（Rack-scale system） 以整座機櫃為單位設計、出貨並完成驗證（qualification）的AI系統，數十顆加速器（accelerator）在同一個垂直擴展網路（scale-up network）中相連，並內建電力轉換、液冷與網路，例如NVIDIA的Vera Rubin NVL72或AMD的Helios[S-010; S-179]。它使機櫃而非伺服器成為設計與採購的單位。見第4.1節。

RAND與FRAND（RAND and FRAND） 標準必要專利（standard-essential patent）的授權（licence）條件，其內容為合理且非歧視（RAND）或公平、合理且非歧視（FRAND），並可能包含權利金。例如JEDEC要求參與者依RAND條件授權其必要請求項，而OCP貢獻者則承諾免權利金（royalty-free），因此單一產品可同時結合兩種體制[S-391; S-047]。見第6.6節。

參考設計（Reference design） 供應商公開的設計藍圖，由夥伴將其轉化為各自的產品；它既非待售產品，亦非開放規格（specification）。AMD稱其Helios機櫃為「a reference design, not a product for sale」，而其符合某項開放機櫃標準之說，係AMD自身的陳述[S-179]。見第4.5節。

預留協議（Reservation agreement） 在正式訂單之前，先於供應商未來生產中預留位置的合約，通常附帶預付款，例如燃氣渦輪機的生產檔期預留與功率半導體的產能預留[S-301; S-021]。預留情形顯示稀缺設備必須提前多久確保取得：截至2026年6月30日，Siemens Energy持有26 GW的生產檔期預留協議，另有69 GW的正式未出貨訂單（backlog）[S-301]。見第5.5節。

光罩（Reticle） 微影設備單次曝光所能製作圖形的最大面積，約為一顆大型晶片的大小；最大的AI封裝體現已以光罩倍數計量。台積電（TSMC）排定於2028年量產的14倍光罩尺寸CoWoS，在單一基板上接合約10顆大型運算晶粒與20座HBM堆疊[S-024]。見第3.4節。

故障穿越 負載在電網故障依正常程序切除期間，仍能維持併網而不脫離並切換至備用電力的能力。2026年7月PJM系統中約3,800 MW的資料中心負載脫離電網後，PJM即針對大型負載（large load）提出故障穿越（ride-through）標準[S-229; S-061]。見第4.4節。

風險試產（Risk production） 新晶片製程在良率（yield）與參數尚未確定、量產尚未開始之前，以低產量進行的首次生產。台積電（TSMC）將其A14製程的風險試產定於2027年，量產定於2028年[S-084; S-024]。見第3.4節。

安全認證（Safety certification） 由獲認可機構依已發布的安全標準對產品進行測試，並據以取得檢驗人員與保險業者所接受的列名。對800 VDC而言，這是介於一份已獲接受的規格（specification）與一項可安裝產品之間的環節：UL 489I涵蓋最高1,500 VDC的固態斷路器（solid-state circuit breaker），而UL Solutions已於2026年9月24日啟動中壓轉800 VDC電源單元的認證（certification）計畫[S-187; S-188]。見第7.4節。

水平擴展網路（Scale-out network） 將眾多機櫃或pod連成數千顆加速器（accelerator）叢集的網路，以延遲換取傳輸距離與路徑多樣性。其傳輸層正收斂至把每條連線的封包分散於多條路徑的機制，見於Ultra Ethernet規格（specification）與OpenAI的Multipath Reliable Connection（MRC）[S-258; S-259]。見第4.6節。

垂直擴展網路（Scale-up network） 將同一機櫃或pod內的加速器（accelerator）緊密相連，使彼此都能以極低延遲讀寫對方記憶體的網路。它把加速器、交換器與網路卡綁在一起，因此鎖定（lock-in）與議價能力（bargaining power）均落於此處；NVIDIA的NVLink、由聯盟撰寫的UALink與各種以乙太網路為基礎的設計正在此競逐[S-244; S-245; S-246]。見第4.6節。

第二供應來源（Second source） 同一零組件的另一家已通過驗證（qualification）的供應商，使買方得以轉換或分拆其訂單。開放介面使在機櫃、電力、冷卻與網路之間各交接點上驗證第二供應來源（second source）的成本降低，但對記憶體或渦輪機生產檔期之類的稀缺投入要素，並未增添任何產能。見第6.1節。

Security Protocol and Data Model（SPDM） DMTF制定的協定，裝置藉以證明其身分並報告其所運行的韌體，為採購中可信證明（attestation）的基礎。2025年5月的SPDM 1.4.0版加入NIST的後量子演算法，但未納入混合模式；DMTF則於2026年就1.5版的混合模式支援徵求回饋意見[S-370; S-357]。見第6.4節。

側置電源機櫃（Sidecar power rack） 置於IT機櫃旁側的機櫃，內含電力轉換裝置，有時並含儲能，並以較高電壓向IT機櫃供電，從而在IT機櫃內釋出空間供運算使用。OCP的Diablo 400規格（specification）界定了一款±400 VDC側置電源機櫃（sidecar），而供應商已宣布推出800 VDC的側置電源機櫃，例如Schneider Electric正與NVIDIA共同開發、適用於最高1.2 MW機櫃者[S-027; S-195]。見第4.2節。

矽光子（Silicon photonics） 以半導體製程在矽晶圓上製作的光學元件，例如調變器、波導與偵測器；雷射通常另行於磷化銦（indium phosphide）上製作。晶圓代工廠正在需求之前即預售此類產能：Tower Semiconductor已簽訂客戶合約，就2027年的矽光子營收達13億美元[S-292]。見第4.7節。

固態斷路器（Solid-state circuit breaker (SSCB)） 以半導體開關而非移動接點切斷電流的斷路器，速度足以清除直流故障。ABB將其SACE Infinitus描述為全球首款取得IEC認證的固態斷路器（solid-state circuit breaker），而SolarEdge與Infineon針對800 VDC、以碳化矽（silicon carbide）JFET打造的斷路器，則屬一項未載明供貨日期的開發合作[S-197; S-055]。見第4.2節。

固態變壓器（Solid-state transformer (SST)） 一種電力電子轉換器，取代傳統的線頻變壓器與整流器，將中壓交流直接轉換為直流。OCP的SST規格（specification）0.3.0版自2026年6月22日生效，為Google、Microsoft與NVIDIA所撰寫的首版草案，訂定諸如轉換至800 VDC之轉換器效率至少98%等要求，而將電路拓撲留給供應商決定[S-054]。見第4.4節。

階段標籤（Stage label） 本報告用以表示一家企業的角色或一項技術在公開紀錄上進展程度的五個標籤：產品、試點階段、研發、標準與已宣布。標籤依循企業自身所用的動詞，因此在量產或出貨獲得報導之前，計畫維持為已宣布，展示維持為研發。見第B.1節。

供應承諾（採購承諾，Supply commitments） 企業在未來期間向其供應商購買庫存、零組件或產能的合約義務，與其對自身客戶的供應承諾有別。NVIDIA的此項金額在2027財年的一個季度內自1,190億美元增至2,790億美元，「primarily related to the procurement of memory」，顯示加速器（accelerator）需求正沿供應鏈向上傳遞[S-002]。見第2.3節。

開關設備（Switchgear） 在電網、發電廠與資料中心機房之間，用以接通、斷開並保護電路的開關、斷路器與保護設備。電網設備訂單正隨資料中心需求上升：Siemens Energy於2026財年上半年在其電網業務中取得約20億歐元的資料中心相關訂單[S-026]。見第5.5節。

轉換成本（Switching cost） 買方以另一家供應商取代原供應商所須負擔的成本與時間，包括重新驗證（qualification）與重新設計；此項成本高時，買方即受鎖定（lock-in）。共通介面可在其所涵蓋的邊界上降低此項成本，而本報告認為開放標準的價值正在於此。見第6.1節。

單晶片系統（SoC）測試機（System-on-chip tester） 在複雜邏輯晶片——包括AI加速器（accelerator）與CPU——製造與封裝過程中對其進行檢測的自動測試設備。測試機產能係依封裝產能規劃：Advantest將其加速器測試機的規劃與封裝擴產相連結，並正提前其目標，即2027年3月前達5,000台、2028年底或2029年初達10,000台[S-116]。見第3.5節。

照付不議協議（Take-or-pay agreement） 一種供應合約，買方承諾就一段期間內的約定數量付款，無論其是否實際提取。Micron的16份策略客戶協議期間通常自2026年至2030年，均「structured as take-or-pay agreements」，其中規模最大者並訂有價格上下限[S-129]。此類協議使供應商取得可見度、買方取得確保的供應，並將價格風險分配於雙方之間。見第5.3節。

token 語言模型讀取與產生的文字單位，可為一個詞或詞的一部分；服務速度以每秒token數衡量。對話上下文中保留的每一個token都會增加鍵值快取（key-value cache）的量，因此上下文愈長、使用者愈多，記憶體需求即愈高[S-142]。Google於2026年5月在其各項服務介面上每月處理逾3,200兆個token[S-491]。見第3.3節與第8.4節。

總處理效能（Total processing performance (TPP)） 美國出口管制（export control）用以衡量晶片運算能力的指標，用於設定許可（licence）門檻。自2026年1月15日起，輸往中國與澳門、總處理效能低於21,000且DRAM頻寬低於6,500 GB/s的晶片——即NVIDIA H200所屬的類別——均逐案審查[S-412]。見第7.1節。

兩相冷卻 一種液冷方式，冷卻液在晶片上或其附近沸騰，並於他處凝結，藉相變吸收熱量，可在冷板（cold plate）或浸沒槽中進行。在OCP內，此法仍處於指引與展示階段，而含氟的兩相流體則面臨歐盟針對全氟及多氟烷基物質（PFAS）的限制程序[S-220; S-450]。見第7.4節。

Ultra Accelerator Link（UALink） 由UALink Consortium撰寫的垂直擴展互連技術，具記憶體語意（memory semantics），可支援單一pod內最多1,024顆加速器（accelerator）；1.0版於2025年4月發布，Common Specification 2.0版則於2026年4月獲批准[S-245; S-251]。規格（specification）的進度領先於產品：The Register報導1.0版晶片於2026年下半年進入實驗室，產品則於2027年較晚時推出[S-252]。見第4.6節。

Ultra Ethernet Ultra Ethernet Consortium就乙太網路上水平擴展傳輸所制定的規格（specification），1.0版於2025年6月發布，現行版本為2026年7月的1.0.3版[S-256; S-257]。其文本將取得任何必要授權（licence）的責任留給實作者，而一家專利池管理機構已徵求與之對應的專利，此事關係到授權規劃[S-267; S-268]。見第4.6節。

Universal Chiplet Interconnect Express（UCIe） 一項開放的晶粒間介面（die-to-die interface），用於在單一封裝體內連接小晶片（chiplet）；2025年8月發布的UCIe 3.0版支援最高64 GT/s[S-167]。公開的速率階梯使結合不同供應商的晶粒成本降低，而UALink則將其小晶片規格（specification）描述為完全符合該介面[S-168]。見第3.4節。

Universal Quick Disconnect（UQD） OCP的一系列液體接頭，使伺服器的冷板（cold plate）迴路得以連接另一家供應商的機櫃歧管；2020年的1.0修訂版訂定最高工作壓力100 psi，其盲插版本則為UQDB[S-212; S-215]。接頭是不同供應商零件最先無法對接之處，因此OCP將其標準化，而迴路本身的設計則仍屬競爭範疇；第2版目前處於beta草案階段[S-060]。見第4.3節。

已驗證最終使用者（Validated end-user, VEU） 美國的一項出口授權（authorisation），使經核准的外國設施得以在無個別許可（licence）的情況下接收特定項目。美國商務部產業安全局將Samsung China Semiconductor、SK hynix Semiconductor（China）與Intel Semiconductor（Dalian）自該名單中移除，自2025年12月31日起生效，並預期此舉每年將使許可申請增加約1,000件[S-510]。見第7.1節。

Vera Rubin NVIDIA在Blackwell之後的加速器（accelerator）平台，整合於Vera Rubin NVL72之類的機櫃級系統（rack-scale system）之中，NVIDIA將其機櫃設計描述為100%液冷[S-178]。NVIDIA於2026年8月表示該平台正「ramping into full production」，機櫃已在指名的雲端夥伴處運行[S-010]。見第4.5節。

白皮書（OCP） 一種資訊性的OCP文件，記載共通的要求與實務做法，與含有符合性條文、可供產品據以測試的規格（specification）有別；指引的作用方式亦同。OCP在電力與設施方面的大部分內容——例如其低壓直流（low-voltage direct current）與儲能文件——均置於白皮書（white paper）之中，而招標文件僅能將其引為參考資料[S-059; S-205]。見第A.1節。

良率（Yield） 製造單位中通過測試者所占的比率。在堆疊記憶體與大型多晶粒封裝中，良率（yield）偏低會使成本倍增，因為組裝後才發現的一顆不良晶粒即可能使整個封裝體報廢，這也是測試產能隨封裝複雜度一併成長的原因[S-116]。見第3.5節。

C.8 縮寫

1H, 2H 一年的上半年與下半年（first half, second half of a year，如2H 2029）

1Q26, 2Q26 2026年第1季、2026年第2季；其他季度依同一方式表示

1Y, 3Y, 5Y 一年、三年與五年的預測期限（horizon；見Interpretation Registry）

200G, 1.6T 每秒200吉位元與每秒1.6太位元：通道、模組與交換器資料速率的簡寫

2.3D, 2.5D, 3D 先進封裝（advanced packaging）的配置方式：晶粒並列於中介層或橋接結構之上（2.3D、2.5D），或彼此垂直堆疊（3D）

3GPP 第三代合作夥伴計畫（3rd Generation Partnership Project）

400ZR 一類用於資料中心之間鏈路的400 Gb/s同調光學介面

5G, 6G 第五代與第六代行動網路

5G-Advanced 3GPP在6G之前的後期5G版本

AC 交流電（alternating current）

ACIE AI雲端、工業與企業（AI Clouds, Industrial, & Enterprise，NVIDIA Data Center的一個次市場）

AI 人工智慧（artificial intelligence）

AI/ML 人工智慧與機器學習（artificial intelligence and machine learning）

AI-RAN 無線接取網路人工智慧（artificial intelligence in the radio access network）

AIME 美國邀請數學考試（American Invitational Mathematics Examination，一組基準題目）

AMBA Advanced Microcontroller Bus Architecture（Arm）

AMD Advanced Micro Devices

AMI American Megatrends International

AP7 台積電先進後端七廠（TSMC Advanced Backend Fab 7），位於台灣的先進封裝（advanced packaging）廠區

APAC Asia-Pacific

API 應用程式介面（application programming interface）

ASE 日月光投資控股（ASE Technology Holding，Advanced Semiconductor Engineering）

ASHRAE 美國暖氣、製冷與空調工程師學會（American Society of Heating, Refrigerating and Air-Conditioning Engineers）

ASIC 特定應用積體電路（application-specific integrated circuit）

AWS Amazon Web Services

BESS 電池儲能系統（battery energy storage system）

BF16 bfloat16，一種16位元浮點數值格式

BIS 美國商務部產業安全局（Bureau of Industry and Security，U.S. Department of Commerce）

BMC 基板管理控制器（baseboard management controller）

bn 十億（billion；US$ bn；圖表標籤中以B表示，例如$132B）

BoF Birds of a Feather（IETF的一種會議形式）

bp 基點（basis point），即一個百分點的百分之一

C 產能承諾，一種前瞻性陳述類型（已安裝、已規劃、已預留、已簽約或在建）

CAGR 複合年成長率（compound annual growth rate）

CAISI AI標準與創新中心（Center for AI Standards and Innovation，美國NIST）

capex, CapEx 資本支出（capital expenditure）

CBFC 基於信用額度的流量控制（Credit Based Flow Control）

CC BY-ND Creative Commons Attribution-NoDerivatives（授權條款）

CDU 冷卻液分配單元（coolant distribution unit；於Project Deschutes中稱Cooling Distribution Unit）

CEI Common Electrical I/O（OIF的實作協議，如CEI-224G）

CHI Coherent Hub Interface（Arm AMBA）

CHIPS Common Hardware for Interfaces, Processors and Systems (CHIPS Alliance)

CLA 貢獻者授權協議（Contributor License Agreement）

CMS 可組合記憶體系統（Composable Memory Systems，OCP子專案）

CNSA 2.0 Commercial National Security Algorithm Suite 2.0

COUPE Compact Universal Photonic Engine (TSMC)

CoWoS Chip-on-Wafer-on-Substrate（台積電）

CoWoS-L 具局部矽互連橋接（local silicon interconnect bridge）之CoWoS

CPC Colder Products Company

CPO 共同封裝光學（co-packaged optics）

CPU 中央處理器（central processing unit）

CRU 愛爾蘭公用事業管制委員會（Commission for Regulation of Utilities, Ireland）

CSD 標準制定準則（Criteria for Standards Development，IEEE）

CSP 雲端服務供應商（cloud service provider）

CTAM Compliance Tool for Accelerator Management

CXL Compute Express Link

CXMT 長鑫存儲技術（ChangXin Memory Technologies；上市名稱為ChangXin Technology Group）

CY 日曆年（calendar year）

DC 直流（direct current）；在NVIDIA的財報中則指Data Center

DC-MHS Data Center Modular Hardware System

DC-SCM Datacenter-ready Secure Control Module

DDR5 Double Data Rate 5，一個DRAM世代

DLC Direct Liquid Cooled（AMD對直接液冷的用語）

DMTF DMTF（前身為Distributed Management Task Force）

DRAM 動態隨機存取記憶體（dynamic random-access memory）

DSP 數位訊號處理器（digital signal processor）

DSX NVIDIA對其AI工廠（AI factory）參考設計（reference design）及其DSX Ready驗證（qualification）計畫的稱法

DUV 深紫外光（deep ultraviolet，微影）

EAR 出口管理條例（Export Administration Regulations，美國）

ECCN Export Control Classification Number

ECOC European Conference on Optical Communication

EDA 電子設計自動化（electronic design automation）

Edge-MHS 邊緣模組化硬體系統（Edge Modular Hardware System，OCP）

EIA 美國電子工業協會（Electronic Industries Alliance，如19吋EIA機架）

EMIB 嵌入式多晶粒互連橋接（Embedded Multi-die Interconnect Bridge，Intel）

EML 電吸收調變雷射（electro-absorption modulated laser）

EPRI Electric Power Research Institute

EPS 電子封裝學會（Electronics Packaging Society，IEEE）

ERCOT 德州電力可靠度委員會（Electric Reliability Council of Texas）

ESUN Ethernet for Scale-Up Networking

EU European Union

EUV 極紫外光（extreme ultraviolet，微影）

FCSA Foundation Chiplet System Architecture

FERC 聯邦能源管理委員會（Federal Energy Regulatory Commission）

F-gas 含氟溫室氣體（fluorinated greenhouse gas，如歐盟F-gas法規）

FIPS Federal Information Processing Standards

FP4 4位元浮點（4-bit floating point，數值格式）

FP8 8位元浮點（8-bit floating point，數值格式）

FRAND 公平、合理且非歧視（fair, reasonable and non-discriminatory）

FSA 最終規格協議（Final Spec Agreement，OCP）

FY 財年，依各企業自身的年度曆（FY2027或FY27即2027財年）

G 財測指引，一種前瞻性陳述類型

GB 吉位元組（gigabyte）；在GB200、GB300之類的NVIDIA產品名稱中則指Grace Blackwell

GbE 吉位元乙太網路（gigabit Ethernet，如800GbE）

GB/s 吉位元組每秒（gigabytes per second）

Gb/s, Gbps 吉位元每秒（gigabits per second）

GCS 玻璃核心基板（glass-core substrate）

GPM 每分鐘美制加侖數（US gallons per minute）

GPU 圖形處理器（graphics processing unit）

GQA 分組查詢注意力（grouped-query attention）

GTCC 燃氣渦輪機複循環（gas-turbine combined cycle）

GT/s 每秒十億次傳輸（gigatransfers per second）

GW 吉瓦（gigawatt）

H1 一年或一個財年的上半年

H2O Heavy-Hitter Oracle（一種KV快取逐出方法）

HBF High Bandwidth Flash

HBM 高頻寬記憶體（high-bandwidth memory）

HBM4, HBM4E 第四代高頻寬記憶體（high-bandwidth memory）及其擴充版本

HPC 高效能運算（high-performance computing）

HPE Hewlett Packard Enterprise

HPR 高功率機櫃（High Power Rack，ORv3）

IC 積體電路（integrated circuit）

ID 識別編號（identifier，如來源編號）

IDC International Data Corporation

IEA 國際能源總署（International Energy Agency）

IEC International Electrotechnical Commission

IEEE Institute of Electrical and Electronics Engineers

IEEE SA IEEE Standards Association

IETF Internet Engineering Task Force

IFRS 國際財務報導準則（International Financial Reporting Standards）

InP 磷化銦（indium phosphide）

I/O 輸入/輸出（input/output）

INT4 4位元整數（4-bit integer，數值格式）

IOWN Innovative Optical and Wireless Network

IP 智慧財產權；在網路領域則指網際網路協定（intellectual property; in networking, Internet Protocol）

IPR 智慧財產權（intellectual property rights）

IT 資訊技術（information technology）

JEDEC JEDEC Solid State Technology Association

JFET 接面場效電晶體（junction field-effect transistor）

JPY 日圓（Japanese yen）

KRW 韓元（Korean won）

KV 鍵值（key-value，如KV快取）

kV 千伏（kilovolt）

kW 千瓦（kilowatt）

L2 第2層（Layer 2），即網路的資料鏈結層

LBNL Lawrence Berkeley National Laboratory

LED 發光二極體（light-emitting diode，如Micro LED）

LLM 大型語言模型（large language model）

LLR 鏈路層重傳（Link Level Retry）

LMS Leighton-Micali Signatures

L.O.C.K. 分層開源密碼金鑰管理（Layered Open-source Cryptographic Key management，OCP）

LPDDR 低功耗雙倍資料速率DRAM（low-power double data rate DRAM，如LPDDR5X）

LPO 線性可插拔光學（Linear Pluggable Optics）

LRO 線性接收光學（linear receive optics）

LTA 長期協議（long-term agreement）

LVDC 低壓直流（low-voltage direct current）

M-CRPS 共通冗餘電源（Common Redundant Power Supply，一項DC-MHS模組規格）

MGX NVIDIA的模組化伺服器與機櫃架構

MHS 模組化硬體系統（Modular Hardware System，OCP）

ML 機器學習（machine learning）

MLA 多頭潛在注意力（multi-head latent attention）

ML-DSA Module-Lattice-Based Digital Signature Algorithm

ML-KEM Module-Lattice-Based Key-Encapsulation Mechanism

MoE 混合專家（mixture of experts）

MOU 合作備忘錄（memorandum of understanding）

MRC Multipath Reliable Connection

MSA 多源協議（multi-source agreement）

MW 百萬瓦（megawatt）

MX 微縮放（Microscaling，數值格式）

MXFP4, MXFP8 4位元與8位元微縮放浮點格式

NA 數值孔徑（numerical aperture，如低數值孔徑EUV）

NAND NAND快閃記憶體（NAND flash memory，源自NOT-AND邏輯）

NDAA 國防授權法（National Defense Authorization Act，美國）

NERC 北美電力可靠度公司（North American Electric Reliability Corporation）

NFPA National Fire Protection Association

NIC 網路介面卡（network interface card）

NIST 國家標準與技術研究院（National Institute of Standards and Technology，美國）

nm 奈米（nanometre，製程節點級別）

NPU 神經網路處理器（neural processing unit）

NRE 一次性工程（non-recurring engineering）

NSA 國家安全局（National Security Agency，美國）

NT$ 新台幣（New Taiwan dollar）

NVFP4 NVIDIA的4位元浮點格式（NVIDIA 4-bit floating point）

NVL72 NVIDIA的機櫃級系統（rack-scale system），在單一NVLink網域內含72顆GPU

OAI 開放加速器基礎設施（Open Accelerator Infrastructure，OCP）

OAM OCP Accelerator Module

OBMF 開放開機與管理框架（Open Boot and Management Framework，OCP）

OCP Open Compute Project

OCS 光路交換（optical circuit switching），或指光路交換機

ODM 原始設計製造商（original design manufacturer）

OFC 光纖通訊會議暨展覽會（Optical Fiber Communication Conference and Exhibition）

OIF Optical Internetworking Forum

O-RAN 開放無線接取網路（Open Radio Access Network，O-RAN ALLIANCE）

ORv3 Open Rack V3

ORW Open Rack Wide

OSAT 委外半導體封裝測試（outsourced semiconductor assembly and test）

OU Open Rack單位（Open Rack unit）

OWF Open Web Foundation

OWFa Open Web Foundation Agreement

PAR 專案授權申請（Project Authorization Request，IEEE）

PB 拍位元組（petabyte）

PCI 周邊元件互連（Peripheral Component Interconnect）

PCIe PCI Express

PCI-SIG PCI Special Interest Group

PFAS 全氟及多氟烷基物質（per- and polyfluoroalkyl substances）

PFC 優先權流量控制（Priority Flow Control）

PG25 體積比約25%的丙二醇冷卻液（propylene glycol coolant at about 25% by volume）

PJM PJM Interconnection（美國一家區域電網營運機構）

PLDM Platform Level Data Model

PQC 後量子密碼學（post-quantum cryptography）

PRC 中華人民共和國（People's Republic of China）

psi 磅力每平方英寸（pounds per square inch）

P&T7 SK hynix的封裝測試七號廠（SK hynix Package & Test site 7），一座先進封裝（advanced packaging）廠

PUCT 德州公用事業委員會（Public Utility Commission of Texas）

PUE 電力使用效率（power usage effectiveness）

Q1–Q4 第1季至第4季（Q2 FY2027即2027財年第2季）

QCT Quanta Cloud Technology

QLC 四階儲存單元（quad-level cell，每個單元儲存四位元的NAND快閃記憶體）

R 藍圖日期，一種前瞻性陳述類型

RAND 合理且非歧視（reasonable and non-discriminatory）

RAS 可靠性、可用性與可維護性（reliability, availability and serviceability）

RC 發布候選版（release candidate；RC1、RC4即編號候選版）

R&D 研究與開發（research and development）

RDL 重佈線層（redistribution layer）

Rev 修訂版（revision，用於文件版本編號）

RMB 人民幣（renminbi，中國元）

RoCE 融合乙太網路上的遠端直接記憶體存取（RDMA，remote direct memory access over Converged Ethernet）

RPO 剩餘履約義務（remaining performance obligations）

RTL 暫存器傳輸級（register-transfer level）

RU 機架單位（rack unit）

Rx 接收器（receiver）

S 情境或期望，一種前瞻性陳述類型

S.A.F.E. Security Appraisal Framework and Enablement (OCP)

SAI Switch Abstraction Interface

SiC 碳化矽（silicon carbide）

SMIC Semiconductor Manufacturing International Corporation

SNIA Storage Networking Industry Association

SoC 單晶片系統（system on chip）

SOCAMM 小型壓縮附接記憶體模組（small outline compression attached memory module，用於伺服器的LPDDR記憶體）

SoIC System on Integrated Chips（台積電）

SONiC Software for Open Networking in the Cloud

SoW-X System-on-Wafer（台積電）

SPDM Security Protocol and Data Model

SPHBM4 標準封裝高頻寬記憶體（Standard Package High Bandwidth Memory，JEDEC JESD330-4）

SPIL Siliconware Precision Industries

SRA 產能時段預留協議（slot reservation agreement）

SSCB 固態斷路器（solid-state circuit breaker）

SSD 固態硬碟（solid-state drive）

SST 固態變壓器（solid-state transformer）

STAR Market 上海證券交易所科技創新板（Science and Technology Innovation Board of the Shanghai Stock Exchange）

SU 垂直擴展（scale-up，見ESUN的規格文本）

SUE Scale Up Ethernet

SUE-T Scale-Up Ethernet Transport

T 管理層目標，一種前瞻性陳述類型

TAM 總體潛在市場（total addressable market）

TB 太位元組（terabyte）

TBA 待公布（to be announced）

TB/s 太位元組每秒（terabytes per second）

Tb/s 太位元每秒（terabits per second）

TC 技術委員會（Technical Committee，如ASHRAE TC 9.9）

TCS 技術冷卻系統（technology cooling system）

TDP 熱設計功耗（thermal design power）

TEL Tokyo Electron

tn 兆（trillion，如KRW 2,450tn）

TPP 總處理效能（total processing performance）

TPU 張量處理器（Tensor Processing Unit，Google）

TSMC Taiwan Semiconductor Manufacturing Company

TWh 太瓦時（terawatt-hour）

Tx 發送器（transmitter）

UAE 阿拉伯聯合大公國（United Arab Emirates）

UALink Ultra Accelerator Link

UALoE 乙太網路上的UALink（UALink over Ethernet）

UBB Universal Baseboard

UC Berkeley University of California, Berkeley

UCIe Universal Chiplet Interconnect Express

UE8M0 一種8位元、僅含指數的縮放格式，搭配FP8與微縮放格式使用

UEC Ultra Ethernet Consortium

UL UL Solutions；其標準的前綴，例如UL 489I、UL 2877與UL 9540

UMC United Microelectronics Corporation

UPS 不斷電系統（uninterruptible power supply）

UQD Universal Quick Disconnect

UQDB Universal Quick Disconnect，盲插版（blind-mate）

U.S. 美國（United States）

US$ 美元（United States dollar）

USB Universal Serial Bus

VAC 交流電壓（volts alternating current）

VDC 直流電壓（volts direct current）

Ver 版本（version，用於文件版本編號）

VR Vera Rubin（用於VR200、VR300之類的NVIDIA產品名稱）

W17 to W45, W+ ASHRAE設施水等級，各以其水溫上限（°C）命名

WFE 晶圓製造設備（wafer-fabrication equipment）

XMSS 擴展梅克爾簽章方案（eXtended Merkle Signature Scheme）

YMTC Yangtze Memory Technologies

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-002 NVIDIA (26 Aug 2026); S-003 SK hynix (29 Jul 2026); S-010 NVIDIA (26 Aug 2026); S-011 CoreWeave (11 Aug 2026); S-012 GE Vernova (22 Jul 2026); S-014 Alphabet (4 Feb 2026); S-015 Alphabet (29 Apr 2026); S-016 Alphabet (22 Jul 2026); S-021 Infineon Technologies (5 Aug 2026); S-022 Hon Hai (12 Aug 2026); S-024 TSMC (23 Apr 2026); S-025 ASML (15 Jul 2026); S-026 Siemens Energy (29 Jun 2026); S-028 NVIDIA Technical Blog (20 May 2025); S-045 Microsoft (29 Apr 2026); S-048 NVIDIA (21 Sep 2026); S-052 AUO (31 Aug 2026); S-055 SolarEdge and Infineon (9 Sep 2026); S-057 Besi (18 Jun 2026); S-063 Wiwynn (7 Aug 2026); S-064 Wiwynn (7 May 2026); S-065 CoreWeave (2 Mar 2026); S-066 Nebius (17 Jul 2026); S-075 Microsoft (29 Jul 2026); S-084 TSMC (16 Jul 2026); S-102 NVIDIA (28 May 2025); S-112 AMD (24 Feb 2026); S-115 Samsung and Broadcom (25 Jul 2026); S-116 Advantest (29 Jul 2026); S-125 SK hynix (22 Apr 2026); S-128 Micron Technology (24 Jun 2026); S-129 Micron Technology (24 Jun 2026); S-140 Samsung Semiconductor (2026); S-156 NVIDIA (18 Mar 2025); S-170 Besi (23 Jul 2026); S-176 Engineering at Meta (15 Oct 2024); S-178 NVIDIA Blog (13 Oct 2025); S-179 AMD (未註明日期); S-188 UL Solutions (24 Sep 2026); S-195 Schneider Electric (13 Oct 2025); S-197 ABB (13 Oct 2025); S-226 NVIDIA (未註明日期); S-231 Bloom Energy (13 Apr 2026); S-235 AMD Newsroom (14 Oct 2025); S-237 NVIDIA Newsroom (15 Oct 2024); S-244 NVIDIA (18 May 2025); S-259 OpenAI (5 May 2026); S-268 Sisvel (未註明日期); S-292 Tower Semiconductor (13 May 2026); S-301 Siemens Energy (5 Aug 2026); S-307 Vertiv (11 Feb 2026); S-328 ASML (14 Nov 2024); S-483 Wiwynn (7 Aug 2026); S-489 Moonshot AI (Jul 2026); S-490 Alibaba Qwen team (21 Jul 2025); S-491 Google (19 May 2026); S-519 Alibaba Qwen team (12 Aug 2026); S-520 Z.ai (18 Aug 2026); S-522 Alibaba Qwen team (16 Feb 2026); S-524 Moonshot AI (Sep 2025); S-526 DeepSeek-AI (21 Aug 2025); S-528 DeepSeek-AI (26 Dec 2024); S-529 DeepSeek-AI (10 Sep 2026); S-542 Amazon Web Services (18 Sep 2026); S-564 DeepSeek (10 Sep 2026)

標準組織與聯盟. S-001 OCP (27 Aug 2026); S-004 OCP (1 Dec 2025); S-019 OCP (11 Aug 2026); S-020 OCP (10 Mar 2026); S-027 OCP (1 Mar 2026); S-030 OCP (26 Feb 2026); S-047 OCP (14 Aug 2026); S-054 OCP (22 Jun 2026); S-058 JEDEC Solid State Technology Association (Dec 2025); S-059 OCP (2 Sep 2026); S-060 OCP (6 Aug 2026); S-067 OCP (未註明日期); S-109 OCP (11 May 2026); S-139 JEDEC (16 Apr 2025); S-158 CXL Consortium (18 Nov 2025); S-161 OCP (19 Dec 2025); S-165 OCP (12 Feb 2026); S-167 UCIe Consortium (5 Aug 2025); S-168 Ultra Accelerator Link Consortium (7 Apr 2026); S-180 OCP (22 Sep 2022); S-181 OCP (13 Oct 2025); S-183 OCP (14 Sep 2023); S-187 ANSI / UL (22 Oct 2025); S-205 OCP (未註明日期); S-207 OCP (13 Oct 2025); S-212 OCP (4 Sep 2020); S-215 OCP (4 Sep 2020); S-216 OCP (未註明日期); S-217 OCP (6 Aug 2026); S-220 OCP (22 Sep 2026); S-227 OCP (Apr 2026); S-228 OCP (28 Apr 2026); S-245 UALink Consortium (8 Apr 2025); S-246 OCP (9 Feb 2026); S-249 OCP (13 Jun 2026); S-251 UALink Consortium (7 Apr 2026); S-256 Ultra Ethernet Consortium (11 Jun 2025); S-257 Ultra Ethernet Consortium (16 Jul 2026); S-258 Ultra Ethernet Consortium (16 Jul 2026); S-262 OCP (Apr 2026); S-267 Ultra Ethernet Consortium (28 Jan 2026); S-272 Linear Pluggable Optics MSA (未註明日期); S-329 OCP (未註明日期); S-331 OCP (未註明日期); S-333 OCP (未註明日期); S-357 DMTF (24 Sep 2026); S-362 OCP (18 Apr 2024); S-363 CHIPS Alliance (未註明日期); S-364 OCP (31 Jul 2025); S-365 CHIPS Alliance (15 Oct 2025); S-370 DMTF (25 May 2025); S-388 Open Web Foundation (未註明日期); S-391 JEDEC Solid State Technology Association (未註明日期)

政府與監管機關. S-005 IEA (16 Apr 2026); S-032 ERCOT (18 Jun 2026); S-040 Texas Legislature (2025); S-042 PJM Inside Lines (13 Aug 2026); S-061 PJM Inside Lines (10 Sep 2026); S-229 PJM Inside Lines (11 Aug 2026); S-373 NIST (13 Aug 2024); S-374 U.S. National Security Agency (Apr 2024); S-412 Federal Register (BIS) (15 Jan 2026); S-414 Federal Register (BIS) (12 Nov 2025); S-430 Federal Energy Regulatory Commission (22 Jul 2026); S-438 IMDA (1 Dec 2025); S-440 METI (20 May 2026); S-450 European Chemicals Agency (未註明日期); S-484 European Commission (9 Nov 2017); S-510 U.S. Department of Commerce (2 Sep 2025)

研究機構. S-053 Nature Electronics (19 Aug 2026); S-136 Gholami et al. (21 Mar 2024); S-137 Patel et al. (30 Nov 2023); S-142 Meta AI (31 Jul 2024); S-146 DeepSeek-AI (7 May 2024); S-152 Bita Darvish Rouhani et al., arXiv (Oct 2023); S-155 Qin et al. (24 Jun 2024); S-293 Jouppi et al. (Apr 2023)

市場研究. S-006 TrendForce (17 Aug 2026); S-007 TrendForce (25 Aug 2026); S-008 TrendForce (20 Jan 2026); S-009 TrendForce (18 Mar 2026); S-017 TrendForce (30 Apr 2026); S-036 TrendForce (18 Sep 2026); S-106 IDC for OCP (12 Nov 2025); S-108 OCP / IDC (22 Jan 2026); S-132 TrendForce (9 Jul 2026); S-276 Cignal AI (7 Jan 2026)

媒體與法律評論. S-033 TrendForce (15 Jun 2026); S-037 TrendForce (6 Aug 2026); S-186 Tom's Hardware (6 Jul 2026); S-252 The Register (7 Apr 2026)

附錄D

來源查核登錄表

本報告所引用的每一項來源，均載明其原始網址、其原文、其所支持的內容與其無法據以確立的內容；並列出十四項主要研判（reading），以及各項背後的證據、推論邏輯與適用邊界。

適用讀者： 希望以來源檢驗論述的專家審閱者、法律與政府分析人員、研究人員及投資專業人士。

本附錄是本報告可受檢驗之處。正文所引用的635項來源各有一張卡片：載明其來源編號、原始公開網址、以該來源原語言呈現的簡短逐字摘錄、該來源非英文者的英文譯文、該來源所支持的內容、其無法據以確立的內容，以及本報告引用該來源的每一處。在各卡片之前，詮釋登錄表列出本報告的14項主要研判（reading），以及各項所依據的證據。本附錄回答的問題是：就任一陳述而言，來源究竟說了什麼？

D.1 如何閱讀來源卡片

每一張卡片均遵循同一條證據鏈。 該鏈自來源出發，先至其確切文字，必要時再至譯文，繼而至本報告自其取得的事實命題，最後至該命題的適用邊界。此一鏈條使讀者得以重建本報告中的任一項發現：自第一頁的某一數字，至深入闡述該數字的章節，至呈現該數字的圖或表，至來源編號，再至原始網址與該來源的原文。

- **表頭。** 來源編號（S-001 …）、出版者，以及出版者所載的文件標題。
- **紀錄列。** 來源類別、發布機構的司法管轄區、原語言、發布日期，以及本報告閱讀該來源的日期。
- **原始網址。** 最具持久性的官方網址；凡穩定的到達頁面與直接文件兩者兼具者，均一併列出。
- **來源原文。** 足以證明該論點的最短段落，以該來源原語言逐字複製，並以綠色斜體印出。各別段落分別印出，絕不併接。取自表格的數值，連同其所出自的表格與列次一併以正體字呈現，因其屬轉錄而非引述。
- **英文譯文。** 來源非英文者，置於原文正下方，並標示為出版者的官方譯文或APEX譯文。
- **支持的研判與限制。** 該來源所支持的內容，依其自身的效力呈現（「expects」、「plans」、「targets」等語均照原文保留），以及其無法據以確立的內容。
- **引用於。** 依憑該來源的每一節、每一圖、每一表與每一項展望，均附連結；另列證據包含該來源的詮釋登錄表條目（A-01 …）。
- **狀態。** 除卡片另有載明外均為現行：較早的財測指引或預測其後經修訂、文件已被較新版本取代，或某一來源的範圍與另一來源不同。另以一行附註說明變動內容。

來源編號具穩定性。 編號依本報告首次引用各來源的順序指派，自第一頁起算，並為後續版次所沿用，因此自本版引用的編號將持續指向同一份文件。綠色斜體僅用於一種用途：自來源複製的文字。本報告自身的分析絕不以綠色呈現。635項來源之中有15項非英文；其摘錄保留原語言，譯文置於其下。

企業來源就該企業所述內容具權威性。 申報文件或業績發布是企業自身業績、財測指引、產能、合約與藍圖的最佳證據。但其本身並不足以確立市場總量、全產業短缺、競爭者地位或未來的採用情形；各企業卡片的限制列即載明其無法承載上述何者。同一規則適用於市場研究——本報告將其標示為其發布機構的估計值或預測值——亦適用於業界媒體，後者僅於無任何第一手來源可用時方予採用。

D.2 詮釋登錄表

A-01 AI基礎設施的限制正從晶片擴及系統

證據	[S-002; S-017; S-037; S-018; S-032]
推論邏輯	對加速器（accelerator）的需求持續上升 → 買方與供應商提前鎖定記憶體、封裝與雷射（NVIDIA的採購承諾升至2,790億美元，主要用於記憶體） → 渦輪機產能時段與電網併接成為最慢的環節 → 交付取決於數項投入要素中最慢者，而非僅取決於加速器
替代解釋	各項限制因單一產品世代轉換而同時堆積，並隨產能追上而消解；或需求放緩，使所有層級同時獲得紓解。
適用邊界	並未為各項限制排序、量化供給缺口，亦未主張任何單一限制對每一項部署均具約束力。
驗證訊號	記憶體、雷射、渦輪機與併網同時出現延伸至2027年以後的產能配額、預訂或長期協議（long-term agreement）。
否證訊號	需求仍在成長之際，數個層級的交期（lead time）同時縮短，且加速器（accelerator）再度成為唯一被報導的限制。
預測期限	當前
適用讀者	財務；產業；研發；政府

A-02 記憶體正同時成為資本配置上的限制與效能上的限制

證據	[S-007; S-003; S-045; S-046; S-002]
推論邏輯	合約價格上漲，供應商並透過長期協議（long-term agreement）配置產能 → 記憶體占雲端資本支出（capex）的比重升高（依TrendForce預測，2026年為47%、2027年為68%） → 買方上調資本支出或變更組態 → 記憶體條件不僅影響系統效能，更形塑預算與合約
替代解釋	價格飆升隨2027年起新增DRAM與HBM產能開出而回歸正常。
適用邊界	並未確立記憶體在每一項部署中均為最大瓶頸；TrendForce的各項占比均為預測值；緯穎（Wiwynn）的採購代理模式屬營收基礎的變動，其本身並非短缺的證據。
驗證訊號	DRAM與HBM占CSP資本支出（capex）的比重上升；長期協議（long-term agreement）、保證金與預付款擴及更多客戶。
否證訊號	合約價格持續逐季下跌，且HBM4與HBM4E廣泛可自多家供應商取得。
預測期限	一年
適用讀者	財務；產業；研發

A-03 電力與冷卻已成為與晶片一併作出的架構決策

證據	[S-005; S-006; S-019; S-028; S-030; S-029]
推論邏輯	AI伺服器功率密度於2020年至2025年間上升十一倍 → 48 V配電與氣冷在密度最高的機櫃中已達其極限 → 液冷（liquid cooling）與±400或800 VDC於設計階段即被選定 → 電力與散熱供應商進入平台的驗證（qualification）途徑
替代解釋	效率提升或低功耗設計減緩機櫃功率的上升，使48 V與氣冷在更大比例的機隊中仍屬可行。

A-03 電力與冷卻已成為與晶片一併作出的架構決策

適用邊界	並未主張800 VDC將於任何特定年度成為廣泛的多供應商量產規範，亦未主張氣冷已然過時；OCP是將800 VDC呈現為與交流並存的一種選項。
驗證訊號	出現具認證保護機制的多供應商800 VDC量產硬體；液冷（liquid cooling）滲透率達到TrendForce所預測的水準。
否證訊號	次世代旗艦機櫃仍以48 V與氣冷出貨，或800 VDC於2028年之後仍為單一供應商的選項。
預測期限	三年
適用讀者	產業；研發；智慧財產權；創新

A-04 即使運算能力已可取得，電網併接與電力設備仍可能左右部署時程

證據	[S-032; S-042; S-012; S-018; S-011; S-043]
推論邏輯	大型負載（large load）的併接申請超出電網依伺服器時程所能併接的範圍 → 業者增設審查研究程序與負載削減（curtailment）條款，且渦輪機產能時段提前數年即已售出 → 已簽約容量等待送電啟用 → 營收實現時程隨電網與設備的時鐘而定
替代解釋	表後（behind the meter）發電、儲能與彈性併接條件，使場址得於電網升級完成之前即行送電。
適用邊界	並未預測具名專案將發生延誤；TrendForce的2030年供需缺口為預測值；已簽約電力並不等於已送電電力。
驗證訊號	併網審查研究期間延長、負載削減（curtailment）條件增加、渦輪機未出貨訂單（backlog）上升，以及已簽約電力與已啟用電力之間的落差擴大。
否證訊號	資料中心需求仍在上升之際，併網交期（lead time）縮短且渦輪機未出貨訂單（backlog）下降。
預測期限	三年
適用讀者	政府；產業；財務

A-05 客製晶片正在GPU之外興起，尚未取代GPU

證據	[S-008; S-009; S-016; S-002; S-115]
推論邏輯	大型買方為自身工作負載設計加速器（accelerator） → ASIC占AI伺服器出貨量的比重上升（依TrendForce預測，2026年為27.8%） → GPU營收在絕對金額上仍持續成長 → 兩條路徑競逐同樣的記憶體、封裝與電力
替代解釋	商用GPU平台擴大其軟體與系統的領先優勢，使ASIC占比停滯；或ASIC在推論領域更快取代GPU。
適用邊界	各項占比均為TrendForce的預測值；Samsung與Broadcom所提延續至2030年逾2,000億美元的數字，係一份合作意向書之下的共同估計金額，而非營收財測指引。
驗證訊號	客製加速器（accelerator）的對外銷售範圍擴大，例如客戶資料中心內的TPU系統；且ASIC占比朝TrendForce所估2030年近40%的水準移動。
否證訊號	後續預測下調ASIC占比，或主要客製專案遭延後或取消。
預測期限	三年

A-05 客製晶片正在GPU之外興起，尚未取代GPU

適用讀者	產業；財務；研發

A-06 開放介面在可藉多供應商驗證降低轉換成本的交接點上最為重要

證據	[S-019; S-020; S-047; S-228; S-048]
推論邏輯	買方共同撰寫介面規格（specification）→ 實施該等規格所需請求項的免權利金（royalty-free）承諾於最終規格（Final Specification）階段生效 → 第二供應來源得以依該文本進行建置並取得驗證（qualification） → 轉換成本（switching cost）在該交接點下降，而稀缺投入要素的配置仍落在任何規格之外
替代解釋	平台擁有者的驗證（qualification）計畫成為實質關卡，因此開放文本對採購（procurement）的改變有限。
適用邊界	規格（specification）發布並不保證採用；OCP無法創造記憶體產能、渦輪機產能時段或電網併接。
驗證訊號	同一介面上出現來自互不相關供應商的OCP認可產品；採購（procurement）文件按版本引用OCP規格（specification）。
否證訊號	所貢獻的設計僅隨貢獻者自身的零組件與驗證（qualification）清單出貨。
預測期限	三年
適用讀者	產業；智慧財產權；創新；政府

A-07 資本結構正成為基礎設施風險的一部分

證據	[S-005; S-044; S-049; S-002; S-065; S-066]
推論邏輯	投資規模超出資產負債表所能承載 → 債務、股權、預付款、保證與以合約為擔保的貸款為產能提供資金 → 曝險集中於少數買方與放款人 → 需求或信用衝擊將不僅經由訂單、亦經由融資傳導
替代解釋	營運現金流與投資同步成長，因此外部融資對最大買方而言仍屬次要。
適用邊界	並未評估任何公司的信用品質，亦未提出投資建議；各項融資結構互異，不予加總。
驗證訊號	擔保型與以合約為擔保的融資額度增加；保證與預付款的成長速度超過資本支出（capex）。
否證訊號	最大買方以營運現金流支應其建置，同時對外發行規模下降。
預測期限	一年
適用讀者	財務；政府；產業

A-08 地緣政治限制已直接進入企業的規劃假設

證據	[S-010; S-002; S-050; S-051; S-503; S-412; S-485; S-486; S-111; S-487; S-504]
推論邏輯	美國出口規則與中國的限制措施制約特定加速器（accelerator）與設備對中國的銷售 → NVIDIA在其2027財年的每一份展望中均未假設任何來自中國的Data Center運算營收，AMD將對中國的MI308銷售先自財測指引中移出、其後再度納入，KLA與Applied Materials並於2025年公布所估計的規則影響 → 政策限制可成為財測指引中明示的規劃假設
替代解釋	許可規定變動使銷售恢復，該項假設隨之自展望中消失；或企業吸收相關限制而不予載明，一如AMD與KLA最新展望對中國隻字未提，可能即屬此情形。
適用邊界	該項假設所涵蓋者為NVIDIA展望期間內的Data Center運算營收；並非指NVIDIA預期全公司來自中國的營收為零。
驗證訊號	更多企業於財測指引中載明因司法管轄區而異的假設；新的許可條件或內容規定進入申報文件。
否證訊號	NVIDIA的展望再度納入來自中國的Data Center運算營收，或其他企業的財測指引在出口規則趨嚴之際仍對該等規則隻字未提。

A-08 地緣政治限制已直接進入企業的規劃假設

預測期限	當前
適用讀者	政府；財務；產業

A-09 台灣的角色正從「組裝」這一簡略代稱，轉向系統整合、先進封裝、光學與驗證

證據	[S-022; S-024; S-052; S-178; S-163]
推論邏輯	機櫃級系統（rack-scale system）必須以整體為單位進行整合與測試 → 台灣系統整合商擴產Vera Rubin機櫃與CPO交換器，同時台積電（TSMC）排定CoWoS與COUPE各階段進程，並有一處800 VDC場址正在高雄興建 → 技術知識與價值集中於封裝、機櫃整合與驗證（qualification），同時新增產能亦在美國成長
替代解釋	整合業務的利潤率持續微薄，價值歸於平台擁有者與零組件供應商，而非台灣的系統整合商。
適用邊界	友達（AUO）的Micro LED CPO與玻璃核心基板（glass-core substrate）工作處於研發階段：量產、客戶、設計導入與營收均未獲確立。
驗證訊號	機櫃層級的驗證（qualification）與測試於台灣系統整合商處進行；CPO交換器量產獲得確認；友達（AUO）的模組通過客戶驗證。
否證訊號	整合與測試移往平台擁有者自身的計畫，或移往台灣企業並無角色的廠區。
預測期限	三年
適用讀者	政府；產業；財務；研發

A-10 下一波研發與智慧財產權機會，坐落於尚未解決的實作、材料、控制、測試與服務邊界

證據	[S-053; S-054; S-059; S-055; S-047; S-060]
推論邏輯	介面透過規格（specification）而趨於共通 → 實施最終規格（Final Specification）所需的請求項於發布時即以免權利金（royalty-free）方式授權（licence） → 差異化移往介面如何建置、控制、測試與維修 → 研究與專利請求項隨之集中於該處
替代解釋	規格（specification）轉而規定實作細節，壓縮該空間；或專有平台使介面本身維持封閉。
適用邊界	並非每一項未解問題都構成專利機會；任何專利結論之前，均須進行請求項層級的新穎性與自由實施（freedom to operate）審查。
驗證訊號	已發布的規格（specification）訂定需求，卻將拓樸、材料與控制留待開放，並有數家公司圍繞該等規格提出實作層級的工作成果。
否證訊號	規格（specification）固定了控制演算法、流體配方或轉換器拓樸，使文本之外幾無可供差異化之處。
預測期限	三年
適用讀者	研發；智慧財產權；創新

A-11 各層級依不同時鐘擴張，延誤風險即坐落於彼此之間的落差

證據	[S-010; S-003; S-023; S-024; S-012; S-018]
推論邏輯	晶片與伺服器以季為單位擴產 → 記憶體晶圓廠與封裝產線需時一至數年 → 渦輪機與電網併接所需時間更長 → 依快時鐘下單的產能須等待慢時鐘上的投入要素，而能彌合該落差者即取得籌碼（leverage）
替代解釋	模組化電力、儲能與場內發電縮短慢時鐘；或需求放緩至快時鐘不再超前於慢時鐘的程度。
適用邊界	各時鐘的時長為本報告的分析；企業所提日期為計畫、藍圖與產能里程碑，仍有待執行。

A-11 各層級依不同時鐘擴張，延誤風險即坐落於彼此之間的落差

驗證訊號	企業里程碑（2027年初的龍仁、2028年的14倍光罩尺寸CoWoS、2029年下半年的印第安納州HBM、2030年30 GW的渦輪機產出）如期達成，而需求仍超出其供應。
否證訊號	電網併接與電力設備的交期（lead time）縮短至伺服器週期的時間尺度。
預測期限	五年
適用讀者	財務；產業；政府

A-12 稀缺使議價能力移轉至掌握稀缺投入要素、驗證、場址、能源與資本者

證據	[S-003; S-021; S-012; S-022; S-063; S-048]
推論邏輯	稀缺投入要素透過長期協議（long-term agreement）與產能或時段預訂進行配置 → 持有者得以訂定條件、預付款與優先順序 → 利潤率較低的組裝業務轉嫁成本或變更其營收基礎 → 價值移往仍屬稀缺之處
替代解釋	新產能到位，預訂失效，定價權回歸買方。
適用邊界	各項利潤率均為各公司自身的衡量方式，用以例示價值池所在，並非跨公司的獲利能力排名；緯穎（Wiwynn）的採購代理模式屬營收基礎的變動，其本身並非短缺的證據。
驗證訊號	附帶預付款的預訂協議增加；更多系統整合商就稀缺零組件採行成本轉嫁或代理模式。
否證訊號	預訂協議遭取消或向下重新議定，且供應商利潤率在數量上升之際仍下滑。
預測期限	一年
適用讀者	財務；產業；智慧財產權

A-13 更精簡的模型使記憶體需求在各層級之間移動，而非減少需求

證據	[S-518; S-529; S-489; S-491; S-003; S-129; S-135]
推論邏輯	開放權重模型（open-weight model）設計使每個token的KV快取在各次發布之間以大幅倍數下降（DeepSeek 2026年各次發布先降至十分之一，繼而降至四分之一；DeepSeek-V4.1-Flash為每個token 890位元組） → 單一token的成本下降，長上下文（long context）、推理與代理因而變得負擔得起 → token量在封閉與開放模型中同樣倍增（以Google作為使用量的代理指標，一年內約為七倍） → 記憶體需求自每位使用者的HBM容量，轉向頻寬、超大型模型的記憶體占用，以及主機記憶體與快閃記憶體中成本較低的快取層級，而HBM供給則持續短缺
替代解釋	一旦最精簡的設計承載大部分流量，效率提升即超越使用量成長，記憶體需求因而下降；或代理的採用趨於平緩，使用量成長隨之放慢。
適用邊界	係依據開發者自身的數字，以及本報告自己公布組態所作的計算；並未衡量各項設計在實際服務流量中的組成，亦未確立任何記憶體製造商的數量。
驗證訊號	token量上升，同時每個token的快取量下降；記憶體製造商提及上下文儲存；為KV快取卸載而生的企業級SSD與主機記憶體需求，其成長速度超過HBM容量。
否證訊號	token量仍在成長之際，HBM或伺服器DRAM合約價格連續兩季下跌。
預測期限	三年
適用讀者	財務；產業；研發；政府

A-14 企業數字如今附帶明示的界線，在被解讀為需求之前必須先予查核	
證據	[S-483; S-063; S-494; S-129; S-500; S-010; S-487]
推論邏輯	記憶體價格上漲、出口規則變動，且大型合約分階段議定 → 企業改變其數字所涵蓋的內容（代理條件、用於回收成本的附加費、取或付合約、未納入最終條件尚在議定中之已承諾數量的未出貨訂單（backlog）、排除或納入中國的展望）→ 公布的營收、未出貨訂單與財測指引，與零組件的實體流動漸行漸遠 → 解讀需求時必須先查核每一項數字的界線
替代解釋	該等界線僅屬一次性調整，隨價格回歸正常而消失，公布的數字再度與出貨量同步。
適用邊界	各項實例取自企業自身的陳述，而非問卷調查；本報告並未量化被排除或納入的金額，企業亦多未予揭露。
驗證訊號	更多企業採行代理條件、附加費或取或付合約，或於財測指引中載明因司法管轄區而異的假設。
否證訊號	企業將較早期間依單一基礎重新表述，或量化被排除的金額，使各項數字得以直接比較。
預測期限	一年
適用讀者	財務；產業；政府

D.3 來源卡片

各卡片依來源編號順序排列，亦即首次引用的順序。正文中的每一編號均連結至其卡片，而每一張卡片的「引用於」列則回連至引用該來源之處。

S-001 Open Compute Project Foundation — Open Data Center: Next-Generation ML Infrastructure Design Principles, Revision 0.7.0
標準組織 · Global · 英文 · 27 August 2026 · 閱讀 26 September 2026
https://www.opencompute.org/documents/open-data-center-spec-revision-0-7-0-4-pdf
“Collaborative 2027–2028 guidance unified across AMD, Google, Meta, and Nvidia, generalizing the scope for industry AI/ML systems.”
支持的研判. Open Data Center Design Principles Revision 0.7.0 (effective 27 August 2026) by AMD, Google, Meta and NVIDIA gives 'Collaborative 2027–2028 guidance' for industry AI/ML systems.
無法據以確立. A design specification below 1.0; its version history dates 0.7.0 to 23 June 2026, and it does not show adoption by operators.
引用於: 第一頁; 第節 第1.1節, 第4.4節, 第6.1節, 第6.2節, 第9.2節, 第A.1節, 第C.7節; 展望 16; 表 20, 21

S-002 NVIDIA Corporation — CFO Commentary on Second Quarter Fiscal 2027 Results
企業新聞稿 · U.S. · 英文 · 26 August 2026 · 閱讀 27 September 2026
https://s201.q4cdn.com/141608511/files/doc_financials/2027/Q227/Q2FY27-CFO-Commentary.pdf
“Our commitments increased from $119 billion last quarter to $279 billion, primarily related to the procurement of memory.”
“ACIE revenue increased 138% from a year ago and 25% sequentially driven by end-demand from AI natives, enterprises, and sovereign customers”
“Shipments of Data Center Hopper products to China during the quarter were less than 1% of Data Center revenue.”
支持的研判. NVIDIA's CFO reports Q2 FY2027 commitments up from US$119 billion to US$279 billion, mainly for memory; Hopper China shipments below 1% of Data Center revenue; guarantees capped at US$105 billion.
無法據以確立. Company disclosures; commitments are obligations to buy, not a measure of shortage, and guarantees are phased contingent obligations, not spending.
引用於: 第一頁; 第節 第1.1節, 第1.3節, 第2.2節, 第2.3節, 第3.1節, 第3.2節, 第5.6節, 第7.1節, 第C.7節; 圖 3, 7, 24; 表 7, 8, 14, 24 · **研判:** A-01, A-02, A-05, A-07, A-08

S-003 SK hynix Inc. — SK hynix Announces 2Q26 Financial Results
企業新聞稿 · Korea · 英文 · 29 July 2026 · 閱讀 27 September 2026
https://news.skhynix.com/en/q2-2026-business-results/
“In a market environment where customer demand exceeds supply capabilities, the ability to deliver requested volumes in a timely manner has emerged as a core business competitiveness.”
“Mid-to-long-term investment plans—including the recently announced P&T7 advanced packaging facility, M17 NAND production base, and the development of a new semiconductor cluster—will be executed in phases based on customer demand and investment efficiency.”
“As AI evolves into agentic forms that perform complex tasks on behalf of users and expands across various services, the underlying demand base for memory is broadening.”
“The company has finalized Long-Term Agreements (LTAs) with around 10 customers, including key strategic partners, and is continuing further discussions with major industry clients.”
支持的研判. SK hynix describes customer demand exceeding supply capabilities, reports a 76% 2Q26 operating margin on KRW 79.3 trillion revenue, LTAs with around 10 customers, and HBM4 mass shipments.
無法據以確立. Company statements; the demand remark covers the memory market in general, not HBM or SK hynix alone, and no shortfall size is given.
引用於: 第節 第1.1節, 第1.2節, 第1.3節, 第1.5節, 第3.2節, 第3.3節, 第5.1節, 第5.6節, 第5.7節, 第5.8節, 第6.1節, 第6.7節, 第7.5節, 第8.4節, 第8.6節, 第9.1節, 第9.2節, 第9.3節, 第B.1節, 第B.2節, 第C.7節; 圖 2, 3, 15, 24, 25, 26, 32; 表 3, 7, 9, 10, 13, 19, 20, 23, 25; 展望 1 · **研判:** A-02, A-11, A-12, A-13

S-004 Open Compute Project Foundation — OCP Adoption page: "Updated Market Studies Show OCP Community Impact Is Accelerating Propelled by AI!"
標準組織 · Global · 英文 · 1 December 2025 · 閱讀 27 September 2026
https://www.opencompute.org/about/ocp-adoption
“The recently completed study found that total spending on OCP-recognized IT infrastructure and solutions will grow to US$295 billion in 2029, up from US$132 billion in 2025, reflecting a compound annual growth rate (CAGR) of 22.2%.”
“The deployment of AI infrastructure is predicted to experience robust growth, with total spending projected to reach around $155 billion by 2029.”
支持的研判. OCP's adoption page reports an IDC study (1 December 2025) projecting OCP-recognized infrastructure spending from US$132 billion in 2025 to US$295 billion in 2029, with AI around US$155 billion.
無法據以確立. Forecast of equipment bought to OCP-recognized specifications, members or not; not OCP revenue, total AI-infrastructure spending or capex; servers (35.7% CAGR) outgrow power and racks.
引用於: 第一頁; 第節 第2.4節, 第9.3節, 第C.2節, 第C.4節, 第C.7節, 第E.2節; 圖 9; 表 2, 34, 36; 展望 11

S-005 International Energy Agency (IEA) — Executive summary – Key Questions on Energy and AI
政府 · Global · 英文 · 16 April 2026 · 閱讀 27 September 2026
https://www.iea.org/reports/key-questions-on-energy-and-ai/executive-summary
直接文件: https://iea.blob.core.windows.net/assets/3179f7f8-01f6-4dd6-bffa-c9f7b73f1dc9/KeyQuestionsonEnergyandAI.pdf
“Data centre investments have grown too large to be funded from company balance sheets alone, and large amounts of funding from capital markets will be critical for their buildout.”
“The largest technology companies are contributing to a surge in data centre investment, as their capital expenditure exceeded USD 400 billion in 2025 – and is expected to jump by another 75% in 2026.”
“Social acceptability is also a growing issue, as communities push back against data centre projects”
“Measured per individual task, the energy efficiency of AI is improving at a rate unprecedented in energy history. Software and hardware advances have resulted in the energy use per AI task dropping by at least an order of magnitude annually in recent years.”
“However, new energy-intensive AI applications are increasingly being launched and used, such as those for video generation, reasoning and agentic tasks. These kinds of tasks can consume hundreds or thousands of times more energy per query than simple text generation.”
“Our updated projections see electricity consumption from data centres roughly doubling from 485 TWh in 2025 to 950 TWh in 2030, accounting for around 3% of global electricity demand by that date.”
支持的研判. The IEA projects global data-centre electricity rising from 485 TWh in 2025 to 950 TWh in 2030 (about 3% of demand) and says investments have outgrown company balance sheets.
無法據以確立. Base Case projections under IEA assumptions; the AI-focused share for 2025 is inferred from growth multiples, not stated.
引用於: 第一頁; 第節 第1.1節, 第1.3節, 第1.5節, 第2.1節, 第2.2節, 第2.5節, 第5.6節, 第7.3節, 第8.4節, 第9.1節, 第9.3節, 第C.2節, 第E.2節, 第E.8節; 圖 3, 11; 展望 8, 9; 表 20, 34, 38 · **研判:** A-03, A-07

S-006 TrendForce — Liquid Cooling Becomes Standard for High-End AI Infrastructure, with Penetration Expected to Reach 53% in 2026, Says TrendForce
市場研究 · Taiwan · 英文 · 17 August 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260817-13183.html
“Liquid cooling penetration among AI chips is expected to increase from approximately 33% in 2025 to 53% in 2026.”
“Liquid cooling penetration among AI chips is projected to rise to 53% in 2026 and approach 60% in 2027.”
支持的研判. TrendForce projects liquid cooling on 53% of AI chips in 2026 (about 33% in 2025), nearing 60% in 2027; it describes all-liquid Vera Rubin and names a sole heat-spreader supplier.
無法據以確立. Market-research forecasts; the denominator is AI chips, not servers or facilities, and supplier status is as of August 2026.
引用於: 第一頁; 第節 第1.2節, 第1.3節, 第4.3節, 第C.7節, 第E.2節; 圖 3, 18, 26; 表 1, 13, 20, 36; 展望 3 · **研判:** A-03

S-007 TrendForce — Memory Prices Soar; DRAM and NAND Flash to Account for 68% of Major CSP CapEx in 2027, Says TrendForce
市場研究 · Taiwan · 英文 · 25 August 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260825-13198.html
“Nevertheless, HBM contract prices could still rise by 70–140% in 2027.”
“TrendForce estimates that DRAM and NAND Flash combined will account for 47% of CSPs' total CapEx in 2026, with their share rising further to 68% in 2027.”
“Server DRAM—a key procurement category for CSPs—saw contract prices rise by a cumulative 64% in 2H25, with a further jump of approximately 270% expected in 2026.”
支持的研判. TrendForce estimates DRAM and NAND at 47% of major CSPs' capex in 2026 and 68% in 2027 (capex +98%, +50%), with server DRAM contract prices up about 270% in 2026.
無法據以確立. Market-research estimates for unnamed major CSPs, not company guidance; it does not say how fast a price-led share would reverse.
引用於: 第一頁; 圖 3, 13; 第節 第1.3節, 第1.5節, 第2.1節, 第3.2節, 第5.3節, 第8.4節, 第9.1節, 第9.6節, 第C.7節, 第E.11節; 表 18, 20, 34, 37, 41; 展望 1 · **研判:** A-02

S-008 TrendForce — Global AI Server Shipments Forecast to Grow Over 28% YoY in 2026, with a Rising Share of ASIC-Based Systems, Says TrendForce
市場研究 · Taiwan · 英文 · 20 January 2026 · 閱讀 27 September 2026 · **預測已修訂**
https://www.trendforce.com/presscenter/news/20260120-12887.html
“GPUs will remain the leading category, accounting for 69.7% of shipments.”
“The share of ASIC-based AI servers is expected to reach 27.8% by 2026, the highest since 2023, as North American firms like Google and Meta expand their own ASIC efforts.”
支持的研判. TrendForce forecasts 2026 AI-server shipments up more than 28%, with GPU-based systems at 69.7% and ASIC-based systems at 27.8%, the highest ASIC share since 2023.
無法據以確立. Forecast of unit-shipment shares, not revenue or accelerator counts; the growth figure was later raised to nearly 31% (August 2026).
預測已修訂. The 2026 shipment-growth forecast of over 28% was raised to nearly 31% in August 2026 [S-013]; the 2026 accelerator shares on this card were not revised there.
引用於: 第節 第1.3節, 第2.3節, 第3.1節, 第5.6節, 第C.7節, 第E.2節; 圖 12, 33; 表 37 · **研判:** A-05

S-009 TrendForce — NVIDIA Expands Product Portfolio to Address AI Training and Inference, Responding to an ASIC Push from CSPs, Says TrendForce
市場研究 · Taiwan · 英文 · 18 March 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260318-12976.html
"TrendForce reports that as CSPs like Google and Amazon increase their internal chip development, ASIC-based AI servers are forecast to represent 27.8% of all AI server shipments in 2026. This percentage is expected to grow to nearly 40% by 2030."
支持的研判. TrendForce forecasts ASIC-based AI servers at 27.8% of all AI-server shipments in 2026, rising to nearly 40% by 2030, as CSPs such as Google and Amazon expand in-house chips.
無法據以確立. Market-research forecast of unit-shipment shares, not revenue or compute; it does not show ASICs displacing GPU volumes in absolute terms.
引用於: 第節 第1.3節, 第3.1節, 第9.1節, 第C.7節; 圖 12; 展望 10; 表 20 · **研判:** A-05

S-010 NVIDIA Corporation — NVIDIA Announces Financial Results for Second Quarter Fiscal 2027
企業新聞稿 · U.S. · 英文 · 26 August 2026 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-second-quarter-fiscal-2027
直接文件: https://investor.nvidia.com/news/press-release-details/2026/NVIDIA-Announces-Financial-Results-for-Second-Quarter-Fiscal-2027/default.aspx
"Announced the NVIDIA Vera Rubin platform is ramping into full production with racks running at partners including CoreWeave, Google Cloud, Microsoft Azure, Oracle Cloud Infrastructure and Nebius."
"Revenue is expected to be $108.0 billion, plus or minus 2%. NVIDIA is not assuming any Data Center compute revenue from China in its outlook."
"Revealed that NVIDIA Spectrum™-6 switch systems — supporting both pluggable and co-packaged optics as part of the NVIDIA Vera Rubin platform — are arriving across the world's gigascale AI factories."
支持的研判. NVIDIA reports Q2 FY2027 Data Center revenue of US$89.0 billion and Vera Rubin 'ramping into full production'; its US$108.0 billion Q3 outlook assumes no China Data Center compute revenue.
無法據以確立. Company results and one-quarter outlook; the China assumption covers Data Center compute revenue in the outlook period only, not all China revenue.
引用於: 第一頁; 第節 第1.1節, 第1.2節, 第1.3節, 第2.3節, 第3.1節, 第4.1節, 第4.5節, 第4.6節, 第4.7節, 第5.1節, 第6.7節, 第7.1節, 第9.2節, 第C.7節; 圖 2, 7, 26; 表 6, 7, 8, 13, 14, 19, 20, 21, 24, 32, 34, 35, 38; 展望 4, 5, 6, 10; 附錄A · **研判:** A-08, A-11, A-14

S-011 CoreWeave, Inc. (Investor Relations) — CoreWeave Reports Strong Second Quarter 2026 Results
企業新聞稿 · U.S. · 英文 · 11 August 2026 · 閱讀 27 September 2026
https://investors.coreweave.com/news/news-details/2026/CoreWeave-Reports-Strong-Second-Quarter-2026-Results/default.aspx
"Expanded active power by nearly 500 MWs to reach 1.5 GW"
"Grew total contracted power to approximately 3.7 GW while further diversifying portfolio of providers and expanding powered land footprint"
支持的研判. CoreWeave reports about 1.5 GW of active power and about 3.7 GW of contracted power at 30 June 2026, and a revenue backlog of about US$104 billion.
無法據以確立. Company measures; contracted power is not energised capacity, and revenue backlog is broader than RPO; the release does not explain the gap's causes.
引用於: 第一頁; 第節 第1.3節, 第2.2節, 第2.5節, 第5.3節, 第9.3節, 第C.7節; 圖 8; 展望 9; 表 24 · **研判:** A-04

S-012 GE Vernova Inc. — GE Vernova reports second quarter 2026 financial results and raises 2026 financial guidance
企業新聞稿 · U.S. · 英文 · 22 July 2026 · 閱讀 27 September 2026
https://www.gevernova.com/news/press-releases/ge-vernova-reports-second-quarter-2026-financial-results-raises-2026-financial
直接文件: https://www.gevernova.com/sites/default/files/gev_webcast_pressrelease_07222026.pdf
"To meet this demand, we remain on track to deliver 20 GW of annual gas turbine output in the third quarter of 2026, with 24 GW in 2028, and we are implementing actions to produce 30 GW in 2030."
"Gas Power equipment backlog and slot reservation agreements grew from 100 to 116 GW; now anticipate reaching at least 125 GW by year-end 2026"
支持的研判. GE Vernova is on track for 20 GW a year of gas-turbine output in Q3 2026 and 24 GW in 2028; gas backlog and slot reservations reached 116 GW.
無法據以確立. Company plans and order measures; reservations are not firm orders, and output is manufacturing capacity, not delivered generation or data-centre supply.
引用於: 第一頁; 第節 第1.1節, 第1.2節, 第1.3節, 第1.5節, 第5.5節, 第5.6節, 第5.8節, 第6.1節, 第6.7節, 第7.3節, 第9.1節, 第9.2節, 第9.3節, 第9.5節, 第9.6節, 第C.4節, 第C.7節; 圖 2, 3, 24, 32; 表 7, 9, 10, 13, 19, 29, 34, 35; 展望 8 · **研判:** A-04, A-11, A-12

S-013 TrendForce — AI Server Shipments Forecast Raised to Nearly 31% YoY in 2026 as 90% Surge in CSP CapEx Fuels Infrastructure Expansion, Says TrendForce
市場研究 · Taiwan · 英文 · 3 August 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260803-13161.html
"TrendForce has therefore revised its 2026 AI server shipment forecast upward from 28% to nearly 31% YoY."
"Surging investment in AI infrastructure is expected to drive the combined CapEx of the world's nine largest CSPs up by approximately 90% YoY in 2026, according to TrendForce's latest AI server research."
支持的研判. TrendForce raised its 2026 AI-server growth forecast from 28% to nearly 31%, expects top-nine CSP capex up about 90% to over US$886.7 billion in 2026, about US$1.3 trillion in 2027.
無法據以確立. Market-research estimates, not company guidance; capex is TrendForce's aggregate for nine named CSPs, and shipments are units, not value.
引用於: 第節 第1.1節, 第2.1節, 第2.3節, 第3.1節, 第E.2節, 第E.6節; 表 2, 41; 圖 12, 33

S-014 Alphabet Inc. (Investor Relations) — Alphabet 2025 Q4 Earnings Call
企業新聞稿 · U.S. · 英文 · 4 February 2026 · 閱讀 27 September 2026 · **財測指引已修訂**
https://abc.xyz/investor/events/event-details/2026/2025-Q4-Earnings-Call-2026-Dr_C033hS6/default.aspx
"I do expect to go through the year in a supply-constrained way."
"For the full year 2026, we expect CapEx to be in the range of $175 billion to $185 billion, with investments ramping over the course of the year."
"Approximately 60% of our investment in 2025, and it's going to be fairly similar in 2026, went towards machines; so, the servers. And then 40% is what you would refer to as long duration assets, which is our data centers and networking equipment."
支持的研判. Alphabet guided 2026 capex of US$175–185 billion in February 2026, after US$91.4 billion in 2025, and put about 60% of 2025 investment into servers, expecting 2026 to be similar.
無法據以確立. Company guidance later revised; backlog and supply statements are management's own, not independent measures of demand.
財測指引已修訂. Its 2026 capital-expenditure guidance of US$175–185 billion was raised to US$180–190 billion in April [S-015] and to US$195–205 billion in July 2026 [S-016]. The other statements on this card stand as made.
引用於: 第節 第1.1節, 第2.1節, 第2.2節, 第C.2節; 圖 5, 8

S-015 Alphabet Inc. (Investor Relations) — Alphabet 2026 Q1 Earnings Call
企業新聞稿 · U.S. · 英文 · 29 April 2026 · 閱讀 27 September 2026 · **財測指引已修訂**
https://abc.xyz/investor/events/event-details/2026/2026-Q1-Earnings-Call-2026-nW8kCrBAKS/default.aspx
"Moving to investment, we are updating our full year 2026 CapEx guidance range to $180 to $190 billion, up from our previous estimate of $175 to $185 billion, to now include investment related to the acquisition of Intersect, which closed in March."
"Google Cloud's backlog nearly doubled sequentially, reaching $462 billion at the end of the first quarter."
支持的研判. Alphabet raised its 2026 capex guidance in April 2026 to US$180–190 billion from US$175–185 billion to include Intersect-related investment, and reported Google Cloud backlog of US$462 billion.
無法據以確立. Guidance as of April 2026, raised again in July; backlog is Google Cloud's own measure.
財測指引已修訂. Its 2026 range of US$180–190 billion was raised to US$195–205 billion in July 2026 [S-016].
引用於: 第節 第1.1節, 第2.1節, 第2.2節, 第C.2節; 圖 5, 8

S-016 Alphabet Inc. (Investor Relations) — Alphabet 2026 Q2 Earnings Call
企業新聞稿 · U.S. · 英文 · 22 July 2026 · 閱讀 27 September 2026
https://abc.xyz/investor/events/event-details/2026/2026-Q2-Earnings-Call-2026-GgTAq7Is0z/default.aspx
"And given the supply constraint environment, we plan to expand the use of third-party capacity in Q3 as a bridging strategy while we build out more internal capacity."
"The increase in the range is primarily due to an acceleration in the delivery of capacity to meet growing demand."
"As we previously shared, we continue to expect our CapEx to increase significantly in 2027, and we'll provide more details at a later date."
"Moving to investments, we are updating our full year 2026 CapEx guidance range to $195-205 billion, up from our previous estimate of $180-190 billion."
支持的研判. Alphabet raised 2026 capex guidance to US$195–205 billion in July 2026, mainly to accelerate capacity delivery, and continues to expect capex to increase significantly in 2027.
無法據以確立. Management guidance, not spending; no 2027 figure is given, and backlog is Alphabet's own measure.
引用於: 第節 第1.1節, 第1.2節, 第1.3節, 第2.1節, 第2.2節, 第2.4節, 第3.1節, 第9.1節, 第9.6節, 第B.2節, 第C.2節; 圖 4, 5, 8; 表 24, 34 · **研判:** A-05

S-017 TrendForce — AI Competition Turns into a Supply Chain Arms Race, Tightening Advanced Packaging and 3nm Capacity, Says TrendForce
市場研究 · Taiwan · 英文 · 30 April 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260430-13028.html
"Industry leader NVIDIA, leveraging its experience and strong supply chain control, was among the first to identify tightening capacity and preemptively secured large volumes of 4/3 nm wafer capacity, CoWoS packaging, as well as key materials such as T-glass, substrates, PCBs, HBM, and SSDs."
"CoWoS has remained in short supply since 2023, prompting customers to seek alternative capacity sources."
支持的研判. TrendForce says CoWoS has been in short supply since 2023 and that NVIDIA pre-emptively secured large volumes of 4/3 nm wafers, CoWoS, T-glass, substrates, PCBs, HBM and SSDs.
無法據以確立. Market-research commentary; it names NVIDIA as the company that secured supply early, not buyers generally, and gives no volumes.
引用於: 第節 第1.1節, 第3.4節, 第C.7節; 圖 3 · **研判:** A-01

S-018 Siemens Energy AG — Siemens Energy Q2 FY26 Pre-Close Group Call
企業新聞稿 · Germany · 英文 · 30 March 2026 · 閱讀 27 September 2026
https://assets.siemens-energy.com/dam/70568add-abb0-4cb7-91bc-b41d01144be8/2026-03-30---Q2-FY26-Pre-close-group-call-script_final-pdf_Original%20file.pdf
“On the grid side, we see a once-in-a-generation investment cycle over the next fifteen years, driven by the need to connect renewables and replace aging assets.”
“With respect to Gas Turbine capacity, we are sold-out until fiscal year 2028, with fiscal year 2029 slots filling up very quickly and in high demand.”
“We expect an elevated gas-turbine market at least until 2035.”
支持的研判. Siemens Energy's head of IR said in March 2026 gas-turbine capacity is sold out until FY2028, FY2029 slots are filling fast, and grids face a fifteen-year investment cycle.
無法據以確立. Pre-close call remarks, not formal guidance; fiscal years end 30 September; market views are the company's own.
引用於: 第節 第1.1節, 第1.3節, 第1.5節, 第5.5節, 第9.4節, 第B.2節; 圖 2, 3; 表 10, 29 · **研判:** A-01, A-04, A-11

S-019 Open Compute Project Foundation — Powering the Next Era of AI: How Google, Microsoft and Nvidia Are Standardizing and Accelerating the Industry Transition to LVDC
標準組織 · Global · 英文 · 11 August 2026 · 閱讀 27 September 2026
https://www.opencompute.org/blog/powering-the-next-era-of-ai-how-google-microsoft-and-nvidia-are-standardizing-and-accelerating-the-industry-transition-to-lvdc
“The published Mt Diablo +/-400V sidecar spec enables current solutions, while a Mt Diablo 2.0 will be coming later to support native 800V.”
“800 VDC is not a replacement for existing AC infrastructure. It provides an additional option that can coexist with existing AC systems.”
“The goal is not a single prescriptive design. The goal is a common set of interfaces and system requirements that lets operators deploy safely, suppliers build interoperable products and the industry move faster together.”
“We are leveraging an ecosystem from other industries—including connectors, protection devices, breakers, solid-state breakers that clear faults in milliseconds, and advanced monitoring systems.”
“We are engaging in parallel with UL Solutions, NFPA, IEEE, and IEC to advance the safety certifications and regulatory frameworks that are prerequisites for global deployment.”
支持的研判. Google, Microsoft and NVIDIA's August 2026 OCP post presents 800 VDC as an addition to AC, says Mt Diablo 2.0 'will be coming later', and reports safety-certification engagement.
無法據以確立. Company authors' statement of direction: it gives no Mt Diablo 2.0 date, names no certified products, and 'developing' partners are not announced or shipping products.
引用於: 第節 第1.1節, 第1.2節, 第1.3節, 第1.5節, 第4.2節, 第4.4節, 第5.6節, 第6.2節, 第6.7節, 第7.4節, 第9.1節, 第9.2節, 第9.6節, 第A.1節, 第C.7節; 表 1, 5, 13, 21, 23, 24; 圖 17, 24, 26; 展望 2, 13 · **研判:** A-03, A-06

S-020 Open Compute Project Foundation (OCP Networking Project) — The OCP ESUN 1.0 Specification has been released
標準組織 · Global · 英文 · 10 March 2026 · 閱讀 26 September 2026
https://www.opencompute.org/blog/the-ocp-esun-10-specification-has-been-released
“the ESUN initiative has grown to include over 175 participating companies, up from the 12 founding members”
“ESUN1.0 specifies use of congestion management (Priority Flow Control and Credit Based Flow Control) and link level reliability (LLR: Link Level Retry) protocols”
支持的研判. OCP's 10 March 2026 post announces the ESUN 1.0 release, with over 175 participating companies, specifying PFC and CBFC congestion management and Link Level Retry.
無法據以確立. Participation count as of March 2026; participation is not implementation, and the post does not show multi-vendor products or conformance testing.
引用於: 第節 第1.1節, 第4.6節, 第6.1節, 第6.7節, 第C.7節; 圖 26; 展望 16; 表 20, 21 · **研判:** A-06

S-021 Infineon Technologies AG — Q3 FY 2026 concluded with record sales driven by strong AI business; further significant increase in revenue and margin expected in Q4 FY 2026
企業新聞稿 · Germany · 英文 · 5 August 2026 · 閱讀 27 September 2026
https://www.infineon.com/press-release/2026/infpr202608-125
直接文件: https://www.infineon.com/row/public/documents/corporate/press/2026/infpr202608-125e.pdf
“These agreements cover a cumulative revenue volume of a high single-digit billion euro amount and feature also certain prepayments.”
“Several leading customers across the AI data center ecosystem have entered into multi-year capacity reservation agreements with us or are currently in negotiations for such.”
支持的研判. Infineon says AI data-centre customers have signed or are negotiating multi-year capacity reservation agreements, cumulatively a high single-digit billion euros with certain prepayments; Q3 FY2026 Segment Result Margin was 19.1%.
無法據以確立. Company statements; some agreements are still in negotiation, amounts are cumulative over several years, and counterparties are unnamed.
引用於: 第節 第1.1節, 第4.2節, 第5.1節, 第5.4節, 第9.3節, 第B.2節, 第C.7節; 圖 3, 25; 表 7, 9, 10, 28 · **研判:** A-12

S-022 Hon Hai Precision Industry Co., Ltd. (Hon Hai Technology Group / Foxconn) — Hon Hai Technology Group (Foxconn) Announces Second Quarter 2026 Financial Results
企業新聞稿 · Taiwan · 英文 · 12 August 2026 · 閱讀 27 September 2026
https://www.foxconn.com/en-us/press-center/press-releases/latest-news/2095
“Shipments of high-performance networking products are moving ahead and full-year revenue from 800G and higher switches is expected to double, while co-packaged optics (CPO) all-optical switches are on track for mass production and shipment in the third quarter.”
“Looking ahead to next year, Chiang said customer demand for AI production capacity will still be very strong.”
“With Vera Rubin (VR) racks entering mass production in the third quarter, overall AI server rack revenue is expected to increase quarter-on-quarter, and shipments are projected to achieve high double-digit growth.”
支持的研判. Hon Hai says Vera Rubin racks enter mass production in Q3 2026, expects AI rack shipments to more than double in 2026, and has CPO switches on track for Q3.
無法據以確立. Company expectations for its own products; not an ODM-market forecast, and it gives no 2027 shipment forecast beyond 'very strong' demand.
引用於: 第節 第1.1節, 第1.2節, 第1.3節, 第4.5節, 第4.7節, 第5.1節, 第5.2節, 第5.4節, 第5.5節, 第5.6節, 第5.7節, 第5.8節, 第6.7節, 第9.2節, 第B.2節, 第C.7節; 表 6, 7, 9, 13, 19, 20, 27; 圖 25, 26, 32; 展望 4, 5, 12 · **研判:** A-09, A-12

S-023 SK hynix Inc. — SK hynix Holds Groundbreaking Ceremony for HBM Production Base in Indiana, ' Beginning a New Future for US-Korea AI'
企業新聞稿 · Korea · 英文 · 28 August 2026 · 閱讀 27 September 2026
https://news.skhynix.com/en/groundbreaking-ceremony-in-indiana/
“The Indiana fab, with a total expected investment of over $4 billion, is scheduled to open its cleanroom by October 2028 and initiate mass production of next-generation HBM in the second half of 2029.”
“Notably, the Indiana fab marks SK hynix's first HBM production base established in the United States, featuring an operation framework closely integrated with production in South Korea.”
支持的研判. SK hynix's Indiana fab, its first U.S. HBM production base (over US$4 billion), is scheduled to open its cleanroom by October 2028 and mass-produce next-generation HBM from 2H 2029.
無法據以確立. Company schedule at the August 2026 groundbreaking; dates may slip, and capacity and customers are not stated.
引用於: 第節 第1.1節, 第1.2節, 第3.2節, 第5.7節, 第5.8節, 第9.3節, 第9.4節, 第9.5節, 第B.1節, 第B.2節; 圖 2, 3, 25, 32; 表 3, 9, 10, 13, 19, 23, 25 · **研判:** A-11

S-024 Taiwan Semiconductor Manufacturing Company Limited (TSMC) — TSMC Debuts A13 Technology at 2026 North America Technology Symposium
企業新聞稿 · Taiwan · 英文 · 23 April 2026 · 閱讀 27 September 2026
https://pr.tsmc.com/english/news/3302
“A 14-reticle size CoWoS, capable of integrating approximately 10 large compute dies and 20 HBM stacks, is slated for production in 2028.”
“In addition, A13 delivers increased power efficiency and performance gains through design-technology co-optimization, and is scheduled to enter production in 2029, one year after A14.”
“TSMC-COUPE™ is set to reach a key milestone with a true co-packaged optics solution using COUPE on substrate beginning production in 2026.”
支持的研判. TSMC's April 2026 symposium release slates 14-reticle CoWoS (about 10 compute dies, 20 HBM stacks) for 2028, larger CoWoS and SoW-X for 2029, and COUPE on substrate from 2026.
無法據以確立. Roadmap dates, not delivered capacity; COUPE efficiency and latency gains are TSMC's own claims.
引用於: 第節 第1.1節, 第1.2節, 第1.3節, 第3.4節, 第3.5節, 第4.7節, 第5.6節, 第5.7節, 第5.8節, 第6.7節, 第9.1節, 第9.3節, 第9.5節, 第9.6節, 第C.2節, 第C.4節, 第C.7節; 圖 2, 3, 25, 32; 表 1, 4, 6, 9, 10, 13, 19, 20, 26, 35; 展望 12 · **研判:** A-09, A-11

S-025 ASML Holding N.V. — ASML reports €9.3 billion total net sales and €2.9 billion net income in Q2 2026
企業新聞稿 · Netherlands · 英文 · 15 July 2026 · 閱讀 27 September 2026
https://www.asml.com/en/news/press-releases/2026/q2-2026-financial-results
“Given the business dynamics discussed above, we now expect total net sales for 2026 to be between €43 billion and €45 billion, with a gross margin between 54% and 56%.”
“Based on this momentum, we are planning to add 30% to our 2026 low NA EUV capacity of around 65 for 2027, and we are investigating to increase capacity with another 30% for 2028.”
“At our next Capital Markets Day, which will be held on June 10, 2027, we will update our longer-term views to reflect the market and technology dynamics since our last Capital Markets Day.”
支持的研判. ASML plans to add 30% to 2026 low-NA EUV (~65) and DUV immersion (~130) capacity for 2027, is investigating another 30% for 2028, and updates longer-term views 10 June 2027.
無法據以確立. Company plans; the 2028 step is only under investigation, and the 30% applies to low-NA EUV and DUV immersion, not High-NA.
引用於: 圖 2, 25, 32; 第節 第3.5節, 第5.7節, 第5.8節, 第7.1節, 第9.2節, 第9.3節, 第B.2節, 第C.5節, 第C.7節; 表 7, 9, 10, 14, 19, 26

S-026 Siemens Energy AG — Siemens Energy Q3 FY26 Pre-Close Group Call
企業新聞稿 · Germany · 英文 · 29 June 2026 · 閱讀 27 September 2026
https://assets.siemens-energy.com/dam/630d3b4f-2d26-4803-804e-b479007691c5/2026-06-29_Q3FY26_Pre-Close-Call_Transcript-pdf_Original%20file.pdf
“From a structural perspective, the market remains supply-constrained rather than demand-constrained, with strong backlog visibility and a growing number of commitments including reservation agreements.”
“I mean on the one hand, as we are booked out onto ' 28 and let's say, '29 and 2030 is filling up very quickly.”
支持的研判. Siemens Energy's head of IR said in June 2026 the gas-turbine market 'remains supply-constrained', it is ' booked out onto '28', and reservations convert to orders within 6–12 months on average.
無法據以確立. Pre-close call remarks, not guidance; the 110–120 GW a year market view is Siemens Energy's own estimate.
引用於: 圖 2, 3, 24; 第節 第1.2節, 第5.5節, 第5.8節, 第7.3節, 第9.1節, 第9.3節, 第9.4節, 第C.4節, 第C.7節; 表 9, 10, 19, 20

S-027 Open Compute Project Foundation — Diablo 400 Project: Rack and Power (v0.7.0)
標準組織 · Global · 英文 · 1 March 2026 · 閱讀 26 September 2026
https://www.opencompute.org/documents/ocp-specification-diablo-400-v0-7-0-final-pdf
“The disaggregated power rack can be scalable and updated to meet various power requirements from 800kW to 1MW+”
“the disaggregated power rack (also known as sidecar) provides a flexible and evolving infrastructure solution”
支持的研判. Diablo 400 Rack and Power 0.7.0 (effective 1 March 2026; Microsoft, Meta, Google) specifies a ±400 VDC disaggregated power rack (sidecar) scalable from 800 kW to 1 MW+.
無法據以確立. A base specification below 1.0 that dropped the 'HVDC' term; it records no listed protection hardware or multi-vendor products.
引用於: 第節 第1.2節, 第4.2節, 第4.4節, 第7.4節, 第9.1節, 第A.1節, 第C.7節; 圖 17; 表 5, 20, 21, 32; 展望 2, 13

S-028 NVIDIA Technical Blog — NVIDIA 800 V HVDC Architecture Will Power the Next Generation of AI Factories
企業新聞稿 · U.S. · 英文 · 20 May 2025 · 閱讀 27 September 2026
https://developer.nvidia.com/blog/nvidia-800-v-hvdc-architecture-will-power-the-next-generation-of-ai-factories
“Traditional 54 V in-rack power distribution, designed for kilowatt (KW)-scale racks, isn't designed to support the megawatt (MW)-scale racks coming soon to modern AI factories.”
“With lower current, thinner conductors can handle the same load, reducing copper requirements by 45%.”
“Full-scale production of 800 VDC data centers will coincide with NVIDIA Kyber rack-scale systems in 2027”
支持的研判. NVIDIA said in May 2025 that full-scale 800 VDC data-centre production will coincide with Kyber racks in 2027, and claims 45% less copper and up to 5% higher efficiency.
無法據以確立. Vendor roadmap and design claims; the 2027 timing is NVIDIA's own statement, not an industry adoption date, and the gains are system-level vendor figures.
引用於: 第節 第1.2節, 第4.2節, 第4.5節, 第9.2節, 第C.6節, 第C.7節; 圖 17; 表 5, 19, 24; 展望 2 · **研判:** A-03

S-029 TrendForce — NVIDIA's 800V Power Rack to Debut as an Optional Configuration for Vera Rubin, with Broader Adoption Expected in the Rubin Ultra Generation, Says TrendForce
市場研究 · Taiwan · 英文 · 25 June 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260625-13121.html
“Widespread deployment is likely to follow in 2028.”
“The platform is expected to be ready for customer shipments in 3Q26 and will be offered as an optional configuration—rather than a standard feature—for customers deploying the Vera Rubin platform.”
支持的研判. TrendForce expects NVIDIA's 800 V power rack to be ready for customer shipments in 3Q26 as an optional Vera Rubin configuration, with widespread deployment likely in 2028.
無法據以確立. Market-research forecast, not NVIDIA's own schedule; it gives no shipment volumes or customer names.
引用於: 第節 第1.2節, 第4.2節; 圖 17; 表 5; 展望 2, 13 · **研判:** A-03

S-030 Open Compute Project Foundation — Project Deschutes: Data Center Facilities 1.0.0 Specification
標準組織 · Global · 英文 · 26 February 2026 · 閱讀 27 September 2026
https://www.opencompute.org/documents/ocp-specification-deschutes-v1-0-pdf
“This specification defines the requirements for the 2MW Next-Generation Cooling Distribution Unit (CDU), intended for high-power, dense IT pods, including AI/ML ASICs.”
“IT dP available (psi) 80-90”
支持的研判. Google's Project Deschutes 1.0.0 (effective 26 February 2026) defines a 2 MW CDU: 500 GPM IT flow, 80–90 psi available differential pressure, 18–55 °C liquids, DI water or PG25.
無法據以確立. One contributor's unit specification; the 80–90 psi is IT-side differential pressure, not static pressure (operating range 0–130 psig), and conformance of listed products is unverified.
引用於: 第節 第1.2節, 第4.3節, 第4.4節, 第6.7節, 第A.1節, 第C.7節; 表 1, 13, 21, 32; 圖 19; 展望 3 · **研判:** A-03

S-031 Advantest Corporation — Advantest Corporation Q1 FY2026 (Three months ended June 30th, 2026)
企業新聞稿 · Japan · 英文 · 29 July 2026 · 閱讀 27 September 2026
https://www.advantest.com/en/investors/ir-library/result/
直接文件: https://www.advantest.com/document/en/investors/ir-library/result/JE_BIZ_260729_note.pdf
“At the same time, external uncertainties remain within the semiconductor industry, including advanced packaging capacity constraints and memory supply dynamics.”
“In particular, tester demand for inference AI-related ASICs, CPUs, and DRAM is growing faster than we previously expected.”
“For the SoC tester market, reflecting strong demand centered on AI-related applications, we are raising our CY2026 market forecast range to 10.5 to 11.5 billion USD.”
支持的研判. Advantest says tester demand for inference-AI ASICs, CPUs and DRAM is outpacing its expectations, raises CY2026 SoC- and memory-tester market forecasts, and names packaging constraints an external uncertainty.
無法據以確立. Company outlook for tester markets it serves; the faster-than-expected demand is scoped to inference-AI ASICs, CPUs and DRAM, not all tester demand.
引用於: 第節 第1.2節, 第3.5節, 第5.6節, 第5.7節, 第B.2節; 表 7, 26; 圖 25

S-032 ERCOT — PUCT Approves ERCOT's Batch Zero Process for Connecting Large Electricity Users While Protecting System Reliability for Texans
監管機關 · U.S. · 英文 · 18 June 2026 · 閱讀 26 September 2026
https://www.ercot.com/news/release/06182026-puct-approves-ercots
“ERCOT is tracking more than 438,000 MW of large load requests, nearly 89 percent from data centers alone.”
支持的研判. PUCT approved ERCOT's Batch Zero large-load process on 18 June 2026; ERCOT was tracking more than 438,000 MW of large-load requests, nearly 89% from data centres.
無法據以確立. Counts requests, not connections or firm demand; it gives no connection lead time or approved capacity.
引用於: 第節 第1.2節, 第5.6節, 第7.2節, 第C.7節; 圖 3, 24 · **研判:** A-01, A-04

S-033 TrendForce — [News] TSMC CoWoS Supply-Demand Gap Reportedly Seen Narrowing from 20% to 10% by End-2026
媒體 · Taiwan · 英文 · 15 June 2026 · 閱讀 27 September 2026
https://www.trendforce.com/news/2026/06/15/news-tsmc-cowos-supply-demand-gap-reportedly-seen-narrowing-from-20-to-10-by-end-2026-as-capacity-expands
“TSMC's monthly CoWoS capacity could reach a record 120,000 to 140,000 wafers in 2026”
“the CoWoS supply-demand gap is expected to narrow significantly from around 20% currently to about 10% by the end of 2026”
支持的研判. TrendForce News reports TSMC monthly CoWoS capacity could reach 120,000–140,000 wafers in 2026, OSATs adding 50,000–60,000, with the supply-demand gap expected to narrow from about 20% to 10% by end-2026.
無法據以確立. Relays press reports (Economic Daily News, Reuters, Commercial Times), not TrendForce research or TSMC disclosure; the gap's basis is undefined.
引用於: 圖 3, 24; 第節 第3.2節, 第3.4節, 第5.6節, 第5.8節, 第C.7節

S-034 TrendForce — Rubin Faces Delay Risks Amid Ongoing Supply Chain Adjustments; Blackwell to Account for Over 70% of NVIDIA's High-End GPU Shipments in 2026, Says TrendForce
市場研究 · Taiwan · 英文 · 8 April 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260408-13003.html
“This downward revision is primarily due to shipment delays facing the Rubin series.”
“In addition to the time required for HBM4 validation, challenges include transitioning network interconnects from CX8 to CX9, managing significantly higher power consumption, and optimizing performance under more advanced liquid cooling solutions.”
支持的研判. TrendForce trims its forecast of 2026 high-end GPU shipment growth from 26.8% to about 26% on Rubin shipment delays, citing HBM4 validation time, the CX8-to-CX9 move, power and cooling.
無法據以確立. Market-research forecast of NVIDIA's shipment mix; HBM4 validation time is one of four named causes of Rubin's delay, not a measured HBM4 shortage.
引用於: 圖 3

S-035 TrendForce — Rapid Contract Price Surge Drives 1Q26 DRAM Industry Up 81% QoQ, Says TrendForce
市場研究 · Taiwan · 英文 · 1 June 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260601-13070.html
“This surge contributed to an overall industry revenue increase of 81% QoQ, reaching $97 billion.”
“TrendForce's latest survey reveals that the memory industry experienced a significant boost in 1Q26 due to rapidly rising contract prices for conventional DRAM, which increased by approximately 93% to 98% QoQ.”
支持的研判. TrendForce estimates 1Q26 DRAM industry revenue rose 81% QoQ to US$97 billion, as conventional DRAM contract prices rose by about 93–98% QoQ.
無法據以確立. Market-research estimates for a past quarter; revenue and price data, not bit volumes, and not a company plan or milestone.
引用於: 第一頁; 圖 3, 14; 第節 第2.3節, 第3.2節, 第9.1節, 第E.2節; 表 36, 37

S-036 TrendForce — AI Chip Packaging Pushes Beyond Reticle Limits, with CoWoS-L Set to Remain the Mainstream Advanced Packaging Solution Through 2028, Says TrendForce
市場研究 · Taiwan · 英文 · 18 September 2026 · 閱讀 26 September 2026
https://www.trendforce.com/presscenter/news/20260918-13241.html
“Persistent constraints on TSMC's packaging capacity have further increased interest in Intel's EMIB technology as an alternative.”
支持的研判. TrendForce says persistent constraints on TSMC's packaging capacity have increased interest in Intel's EMIB as an alternative, and expects CoWoS-L to remain the mainstream solution through 2028.
無法據以確立. Market-research view; the CoWoS-L expectation is bounded to 2028 (per the headline), and interest in EMIB is not adoption or orders.
引用於: 圖 3; 第節 第3.4節, 第C.7節

S-037 TrendForce — InP Shortage Emerges as AI Optical Interconnect Bottleneck
媒體 · Taiwan · 英文 · 6 August 2026 · 閱讀 27 September 2026
https://www.trendforce.com/news/2026/08/06/news-inp-shortage-emerges-as-ai-optical-interconnect-bottleneck
“Lumentum CEO Michael Hurlston has warned that the supply-demand gap for indium phosphide (InP) has now surpassed that of DRAM and NAND, making it one of the most critical constraints on AI data center optical interconnect expansion.”
“industry analysts expect the supply imbalance to persist beyond 2027”
支持的研判. TrendForce News reports Lumentum's CEO saying InP's supply-demand gap has surpassed DRAM's and NAND's, notes China's roughly 70% share of refined indium and export curbs, and NVIDIA's US$2 billion commitments.
無法據以確立. News report; the 'beyond 2027' expectation is attributed to industry analysts, not TrendForce, and the gap is not quantified.
引用於: 圖 3, 25; 第節 第1.3節, 第4.7節, 第5.4節, 第5.7節, 第9.3節, 第C.7節; 表 1; 展望 4 · **研判:** A-01

S-038 TrendForce — Extended Component Lead Times Weigh on General Server Growth; 2026 Server Shipments Forecast to Grow 13% YoY, Says TrendForce
市場研究 · Taiwan · 英文 · 15 April 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260415-13013.html
"For PMICs, AI servers require substantially higher power density than general servers and are therefore prioritized by suppliers. Lead times are therefore expected to extend from 21–26 weeks to 35–40 weeks."
支持的研判. TrendForce expects server PMIC lead times to extend from 21–26 weeks to 35–40 weeks because suppliers prioritise AI servers, which require substantially higher power density.
無法據以確立. Market-research forecast; covers server PMICs, not power-conversion or grid equipment, and the longer lead times fall mainly on general servers, as AI servers get priority.
引用於: 圖 3, 33; 第節 第E.2節

S-039 GE Vernova Inc. — GE Vernova raises multi-year financial outlook, doubles dividend and increases buyback authorization
企業新聞稿 · U.S. · 英文 · 9 December 2025 · 閱讀 27 September 2026
https://www.gevernova.com/news/press-releases/ge-vernova-raises-multi-year-financial-outlook-doubles-dividend-increases-buyback-authorization
"Expects to grow total backlog from $135B to approximately $200B by year-end 2028, inclusive of doubling the size of Electrification backlog from $30B to $60B"
"18 GW of gas turbine contracts signed quarter-to-date; expects to reach 80 GW of combined slot reservation agreements and backlog by year-end"
支持的研判. GE Vernova expected in December 2025 total backlog of about US$200 billion by end-2028, with Electrification doubling to US$60 billion, and 80 GW of gas slots plus backlog by end-2025.
無法據以確立. Company expectations as of December 2025; the 80 GW was a year-end target, not a reported outcome.
引用於: 圖 3; 第節 第5.5節; 表 10, 19

S-040 Texas Legislature — S.B. No. 6 (enrolled), 89th Legislature, Regular Session
政府 · U.S. (Texas) · 英文 · 2025 · 閱讀 27 September 2026
https://capitol.texas.gov/tlodocs/89R/billtext/html/SB00006F.htm
"The commission shall establish a demand threshold of 75 megawatts unless the commission determines that a lower threshold is necessary to accomplish the purposes described by Subsection (b)."
"After the independent organization deploys all available market services, except for frequency responsive services, the independent organization may direct the applicable electric utility or municipally owned utility to require the large load customer to either deploy the customer's on-site backup generating facilities or curtail load."
支持的研判. Texas SB 6 sets a 75 MW large-load threshold, lets ERCOT have large loads deploy backup generation or curtail in emergencies, and requires amended transmission-cost rules by 31 December 2026.
無法據以確立. Statute text; PUCT may set a lower threshold, ERCOT acts through utilities after market services, and the 2026 rules cover wholesale transmission charges generally.
引用於: 圖 3; 第節 第7.2節, 第C.7節; 展望 17

S-041 Federal Energy Regulatory Commission — Fact Sheet: FERC Directs Nation's Largest Grid Operator to Create New Rules to Embrace …
監管機關 · U.S. · 英文 · 18 December 2025 · 閱讀 26 September 2026
https://www.ferc.gov/news-events/news/fact-sheet-ferc-directs-nations-largest-grid-operator-create-new-rules-embrace
"PJM must revise its tariff to detail the terms and conditions for interconnection customers to follow when using generating facilities to serve co-located load."
支持的研判. FERC's December 2025 order requires PJM to set tariff terms for generation serving co-located load and to offer four transmission services for such load, two non-firm.
無法據以確立. A fact sheet on an order requiring compliance filings; final PJM terms, prices and timelines are set later.
引用於: 圖 3; 第節 第7.2節

S-042 PJM Inside Lines — PJM Proposes Framework To Connect Data Centers Without Compromising Reliability, Affordability
監管機關 · U.S. · 英文 · 13 August 2026 · 閱讀 27 September 2026
https://insidelines.pjm.com/pjm-proposes-framework-to-connect-data-centers-without-compromising-reliability-affordability
"Of 32 GW of growth in forecasted electricity demand between 2024 and 2030, 30 GW is attributed to data centers."
"IRAS would apply to new Large Loads that have not brought their own new capacity or had their capacity needs covered through the Reliability Backstop Procurement, a special process designed to fill forecasted capacity shortfalls starting in June 2027."
支持的研判. PJM's August 2026 filing would curtail or transfer new large loads lacking own capacity before other consumers in emergencies; it attributes 30 of 32 GW forecast growth to data centres.
無法據以確立. A proposal filed with FERC, not an approved tariff; it covers new large loads (with a backstop-procurement exception), not existing data centres.
引用於: 圖 3, 24; 第節 第1.3節, 第1.5節, 第5.6節, 第7.2節, 第C.7節; 表 1 · **研判:** A-04

S-043 TrendForce — Global Power Demand Capacity for Data Centers Expected to Rise 31% YoY in 2026; Grid Supply Gap to Widen in 2028, Says TrendForce
市場研究 · Taiwan · 英文 · 16 September 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260916-13237.html
"According to current grid data, TrendForce projects that the U.S. data center power supply-demand gap will surpass 170 GW by 2030."
"Global data center power demand capacity reached 122.9 GW in 2025 and is estimated to rise to 161 GW in 2026."
"However, after 2028, delays in grid power delivery are expected to become more frequent, substantially contributing to the growing gap between data center power supply and demand capacity."
支持的研判. TrendForce puts data-centre power demand capacity at 161 GW in 2026 (122.9 GW in 2025) and 490.7 GW in 2030, against 222.6 GW of grid capacity: a 268 GW gap.
無法據以確立. Capacity (GW) forecasts, not energy; the 2030 grid figure assumes other demand fully met, and ' grid-connected through 2028' refers to the U.S.
引用於: 圖 3, 28; 第節 第1.5節, 第2.5節, 第7.3節, 第9.1節, 第9.3節; 展望 8; 表 20 · **研判:** A-04

S-044 Oracle Corporation (Investor Relations) — Oracle Announces Record Q4 and FY 2026 Results Driven by Cloud Infrastructure & Cloud Applications
企業新聞稿 · U.S. · 英文 · 10 June 2026 · 閱讀 27 September 2026
https://investor.oracle.com/investor-news/news-details/2026/Oracle-Announces-Record-Q4-and-FY-2026-Results-Driven-by-Cloud-Infrastructure--Cloud-Applications/default.aspx
"The prepaid and customer supplied hardware portions of our large AI contracts now total $75 billion."
"In fiscal year 2027, Oracle expects to raise approximately $40 billion through a combination of debt and equity financing including its previously announced $20 billion at-the-market equity issuance."
"In fiscal year 2026, Oracle raised $43 billion in debt financing and $5 billion in equity financing."
支持的研判. Oracle reports FY2026 capex of US$55.7 billion (FY2025 US$21.2 billion), US$43 billion debt and US$5 billion equity raised, negative FCF of US$23.7 billion, and expects about US$40 billion FY2027 financing.
無法據以確立. Company results and plans; capex is gross of customer prepayments, and the FY2027 financing is an expectation.
引用於: 圖 3, 4, 8; 第節 第1.3節, 第1.5節, 第2.1節, 第2.2節, 第9.1節; 展望 9; 表 20, 39 · **研判:** A-07

S-045 Microsoft Corporation (Investor Relations) — Microsoft Fiscal Year 2026 Third Quarter Earnings Conference Call
企業新聞稿 · U.S. · 英文 · 29 April 2026 · 閱讀 27 September 2026 · **財測指引已修訂**
https://www.microsoft.com/en-us/investor/events/fy-2026/earnings-fy-2026-q3
"For calendar year 2026, we expect to invest roughly $190 billion in capital expenditures which includes approximately $25 billion from the impact of higher component pricing."
"And this quarter, roughly two thirds of our capex was for short-lived assets, primarily GPUs and CPUs."
支持的研判. Microsoft said in April 2026 it expects roughly US$190 billion of calendar-2026 capex, including about US$25 billion from higher component pricing; about two-thirds of the quarter's capex was short-lived assets.
無法據以確立. Expectation on a basis that includes finance leases; re-based to about US$175 billion in July 2026.
財測指引已修訂. The calendar-2026 expectation of about US$190 billion was re-based to about US$175 billion in July 2026, with investment unchanged, because longer useful lives move leases outside capital expenditure [S-075]: a change of basis, not a cut.
引用於: 第一頁; 第節 第1.3節, 第2.1節, 第2.2節, 第3.2節, 第C.5節, 第C.7節; 圖 4, 5; 表 38 · **研判:** A-02

S-046 Meta Platforms, Inc. (Investor Relations) — META Q1 2026 Earnings Call Transcript
企業新聞稿 · U.S. · 英文 · 29 April 2026 · 閱讀 27 September 2026
https://s21.q4cdn.com/399680738/files/doc_financials/2026/q1/META-Q1-2026-Earnings-Call-Transcript.pdf
"On that note, we are increasing our infrastructure capex forecast for this year. Most of that is due to higher component costs, particularly memory pricing."
支持的研判. On its April 2026 call Meta said it is increasing its 2026 infrastructure capex forecast and that most of the increase is due to higher component costs, particularly memory pricing.
無法據以確立. Attributes most of the increase, not most of the forecast, to component costs; no amount given.
引用於: 第一頁; 第節 第1.3節, 第2.1節 · **研判:** A-02

S-047 Open Compute Project Foundation — Open Compute Project Foundation Intellectual Property Rights Management Policy
標準組織 · Global · 英文 · 14 August 2026 · 閱讀 27 September 2026
https://www.opencompute.org/documents/ocp-ip-rights-management-policy
"a 30-day period starting on the date the final document is uploaded to the OCP Contribution Hub to initiate the Steering Committee vote"
"Each Contributor to the Proposed Specification will be deemed to have agreed to the Specification License upon the Final Specification's release as set forth in Section 3.4, except for materials identified in a valid Exclusion Notice."
支持的研判. OCP's IPR policy (14 August 2026) deems each contributor to accept the royalty-free OWFa 0.9 when a Final Specification is released, except claims excluded during the 30-day opt-out.
無法據以確立. Binds contributors only, under the policy's text; it does not show how many Exclusion Notices are filed or how courts would read the commitment.
引用於: 第節 第1.3節, 第1.4節, 第6.1節, 第6.2節, 第6.6節, 第6.7節, 第C.7節; 表 12; 附錄A · **研判:** A-06, A-10

S-048 NVIDIA — NVIDIA Launches DSX Ready to Qualify Power and Cooling Products for AI Factories
企業新聞稿 · U.S. · 英文 · 21 September 2026 · 閱讀 26 September 2026
https://blogs.nvidia.com/blog/dsx-ready-ai-factories-power-cooling
"At launch, DSX Ready includes qualified BESS solutions Hitachi Energy, LG Energy Solution and Tesla, and qualified CDU solutions from LG Electronics, LiquidStack and Vertiv."
支持的研判. NVIDIA launched DSX Ready on 21 September 2026 to qualify AI-factory power and cooling products, starting with three BESS makers (Hitachi Energy, LG Energy Solution, Tesla) and three CDU makers.
無法據以確立. A vendor programme days old at the cut-off; it does not show that buyers must use qualified products or exclude others.
引用於: 第節 第1.3節, 第4.3節, 第4.4節, 第5.6節, 第C.7節; 表 5, 28; 圖 24 · **研判:** A-06, A-12

S-049 Oracle Corporation (Investor Relations) — Oracle Announces Q1 Results Driven by Triple-Digit Growth in Cloud Infrastructure Revenues
企業新聞稿 · U.S. · 英文 · 10 September 2026 · 閱讀 27 September 2026
https://investor.oracle.com/investor-news/news-details/2026/Oracle-Announces-Q1-Results-Driven-by-Triple-Digit-Growth-in-Cloud-Infrastructure-Revenues/default.aspx
"Customer demand for AI Cloud Training and Inferencing Services continues to grow faster than supply."
"Oracle booked more than $30 billion of additional AI cloud contracts in Q1 increasing its RPO to $664 billion."
支持的研判. Oracle reports RPO of US$664 billion after over US$30 billion of Q1 FY2027 AI cloud contracts, 850 MW delivered, a US$20 billion ATM sale completed, and AI-cloud demand outpacing supply.
無法據以確立. Company statements; the demand claim covers AI cloud training and inferencing services only, and RPO is not revenue.
引用於: 第節 第1.3節, 第2.2節, 第9.1節; 圖 8; 展望 9; 表 20, 24 · **研判:** A-07

S-050 NVIDIA Corporation — NVIDIA Announces Financial Results for Fourth Quarter and Fiscal 2026
企業新聞稿 · U.S. · 英文 · 25 February 2026 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-fourth-quarter-and-fiscal-2026
"Revenue is expected to be $78.0 billion, plus or minus 2%. NVIDIA is not assuming any Data Center compute revenue from China in its outlook."
"Fourth-quarter revenue was a record $62.3 billion, up 22% from the previous quarter and up 75% from a year ago, driven by the major platform shifts — accelerated computing and AI."
支持的研判. NVIDIA reported record Q4 FY2026 Data Center revenue of US$62.3 billion and guided Q1 FY2027 revenue of US$78.0 billion ±2%, assuming no Data Center compute revenue from China.
無法據以確立. Company results and outlook; the China assumption applies to the outlook period, not all China revenue.
引用於: 第節 第1.3節, 第2.3節, 第7.1節; 圖 7; 表 14 · **研判:** A-08

S-051 NVIDIA Corporation — NVIDIA Announces Financial Results for First Quarter Fiscal 2027
企業新聞稿 · U.S. · 英文 · 20 May 2026 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-first-quarter-fiscal-2027
"Revenue is expected to be $91.0 billion, plus or minus 2%. NVIDIA is not assuming any Data Center compute revenue from China in its outlook."
"First-quarter revenue was a record $75.2 billion, up 21% from the previous quarter and up 92% from a year ago."
支持的研判. NVIDIA reported Q1 FY2027 revenue of US$81.6 billion (Data Center US$75.2 billion) and guided Q2 revenue of US$91.0 billion ±2%, assuming no Data Center compute revenue from China.
無法據以確立. Company results and outlook; the China assumption applies to Data Center compute in the outlook, not all China revenue.
引用於: 第節 第1.3節, 第2.3節, 第7.1節; 圖 7; 表 14 · **研判:** A-08

S-052 AUO Corporation — AUO Showcases Micro LED Optical Communication and Glass Core Substrates Technologies at SEMICON Taiwan 2026
企業新聞稿 · Taiwan · 英文 · 31 August 2026 · 閱讀 27 September 2026
https://www.auo.com/en-global/New_Archive/detail/News_Archive_Product_20260831
"Targeting high-speed, short-reach interconnects of up to 10 meters in AI data centers, AUO provides key integration capabilities including mass transfer, redistribution layer (RDL) packaging, optical coupling, and system architecture design."
"The solution integrates Micro LED transmitters (Tx) from group company Ennostar and Micro photodetector receivers (Rx) from Tyntek, with Daxin's photosensitive dielectric materials expected to be incorporated into the RDL packaging structure."
"Together with optical fiber solutions from ecosystem partner Corning, AUO is developing a system-level Micro LED CPO optical module designed for high-speed interconnect applications such as co-packaged optics (CPO) and active optical cables (AOC)."
支持的研判. AUO said (31 August 2026) it is developing with partners a Micro LED CPO module for AI data-centre links up to 10 m, and would show glass-core substrates with Corning.
無法據以確立. Pre-event announcement of R&D exhibits; no production, customers, link performance or revenue; component figures are AUO's own.
引用於: 第節 第1.3節, 第3.5節, 第4.7節, 第5.7節, 第B.1節, 第B.2節, 第C.7節; 表 1, 6, 13, 20, 23, 30; 圖 26; 展望 12 · **研判:** A-09

S-053 Nature Electronics — Co-packaged optics for high-performance computing and artificial intelligence
研究 · Global · 英文 · 19 August 2026 · 閱讀 26 September 2026
https://www.nature.com/articles/s41928-026-01681-6
"thermal management, manufacturability and standardization"
"will need to be addressed to establish optical compute interconnects as a foundational communication technology for high-performance computing infrastructure"
支持的研判. A Nature Electronics review (August 2026) says thermal management, manufacturability and standardization must be addressed before optical compute interconnects become foundational for high-performance computing.
無法據以確立. An academic review; it identifies obstacles without ranking them or assessing any company's products or patents.
引用於: 第節 第1.3節, 第4.7節, 第6.7節, 第C.7節; 表 1, 13 · **研判:** A-10

S-054 Open Compute Project Foundation — Solid State Transformer (SST) Specification, Revision 0.3.0
標準組織 · Global · 英文 · 22 June 2026 · 閱讀 27 September 2026
https://www.opencompute.org/documents/ocp-sst-design-specification-v0-3-final-pdf
"Efficiency >= 98% between 50-100% utilization (power train loss only)"
"This document defines the product using a 'black-box' approach, with the focus on interfaces as application level guidance (e.g. electrical, mechanical, thermal, environmental, control, communications, and behavior expectation etc)."
取自來源表格之數值（轉錄）： 5/20/26 | 0.3 | Google: Jing Wang, Sangyoun "SY" Kim, Satarupa Bal, Kei Hao, Dragan Ristanovic, Ashish Upreti, Jason Adrian, Christina Peabody Microsoft: Ehsan Nasr, Pritesh Patel, Xianyong Feng Nvidia: Mike Tu | first draft of SST product for OCP community
支持的研判. The SST Specification 0.3.0 (effective 22 June 2026; Google, Microsoft, NVIDIA), a first draft, defines a black-box MV-to-800 VDC converter by interfaces, requiring ≥98% efficiency at 50–100% utilisation.
無法據以確立. A first draft leaving topology to vendors and grounding details to later revisions; it shows no certified or shipping products.
引用於: 第節 第1.3節, 第1.4節, 第4.4節, 第6.7節, 第7.4節, 第9.2節, 第9.4節, 第A.1節, 第C.7節; 表 1, 5, 13, 20, 21, 32; 圖 17, 26; 展望 13, 17 · **研判:** A-10

S-055 SolarEdge Technologies, Inc. and Infineon Technologies AG (joint announcement; SolarEdge IR copy) — SolarEdge Extends Collaboration with Infineon to Enable Solid-State Protection for 800 VDC AI Data Centers
企業新聞稿 · U.S. / Germany · 英文 · 9 September 2026 · 閱讀 27 September 2026
https://investors.solaredge.com/news-releases/news-release-details/solaredge-extends-collaboration-infineon-enable-solid-state
"Companies extend collaboration into solid-state circuit breaker (SSCB) technology in line with leading industry framework for 800 Volts direct current (VDC) architecture"
"SSCBs are designed to interrupt faults within a few microseconds, with no mechanical contacts to separate, while Infineon's CoolSiC™ JFET devices deliver efficient operation and superior robustness."
"SolarEdge is leading the design of its SSCB solution, with Infineon providing its silicon carbide (SiC) JFET technology at the core of the protection devices."
支持的研判. SolarEdge and Infineon extended collaboration on 9 September 2026 to SSCBs for 800 VDC, SolarEdge leading design, Infineon supplying SiC JFETs; SSCBs are designed to interrupt faults within microseconds.
無法據以確立. Development collaboration: no availability, production or demonstration date, no named platform; dateline Milpitas and Munich, not Israel.
引用於: 第節 第1.3節, 第4.2節, 第5.6節, 第6.7節, 第B.2節, 第C.7節; 表 1, 5, 20, 28; 圖 26; 展望 2 · **研判:** A-10

S-056 Samsung Electronics Co., Ltd. — Samsung Ships Industry-First Commercial HBM4 With Ultimate Performance for AI Computing
企業新聞稿 · Korea · 英文 · 12 February 2026 · 閱讀 27 September 2026
https://news.samsung.com/global/samsung-ships-industry-first-commercial-hbm4-with-ultimate-performance-for-ai-computing
"Following the successful introduction of HBM4 to market, sampling for HBM4E is expected to begin in the second half of 2026, while custom HBM samples will start reaching customers in 2027, according to their respective specifications."
"Samsung anticipates that its HBM sales will more than triple in 2026 compared to 2025, and is proactively expanding its HBM4 production capacity."
支持的研判. Samsung says it began HBM4 mass production and shipped commercial products (February 2026), anticipates 2026 HBM sales more than triple 2025's, and expects custom HBM samples in 2027.
無法據以確立. Company statements; the sales outlook is an expectation and 'industry-first' is Samsung's claim; no volumes, customers or start date given.
引用於: 表 1, 3, 9, 10, 13, 19, 22, 25; 第節 第1.5節, 第3.2節, 第3.4節, 第5.7節, 第6.7節, 第7.5節, 第A.2節; 圖 25

S-057 BE Semiconductor Industries N.V. (Besi) — BE Semiconductor Industries N.V. Increases Long-Term Financial Targets at 2026 Investor Day
企業新聞稿 · Netherlands · 英文 · 18 June 2026 · 閱讀 27 September 2026
https://www.besi.com/investor-relations/press-releases/details/be-semiconductor-industries-nv-increases-long-term-financial-targets-at-2026-investor-day/
"Accordingly, we are raising Besi's long-term revenue and operating margin targets as set forth below."
"€ 1.7 – € 2.2 billion"
支持的研判. At its June 2026 Investor Day Besi raised long-term targets to €1.7–2.2 billion revenue (from €1.5–1.9 billion) and 45–55% operating margin (from 40–55%), citing AI packaging and hybrid-bonding demand.
無法據以確立. Management targets with no target year; not a forecast or guidance for any period.
引用於: 表 1, 9, 10, 26, 34, 35; 第節 第3.5節, 第5.8節

S-058 JEDEC Solid State Technology Association — JESD270-4A: High Bandwidth Memory (HBM4) DRAM
標準組織 · U.S. · 英文 · December 2025 · 閱讀 27 September 2026
https://www.jedec.org/standards-documents/docs/jesd270-4
"The HBM4 DRAM is tightly coupled to the host compute die with a distributed interface. The interface is divided into independent channels."
"Each channel interface maintains a 64 bit data bus operating at double data rate (DDR)."
支持的研判. JEDEC's JESD270-4A (HBM4 DRAM, December 2025) describes an interface tightly coupled to the host compute die, divided into independent channels with 64-bit DDR data buses.
無法據以確立. The catalogue abstract describes the channel interface; it does not describe base-die logic, supplier speeds or shipments.
引用於: 表 1, 11, 13, 22; 第節 第3.4節, 第6.3節, 第6.7節, 第C.2節, 第C.7節; 圖 26

S-059 Open Compute Project Foundation — Data Center Facility – Low Voltage Direct Current Power Distribution, White Paper v1.1.0
標準組織 · Global · 英文 · 2 September 2026 · 閱讀 27 September 2026
https://www.opencompute.org/documents/dcf-power-distribution-lvdc-white-paper-version-1-1-0-pdf
"In the context of power distribution, low voltage is defined by international product standards and regulations along with installation codes as operation at or below 1000 VAC or 1500 VDC."
"Added new chapters 9, 10 and 11 (IT rack, Control&Operation, EMC), major expansion for all other chapters."
支持的研判. OCP's LVDC white paper v1.1.0 (2 September 2026) notes that standards and codes define low voltage as up to 1000 VAC or 1500 VDC, adding IT-rack, control and EMC chapters.
無法據以確立. A white paper, not a specification or code: it sets no protection timings and creates no listing; its ride-through sections are guidance.
引用於: 表 1, 21; 第節 第4.2節, 第4.4節, 第6.7節, 第7.4節, 第9.4節, 第A.1節, 第C.7節; 展望 16 · **研判:** A-10

S-060 Open Compute Project Foundation — Cooling Environments/Cold Plate (wiki); PG Guidelines; Base Specification for Single Phase Water-Based Cold Plate Coolants v1.3
標準組織 · Global · 英文 · 6 August 2026 · 閱讀 26 September 2026
https://www.opencompute.org/wiki/Cooling_Environments/Cold_Plate
"Universal Quick Disconnect (UQD) V2 Specification revision for both handmate and blindmate versions."
"Guidelines for single & two phase coolant fluids: fluid spec, testing, material compatibility, etc."
支持的研判. OCP's Cold Plate wiki lists a UQD v2 revision (beta draft, CPC and Etched) and guidelines for single- and two-phase fluids, with fluid specification and testing as work items.
無法據以確立. A project wiki: drafts and work items are not released specifications, and it gives no completion dates.
引用於: 表 1, 21; 第節 第4.3節, 第6.7節, 第C.7節; 圖 26 · **研判:** A-10

S-061 PJM Inside Lines — PJM Proposes Reliability Standards to Manage Large Load Disconnection Events
監管機關 · U.S. · 英文 · 10 September 2026 · 閱讀 27 September 2026
https://insidelines.pjm.com/reliability-standards-to-manage-large-load-disconnection-events-proposed-by-pjm
"the North American Electric Reliability Corporation (NERC) is developing Large Load reliability guidance and ride-through recommendations and standards."
"PJM plans to file proposed solutions to FERC in November."
支持的研判. PJM (10 September 2026) proposes that large loads ride through normally cleared transmission disturbances, plans to file with FERC in November, and notes NERC is developing large-load ride-through standards.
無法據以確立. A proposal with a planned filing date; nothing is filed or approved yet, and the requirements bind only after FERC acceptance.
引用於: 表 1; 第節 第1.5節, 第4.4節, 第5.6節, 第7.2節, 第C.7節

S-062 Open Compute Project Foundation — Delivering an Open Data Center Ecosystem for AI
標準組織 · Global · 英文 · 29 April 2026 · 閱讀 26 September 2026
https://www.opencompute.org/blog/delivering-an-open-data-center-ecosystem-for-ai
"today announced newly approved OCP contributions, new projects, and recently formed alliances to deliver on its Open Data Center Ecosystem for AI vision announced in October of 2025"
"EPRI and OCP aim to accelerate digital innovation and further develop the potential of data centers to serve as flexible resources for the power system."
支持的研判. OCP's 29 April 2026 post lists eight newly approved contribution items, including ORW, LVDC, energy storage, telemetry and FCSA, and announces an EPRI alliance on flexible data centres.
無法據以確立. Announcement of approvals and alliances; version numbers come from linked files, and it gives no deliverables or dates for the EPRI or Current/OS work.
引用於: 表 1, 13, 21; 第節 第4.4節, 第4.5節, 第6.1節, 第6.2節, 第9.4節; 展望 17

S-063 Wiwynn Corporation — Wiwynn Reports Second Quarter 2026 Financial Results
企業新聞稿 · Taiwan · 英文 · 7 August 2026 · 閱讀 27 September 2026
https://www.wiwynn.com/news/wiwynn-reports-second-quarter-2026-financial-results
"In response to changes in the AI product mix and rising component costs, the Company has continued to work with customers to adjust its business model."
"Combined with higher NRE income from a new product introduction during the quarter, gross margin improved to 9.3%."
"Beginning in April, memory procurement for certain customers shifted to a procurement agency model, with the related amounts excluded from revenue and cost of goods sold."
支持的研判. Wiwynn says memory for certain customers moved from April 2026 to a procurement agency model, excluded from revenue and COGS; Q2 2026 revenue NT$278.153 billion, gross margin 9.3%.
無法據以確立. Company disclosure for certain customers only; it does not say buyers procure memory themselves, name the customers or quantify the excluded amounts.
引用於: 第一頁; 第節 第1.5節, 第3.2節, 第5.1節, 第5.2節, 第5.3節, 第C.3節, 第C.7節; 表 7, 8, 27, 39; 圖 22 · **研判:** A-12, A-14

S-064 Wiwynn Corporation — Wiwynn Reports First Quarter 2026 Financial Results
企業新聞稿 · Taiwan · 英文 · 7 May 2026 · 閱讀 27 September 2026
https://www.wiwynn.com/news/wiwynn-reports-first-quarter-2026-financial-results
"Beginning in April, memory procurement for certain customers shifted to a procurement agency model, under which memory revenue is excluded from reported revenue."
"Looking ahead to the full year, total profit is expected to remain unaffected by partial changes in business models, with shipment volume continuing to achieve double-digit growth."
支持的研判. Wiwynn says memory procurement for certain customers moved to an agency model from April 2026, excluding it from revenue, and expects full-year profit unaffected with double-digit shipment growth.
無法據以確立. Company statement; does not quantify the memory excluded or how many customers use the agency model.
引用於: 第節 第1.5節, 第3.2節, 第5.2節, 第C.3節; 圖 22; 表 27

S-065 CoreWeave / SEC EDGAR — CoreWeave, Inc. Form 10-K for fiscal year ended December 31, 2025
申報文件 · U.S. · 英文 · 2 March 2026 · 閱讀 26 September 2026
https://www.sec.gov/Archives/edgar/data/1769628/000176962826000104/crwv-20251231.htm
"We recognized an aggregate of approximately 67% of our revenue from our top customer, Microsoft, for the year ended December 31, 2025."
支持的研判. CoreWeave's 2025 Form 10-K states it recognized about 67% of 2025 revenue from its top customer, Microsoft.
無法據以確立. Revenue concentration for 2025; the cited passage does not give the customer split of backlog or RPO.
引用於: 第節 第1.5節, 第2.2節, 第5.6節, 第C.7節; 圖 24 · **研判:** A-07

S-066 Nebius — Nebius raises $775 million in first secured debt financing to accelerate global buildout
企業新聞稿 · Netherlands · 英文 · 17 July 2026 · 閱讀 26 September 2026
https://nebius.com/newsroom/nebius-raises-775-million-in-first-secured-debt-financing-to-accelerate-global-buildout
"Senior secured debt backed by GPU infrastructure and contracted cash flows from an agreement with an investment-grade customer."
支持的研判. Nebius says its first secured debt, US$775 million, is backed by GPU infrastructure and contracted cash flows from an agreement with an investment-grade customer.
無法據以確立. One financing by one company; it does not show that contract-backed debt is typical across the sector.
引用於: 第節 第1.5節, 第5.6節, 第C.7節; 圖 24 · **研判:** A-07

S-067 Open Compute Project Foundation — 2026 OCP Global Summit » Open Compute Project; Registration » 2026 OCP Global Summit
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/summit/global-summit
"Scaling Innovation for the AI Era"
"October 12–15, 2026"
支持的研判. OCP's Summit pages set the 2026 Global Summit for 12–15 October 2026 in San Jose under the theme 'Scaling Innovation for the AI Era', with sessions to be recorded.
無法據以確立. Event announcement before the Summit; it does not show which specifications will be released or accepted there.
引用於: 第節 第1.5節, 第6.2節, 第9.2節, 第A.1節, 第C.6節; 展望 2, 3, 5; 圖 32

S-068 Federal Energy Regulatory Commission — FERC launches aggressive, targeted action to speed large load integration
監管機關 · U.S. · 英文 · 18 June 2026 · 閱讀 26 September 2026
https://www.ferc.gov/news-events/news/ferc-launches-aggressive-targeted-action-speed-large-load-integration
"issued tailored show cause orders under section 206 of the Federal Power Act to each of the six regional grid operators under its jurisdiction"
支持的研判. On 18 June 2026 FERC issued section 206 show-cause orders to the six regional grid operators it regulates, requiring them within 60 days to justify tariffs or propose large-load reforms.
無法據以確立. Orders to show cause, not final rules or tariffs; they do not yet set terms for flexible or co-located loads.
引用於: 第節 第1.5節, 第7.2節, 第9.1節; 展望 8

S-069 Amazon.com, Inc. (Investor Relations) — Amazon.com Announces Fourth Quarter Results
企業新聞稿 · U.S. · 英文 · 5 February 2026 · 閱讀 27 September 2026 · **財測指引已修訂**
https://ir.aboutamazon.com/news-release/news-release-details/2026/Amazon-com-Announces-Fourth-Quarter-Results/default.aspx
"With such strong demand for our existing offerings and seminal opportunities like AI, chips, robotics, and low earth orbit satellites, we expect to invest about $200 billion in capital expenditures across Amazon in 2026, and anticipate strong long-term return on invested capital"
"Free cash flow decreased to $11.2 billion for the trailing twelve months, driven primarily by a year-over-year increase of $50.7 billion in purchases of property and equipment, net of proceeds from sales and incentives."
取自來源表格之數值（轉錄）: Consolidated Statements of Cash Flows, line 'Purchases of property and equipment': Twelve Months Ended December 31, 2024 (82,999); Twelve Months Ended December 31, 2025 (131,819) (in millions)
取自來源表格之數值（轉錄）: Consolidated Statements of Cash Flows, line 'Proceeds from property and equipment sales and incentives': Twelve Months Ended December 31, 2024 5,341; Twelve Months Ended December 31, 2025 3,499 (in millions)
支持的研判. Amazon expected in February 2026 about US$200 billion of company-wide 2026 capex; 2025 purchases of property and equipment were US$131.8 billion, with US$3.5 billion of proceeds and incentives.
無法據以確立. February 2026 expectation, later raised; company-wide, not AWS alone.
財測指引已修訂. Its expectation of about US$200 billion of 2026 capital expenditure became about US$220 billion of cash capital expenditure on the July 2026 call [S-070].
引用於: 第節 第2.1節; 圖 4, 5

S-070 StockAnalysis (transcript of Amazon.com, Inc. Q2 2026 earnings call) — Amazon (AMZN) Q2 2026 Earnings Call Transcript & Audio
媒體文字紀錄 · U.S. · 英文 · 30 July 2026 · 閱讀 28 September 2026
https://stockanalysis.com/stocks/amzn/transcripts/657334-q2-2026/
"We now believe we will spend approximately $220 billion in cash CapEx in 2026. The higher cost of memory pushing this number up from our prior estimate of about $200 billion."
"I'll start with AWS, which is booming right now, and I'll share the numbers, what we think is going on, and why we're enthusiastic about the ROIC equation, even with heavy CapEx the next few years."
"Even at that amount, we will still not have enough capacity to meet all the demand we have in 2026, and I believe this dynamic will also be true in 2027, too."
支持的研判. Amazon expects about US$220 billion of 2026 cash capex (from about US$200 billion) on higher memory costs, says capacity will still trail demand, and that 2027 capacity is largely reserved; in its AWS remarks it gives a backlog of US$496 billion.
無法據以確立. Third-party transcript of management remarks; capex is cash capex, and 'largely reserved' and the capacity shortfall are unquantified.
引用於: 第節 第2.1節, 第2.2節, 第3.2節, 第9.1節, 第B.2節; 圖 4, 5, 8; 表 19, 24

S-071 Meta Platforms, Inc. (Investor Relations) — Meta Reports Fourth Quarter and Full Year 2025 Results
企業新聞稿 · U.S. · 英文 · 28 January 2026 · 閱讀 27 September 2026 · **財測指引已修訂**
https://investor.atmeta.com/investor-news/press-release-details/2026/Meta-Reports-Fourth-Quarter-and-Full-Year-2025-Results/default.aspx
"We anticipate 2026 capital expenditures, including principal payments on finance leases, to be in the range of $115-135 billion, with year-over-year growth driven by increased investment to support our Meta Superintelligence Labs efforts and core business."
"Capital expenditures, including principal payments on finance leases, were $22.14 billion and $72.22 billion for the fourth quarter and full year 2025, respectively."
支持的研判. In January 2026 Meta anticipated 2026 capex, including finance-lease principal payments, of US$115–135 billion; its 2025 figure on that basis was US$72.22 billion.
無法據以確立. Guidance later raised; Meta's definition adds finance-lease principal payments.
財測指引已修訂. Its 2026 range of US$115–135 billion was raised to US$125–145 billion in April [S-072] and narrowed to US$130–145 billion in July 2026 [S-073].
引用於: 第節 第2.1節; 圖 4, 5

S-072 Meta Platforms, Inc. (Investor Relations) — Meta Reports First Quarter 2026 Results
企業新聞稿 · U.S. · 英文 · 29 April 2026 · 閱讀 27 September 2026 · **財測指引已修訂**
https://investor.atmeta.com/investor-news/press-release-details/2026/Meta-Reports-First-Quarter-2026-Results/default.aspx
"We anticipate 2026 capital expenditures, including principal payments on finance leases, to be in the range of $125-145 billion, increased from our prior range of $115-135 billion. This reflects our expectations for higher component pricing this year and, to a lesser extent, additional data center costs to support future year capacity."
支持的研判. In April 2026 Meta raised its 2026 capex range, including finance-lease principal payments, to US$125–145 billion, citing higher component pricing and, less so, data-centre costs.
無法據以確立. Guidance as of April 2026, narrowed in July; Meta-specific definition.
財測指引已修訂. Its 2026 range of US$125–145 billion was narrowed to US$130–145 billion in July 2026 [S-073].
引用於: 第節 第2.1節; 圖 5

S-073 Meta Platforms / SEC EDGAR — Meta Reports Second Quarter 2026 Results (Exhibit 99.1)
企業新聞稿 · U.S. · 英文 · 29 July 2026 · 閱讀 26 September 2026
https://investor.atmeta.com/investor-news/press-release-details/2026/Meta-Reports-Second-Quarter-2026-Results/default.aspx
"Anticipate 2026 capital expenditures, including principal payments on finance leases, to be in the range of $130-145 billion, narrowed from our prior outlook of $125-145 billion."
支持的研判. Meta narrowed its 2026 capital expenditure outlook, including principal payments on finance leases, to US$130–145 billion from US$125–145 billion in July 2026.
無法據以確立. Company guidance on its own capex definition, including finance-lease principal; not actual spending and not comparable with other companies' capex.
引用於: 第節 第2.1節; 圖 4, 5; 表 24

S-074 Meta Platforms, Inc. (Investor Relations) — META Q2 2026 Earnings Call Transcript
企業新聞稿 · U.S. · 英文 · 29 July 2026 · 閱讀 27 September 2026
https://s21.q4cdn.com/399680738/files/doc_financials/2026/q2/META-Q2-2026-Earnings-Call-Transcript.pdf
"Brian, on your second question, we aren't providing a specific outlook for 2027 CapEx at this time. Infrastructure planning remains highly dynamic and even this year, there are a range of outcomes embedded in our outlook."
"Yesterday, as part of our Meta Compute effort, we announced a new strategic venture with BlackRock to develop a new 1GW data center in El Paso, Texas."
支持的研判. In July 2026 Meta gave no specific 2027 capex outlook and said it had announced a venture with BlackRock to develop a 1 GW El Paso, Texas data centre.
無法據以確立. Call remarks; venture terms, funding shares and timing are not established here.
引用於: 第節 第2.1節, 第2.2節, 第9.1節

S-075 Microsoft Corporation (Investor Relations) — Microsoft Fiscal Year 2026 Fourth Quarter Earnings Conference Call
企業新聞稿 · U.S. · 英文 · 29 July 2026 · 閱讀 27 September 2026
https://www.microsoft.com/en-us/investor/events/fy-2026/earnings-fy-2026-q4
"Outside of this useful life impact, our calendar year 2026 CapEx investment expectations remain unchanged. However, the shift from finance to operating leases adjusts our expectation to approximately $175 billion."
"we remain focused on delivering efficiencies that help us bridge the gaps we see as customer demand continues to exceed supply"
"All up, we added another gigawatt of capacity this quarter and remain on track to roughly double our overall capacity in just two years."
"The greater impact is on capital expenditures as more of our future datacenter leases will shift from finance leases to operating leases as a result of this update. Finance leases are included in capital expenditures while operating leases are not."
支持的研判. Microsoft re-bases its calendar-2026 capex expectation to about US$175 billion, investment unchanged, as longer useful lives shift leases outside capex; it remains on track to roughly double capacity.
無法據以確立. Management statements; the two-year doubling goal dates from October 2025 (S-462), and the capex figure is a basis change, not a cut.
引用於: 第節 第2.1節, 第2.2節, 第3.2節, 第9.1節, 第C.5節, 第C.7節; 圖 4, 5, 8; 表 8, 19, 24, 38, 39; 展望 9

S-076 StockAnalysis (transcript of Oracle Corporation Q1 FY2027 earnings call) — Oracle (ORCL) Q1 2027 Earnings Call Transcript
媒體文字紀錄 · U.S. · 英文 · 10 September 2026 · 閱讀 27 September 2026
https://stockanalysis.com/stocks/orcl/transcripts/689425-q1-2027/
"We continue to anticipate $90 billion-$95 billion in CapEx for the full year, with not more than $70 billion in net cash CapEx."
"We delivered 850 megawatts of AI capacity containing more than 300,000 GPUs to customers since the end of Q4."
支持的研判. Oracle anticipates FY2027 capex US$90–95 billion with at most US$70 billion net cash, expects around half of RPO to convert within 36 months, and reports 850 MW delivered since Q4.
無法據以確立. Third-party transcript; the 850 MW covers the period since the end of fiscal Q4 (31 May) to the 10 September call, not exactly the quarter.
引用於: 第節 第2.1節, 第2.2節; 圖 4

S-077 CoreWeave, Inc. (Investor Relations; FactSet CallStreet corrected transcript) — Corrected Transcript: CoreWeave, Inc. (CRWV-US) Q4 2025 Earnings Call, 26-February-2026
企業新聞稿 · U.S. · 英文 · 26 February 2026 · 閱讀 27 September 2026 · **財測指引已修訂**
https://s205.q4cdn.com/133937190/files/doc_financials/2025/q4/CORRECTED-TRANSCRIPT-CoreWeave-Inc-CRWV-US-Q4-2025-Earnings-Call-26-February-2026-5-00-PM-ET.pdf
"CapEx in Q4 totaled $8.2 billion and $14.9 billion for the full year, higher than anticipated due to our team's ability to put infrastructure in service ahead of our expectations."
"We expect 2026 CapEx of $30 billion to $35 billion, which is more than double our 2025 investment."
支持的研判. In February 2026 CoreWeave said 2025 capex was US$14.9 billion, expected 2026 capex of US$30–35 billion, and expected virtually all its 3.1 GW contracted at end-2025 online by end-2027.
無法據以確立. Call statements; the 2026 range was raised twice; call capex differs from the cash-flow line.
財測指引已修訂. Its 2026 range of US$30–35 billion was raised to US$31–35 billion in May [S-078] and to US$35–39 billion in August 2026 [S-079].
引用於: 第節 第2.1節, 第2.2節; 圖 4, 5; 表 19

S-078 CoreWeave, Inc. (Investor Relations; FactSet CallStreet corrected transcript) — Corrected Transcript: CoreWeave, Inc. (CRWV-US) Q1 2026 Earnings Call, 7-May-2026
企業新聞稿 · U.S. · 英文 · 7 May 2026 · 閱讀 27 September 2026 · **財測指引已修訂**
https://s205.q4cdn.com/133937190/files/doc_financials/2026/q1/CoreWeave-Inc-CRWV-US-Q1-2026-Earnings-Call-7-May-2026-5_00-PM-ET.pdf
"For the full year, we now expect CapEx of $31 billion to $35 billion. The increase on the low end from our previous guidance is related to increases in component pricing."
"Active power now exceeds 1 gigawatt and our contracted power is more than 3.5 gigawatts, leaving us strongly positioned to meet our goal of reaching more than 8 gigawatts of active power by 2030."
支持的研判. CoreWeave said in May 2026 it expects 2026 capex of US$31–35 billion, the low end raised for component pricing, and targets more than 8 GW of active power by 2030.
無法據以確立. Guidance, later raised again in August; the 8 GW is a goal, not contracted capacity.
財測指引已修訂. Its 2026 range of US$31–35 billion was raised to US$35–39 billion in August 2026 [S-079].
引用於: 第節 第2.1節, 第2.2節, 第9.3節; 圖 5; 表 19

S-079 CoreWeave, Inc. (Investor Relations; FactSet CallStreet corrected transcript) — Corrected Transcript: CoreWeave, Inc. (CRWV-US) Q2 2026 Earnings Call, 11-August-2026
企業新聞稿 · U.S. · 英文 · 11 August 2026 · 閱讀 27 September 2026
https://s205.q4cdn.com/133937190/files/doc_financials/2026/q2/CRWV-US-CORRECTED-TRANSCRIPT-CoreWeave-Q2-2026-Earnings-Call-11August2026.pdf
"Contracted power grew to 3.7 gigawatts in Q2. Since quarter end, we have added roughly 500 megawatts, bringing contracted power to 4.2 gigawatts as of today."
"We remain firmly on track to reach at least 8 gigawatts by 2030."
支持的研判. CoreWeave's Q2 2026 call gives 4.2 GW contracted on 11 August, 2026 capex guidance of US$35–39 billion, over 1.85 GW active by year-end, and 'at least 8 gigawatts by 2030'.
無法據以確立. Management guidance and goals; the 8 GW remark does not say 'active power' (the Q1 call, S-078, does), and none is delivered capacity.
引用於: 第節 第2.1節, 第2.2節, 第9.3節, 第9.5節; 圖 4, 5, 8, 32; 展望 9; 表 19

S-080 Alphabet Inc. / U.S. SEC (EDGAR XBRL API) — Alphabet Inc. - company concept us-gaap:PaymentsToAcquirePropertyPlantAndEquipment (10-K FY2025; 10-Q Q1 and Q2 2026)
申報文件 · U.S. · 英文 · 23 July 2026 · 閱讀 27 September 2026
https://data.sec.gov/api/xbrl/companyconcept/CIK0001652044/us-gaap/PaymentsToAcquirePropertyPlantAndEquipment.json
"PaymentsToAcquirePropertyPlantAndEquipment: 2025-01-01 to 2025-12-31, val 91447000000, form 10-K, filed 2026-02-05 (accn 0001652044-26-000018)"
"PaymentsToAcquirePropertyPlantAndEquipment: 2026-01-01 to 2026-06-30, val 80598000000, form 10-Q, filed 2026-07-23 (accn 0001652044-26-000071)"
支持的研判. SEC XBRL data give Alphabet's purchases of property and equipment as US$91.447 billion for 2025 (US$52.535 billion for 2024) and US$80.598 billion for H1 2026.
無法據以確立. Cash-flow purchases only; excludes assets acquired under finance leases and is not comparable with other companies' capex definitions.
引用於: 圖 4; 第節 第2.1節

S-081 Microsoft Corporation (Investor Relations) — Microsoft Cloud and AI Strength Fuels Fourth Quarter Results
企業新聞稿 · U.S. · 英文 · 29 July 2026 · 閱讀 27 September 2026
https://www.microsoft.com/en-us/investor/earnings/fy-2026-q4/press-release-webcast
取自來源表格之數值（轉錄）: CASH FLOWS STATEMENTS, line 'Additions to property and equipment': Twelve Months Ended June 30, 2026 (115,948); Twelve Months Ended June 30, 2025 (64,551) (In millions)
支持的研判. Microsoft's cash-flow statement shows additions to property and equipment of US$115.9 billion in fiscal 2026 (to June) against US$64.6 billion in fiscal 2025.
無法據以確立. Cash additions exclude finance leases, so they differ from Microsoft's quoted capex; company-wide, not AI-only.
引用於: 圖 4

S-082 Taiwan Semiconductor Manufacturing Company Limited (TSMC) — 2330.TW - Q4 2025 Taiwan Semiconductor Manufacturing Co Ltd Earnings Call (Chinese, English)
企業新聞稿 · Taiwan · 英文 · 15 January 2026 · 閱讀 27 September 2026 · **財測指引已修訂**
https://investor.tsmc.com/english/quarterly-results/2025/q4
直接文件: https://investor.tsmc.com/english/encrypt/files/encrypt_file/reports/2026-01/51d09df96cd89ac19d65af39032b038dc2896a24/TSMC%204Q25%20Transcript.pdf
"Based on our planning framework, we raise our forecast for the revenue growth from AI Accelerators to approach a mid-to-high-fifties % CAGR for the (added by company after the call) five-year period from 2024 to 2029."
"In 2026, we expect our capital budget to be between USD 52 billion and USD 56 billion as we continue to invest to support our customers' growth."
支持的研判. TSMC said in January 2026 it expects a 2026 capital budget of US$52–56 billion and raised its AI-accelerator revenue growth forecast to approach a mid-to-high-fifties % CAGR for 2024–2029.
無法據以確立. Company guidance and forecast for its own business; the capital budget was raised in April and July 2026.
財測指引已修訂. Its 2026 capital budget of US$52–56 billion moved to the high end of that range in April [S-083] and to US$60–64 billion in July 2026 [S-084].
引用於: 第節 第2.1節, 第3.4節; 圖 5; 表 2, 4, 26

S-083 Taiwan Semiconductor Manufacturing Company Limited (TSMC) — 2330.TW - Q1 2026 Taiwan Semiconductor Manufacturing Co Ltd Earnings Call
企業新聞稿 · Taiwan · 英文 · 16 April 2026 · 閱讀 27 September 2026 · **財測指引已修訂**
https://investor.tsmc.com/english/encrypt/files/encrypt_file/reports/2026-04/3cef85204275f94fd111485cfdf4adb3c0263c45/TSMC%201Q26%20Transcript.pdf
"We now expect our 2026 capital budget to be towards the high end of our range of between USD52 billion and USD56 billion as we continue to invest heavily to support our customers' growth."
"In Arizona, our second fab will also utilize 3-nanometer technologies. Construction is already complete and volume production will begin in the second half of 2027."
支持的研判. In April 2026 TSMC said its second Arizona fab will begin volume production in 2H 2027 and its 2026 capital budget would be toward the high end of US$52–56 billion.
無法據以確立. Company schedule and guidance; dates can slip, and the capital budget was raised again in July 2026.
財測指引已修訂. The 2026 capital budget was raised to US$60–64 billion in July 2026 [S-084].
引用於: 第節 第2.1節, 第3.4節, 第5.7節, 第7.5節, 第9.2節, 第B.2節; 圖 5, 32; 表 4, 9, 10, 19, 26; 展望 12

S-084 Taiwan Semiconductor Manufacturing Company Limited (TSMC) — 2330.TW - Q2 2026 Taiwan Semiconductor Manufacturing Co Ltd Earnings Call (Chinese, English)
企業新聞稿 · Taiwan · 英文 · 16 July 2026 · 閱讀 27 September 2026
https://investor.tsmc.com/english/quarterly-results/2026/q2
直接文件: https://investor.tsmc.com/english/encrypt/files/encrypt_file/reports/2026-08/3e494f0c14dd0890f897aa044415e21d93486cc4/TSMC%202Q26%20Transcript.pdf
"Given the continued strong structural demand from our customers, including the newly emerging agentic AI market, we have decided to raise our full-year 2026 capital budget to be between USD60 billion and USD64 billion as we continue to invest heavily to support our customers' growth."
"Last time, we said our CapEx in the next three years will be significantly higher than the CapEx in the past three years. Now the CapEx in the next three years will be even more significantly higher than the past three years."
"But stronger than what we said before."
"We started the production right now, and it will be ramped up."
支持的研判. In July 2026 TSMC raised its 2026 capital budget to US$60–64 billion, said COUPE production had started, to be ramped, and announced US$100 billion more Arizona fab and packaging investment.
無法據以確立. Earnings-call guidance and plans; no COUPE volumes; the US$265 billion total is the Commerce Department's figure, not TSMC's.
引用於: 第節 第2.1節, 第3.4節, 第4.7節, 第5.1節, 第5.7節, 第5.8節, 第6.7節, 第7.5節, 第9.4節, 第C.7節; 圖 5, 32; 表 4, 6, 7, 9, 10, 13, 19; 展望 4, 12

S-085 The Motley Fool (transcript of Amazon.com, Inc. Q1 2026 earnings call) — Amazon (AMZN) Q1 2026 Earnings Call Transcript
媒體文字紀錄 · U.S. · 英文 · 29 April 2026 · 閱讀 28 September 2026
https://www.fool.com/earnings/call-transcripts/2026/04/29/amazon-amzn-q1-2026-earnings-call-transcript/
"can you just give us an update on what the AWS backlog looks like and any visibility on the breadth of that backlog beyond the big labs?"
"Our cash CapEx is $43.2 billion in Q1. This primarily relates to AWS and generative AI, as we invest to support strong customer demand."
"the backlog for Q1 is $364 billion. That does not include the recent deal that we announced with Anthropic for over $100 billion."
支持的研判. Amazon reported Q1 2026 cash capex of US$43.2 billion and, answering a question on the AWS backlog, a Q1 backlog of US$364 billion, excluding a recently announced Anthropic deal of over US$100 billion.
無法據以確立. Third-party transcript; it does not date the Anthropic announcement, only that it was recent and excluded from the Q1 backlog.
引用於: 第節 第2.1節, 第2.2節; 圖 8

S-086 International Energy Agency (IEA) — Data centre electricity use surged in 2025, even with tightening bottlenecks driving a scramble for solutions
政府 · Global · 英文 · 16 April 2026 · 閱讀 27 September 2026
https://www.iea.org/news/data-centre-electricity-use-surged-in-2025-even-with-tightening-bottlenecks-driving-a-scramble-for-solutions
"As a result, electricity consumption from data centres is set to double by 2030, and power use from those focused on AI is poised to triple."
"Electricity demand from data centres soared by 17% in 2025, and that of AI-focused data centres climbed even faster – well outpacing growth in global electricity demand of 3%."
支持的研判. IEA's April 2026 release says data-centre electricity demand rose 17% in 2025 against 3% globally, is set to double by 2030, and five tech companies' capex exceeded US$400 billion.
無法據以確立. A press summary of IEA projections; the capex figure is five companies' total capital spending, not data-centre spending alone.
引用於: 第一頁; 第節 第2.1節, 第2.5節, 第E.1節, 第E.6節, 第E.8節; 圖 11; 表 36, 38

S-087 Microsoft Corporation (Investor Relations) — Microsoft Fiscal Year 2025 Fourth Quarter Earnings Conference Call
企業新聞稿 · U.S. · 英文 · 30 July 2025 · 閱讀 27 September 2026
https://www.microsoft.com/en-us/investor/events/fy-2025/earnings-fy-2025-q4
"Capital expenditures were $24.2 billion, including $6.5 billion of finance leases where we recognize the full value at the time of lease commencement."
支持的研判. Microsoft's FY2025 Q4 call shows its capital expenditures include finance leases recognised at full value at lease commencement: US$6.5 billion of US$24.2 billion that quarter.
無法據以確立. Documents Microsoft's reporting basis for one quarter; says nothing about other companies' definitions.
引用於: 第節 第2.1節

S-088 CoreWeave, Inc. / U.S. SEC (EDGAR, Exhibit 99.1) — CoreWeave Reports Strong Fourth Quarter and Fiscal Year 2025 Results
申報文件 · U.S. · 英文 · 26 February 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/1769628/000176962826000094/coreweave4q25earningspress.htm
"Revenue backlog was $66.8 billion as of December 31, 2025."
"Expanded total contracted power to approximately 3.1 GW while further diversifying our portfolio"
支持的研判. CoreWeave reports end-2025 revenue backlog of US$66.8 billion, active power above 850 MW, contracted power of about 3.1 GW, and 2025 cash purchases of property and equipment of US$10.3 billion.
無法據以確立. Company results; contracted power is not active capacity; the capex stated on the call is not in this release.
引用於: 第節 第2.1節, 第2.2節; 圖 8

S-089 The Motley Fool (transcript of Amazon.com, Inc. Q4 2025 earnings call) — Amazon (AMZN) Q4 2025 Earnings Call Transcript
媒體文字紀錄 · U.S. · 英文 · 5 February 2026 · 閱讀 27 September 2026
https://www.fool.com/earnings/call-transcripts/2026/02/05/amazon-amzn-q4-2025-earnings-call-transcript/
"our backlog is $244 billion. That's up 40% year over year."
支持的研判. Amazon's CEO said on the Q4 2025 call that ' our backlog is $244 billion', up 40% year over year.
無法據以確立. Third-party transcript; the answer does not name AWS (the later calls give the series as the AWS backlog [S-085; S-070]); backlog is contracted future revenue, not recognised revenue or capacity.
引用於: 第節 第2.2節; 圖 8

S-090 Oracle Corporation (Investor Relations) — Oracle Announces Fiscal Year 2026 Second Quarter Financial Results
企業新聞稿 · U.S. · 英文 · 10 December 2025 · 閱讀 27 September 2026
https://investor.oracle.com/investor-news/news-details/2025/Oracle-Announces-Fiscal-Year-2026-Second-Quarter-Financial-Results/default.aspx
"Total Remaining Performance Obligations were up 438% year-over-year in USD to $523 billion."
支持的研判. Oracle reports remaining performance obligations of US$523 billion at the end of its fiscal Q2 2026 (November 2025), up 438% year on year.
無法據以確立. RPO is contracted, not recognised revenue; conversion timing is not given here.
引用於: 圖 8

S-091 Oracle Corporation (Investor Relations) — Oracle Announces Fiscal Year 2026 Third Quarter Financial Results
企業新聞稿 · U.S. · 英文 · 10 March 2026 · 閱讀 27 September 2026
https://investor.oracle.com/investor-news/news-details/2026/Oracle-Announces-Fiscal-Year-2026-Third-Quarter-Financial-Results/default.aspx
直接文件: https://s23.q4cdn.com/440135859/files/doc_financials/2026/q3/3q26-pressrelease-March-FINAL.pdf
"Remaining Performance Obligations, or RPO, ended the quarter at $553 billion, up 325% from last year and up $29 billion from last quarter."
支持的研判. Oracle reports RPO of US$553 billion at the end of fiscal Q3 2026 (February 2026), up 325% year on year and US$29 billion on the prior quarter.
無法據以確立. RPO is contracted, not recognised revenue; conversion timing is not given here.
引用於: 圖 8

S-092 CoreWeave, Inc. (Investor Relations) — CoreWeave Reports Strong First Quarter 2026 Results
企業新聞稿 · U.S. · 英文 · 7 May 2026 · 閱讀 27 September 2026
https://investors.coreweave.com/news/news-details/2026/CoreWeave-Reports-Strong-First-Quarter-2026-Results/default.aspx
"Revenue backlog was $99.4 billion as of March 31, 2026."
"Expanded total contracted power by more than 400 MW to over 3.5 GW while further diversifying portfolio of providers"
支持的研判. CoreWeave reports revenue backlog of US$99.4 billion at 31 March 2026, active power above 1 GW and contracted power above 3.5 GW.
無法據以確立. Company measures; revenue backlog adds amounts beyond RPO, and contracted power is not active power.
引用於: 圖 8

S-093 Investing.com (transcript of Amazon.com, Inc. Q4 2025 earnings call) — Earnings call transcript: Amazon's Q4 2025 revenue beats forecast, stock dips
媒體文字紀錄 · U.S. · 英文 · 5 February 2026 · 閱讀 27 September 2026
https://www.investing.com/news/transcripts/earnings-call-transcript-amazons-q4-2025-revenue-beats-forecast-stock-dips-93CH-4489218
"If you look, you know, in the last 12 months, we added 3.99 gigawatts of power."
"We expect to double it again by the end of 2027."
支持的研判. Amazon's CEO said on the Q4 2025 call it added 3.99 GW of power in the last 12 months and expects 'to double it again' by end-2027.
無法據以確立. Third-party transcript of management remarks; power added is not IT capacity in service, and the doubling is an expectation.
引用於: 第節 第2.2節

S-094 Microsoft Corporation / U.S. SEC (EDGAR) — Microsoft Corporation Form 10-K for the fiscal year ended June 30, 2026 - Supplemental Cash Flow Information Related to Leases (Detail); Leases note
申報文件 · U.S. · 英文 · 29 July 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/789019/000119312526323660/R91.htm
直接文件: https://www.sec.gov/Archives/edgar/data/789019/000119312526323660/msft-20260630.htm
"As of June 30, 2026, we had additional leases, primarily for datacenters, that had not yet commenced of $329.1 billion, with some arrangements subject to certain contractual conditions being met."
"These leases will commence between fiscal year 2027 and fiscal year 2033 with lease terms of 1 year to 20 years."
支持的研判. Microsoft's FY2026 Form 10-K reports US$329.1 billion of leases, primarily data centres, not yet commenced at 30 June 2026, some conditional, commencing FY2027–FY2033 with 1–20-year terms.
無法據以確立. Contractual lease commitments, some conditional; not capital expenditure and not delivered capacity.
引用於: 第節 第2.2節

S-095 CoreWeave, Inc. (Investor Relations) / NVIDIA Corporation — NVIDIA and CoreWeave Strengthen Collaboration to Accelerate Buildout of AI Factories
企業新聞稿 · U.S. · 英文 · 26 January 2026 · 閱讀 27 September 2026
https://investors.coreweave.com/news/news-details/2026/NVIDIA-and-CoreWeave-Strengthen-Collaboration-to-Accelerate-Buildout-of-AI-Factories/default.aspx
直接文件: https://nvidianews.nvidia.com/news/nvidia-and-coreweave-strengthen-collaboration-to-accelerate-buildout-of-ai-factories
"NVIDIA has invested $2 billion in CoreWeave Class A common stock at a purchase price of $87.20 per share."
"NVIDIA and CoreWeave, Inc. (NASDAQ: CRWV) today announced an expansion of their long-standing complementary relationship to enable CoreWeave to accelerate the buildout of more than 5 gigawatts of AI factories by 2030 to advance AI adoption at global scale."
支持的研判. NVIDIA invested US$2 billion in CoreWeave Class A stock at US$87.20 a share in January 2026 as they expanded collaboration toward more than 5 GW of AI factories by 2030.
無法據以確立. Announcement; the 5 GW is a 2030 build-out goal, not contracted or delivered capacity; guarantees are not covered.
引用於: 第節 第2.2節

S-096 Amazon.com, Inc. (Investor Relations) — Amazon.com Announces Second Quarter Results
企業新聞稿 · U.S. · 英文 · 30 July 2026 · 閱讀 27 September 2026
https://ir.aboutamazon.com/news-release/news-release-details/2026/Amazon-com-Announces-Second-Quarter-Results/default.aspx
"Free cash flow decreased to an outflow of $7.6 billion for the trailing twelve months, driven primarily by a year-over-year increase of $66.1 billion in purchases of property and equipment, net of proceeds from sales and incentives."
支持的研判. Amazon reports trailing-twelve-month free cash flow turned to an outflow of US$7.6 billion to June 2026, driven by a US$66.1 billion rise in net purchases of property and equipment.
無法據以確立. Company-wide cash measure; it does not separate AI or AWS capital spending, and free cash flow is Amazon's own non-GAAP definition.
引用於: 第節 第2.2節

S-097 NVIDIA Corporation — CFO Commentary on First Quarter Fiscal 2027 Results
企業新聞稿 · U.S. · 英文 · 20 May 2026 · 閱讀 27 September 2026
https://s201.q4cdn.com/141608511/files/doc_financials/2027/Q127/Q1FY27-CFO-Commentary.pdf
"No shipments of Data Center Hopper products to China occurred during the quarter, compared with $4.6 billion in the first quarter of fiscal year 2026."
"We will have two market platforms – Data Center and Edge Computing. Within Data Center, we will report two sub-markets, Hyperscale and ACIE which incorporates AI Clouds, Industrial, and Enterprise."
支持的研判. NVIDIA's CFO commentary reports no Data Center Hopper shipments to China in Q1 FY2027 (US$4.6 billion a year earlier) and a new framework with Data Center and Edge Computing platforms.
無法據以確立. Company commentary; covers Hopper shipments only, not all China revenue or other products.
引用於: 第節 第2.3節, 第7.1節; 圖 7

S-098 NVIDIA Corporation — NVIDIA Announces Financial Results for First Quarter Fiscal 2025
企業新聞稿 · U.S. · 英文 · 22 May 2024 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-first-quarter-fiscal-2025
"First-quarter revenue was a record $22.6 billion, up 23% from the previous quarter and up 427% from a year ago."
支持的研判. NVIDIA reported record Q1 FY2025 Data Center revenue of US$22.6 billion, up 23% on the quarter and 427% on the year.
無法據以確立. As first reported, company segment measure; later recasts are not reflected.
引用於: 第節 第2.3節; 圖 7

S-099 NVIDIA Corporation — NVIDIA Announces Financial Results for Second Quarter Fiscal 2025
企業新聞稿 · U.S. · 英文 · 28 August 2024 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-second-quarter-fiscal-2025
"Second-quarter revenue was a record $26.3 billion, up 16% from the previous quarter and up 154% from a year ago."
支持的研判. NVIDIA reported record Q2 FY2025 Data Center revenue of US$26.3 billion, up 16% on the quarter and 154% on the year.
無法據以確立. As first reported, company segment measure; later recasts are not reflected.
引用於: 圖 7

S-100 NVIDIA Corporation — NVIDIA Announces Financial Results for Third Quarter Fiscal 2025
企業新聞稿 · U.S. · 英文 · 20 November 2024 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-third-quarter-fiscal-2025
"Third-quarter revenue was a record $30.8 billion, up 17% from the previous quarter and up 112% from a year ago."
支持的研判. NVIDIA reported record Q3 FY2025 Data Center revenue of US$30.8 billion, up 17% on the quarter and 112% on the year.
無法據以確立. As first reported, company segment measure; later recasts are not reflected.
引用於: 圖 7

S-101 NVIDIA Corporation — NVIDIA Announces Financial Results for Fourth Quarter and Fiscal 2025
企業新聞稿 · U.S. · 英文 · 26 February 2025 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-fourth-quarter-and-fiscal-2025
"Fourth-quarter revenue was a record $35.6 billion, up 16% from the previous quarter and up 93% from a year ago."
支持的研判. NVIDIA reported record Q4 FY2025 Data Center revenue of US$35.6 billion, up 16% on the quarter and 93% on the year.
無法據以確立. As first reported, company segment measure; later recasts are not reflected.
引用於: 圖 7

S-102 NVIDIA Corporation — NVIDIA Announces Financial Results for First Quarter Fiscal 2026
企業新聞稿 · U.S. · 英文 · 28 May 2025 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-first-quarter-fiscal-2026
"First-quarter revenue was $39.1 billion, up 10% from the previous quarter and up 73% from a year ago."
"On April 9, 2025, NVIDIA was informed by the U.S. government that a license is required for exports of its H20 products into the China market."
支持的研判. NVIDIA reported Q1 FY2026 Data Center revenue of US$39.1 billion, said H20 exports to China needed a licence from 9 April 2025, and took a US$4.5 billion H20 charge.
無法據以確立. Company account; the charge covers H20 excess inventory and purchase obligations, not lost sales.
引用於: 圖 7; 第節 第7.1節, 第C.7節; 表 14

S-103 NVIDIA Corporation — NVIDIA Announces Financial Results for Second Quarter Fiscal 2026
企業新聞稿 · U.S. · 英文 · 27 August 2025 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-second-quarter-fiscal-2026
"Revenue is expected to be $54.0 billion, plus or minus 2%. The company has not assumed any H20 shipments to China in the outlook."
"There were no H20 sales to China-based customers in the second quarter."
支持的研判. NVIDIA's August 2025 outlook for Q3 FY2026 assumed no H20 shipments to China; Q2 FY2026 had no H20 sales to China-based customers and Data Center revenue of US$41.1 billion.
無法據以確立. One outlook, for Q3 FY2026 only; it does not show the wording of other fiscal-2026 outlooks, and covers H20, not all China revenue.
引用於: 圖 7; 第節 第2.3節, 第7.1節; 表 14

S-104 NVIDIA Corporation — NVIDIA Announces Financial Results for Third Quarter Fiscal 2026
企業新聞稿 · U.S. · 英文 · 19 November 2025 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-announces-financial-results-for-third-quarter-fiscal-2026
"Third-quarter revenue was a record $51.2 billion, up 25% from the previous quarter and up 66% from a year ago."
支持的研判. NVIDIA reported record Q3 FY2026 Data Center revenue of US$51.2 billion for the quarter ended 26 October 2025.
無法據以確立. As first reported, company segment measure; later recasts are not reflected.
引用於: 圖 7

S-105 TrendForce — NVIDIA's AI Rack Transition Accelerates; NVL72 Output Value to Exceed US$710 Billion in 2027, Says TrendForce
市場研究 · Taiwan · 英文 · 28 August 2026 · 閱讀 26 September 2026
https://www.trendforce.com/presscenter/news/20260828-13204.html
"their share of procurement is expected to decline from approximately 70% in 2025 to 60% in 2026"
"combined output value of GB300, VR200, and VR300 NVL72 rack systems is projected to exceed US$710 billion in 2027"
支持的研判. TrendForce expects North American CSPs' share of NVIDIA rack procurement to fall from about 70% (2025) to 60% (2026), and NVL72 rack output value to exceed US$710 billion in 2027.
無法據以確立. Market-research projections; output value is not units or NVIDIA revenue, and the excerpt's 'their' means the five major North American CSPs.
引用於: 第節 第2.3節, 第4.5節

S-106 IDC for the Open Compute Project Foundation (hosted on OCP's CDN) — IDC Market Impact Study (cover: OCP Global Summit 2025; Chris Drake, Senior Research Director, Compute Infrastructure and Service Provider Trends)
市場研究 · Global · 英文 · 12 November 2025 · 閱讀 27 September 2026 · **範圍不同**
https://146a55aca6f00848c565-a7635525d40ac1c70300198708936b4e.ssl.cf1.rackcdn.com/images/2538f9e7afe4473fb199c4f688b9d954a101db6a.pdf
"Total spending on OCP-recognized IT infrastructure and solutions will grow to US$295 billion in 2029, up from US$132 billion in 2025, reflecting a compound annual growth rate (CAGR) of 22.2%."
支持的研判. IDC's study deck for OCP projects spending on OCP-recognized IT infrastructure and solutions rising from US$132 billion in 2025 to US$295 billion in 2029, a 22.2% CAGR.
無法據以確立. IDC forecast of standards-linked equipment spending, not the AI data-centre market; the deck's only AI figure is about US$90 billion by 2028, not US$155 billion.
範圍不同. The deck's only AI figure is about US$90 billion for 2028; the adoption page gives around US$155 billion for 2029 [S-004]. Different years and presentations of IDC's study; the report uses the 2029 figure for the AI use case.
引用於: 第節 第2.4節, 第C.4節; 圖 9

S-107 IDC — IDC Market Impact Study — OCP Global Summit 2025 (IDC document US53932625)
市場研究 · Global · 英文 · December 2025 · 閱讀 27 September 2026
https://my.idc.com/getdoc.jsp?containerId=US53932625
"The study included a series of in-depth interviews with IT and datacenter infrastructure vendors that have incorporated OCP-recognized standards into their product and solution offerings."
"In addition to its qualitative analysis of the market for OCP-specified infrastructure, IDC has developed a market model and growth forecast for OCP technology adoption, the top-level highlights of which are included in this presentation."
支持的研判. IDC says its OCP market study, done for the Open Compute Foundation, combines interviews with vendors using OCP-recognized standards and an IDC market model and adoption forecast.
無法據以確立. Commissioned by the Open Compute Foundation; the presentation gives only top-level highlights, not the model's definitions or data.
引用於: 第節 第2.4節

S-108 Open Compute Project Foundation / IDC (presenters: Chris Drake and Kuba Stolarski, IDC; Steve Helvie, OCP Foundation) — Understanding the Global Adoption of OCP Designs: A view from IDC on adoption, spend, & buyer intent (OCP Webinar, January 22, 2026)
市場研究 · Global · 英文 · 22 January 2026 · 閱讀 27 September 2026 · **範圍不同**
https://www.opencompute.org/events/past-events/webinar-understanding-the-global-adoption-of-ocp-designs-in-infrastructure-build-outs-a-view-from-idc-on-adoption-spend-and-buyer-intent
直接文件: https://146a55aca6f00848c565-a7635525d40ac1c70300198708936b4e.ssl.cf1.rackcdn.com/images/ef1c0b9cdbf6d8b94de6de7807b988db0f2de920.pdf
"OCP infrastructure spending by use case, 2019-2029 (USD $B)"
"Total spending on OCP-recognized IT infrastructure and solutions will grow to US$295 billion in 2029, up from US$132 billion in 2025, reflecting a CAGR of 22.2%."
支持的研判. IDC's January 2026 OCP webinar deck charts OCP infrastructure spending by use case, AI being one slice (labelled 156 in 2029) of a total rising to US$295 billion.
無法據以確立. OCP-commissioned IDC forecast; the 156 is a chart label, not a sentence, and the use-case slice is not the whole AI-infrastructure market.
範圍不同. Labels the AI use case 156 (US$ billion, 2029) inside the OCP-recognized total, where the adoption page says around US$155 billion [S-004]: the same slice, rounded differently, and not the whole AI-infrastructure market.
引用於: 第節 第2.4節, 第9.3節, 第C.2節, 第C.4節, 第C.7節; 圖 9; 展望 11

S-109 Open Compute Project Foundation (blog, Ewa McCurdy) — Evolving with Purpose: Streamlining OCP Product Recognition
標準組織 · Global · 英文 · 11 May 2026 · 閱讀 27 September 2026
https://www.opencompute.org/blog/evolving-with-purpose-streamlining-ocp-product-recognition
"Solution Providers may continue using the OCP Inspired™ logo through July 31, 2026, after which it will be formally retired."
"A 30-day feedback period begins today, during which we welcome input from the OCP Community."
支持的研判. OCP announced on 11 May 2026 that the OCP Inspired mark would be retired after 31 July 2026, leaving OCP Accepted, which requires review against approved contributions.
無法據以確立. A plan opened for 30 days' feedback; the post does not confirm the retirement took effect, nor how recognised products are counted in market studies.
引用於: 第節 第2.4節, 第6.1節, 第6.2節, 第C.7節; 附錄A

S-110 International Energy Agency (IEA) — Executive summary – Energy and AI
政府 · Global · 英文 · 10 April 2025 · 閱讀 27 September 2026
https://www.iea.org/reports/energy-and-ai/executive-summary
直接文件: https://iea.blob.core.windows.net/assets/de9dea13-b07d-42c5-a398-d1b3ae17d866/EnergyandAI.pdf
"Data centre electricity consumption is set to more than double to around 945 TWh by 2030."
支持的研判. The IEA's April 2025 Energy and AI report projected data-centre electricity consumption more than doubling to around 945 TWh by 2030.
無法據以確立. Base Case projection from the 2025 edition; the IEA updated its projections in April 2026.
引用於: 第節 第2.5節

S-111 AMD Investor Relations — AMD Reports Second Quarter 2026 Financial Results
企業新聞稿 · U.S. · 英文 · 4 August 2026 · 閱讀 26 September 2026
https://ir.amd.com/news-events/press-releases/detail/1295/amd-reports-second-quarter-2026-financial-results
"AMD Helios delivers leadership inference tokens per dollar and is being deployed by leading AI labs and cloud providers"
"For the third quarter of 2026, AMD expects revenue to be approximately $13 billion, plus or minus $300 million."
"Data Center segment revenue was $6.7 billion, up 107% year-over-year, driven by strong demand for AMD EPYC™ processors and AMD Instinct™ GPUs."
支持的研判. AMD reports Q2 2026 Data Center segment revenue of US$6.7 billion, up 107%, and says Helios 'is being deployed by leading AI labs and cloud providers'.
無法據以確立. Company results; the segment includes EPYC CPUs as well as GPUs, and no Helios volumes or revenue are disclosed.
引用於: 第節 第1.3節, 第3.1節, 第7.1節, 第A.2節; 表 8, 14, 20, 21, 24; 展望 6; 附錄A · **研判:** A-08

S-112 AMD — AMD and Meta Announce Expanded Strategic Partnership to Deploy 6 Gigawatts of AMD GPUs
企業新聞稿 · U.S. · 英文 · 24 February 2026 · 閱讀 27 September 2026
https://newsroom.amd.com/news/amd-and-meta-announce-expanded-strategic-partnersh
"developed jointly by AMD and Meta through the Open Compute Project"
"AMD and Meta agree to a definitive multi-year, multi-generation partnership to deploy up to 6 gigawatts of AMD Instinct™ GPUs."
支持的研判. AMD and Meta agreed to deploy up to 6 GW of AMD GPUs, with first-gigawatt shipments scheduled from 2H 2026; AMD says Helios was developed jointly with Meta through OCP.
無法據以確立. Company agreement with an 'up to' ceiling; not delivered capacity, and 'developed through OCP' is AMD's description, not an OCP specification.
引用於: 第節 第3.1節, 第4.5節, 第6.2節, 第C.7節; 展望 5; 表 24

S-113 Broadcom — Broadcom Inc. Announces Third Quarter Fiscal Year 2026 Financial Results and Quarterly Dividend
企業新聞稿 · U.S. · 英文 · 2 September 2026 · 閱讀 26 September 2026
https://investors.broadcom.com/news-releases/news-release-details/broadcom-inc-announces-third-quarter-fiscal-year-2026-financial
"Demand for our custom AI accelerators and networking continues to be very strong."
"Q3 AI semiconductor revenue of $16.7 billion grew 221% year-over-year, and 54% quarter-over-quarter"
支持的研判. Broadcom reports Q3 FY2026 AI semiconductor revenue of US$16.7 billion (up 221%), guides US$21.7 billion for Q4, and says custom-accelerator and networking demand 'continues to be very strong'.
無法據以確立. Company results and one-quarter guidance; its 'AI semiconductor' category is Broadcom's own definition, not a market total or unit volume.
引用於: 第節 第3.1節; 表 24

S-114 TrendForce — Aggressive Procurement of NVIDIA GB and Rubin Racks by North American CSPs Will Drive a 1.2x Surge in AI Inference Computing Power in 2026, Says TrendForce
市場研究 · Taiwan · 英文 · 20 May 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260520-13053.html
"Trainium series will account for over 40% of Amazon's own AI server shipments in 2026"
"These five companies are expected to account for over 60% of global demand for NVIDIA's GB and VR series servers this year."
支持的研判. TrendForce forecasts the top five North American CSPs at over 60% of 2026 demand for NVIDIA GB- and VR-series servers, and Trainium at over 40% of Amazon's own AI-server shipments.
無法據以確立. Market-research forecasts for 2026 of unit demand and shipment shares, not spending or delivered systems.
引用於: 第節 第3.1節, 第5.7節; 圖 25

S-115 Samsung Electronics Co., Ltd. (joint announcement with Broadcom Inc.) — Samsung Electronics and Broadcom Expand Strategic Collaboration Across Memory and Foundry Technologies
企業新聞稿 · Korea / U.S. · 英文 · 25 July 2026 · 閱讀 27 September 2026
https://news.samsung.com/global/samsung-electronics-and-broadcom-expand-strategic-collaboration-across-memory-and-foundry-technologies
"The companies expect the collaboration to be estimated at more than $200 billion across memory and foundry over the next five years through 2030."
"On the memory front, Samsung and Broadcom plan to pursue a strategic collaboration for the supply of industry-leading memory solutions, including High Bandwidth Memory (HBM), supporting Broadcom's next-generation AI accelerators."
支持的研判. Samsung and Broadcom signed an MOU (25 July 2026) and estimate their memory-and-foundry collaboration at more than US$200 billion through 2030, including planned HBM supply for Broadcom AI accelerators.
無法據以確立. MOU-stage joint estimate; not revenue, orders or a definitive supply agreement, and the value's basis is undefined.
引用於: 第節 第3.1節, 第5.6節, 第9.3節; 表 3, 25, 35; 展望 10 · **研判:** A-05

S-116 Advantest Corporation — Advantest Corporation Q1 FY2026 (Three months ended June 30th, 2026) Financial Briefing Q&A Summary
企業新聞稿 · Japan · 英文 · 29 July 2026 · 閱讀 27 September 2026
https://www.advantest.com/en/investors/ir-library/result/
直接文件: https://www.advantest.com/document/en/investors/ir-library/result/JE_BIZ_260729_QA.pdf
"And, while the market size for GPUs remains the largest today, custom ASICs could become as large as, if not significantly larger than, the GPU market over time."
"The next step could be at a scale of 15,000 systems or more."
"We are seeing very strong SoC tester demand across GPUs, custom ASICs and CPUs, and the increase in the SoC tester TAM is coming from more inference AI applications, particularly in the CPU business."
支持的研判. Advantest reports very strong SoC-tester demand across GPUs, custom ASICs and CPUs, ties tester capacity to packaging expansion, and is pulling in its 5,000- and 10,000-system targets; 15,000 'could' follow.
無法據以確立. Earnings-call remarks; the figures are capacity targets and a scenario, not orders, allocations or delivered systems; 'this year' may mean fiscal year.
引用於: 第節 第3.1節, 第3.5節, 第5.6節, 第5.7節, 第5.8節, 第9.1節, 第9.3節, 第C.5節, 第C.7節; 表 7, 9, 10, 19, 34; 圖 24, 25; 展望 10

S-117 TrendForce — 2Q25 DRAM Revenue Jumps 17.1%, SK hynix Expands Market Share as Taiwanese Vendors Deliver Strong Growth, Says TrendForce
市場研究 · Taiwan · 英文 · 2 September 2025 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20250902-12694.html
"TrendForce's latest investigations demonstrate that global DRAM industry revenue reached US$31.63 billion in the second quarter of 2025, up 17.1% QoQ."
支持的研判. TrendForce estimates global DRAM industry revenue at US$31.63 billion in 2Q25, up 17.1% QoQ.
無法據以確立. Market-research estimate of revenue; later quarters come from separate TrendForce releases, each with its own basis.
引用於: 第一頁; 第節 第3.2節, 第E.2節; 圖 14; 表 36

S-118 TrendForce — DRAM Industry Revenue Rises 59.5% QoQ in 2Q26 as Supply Expansion Continues to Lag Demand Growth, Says TrendForce
市場研究 · Taiwan · 英文 · 7 September 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260907-13219.html
"TrendForce's latest findings on the memory industry reveal that sharp rises in traditional DRAM contract prices have led to a 59.5% QoQ increase in total DRAM industry revenue, reaching nearly US$154.73 billion in 2Q26."
支持的研判. TrendForce estimates total DRAM industry revenue rose 59.5% QoQ to nearly US$154.73 billion in 2Q26, driven by sharp rises in conventional DRAM contract prices.
無法據以確立. Market-research estimate of revenue, not bits shipped; the rise reflects prices, so it does not measure volume growth.
引用於: 第一頁; 第節 第3.2節, 第E.2節; 圖 14; 表 36

S-119 TrendForce — 2Q26 Revenue Ranking among Server DRAM Suppliers
市場研究 · Taiwan · 英文 · 9 September 2026 · 閱讀 27 September 2026
https://www.trendforce.com/research/download/RP260907MR
"In 2Q26, Server DRAM revenue surged 53% QoQ to $75.58B."
支持的研判. TrendForce estimates server DRAM revenue rose 53% QoQ to US$75.58 billion in 2Q26.
無法據以確立. Public abstract of a paid report; the share of total DRAM revenue is the report's calculation against a separate TrendForce total.
引用於: 第節 第3.2節; 圖 14

S-120 TrendForce — Global DRAM Revenue Jumps 30.9% in 3Q25, Micron's Market Share Climbs by 3.7 Percentage Points, Says TrendForce
市場研究 · Taiwan · 英文 · 26 November 2025 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20251126-12802.html
"TrendForce's latest research shows that significant increases in conventional DRAM contract prices, higher bit shipments, and growing HBM volumes drove the global DRAM industry revenue to US$41.4 billion in 3Q25, marking a strong 30.9% QoQ growth."
支持的研判. TrendForce estimates global DRAM industry revenue rose 30.9% QoQ to US$41.4 billion in 3Q25, on higher contract prices, bit shipments and HBM volumes.
無法據以確立. Market-research estimate of revenue; the series includes HBM and mixes price and volume effects.
引用於: 第一頁; 圖 14; 第節 第E.2節

S-121 TrendForce — Price Rally Drives 4Q25 DRAM Revenue Up 29.4%; Samsung Regains No. 1 Market Share, Says TrendForce
市場研究 · Taiwan · 英文 · 26 February 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260226-12937.html
"Aggressive additional orders have driven a sharp increase in conventional DRAM contract prices, lifting total DRAM industry revenue to $53.58 billion in 4Q25, up 29.4% QoQ."
支持的研判. TrendForce estimates total DRAM industry revenue rose 29.4% QoQ to US$53.58 billion in 4Q25 on sharply higher conventional DRAM contract prices.
無法據以確立. Market-research estimate of revenue, not bit volumes; one quarter of the chart series.
引用於: 第一頁; 圖 14; 第節 第E.2節

S-122 SK hynix Inc. — SK hynix Announces 2Q25 Financial Results
企業新聞稿 · Korea · 英文 · 24 July 2025 · 閱讀 27 September 2026
https://news.skhynix.com/sk-hynix-announces-2q25-financial-results/
"SK hynix Inc. (or "the company", www.skhynix.com) announced today that it has recorded 22.232 trillion won in revenues, 9.2129 trillion won in operating profit(with an operating margin of 41%), and 6.9962 trillion won in net profit(with a net margin of 31%) in the second quarter."
支持的研判. SK hynix reports 2Q25 revenue of KRW 22.232 trillion and operating profit of KRW 9.2129 trillion, a 41% operating margin.
無法據以確立. Company-wide results, not HBM-specific.
引用於: 第節 第3.2節; 圖 15

S-123 SK hynix Inc. — SK hynix Announces 3Q25 Financial Results
企業新聞稿 · Korea · 英文 · 29 October 2025 · 閱讀 27 September 2026
https://news.skhynix.com/en/sk-hynix-announces-3q25-financial-results/
"SK hynix Inc. (or "the company", www.skhynix.com) announced today that it has recorded 24.4489 trillion won in revenues, 11.3834 trillion won in operating profit (with an operating margin of 47%), and 12.5975 trillion won in net profit (with a net margin of 52%) in the third quarter."
支持的研判. SK hynix reports 3Q25 revenue of KRW 24.4489 trillion and operating profit of KRW 11.3834 trillion, a 47% operating margin.
無法據以確立. Company-wide results, not HBM-specific.
引用於: 圖 15

S-124 SK hynix Inc. — SK hynix Announces FY25 Financial Results
企業新聞稿 · Korea · 英文 · 28 January 2026 · 閱讀 27 September 2026
https://news.skhynix.com/en/sk-hynix-announces-fy25-financial-results/
"As a result, the company achieved record-high quarterly performance across all three indicators, with revenue rising 34% to 32.8267 trillion won, operating profit surging 68% to 19.1696 trillion won, and operating margin reaching 58% quarter-on-quarter."
支持的研判. SK hynix reports 4Q25 revenue of KRW 32.8267 trillion and operating profit of KRW 19.1696 trillion, a 58% operating margin.
無法據以確立. Company-wide results, not HBM-specific.
引用於: 圖 15

S-125 SK hynix Inc. — SK hynix Announces 1Q26 Financial Results
企業新聞稿 · Korea · 英文 · 22 April 2026 · 閱讀 27 September 2026
https://news.skhynix.com/en/q1-2026-business-results/
"Meanwhile, SK hynix emphasized that within the environment where customer demand exceeds supply capacity, securing stable supply capability to meet the structural demand growth of the AI era has emerged as a key competitive advantage."
"SK hynix Inc. (or "the company", www.skhynix.com) announced today that it has recorded 52.5763 trillion won in revenue, 37.6103 trillion won in operating profit (with an operating margin of 72%), and 40.3459 trillion won in net profit (with a net margin of 77%) in the first quarter."
支持的研判. SK hynix reports 1Q26 revenue of KRW 52.5763 trillion, a 72% operating margin, and describes an environment where customer demand exceeds supply capacity.
無法據以確立. Company-wide results; the demand statement is SK hynix's account of its own market, not memory as a whole.
引用於: 圖 15; 第節 第C.2節

S-126 SK hynix Inc. — SK hynix and NVIDIA Announce Multi-year Technology Partnership to Advance Memory for AI Factories
企業新聞稿 · Korea · 英文 · 8 June 2026 · 閱讀 27 September 2026
https://news.skhynix.com/en/multi-year-tech-partnership-with-nvidia/
"The multi-year agreement supports supply to address the extended development cycles of advanced memory."
"SK hynix Inc. (or "the company", www.skhynix.com) and NVIDIA today announced a multi-year technology partnership to advance next-generation memory for the global AI factory buildout and accelerate semiconductor design and manufacturing."
支持的研判. SK hynix and NVIDIA announced in June 2026 a multi-year technology partnership whose agreement 'supports supply' of advanced memory for AI factories.
無法據以確立. States no volumes, duration, value or named products; a technology partnership, not a disclosed supply contract.
引用於: 第節 第3.2節; 表 3, 25

S-127 Samsung Electronics Co., Ltd. — Samsung Electronics Announces Second Quarter 2026 Results
企業新聞稿 · Korea · 英文 · 30 July 2026 · 閱讀 27 September 2026
https://news.samsung.com/global/samsung-electronics-announces-second-quarter-2026-results
"This is projected to keep the market undersupplied, despite partial demand moderation in mobile and PCs."
"In H2 2026, the Memory Business expects robust demand centered on servers stemming from continued AI infrastructure capex and broader adoption of agentic AI."
"Meanwhile, growth in demand for server DRAM, eSSDs, and HBM is expected to accelerate."
支持的研判. Samsung expects accelerating demand growth for server DRAM, eSSDs and HBM in 2H 2026, 'projected to keep the market undersupplied', and says it shipped the industry's first HBM4E samples.
無法據以確立. Company projection for memory overall in 2H 2026; no bit or dollar size of the shortfall, and 'first' is Samsung's claim.
引用於: 第節 第3.2節, 第3.3節, 第5.7節, 第8.4節, 第B.2節; 表 3, 23, 25

S-128 Micron Technology — Micron Technology, Inc. Reports Record Results for the Third Quarter of Fiscal 2026
企業新聞稿 · U.S. · 英文 · 24 June 2026 · 閱讀 26 September 2026
https://investors.micron.com/news/press-release/2026/Micron-Technology-Inc--Reports-Record-Results-for-the-Third-Quarter-of-Fiscal-2026/default.aspx
"We believe our multi-year Strategic Customer Agreements will significantly enhance the durability and predictability of Micron's strong financial performance."
"Development of HBM4E, built on 1-gamma DRAM technology, is well underway, with volume production expected in calendar 2027."
支持的研判. Micron expects HBM4E volume production in calendar 2027, reports HBM4 in high-volume shipments for its lead customer's platform, and believes its multi-year Strategic Customer Agreements will enhance predictability.
無法據以確立. Company expectations; HBM4 shipments are for one lead customer's platform, and no HBM revenue, bit volumes or agreement terms are given.
引用於: 第節 第3.2節, 第6.7節, 第7.5節, 第9.2節, 第A.2節, 第C.7節; 表 3, 13, 19, 25; 展望 1

S-129 Micron Technology — Micron FQ3 2026 Prepared Remarks
企業新聞稿 · U.S. · 英文 · 24 June 2026 · 閱讀 27 September 2026
https://s25.q4cdn.com/621799436/files/doc_financials/2026/q3/Q3-FY26-Prepared-Remarks.pdf
"These SCAs are structured as take-or-pay agreements, with binding commitments to purchase specific volumes over this multi-year term."
"Agentic AI is structurally reshaping data center infrastructure, extending beyond accelerator-only racks to include CPU racks for the agent control plane and program execution, and storage racks for rapidly expanding context store."
"we project to receive cash deposits and related financial commitments of $22 billion"
"The largest agreements generally have a ceiling price for existing products at the current CQ2 market price, and a floor price through the term of the agreement."
支持的研判. Micron projects US$22 billion of cash deposits and related financial commitments under its signed Strategic Customer Agreements, the largest of which carry floor prices through their term.
無法據以確立. Company projection; deposits were expected on the balance sheet largely in FQ4 2026, and terms vary by agreement.
引用於: 第節 第1.3節, 第3.2節, 第5.1節, 第5.3節, 第5.6節, 第5.8節, 第8.4節, 第8.6節, 第C.7節; 表 3, 7, 8, 25; 展望 15 · **研判:** A-13, A-14

S-130 KIOXIA Holdings Corporation — Investor Day (June 2, 2026): Presentation Slides (with Script)
企業新聞稿 · Japan · 英文 · 2 June 2026 · 閱讀 27 September 2026
https://www.kioxia-holdings.com/en-jp/ir/library/event.html
直接文件: https://www.kioxia-holdings.com/content/dam/kioxia-hd/en-jp/ir/library/event/asset/Kioxia_Investor_Day_2026_en_script.pdf
"On a portfolio basis, we plan to increase the share of our overall revenues delivered by data center products to over 60% by FY2028."
"We are moving from one-year LTAs to multi-year LTAs to enable greater revenue visibility and stability."
支持的研判. Kioxia's CFO said in June 2026 it plans to raise data-centre products above 60% of revenue by FY2028 and is moving from one-year to multi-year LTAs.
無法據以確立. Management plans; does not report signed multi-year agreements, customers or volumes.
引用於: 第節 第3.2節, 第3.3節, 第5.7節, 第9.3節; 圖 25; 表 25

S-131 KIOXIA Holdings Corporation — Investor Day (June 2, 2026): Presentation Slides
企業新聞稿 · Japan · 英文 · 2 June 2026 · 閱讀 27 September 2026
https://www.kioxia-holdings.com/en-jp/ir/library/event.html
直接文件: https://www.kioxia-holdings.com/content/dam/kioxia-hd/en-jp/ir/library/event/asset/Kioxia-Investor-Day-2026-en.pdf
"Portfolio Ratio of Data Center Revenue Over 60%(by FY28)"
支持的研判. Kioxia's June 2026 Investor Day slides set a data-centre revenue share of over 60% by FY2028.
無法據以確立. Management target; FY2028 ends March 2029; not a revenue amount or forecast.
引用於: 第節 第3.2節; 表 19

S-132 TrendForce — Long-Term Agreements Cap Price Increases; Server DRAM Contract Prices Expected to Rise 13-18% QoQ in 3Q26, Says TrendForce
市場研究 · Taiwan · 英文 · 9 July 2026 · 閱讀 26 September 2026
https://www.trendforce.com/presscenter/news/20260709-13140.html
"several U.S.-based CSPs have entered into multi-year long-term agreements (LTAs), which restrict suppliers from raising prices for these clients"
支持的研判. TrendForce reports several U.S.-based CSPs hold multi-year DRAM LTAs restricting suppliers from raising their prices, leaving increases to fall on buyers and volumes outside LTAs.
無法據以確立. Market-research report; it does not name the CSPs or disclose contract terms, volumes or durations.
引用於: 第節 第3.2節, 第5.6節, 第C.7節

S-133 Investing.com via Yahoo Finance (citing Bloomberg) — Nvidia certifies Samsung, SK Hynix and Micron for Vera Rubin HBM4 supply
媒體 · U.S. · 英文 · 5 June 2026 · 閱讀 27 September 2026
https://finance.yahoo.com/sectors/technology/articles/nvidia-certifies-samsung-sk-hynix-133001560.html
"Samsung Electronics (LON:0593xq), SK Hynix Inc (KS:000660), and Micron Technology Inc (NASDAQ:MU) have all passed certification to supply HBM4 high-bandwidth memory for the company's next-generation Vera Rubin AI accelerator platform"
"The announcement, reported by Bloomberg, ends months of supply-chain speculation about which chipmakers would qualify for Vera Rubin"
支持的研判. Investing.com, citing Bloomberg, reported on 5 June 2026 that Samsung, SK Hynix and Micron had all passed certification to supply HBM4 for NVIDIA's Vera Rubin platform.
無法據以確立. Secondary report of Bloomberg; certification is not a supply contract, and it gives no volumes or allocation among the three suppliers.
引用於: 表 3

S-134 Samsung Electronics Co., Ltd. — Future Business 、Management Plan
申報文件 · Korea · 英文 · 29 June 2026 · 閱讀 27 September 2026
https://www.samsung.com/global/ir/reports-disclosures/public-disclosure-view.84685/
"Approximately KRW 2,450 trillion (including approximately KRW 2,100 trillion for the semiconductor business)"
"The above mid- to long-term investment plan is provided as a guideline for reference based on the current market conditions."
支持的研判. Samsung's disclosed mid- to long-term plan sets investment of about KRW 2,450 trillion over 2026–2040, about KRW 2,100 trillion for semiconductors, as a reference guideline.
無法據以確立. A reference guideline based on current market conditions, not a commitment, annual capex guidance or capacity plan.
引用於: 表 3, 19; 第節 第3.2節, 第9.4節, 第9.5節

S-135 TrendForce — DRAM Supply to Remain Tight in 2027, Prompting NVIDIA to Lower HBM Configurations for Rubin Ultra, Says TrendForce
市場研究 · Taiwan · 英文 · 4 August 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260804-13166.html
"From a supply-demand perspective, HBM bit shipments are projected to grow by 50–60% YoY in 2027, which will still be insufficient to keep pace with demand growth."
"DRAM supply will remain tight through 2027"
"Since the third quarter of 2026, NVIDIA has begun expanding its evaluation of the Rubin Ultra's HBM configuration beyond its original 12-Hi HBM4e design to include 8-Hi HBM4e, 12-Hi HBM4, and 8-Hi HBM4 alternatives."
支持的研判. TrendForce expects DRAM supply to remain tight through 2027 and reports NVIDIA evaluating smaller HBM configurations for Rubin Ultra than its original 12-Hi HBM4e design.
無法據以確立. Market-research view; NVIDIA's final Rubin Ultra HBM specification is undetermined, and tightness is not quantified as a supply-demand gap.
引用於: 第節 第3.2節, 第3.3節, 第8.3節, 第8.4節, 第8.6節, 第9.1節; 圖 29; 展望 1, 7 · **研判:** A-13

S-136 Amir Gholami, Zhewei Yao, Sehoon Kim, Coleman Hooper, Michael W. Mahoney, Kurt Keutzer — AI and Memory Wall
研究 · U.S. · 英文 · 21 March 2024 · 閱讀 27 September 2026
https://arxiv.org/abs/2403.14123
"peak server hardware FLOPS has been scaling at 3.0x/2yrs, outpacing the growth of DRAM and interconnect bandwidth, which have only scaled at 1.6 and 1.4 times every 2 years"
"This disparity has made memory, rather than compute, the primary bottleneck in AI applications, particularly in serving."
支持的研判. Gholami et al. report peak server FLOPS scaling 3.0x every two years against 1.6x for DRAM and 1.4x for interconnect bandwidth, making memory the primary bottleneck, especially in serving.
無法據以確立. Long-run trend rates from the authors' selected hardware data; not a measurement of any current product.
引用於: 第節 第3.3節, 第C.7節

S-137 Pratyush Patel, Esha Choukse, Chaojie Zhang, Aashaka Shah, Inigo Goiri, Saeed Maleki, Ricardo Bianchini — Splitwise: Efficient generative LLM inference using phase splitting
研究 · U.S. · 英文 · 30 November 2023 · 閱讀 27 September 2026
https://arxiv.org/abs/2311.18677
"Based on our extensive characterization, we find that there are two main phases during an LLM inference request: a compute-intensive prompt computation, and a memory-intensive token generation, each with distinct latency, throughput, memory, and power characteristics."
"With Splitwise, we propose splitting the two phases of a LLM inference request on to separate machines."
支持的研判. Splitwise characterises LLM inference as a compute-intensive prompt phase and a memory-intensive token-generation phase, and runs them on separately provisioned machines.
無法據以確立. Author-reported characterisation on their hardware and traces; results depend on cluster design and cost assumptions.
引用於: 第節 第3.3節, 第C.7節

S-138 arXiv — FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness
研究 · U.S. · 英文 · 27 May 2022 · 閱讀 26 September 2026
https://arxiv.org/abs/2205.14135
"uses tiling to reduce the number of memory reads/writes between GPU high bandwidth memory (HBM) and GPU on-chip SRAM"
支持的研判. FlashAttention is an IO-aware exact attention algorithm that uses tiling to reduce memory reads and writes between GPU HBM and on-chip SRAM.
無法據以確立. Speedups are measured on the authors' 2022 benchmarks and GPUs; it reduces memory traffic, not model or KV-cache size.
引用於: 第節 第3.3節

S-139 JEDEC — JEDEC and Industry Leaders Collaborate to Release JESD270-4 HBM4 Standard
標準組織 · U.S. · 英文 · 16 April 2025 · 閱讀 26 September 2026 · **已被取代**
https://www.jedec.org/news/pressreleases/jedec%C2%AE-and-industry-leaders-collaborate-release-jesd270-4-hbm4-standard-advancing
"With transfer speeds up to 8 Gb/s across a 2048-bit interface, HBM4 boosts total bandwidth up to 2 TB/s."
支持的研判. JEDEC's 16 April 2025 release announces JESD270-4 HBM4, with transfer speeds up to 8 Gb/s across a 2048-bit interface and total bandwidth up to 2 TB/s.
無法據以確立. An April 2025 release: it cannot show whether JEDEC later announced an HBM4E standard, nor supplier speeds or shipments.
已被取代. The April 2025 release announced JESD270-4; JEDEC's revision JESD270-4A (December 2025) is the current HBM4 document [S-058].
引用於: 第節 第3.3節, 第6.3節, 第A.2節, 第C.2節, 第C.7節; 表 11, 22

S-140 Samsung Semiconductor — Samsung HBM (High Bandwidth Memory) product page
企業新聞稿 · Korea · 英文 · 2026 · 閱讀 27 September 2026
https://semiconductor.samsung.com/dram/hbm
“Samsung HBM4E delivers greater capacity and performance for next-generation AI systems. Powered by the industry's first 1c DRAM, a 4 nm foundry-based logic base die, optimized design and process technologies, and advanced packaging, it delivers up to 64 GB of capacity and 4 TB/s of bandwidth in a single 16-high stack, with I/O speeds of up to 16 Gbps per pin - more than 20% faster than HBM4.”
“Specific timelines for HBM customer qualification have not been publicly disclosed.”
支持的研判. Samsung lists HBM4E at up to 64 GB and 4 TB/s per 16-high stack (16 Gbps per pin), HBM4 on a 4 nm base die, and undisclosed qualification timelines.
無法據以確立. Vendor product-page maxima; not shipping volumes, customer qualification or delivered bandwidth in a system.
引用於: 第節 第3.3節, 第3.4節, 第C.7節

S-141 JEDEC — New JEDEC SPHBM4 Standard Enables HBM4-Class Bandwidth on Organic Substrates
標準組織 · U.S. · 英文 · 13 July 2026 · 閱讀 26 September 2026
https://www.jedec.org/news/pressreleases/new-jedec%C2%AE-sphbm4-standard-enables-hbm4-class-bandwidth-organic-substrates
“SPHBM4 uses the same DRAM dies as HBM4”
支持的研判. JEDEC's 13 July 2026 release says the new SPHBM4 standard uses the same DRAM dies as HBM4 and enables HBM4-class bandwidth on organic substrates.
無法據以確立. Standards-body announcement; the base-die change and the JESD330-4 number are in the release, not this excerpt. Shows no products or capacity.
引用於: 第節 第3.3節, 第3.4節, 第6.7節; 表 11, 13, 22

S-142 Llama Team, Meta AI — The Llama 3 Herd of Models
研究 · U.S. · 英文 · 31 July 2024 · 閱讀 27 September 2026
https://arxiv.org/abs/2407.21783
“We use grouped query attention (GQA; Ainslie et al. (2023)) with 8 key-value heads to improve inference speed and to reduce the size of key-value caches during decoding.”
支持的研判. The Llama 3 paper uses GQA with 8 key-value heads; its Table 3 gives the 70B model 80 layers, model dimension 8,192 and 64 attention heads.
無法據以確立. Head dimension 128 is derived (8,192 ÷ 64), not stated; the 42.9 GB cache figure is APEX arithmetic at 16-bit precision.
引用於: 第節 第3.3節, 第C.7節

S-143 MiniMax — MiniMax-M1: Scaling Test-Time Compute Efficiently with Lightning Attention
研究 · China · 英文 · 16 June 2025 · 閱讀 27 September 2026
https://arxiv.org/abs/2506.13585
“The M1 model natively supports a context length of 1 million tokens, 8x the context size of DeepSeek R1.”
“The model is developed based on our previous MiniMax-Text-01 model, which contains a total of 456 billion parameters with 45.9 billion parameters activated per token.”
支持的研判. MiniMax-M1, built on MiniMax-Text-01 (456 billion parameters, 45.9 billion activated per token), natively supports a context length of 1 million tokens.
無法據以確立. A maximum supported context, not typical usage; memory cost at that length depends on its lightning-attention design.
引用於: 第節 第3.3節

S-144 arXiv — InferCept: Efficient Intercept Support for Augmented Large Language Model Inference
研究 · U.S. · 英文 · 2 February 2024 · 閱讀 26 September 2026
https://arxiv.org/abs/2402.01869
“unnecessary recomputation of already computed contexts, which accounts for 37-40% of total model forwarding time”
支持的研判. InferCept finds that recomputing already-computed context after LLM interceptions such as tool calls accounts for 37–40% of total model forwarding time in current systems.
無法據以確立. Measured on the authors' workloads and systems; the paper's remedy mixes preserving, swapping and discarding cache, not only keeping it.
引用於: 第節 第3.3節

S-145 arXiv (Qwen Team, Alibaba) — Qwen3 Technical Report
研究 · China · 英文 · 14 May 2025 · 閱讀 27 September 2026
https://arxiv.org/abs/2505.09388
“The Qwen3 series includes models of both dense and Mixture-of-Expert (MoE) architectures, with parameter scales ranging from 0.6 to 235 billion.”
“Meanwhile, Qwen3 introduces a thinking budget mechanism, allowing users to allocate computational resources adaptively during inference, thereby balancing latency and performance based on task complexity.”
支持的研判. The Qwen3 technical report describes dense and MoE models from 0.6 to 235 billion parameters and a thinking budget mechanism for allocating inference compute adaptively.
無法據以確立. An author description of the mechanism; tokens added per request depend on the budget set and the task.
引用於: 第節 第3.3節

S-146 DeepSeek-AI — DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model
研究 · China · 英文 · 7 May 2024 · 閱讀 27 September 2026
https://arxiv.org/abs/2405.04434
“MLA guarantees efficient inference through significantly compressing the Key-Value (KV) cache into a latent vector”
“reduces the KV cache by 93.3%”
支持的研判. DeepSeek-V2 reports that Multi-head Latent Attention compresses the KV cache into a latent vector and that, compared with DeepSeek 67B, it reduces the KV cache by 93.3%.
無法據以確立. Author-reported comparison with its own predecessor model; the reduction reflects the whole architecture change, not MLA in isolation.
引用於: 第節 第3.3節, 第C.7節

S-147 arXiv (DeepSeek-AI) — DeepSeek-V3 Technical Report
研究 · China · 英文 · 27 December 2024 · 閱讀 27 September 2026
https://arxiv.org/abs/2412.19437
“We present DeepSeek-V3, a strong Mixture-of-Experts (MoE) language model with 671B total parameters with 37B activated for each token.”
支持的研判. DeepSeek-V3 is a Mixture-of-Experts model with 671 billion total parameters, 37 billion activated per token, using Multi-head Latent Attention.
無法據以確立. Author-reported architecture; activating fewer parameters per token does not remove the need to hold all 671B parameters.
引用於: 第節 第3.3節

S-148 Kimi Team — Kimi K2: Open Agentic Intelligence
研究 · China · 英文 · 28 July 2025 · 閱讀 27 September 2026
https://arxiv.org/abs/2507.20534
“We introduce Kimi K2, a Mixture-of-Experts (MoE) large language model with 32 billion activated parameters and 1 trillion total parameters.”
“Based on MuonClip, K2 was pre-trained on 15.5 trillion tokens with zero loss spike.”
支持的研判. The Kimi K2 report describes a Mixture-of-Experts model with 1 trillion total parameters and 32 billion activated parameters, pre-trained on 15.5 trillion tokens.
無法據以確立. Author-reported architecture of one open model; it says nothing about memory needed to serve it at scale.
引用於: 第節 第3.3節

S-149 arXiv — Efficient Memory Management for Large Language Model Serving with PagedAttention
研究 · U.S. · 英文 · 12 September 2023 · 閱讀 27 September 2026
https://arxiv.org/abs/2309.06180
“To address this problem, we propose PagedAttention, an attention algorithm inspired by the classical virtual memory and paging techniques in operating systems.”
“On top of it, we build vLLM, an LLM serving system that achieves (1) near-zero waste in KV cache memory and (2) flexible sharing of KV cache within and across requests to further reduce memory usage.”
支持的研判. The vLLM paper introduces PagedAttention, which applies operating-system paging to the KV cache for near-zero wasted cache memory and cache sharing within and across requests.
無法據以確立. Gains are the authors' own measurements on popular open LLMs; paging cuts wasted memory, not bytes per token.
引用於: 第節 第3.3節

S-150 arXiv — KIVI: A Tuning-Free Asymmetric 2bit Quantization for KV Cache
研究 · U.S. · 英文 · 5 February 2024 · 閱讀 27 September 2026
https://arxiv.org/abs/2402.02750
“A straightforward and effective solution to reduce KV cache size is quantization, which decreases the total bytes taken by KV cache.”
“From this analysis, we developed a tuning-free 2bit KV cache quantization algorithm named KIVI.”
支持的研判. KIVI proposes tuning-free asymmetric 2-bit KV-cache quantization, keys per channel and values per token, reporting 2.6x less peak memory and up to 4x larger batches.
無法據以確立. Author-reported results on Llama, Falcon and Mistral models; quality and speed gains depend on model and workload.
引用於: 第節 第3.3節

S-151 arXiv — H2O: Heavy-Hitter Oracle for Efficient Generative Inference of Large Language Models
研究 · U.S. · 英文 · 24 June 2023 · 閱讀 27 September 2026
https://arxiv.org/abs/2306.14048
“Our approach is based on the noteworthy observation that a small portion of tokens contributes most of the value when computing attention scores.”
“We formulate the KV cache eviction as a dynamic submodular problem and prove (under mild assumptions) a theoretical guarantee for our novel eviction algorithm which could help guide future work.”
支持的研判. H2O keeps a small set of heavy-hitter tokens that contribute most attention value plus recent tokens, evicting the rest of the KV cache, and reports up to 29x throughput gains.
無法據以確立. Author-reported results on OPT, LLaMA and GPT-NeoX models against specific baselines; accuracy effects vary by task.
引用於: 第節 第3.3節

S-152 Bita Darvish Rouhani et al. (Microsoft, AMD, Intel, Meta, NVIDIA, Qualcomm), arXiv — Microscaling Data Formats for Deep Learning
研究 · U.S. · 英文 · October 2023 · 閱讀 27 September 2026
https://arxiv.org/abs/2310.10537
“This paper evaluates Microscaling (MX) data formats that combine a per-block scaling factor with narrow floating-point and integer types for individual elements.”
“We also show the first instance of training generative language models at sub-8-bit weights, activations, and gradients with minimal accuracy loss and no modifications to the training recipe.”
支持的研判. The multi-company Microscaling paper reports that MX formats allow training generative language models with sub-8-bit weights, activations and gradients at minimal accuracy loss.
無法據以確立. Author-reported experiments on selected benchmarks and models; it does not show production training at scale.
引用於: 第節 第3.3節, 第C.7節

S-153 NVIDIA (Developer Blog) — Introducing NVFP4 for Efficient and Accurate Low-Precision Inference
企業新聞稿 · U.S. · 英文 · 24 June 2025 · 閱讀 26 September 2026
https://developer.nvidia.com/blog/introducing-nvfp4-for-efficient-and-accurate-low-precision-inference
“1% or less accuracy degradation versus FP8 on DeepSeek-R1-0528 benchmarks”
支持的研判. NVIDIA reports NVFP4 shows '1% or less accuracy degradation versus FP8 on DeepSeek-R1-0528 benchmarks'.
無法據以確立. Vendor benchmark result for named models; not a general accuracy guarantee across models and tasks.
引用於: 第節 第3.3節

S-154 Yinmin Zhong, Shengyu Liu, Junda Chen, Jianbo Hu, Yibo Zhu, Xuanzhe Liu, Xin Jin, Hao Zhang — DistServe: Disaggregating Prefill and Decoding for Goodput-optimized Large Language Model Serving
研究 · China · 英文 · 18 January 2024 · 閱讀 27 September 2026
https://arxiv.org/abs/2401.09670
“DistServe assigns prefill and decoding computation to different GPUs, hence eliminating prefill-decoding interferences.”
“Our evaluations show that on various popular LLMs, applications, and latency requirements, DistServe can serve 7.4x more requests or 12.6x tighter SLO, compared to state-of-the-art systems, while staying within latency constraints for > 90% of requests.”
支持的研判. DistServe runs prefill and decoding on different GPUs and reports serving 7.4x more requests, or 12.6x tighter SLOs, than prior systems while meeting latency constraints for over 90% of requests.
無法據以確立. Author-reported results against selected baselines; gains vary by model, application and latency target, and both phases still use GPUs.
引用於: 第節 第3.3節

S-155 Ruoyu Qin, Zheming Li, Weiran He, Mingxing Zhang, Yongwei Wu, Weimin Zheng, Xinran Xu — Mooncake: A KVCache-centric Disaggregated Architecture for LLM Serving
研究 · China · 英文 · 24 June 2024 · 閱讀 27 September 2026
https://arxiv.org/abs/2407.00079
“It also leverages the underutilized CPU, DRAM, and SSD resources of the GPU cluster to implement a disaggregated cache of KVCache.”
“Under real workloads, Mooncake's innovative architecture enables Kimi to handle 75% more requests.”
支持的研判. Mooncake, Kimi's serving platform, disaggregates prefill and decoding and uses the GPU cluster's underused CPU, DRAM and SSD as a KV cache, letting Kimi handle 75% more requests.
無法據以確立. Author-reported results on Moonshot's own workloads; the 525% figure is simulated, and decoding still runs on GPUs.
引用於: 第節 第3.3節, 第C.7節

S-156 NVIDIA — NVIDIA Dynamo Open-Source Library Accelerates and Scales AI Reasoning Models
企業新聞稿 · U.S. · 英文 · 18 March 2025 · 閱讀 26 September 2026
https://nvidianews.nvidia.com/news/nvidia-dynamo-open-source-library-accelerates-and-scales-ai-reasoning-models
“intelligently offloads and reloads inference data to and from lower-cost memory and storage devices”
支持的研判. NVIDIA says its open-source Dynamo inference framework offloads and reloads inference data to and from lower-cost memory and storage devices.
無法據以確立. Software capability described by the vendor; it does not quantify HBM savings or show HBM leaving the decode path.
引用於: 第節 第3.3節, 第C.7節

S-157 SK hynix Newsroom — SK hynix Honored with Gold Tower Order and Presidential Citation at the 61st Invention Day Ceremony
企業新聞稿 · Korea · 英文 · 20 May 2026 · 閱讀 26 September 2026
https://news.skhynix.com/en/invention-day-2026
“AIN-B is being brought to market under the product name HBF through a strategic partnership with SanDisk, with both companies jointly working on technical standardization.”
支持的研判. SK hynix says its AIN-B product is being brought to market as HBF through a strategic partnership with SanDisk, with joint work on technical standardization.
無法據以確立. Company statement of a partnership; no product date, specification or standards-body adoption of HBF is established.
引用於: 第節 第3.3節

S-158 CXL Consortium — CXL Consortium Releases the Compute Express Link 4.0 Specification Increasing Speed and Bandwidth
標準組織 · U.S. · 英文 · 18 November 2025 · 閱讀 26 September 2026
https://computeexpresslink.org/wp-content/uploads/2025/11/CXL_4.0-Specification-Release_FINAL_Website-Copy.pdf
“Doubles the bandwidth to 128GTs with zero added latency”
支持的研判. The CXL Consortium's 18 November 2025 release says CXL 4.0 doubles the bandwidth to 128 GT/s with zero added latency.
無法據以確立. The consortium's own announcement; it shows a specification release, not products or pooled-memory deployments. 'Bundled ports' rests on the release's feature list, not this excerpt.
引用於: 第節 第3.3節, 第C.7節; 表 22

S-159 Huaicheng Li, Daniel S. Berger, Stanko Novakovic, Lisa Hsu, Dan Ernst, Pantea Zardoshti, Monish Shah, Samir Rajadnya, Scott Lee, Ishwar Agarwal, Mark D. Hill, Marcus Fontoura, Ricardo Bianchini — Pond: CXL-Based Memory Pooling Systems for Cloud Platforms
研究 · U.S. · 英文 · 1 March 2022 · 閱讀 27 September 2026
https://arxiv.org/abs/2203.00241
“First, our analysis of cloud production traces shows that pooling across 8-16 sockets is enough to achieve most of the benefits.”
“Our evaluation with 158 workloads shows that Pond reduces DRAM costs by 7% with performance within 1-5% of same-NUMA-node VM allocations.”
支持的研判. Microsoft's Pond CXL memory-pooling study finds pooling across 8–16 sockets achieves most benefits and cuts DRAM cost 7% with performance within 1–5% of same-NUMA-node allocation.
無法據以確立. Cloud VM workloads studied in 2022, not AI accelerators; results depend on the authors' prediction models and CXL latency assumptions.
引用於: 第節 第3.3節, 第6.7節

S-160 Open Compute Project Foundation — Server/CMS - OpenCompute wiki
標準組織 · Global · 英文 · 22 August 2026 · 閱讀 26 September 2026
https://www.opencompute.org/wiki/Server/CMS
“Meta/Google Contrib- Draft under review - Hyperscale CXL Tiered Memory Expander for OCP - Base Specification”
“The OCP Composable Memory System is a sub-project within the Server project.”
支持的研判. OCP's CMS wiki lists a Meta/Google ' Hyperscale CXL Tiered Memory Expander for OCP - Base Specification' as a draft under review.
無法據以確立. No version or date is shown; a draft under review is not an accepted specification or a product.
引用於: 第節 第3.3節; 表 21

S-161 Open Compute Project Foundation — CMS: Memory Fabric Orchestration Architecture (White Paper v1.0)
標準組織 · Global · 英文 · 19 December 2025 · 閱讀 26 September 2026
https://www.opencompute.org/documents/ocp-cms-memory-fabric-orchestration-architecture-white-paper-12192025-pdf
“they do not require the hardware-based cache-coherence or the low-latency features that CXL specification offers”
支持的研判. OCP's CMS white paper v1.0 (19 December 2025) says AI scale-up workloads 'do not require the hardware-based cache-coherence or the low-latency features that CXL specification offers'.
無法據以確立. A white paper by eight contributors, not an OCP specification; it does not measure CXL adoption or rule out coherent memory uses.
引用於: 第節 第3.3節, 第C.7節; 表 21

S-162 TrendForce — [News] TSMC Reportedly Targets 22% 2nm, 16%+ 3nm Capacity Boost by Mid-2027; CoWoS to Double by 2028
媒體 · Taiwan · 英文 · 14 September 2026 · 閱讀 27 September 2026
https://www.trendforce.com/news/2026/09/14/news-tsmc-reportedly-targets-22-2nm-16-3nm-capacity-boost-by-mid-2027-cowos-to-double-by-2028
“Wccftech, citing the Economic Daily News, reports that CoWoS capacity could double from roughly 130,000 wafers per month by end-2026 to 260,000 by end-2028.”
“The ramp is expected to center on AP7 in Taiwan and the Arizona campus, where local packaging would reduce the need to ship advanced chips back to Taiwan, the report adds.”
支持的研判. TrendForce News, relaying Wccftech and Economic Daily News, reports TSMC CoWoS capacity could double from about 130,000 wafers monthly at end-2026 to 260,000 by end-2028, centred on AP7 and Arizona.
無法據以確立. A press-reported target ('reportedly', ' could'), not a TSMC disclosure or TrendForce estimate; wafer capacity is not package or accelerator output.
引用於: 第節 第3.4節, 第7.5節

S-163 U.S. Department of Commerce — Trump Administration Secures an Additional $100 Billion U.S. Semiconductor Manufacturing Investment for a Total of $265 Billion from TSMC
政府 · U.S. · 英文 · 16 July 2026 · 閱讀 27 September 2026
https://www.commerce.gov/news/press-releases/2026/07/trump-administration-secures-additional-100-billion-us-semiconductor
“TSMC has just announced an incremental $100 billion investment for a total of $265 billion in the U.S.”
“This additional $100 billion investment marks a significant expansion of TSMC's commitment to build multiple fabrication plants and advanced packaging facilities in the United States.”
支持的研判. Commerce says TSMC's additional US$100 billion, announced in July 2026, brings its U.S. investment to US$265 billion, covering multiple fabs and advanced-packaging facilities.
無法據以確立. A government statement of planned investment, not spending or output; it does not show Taiwan capacity being reduced or relocated.
引用於: 第節 第3.4節, 第5.7節, 第7.5節, 第9.4節; 表 4; 圖 25; 展望 12 · **研判:** A-09

S-164 Arm Newsroom — Arm Sets the Standard for Open, Converged AI Data Centers
企業新聞稿 · U.K. · 英文 · 14 October 2025 · 閱讀 27 September 2026
https://newsroom.arm.com/news/arm-sets-the-standard-for-open-converged-ai-data-centers
“Today we're taking a step toward deeper industry collaboration around chiplets by contributing the Foundation Chiplet System Architecture (FCSA) specification to the Open Compute Project (OCP).”
支持的研判. Arm announced on 14 October 2025 that it is contributing its Foundation Chiplet System Architecture (FCSA) specification to OCP, extending the CSA it introduced in February 2024.
無法據以確立. Arm's contribution announcement; it does not establish the FCSA release, adoption or any product built to it.
引用於: 第節 第3.4節

S-165 Open Compute Project Foundation (contributed by Arm) — Foundation Chiplet System Architecture 1.0.0
標準組織 · Global · 英文 · 12 February 2026 · 閱讀 26 September 2026
https://www.opencompute.org/documents/fcsa-1-0-0-pdf
“This Foundation Chiplet System Architecture document specifies a hardware system architecture for a system that is distributed over multiple chiplets.”
“otherwise the FCSA maps the interface component as TO BE DEFINED”
支持的研判. FCSA 1.0.0, contributed by Arm, specifies a multi-chiplet system architecture, using protocols and transports only where industry standards exist and otherwise mapping interfaces as 'TO BE DEFINED'.
無法據以確立. The PDF is dated 'February 2026'; the 12 February date is from OCP's FCSA wiki. It shows no conformant products or TBD count.
引用於: 第節 第3.4節, 第6.3節, 第6.7節, 第C.7節; 表 11, 13, 21; 圖 26

S-166 Open Compute Project Foundation — Building the Foundation for an Open Chiplet Future: Foundation Chiplet System Architecture (FCSA)
標準組織 · Global · 英文 · 8 April 2026 · 閱讀 26 September 2026
https://www.opencompute.org/blog/building-the-foundation-for-an-open-chiplet-future-foundation-chiplet-system-architecture-fcsa
“standard interconnects (e.g., UCIe, AMBA CHI-C2C, CXL, PCIe)”
支持的研判. OCP's April 2026 blog says FCSA maps chiplet interfaces to standard interconnects such as UCIe, AMBA CHI-C2C, CXL and PCIe.
無法據以確立. A promotional blog, not the specification; it does not establish conformance, adoption or which mappings remain 'TO BE DEFINED'.
引用於: 第節 第3.4節, 第A.2節

S-167 UCIe Consortium — UCIe Consortium Introduces 3.0 Specification With 64 GT/s Performance and Enhanced Manageability
標準組織 · U.S. · 英文 · 5 August 2025 · 閱讀 26 September 2026
https://www.uciexpress.org/_files/ugd/8dc731_ae67289d0ec646cdba5c1aee245538b3.pdf
“Support for 48 GT/s and 64 GT/s data rates, doubling the bandwidth of UCIe 2.0 (32 GT/s)”
支持的研判. The UCIe Consortium's 5 August 2025 release introduces UCIe 3.0 with support for 48 GT/s and 64 GT/s data rates, doubling UCIe 2.0's 32 GT/s bandwidth.
無法據以確立. The consortium's announcement of a specification; it shows no products, and implementation terms are set in member agreements.
引用於: 第節 第3.4節, 第6.3節, 第C.7節; 表 11, 13, 22

S-168 Ultra Accelerator Link Consortium, Inc. — UALink Specifications page; UALink white paper
標準組織 · U.S. · 英文 · 7 April 2026 · 閱讀 26 September 2026
https://ualinkconsortium.org/specification
“Fully compliant with the UCIe® 3.0 Specification for simplified integration into existing chiplet ecosystems”
“The fabric makes use of a Switch Abstraction Interface (SAI), which will enable operators to monitor and configure UALink networks”
支持的研判. UALink's specifications material describes its chiplet specification as fully compliant with UCIe 3.0 and says the fabric uses the Switch Abstraction Interface (SAI) for monitoring and configuration.
無法據以確立. Consortium self-description (SAI sentence from the January 2026 white paper); no products shown. The page lists Chiplet 1.01, not the 7 April 1.0.
引用於: 第節 第3.4節, 第4.6節, 第6.3節, 第C.7節; 表 11, 22

S-169 友達光電 AUO Corporation — 瞄準AI世代核心 友達SEMICON Taiwan 秀關鍵實力
企業新聞稿 · Taiwan · 繁體中文 · 31 August 2026 · 閱讀 27 September 2026
https://www.auo.com/zh-TW/News_Archive/detail/News_Archive_Product_20260831
“並規劃運用達興材料之感光性介電絕緣材於RDL 封裝中”
APEX譯文: and plans to use Daxin Materials' photosensitive dielectric insulating material in the RDL packaging
“並結合生態圈夥伴康寧的光纖解決方案，打造系統級Micro LED CPO光學模組，適合應用於CPO共封裝光學及AOC主動式光纜等高速互連情境。”
APEX譯文: and, combining the optical-fibre solutions of ecosystem partner Corning, builds a system-level Micro LED CPO optical module suited to high-speed interconnect scenarios such as CPO co-packaged optics and AOC active optical cables.
支持的研判. AUO's Chinese release says it builds, with Corning fibre, a system-level Micro LED CPO module for high-speed interconnects and plans to use Daxin's photosensitive dielectric in RDL packaging.
無法據以確立. Pre-event company release (31 August 2026); plans and interim results only, with no production, customer or performance data.
引用於: 第節 第3.5節, 第4.7節, 第5.7節; 表 6, 30

S-170 BE Semiconductor Industries N.V. (Besi) — BE Semiconductor Industries N.V. Announces Q2-26 and H1-26 Results
企業新聞稿 · Netherlands · 英文 · 23 July 2026 · 閱讀 27 September 2026
https://www.besi.com/investor-relations/press-releases/details/be-semiconductor-industries-nv-announces-q2-26-and-h1-26-results/
直接文件: https://www.globenewswire.com/news-release/2026/07/23/3331898/0/en/be-semiconductor-industries-n-v-announces-q2-26-and-h1-26-results.html
“Hybrid bonding customer adoption increased from 15 at year end 2025 to 21 at the end of Q2-26, use cases increased for logic, memory, co-packaged optics and consumer applications and orders increased materially versus H1-25 as significant new capacity was added.”
“Overall, we estimate that system orders for AI applications rose to approximately 60% in H1-26 versus approximately 50% in H1-25.”
支持的研判. Besi reports hybrid-bonding customers up from 15 to 21 in H1 2026, AI about 60% of H1 system orders, Q2 orders €292.9 million (+8.6% on Q1), gross margin 65.7%.
無法據以確立. Company results and estimates; the AI share is Besi's estimate, and the 8.6% order growth is sequential, not year on year.
引用於: 第節 第3.5節, 第5.1節, 第5.7節, 第C.7節; 表 7, 10, 26; 圖 25

S-171 Tokyo Electron Limited — Q1 FY2027 Financial Announcement
企業新聞稿 · Japan · 英文 · 30 July 2026 · 閱讀 27 September 2026
https://www.tel.com/ir/library/report/index.html
直接文件: https://www.tel.com/ir/library/report/i242su0000000goz-att/fy27q1transcript-e.pdf
“For advanced packaging, prober sales show an outstanding growth and the full-year sales are likely to go significantly beyond 100 billion yen.”
“For HBM, the number of stacked DRAMs will increase from 12 to 16.”
支持的研判. Tokyo Electron says HBM stacks will rise from 12 to 16 DRAM dies and advanced-packaging prober sales show outstanding growth, with full-year sales likely significantly above JPY 100 billion.
無法據以確立. Management remarks; 'likely' is an outlook, the 100-billion figure's scope (probers or all packaging) is not explicit, and the 16-high timing is unstated.
引用於: 第節 第3.5節; 表 26

S-172 Tokyo Electron Limited — Q1 FY2027 (April - June 2026) Financial Announcement
企業新聞稿 · Japan · 英文 · 30 July 2026 · 閱讀 27 September 2026
https://www.tel.com/ir/library/report/index.html
直接文件: https://www.tel.com/ir/library/report/i242su0000000goz-att/fy27q1presentations-e.pdf
“WFE for AI data centers are growing to exceed 50% of total.”
“Bonding-related equipment: 500B yen”
支持的研判. Tokyo Electron: AI data-centre WFE growing to exceed 50% of total; FY2027 advanced-packaging sales expected well above JPY 100bn; JPY 500bn cumulative bonding sales to 2030 listed as an opportunity.
無法據以確立. Company slides; the 2030 figures are presented as business opportunities, not targets or guidance; the WFE share is TEL's market view.
引用於: 第節 第3.5節, 第5.7節, 第B.2節; 圖 25; 表 9, 26

S-173 Tokyo Electron Limited — Summary of Consolidated Financial Results for the First Quarter Ended June 30, 2026 (Japanese GAAP)
申報文件 · Japan · 英文 · 30 July 2026 · 閱讀 27 September 2026
https://www.tel.com/ir/library/report/index.html
直接文件: https://www.tel.com/ir/library/report/i242su0000000goz-att/fy27q1tanshin-e.pdf
“The full-year consolidated financial forecasts are scheduled to be disclosed at the time of the financial results announcement for the second quarter ending September 30, 2026.”
支持的研判. Tokyo Electron says its full-year consolidated forecasts are scheduled for disclosure with its results for the second quarter ending 30 September 2026.
無法據以確立. Establishes only the disclosure schedule; it gives no forecast figures, and the date remains subject to change.
引用於: 第節 第3.5節

S-174 NVIDIA Technical Blog — Inside NVIDIA Blackwell Ultra: The Chip Powering the AI Factory Era
企業新聞稿 · U.S. · 英文 · 22 August 2025 · 閱讀 27 September 2026
https://developer.nvidia.com/blog/inside-nvidia-blackwell-ultra-the-chip-powering-the-ai-factory-era
“With 288 GB of HBM3e per GPU, it offers 3.6x more on-package memory than H100 and 50% more than Blackwell, as shown in Figure 5.”
取自來源表格之數值（轉錄）: NVIDIA GPU chip comparison: Max power (TGP) | Blackwell: Up to 1,200W | Blackwell Ultra: Up to 1,400W
支持的研判. NVIDIA's comparison table gives maximum power (TGP) of up to 1,200 W for Blackwell and up to 1,400 W for Blackwell Ultra.
無法據以確立. Vendor maximum total graphics power per GPU; not typical operating draw, rack power or a cooling threshold.
引用於: 第節 第4.1節, 第4.3節, 第4.5節, 第8.3節; 圖 19, 29

S-175 NVIDIA — NVIDIA DSX Facilities Infrastructure Design Guide v2.0 - Reference Design Overview
企業新聞稿 · U.S. · 英文 · 19 August 2026 · 閱讀 27 September 2026
https://docs.nvidia.com/dsx/facilities-infra/reference-design-overview
“Cabinet TDP scales from MGX Gen 1.1 at 198 kW to Vera Rubin NVL72 at 330 kW.”
“In this reference site, BESS sits on the 34.5 kV switchgear bus as a controllable buffer between them.”
支持的研判. NVIDIA's DSX facilities design guide scales cabinet TDP from 198 kW (MGX Gen 1.1) to 330 kW (Vera Rubin NVL72) and places BESS on the reference site's 34.5 kV bus.
無法據以確立. A vendor reference design for one site type; not the power of deployed cabinets, nor an electrical or grid-code requirement.
引用於: 第節 第4.1節, 第4.2節, 第4.4節; 圖 19

S-176 Engineering at Meta — Meta's open AI hardware vision
企業新聞稿 · U.S. · 英文 · 15 October 2024 · 閱讀 26 September 2026
https://engineering.fb.com/2024/10/15/data-infrastructure/metas-open-ai-hardware-vision
"With Catalina we're introducing the ORv3, a high-power rack (HPR) capable of supporting up to 140kW."
支持的研判. Meta introduced the ORv3 High Power Rack (HPR) with its liquid-cooled Catalina system on 15 October 2024, capable of supporting up to 140 kW.
無法據以確立. Meta's product announcement; not an OCP-released HPR specification (folders still under review) and not a count of deployed racks.
引用於: 第節 第4.1節, 第4.5節, 第A.1節, 第C.7節; 表 5, 21

S-177 Google Cloud Blog — AI infrastructure is hot. New power distribution and liquid cooling infrastructure can help
企業新聞稿 · U.S. · 英文 · 29 April 2025 · 閱讀 27 September 2026
https://cloud.google.com/blog/topics/systems/enabling-1-mw-it-racks-and-liquid-cooling-at-ocp-emea-summit
"We are excited to introduce +/-400 VDC power delivery that can support up to 1 MW per rack."
"we've deployed liquid cooling at GigaWatt scale across more than 2000 TPU Pods in the past seven years with remarkable uptime — consistently at about 99.999%."
支持的研判. Google introduces ±400 VDC for up to 1 MW per rack, says ML will need over 500 kW per IT rack before 2030, and reports ~99.999% liquid-cooling uptime since 2018.
無法據以確立. Google's own fleet and forecast; the 500 kW figure is a projection, and ±400 VDC was then a Mt Diablo draft, not a released specification.
引用於: 第節 第4.1節, 第4.2節, 第4.3節

S-178 NVIDIA Blog — NVIDIA, Partners Drive Next-Gen Efficient Gigawatt AI Factories in Buildup for Vera Rubin
企業新聞稿 · U.S. · 英文 · 13 October 2025 · 閱讀 27 September 2026
https://blogs.nvidia.com/blog/gigawatt-ai-factories-ocp-vera-rubin
"NVIDIA plans to contribute the upgraded rack as well as the compute tray innovations as an open standard for the OCP consortium."
"Foxconn provided details on its 40-megawatt Taiwan data center, Kaohsiung-1, being built for 800 VDC."
支持的研判. NVIDIA (October 2025) plans to contribute the Vera Rubin rack and compute tray to OCP, calls the tray 100% liquid-cooled, and reports Kaohsiung-1 being built for 800 VDC.
無法據以確立. A plan and status as of October 2025; it does not show OCP acceptance, a whole-rack liquid-cooling claim or later construction status.
引用於: 第節 第4.1節, 第4.2節, 第4.3節, 第4.5節, 第B.2節, 第C.7節; 附錄A; 表 21 · **研判:** A-09

S-179 AMD — AMD Helios Rackscale Solution
企業新聞稿 · U.S. · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.amd.com/en/products/rackscale-solutions/helios.html
"It's a reference design, not a product for sale."
"Two 512 lane 200G UALoE switch ASICs with 216 UALoE (x2) active links per ASIC"
"The AMD Helios rackscale design integrates 72 AMD Instinct™ MI455X GPUs with AMD EPYC™ "Venice" CPUs and AMD Pensando™ Vulcano AI NICs networking using UALink"
支持的研判. AMD describes Helios as 'a reference design, not a product for sale': 72 MI455X GPUs with 432 GB HBM4 each, UALink over Ethernet scale-up, liquid-cooled trays, on Open Rack Wide.
無法據以確立. Vendor reference-design page; ORW conformance is AMD's statement, it says liquid cooling reaches compute and switch trays, not 100%, and NICs are 'UEC-ready'.
引用於: 第節 第4.1節, 第4.3節, 第4.5節, 第4.6節, 第A.2節, 第C.7節; 附錄A; 表 21, 22

S-180 Open Compute Project Foundation — Open Rack V3 Base Specification Revision 1.0
標準組織 · Global · 英文 · 22 September 2022 · 閱讀 27 September 2026
https://www.opencompute.org/documents/open-rack-base-specification-version-3-pdf
"The expected nominal input voltage delivered by the Open Rack V3 Power Shelf is 51VDC."
"There are no formal requirements for external dimensions such as depth, height, & width of the frame."
支持的研判. The ORv3 Base Specification Revision 1.0 (released 22 September 2022; Meta and Google) sets a 51 VDC nominal shelf output, 46–52 V IT input tolerance, and no frame-dimension requirements.
無法據以確立. Its '100A continuous' figures rate power and ground paths, not total rack power; revision 1.1 (December 2023) is listed on OCP's wiki.
引用於: 第節 第4.1節, 第4.2節, 第4.5節, 第C.7節; 圖 17; 表 5, 21, 32

S-181 Open Compute Project Foundation (PR Newswire) — Realizing the Open Data Center Ecosystem Vision
標準組織 · Global · 英文 · 13 October 2025 · 閱讀 27 September 2026
https://www.prnewswire.com/news-releases/realizing-the-open-data-center-ecosystem-vision-302582596.html
"announces an expansion to its Open Systems for AI, an umbrella initiative with the new Open Data Center for AI Strategic Initiative"
"Mt Diablo (Diablo 400) power-rack sidecar for powering AI clusters, co-authored by Google, Meta and Microsoft"
支持的研判. OCP announced (13 October 2025) the Open Data Center for AI Strategic Initiative, citing contributions including the Mt Diablo (Diablo 400) power-rack sidecar and Google's Deschutes CDU.
無法據以確立. OCP's own announcement; it does not mention the ASHRAE alliance (S-182) or establish released specifications or adoption.
引用於: 第節 第4.1節, 第C.7節

S-182 Open Compute Project Foundation / ASHRAE — Open Compute Project Foundation and ASHRAE Form New Alliance
標準組織 · Global · 英文 · 13 October 2025 · 閱讀 26 September 2026
https://www.opencompute.org/blog/open-compute-project-foundation-and-ashrae-form-new-alliance
"announce a new alliance focused on data center liquid cooling technology standards and best practices."
支持的研判. OCP and ASHRAE announced on 13 October 2025 an alliance, drawing on ASHRAE TC 9.9, focused on data-centre liquid-cooling technology standards and best practices.
無法據以確立. Announcement of intent; it sets no dated joint deliverables, and ASHRAE re-publishing OCP guidance is described as possible, not planned.
引用於: 第節 第4.1節, 第4.3節, 第4.4節, 第6.3節; 表 11, 13

S-183 Open Compute Project Foundation (OAI sub-project; AMD, H3C, Wiwynn contributors) — OAI-OAM Base Specification r2.0 v1.0
標準組織 · Global · 英文 · 14 September 2023 · 閱讀 27 September 2026
https://www.opencompute.org/documents/oai-oam-base-specification-r2-0-v1-0-20230919-pdf
"The module supports up to 1000W with input voltage of 44V to 59.5V."
"An air-cooled solution recommends TDP equal or less than 600W modules. For modules that are over 600W, other cooling solution is recommended such as liquid cooling."
支持的研判. OAM r2.0 v1.0 (September 2023) supports modules up to 1000 W at 44–59.5 V and recommends air cooling up to 600 W TDP, liquid or other cooling above.
無法據以確立. A 2023 module specification; it does not cover newer vendor-specific accelerator modules or show how many products use OAM.
引用於: 第節 第4.1節, 第4.3節, 第4.5節, 第C.6節; 表 21

S-184 Open Compute Project Foundation (Meta) — Open Rack V3 48V BBU Revision 1.4
標準組織 · Global · 英文 · 12 September 2023 · 閱讀 26 September 2026
https://www.opencompute.org/documents/open-rack-v3-bbu-module-spec-1-4-pdf
"6 BBUs with 5+1 redundancy are included in the BBU shelf."
支持的研判. Meta's Open Rack V3 48V BBU specification Revision 1.4 (12 September 2023) uses lithium-ion cells and places six BBUs with 5+1 redundancy in a BBU shelf.
無法據以確立. A single-vendor module specification; it does not show shelf deployments or suitability for megawatt-class racks.
引用於: 第節 第4.2節

S-185 Microsoft Azure Blog — Accelerating industry-wide innovations in datacenter infrastructure and security
企業新聞稿 · U.S. · 英文 · 15 October 2024 · 閱讀 27 September 2026
https://azure.microsoft.com/en-us/blog/accelerating-industry-wide-innovations-in-datacenter-infrastructure-and-security
"a disaggregated 400 High Voltage Direct Current (VDC) unit that scales from hundreds of kW up to 1MW, enabling 15% to 35% more AI accelerators in each server rack."
"We are solving the increased power infrastructure demands in the age of AI with Mt. Diablo, our latest collaboration with Meta."
支持的研判. Microsoft's 15 October 2024 post announces Mt. Diablo, a collaboration with Meta: a disaggregated 400 VDC power unit scaling up to 1 MW, enabling 15–35% more AI accelerators per rack.
無法據以確立. Microsoft's announcement of a design; the accelerator gain is a vendor claim, and specification status is set by OCP documents.
引用於: 第節 第4.2節; 圖 17

S-186 Tom's Hardware (citing SemiAnalysis) — Nvidia's Kyber rack for Rubin Ultra reportedly delayed to 2028, stopgap solution also axed due to customer pushback
媒體 · U.S. · 英文 · 6 July 2026 · 閱讀 26 September 2026
https://www.tomshardware.com/pc-components/gpus/nvidias-kyber-rack-for-rubin-ultra-slips-to-2028
"Our roadmap is intact"
支持的研判. Tom's Hardware, citing SemiAnalysis, reported in July 2026 that NVIDIA's Kyber rack for Rubin Ultra was delayed to 2028; NVIDIA replied 'Our roadmap is intact'.
無法據以確立. Second-hand report of an analyst claim; NVIDIA did not confirm a delay, and the midplane cause rests on SemiAnalysis.
引用於: 第節 第4.2節, 第4.5節, 第C.7節; 圖 17; 表 5; 展望 2, 5

S-187 ANSI / UL — UL 489I Ed. 1-2025 - Solid State Circuit Breakers
標準組織 · U.S. · 英文 · 22 October 2025 · 閱讀 27 September 2026
https://webstore.ansi.org/standards/ul/ul489ied2025
"These devices are intended to provide service entrance, feeder, and branch circuit protection in accordance with NFPA 70, National Electrical Code (NEC)."
"These requirements cover solid state circuit breakers (SSCB), and solid state hybrid circuit breakers (SSHCB) rated up to 1000 Vac and 1500 Vdc"
支持的研判. UL 489I Ed. 1 (22 October 2025) covers solid-state and hybrid circuit breakers and switches rated up to 1000 Vac and 1500 Vdc for NEC circuit protection.
無法據以確立. A product-safety standard's catalogue entry; it does not name 800 VDC data centres, certified products or coordination schemes.
引用於: 圖 17; 第節 第4.2節, 第A.2節, 第C.7節; 表 5, 13, 22

S-188 UL Solutions (via StockTitan) — UL Solutions Launches Certification Program for Emerging Power Technology Supporting AI Data Centers
企業新聞稿 · U.S. · 英文 · 24 September 2026 · 閱讀 27 September 2026
https://www.stocktitan.net/news/ULS/ul-solutions-launches-certification-program-for-emerging-power-j52svg4cno25.html
"UL Solutions Inc. (NYSE: ULS), a global leader in applied safety science, today announced a new certification program for medium-voltage solid-state power units, an emerging technology that can deliver electricity more directly and efficiently to the data center systems powering AI."
"The certification program is based on UL 2877, the Outline for Investigation for Power Supplies, Medium Voltage."
支持的研判. UL Solutions announced 24 September 2026 a certification program for medium-voltage solid-state power units for AI data centres, based on UL 2877, its outline of investigation for medium-voltage power supplies.
無法據以確立. Launch announcement; it records no certified unit, and an outline of investigation is not a consensus standard.
引用於: 圖 17; 第節 第4.2節, 第A.1節, 第A.2節, 第C.7節; 表 5, 22, 32

S-189 Vertiv — From vision to readiness: Vertiv collaborates with NVIDIA to advance 800 VDC platform designs…
企業新聞稿 · U.S. · 英文 · 13 October 2025 · 閱讀 26 September 2026
https://www.vertiv.com/en-us/about/news-and-insights/corporate-news/from-vision-to-readiness-vertiv-collaborates-with-nvidia-to-advance-800-vdc-platform-designs-to-power-the-next-generation-of-ai-factories
"The 800 VDC power portfolio is planned to release in the second half of 2026, aligning to support the 2027 rollout of NVIDIA Rubin Ultra platforms."
支持的研判. Vertiv said (13 October 2025) its 800 VDC power portfolio is planned for release in 2H 2026, aligned with NVIDIA's 2027 Rubin Ultra rollout.
無法據以確立. A planned release window; not shipment, certification or customer deployment of 800 VDC products.
引用於: 圖 17; 第節 第4.2節, 第B.1節, 第B.2節; 表 5, 28

S-190 NVIDIA — Why Scaling AI Compute Performance Requires a New Power Architecture
企業新聞稿 · U.S. · 英文 · 11 August 2026 · 閱讀 27 September 2026
https://blogs.nvidia.com/blog/800-vdc-power-architecture-ai-factory
"supporting up to 2 megawatts per row, with availability expected in 2027"
"NVIDIA, Google and Microsoft have been developing the 800 VDC architecture together through the Open Compute Project (OCP), and published a joint white paper March 2026 and the LVDC Solid-State Transformer Specification v0.3 July 2026. More than 80 equipment manufacturers and infrastructure companies are already building products to this specification."
支持的研判. NVIDIA says over 80 companies are building products to the 800 VDC specification developed with Google and Microsoft through OCP, and expects 2 MW-per-row power centres in 2027.
無法據以確立. A vendor blog; the company count is NVIDIA's own claim, and row power-centre availability in 2027 is an expectation.
引用於: 圖 17; 第節 第4.2節, 第4.4節; 表 5

S-191 GlobeNewswire / ABB — ABB's new direct current portfolio aims to rewire AI data center energy infrastructure
企業新聞稿 · Switzerland · 英文 · 21 September 2026 · 閱讀 27 September 2026
https://www.globenewswire.com/news-release/2026/09/21/3365178/0/en/abb-s-new-direct-current-portfolio-aims-to-rewire-ai-data-center-energy-infrastructure.html
"the only major electrification supplier with both the world's first fully IEC-certified solid-state circuit breaker and an industry-first static medium-voltage UPS in production"
"ABB introduces Infinitus, the industry's first source-to-rack direct current portfolio for AI data centers designed for architecture flexibility"
支持的研判. ABB introduced its Infinitus source-to-rack DC portfolio for AI data centres on 21 September 2026, expecting first DC-native facilities within two to three years; it calls its SSCB world-first IEC-certified.
無法據以確立. Company claims; several main components are still to be produced over the next 12 months; 'first' claims are ABB's own.
引用於: 圖 17; 第節 第4.2節, 第5.6節; 表 5, 28

S-192 Infineon Technologies AG (distributed via PR Newswire) — Infineon Joins NVIDIA's MGX™ AI Factory Ecosystem to Transform Power Delivery Architecture for Next-Generation AI Server Racks
企業新聞稿 · Germany · 英文 · 28 May 2026 · 閱讀 27 September 2026
https://www.prnewswire.com/news-releases/infineon-joins-nvidias-mgx-ai-factory-ecosystem-to-transform-power-delivery-architecture-for-next-generation-ai-server-racks-302785037.html
"The combination of Infineon's proprietary SiC JFET technology and dedicated control ICs are the perfect match for protection and hot-swap functionality of native 800 V server boards."
"Infineon's power management solutions convert power from 800 V to 50 V, 12 V or even down to 6 V."
支持的研判. Infineon says its solutions will support NVIDIA's MGX 800 VDC architecture, converting 800 V to 50, 12 or 6 V, with SiC JFETs for 800 V board protection and hot-swap.
無法據以確立. Company announcement (May 2026); gives no shipment timing, customers or volumes; 'perfect match' is marketing language.
引用於: 第節 第4.2節, 第5.4節, 第6.7節; 表 5, 28

S-193 Infineon Technologies AG — Infineon Analyst Call Q3 FY26 Intro Statement; 5 August 2026
企業新聞稿 · Germany · 英文 · 5 August 2026 · 閱讀 27 September 2026
https://www.infineon.com/row/public/documents/corporate/investors/presentations/2026/2026-08-05-q3-fy26-intro-statement-analyst-call-v01-00-en.pdf
"Successful execution of capacity ramps and conversions from other areas will help us achieve more than 1.6 billion Euros of dedicated AI power revenues in the current fiscal year, ahead of the so-far planned 1.5 billion."
"We will revise our projection of 2.5 billion Euros plus AI Datacenter revenues for our 2027 fiscal year upwards as part of the annual guidance to be given in our November earnings call."
"These agreements encompass a total cumulative sales volume of a high single-digit billion Euro amount over multiple years."
支持的研判. Infineon expects over €1.6 billion of AI power revenue in FY2026 (vs €1.5 billion planned), will raise its FY2027 '€2.5 billion plus' projection in November, and reports reservations with prepayments.
無法據以確立. Management expectations in an analyst-call statement; the FY2027 figure awaits revision, and some agreements are still being negotiated.
引用於: 第節 第4.2節, 第5.4節, 第5.6節, 第5.8節; 表 7, 10, 13, 19, 28; 展望 13

S-194 Eaton — Eaton unveils next-generation architecture to advance 800 VDC power infrastructure for AI factories
企業新聞稿 · Ireland · 英文 · 13 October 2025 · 閱讀 26 September 2026
https://www.eaton.com/us/en-us/company/news-insights/news-releases/2025/eaton-unveils-next-generation-architecture.html
"a new reference architecture designed to accelerate the adoption of 800 VDC power in artificial intelligence (AI) data centers"
支持的研判. Eaton unveiled (13 October 2025) an 800 VDC reference architecture for AI data centres, including supercapacitors for fast-cycle backup.
無法據以確立. A reference architecture announcement; no product availability dates, certifications or customer deployments are given.
引用於: 第節 第4.2節; 表 28

S-195 Schneider Electric — Schneider Electric Highlights Innovation in 800 VDC Power Systems in support of NVIDIA's next generation GPUs
企業新聞稿 · France · 英文 · 13 October 2025 · 閱讀 26 September 2026
https://www.se.com/ww/en/about-us/newsroom/news/press-releases/schneider-electric-highlights-innovation-in-800-vdc-power-systems-in-support-of-nvidia%E2%80%99s-next-generation-gpus-68ecb631098fd4ea49024353
"Schneider Electric is collaborating with NVIDIA to develop an 800 VDC sidecar capable of powering racks of up to 1.2 MW"
支持的研判. Schneider Electric is collaborating with NVIDIA to develop an 800 VDC sidecar for racks of up to 1.2 MW, with modular energy storage for backup and load smoothing.
無法據以確立. A development announcement with no availability date, certification or customer deployment.
引用於: 第節 第4.2節, 第C.7節; 表 5, 28

S-196 Vertiv Holdings Co — Vertiv Reports Strong Second Quarter 2026 with Diluted EPS Growth of 53% (Adjusted Diluted EPS Growth of +60%); Raises Full Year 2026 Guidance Across All Key Metrics
企業新聞稿 · U.S. · 英文 · 29 July 2026 · 閱讀 27 September 2026
https://investors.vertiv.com/news/news-details/2026/Vertiv-Reports-Strong-Second-Quarter-2026-with-Diluted-EPS-Growth-of-53-Adjusted-Diluted-EPS-Growth-of-60-Raises-Full-Year-2026-Guidance-Across-All-Key-Metrics/default.aspx
直接文件: https://s205.q4cdn.com/554782763/files/doc_financials/2026/q2/Vertiv-Second-Quarter-2026-Results-Presentation.pdf
"The strong demand environment, growing pipelines, and the continual capacity expansions underway to serve customers give Vertiv the confidence to raise full year net sales guidance, along with increases to other associated financial metrics."
"Path to 800 VDC validation in action with VisionBay.AI K1 site in Kaohsiung, Taiwan"
支持的研判. Vertiv raised full-year 2026 guidance to net sales of US$13.8–14.2 billion and adjusted operating margin of 23.3–24.3%, citing growing pipelines, and shows an 800 VDC validation path in Kaohsiung.
無法據以確立. Company guidance for the whole group; it gives no backlog or orders figure, and the Kaohsiung item is a validation path, not a product release.
引用於: 第節 第4.2節, 第5.1節, 第5.4節, 第5.7節, 第5.8節; 表 7, 9, 10, 13, 19

S-197 ABB — ABB to develop next-generation AI data centers with NVIDIA
企業新聞稿 · Switzerland · 英文 · 13 October 2025 · 閱讀 26 September 2026
https://new.abb.com/news/detail/129805/abb-to-develop-next-generation-ai-data-centers-with-nvidia
"ABB's SACE Infinitus is the world's first IEC-certified solid-state circuit breaker"
支持的研判. ABB states that its SACE Infinitus is 'the world's first IEC-certified solid-state circuit breaker', as it partners with NVIDIA on 800 VDC AI data centres.
無法據以確立. ABB's own 'world's first' claim, not independently verified here; IEC certification is not a UL listing for North American markets.
引用於: 第節 第4.2節, 第7.4節, 第C.7節; 表 5, 28

S-198 ABB — SACE Infinitus - Circuit breakers | ABB
企業新聞稿 · Switzerland · 英文 · 未註明日期 · 閱讀 26 September 2026
https://new.abb.com/low-voltage/products/circuit-breakers/sace-infinitus
“SACE Infinitus's 25 microsecond breaking time and immediate fault isolation could mean up to $100K USD cost savings”
支持的研判. ABB's product page gives SACE Infinitus a 25-microsecond breaking time, with a 2,500 A frame and operation up to 1,250 V DC.
無法據以確立. Vendor specification of a device's breaking time; not a system fault-clearing time, and the cost-saving figure is a vendor claim.
引用於: 第節 第4.2節, 第7.4節; 表 5

S-199 Open Compute Project Foundation — OCP wiki: Data_Center_Facility/Power
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/w/index.php?title=Data_Center_Facility/Power
“Codes and Standards --> Lisa Grams (UL Solutions)”
支持的研判. OCP's Data Center Facility Power wiki names UL Solutions (Lisa Grams) as lead of the LVDC Codes and Standards workstream; ABB leads Protection and Safety.
無法據以確立. Workstream leadership as read in September 2026; it does not show outputs or certification decisions by UL Solutions.
引用於: 第節 第4.2節

S-200 ServeTheHome — NVIDIA Starts to Tackle GPU Power Smoothing with the NVIDIA GB300 NVL72
媒體 · U.S. · 英文 · 28 July 2025 · 閱讀 27 September 2026
https://www.servethehome.com/nvidia-starts-to-tackle-gpu-power-smoothing-with-the-nvidia-gb300-nvl72
“One of the interesting ones that NVIDIA is doing is now a GPU burn.”
“NVIDIA also has added things like more capacitance in its power supplies.”
支持的研判. ServeTheHome reports NVIDIA smoothing GB300 NVL72 power swings with a 'GPU burn' workload technique and more capacitance in its power supplies.
無法據以確立. Trade-press description of NVIDIA's presentation; no quantified smoothing or wattage figures are given.
引用於: 第節 第4.2節

S-201 Open Compute Project Foundation — Energy (OCP community project page)
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/community/energy
“AI workload transient response, buffering, resilience, and power-quality considerations”
支持的研判. OCP's Energy sub-project lists AI workload transient response, buffering, resilience and power-quality considerations within the scope of its energy-storage work.
無法據以確立. Scope list only; it gives no load-swing magnitudes or timescales and no requirements.
引用於: 第節 第4.2節

S-202 Open Compute Project Foundation — Open Rack/SpecsAndDesigns - OpenCompute (wiki)
標準組織 · Global · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.opencompute.org/wiki/Open_Rack/SpecsAndDesigns
“These submissions are available for public comment while the Incubation Committee considers including them in the RACK and POWER project.”
“Open Rack V3 Base Specification 1.1 DEC23 META Google”
“High Power ORV3 System Specs Folder Varies OCT24 META”
支持的研判. OCP's Open Rack wiki lists ORv3 Base Specification 1.1 (December 2023), ORW specifications, and HPR folders and a preliminary ORW Canister guide as proposals under review.
無法據以確立. Wiki last edited April 2026; listings under review are not accepted specifications, and folder contents were not readable.
引用於: 表 5, 21; 第節 第4.5節, 第A.1節; 附錄A

S-203 Business Wire / Amperesand — Amperesand Raises $80M Series A Co-led by Walden Catalyst Ventures and Temasek to Redefine Power Infrastructure for AI Data Centers
企業新聞稿 · Singapore · 英文 · 18 November 2025 · 閱讀 27 September 2026
https://www.businesswire.com/news/home/20251118244299/en/Amperesand-Raises-$80M-Series-A-Co-led-by-Walden-Catalyst-Ventures-and-Temasek-to-Redefine-Power-Infrastructure-for-AI-Data-Centers
“30MW of commercial systems to be delivered in 2026 to Hyperscale AI data center and critical power customers”
支持的研判. Amperesand said in November 2025, raising a US$80 million Series A, that 30 MW of commercial SST systems are to be delivered in 2026 to hyperscale AI and critical-power customers.
無法據以確立. Company plan at a funding round; delivery, named customers and certification are not established.
引用於: 表 5; 第節 第4.4節

S-204 Heron Power — Heron Power (company site, news section)
企業新聞稿 · U.S. · 英文 · 18 February 2026 · 閱讀 27 September 2026
https://heronpower.com
“Next-gen transformer startup Heron Power is building its first 10 engineering prototypes this summer and will deploy some of those with early customers, Heron Power founder and CEO Drew Baglino tells Axios Pro.”
支持的研判. Heron Power's site relays an Axios report that it is building its first 10 engineering prototypes of its next-generation transformer in summer 2026, some for early customers.
無法據以確立. Company site quoting Axios; 'this summer' is dated 2026 by the item, and prototypes are plans, not deliveries or certified products.
引用於: 表 5; 第節 第4.4節

S-205 Open Compute Project Foundation — White Paper: Requirements for Energy Storage Systems Used in Data Centers (Rev 1.0, Ver 1.0)
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/documents/white-paper-ocp-ready-requirements-for-energy-storage-systems-final-pdf
“NFPA 855 (2026) Annex G provides guidance on the scope and implementation of the Large Scale Fire Testing now required by NFPA 855 in 9.2.1.”
“This white paper explores the requirements for Energy Storage Systems (ESS) deployments within data centers, focusing on critical considerations for safe and reliable deployment in hyperscale wholesale environments.”
支持的研判. OCP's white paper on energy storage systems in data centres (Revision 1.0, January 2026) records that NFPA 855 (2026) now requires large-scale fire testing.
無法據以確立. A white paper summarising codes and practice, not a code; the NFPA 855 requirement binds only where that code is adopted.
引用於: 表 5, 21; 第節 第4.4節, 第7.4節, 第C.7節

S-206 NVIDIA — NVIDIA DSX - BESS Self-Qualification Guidelines
企業新聞稿 · U.S. · 英文 · 28 May 2026 · 閱讀 26 September 2026
https://docs.nvidia.com/dsx/facilities-infra/bess/overview
“the partner-run qualification process for BESS systems that support AI load buffering, demand response (DR), and low voltage ride-through (LVRT) in grid-connected and islanded onsite generation modes”
支持的研判. NVIDIA's DSX BESS Self-Qualification Guidelines v1.0 (28 May 2026) define partner-run qualification of BESS for AI load buffering, demand response and low-voltage ride-through.
無法據以確立. Vendor qualification guidelines; a partner self-qualification route, not an independent certification or a safety standard such as NFPA 855.
引用於: 表 5; 第節 第4.4節

S-207 Open Compute Project Foundation — Realizing the Open Data Center Ecosystem Vision
標準組織 · Global · 英文 · 13 October 2025 · 閱讀 26 September 2026
https://www.opencompute.org/blog/realizing-the-open-data-center-ecosystem-vision
“targeted to support ~2 MW heat loads, with hydraulic capacity targets of 500 GPM at 80-90 psi”
支持的研判. OCP's October 2025 post says Google's Deschutes CDU is targeted to support about 2 MW heat loads, with hydraulic capacity targets of 500 GPM at 80–90 psi.
無法據以確立. Pre-release targets from October 2025; the released 1.0.0 specification (S-030) is the authority, and the post names no vendors building the unit.
引用於: 第節 第4.3節, 第C.7節; 圖 19; 表 21

S-208 Open Compute Project Foundation Marketplace (vendor: Boyd Thermal by Eaton) — Eaton ROL4000-48U65 Deschutes CDU » Open Compute Project
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/ai-marketplace/products/728/boyd-rol4000-48u65-deschutes-cdu
“Inspired by OCP's fifth generation Project Deschutes CDU specification.”
支持的研判. OCP's Marketplace lists Eaton's (Boyd Thermal) ROL4000-48U65 CDU as 'Inspired by' Project Deschutes, quoting 2 MW at a 3 °C approach, with no OCP Accepted designation shown.
無法據以確立. Vendor-supplied, undated listing; performance figures are the vendor's claims, not verified conformance to the Deschutes specification.
引用於: 第節 第4.3節, 第5.4節; 表 21, 28

S-209 LG Electronics — LG Electronics' 2.6-Megawatt Coolant Distribution Unit Qualifies as NVIDIA DSX Ready CDU
企業新聞稿 · Korea · 英文 · 22 September 2026 · 閱讀 27 September 2026
https://www.lg.com/global/newsroom/news/eco-solution/lg-electronics-2-5-megawatt-coolant-distribution-unit-qualifies-as-nvidia-dsx-ready-cdu
“LG Electronics has completed qualification for its 2.6-megawatt (MW) Coolant Distribution Unit (CDU), meeting the applicable NVIDIA DSX reference design requirements for DSX AI Factory Platform.”
支持的研判. LG Electronics says (22 September 2026) its 2.6 MW CDU completed NVIDIA DSX Ready qualification, meeting the applicable DSX reference design requirements.
無法據以確立. Vendor's own qualification announcement; no customers, shipments or independent test data; the URL slug says 2.5 MW, headline and body 2.6 MW.
引用於: 第節 第4.3節, 第5.6節, 第B.2節; 展望 3; 表 28

S-210 Dell'Oro Group — Market at the Boiling Point: Is the CDU Space Becoming Saturated?
市場研究 · U.S. · 英文 · 30 September 2025 · 閱讀 26 September 2026
https://www.delloro.com/market-at-the-boiling-point-is-the-cdu-space-becoming-saturated
“around 40 companies with CDUs within their product portfolios”
支持的研判. Dell'Oro counts around 40 companies with CDUs in their product portfolios (September 2025) and expects consolidation, with fewer than 10 vendors eventually holding most of the market.
無法據以確立. Analyst blog view of September 2025; the count is not a market-share measure, and 'fewer than 10' is an expectation.
引用於: 第節 第4.3節

S-211 Boyd — Boyd Completes the Sale of its Thermal Business to Eaton
企業新聞稿 · U.S. · 英文 · 12 March 2026 · 閱讀 27 September 2026
https://www.boydcorp.com/boyd-completes-the-sale-of-its-thermal-business-to-eaton-for-9-5-billion.html
“Boyd Corporation ("Boyd"), the trusted global innovator in engineered materials and thermal solutions, today announced it has completed the sale of its Thermal business to Eaton, an intelligent power management company, for $9.5 billion”
支持的研判. Boyd announced on 12 March 2026 that it completed the sale of its Thermal business to Eaton for US$9.5 billion.
無法據以確立. The release does not describe the Thermal business's products (such as CDUs or cold plates) or its market share.
引用於: 第節 第4.3節, 第5.4節; 表 7, 28

S-212 Open Compute Project Foundation (contributed by Intel) — Universal Quick Disconnect (UQD) Specification Revision 1.0; Universal Quick Disconnect Blind-Mate (UQDB) Specification Revision 1.0
標準組織 · Global · 英文 · 4 September 2020 · 閱讀 26 September 2026
https://www.opencompute.org/documents/ocp-universal-quick-disconnect-uqd-specification-rev-1-0-2-pdf
"Maximum operating pressure 100 psi"
"Universal Quick Disconnect (UQD) Specification Revision 1.0"
支持的研判. Intel's UQD Specification Revision 1.0 (4 September 2020; filed as 1.0.2) defines UQD02–UQD08 couplings with a 100 psi maximum operating pressure and a 17–65 °C range.
無法據以確立. A 2020 coupling specification; a v2 revision is still in draft, and it does not set the loop pressures a CDU delivers.
引用於: 圖 19; 第節 第4.3節, 第C.7節; 表 21

S-213 ASHRAE — Emergence and Expansion of Liquid Cooling in Mainstream Data Centers (ASHRAE TC 9.9 white paper)
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.ashrae.org/file%20library/technical%20resources/bookstore/emergence-and-expansion-of-liquid-cooling-in-mainstream-data-centers_wp.pdf
"W17 (previously W1), W27 (W2), W32 (W3), W40 (new), W45 (W4), and W+ (W5)."
支持的研判. ASHRAE's TC 9.9 liquid-cooling white paper renames the facility-water classes W17 (previously W1), W27, W32, W40 (new), W45 and W+ (previously W5).
無法據以確立. Class names only; temperature bounds and the 2021 date are not in the excerpt, and the paper sets no OCP coupling or CDU requirement.
引用於: 圖 19; 第節 第4.3節; 表 11

S-214 Open Compute Project Foundation — Cold Plate » Open Compute Project
標準組織 · Global · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.opencompute.org/community/cold-plate
"The sub-project produces requirements/guidelines, specs, best practices and products which are related to the technology cooling system (TCS) from the cold plate to the CDU."
支持的研判. OCP's Cold Plate sub-project produces requirements, specifications, best practices and products for the technology cooling system (TCS) from the cold plate to the CDU.
無法據以確立. Scope statement on an undated project page; it lists no versioned documents and does not show which parts of the loop are specified.
引用於: 第節 第4.3節

S-215 Open Compute Project Foundation — Universal Quick Disconnect Blind-Mate (UQDB) Specification Revision 1.0
標準組織 · Global · 英文 · 4 September 2020 · 閱讀 26 September 2026
https://www.opencompute.org/documents/uqdb-spec-1-0-pdf
"Universal Quick Disconnect Blind-Mate (UQDB) Specification"
支持的研判. Intel's UQDB (blind-mate) Specification Revision 1.0 (4 September 2020) covers UQDB02–08 couplings at 100 psi and 17–65 °C, with a 1.0 mm radial self-alignment allowance.
無法據以確立. A 2020 specification; a v2 hand-mate and blind-mate revision is still in draft, and it shows no multi-vendor mating tests.
引用於: 第節 第4.3節, 第C.7節; 表 21

S-216 Open Compute Project Foundation — Guidelines for Using Propylene Glycol-Based Heat Transfer Fluids in Single-Phase Cold Plate-Based Liquid Cooled Racks
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/documents/guidelines-for-using-propylene-glycol-based-heat-transfer-fluids-in-single-phase-cold-plate-based-liquid-cooled-racks-final-pdf
"PG25: 9 to 15°F (-13 to -9 °C)"
"The typical operating temperature of closed loop cooling systems is <49°C (120°F) and is not expected to be >66°C (150°F)."
支持的研判. OCP's final, undated PG guideline (Dow, Third Coast Chemicals, Intel, Meta) gives PG25 a freeze point of −13 to −9 °C and closed-loop operating temperatures typically below 49 °C.
無法據以確立. A guideline, not a specification: its values are recommendations, and it does not address fluorinated fluids or regulatory status.
引用於: 第節 第4.3節, 第7.4節, 第C.7節; 表 13, 21

S-217 Open Compute Project Foundation — Base Specification for Single Phase Water-Based Cold Plate Coolants, Version 1.3
標準組織 · Global · 英文 · 6 August 2026 · 閱讀 26 September 2026
https://www.opencompute.org/documents/base-specification-for-single-phase-water-based-cold-plate-coolant-v1-3-final-pdf
"This specification defines the technical details on the use of water-based coolants for single-phase cold plate cooling systems within the OCP ecosystem."
支持的研判. OCP's Base Specification for Single Phase Water-Based Cold Plate Coolants v1.3, effective 6 August 2026 and ChemTreat-led, sets pH 8.0–10.5 and conductivity below 2,000 µS/cm.
無法據以確立. Covers water-based single-phase cold-plate coolants only, not PG25 glycol or two-phase dielectric fluids.
引用於: 第節 第4.3節, 第6.7節, 第A.1節, 第C.7節; 圖 26; 表 13, 21

S-218 ASHRAE TC 9.9 — TC 9.9 Mission Critical Facilities, Data Centers, Technology Spaces and Electronic Equipment
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 27 September 2026
https://tpc.ashrae.org/?cmtKey=fd4a4ee6-96a3-4f61-8b85-43418dfa988d
"Technical Bulletin - TCS Coolant Integrity and System Readiness Best Practices"
支持的研判. ASHRAE TC 9.9's committee page lists, among new TC 9.9 publications, a Technical Bulletin on TCS Coolant Integrity and System Readiness Best Practices.
無法據以確立. The page gives no publication date, so '2026 outputs' is not shown; it does not describe the bulletin's content.
引用於: 第節 第4.3節

S-219 Microsoft Source — AI chips are getting hotter. A microfluidics breakthrough goes straight to the silicon to cool up to three times better.
企業新聞稿 · U.S. · 英文 · 24 September 2025 · 閱讀 26 September 2026
https://news.microsoft.com/source/features/innovation/microfluidics-liquid-cooling-ai-chips
"The company's lab-scale tests showed microfluidics performed up to three times better than cold plates at removing heat."
支持的研判. Microsoft reports that in-chip microfluidic cooling, developed with Corintis, removed heat up to three times better than cold plates in lab-scale tests.
無法據以確立. Lab-scale company tests; not a product, deployment date or independent measurement.
引用於: 第節 第4.3節

S-220 Open Compute Project Foundation — OCP Global Summit 2026: Innovation Village & Future Technologies Symposium (FTS) Demo Lineup
標準組織 · Global · 英文 · 22 September 2026 · 閱讀 26 September 2026
https://www.opencompute.org/blog/ocp-global-summit-2026-innovation-village-and-future-technologies-symposium-fts-demo-lineup
"a fully operational Blackwell server with a transparent window showing live flow-boiling over the GPU under real AI workloads."
"An open, vendor-neutral rack-level coordination layer for meeting grid codes without a central UPS, extending several OCP power specs."
支持的研判. OCP's 22 September 2026 lineup lists Summit demos: two-phase flow boiling on a working Blackwell server, a Deschutes-sized 1 MW heat exchanger, rack-level grid-code coordination, liquid-cooled NIC 3.0 mock-ups.
無法據以確立. Scheduled demonstrations for 12–15 October 2026, not yet shown at cut-off; they are prototypes, not specifications, and the post does not cover OCP's fluid guidelines.
引用於: 第節 第4.3節, 第4.5節, 第6.2節, 第6.7節, 第C.7節; 表 13, 21

S-221 Open Compute Project Foundation — Cooling Environments/Immersion - OpenCompute (wiki); OCP Immersion Requirements Rev 2.10
標準組織 · Global · 英文 · 18 August 2023 · 閱讀 27 September 2026
https://www.opencompute.org/w/index.php?title=Cooling_Environments/Immersion
"Aug 18, 2023: Immersion Requirements rev2.1"
"2.10 August 18, 2023 Amended publication"
支持的研判. OCP's Immersion wiki lists Immersion Requirements as an accepted contribution dated 18 August 2023; the linked document's revision history records revision 2.10, an 'Amended publication'.
無法據以確立. The wiki labels it 'rev2.1'; neither source shows newer immersion work, two-phase immersion requirements or adoption.
引用於: 第節 第4.3節; 表 21

S-222 Carrier — Carrier Ventures Expands Investment in ZutaCore to Scale Liquid Cooling for AI Data Centers
企業新聞稿 · U.S. · 英文 · 29 April 2026 · 閱讀 27 September 2026
https://www.carrier.com/us/en/news/carrierventures-expands-investment-in-zutacore-to-scale-liquid-cooling-for-ai-data-centers
"today announced that its venture group, Carrier Ventures, has expanded its investment in ZutaCore, a provider of direct-to-chip, waterless liquid cooling solutions."
支持的研判. Carrier says its venture group, Carrier Ventures, expanded its investment in ZutaCore, a direct-to-chip waterless liquid-cooling provider, on 29 April 2026.
無法據以確立. Investment announcement; does not establish ZutaCore's deployments, customers or market position.
引用於: 第節 第4.3節

S-223 Data Center Frontier — Cooling Consolidation Hits AI Scale: LiquidStack, Submer, and the Future of Data Center Thermal Strategy
媒體 · U.S. · 英文 · 11 February 2026 · 閱讀 27 September 2026
https://www.datacenterfrontier.com/cooling/article/55356895/cooling-consolidation-hits-ai-scale-liquidstack-submer-and-the-future-of-data-center-thermal-strategy
"on February 10 announced it has entered into a definitive agreement to acquire LiquidStack"
支持的研判. Data Center Frontier reports that Trane Technologies announced on 10 February 2026 a definitive agreement to acquire LiquidStack.
無法據以確立. Press report of a signed agreement, not a completed acquisition; terms and closing are not given.
引用於: 第節 第4.3節

S-224 AMD — AMD Instinct MI350 Series
企業新聞稿 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.amd.com/en/products/accelerators/instinct/mi350.html
"from standard air cooled UBB-based servers to ultra dense Direct Liquid Cooled (DLC) platforms"
支持的研判. AMD says its MI350 Series platforms range ' from standard air cooled UBB-based servers to ultra dense Direct Liquid Cooled (DLC) platforms'.
無法據以確立. Vendor product-range description; no split of shipments between air- and liquid-cooled platforms.
引用於: 第節 第4.3節, 第4.5節; 表 21

S-225 Google Cloud Blog — Introducing Brazos: Bringing liquid cooling to air-cooled data centers
企業新聞稿 · U.S. · 英文 · 16 June 2026 · 閱讀 27 September 2026
https://cloud.google.com/blog/topics/systems/brazos-liquid-cooling-system-for-air-cooled-data-centers
"The Google-developed Brazos is a rack-mounted, closed-loop liquid-to-air cooling system that you can deploy in an existing air-cooled environment."
"Each modular chassis occupies 11 Open Units (OU) of rack height and interfaces with standard Open Compute Project (OCP) ORv3 form-factor racks."
支持的研判. Google says Brazos is a closed-loop liquid-to-air system for air-cooled halls: three modular units, each 11 OU and ORv3-compatible, rejecting a nominal 60 kW per rack to the hot aisle.
無法據以確立. Vendor description of its own design; not a deployment count, and the 60 kW is nominal per rack across all three units, not per unit.
引用於: 第節 第4.3節

S-226 NVIDIA — NVIDIA at OCP Global Summit 2026 | Oct. 12–15, San Jose
企業新聞稿 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.nvidia.com/en-us/events/ocp-summit
“AI Factories at Scale: Advancing Open Infrastructure for Agentic AI”
“more than 200 ecosystem partners”
支持的研判. NVIDIA's OCP Summit 2026 page lists an Ian Buck keynote and about 30 sessions, including 800 VDC, liquid-cooled rack handling and CDUs, and says 200-plus partners adopted MGX.
無法據以確立. A pre-event agenda; session titles are not published specifications, products or outcomes of the Summit.
引用於: 第節 第4.3節, 第4.4節, 第6.2節, 第C.7節

S-227 Open Compute Project Foundation — Open Data Center for AI: A Strategic Initiative of the Open Compute Project (white paper)
標準組織 · Global · 英文 · April 2026 · 閱讀 26 September 2026
https://www.opencompute.org/documents/ocp-open-data-center-for-ai-whitepaper-final-pdf
“Open and Standard Service Level Agreements with Colocation, Neocloud and other 3rd party facility operators”
支持的研判. OCP's Open Data Center for AI white paper (v0.5, March 2026) lists five priorities, including standard SLAs with colocation and neocloud operators and power estimation to minimise stranded power.
無法據以確立. A strategy white paper at v0.5; it sets priorities and goals, not requirements, and allocates no risk between operators and tenants.
引用於: 第節 第4.4節, 第C.7節; 表 21

S-228 Open Compute Project Foundation (Rittal, Meta) — Open Rack Wide (ORW) Base Specification V1.0.0
標準組織 · Global · 英文 · 28 April 2026 · 閱讀 27 September 2026
https://www.opencompute.org/documents/open-rack-wide-orw-base-specification-v1-0-0-final-pdf
“ORW Base Specification is designed to support the growing needs of densely populated IT gear racks, requiring a larger footprint than a traditional ORv3 based IT gear rack.”
“There are no prescribed requirements in the base specification for external frame dimensions such as depth, height, base geometry, or maximum payload capability.”
支持的研判. ORW Base Specification V1.0.0 (28 April 2026; Rittal and Meta) serves racks needing a larger footprint than ORv3 and prescribes no external frame dimensions or maximum payload.
無法據以確立. Interface and tolerance base only; frame size and payload sit in separate design specifications, and it does not show multi-vendor supply.
引用於: 第節 第4.4節, 第4.5節, 第6.6節, 第6.7節, 第A.1節, 第C.7節; 展望 5; 表 21, 32 · **研判:** A-06

S-229 PJM Inside Lines — PJM, Dominion Review Large Load Transfer Event
監管機關 · U.S. · 英文 · 11 August 2026 · 閱讀 26 September 2026
https://insidelines.pjm.com/pjm-dominion-review-large-load-transfer-event
“In all, approximately 3,800 MW of Large Load disconnected and transferred without warning”
支持的研判. PJM and Dominion report that about 3,800 MW of large load disconnected and transferred without warning after a 230 kV line trip in Northern Virginia on 22 July 2026.
無法據以確立. Operator's own account; date, trigger and 'largest in PJM's history' are in the article, not this excerpt; why loads disconnected is not stated here.
引用於: 第節 第4.4節, 第7.2節, 第C.7節

S-230 NERC — Large Loads Frequently Asked Questions
監管機關 · U.S. · 英文 · August 2026 · 閱讀 27 September 2026
https://www.nerc.com/globalassets/initiatives/large-loads-action-plan/large-loads-faqs.pdf
“managing rapid fluctuations in consumption during all conditions—including fault and normal—which can destabilize the grid”
“Three foundational standards will address: CLO-001-1 Computational Load Interconnection, Studies, & Modeling Data. CLO-002-1 Computational Load Operational Data and Communications. CLO-003-1 Computational Load Protection Coordination and Disturbance Monitoring.”
支持的研判. NERC's August 2026 FAQ says three foundational computational-load standards, including CLO-003-1 on protection coordination and disturbance monitoring, are due to FERC by 31 December 2026, to manage rapid consumption swings.
無法據以確立. Standards under development, not yet approved or in force; the FAQ does not give their final requirements.
引用於: 第節 第4.4節, 第7.2節

S-231 Bloom Energy (investor release) — Bloom Energy and Oracle Expand Strategic Partnership to Deploy up to 2.8 GW to Accelerate AI Infrastructure Build-Out
企業新聞稿 · U.S. · 英文 · 13 April 2026 · 閱讀 27 September 2026
https://investor.bloomenergy.com/press-releases/press-release-details/2026/Bloom-Energy-and-Oracle-Expand-Strategic-Partnership-to-Deploy-up-to-2-8-GW-to-Accelerate-AI-Infrastructure-Build-Out/default.aspx
“Under a master services agreement, Oracle intends to procure up to 2.8 gigawatts (GW) of Bloom's fuel cell systems.”
“As part of this agreement, an initial 1.2 GW of capacity has been contracted, with deployment underway and continuing into next year.”
支持的研判. Under a master services agreement Oracle intends to procure up to 2.8 GW of Bloom fuel cells, 1.2 GW contracted; Bloom says it delivered one Oracle system in 55 days.
無法據以確立. Company release; 2.8 GW is an intended ceiling, not firm orders; the 55-day delivery was one earlier system.
引用於: 第節 第4.4節, 第7.3節, 第C.7節; 表 29

S-232 Vertiv (via StockTitan) — Vertiv Announces Agreement to Acquire UtilityInnovation Group
企業新聞稿 · U.S. · 英文 · 2 September 2026 · 閱讀 26 September 2026
https://www.stocktitan.net/news/VRT/vertiv-announces-agreement-to-acquire-utility-innovation-group-to-1k5ugtxlg1gv.html
“competitive advantage increasingly depends on how quickly they can move from site selection to first token”
支持的研判. Vertiv agreed to acquire UtilityInnovation Group, a microgrid and behind-the-meter power designer, for US$1.45 billion cash plus an earn-out, citing speed 'from site selection to first token'.
無法據以確立. Agreement announced, closing expected later; the time-to-power benefit is Vertiv's claim, not measured.
引用於: 第節 第4.4節, 第5.4節; 表 7

S-233 Open Compute Project Foundation AI Marketplace (vendor: Rittal) — Rittal Open Rack Wide (ORW) – 7844683 » Open Compute Project
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/ai-marketplace/products/860/rittal-open-rack-wide-orw-7844683
“The Rack supports higher powered Liquid cooled Busbars (700KW +) with an adjacent power rack.”
支持的研判. Rittal's OCP Marketplace listing describes a double-width ORW rack based on the ORv3 HPR feature set, supporting liquid-cooled busbars of 700 kW+ with an adjacent power rack.
無法據以確立. Vendor-supplied listing; it cites 'Base Specification v1.0.1' although the base PDF is v1.0.0, and shows no OCP Accepted mark.
引用於: 第節 第4.5節; 附錄A; 表 21

S-234 Open Compute Project Foundation (Rittal, Meta) — Open Rack Wide (ORW) Meta Design Specification V1.0.1
標準組織 · Global · 英文 · 5 May 2026 · 閱讀 26 September 2026
https://www.opencompute.org/documents/open-rack-wide-orw-meta-design-specification-v1-0-1-final-pdf
“Load rating of the frame SHALL be 4700 kg (~10,360 lbs) minimum with allowable support bracing and/or stiffening members included.”
支持的研判. Meta's ORW Design Specification V1.0.1 (effective 5 May 2026) sets a 1,200 mm nominal width, 1,219 mm depth and a frame load rating of at least 4,700 kg.
無法據以確立. Meta's implementation of the ORW base, not a requirement on other frames; it gives no busbar current or power rating.
引用於: 第節 第4.5節; 表 21

S-235 AMD Newsroom (GLOBE NEWSWIRE) — AMD Showcases 'Helios' Rack-Scale Platform Built on the Open Compute Project Open Rack for AI, Introduced by Meta
企業新聞稿 · U.S. · 英文 · 14 October 2025 · 閱讀 26 September 2026
https://newsroom.amd.com/news/amd-showcases-helios-rack-scale-platform-built-o
“showcased a static display of its "Helios," rack scale platform for the first time in public.”
支持的研判. AMD showed a static display of its Helios rack-scale platform for the first time in public on 14 October 2025, developed based on Meta's Open Rack Wide specification.
無法據以確立. A static exhibit announcement; not production, shipment or OCP acceptance of the Helios design.
引用於: 第節 第4.5節, 第6.2節, 第C.7節

S-236 AMD — AAI 2026: AMD Delivers Full-Stack Compute for the Agentic AI Era
企業新聞稿 · U.S. · 英文 · 23 July 2026 · 閱讀 27 September 2026
https://newsroom.amd.com/news/aai-2026-full-stack-compute-agentic-ai
“AMD launches its next-generation AI infrastructure and physical AI portfolio at Advancing AI 2026, led by AMD Helios rackscale solutions, now in production to be deployed by leading AI companies at gigawatt scale.”
“OpenAI expects to bring Helios online beginning in the fourth quarter of 2026, with deployments accelerating throughout 2027.”
支持的研判. AMD says Helios is 'now in production' for gigawatt-scale deployment, that OpenAI expects to bring it online from Q4 2026, and that MI500 GPUs come in 2027.
無法據以確立. Company and customer expectations; no Helios volumes or dates of customer acceptance, and MI500 timing is a roadmap.
引用於: 第節 第4.5節, 第9.2節; 展望 5; 表 19, 24

S-237 NVIDIA Newsroom — NVIDIA Contributes Blackwell Platform Design to Open Hardware Ecosystem, Accelerating AI Infrastructure Innovation
企業新聞稿 · U.S. · 英文 · 15 October 2024 · 閱讀 26 September 2026
https://nvidianews.nvidia.com/news/nvidia-contributes-blackwell-platform-design-to-open-hardware-ecosystem-accelerating-ai-infrastructure-innovation
“the rack architecture, compute and switch tray mechanicals, liquid-cooling and thermal environment specifications”
支持的研判. NVIDIA contributed elements of its GB200 NVL72 design to OCP on 15 October 2024, including rack architecture, tray mechanicals, and liquid-cooling and thermal environment specifications.
無法據以確立. A company design contribution; it does not establish an OCP-released versioned specification, OCP acceptance or multi-vendor products.
引用於: 第節 第4.5節, 第A.1節, 第C.7節; 附錄A; 表 21

S-238 Open Compute Project Foundation — Open Compute Project Foundation Expands Its Open Systems for AI Initiative
標準組織 · Global · 英文 · 15 October 2024 · 閱讀 26 September 2026
https://www.opencompute.org/blog/open-compute-project-foundation-expands-its-open-systems-for-ai-initiative
“NVIDIA's contributions to OCP helps ensure high compute density racks and compute trays from multiple vendors are interoperable in power, cooling and mechanical interfaces”
支持的研判. OCP recorded NVIDIA's MGX-based GB200 NVL72 rack contribution (19-inch EIA support, 1,400 A busbar), saying it helps multi-vendor racks and trays interoperate in power, cooling and mechanical interfaces.
無法據以確立. October 2024 announcement; it describes a contributed product design, not adoption by other vendors or a multi-vendor rack standard.
引用於: 第節 第4.5節, 第A.1節

S-239 Open Compute Project Foundation — MHS » Open Compute Project
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/community/mhs
“interoperability between key elements of data center, edge and enterprise infrastructure by providing consistent interfaces and form factors among modular building blocks”
支持的研判. OCP's MHS project seeks interoperability 'by providing consistent interfaces and form factors among modular building blocks' across data-centre, edge and enterprise infrastructure.
無法據以確立. Goal statement on an undated project page; it does not report specification status, release pace or adoption.
引用於: 第節 第4.5節

S-240 Open Compute Project Foundation — Server/MHS/DC-MHS-Specs-and-Designs - OpenCompute wiki
標準組織 · Global · 英文 · 24 April 2026 · 閱讀 26 September 2026
https://www.opencompute.org/w/index.php?title=Server%2FMHS%2FDC-MHS-Specs-and-Designs
“DC-MHS workstreams are still working the MHS R1 v1.x specifications.”
“The Final Spec Agreement (FSA) for the latest specs has not yet been signed.”
支持的研判. OCP's DC-MHS wiki (last modified 24 April 2026) says workstreams are still on the MHS R1 v1.x specifications and the Final Spec Agreement for the latest specs is unsigned.
無法據以確立. Wiki status as of April 2026; it does not explain the delay or show product adoption of DC-MHS modules.
引用於: 第節 第4.5節; 表 21

S-241 Open Compute Project Foundation — Server/NIC - OpenCompute wiki
標準組織 · Global · 英文 · 6 August 2026 · 閱讀 27 September 2026
https://www.opencompute.org/wiki/Server/NIC
取自來源表格之數值（轉錄）: OCP NIC 3.0 Specification | PDF | 1.6.0 Released | 03/31/2025
取自來源表格之數值（轉錄）: OCP NIC 3.0 Specification (Draft) | PDF | 1.6.1 Draft | 08/05/2026
支持的研判. OCP's NIC wiki lists OCP NIC 3.0 Specification 1.6.0 as the latest released version (31 March 2025) and a 1.6.1 draft dated 5 August 2026.
無法據以確立. Wiki status table (edited 6 August 2026); it does not mention liquid-cooled mock-ups or the 2026 Summit, which rest on S-220.
引用於: 第節 第4.5節

S-242 Open Compute Project Foundation — Server/MHS/DC-SCM-Specs-and-Designs - OpenCompute wiki; Datacenter Secure Control Module Specification (DC-SCM) Rev 2.2 Ver 1.0
標準組織 · Global · 英文 · 27 July 2022 · 閱讀 26 September 2026
https://www.opencompute.org/w/index.php?title=Server%2FMHS%2FDC-SCM-Specs-and-Designs
“Contributions to this 2.2 revision of the specification are made under the terms and conditions set forth in Open Web Foundation Contributor License Agreement ("052021 OWF CLA 1.0")”
支持的研判. OCP's DC-SCM wiki lists Rev 2.2 Ver 1.0 (submitted 16 October 2025), contributed under the OWF CLA 1.0 rather than the current OWF 0.9 pair.
無法據以確立. Wiki listing; it does not show multi-vendor interoperability of DC-SCM modules or the status of the 3.0 workstream.
引用於: 第節 第4.5節, 第6.4節; 表 21

S-243 Broadcom Investor Relations — OpenAI and Broadcom announce strategic collaboration to deploy 10 gigawatts of OpenAI-designed AI accelerators
企業新聞稿 · U.S. · 英文 · 13 October 2025 · 閱讀 27 September 2026
https://investors.broadcom.com/news-releases/news-release-details/openai-and-broadcom-announce-strategic-collaboration-deploy-10
“racks, scaled entirely with Ethernet and other connectivity solutions from Broadcom”
“Broadcom to deploy racks of AI accelerator and network systems targeted to start in the second half of 2026, to complete by end of 2029”
支持的研判. OpenAI and Broadcom plan 10 GW of OpenAI-designed accelerators in racks scaled entirely with Broadcom Ethernet and connectivity, targeted to start in 2H 2026 and complete by end-2029.
無法據以確立. A collaboration announcement with target dates; not a deployment record, and the 10 GW is a plan, not delivered capacity.
引用於: 第節 第4.5節, 第4.6節; 展望 10; 表 19, 24

S-244 NVIDIA — NVIDIA Unveils NVLink Fusion for Industry to Build Semi-Custom AI Infrastructure With NVIDIA Partner Ecosystem
企業新聞稿 · U.S. · 英文 · 18 May 2025 · 閱讀 26 September 2026
https://nvidianews.nvidia.com/news/nvidia-nvlink-fusion-semi-custom-ai-infrastructure-partner-ecosystem
“NVIDIA Unveils NVLink Fusion for Industry to Build Semi-Custom AI Infrastructure With NVIDIA Partner Ecosystem”
支持的研判. NVIDIA announced NVLink Fusion on 18 May 2025 so that industry can build semi-custom AI infrastructure with NVIDIA's partner ecosystem.
無法據以確立. NVIDIA's own programme; it is not a consortium or open specification, and names no deployed third-party NVLink systems.
引用於: 第節 第4.6節, 第C.7節; 圖 20

S-245 UALink Consortium — UALink Consortium Releases the Ultra Accelerator Link 200G 1.0 Specification
標準組織 · U.S. · 英文 · 8 April 2025 · 閱讀 26 September 2026
https://ualinkconsortium.org/wp-content/uploads/2025/04/UALink-1.0-Specification-PR_FINAL.pdf
“up to 1,024 accelerators within an AI computing pod”
“UALink is the only memory semantic solution for scale-up AI optimized for lower power, latency and cost while increasing effective bandwidth.”
支持的研判. The UALink Consortium's 8 April 2025 release announces the UALink 200G 1.0 specification for up to 1,024 accelerators within an AI computing pod, as a memory-semantic scale-up solution.
無法據以確立. Consortium announcement of a specification; 'the only memory semantic solution' is its own claim, and no silicon or deployment is shown.
引用於: 第節 第4.6節, 第6.3節, 第C.7節; 表 11, 22

S-246 Open Compute Project Foundation — OCP ESUN - Network Operator Requirements - Base Specification Rev 1.0
標準組織 · Global · 英文 · 9 February 2026 · 閱讀 27 September 2026
https://www.opencompute.org/documents/ocp-esun-network-operator-requirements-base-specification-rev-1-0-final-pdf
“allows use of L2 encapsulation to carry all the SU traffic with additional 4B and removal of IP headers (saving 20-40B)”
“ESUN proposes use of a feature, Link Level Retry, defined by Ultra Ethernet Consortium (UEC™) to achieve link resiliency”
支持的研判. ESUN's base specification (effective 9 February 2026) adopts UEC-defined Link Level Retry and CBFC, and allows L2 encapsulation adding 4 bytes while removing IP headers (saving 20–40 bytes).
無法據以確立. Operator requirements (Meta, Microsoft) under Modified OWFa 0.9; LLR and CBFC are required on switches, optional on hosts; it defines no endpoint transport.
引用於: 第節 第4.6節, 第6.3節, 第6.6節, 第6.7節, 第A.2節, 第C.7節; 圖 20; 表 11, 21

S-247 Open Compute Project Foundation (contributed by Broadcom) — Scale Up Ethernet (SUE) Specification 1.0.0
標準組織 · Global · 英文 · 5 September 2025 · 閱讀 26 September 2026
https://www.opencompute.org/documents/ocp-sue-spec-final-pdf-1
“This specification describes a transport over Ethernet for Load / Store and Single sided memory operations between XPUs.”
支持的研判. Broadcom's Scale Up Ethernet 1.0.0, approved 5 September 2025 under OWFa 0.9, describes an Ethernet transport for load/store and single-sided memory operations between XPUs.
無法據以確立. A single-company contribution; it is not a multi-party standard and does not show adoption beyond Broadcom or conformance testing.
引用於: 第節 第4.6節; 圖 20; 表 21

S-248 Open Compute Project Foundation — Introducing ESUN: Advancing Ethernet for Scale-Up AI Infrastructure at OCP
標準組織 · Global · 英文 · 13 October 2025 · 閱讀 26 September 2026
https://www.opencompute.org/blog/introducing-esun-advancing-ethernet-for-scale-up-ai-infrastructure-at-ocp
“ESUN focuses solely on open, standards-based Ethernet switching and framing for scale-up networking”
“AMD, Arista, ARM, Broadcom, Cisco, HPE Networking, Marvell, Meta, Microsoft, Nvidia, OpenAI and Oracle”
支持的研判. OCP launched ESUN on 13 October 2025 with 12 initial members, including AMD, Broadcom, Meta, Microsoft, NVIDIA and OpenAI, focused 'solely' on Ethernet switching and framing for scale-up.
無法據以確立. Launch announcement of intent; it excludes endpoint (host-side) stacks, publishes no specification, and its engagement with UEC and IEEE 802.3 is a stated intention.
引用於: 第節 第4.6節, 第6.1節; 圖 20; 表 11

S-249 Open Compute Project Foundation — Networking/SUE-T - OpenCompute wiki
標準組織 · Global · 英文 · 13 June 2026 · 閱讀 26 September 2026
https://www.opencompute.org/wiki/Networking/SUE-T
“SUE has been renamed and clarified now as SUE-T”
支持的研判. OCP's SUE-T wiki says SUE 'has been renamed and clarified now as SUE-T', an endpoint scale-up workstream whose participants sign a feedback agreement after kickoff.
無法據以確立. Wiki as of June 2026; it lists no SUE-T specification release or schedule.
引用於: 第節 第4.6節, 第6.2節, 第6.7節, 第C.7節; 圖 20; 表 21

S-250 UALink Consortium — UALink: An Open, High-Efficiency ScaleUp Interconnect for AI (white paper); UALink 200G 1.0 press release
標準組織 · U.S. · 英文 · 8 April 2025 · 閱讀 26 September 2026
https://ualinkconsortium.org/wp-content/uploads/2026/01/UALink_White_Paper_Publication_Candidate_FINAL_VERSION.pdf
“supports memory semantics (read, write, and atomic operations)”
支持的研判. UALink's white paper says UALink supports memory semantics (read, write, and atomic operations).
無法據以確立. The white paper was uploaded in January 2026, later than the 1.0 release it describes; it does not describe cache coherence or products.
引用於: 第節 第4.6節

S-251 UALink Consortium (Business Wire) — Ultra Accelerator Link (UALink) Consortium Publishes Four Specifications Defining In-Network Compute, Chiplets, Manageability and 200G Performance
標準組織 · Global · 英文 · 7 April 2026 · 閱讀 27 September 2026
https://www.businesswire.com/news/home/20260407620696/en
“encompasses three major additions – In-Network Compute, Chiplet Definition, and Manageability”
“Introduces UALink as a system with centralized control and management planes. Utilizes standardized protocols, modeling and APIs like gNMI, Yang, SAI and Redfish.”
支持的研判. UALink Consortium published four specifications on 7 April 2026: Common 2.0 with in-network compute, 200G DL/PL 2.0, Manageability 1.0 (gNMI, YANG, SAI, Redfish) and Chiplet 1.0 compliant with UCIe 3.0.
無法據以確立. Consortium announcement of ratified documents; not shipping silicon, interoperability testing or adoption by OCP.
引用於: 第節 第4.6節, 第A.2節, 第C.7節; 圖 20; 表 22

S-252 The Register (Simon Sharwood) — No-Nvidia interconnect club delivers 2.0 spec before v1.0 silicon ships
媒體 · U.K. · 英文 · 7 April 2026 · 閱讀 26 September 2026
https://www.theregister.com/2026/04/07/ualink_2_specs
“Chips for the group's 1.0 spec will reach labs in the second half of 2026, appear in 2027, and reach products later that year.”
支持的研判. The Register reported that chips for UALink's 1.0 specification would reach labs in the second half of 2026, appear in 2027 and reach products later that year.
無法據以確立. Trade-press report of expected timing, not vendor-committed dates or shipments.
引用於: 第節 第4.6節, 第C.7節

S-253 Credo Technology Group (company press release) — Credo Introduces Industry's First 224G Multiprotocol AI Scale-Up Retimer Supporting UALink, ESUN and Ethernet
企業新聞稿 · U.S. · 英文 · 29 January 2026 · 閱讀 27 September 2026
https://investors.credosemi.com/news-events/news/news-details/2026/Credo-Introduces-Industrys-First-224G-Multiprotocol-AI-Scale-Up-Retimer-Supporting-UALink-ESUN-and-Ethernet/default.aspx
“"The fast-growing scale-up market consists of a diversity of scale-up protocols such as UALink, ESUN and Ethernet which converge at layer 1 on the 224G 802.3 standard," said Sandeep Shah, AVP of Product for Credo's Retimer Products.”
“today announced its Blue Heron 224G AI scale-up retimer, optimized to support the growing need for extended cable and PCB backplane links using UALink, ESUN, and Ethernet protocols.”
“Blue Heron is sampling now with production quantities available in CQ3 2026.”
支持的研判. Credo announced on 29 January 2026 its Blue Heron 224G AI scale-up retimer for UALink, ESUN and Ethernet links, sampling then, with production quantities due in CQ3 2026.
無法據以確立. Company announcement; shows sampling of one multiprotocol part, not shipping UALink 200G systems, design wins or volumes.
引用於: 第節 第4.6節, 第4.7節, 第6.3節, 第A.2節

S-254 NVIDIA — NVIDIA Kicks Off the Next Generation of AI With Rubin — Six New Chips, One Incredible AI Supercomputer
企業新聞稿 · U.S. · 英文 · 5 January 2026 · 閱讀 26 September 2026
https://nvidianews.nvidia.com/news/rubin-platform-ai-supercomputer
“NVIDIA Rubin is in full production, and Rubin-based products will be available from partners the second half of 2026.”
支持的研判. NVIDIA said on 5 January 2026 that Rubin 'is in full production', with Rubin-based products available from partners in 2H 2026, across six chips including NVLink 6 and ConnectX-9.
無法據以確立. Vendor launch statement; partner products were expected in 2H 2026, and no volumes are given.
引用於: 第節 第4.6節; 圖 20

S-255 NVIDIA — NVIDIA Spectrum-X Ethernet Switches Speed Up Networks for Meta and Oracle
企業新聞稿 · U.S. · 英文 · 13 October 2025 · 閱讀 26 September 2026
https://nvidianews.nvidia.com/news/nvidia-spectrum-x-ethernet-switches-speed-up-networks-for-meta-and-oracle
“By integrating NVIDIA Spectrum Ethernet into the Minipack3N switch and FBOSS, we can extend our open networking approach.”
支持的研判. NVIDIA's release quotes Meta integrating NVIDIA Spectrum Ethernet into its Minipack3N switch and FBOSS to extend its open networking approach.
無法據以確立. Vendor release with a customer quote; no volumes, deployment dates or share of Meta's network are given.
引用於: 第節 第4.6節

S-256 Ultra Ethernet Consortium — Ultra Ethernet Consortium (UEC) Launches Specification 1.0 Transforming Ethernet for AI and HPC at Scale
標準組織 · U.S. · 英文 · 11 June 2025 · 閱讀 26 September 2026
https://ultraethernet.org/ultra-ethernet-consortium-uec-launches-specification-1-0-transforming-ethernet-for-ai-and-hpc-at-scale
“UEC Spec 1.0 ignites a new era of interoperable, high performance, Ethernet Innovation.”
支持的研判. UEC's 11 June 2025 release announces Ultra Ethernet Specification 1.0, saying it 'ignites a new era of interoperable, high performance, Ethernet Innovation'.
無法據以確立. The consortium's promotional release; the date is the release date, and the excerpt shows no product or interoperability test result.
引用於: 第節 第4.6節, 第6.3節, 第C.7節; 表 11, 22

S-257 Ultra Ethernet Consortium — Ultra Ethernet Specification History; UE Specification 1.0.3 Release Notes
標準組織 · U.S. · 英文 · 16 July 2026 · 閱讀 26 September 2026
https://ultraethernet.org/specification-history
“1.0.3 (July 16, 2026) - Current published version”
“200 Gb/s per lane signaling”
支持的研判. UEC's specification history lists version 1.0.3 of 16 July 2026 as the current published version, and its release notes add 200 Gb/s per lane signaling.
無法據以確立. A version history, not a product or conformance record; 'current' is as read in September 2026.
引用於: 第節 第4.6節, 第4.7節, 第6.3節, 第C.7節; 圖 20, 26; 表 11, 22; 展望 16

S-258 Ultra Ethernet Consortium — Ultra Ethernet Specification v1.0.3
標準組織 · U.S. · 英文 · 16 July 2026 · 閱讀 26 September 2026
https://ultraethernet.org/wp-content/uploads/sites/20/2026/08/UE-Specification-1.0.3.pdf
“Packets of the same PDC traversing different paths may arrive in a different order than how they were sent.”
支持的研判. Ultra Ethernet Specification 1.0.3 states that packets of the same PDC traversing different paths may arrive in a different order than how they were sent.
無法據以確立. Specification text on transport behaviour; it does not show deployment, performance or which products implement multipath spraying.
引用於: 第節 第4.6節, 第C.7節

S-259 OpenAI — Supercomputer networking to accelerate large scale AI training
企業新聞稿 · U.S. · 英文 · 5 May 2026 · 閱讀 26 September 2026
https://openai.com/index/mrc-supercomputer-networking
“Today, the MRC specification is available as an Open Compute Project (OCP) contribution for the community to use.”
支持的研判. OpenAI announced on 5 May 2026 the MRC specification, co-developed with AMD, Broadcom, Intel, Microsoft and NVIDIA, available as an OCP contribution and used on its largest GB200 clusters.
無法據以確立. Author's own account of deployment; an OCP contribution is not an adopted multi-vendor standard or evidence of use elsewhere.
引用於: 第節 第4.6節, 第A.1節, 第C.7節; 表 21, 24

S-260 OpenAI; Open Compute Project Foundation — Supercomputer networking to accelerate large scale AI training (OpenAI); Multipath Reliable Connection (MRC) Specification Rev 1.0 (OCP)
標準組織 · Global · 英文 · 5 May 2026 · 閱讀 27 September 2026
https://www.opencompute.org/documents/ocp-mrc-1-0-pdf
“MRC takes the packets from a single transfer and sprays them across hundreds of paths through our network”
“Multipath Reliable Connection (MRC) Specification Revision 1.0 Publication Date: 03/21/26”
“MRC extends RC transport with explicit multipath operation, path health monitoring, and failure recovery so that endpoints can sustain effective throughput in the presence of congestion, link failures, and other fabric impairments.”
支持的研判. OCP MRC Specification Revision 1.0 (publication date 21 March 2026), a joint AMD, Broadcom, Intel, Microsoft, NVIDIA and OpenAI contribution, extends RC transport with explicit multipath operation and failure recovery.
無法據以確立. The specification says nothing about deployment; OpenAI's claims that MRC runs on its largest GB200 clusters are in its announcement (S-259).
引用於: 第節 第4.6節, 第A.1節; 圖 20; 表 21

S-261 Open Compute Project Foundation (Hedgehog, Celestica) — Open Cluster Designs Aligned AI Inference Fabric Reference Architecture (v1.2)
標準組織 · Global · 英文 · 26 April 2026 · 閱讀 26 September 2026
https://www.opencompute.org/documents/open-cluster-designs-aligned-ai-inference-fabric-reference-architecture-pdf
“Dual-plane Network design using two isolated redundant fabrics (Plane A and Plane B)”
支持的研判. OCP's AI Inference Fabric Reference Architecture v1.2 (26 April 2026; Hedgehog, Celestica) packages a SONiC-based Ethernet design for 64–1,024 xPUs with a lossless back-end and two isolated redundant fabrics.
無法據以確立. A two-company reference architecture, not a specification; it does not report deployments or performance measurements.
引用於: 第節 第4.6節; 圖 20

S-262 Open Compute Project — Optical Circuit Switching for AI and Machine Learning Infrastructure (OCP OCS white paper)
標準組織 · Global · 英文 · April 2026 · 閱讀 26 September 2026
https://www.opencompute.org/documents/ocp-ocs-white-paper-april-2026-final-pdf
“While this document is a technical white paper rather than a hardware or software specification, it is intentionally structured to align with OCP's philosophy.”
支持的研判. OCP's April 2026 OCS white paper calls itself a technical white paper rather than a hardware or software specification; it attributes OCS adoption so far mainly to Google's deployment.
無法據以確立. Guidance, not a specification; Google's production OCS is proprietary and is not an implementation of this paper. Gives no vendor timelines.
引用於: 圖 20; 第節 第4.7節, 第C.7節; 表 6, 21; 附錄A

S-263 IEEE 802.3 Working Group — IEEE 802.3 Ethernet Working Group
標準組織 · Global · 英文 · 6 August 2026 · 閱讀 27 September 2026
https://www.ieee802.org/3/index.html
"IEEE P802.3dj 200 Gb/s, 400 Gb/s, 800 Gb/s, and 1.6 Tb/s Ethernet Task Force."
"Last Update: 6 August 2026"
支持的研判. The IEEE 802.3 Working Group page (updated 6 August 2026) lists IEEE P802.3dj, the 200 Gb/s to 1.6 Tb/s Ethernet Task Force, as an active project.
無法據以確立. Shows a task force still at work on the page's update date, not an approved standard or a completion date.
引用於: 圖 20; 第節 第4.7節; 表 22

S-264 OIF — Current Work
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.oiforum.com/technical-work/current-work
"Common Electrical I/O – 224G-VSR Common Electrical Chip-to-Module Interface Project"
"Common Electrical I/O – 448G-VSR Common Electrical Chip-to-Module Project"
支持的研判. OIF's Current Work page lists Common Electrical I/O 224G projects (Linear, XSR, VSR, MR, LR) and 448G projects (VSR, LR) among its active work.
無法據以確立. Lists projects in progress, not published implementation agreements; it gives no completion dates.
引用於: 圖 20; 第節 第4.7節, 第6.3節; 表 6, 11

S-265 IEEE 802.1 Working Group — P802.1Qdt - Priority-based Flow Control Enhancements
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://1.ieee802.org/tsn/802-1qdt
"places emphasis on the requirements for low latency and lossless transmission in large-scale and geographically dispersed data centers"
支持的研判. IEEE P802.1Qdt, Priority-based Flow Control Enhancements, is an IEEE 802.1 project aimed at low-latency, lossless transmission in large and geographically dispersed data centres.
無法據以確立. A project in draft (editor's draft stage per the page); not an approved standard. Its placement at the link layer is the report's reading.
引用於: 圖 20

S-266 SONiC Foundation / Linux Foundation — sonic-net/SONiC (GitHub)
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://github.com/sonic-net/SONiC
"SONiC is licensed under the Apache License 2.0."
支持的研判. The sonic-net/SONiC repository states that SONiC is licensed under the Apache License 2.0.
無法據以確立. Licence statement only; it does not describe SONiC's functions or deployment. Placing SONiC in the cluster-operations layer is the report's reading.
引用於: 圖 20

S-267 Ultra Ethernet Consortium — Ultra Ethernet Specification v1.0.2
標準組織 · U.S. · 英文 · 28 January 2026 · 閱讀 26 September 2026
https://ultraethernet.org/wp-content/uploads/sites/20/2026/01/UE-Specification-1.0.2-1.pdf
"Implementers of this Specification are solely responsible for obtaining any necessary licenses or permissions from the appropriate rights holders."
支持的研判. The Ultra Ethernet Specification v1.0.2 legal notice says implementers are solely responsible for obtaining any necessary licences or permissions from the rights holders; it grants no patent rights.
無法據以確立. Specification notice only; participants' RAND duties come from the charter (S-393). The CC BY-ND 4.0 licence is in the notice, not this excerpt.
引用於: 第節 第4.6節, 第6.6節, 第C.7節; 表 12

S-268 Sisvel — Sisvel | Ultra Ethernet (programme in development)
企業新聞稿 · Luxembourg · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.sisvel.com/licensing-programmes/programmes-in-development/ultra-ethernet
"at least one claim chart mapping a patent to the Ultra Ethernet specification, version 1.0.2"
支持的研判. Sisvel's programme page invites patent owners to submit at least one claim chart mapping a patent to Ultra Ethernet specification v1.0.2 for a pool in development.
無法據以確立. A call for submissions by a pool administrator; not an operating pool, licence terms, essentiality findings or royalties.
引用於: 第節 第4.6節, 第6.6節, 第C.7節

S-269 The Register — Nvidia embraces optical scale-up as copper reaches limits
媒體 · U.K. · 英文 · 5 April 2026 · 閱讀 26 September 2026
https://www.theregister.com/2026/04/05/nvidia_optical_scale_up
"We need a lot more capacity for copper; we need a lot more capacity for optics; we need a lot more capacity for CPO."
支持的研判. The Register reports NVIDIA's CEO saying 'we need a lot more capacity' for copper, optics and CPO, and its networking head describing copper as cheap but short-reach.
無法據以確立. Press account of conference remarks; the excerpt does not name its speaker, and no capacity figures are given.
引用於: 第節 第4.7節

S-270 IEEE 802.3 Ethernet Working Group — IEEE 802.3 400 Gb/s/Lane Signaling Study Group
標準組織 · Global · 英文 · 3 August 2026 · 閱讀 26 September 2026
https://grouper.ieee.org/groups/802/3/400GPL/index.html
"400 Gb/s per lane signaling for electrical interconnects as well as single-mode fiber optical interconnects with reaches up to 500 meters"
支持的研判. The IEEE 802.3 400 Gb/s/Lane Signaling Study Group covers 400 Gb/s-per-lane electrical interconnects and single-mode fibre optical interconnects with reaches up to 500 metres.
無法據以確立. Study group, not a task force; the 500 m reach applies to fibre. Charter and PAR dates are from the page, not this excerpt.
引用於: 第節 第4.7節, 第6.3節; 表 6, 11, 22

S-271 Converge Digest — OIF Brings 448G, 800ZR, CMIS and AI Interoperability to ECOC 2026
媒體 · U.S. · 英文 · 26 August 2026 · 閱讀 26 September 2026
https://convergedigest.com/oif-448g-800zr-cmis-ai-interoperability-ecoc-2026
"OIF began its CEI-448G Framework work in 2024, and its 2026 demonstrations move that effort toward multi-vendor physical validation."
支持的研判. Converge Digest reports OIF began CEI-448G framework work in 2024 and that its 2026 (ECOC) demonstrations move the effort toward multi-vendor physical validation.
無法據以確立. Trade-press preview; demonstrations are not an implementation agreement, and no completion date for CEI-448G is given.
引用於: 第節 第4.7節; 表 22

S-272 Linear Pluggable Optics MSA — LPO MSA FAQs
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.lpo-msa.org/home/faqs.html
"lower power, lower cost, and lower latency than competing solutions"
支持的研判. The Linear Pluggable Optics MSA's FAQ claims LPO offers lower power, lower cost, and lower latency than competing solutions.
無法據以確立. Qualitative claims by the MSA itself; the FAQ gives no power, cost or latency figures and no independent measurement.
引用於: 第節 第4.7節, 第C.7節; 表 6

S-273 IEEE Electronics Packaging Society, Photonics Technical Committee (authors: Sujit Ramachandra, Celestial AI; Farnood Rezaie, Cisco Systems) — Linear Pluggable Optics – An Overview
研究 · Global · 英文 · 未註明日期 · 閱讀 27 September 2026
https://eps.ieee.org/wp-content/uploads/2026/03/Linear-Pluggable-Optics_V2-UPDATED.pdf
"Similar to LPO, LRO systems (Fig. 3) eliminate the DSP on the receiver but retain it in the transmit path to meet integrity standards (IEEE 802.3)."
"At 200G/lane or 1.6Tbps, LRO is expected to be a better choice since the initial solutions require > 30W of power, presenting challenges to efficiently managing the thermals."
支持的研判. An IEEE EPS technical-committee overview says LRO drops only the receive-side DSP and is expected to be the better choice at 200G/lane, where initial 1.6T solutions exceed 30 W.
無法據以確立. Authors' views in an undated overview (March 2026 upload), not a standard or measured data; the 30 W figure is a general estimate.
引用於: 第節 第4.7節; 表 6

S-274 Coherent Corp. — Coherent Demonstrates 1.6T Optical Transceivers Based on 200G VCSELs
企業新聞稿 · U.S. · 英文 · 1 April 2025 · 閱讀 26 September 2026
https://www.coherent.com/news/press-releases/optical-transceivers-based-on-200g-vcsels
"We are proactively demonstrating this 200G-per-lane multimode technology even as the standards are still being written"
支持的研判. Coherent demonstrated a 1.6T multimode transceiver on 200G-per-lane VCSELs at OFC 2025, 'even as the standards are still being written'.
無法據以確立. A trade-show demonstration; not a shipping product or a completed IEEE standard.
引用於: 第節 第4.7節

S-275 Coherent Corp. — Coherent to Unveil Breakthrough AI-Scale Optical Innovations and Industry Leadership at OFC 2026
企業新聞稿 · U.S. · 英文 · 17 March 2026 · 閱讀 26 September 2026
https://www.coherent.com/news/press-releases/coherent-ai-scale-optical-innovations-ofc-2026
"400G/lane, 3.2T transceivers and emerging architectures for 12.8T"
支持的研判. Coherent announced (17 March 2026) it would show 400G-per-lane 3.2T transceivers and emerging architectures for 12.8T at OFC 2026.
無法據以確立. Pre-show announcement; names no customers, sampling dates or production timeline.
引用於: 第節 第4.7節; 表 6, 30

S-276 Cignal AI — Optical Component Revenue Reaches Nearly $25B in 2025
市場研究 · U.S. · 英文 · 7 January 2026 · 閱讀 27 September 2026
https://cignal.ai/2026/01/optical-component-revenue-reaches-nearly-25b-in-2025
"Innolight remains the leader in 800GbE module shipments but the gap is narrowing with Eoptolink, Coherent, and Nvidia. Marvell and Acacia remain the leaders in 400ZRx"
"1.6TbE forecast will exceed 5 million in 2026."
支持的研判. Cignal AI forecasts 1.6TbE module shipments above 5 million in 2026 and says Innolight leads 800GbE shipments, with Eoptolink, Coherent and Nvidia narrowing the gap; Marvell and Acacia lead 400ZRx.
無法據以確立. Market-research summary; rankings are by shipments, not revenue, and the page gives no individual market shares.
引用於: 第節 第4.7節, 第C.7節; 表 6, 30

S-277 TrendForce — China Accounts for 56% of Global Optical Module Manufacturing; Short-Term Supply Chain Decoupling Unlikely Under Potential U.S. Restrictions, Says TrendForce
市場研究 · Taiwan · 英文 · 5 August 2026 · 閱讀 26 September 2026
https://www.trendforce.com/presscenter/news/20260805-13169.html
"Chinese optical module makers will account for approximately 56% of global contract manufacturing capacity in 2026"
支持的研判. TrendForce estimates Chinese optical-module makers will account for about 56% of global optical-module contract manufacturing capacity in 2026.
無法據以確立. A share of contract manufacturing capacity, not of module shipments, revenue or laser supply; a 2026 estimate.
引用於: 第節 第4.7節, 第B.2節

S-278 TrendForce — NVIDIA Compute Architecture Paves the Way for Scale-Up Optical Interconnects; CPO Penetration in AI Data Centers Expected to Rise Steadily, Says TrendForce
市場研究 · Taiwan · 英文 · 11 March 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260311-12962.html
"It is estimated that CPOs will account for only about 0.5% of optical transceiver modules used in AI data centers in 2026."
"TrendForce forecasts that silicon-photonics-based CPO solutions could reach approximately 35% penetration of AI data centers by around 2030."
支持的研判. TrendForce puts CPO at about 0.5% of optical transceiver modules in AI data centres in 2026 and says CPO could reach about 35% penetration of AI data centres around 2030.
無法據以確立. Two different measures: the 2026 share is of optical modules, the 2030 figure is penetration of AI data centres; 'could' marks a scenario.
引用於: 第節 第4.7節; 展望 14

S-279 NVIDIA Newsroom — NVIDIA Announces Spectrum-X Photonics, Co-Packaged Optics Networking Switches to Scale AI Factories to Millions of GPUs
企業新聞稿 · U.S. · 英文 · 18 March 2025 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-spectrum-x-co-packaged-optics-networking-switches-ai-factories
"They integrate optics innovations with 4x fewer lasers to deliver 3.5x more power efficiency"
支持的研判. NVIDIA announced Spectrum-X Photonics and Quantum-X Photonics CPO switches on 18 March 2025, saying they integrate optics innovations with 4x fewer lasers.
無法據以確立. Vendor launch claims; availability was then expected later in 2025 (Quantum-X) and in 2026 (Spectrum-X), not shipped.
引用於: 第節 第4.7節

S-280 NVIDIA — NVIDIA Silicon Photonics Networking; NVIDIA Announces Spectrum-X Photonics, Co-Packaged Optics Networking Switches (2025-03-18)
企業新聞稿 · U.S. · 英文 · 18 March 2025 · 閱讀 26 September 2026
https://www.nvidia.com/en-us/networking/products/silicon-photonics
"Available in the second half of 2026"
"NVIDIA CPO technology significantly reduces network power, enabling 5x better power efficiency over pluggable transceivers"
支持的研判. NVIDIA's silicon photonics page lists Spectrum-X Photonics as 'Available in the second half of 2026' and claims '5x better power efficiency over pluggable transceivers'.
無法據以確立. Vendor product page; the efficiency figure is NVIDIA's relative claim (3.5x in March 2025), and availability is not shipment.
引用於: 第節 第4.7節

S-281 Lambda — Unbox one of NVIDIA's first co-packaged optics switches with us
企業新聞稿 · U.S. · 英文 · 1 June 2026 · 閱讀 26 September 2026
https://lambda.ai/blog/unbox-one-of-nvidias-first-co-packaged-optics-samples-with-lambda
"Engineering samples are a chance to work through rack design, cooling, power delivery, fiber routing, and the installation process with the vendor while the product is still becoming real."
支持的研判. Lambda describes receiving NVIDIA co-packaged optics switch engineering samples in June 2026 to work through rack, cooling, power, fibre and installation 'while the product is still becoming real'.
無法據以確立. One customer's account of pre-production samples; not NVIDIA shipment volumes or production timing.
引用於: 第節 第4.7節; 表 6

S-282 TrendForce — NVIDIA and Broadcom Begin Volume Ramp of CPO Switches, with Optical Engine Yield and Advanced Packaging Capacity Emerging as Key Expansion Bottlenecks, Says TrendForce
市場研究 · Taiwan · 英文 · 27 July 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260727-13151.html
"NVIDIA has begun shipping its next-generation Spectrum-X CPO switch to select partners"
"Broadcom's 51.2T Bailly CPO switch has entered volume manufacturing with support from Delta Electronics and Micas Networks."
支持的研判. TrendForce says Broadcom's Bailly CPO switch is in volume manufacturing and NVIDIA ships Spectrum-X CPO to select partners; volume ramp comes 2027–2028, paced by optical-engine, silicon-photonics and packaging supply.
無法據以確立. Market-research view; 'volume manufacturing' is not a shipment count, and the yield bottleneck is named in the headline rather than quantified.
引用於: 第節 第4.7節, 第5.4節; 表 6

S-283 ServeTheHome — Broadcom Tomahawk 6 - Davisson 102.4T Switch with Co-Packaged Optics Shipping
媒體 · U.S. · 英文 · 12 October 2025 · 閱讀 26 September 2026
https://www.servethehome.com/broadcom-tomahawk-6-davisson-102-4t-switch-with-co-packaged-optics-shipping
"light sources are usually the part with the highest failure rate, so making them FRUs means switches can be serviced in the field"
支持的研判. ServeTheHome reported Broadcom's 102.4T Tomahawk 6 – Davisson CPO switch shipping (October 2025) and explained its field-replaceable light sources: lasers are usually the highest-failure-rate part.
無法據以確立. Trade-press report; the failure-rate point is ServeTheHome's explanation, not Broadcom-published reliability data.
引用於: 第節 第4.7節; 表 6; 展望 14

S-284 Optical Internetworking Forum (OIF) — Implementation Agreement for a 3.2 Tb/s Co-Packaged (CPO) Module, OIF-Co-Packaging-3.2T-Module-01.0
標準組織 · U.S. · 英文 · 29 March 2023 · 閱讀 27 September 2026
https://www.oiforum.com/wp-content/uploads/OIF-Co-Packaging-3.2T-Module-01.0.pdf
"This document defines the technical specifications for a 3.2 Tb/s Co-packaged Optical (CPO) transceiver module"
"The 3.2 Tb/s transceiver is a building block for a 51.2 Tb/s switch assembly."
支持的研判. OIF's implementation agreement of 29 March 2023 defines a 3.2 Tb/s co-packaged optical transceiver module as a building block for a 51.2 Tb/s switch assembly.
無法據以確立. A 2023 agreement written for 51.2 Tb/s switches; it does not address 100 Tb/s-class switches or show vendor adoption.
引用於: 第節 第4.7節, 第6.3節, 第6.7節; 表 6, 11, 13, 22; 圖 26; 展望 14

S-285 Tech Times — Co-Packaged Optics Standard Gets Q4 Deadline: OCP Blueprint Targets Wasted AI Compute
媒體 · U.S. · 英文 · 15 August 2026 · 閱讀 27 September 2026
https://www.techtimes.com/articles/324563/20260815/co-packaged-optics-standard-gets-q4-deadline-ocp-blueprint-targets-wasted-ai-compute.htm
"The coalition's stated next step is submission of first specifications, expected in Q4 2026."
"An industry coalition of 20 companies formalized its effort to end copper's reign over AI data center interconnects on August 13, publishing a 300-page architectural blueprint within the Open Compute Project"
支持的研判. Tech Times reports a 20-company coalition published a 300-page CPO blueprint within OCP on 13 August 2026 and expects to submit its first specifications in Q4 2026.
無法據以確立. Single press report; Q4 2026 is for submitting first specifications, not for a published or released OCP specification.
引用於: 第節 第4.7節, 第6.7節, 第9.2節, 第A.1節; 展望 4; 圖 32; 表 20, 21

S-286 Tech Times — Nvidia Ships Co-Packaged Optics Switch as Ecosystem Testing Gap Widens
媒體 · U.S. · 英文 · 12 August 2026 · 閱讀 27 September 2026
https://www.techtimes.com/articles/324160/20260812/nvidia-ships-co-packaged-optics-switch-ecosystem-testing-gap-widens.htm
"Co-packaged optics has become the inflection point for AI infrastructure, she said, but the industry has not yet built the shared testing standards and simulation groundwork the technology needs."
"He also named the three challenges the industry has not yet collectively solved: wafer-level testing protocols, fiber array unit integration, and high-speed optical packaging assembly."
支持的研判. Tech Times reports ASE saying CPO lacks shared testing standards and simulation groundwork, and TSMC naming wafer-level test, fibre-array integration and assembly as unsolved; it calls procurement effectively single-source.
無法據以確立. Press account of summit remarks; single-sourcing is the article's reported consequence, not a measured market outcome.
引用於: 第節 第4.7節, 第6.7節, 第A.1節; 展望 4; 表 20

S-287 Open Compute Project Foundation — Co-Located Events - 2026 OCP Global Summit
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/summit/global-summit/co-located-events
"224Gbps and 448Gbps electrical channel specifications, next-generation optical channel specifications, CPO and NPO based architectures"
"AI, Redfish, SPDM, libspdm, PLDM, and more"
支持的研判. OCP lists co-located DMTF (Redfish, SPDM, PLDM) and OIF (224/448 Gbps electrical, optical channels, CPO/NPO) workshops on 13 October 2026 at the San Jose Convention Center.
無法據以確立. An agenda listing; workshops are not specification releases and do not show OCP adoption of DMTF or OIF documents.
引用於: 第節 第4.7節, 第6.2節; 表 11

S-288 NVIDIA Corporation (NVIDIA Newsroom) — NVIDIA Announces Strategic Partnership With Lumentum to Develop State-of-the-Art Optics Technology
企業新聞稿 · U.S. · 英文 · 2 March 2026 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-announces-strategic-partnership-with-lumentum-to-develop-state-of-the-art-optics-technology
"NVIDIA is investing $2 billion in Lumentum to support R&D, future capacity and operations as the company builds out its U.S.-based manufacturing capabilities in a new fab."
"The nonexclusive agreement includes an NVIDIA multibillion purchase commitment and future capacity access rights for advanced laser components."
支持的研判. NVIDIA says it is investing US$2 billion in Lumentum and that their nonexclusive agreement includes a multibillion purchase commitment and future capacity access rights for advanced laser components.
無法據以確立. Company announcement of an agreement; it gives no volumes, prices, lead times or allocation to other buyers, and does not measure demand.
引用於: 第節 第4.7節

S-289 NVIDIA Corporation (NVIDIA Newsroom) — NVIDIA and Coherent Announce Strategic Partnership to Develop Optics Technology to Scale Next-Generation Data Center Architecture
企業新聞稿 · U.S. · 英文 · 2 March 2026 · 閱讀 27 September 2026
https://nvidianews.nvidia.com/news/nvidia-and-coherent-announce-strategic-partnership-to-develop-optics-technology-to-scale-next-generation-data-center-architecture
"The nonexclusive agreement includes an NVIDIA multibillion-dollar purchase commitment and future access and capacity rights for advanced laser and optical networking products."
"NVIDIA is investing $2 billion in Coherent to support research and development, future capacity and operations"
支持的研判. NVIDIA says it is investing US$2 billion in Coherent; the nonexclusive agreement includes a multibillion-dollar purchase commitment and future capacity rights for advanced laser and optical networking products.
無法據以確立. Company announcement of an agreement; it gives no volumes, prices, lead times or allocation to other buyers.
引用於: 第節 第4.7節

S-290 Tech Times — AI Data Center Optical Component Shortage: Nvidia's $4B Laser Lockup Pushes Rivals Past 2027
媒體 · U.S. · 英文 · 27 May 2026 · 閱讀 26 September 2026
https://www.techtimes.com/articles/317281/20260527/ai-data-center-optical-component-shortage-nvidias-4b-laser-lockup-pushes-rivals-past-2027.htm
"lead times at key EML suppliers have stretched beyond 2027 for any buyer who is not Nvidia"
支持的研判. Tech Times reported in May 2026 that lead times at key EML suppliers had stretched beyond 2027 for any buyer other than NVIDIA.
無法據以確立. Single trade-press report; lead times are not quantified by supplier, and NVIDIA's allocation terms are not public.
引用於: 第節 第4.7節

S-291 Tom's Hardware — Lumentum CEO warns of impending bottleneck on critical material used for silicon photonics
媒體 · U.S. · 英文 · 31 July 2026 · 閱讀 27 September 2026
https://www.tomshardware.com/tech-industry/semiconductors/lumentum-ceo-says-the-indium-phosphide-shortage-will-become-worse-than-memory
"On the May earnings call, Hurlston told analysts the company expects its supply line to increase 50% measured from one December quarter to the next, and in the same breath said the supply-demand imbalance on EMLs had widened from the 25% to 30% given a quarter earlier to "somewhere greater than 30%," with pump lasers tighter still."
"While Lumentum runs five indium phosphide fabs, it's still shipping more than 30% below what customers want."
支持的研判. Tom's Hardware reports Lumentum shipping more than 30% below customer demand despite five InP fabs, citing the CEO's May call remark that the EML imbalance exceeded 30%.
無法據以確立. The 30% figure dates from Lumentum's May 2026 earnings call; the July remarks reported (RAISE Summit) concerned InP versus DRAM and NAND.
引用於: 第節 第4.7節, 第5.4節, 第5.8節

S-292 Tower Semiconductor (investor relations press release) — Tower Semiconductor Signs Customer Contracts for $1.3 Billion Silicon Photonics Revenue for 2027
企業新聞稿 · Israel · 英文 · 13 May 2026 · 閱讀 27 September 2026
https://ir.towersemi.com/news-releases/news-release-details/tower-semiconductor-signs-customer-contracts-13-billion-silicon
"a rich and broad SiPho customer base of over 50 active customers"
"today announced it has signed Silicon Photonics (SiPho) contracts for $1.3 billion for 2027 revenue with its largest customers, and the receipt of $290 million in customers' prepayments for capacity reservation."
支持的研判. Tower says it signed US$1.3 billion of silicon-photonics contracts for 2027 revenue, received US$290 million of customer prepayments for capacity reservation, and has over 50 active SiPho customers.
無法據以確立. Company announcement; 2027 revenue is contracted, not earned; does not link the prepayments to specific contracts or name customers.
引用於: 第節 第4.7節, 第B.2節, 第C.7節; 表 30

S-293 Norman P. Jouppi, George Kurian, Sheng Li, Peter Ma, Rahul Nagarajan, Lifeng Nai, Nishant Patil, Suvinay Subramanian, Andy Swing, Brian Towles, Cliff Young, Xiang Zhou, Zongwei Zhou, David Patterson — TPU v4: An Optically Reconfigurable Supercomputer for Machine Learning with Hardware Support for Embeddings
研究 · U.S. · 英文 · April 2023 · 閱讀 26 September 2026
https://arxiv.org/abs/2304.01433
"Deployed since 2020, TPU v4 outperforms TPU v3 by 2.1x and improves performance/Watt by 2.7x."
支持的研判. Google's TPU v4 paper reports the optically reconfigurable supercomputer 'Deployed since 2020' and outperforming TPU v3 by 2.1x with 2.7x better performance per watt.
無法據以確立. Google's own system and figures; system-level gains are not attributed to optical switching alone, and no merchant or open product is implied.
引用於: 第節 第4.7節, 第6.7節, 第C.7節; 表 6, 21, 24; 圖 26

S-294 Open Compute Project Foundation — The Open Compute Project Announces New Optical Circuit Switching (OCS) Project
標準組織 · Global · 英文 · 21 July 2025 · 閱讀 26 September 2026
https://www.opencompute.org/blog/the-open-compute-project-announces-new-optical-circuit-switching-ocs-project
"OCS leverages photonic technology to route data optically, reducing power consumption and enhancing reliability for large-scale AI workloads."
支持的研判. OCP announced a new Optical Circuit Switching (OCS) project on 21 July 2025, describing OCS as routing data optically to cut power and improve reliability for AI workloads.
無法據以確立. Project-formation announcement; it establishes no specification, deployment or performance data.
引用於: 第節 第4.7節

S-295 Lumentum — 300x300 Optical Circuit Switch (OCS) | Lumentum
企業新聞稿 · U.S. · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.lumentum.com/en/products/300x300-optical-circuit-switch-ocs
"300x300 Optical Circuit Switch (OCS)"
"The Lumentum R300 Optical Circuit Switch (OCS) brings a new level of scalability, latency, and power efficiency to cloud-scale and AI/ML data center networks."
支持的研判. Lumentum's product page markets the R300, a 300×300 optical circuit switch, for cloud-scale and AI/ML data-centre networks.
無法據以確立. Undated marketing page; says nothing about the OFC 2026 demonstration, shipments, customers or deployment scale.
引用於: 第節 第4.7節; 表 6

S-296 Marvell Technology (company press release) — Marvell and Lumentum to Demonstrate Optical Circuit Switching for Next-generation AI Scale-up Infrastructure
企業新聞稿 · U.S. · 英文 · 16 March 2026 · 閱讀 27 September 2026
https://investor.marvell.com/news-events/press-releases/detail/1015/marvell-and-lumentum-to-demonstrate-optical-circuit-switching-for-next-generation-ai-scale-up-infrastructure
"By demonstrating Marvell optical connectivity solutions interoperating with Lumentum's OCS platform, we're showcasing how tomorrow's AI networks can achieve massive gains while delivering breakthrough gains in performance, power efficiency and architectural agility."
"The integrated, rack-level system will be showcased in the Marvell booth (#1600) at OFC 2026, taking place March 17-19 at the Los Angeles Convention Center."
支持的研判. Marvell said on 16 March 2026 that a rack-level system with Marvell optics interoperating with Lumentum's R300 OCS would be showcased at OFC 2026, 17–19 March, Los Angeles.
無法據以確立. Pre-event announcement; does not show the demonstration took place, and gives no port count, availability or customers.
引用於: 第節 第4.7節; 表 6, 30

S-297 Ayar Labs — Ayar Labs closes $500M Series E, accelerates volume production of co-packaged optics
企業新聞稿 · U.S. · 英文 · March 2026 · 閱讀 26 September 2026
https://ayarlabs.com/news/ayar-labs-closes-500m-series-e-accelerates-volume-production-of-co-packaged-optics
"Ayar Labs Closes $500M Series E, Accelerates Volume Production of Co-Packaged Optics"
支持的研判. Ayar Labs says it closed a US$500 million Series E in March 2026 to accelerate volume production of co-packaged optics.
無法據以確立. Funding announcement; does not establish volume production, customers, shipments or revenue.
引用於: 表 6, 30

S-298 Marvell Technology (company press release) — Marvell Completes Acquisition of Celestial AI
企業新聞稿 · U.S. · 英文 · 2 February 2026 · 閱讀 27 September 2026
https://www.marvell.com/company/newsroom/marvell-completes-acquisition-of-celestial-ai.html
"Marvell expects initial revenue contributions from Celestial AI to begin in the second half of fiscal 2028, with revenue ramping meaningfully in the fourth quarter to a $500 million annualized run rate."
支持的研判. Marvell says it completed the Celestial AI acquisition (2 February 2026) and expects Celestial AI revenue from 2H fiscal 2028, ramping in Q4 to a US$500 million annualized run rate.
無法據以確立. Company expectation; a run rate is not realised revenue, and Marvell's fiscal 2028 ends in early 2028.
引用於: 表 6, 30

S-299 Eaton Corporation plc — Eaton Reports Record Second Quarter 2026 Results, with Strong Organic Growth, Accelerating Orders and Backlog, and Raises Organic Growth Guidance
企業新聞稿 · Ireland / U.S. · 英文 · 31 July 2026 · 閱讀 27 September 2026
https://www.eaton.com/us/en-us/company/news-insights/news-releases/2026/eaton-reports-record-second-quarter-2026-results.html
直接文件: https://www.sec.gov/Archives/edgar/data/1551182/000155118226000027/etn06302026exhibit99.htm
"Strong year-over-year total backlog growth of 43% in Electrical sector and 28% in Aerospace segment"
"On a rolling twelve-month basis, the book-to-bill ratio for the Electrical businesses remained strong at 1.2."
支持的研判. Eaton reports Electrical sector backlog up 43% year on year and an Electrical book-to-bill of 1.2 (rolling twelve months) at June 2026, with Electrical Global backlog up 103%.
無法據以確立. Company results; no data-centre order figure, and Electrical Global growth is not split between organic and Boyd Thermal.
引用於: 表 7, 10, 28; 第節 第5.4節

S-300 Siemens Energy AG — Earnings Release Q3 FY 2026
企業新聞稿 · Germany · 英文 · 5 August 2026 · 閱讀 27 September 2026
https://assets.siemens-energy.com/dam/30018eed-275d-4e73-811e-b49d004cdc3d/2026-08-05-Earnings_Release_Q3_FY26_EN-pdf_Original%20file.pdf
"Book-to-bill ratio (ratio of orders to revenue) was 1.57, while order backlog increased to €162bn at quarter-end."
"Demand from the U.S., including large orders related to data centers, and orders from the Middle East and Asia for new power plants were the key drivers of the development."
支持的研判. Siemens Energy reports record Q3 FY2026 orders, a €162 billion order backlog at quarter-end with book-to-bill of 1.57, and U.S. demand including large data-centre orders among key drivers.
無法據以確立. Quarter to 30 June 2026, published in August; company-wide figures in which data-centre demand is a named driver, not a measured share.
引用於: 表 7, 29; 第節 第5.5節, 第5.7節; 圖 25

S-301 Siemens Energy AG — Profitable growth accelerated – Siemens Gamesa returning to profitability (Analyst presentation Q3 FY26)
企業新聞稿 · Germany · 英文 · 5 August 2026 · 閱讀 27 September 2026
https://assets.siemens-energy.com/dam/3e846440-66dd-4a56-be75-b49d004e6741/2026-08-05_Q3_Analyst_presentation-pdf_Original%20file.pdf
"High level of firm customer commitments (69 GW backlog / 26 GW SRAs)"
"15 GW booked (12 GW converted from SRAs2), 6 GW shipped"
支持的研判. Siemens Energy's Q3 FY2026 presentation reports gas-turbine firm commitments of 69 GW backlog and 26 GW slot reservation agreements, with 15 GW booked in Q3, 12 GW converted from SRAs.
無法據以確立. Slide figures at 30 June 2026; SRAs are reservations rather than firm orders; GW are equipment ratings, not generation.
引用於: 表 7, 13, 29; 第節 第5.5節, 第9.3節, 第C.7節; 展望 8

S-302 Wiwynn Corporation — Wiwynn Reports First Quarter 2025 Financial Results
企業新聞稿 · Taiwan · 英文 · 8 May 2025 · 閱讀 27 September 2026
https://www.wiwynn.com/news/wiwynn-reports-first-quarter-2025-financial-results
"Following the Board of Directors meeting today (May 8), Wiwynn Corporation announced consolidated revenue was NT$170.655 billion, an increase of 145.1% year-over-year."
"Year-over-year, gross margin decreased by 2.4 ppt to 8.7%, operating margin decreased by 1.4 ppt to 7.0%, and PAT margin decreased by 1.1 ppt to 5.7%."
支持的研判. Wiwynn reports Q1 2025 revenue of NT$170.655 billion, up 145.1% year on year, and an 8.7% gross margin.
無法據以確立. Before the April 2026 agency-model change, so not like-for-like with quarters from April 2026.
引用於: 圖 22

S-303 Wiwynn Corporation — Wiwynn Reports Second Quarter 2025 Financial Results
企業新聞稿 · Taiwan · 英文 · 8 August 2025 · 閱讀 27 September 2026
https://www.wiwynn.com/news/wiwynn-reports-second-quarter-2025-financial-results
"In the second quarter ended June 30, 2025, consolidated revenue reached NT$220.744 billion, an increase of 184.9% year-over-year."
"Driven primarily by strong demand for general and AI servers, hardware product revenue grew significantly faster than product design and development income (one-time NRE engineering fees), resulting in a gross margin of 8.6%."
支持的研判. Wiwynn reports Q2 2025 revenue of NT$220.744 billion, up 184.9% year on year, and an 8.6% gross margin.
無法據以確立. Before the April 2026 agency-model change, so not like-for-like with quarters from April 2026.
引用於: 圖 22

S-304 Wiwynn Corporation — Wiwynn Reports Third Quarter 2025 Financial Results
企業新聞稿 · Taiwan · 英文 · 7 November 2025 · 閱讀 27 September 2026
https://www.wiwynn.com/news/wiwynn-reports-third-quarter-2025-financial-results
"In the third quarter ended September, 2025, consolidated revenue achieved NT$266.824 billion, increased by 172.8% year-over-year."
"Benefiting from economies of scale and improved operational efficiency, gross margin was 8.8%, operating margin was 7.3%, and PAT margin was 5.8%."
支持的研判. Wiwynn reports Q3 2025 revenue of NT$266.824 billion, up 172.8% year on year, and an 8.8% gross margin.
無法據以確立. Before the April 2026 agency-model change, so not like-for-like with quarters from April 2026.
引用於: 圖 22

S-305 Wiwynn Corporation — Wiwynn Reports Fourth Quarter 2025 Unaudited Financial Results
企業新聞稿 · Taiwan · 英文 · 8 January 2026 · 閱讀 27 September 2026
https://www.wiwynn.com/news/wiwynn-reports-fourth-quarter-2025-unaudited-financial-results
"In the fourth quarter ended December 31, 2025, consolidated revenue achieved NT$292.440 billion, increased by 152.9% year-over-year."
"Due to changes in the product portfolio, the gross margin declined to 7.2%, operating margin was 5.6%, and PAT margin was 4.7%."
支持的研判. Wiwynn reports unaudited Q4 2025 revenue of NT$292.440 billion, up 152.9% year on year, and a 7.2% gross margin.
無法據以確立. Unaudited; before the April 2026 agency-model change, so not like-for-like with later quarters.
引用於: 圖 22

S-306 Wiwynn — Amazon and Wiwynn announce expansion of advanced manufacturing facility in Texas, creating nearly 1,000 jobs
企業新聞稿 · Taiwan · 英文 · 10 September 2026 · 閱讀 26 September 2026
https://www.wiwynn.com/news/amazon-wiwynn-data-center-manufacturing-jobs-texas
"The facility brings production of custom hardware to the U.S. and strengthens domestic manufacturing capacity for Amazon's infrastructure."
支持的研判. Amazon and Wiwynn announced (10 September 2026) an expansion of over US$1.6 billion at Wiwynn's Socorro, Texas plant, bringing custom hardware production for Amazon's infrastructure to the U.S.
無法據以確立. A planned expansion with jobs by end-2027; not current output, volumes or Wiwynn's overall U.S. share.
引用於: 第節 第5.2節, 第B.2節; 表 9, 19, 24; 展望 12

S-307 Vertiv Holdings Co — Vertiv Reports Strong Fourth Quarter with Organic Orders Growth of 252% and Diluted EPS Growth of 200% (Adjusted Diluted EPS +37%)
企業新聞稿 · U.S. · 英文 · 11 February 2026 · 閱讀 27 September 2026
https://investors.vertiv.com/news/news-details/2026/Vertiv-Reports-Strong-Fourth-Quarter-with-Organic-Orders-Growth-of-252-and-Diluted-EPS-Growth-of-200-Adjusted-Diluted-EPS-37/default.aspx
"Fourth quarter 2025 book-to-bill ratio was ~2.9x and backlog increased to $15.0 billion, up 109% compared to the same period last year."
支持的研判. Vertiv reports a US$15.0 billion backlog at end-2025, up 109% year on year, and a fourth-quarter 2025 book-to-bill ratio of about 2.9.
無法據以確立. Group-level company figures; they do not isolate AI or 800 VDC orders, and Vertiv's Q2 2026 release gives no updated backlog.
引用於: 第節 第5.4節, 第B.1節, 第C.7節; 表 13, 28

S-308 Schneider Electric SE — Half Year 2026 Results
企業新聞稿 · France · 英文 · 30 July 2026 · 閱讀 27 September 2026
https://www.se.com/ww/en/assets/pdf/release-hy-results-2026
"Data Center & Networks to sustain strong momentum, supported by a robust backlog and continued demand; Industry and Infrastructure to contribute strongly; Buildings to track macroeconomic trends"
"The Data Center & Networks end-market continues to see sustained high demand, up triple-digit in Q2."
支持的研判. Schneider Electric says its Data Center & Networks end-market saw sustained high demand, 'up triple-digit in Q2' 2026, and expects strong momentum supported by a robust backlog.
無法據以確立. Company commentary; does not say whether orders or revenue rose triple-digit, and gives no data-centre order value.
引用於: 第節 第5.4節; 表 28

S-309 BGL (via StockTitan) — BGL Announces the Sale of Strategic Thermal Labs to Vertiv
企業新聞稿 · U.S. · 英文 · 20 July 2026 · 閱讀 26 September 2026
https://www.stocktitan.net/news/VRT/bgl-announces-the-sale-of-strategic-thermal-labs-to-b3glhctu5yso.html
"direct-to-chip liquid cooling, including the design and development of high-performance cold plate solutions"
支持的研判. BGL announced the sale of Strategic Thermal Labs, a direct-to-chip liquid-cooling and cold-plate specialist, to Vertiv.
無法據以確立. Adviser's release via an aggregator; no price, revenue or capacity of the acquired business is disclosed.
引用於: 第節 第5.4節

S-310 Infineon Technologies AG — Growth prospects improved: Infineon raises full-year guidance. AI boom strengthens further. Automotive order intake improves. Revised segment structure effective from the fourth quarter
企業新聞稿 · Germany · 英文 · 6 May 2026 · 閱讀 27 September 2026
https://www.infineon.com/press-release/2026/infxx202605-082
直接文件: https://www.infineon.com/row/public/documents/corporate/press/2026/infxx202605-082e.pdf
"In this area, notwithstanding a weaker currency, revenue of around €1.5 billion is expected for the 2026 fiscal year and around €2.5 billion for the 2027 fiscal year."
支持的研判. In May 2026 Infineon expected AI data-centre power revenue of around €1.5 billion in FY2026 and around €2.5 billion in FY2027.
無法據以確立. Superseded outlook of May 2026, raised in August 2026; covers Infineon's AI data-centre power business only.
引用於: 第節 第5.4節

S-311 Utility Dive — GE Vernova gas turbine backlog climbs to 116 GW
媒體 · U.S. · 英文 · 23 July 2026 · 閱讀 26 September 2026
https://www.utilitydive.com/news/ge-vernova-gas-turbine-backlog-climbs-to-116-gw/826039
"taking reservations for 2031 deliveries now"
支持的研判. Utility Dive reports GE Vernova is already taking reservations for gas-turbine deliveries in 2031.
無法據以確立. Press report; slot reservations are not firm orders or deliveries.
引用於: 第節 第5.5節, 第5.8節

S-312 Huawei — Groundbreaking SuperPoD Interconnect: Leading a New Paradigm for AI Infrastructure
企業新聞稿 · China · 英文 · 18 September 2025 · 閱讀 26 September 2026
https://www.huawei.com/en/news/2025/9/hc-xu-keynote-speech
"With HiBL 1.0, our proprietary, low-cost HBM that is more cost-effective than HBM3E and HBM4E, the Ascend 950PR will help our customers maintain the right level of performance while significantly reducing their investment in the hardware needed for prefill and recommendation systems."
"That's where our HiZQ 2.0 HBM comes in, delivering 144 GB of memory and 4 TB/s memory access bandwidth."
"The Ascend 950DT chip will be available in the fourth quarter of 2026."
"Today, we are releasing its technical specifications. And we hope that industry partners will adopt this protocol"
支持的研判. Huawei's roadmap (18 September 2025) puts the Ascend 950DT in Q4 2026, the 960 in Q4 2027 and the 970 in Q4 2028, and releases UnifiedBus 2.0 specifications for partners.
無法據以確立. Vendor roadmap dates; not shipments, performance verification or adoption of UnifiedBus outside Huawei.
引用於: 圖 25; 第節 第5.7節, 第7.1節, 第7.5節, 第8.5節, 第9.4節, 第B.2節; 表 9, 19, 24; 展望 18, 19

S-313 Shanghai Stock Exchange — Listing announcement of ChangXin Technology Group Co., Ltd. (688825) on the STAR Market
監管機關 · China · 簡體中文 · 24 July 2026 · 閱讀 26 September 2026
https://www.sse.com.cn/disclosure/announcement/listing/ipo/c/c_20260724_10826610.shtml
"长鑫科技集团股份有限公司A股股票将在本所科创板上市交易"
APEX譯文: The A shares of ChangXin Technology Group Co., Ltd. will be listed and traded on this Exchange's STAR Market.
支持的研判. The Shanghai Stock Exchange's notice of 24 July 2026 says ChangXin Technology Group's A shares will list and trade on its STAR Market (stock code 688825).
無法據以確立. A listing notice only; it says nothing about the prospectus's use of proceeds, HBM plans or production capacity.
引用於: 圖 25; 第節 第5.7節; 表 25

S-314 U.S. Department of Commerce — Statement on UAE and Saudi Chip Exports
政府 · U.S. · 英文 · 19 November 2025 · 閱讀 27 September 2026
https://www.commerce.gov/news/press-releases/2025/11/statement-uae-and-saudi-chip-exports
"Both companies are receiving approvals to purchase the equivalent of up to 35,000 Nvidia Blackwell chips (GB300s)."
支持的研判. Commerce says it authorised advanced-chip exports to G42 (UAE) and HUMAIN (Saudi Arabia) in November 2025, with approvals to purchase the equivalent of up to 35,000 GB300 chips.
無法據以確立. The statement does not say whether 35,000 is per company or combined; approvals carry security and reporting conditions; no shipments shown.
引用於: 圖 25; 第節 第5.7節, 第7.5節; 表 15

S-315 Federal Register (BIS) — Enhanced Favorable Treatment for the United Arab Emirates Under the Export Administration Regulations
監管機關 · U.S. · 英文 · 14 July 2026 · 閱讀 26 September 2026
https://www.federalregister.gov/documents/2026/07/14/2026-14132/enhanced-favorable-treatment-for-the-united-arab-emirates-under-the-export-administration
"Absent subsequent notice by BIS, the authorization provided for G42 and Core42 in this supplement shall automatically expire on April 6, 2027."
支持的研判. BIS's July 2026 rule moved the UAE into Country Group A:5 and authorised licence-free advanced-computing receipt by G42 and Core42, expiring 6 April 2027 absent further BIS notice.
無法據以確立. UAE only; end-use and end-user controls still apply, and the rule gives no shipment volumes.
引用於: 圖 25, 32; 第節 第5.7節, 第7.1節, 第7.5節, 第9.2節; 表 15

S-316 Press Information Bureau, Government of India (Ministry of Electronics & Information Technology) — India to Add 20,000 GPUs Beyond Existing 38,000 to Strengthen National AI Infrastructure; Announces MEITY Minister Shri Ashwini Vaishnaw at Day Two of India AI Impact Summit 2026
政府 · India · 英文 · 17 February 2026 · 閱讀 27 September 2026
https://www.pib.gov.in/PressReleasePage.aspx?PRID=2229171&reg=3&lang=1
"India will scale its compute capacity beyond the existing 38,000 GPUs with the addition of 20,000 GPUs in the coming weeks."
支持的研判. India's government said on 17 February 2026 it will add 20,000 GPUs to its existing 38,000 in the coming weeks.
無法據以確立. A February 2026 announcement of planned additions, not a record of delivery or current capacity.
引用於: 圖 25; 第節 第5.7節, 第7.5節; 表 9

S-317 Amkor Technology (investor relations press release) — Amkor Technology Announces Phase 2 of Arizona Advanced Packaging and Test Campus; Expands Investment to $12 Billion
企業新聞稿 · U.S. · 英文 · 8 September 2026 · 閱讀 27 September 2026
https://ir.amkor.com/news-releases/news-release-details/amkor-technology-announces-phase-2-arizona-advanced-packaging
"The expansion increases the project's planned investment to approximately $12 billion."
"Construction of phase 2 is anticipated to begin late 2027, with completion expected by the end of 2029."
支持的研判. Amkor says Phase 2 raises its planned Arizona packaging and test campus investment to about US$12 billion, with construction anticipated from late 2027 and completion expected by end-2029.
無法據以確立. Company plan; construction has not begun, dates may change, and customer commitments are described only in cleanroom area.
引用於: 第節 第5.7節, 第B.2節; 表 9, 26

S-318 UMC — UMC Licenses imec's iSiPP300 Technology to Extend Silicon Photonics Capabilities for Next-Generation Connectivity
企業新聞稿 · Taiwan · 英文 · 8 December 2025 · 閱讀 26 September 2026
https://www.umc.com/en/News/press_release/Content/technology_related/20251208
"UMC is working with several new customers to deliver PIC chips on this new platform for optical transceiver applications, with risk production slated for 2026 and 2027."
支持的研判. UMC says it licensed imec's iSiPP300 silicon-photonics platform and is working with several new customers on PIC chips for optical transceivers, with risk production slated for 2026 and 2027.
無法據以確立. Company plan; risk production is not volume production, and customers, volumes and revenue are not named.
引用於: 第節 第5.7節, 第B.2節; 表 9, 30

S-319 TrendForce — Japanese and Korean MLCC Makers Cut Consumer Supply to Focus on AI, Chinese Suppliers Enter Global Brand Qualification, Says TrendForce
市場研究 · Taiwan · 英文 · 24 September 2026 · 閱讀 26 September 2026
https://www.trendforce.com/presscenter/news/20260924-13252.html
"Murata announced on September 10th that it would discontinue certain products, such as general-purpose GRM types for consumer use"
支持的研判. TrendForce reports Japanese and Korean MLCC makers cutting consumer supply to focus on AI, citing Murata's 10 September plan to discontinue general-purpose consumer GRM types.
無法據以確立. Discontinuation covers certain consumer product lines, not all consumer MLCCs; the captured excerpt names Murata only.
引用於: 第節 第5.7節

S-320 imec — UMC licenses imec's iSiPP300 technology to extend silicon photonics capabilities for next-generation applications
研究 · Belgium · 英文 · 8 December 2025 · 閱讀 26 September 2026
https://www.imec-int.com/en/press/umc-licenses-imecs-isipp300-technology-extend-silicon-photonics-capabilities-next-generation
"UMC is working with several new customers to deliver PIC chips on this new platform for optical transceiver applications, with risk production slated for 2026 and 2027"
支持的研判. imec's December 2025 release says UMC licensed its iSiPP300 silicon-photonics platform and is working with new customers on PIC chips, with risk production slated for 2026 and 2027.
無法據以確立. A licensing announcement; risk production is a plan, and the release names no customers, volumes or licence terms.
引用於: 第節 第5.7節; 表 30

S-321 Tom's Hardware — CXMT closes up 466 percent in Shanghai debut with no HBM project in its IPO prospectus
媒體 · U.S. · 英文 · 27 July 2026 · 閱讀 27 September 2026
https://www.tomshardware.com/tech-industry/cxmt-closes-up-466-percent-in-shanghai-debut-with-no-hbm-project-in-its-ipo-prospectus
"The filing contains no dedicated HBM project and no disclosed funding commitment to a near-term HBM expansion, and CXMT hasn't broken down where the remaining roughly 28 billion yuan goes beyond describing it as working capital."
"Nearly 70% of the named spending goes to wafer lines and DRAM processes."
支持的研判. Tom's Hardware reports CXMT's IPO prospectus contains no dedicated HBM project or near-term HBM funding commitment, with nearly 70% of named spending going to wafer lines and DRAM processes.
無法據以確立. Press reading of a prospectus; absence of a disclosed HBM project does not show that CXMT has no HBM development.
引用於: 第節 第5.7節, 第B.2節; 表 25

S-322 Tom's Hardware, citing Counterpoint Research — YMTC breaks into the top three NAND makers for the first time as AI servers swallow 48% of all flash
媒體 · U.S. · 英文 · 12 August 2026 · 閱讀 27 September 2026
https://www.tomshardware.com/tech-industry/ymtc-breaks-into-the-top-three-nand-makers-for-the-first-time
"Yangtze Memory Technologies (YMTC) shipped 14% of the world's NAND flash in the second quarter of 2026, entering the global top three for the first time and narrowly displacing Kioxia, according to Counterpoint Research's Memory & Storage Tracker published today."
支持的研判. Tom's Hardware, citing Counterpoint Research, reports YMTC shipped 14% of the world's NAND flash in 2Q26, entering the global top three for the first time.
無法據以確立. Shipment share from one tracker; the article does not state the unit as bits, and YMTC ranks lower by revenue.
引用於: 第節 第5.7節; 表 25

S-323 European Commission (EUR-Lex) — AI Continent Action Plan, COM(2025) 165 final
政府 · EU · 英文 · 9 April 2025 · 閱讀 26 September 2026
https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52025DC0165
"mobilise EUR 20 billion investment for AI infrastructure, notably targeting up to 5 AI Gigafactories"
支持的研判. The Commission's AI Continent Action Plan (COM(2025) 165, 9 April 2025) aims to mobilise EUR 20 billion investment for AI infrastructure, notably targeting up to 5 AI Gigafactories.
無法據以確立. A Commission communication stating an aim to mobilise investment, not committed EU budget; the €20 billion is for AI infrastructure, notably gigafactories.
引用於: 第節 第5.7節, 第7.5節

S-324 StorageReview — NVIDIA Spectrum-X Ethernet Photonics Enters Full Production With 4x Fewer Lasers and a Five-Vendor CPO Supply Chain
媒體 · U.S. · 英文 · 15 August 2026 · 閱讀 27 September 2026
https://www.storagereview.com/news/nvidia-spectrum-x-ethernet-photonics-enters-full-production-with-4x-fewer-lasers-and-a-five-vendor-cpo-supply-chain
"TSMC handles silicon photonics fabrication. SPIL performs chip-scale packaging and testing."
"Lumentum fabricates the laser chips, and TFC Communication builds the laser module subassemblies. Foxconn does the final switch system assembly."
支持的研判. StorageReview reports NVIDIA's Spectrum-X Photonics CPO chain: TSMC silicon photonics, SPIL packaging and test, Lumentum laser chips, TFC laser modules and Foxconn final switch assembly.
無法據以確立. Trade-press account; it gives no volumes or shares by supplier, and does not show that no second source exists.
引用於: 第節 第5.7節

S-325 Rapidus — Rapidus Completes 150 Billion Yen Funding Round from Japan Government
企業新聞稿 · Japan · 英文 · 5 June 2026 · 閱讀 27 September 2026
https://www.rapidus.inc/en/news_topics/information/20260605-2
"Rapidus Corporation today announced that it has completed an additional funding round of 150 billion yen (equivalent to $943 million USD) from the Information-Technology Promotion Agency (IPA), Japan, an independent administrative agency under the jurisdiction of the Ministry of Economy, Trade and Industry (METI)"
"To support its evolution from the R&D stage to full-scale 2nm logic semiconductor manufacturing by 2027, Rapidus will continue to raise capital and financing from a mix of public and private sources."
支持的研判. Rapidus completed a ¥150 billion funding round from the Japanese government (5 June 2026) to support its move to full-scale 2 nm logic manufacturing by 2027.
無法據以確立. Company funding news and target; 2027 is a goal, not evidence of yields, customers or volume production.
引用於: 表 9, 26; 第節 第7.5節

S-326 KIOXIA Holdings Corporation — Kioxia Announces Growth Strategy for the AI Inference Era at Investor Day
企業新聞稿 · Japan · 英文 · 2 June 2026 · 閱讀 27 September 2026
https://www.kioxia-holdings.com/en-jp/news/2026/20260602-1.html
"To drive this high-quality profitability, Kioxia will allocate approximately 470 billion yen annually to capital expenditures and 230 billion yen annually to research and development over the next three years, focusing heavily on high-growth, high-profit sectors."
支持的研判. Kioxia says it will allocate about JPY 470 billion a year to capital expenditure and JPY 230 billion a year to R&D over the next three years.
無法據以確立. A spending plan from an investor day, not a stated capacity addition or a commitment to specific fabs; execution may change.
引用於: 表 9

S-327 The National — Stargate UAE's first phase to be completed in third quarter of 2026
媒體 · UAE · 英文 · 5 December 2025 · 閱讀 26 September 2026
https://www.thenationalnews.com/business/2025/12/05/stargate-uaes-first-phase-to-be-completed-in-third-quarter-of-2026
"That first phase of 200MW will be in place and ready at some point in the third quarter of next year."
支持的研判. The National reported in December 2025 that Stargate UAE's first 200 MW phase would be in place and ready in the third quarter of 2026 ('next year').
無法據以確立. A reported statement by Mubadala's chief executive, dated December 2025; it does not confirm commissioning, and the excerpt does not name the speaker.
引用於: 表 9

S-328 ASML Holding N.V. — ASML provides updated view on market opportunities at 2024 Investor Day meeting
企業新聞稿 · Netherlands · 英文 · 14 November 2024 · 閱讀 27 September 2026
https://www.asml.com/en/news/press-releases/2024/asml-investor-day-2024
"At its 2024 Investor Day today, ASML Holding N.V. (ASML) provides an update on its long-term strategy as well as global market and technology trends, confirming potential 2030 scenarios for annual revenue between approximately €44 billion and €60 billion with a gross margin of between approximately 56% and 60%."
支持的研判. ASML's November 2024 Investor Day confirmed potential 2030 scenarios of annual revenue between about €44 billion and €60 billion, with gross margin of about 56–60%.
無法據以確立. Scenarios, not guidance; dated November 2024 and due to be updated at the June 2027 Capital Markets Day.
引用於: 表 10; 第節 第9.3節, 第C.5節

S-329 Open Compute Project Foundation — Home » Open Compute Project
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org
"Though not a standards body, the OCP Community forges new technology norms"
"The Open Compute Project was first initiated by Facebook, now Meta, in 2011."
支持的研判. OCP's home page says Facebook, now Meta, initiated the project in 2011 and that, though not a standards body, the OCP Community forges new technology norms.
無法據以確立. Self-description on an undated web page; it does not establish that OCP documents are adopted or referenced as standards.
引用於: 第節 第6.1節, 第C.7節

S-330 Open Compute Project Foundation — Bylaws of Open Compute Project Foundation As Amended 14 June 2024
標準組織 · Global · 英文 · 14 June 2024 · 閱讀 26 September 2026
https://www.opencompute.org/documents/bylaws-june-2024
"Delaware nonprofit nonstock corporation"
支持的研判. OCP's bylaws, as amended 14 June 2024, organise the Foundation as a Delaware nonprofit nonstock corporation for purposes under Section 501(c)(6) of the Internal Revenue Code.
無法據以確立. Governing document; it does not describe how specifications are approved or how board seats map to member companies today.
引用於: 第節 第6.1節

S-331 Open Compute Project Foundation — Board Members » Open Compute Project; OCP Adds Three New Companies to its Board of Directors (2025-10-13)
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/about/board
"Board Member, OCP CEO"
支持的研判. OCP's board page lists directors from Meta (chair), Google, Microsoft, AMD, Arm, Intel and NVIDIA, plus an individual director and CEO George Tchaparian as 'Board Member, OCP CEO'.
無法據以確立. Board composition as read in September 2026; it does not show how seats are allocated, voting power, or which companies author specific specifications.
引用於: 第節 第6.1節, 第C.7節

S-332 Open Compute Project Foundation — The Open Compute Project Foundation Expands its Advisory Board, Market and Global Coverage
標準組織 · Global · 英文 · 15 January 2026 · 閱讀 26 September 2026
https://www.opencompute.org/blog/the-open-compute-project-foundation-expands-its-advisory-board-market-and-global-coverage
"we are expanding the Foundation's Advisory Board specifically with representatives from organizations that reach new businesses, new technology segments, and geographies relevant to the OCP Community"
支持的研判. On 15 January 2026 OCP expanded its Advisory Board with six companies, among them Delta Electronics, Schneider Electric and Vantage Data Centers, to reach new businesses, segments and geographies.
無法據以確立. It does not state the Advisory Board's powers or any role in approving specifications.
引用於: 第節 第6.1節

S-333 Open Compute Project Foundation — About OCP Accepted
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/sp/about-ocp-accepted
"A product can be recognized as OCP Accepted™ if it complies with an approved OCP specification or other types of contribution; HOWEVER, no design files have been contributed."
支持的研判. OCP says a product can be recognised as OCP Accepted if it complies with an approved OCP specification or other contribution, even though no design files have been contributed.
無法據以確立. Programme page; it does not grant patent licences or show test methods, and another OCP page describes Accepted products inconsistently.
引用於: 第節 第6.1節, 第C.7節; 附錄A

S-334 Open Compute Project Foundation — Solution Provider Program FAQs
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/sp/frequently-asked-questions
"No, if you are seeking OCP Accepted™ you do not have to disclose any design files"
支持的研判. OCP's Solution Provider FAQ says applicants seeking OCP Accepted do not have to disclose any design files.
無法據以確立. FAQ guidance only; applicants must still identify the approved specification or contribution their product aligns with.
引用於: 第節 第6.1節

S-335 Open Compute Project Foundation; IDC — Foundation Overview 2026 (OCP); IDC Market Impact Study slides, November 2025
標準組織 · Global · 英文 · 未註明日期 · 閱讀 27 September 2026
https://146a55aca6f00848c565-a7635525d40ac1c70300198708936b4e.ssl.cf1.rackcdn.com/images/4660bec1bbfec1bf0c0fac4b2baf203c24ad85b2.pdf
"800+ of SKU listings"
"150+ Solution Providers"
支持的研判. OCP's Foundation Overview 2026 brochure counts 500+ members, 150+ Marketplace solution providers and 800+ SKU listings.
無法據以確立. Self-reported, undated promotional counts of participation; they do not measure deployment, contribution or adoption of any specification.
引用於: 第節 第6.1節

S-336 Open Compute Project Foundation — Semi-Private Workstreams » Open Compute Project
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/community/semi-private-workstreams
"Semi-Private Workstreams consist of 'Contributor' organizations (members that have signed a Contribution License Agreement, or CLA), who are collaborating to jointly develop a specification."
支持的研判. OCP defines semi-private workstreams as CLA-signing contributor members jointly developing a specification; other members may join by signing OCP's Feedback Agreement.
無法據以確立. It does not say when or how semi-private drafts become visible to non-participants; that inference is the report's.
引用於: 第節 第6.1節

S-337 Open Compute Project Foundation — Open Compute Project Foundation Launches Collaboration Acceleration Fund to Advance Open Hardware Innovation
標準組織 · Global · 英文 · 7 April 2026 · 閱讀 26 September 2026
https://www.opencompute.org/blog/open-compute-project-foundation-launches-collaboration-acceleration-fund-to-advance-open-hardware-innovation
"By lowering barriers to collaboration and supporting projects that benefit the broader ecosystem, we are ensuring that transformative ideas can move forward in a transparent and inclusive way."
支持的研判. OCP announced its Collaboration Acceleration Fund on 7 April 2026 as a grant programme to lower barriers to collaboration on projects benefiting the wider ecosystem.
無法據以確立. Launch announcement; matching terms and licence conditions are on the fund page (S-338), and no fund size or awards are disclosed.
引用於: 第節 第6.1節

S-338 Open Compute Project Foundation — OCP Foundation Collaboration Acceleration Fund » Open Compute Project
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/ocp-foundation-collaboration-acceleration-fund
"The Collaboration Acceleration Fund is a non-equity, pre-competitive grant program"
支持的研判. OCP's fund page calls the Collaboration Acceleration Fund 'a non-equity, pre-competitive grant program' offering 25%–50% resource matching, with outcomes contributed back under OCP-approved licences.
無法據以確立. No fund size or awards are disclosed; licence terms depend on the type of contribution under OCP's IPR policy.
引用於: 第節 第6.1節

S-339 Open Compute Project Foundation — Theme » 2026 OCP Global Summit » Open Compute Project
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/summit/global-summit/theme
"rethink data center technologies from Chip to Grid"
"AI computation is to be distributed from centralized data centers to regional data centers, to devices closer to the end user."
支持的研判. OCP's 2026 Summit theme page calls for rethinking data-centre technologies 'from Chip to Grid' and distributing AI computation from central data centres to regional sites and devices.
無法據以確立. Event framing, not a work programme or specification commitment.
引用於: 第節 第6.2節

S-340 Open Compute Project Foundation — 2025 OCP Global Summit (Past Events) | Open Compute Project
標準組織 · Global · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.opencompute.org/events/past-events/2025-ocp-global-summit
"Co-located - DMTF"
"Co-located - Misc"
支持的研判. OCP's 2025 Global Summit page (San Jose, from 13 October 2025) groups co-located sessions under DMTF, IPEC, P4, SONiC and 'Misc'; OIF does not appear.
無法據以確立. A recordings index, not the full programme; the 'Misc' group is not itemised, so the page cannot prove that no OIF workshop was held.
引用於: 第節 第6.2節

S-341 Lattice Semiconductor (Business Wire) — Lattice to Deliver Keynote at 2026 OCP Global Summit
企業新聞稿 · U.S. · 英文 · 12 August 2026 · 閱讀 26 September 2026
https://ir.latticesemi.com/news-releases/news-release-details/lattice-deliver-keynote-2026-ocp-global-summit
"the growing need for a common management and control model for AI and compute infrastructure"
支持的研判. Lattice announced a 2026 OCP Global Summit keynote on 'the growing need for a common management and control model for AI and compute infrastructure'.
無法據以確立. An event announcement of a talk topic; not a specification, product or OCP position.
引用於: 第節 第6.2節

S-342 Open Compute Project Foundation — Innovation Village » 2026 OCP Global Summit » Open Compute Project
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/summit/global-summit/innovation-village
"Discover prototype demonstrations from teams working on OCP Accepted™ and OCP Inspired™ innovations."
支持的研判. OCP describes the 2026 Summit's Innovation Village as prototype demonstrations from teams working on OCP Accepted and OCP Inspired innovations.
無法據以確立. Undated event page still naming the retiring OCP Inspired mark; demonstrations are not specifications or recognised products.
引用於: 第節 第6.2節

S-343 Open Compute Project Foundation — OCP IP Rights Management Policy; FAQs about Licenses and OCP's Legal Process
標準組織 · Global · 英文 · 14 August 2026 · 閱讀 27 September 2026
https://146a55aca6f00848c565-a7635525d40ac1c70300198708936b4e.ssl.cf1.rackcdn.com/images/d6e18c0c54567ca4a971ef140f53c531933e946f.pdf
"At the conclusion or waiver of the Opt-Out Period the SC approved contribution will be published."
"OCP staff will announce a 30-day 'opt-out period' via formal written notice to the contributors."
支持的研判. OCP's legal FAQ says staff announce a 30-day opt-out period to contributors, and the Steering Committee-approved contribution is published when that period ends or is waived.
無法據以確立. Undated FAQ; it does not say what a released Final Specification prints (licence notice, contributor names, exclusions); that is in the IP policy (S-047).
引用於: 第節 第6.2節, 第6.6節

S-344 Open Compute Project Foundation — OCP Global Summit is moving to a new location!
標準組織 · Global · 英文 · 1 September 2026 · 閱讀 26 September 2026
https://www.opencompute.org/blog/ocp-global-summit-is-moving-to-a-new-location
"Beginning in 2027, the OCP Global Summit is moving to the Moscone Center in downtown San Francisco, California"
"2027: October 4–7 at Moscone Center; 2028: September 25–28; 2029: October 15–18; 2030: September 30–October 3"
支持的研判. OCP says the Global Summit moves to San Francisco's Moscone Center from 2027, with dates 4–7 October 2027, 25–28 September 2028, 15–18 October 2029 and 30 September–3 October 2030.
無法據以確立. Announced venue and dates, subject to change; it says nothing about agendas or specification releases at those Summits.
引用於: 第節 第6.2節, 第9.2節; 展望 2, 5, 13, 14, 16; 圖 32

S-345 Open Compute Project Foundation — 2026 OCP APAC Summit » Open Compute Project; 2026 OCP APAC Summit: By the Numbers
標準組織 · Global · 英文 · 27 August 2026 · 閱讀 27 September 2026
https://www.opencompute.org/events/past-events/2026-ocp-apac-summit
"2,453 registered participants"
"The 2026 OCP APAC Summit brought nearly 2,500 attendees of the global OCP Community together at TaiNEX2 in Taipei City, Taiwan, on August 11–12"
支持的研判. OCP reports that its 2026 APAC Summit ran on 11–12 August 2026 at TaiNEX2 in Taipei, with 2,453 registered participants.
無法據以確立. Event statistics only; it does not list specification releases timed to the Summit.
引用於: 第節 第6.2節

S-346 UALink Consortium — Ultra Accelerator Link (UALink) Consortium Publishes Four Specifications Defining In-Network Compute
標準組織 · U.S. · 英文 · 7 April 2026 · 閱讀 26 September 2026
https://ualinkconsortium.org/wp-content/uploads/2026/04/UALink-2.0-Specification-PR_FINAL.pdf
"The advancements to UALink technology introduced in this release will enable the industry to quickly and efficiently integrate UALink solutions into their architectures."
支持的研判. The UALink Consortium's 7 April 2026 release, titled as publishing four specifications defining in-network compute, presents them as enabling quick integration of UALink solutions.
無法據以確立. Consortium announcement; the excerpt does not list the four documents, and nothing shows conformant products or volume.
引用於: 第節 第6.3節, 第6.7節; 表 11; 圖 26; 展望 16

S-347 IEEE 802.3 Ethernet Working Group — IEEE 802.3 Ethernet Working Group
標準組織 · Global · 英文 · 6 August 2026 · 閱讀 27 September 2026
https://grouper.ieee.org/groups/802/3/index.html
"The IEEE 802.3 Working Group develops standards for Ethernet networks. We have a number of active projects, study groups, and ad hocs as listed below:"
"IEEE P802.3dj 200 Gb/s, 400 Gb/s, 800 Gb/s, and 1.6 Tb/s Ethernet Task Force."
支持的研判. The IEEE 802.3 Working Group page (updated 6 August 2026) lists IEEE P802.3dj, the 200 Gb/s to 1.6 Tb/s Ethernet Task Force, among its active projects.
無法據以確立. Project listing only: it shows P802.3dj unfinished, not product volumes or multi-vendor deployment of Ethernet scale-out.
引用於: 第節 第6.3節; 表 11; 圖 26

S-348 CXL Consortium (press release via Business Wire) — CXL Consortium Releases the Compute Express Link 4.0 Specification Increasing Speed and Bandwidth
標準組織 · Global · 英文 · 18 November 2025 · 閱讀 27 September 2026
https://www.businesswire.com/news/home/20251118275848/en/CXL-Consortium-Releases-the-Compute-Express-Link-4.0-Specification-Increasing-Speed-and-Bandwidth
"CXL Consortium, the industry standard organization developing and promoting an open coherent interconnection for heterogeneous memory and computing solutions, today announced the release of the Compute Express Link® (CXL®) 4.0 specification to meet the increasing demands of emerging workloads placed on today's data centers."
"Doubles the bandwidth to 128GTs with zero added latency"
支持的研判. The CXL Consortium announced on 18 November 2025 the release of the CXL 4.0 specification, which doubles bandwidth to 128 GT/s with zero added latency.
無法據以確立. Specification release only; does not establish products, multi-vendor interoperability or volume shipments.
引用於: 第節 第6.3節, 第6.7節; 表 11; 圖 26

S-349 PCI-SIG (issuer's release distributed via Business Wire) — PCI-SIG® Releases PCIe® 7.0 Specification to Support the Bandwidth Demands of Artificial Intelligence at 128.0 GT/s Transfer Rates
標準組織 · U.S. · 英文 · 11 June 2025 · 閱讀 27 September 2026
https://www.businesswire.com/news/home/20250611299049/en/PCI-SIG-Releases-PCIe-7.0-Specification-to-Support-the-Bandwidth-Demands-of-Artificial-Intelligence-at-128.0-GTs-Transfer-Rates
"PCI-SIG® today announced the official release of the PCI Express® (PCIe®) 7.0 specification, reaching 128.0 GT/s, to members."
支持的研判. PCI-SIG announced on 11 June 2025 the official release of the PCIe 7.0 specification, reaching 128.0 GT/s, to members.
無法據以確立. Released to members only; the release does not show public availability, products or shipment dates.
引用於: 第節 第6.3節, 第A.2節; 表 11, 22; 圖 26

S-350 PCI-SIG (issuer's release distributed via Business Wire) — PCI-SIG® Announces PCI Express® 8.0 Specification to Reach 256.0 GT/s
標準組織 · U.S. · 英文 · 5 August 2025 · 閱讀 27 September 2026
https://www.businesswire.com/news/home/20250805675479/en/PCI-SIG-Announces-PCI-Express-8.0-Specification-to-Reach-256.0-GTs
"PCI-SIG® today announced the PCI Express® (PCIe®) 8.0 specification will double the data of the PCIe 7.0 specification to 256.0 GT/s and is planned for release to members by 2028."
支持的研判. PCI-SIG announced on 5 August 2025 that the PCIe 8.0 specification will reach 256.0 GT/s and is planned for release to members by 2028.
無法據以確立. A plan ('planned for release'), not a released specification; the release target is to members, not public.
引用於: 第節 第6.3節; 表 11, 22

S-351 Open Compute Project Foundation — Open Compute Project Foundation and UALink Consortium Announce a New Collaboration
標準組織 · Global · 英文 · 29 April 2025 · 閱讀 26 September 2026
https://www.opencompute.org/blog/open-compute-project-foundation-and-ualinktm-consortium-announce-a-new-collaboration
"Partnering with the OCP Community will accelerate the adoption of UALink's innovations into complete systems, delivering transformative performance for AI markets."
支持的研判. OCP and the UALink Consortium announced a collaboration on 29 April 2025 to accelerate adoption of UALink into complete AI systems through OCP work.
無法據以確立. Announcement of intent; it names no joint specification, and UALink's specifications and IP terms remain the consortium's own.
引用於: 第節 第6.3節; 表 11

S-352 Open Compute Project Foundation — Open Compute Project Foundation and SNIA Announce a New Collaboration
標準組織 · Global · 英文 · 13 October 2025 · 閱讀 27 September 2026
https://www.opencompute.org/blog/open-compute-project-foundation-and-snia-announce-a-new-collaboration
"the Open Compute Project Foundation (OCP), the nonprofit international organization bringing at-scale innovations and hyperscale best practices to all, and SNIA, a nonprofit global organization for technologies related to handling and optimizing data, announce a new collaboration as alliance partners"
支持的研判. OCP and SNIA announced a collaboration as alliance partners on 13 October 2025, spanning storage, memory and data pipelines for AI.
無法據以確立. Announcement only; it records no jointly developed specification or document, and each body's work (OCP L.O.C.K., SNIA Storage.AI) stays separate.
引用於: 第節 第6.3節; 表 11

S-353 Ultra Ethernet Consortium — UEC Press Releases
標準組織 · U.S. · 英文 · 17 October 2023 · 閱讀 27 September 2026
https://ultraethernet.org/press-releases
"Open Compute Project Foundation and Ultra Ethernet Consortium Announce a New Collaboration"
支持的研判. UEC's press-release list includes 'Open Compute Project Foundation and Ultra Ethernet Consortium Announce a New Collaboration', dated 17 October 2023.
無法據以確立. A listing title only; the collaboration's scope and any joint deliverables are not shown on the index page.
引用於: 表 11

S-354 OCP Foundation — Alliance Partners - Open Compute Project
標準組織 · Global · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.opencompute.org/about/alliance-partners
“The OCP works with many organizations, and what follows is a list of OCP's alliance partners with formal agreements used to bridge communities.”
“OCP collaboration with JEDEC brings together members from both the OCP and JEDEC communities to share efforts in developing and maintaining global standards needed to advance the electronics industry.”
支持的研判. OCP's alliance-partners page lists JEDEC among partners with formal agreements, describing a collaboration to share efforts on global standards for the electronics industry.
無法據以確立. Undated listing; it names no joint JEDEC–OCP document and does not show OCP authoring memory standards, which remain JEDEC's.
引用於: 表 11

S-355 DMTF — Redfish | DMTF
標準組織 · U.S. · 英文 · 14 September 2026 · 閱讀 27 September 2026
https://www.dmtf.org/standards/redfish
“DMTF's Redfish® is a standard designed to deliver simple and secure management for converged, hybrid IT and the Software Defined Data Center (SDDC).”
“Redfish Release 2026.2 Overview”
支持的研判. DMTF's Redfish page describes Redfish as a standard for simple and secure management of converged, hybrid IT and the software-defined data centre, and lists Release 2026.2 (14 September 2026).
無法據以確立. Standards-body page; it shows releases, not multi-vendor deployment volume. Specification 1.25.0 is in the release listing, not these excerpts.
引用於: 表 11; 圖 26

S-356 DMTF — SPDM | DMTF
標準組織 · U.S. · 英文 · 2 July 2026 · 閱讀 26 September 2026
https://www.dmtf.org/standards/spdm
“hybrid support of traditional and PQC”
支持的研判. DMTF's SPDM page records SPDM (DSP0274) version 1.4.1 of 2 July 2026 and the SPDM Working Group's plan for hybrid support of traditional and PQC in SPDM 1.5.
無法據以確立. The 1.4.1 date is from the page's listing, not this excerpt; a plan for 1.5 is not a published standard.
引用於: 表 11, 22; 第節 第6.4節

S-357 DMTF — Recent News | DMTF
標準組織 · U.S. · 英文 · 24 September 2026 · 閱讀 27 September 2026
https://www.dmtf.org/news
“Help Us Get SPDM 1.5 Right: Feedback Requested on Plan for Hybrid Support of Traditional and PQC (Deadline: August 31, 2026)”
“Redfish Release 2026.2 Now Available”
支持的研判. DMTF's news page requests feedback by 31 August 2026 on a plan for hybrid traditional and PQC support in SPDM 1.5, and announces Redfish Release 2026.2.
無法據以確立. A plan under consultation, not a published SPDM 1.5; the page does not show SPDM 1.4.1's date or any procurement rule.
引用於: 表 11, 22; 第節 第6.4節, 第C.7節

S-358 Open Compute Project Foundation (Test & Validation) — opencomputeproject/ocp-diag-ctam (GitHub)
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://github.com/opencomputeproject/ocp-diag-ctam
“provides acceptance testing for Accelerator Management in cloud data centers”
支持的研判. OCP's Compliance Tool for Accelerator Management (CTAM) repository says the tool 'provides acceptance testing for Accelerator Management in cloud data centers'.
無法據以確立. Repository description; release versions and adoption are not stated. The Redfish basis rests on the README's test description, not this excerpt.
引用於: 表 11; 第節 第6.4節

S-359 CHIPS Alliance (Linux Foundation) — CHIPS Alliance at 2026 OCP Global Summit
標準組織 · U.S. · 英文 · 27 August 2026 · 閱讀 26 September 2026
https://www.chipsalliance.org/news/ocp-global-2026
“OCP is driving this capability toward DMTF SPDM, where it is anticipated for addition in SPDM 1.5”
“Data-center operators and their tenants need to provision their own identity certificates to fleet hardware, independent of vendor-issued birth identities.”
支持的研判. CHIPS Alliance says OCP is driving the attested certificate-request capability of its Device Identity Provisioning work toward DMTF SPDM, where it is anticipated for addition in SPDM 1.5.
無法據以確立. 'Anticipated' is an expectation reported by CHIPS Alliance, not a DMTF decision; SPDM 1.5 was unpublished at cut-off.
引用於: 表 11; 第節 第6.4節, 第6.6節

S-360 IETF — Fast Network Notifications Problem Statement
標準組織 · Global · 英文 · 7 May 2026 · 閱讀 26 September 2026
https://datatracker.ietf.org/doc/draft-ietf-rtgwg-net-notif-ps
“Many network applications, ranging from Artificial Intelligence (AI)/Machine Learning (ML) training or inference to cloud services, require high bandwidth, low delay, low jitter”
支持的研判. The IETF RTGWG draft problem statement on fast network notifications names AI/ML training and inference among applications needing high bandwidth, low delay and low jitter.
無法據以確立. An Internet-Draft (revision of 7 May 2026), not an RFC or standard; it states motivations, not a solution or any deployment.
引用於: 表 11; 第節 第6.5節

S-361 IETF Datatracker — BoF request: FAst Network Notification (FANN)
標準組織 · Global · 英文 · 15 June 2026 · 閱讀 27 September 2026
https://datatracker.ietf.org/doc/bofreq-dong-fast-network-notification-fann
“This is a placeholder for a FANN WG meeting, which is in the process of chartering.”
支持的研判. IETF Datatracker shows the FAst Network Notification (FANN) BoF request in state 'Approved', last updated 15 June 2026, and says a FANN working group is being chartered.
無法據以確立. Process status only; the page gives no approval date separate from its last update and does not itself mention AI (S-360 does).
引用於: 表 11

S-362 Open Compute Project Foundation — Caliptra 1.0 » Open Compute Project Marketplace
標準組織 · Global · 英文 · 18 April 2024 · 閱讀 26 September 2026
https://www.opencompute.org/products/614/caliptra-10
“IP and firmware for an integrated Root of Trust block”
支持的研判. OCP's Marketplace entry for Caliptra 1.0 describes 'IP and firmware for an integrated Root of Trust block' for data-centre system-on-chip designs.
無法據以確立. A 2024 Caliptra 1.0 listing; it states no licence (Apache 2.0 comes from CHIPS Alliance) and no post-quantum features.
引用於: 第節 第6.4節, 第C.7節

S-363 CHIPS Alliance — chipsalliance/Caliptra (GitHub)
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://github.com/chipsalliance/Caliptra
“All code written for Caliptra and found in these repositories is licensed under the Apache Source License 2.0.”
支持的研判. The Caliptra GitHub repository states that all code written for Caliptra in its repositories is licensed under the Apache License 2.0; the project sits in the CHIPS Alliance.
無法據以確立. Covers code only; Caliptra's specification terms and the reach of Apache 2.0's patent grant are not stated in the excerpt.
引用於: 第節 第6.4節, 第6.6節, 第C.7節; 表 12

S-364 Open Compute Project Foundation — Caliptra: A Datacenter System on a Chip (SoC) Root of Trust (RoT) 2.0 Modular Base Specification
標準組織 · Global · 英文 · 31 July 2025 · 閱讀 27 September 2026
https://www.opencompute.org/documents/caliptra-2-0-pdf
“July 2025 2.0 Major changes. Addition of Caliptra Subsystem.”
“To ensure quantum-resistant RTM, each certificate includes a dual signatures based on ECC Secp384r1 and PQC MLDSA-87.”
支持的研判. Caliptra 2.0 (July 2025; AMD, Google, Microsoft, NVIDIA) adds the Caliptra Subsystem and dual ECC secp384r1/ML-DSA-87 certificate signatures; firmware trust extends from the Caliptra ROM.
無法據以確立. A specification, not silicon: it does not show which parts integrate 2.0, and the text read ties dual signatures to certificates, not firmware images.
引用於: 第節 第6.4節, 第C.7節; 表 21

S-365 CHIPS Alliance (Linux Foundation) — CHIPS Alliance's Caliptra Launches 2.1 RTL Release
標準組織 · U.S. · 英文 · 15 October 2025 · 閱讀 27 September 2026
https://www.chipsalliance.org/news/caliptra2-1
“Caliptra 2.1 integrates Adams Bridge 2.0, offering both ML-DSA and ML-KEM with side channel countermeasures.”
“Implemented in Caliptra 2.1, L.O.C.K provides secure and transparent key management in self-encrypting storage devices.”
支持的研判. CHIPS Alliance's 15 October 2025 news item says Caliptra 2.1 integrates Adams Bridge 2.0 with ML-DSA and ML-KEM and implements L.O.C.K key management for self-encrypting storage.
無法據以確立. The project's own announcement; the date is the news date (the RTL tag may be earlier) and it shows no silicon shipping with 2.1.
引用於: 第節 第6.4節, 第C.7節; 表 21

S-366 Open Compute Project Foundation — NVMe Project OCP L.O.C.K. Modular Base Specification (v1.1 RC4)
標準組織 · Global · 英文 · 3 June 2026 · 閱讀 26 September 2026
https://www.opencompute.org/documents/ocp-lock-specification-v1-1-final-pdf
“MEKs are never visible to drive firmware”
“provide secure key management for Data-At-Rest protection in self-encrypting storage devices”
支持的研判. OCP L.O.C.K. 1.1 (RC4, effective 3 June 2026) specifies key management for data-at-rest in self-encrypting drives, in which media encryption keys are never visible to drive firmware.
無法據以確立. A release candidate despite its 'final' URL; it does not show shipping drives or controllers that implement it.
引用於: 第節 第6.4節; 表 21

S-367 Open Compute Project Foundation — OCP S.A.F.E. Program » Open Compute Project
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/community/ocp-safe-program
“standardizes firmware security reviews across the data center supply chain”
支持的研判. OCP's S.A.F.E. page says the programme standardizes firmware security reviews across the data-centre supply chain and lists 11 approved Security Review Providers.
無法據以確立. Programme description as read in September 2026; it does not show which products were reviewed or how findings bind buyers.
引用於: 第節 第6.4節

S-368 Open Compute Project Foundation — Server/MHS - OpenCompute wiki
標準組織 · Global · 英文 · 15 July 2026 · 閱讀 26 September 2026
https://www.opencompute.org/wiki/Server/DC-MHS
“MHS DC-SCM 2.x and 3.0 Workstreams - Current Developments and What's Ahead”
支持的研判. OCP's Server/MHS wiki (last modified 15 July 2026) lists Summit material titled 'MHS DC-SCM 2.x and 3.0 Workstreams', showing a DC-SCM 3.0 workstream under way.
無法據以確立. A session title from the October 2025 Summit, not a specification; no DC-SCM 3.0 draft or date is posted. Rev 2.2 status rests on S-242.
引用於: 第節 第6.4節; 表 21

S-369 Lattice Semiconductor — Lattice Semiconductor Completes Acquisition of AMI
企業新聞稿 · U.S. · 英文 · 27 July 2026 · 閱讀 27 September 2026
https://ir.latticesemi.com/news-releases/news-release-details/lattice-semiconductor-completes-acquisition-ami
“Lattice Semiconductor Corporation (NASDAQ: LSCC), the low power programmable leader, today announced the completion of its acquisition of AMI.”
“AMI, a Lattice Company, is a global leader in dynamic firmware for security, orchestration, and manageability solutions”
支持的研判. Lattice, a programmable-logic vendor, completed its acquisition of AMI, a firmware supplier for security, orchestration and manageability, on 27 July 2026.
無法據以確立. The release does not name MegaRAC or BMC firmware, and gives no market share for AMI's firmware.
引用於: 第節 第6.4節

S-370 DMTF — All Published Versions of DSP0274 (SPDM)
標準組織 · U.S. · 英文 · 25 May 2025 · 閱讀 26 September 2026
https://www.dmtf.org/sites/default/files/SPDM_1.4_Features_Overview.pdf
“No hybrid support in SPDM 1.4”
支持的研判. DMTF's SPDM 1.4 features overview states 'No hybrid support in SPDM 1.4', the release that adds NIST post-quantum algorithms to SPDM.
無法據以確立. A June 2025 overview: it cannot show version 1.4.1 of 2 July 2026; the 25 May 2025 date comes from DMTF's DSP0274 version list.
引用於: 第節 第6.4節, 第6.7節, 第C.7節; 表 22

S-371 Open Compute Project Foundation — OCP GPU & Accelerator RAS Requirements V1.7
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/documents/ocp-gpu-and-accelerators-ras-requirements-v1-7-10-23-2025-pdf
“Crash dumps must be collected at the time of Uncorrectable, Fatal errors automatically”
支持的研判. OCP's GPU and Accelerator RAS Requirements v1.7, from Microsoft, Google, NVIDIA, Meta and AMD, require error injection, CPER error reporting in-band and out-of-band, and automatic crash dumps at fatal errors.
無法據以確立. Buyer-written requirements; the document prints no publication date (the October 2025 date is from its file name) and does not show supplier compliance.
引用於: 第節 第6.4節, 第6.7節; 表 21

S-372 Open Compute Project Foundation (Hardware Management / RAS API workstream) — RAS API Specification Version 0.9
標準組織 · Global · 英文 · 14 December 2025 · 閱讀 27 September 2026
https://www.opencompute.org/documents/2025-ocp-ras-api-v0-9-final-pdf
“The scope of this document is to define the RAS API data exchange and communication protocol that is transport agnostic.”
“Interim release to reflect change of direction, to give visibility to industry partners.”
支持的研判. The RAS API Specification 0.9 (14 December 2025) is an 'Interim release to reflect change of direction', defining a transport-agnostic RAS data exchange and communication protocol.
無法據以確立. Interim draft below 1.0; it does not show implementations, the new direction's final form, or a release date for the next revision.
引用於: 第節 第6.4節; 表 21

S-373 NIST — NIST Releases First 3 Finalized Post-Quantum Encryption Standards
政府 · U.S. · 英文 · 13 August 2024 · 閱讀 26 September 2026
https://www.nist.gov/news-events/news/2024/08/nist-releases-first-3-finalized-post-quantum-encryption-standards
“These finalized standards are the capstone of NIST's efforts to safeguard our confidential electronic information.”
支持的研判. NIST announced on 13 August 2024 the release of its first three finalized post-quantum encryption standards (FIPS 203, 204 and 205).
無法據以確立. A release announcement; it sets no migration deadlines and says nothing about product validation or adoption.
引用於: 第節 第6.4節, 第C.7節

S-374 U.S. National Security Agency — The Commercial National Security Algorithm Suite 2.0 and Quantum Computing FAQ
政府 · U.S. · 英文 · April 2024 · 閱讀 27 September 2026
https://media.defense.gov/2022/Sep/07/2003071836/-1/-1/1/CSI_CNSA_2.0_FAQ_.PDF
“support and prefer CNSA 2.0 by 2025 where available, exclusively use CNSA 2.0 by 2030”
“Web browsers/servers and cloud services: support and prefer CNSA 2.0 by 2025, exclusively use CNSA 2.0 by 2033.”
支持的研判. NSA's CNSA 2.0 FAQ (April 2024) sets national-security-system deadlines: firmware signing exclusive by 2030, cloud services by 2033; it recommends LMS/XMSS but approves ML-DSA for all signing.
無法據以確立. Applies to U.S. national security systems only; April 2024 guidance predating FIPS 204; does not assess Caliptra or any product.
引用於: 第節 第6.4節, 第9.4節, 第9.5節, 第C.7節; 圖 32

S-375 Federal Register — Securing the Nation Against Advanced Cryptographic Attacks
政府 · U.S. · 英文 · 25 June 2026 · 閱讀 26 September 2026
https://www.federalregister.gov/documents/2026/06/25/2026-12909/securing-the-nation-against-advanced-cryptographic-attacks
“transition all HVAs and high impact systems to use PQC for key establishment by December 31, 2030”
支持的研判. Executive Order 14412 (22 June 2026) requires PQC key establishment for high-value and high-impact federal systems by 31 December 2030, signatures by 2031, and directs a contractor rule.
無法據以確立. Binds federal agencies; the contractor rule is directed, not issued, by the order, and its coverage of firmware is not specified.
引用於: 第節 第6.4節; 圖 32

S-376 Office of Management and Budget — Execution of the Migration to Post-Quantum Cryptography (M-26-15)
政府 · U.S. · 英文 · 24 June 2026 · 閱讀 27 September 2026
https://www.whitehouse.gov/wp-content/uploads/2026/06/M-26-15-Execution-of-the-Migration-to-Post-Quantum-Cryptography.pdf
“Agencies should ensure that any requirements for products used by the agency in those categories include PQC integration.”
支持的研判. OMB memorandum M-26-15 (24 June 2026) says agencies should require PQC integration for products in CISA's post-quantum product categories and must submit PQC migration plans within 120 days.
無法據以確立. Applies to federal agencies' purchases in CISA's categories, not all vendor products; it sets no vendor-side obligation directly.
引用於: 第節 第6.4節

S-377 OpenBMC project — openbmc/bmcweb LICENSE
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://raw.githubusercontent.com/openbmc/bmcweb/master/LICENSE
“Apache License Version 2.0, January 2004”
支持的研判. OpenBMC's bmcweb (its Redfish web server) is licensed under the Apache License, Version 2.0, whose text includes an express patent licence from each contributor.
無法據以確立. Covers bmcweb, not every OpenBMC image (built from components under their own licences); the licence text itself, not any firmware supplier's terms.
引用於: 第節 第6.4節, 第6.6節

S-378 CHIPS Alliance (Linux Foundation) — [Public] Caliptra: A silicon root of trust (DAC 2025)
標準組織 · U.S. · 英文 · 2025 · 閱讀 27 September 2026
https://www.chipsalliance.org/events/DAC_Caliptra_2025.pptx.pdf
“NIST PQC (CNSA 2.0 Compliant)”
支持的研判. CHIPS Alliance's DAC 2025 Caliptra deck lists 'NIST PQC (CNSA 2.0 Compliant)' and 'Dilithium/ML-DSA-87' among Caliptra 2.0's features, presenting its ML-DSA-87 signing as CNSA 2.0 compliant.
無法據以確立. A project's self-description of compliance, not an NSA determination; the recorded README URL does not carry the CNSA claim.
引用於: 第節 第6.4節

S-379 IOWN Global Forum — IOWN Global Forum and Open Compute Project Join Forces to Deliver on the Next Wave of AI
標準組織 · Global · 英文 · 10 February 2026 · 閱讀 27 September 2026
https://iowngf.org/iown-global-forum-and-open-compute-project-join-forces-to-deliver-on-the-next-wave-of-ai
“today announce a cooperative framework, called the AI Computing Continuum, to deliver a seamless computational infrastructure from centralized to edge deployments.”
“The IOWN Global Forum will define the communications architecture by leveraging photonics-based optical communication and device technologies. OCP will drive the specifications for the open hardware required to deliver the infrastructure.”
支持的研判. The IOWN Global Forum and OCP announced on 10 February 2026 an 'AI Computing Continuum' framework: IOWN defines a photonics-based communications architecture, OCP drives open hardware specifications.
無法據以確立. A joint announcement of intent; it names no joint specification, deliverable or date. Published by IOWN GF, a party to it.
引用於: 第節 第6.5節

S-380 OCP Foundation wiki — Server/MHS/Edge-Specs-and-Designs - OpenCompute wiki
標準組織 · Global · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.opencompute.org/w/index.php?title=Server%2FMHS%2FEdge-Specs-and-Designs
“designs and specifications enabling interoperable hardware for edge and telco edge platforms”
“Specification Edge-MHS 0.9 RC1 12/08/2025 Dirk Blevins & Vincent Nguyen”
支持的研判. OCP's Edge-MHS wiki lists release candidates 0.5 RC1, 0.7 RC1 and 0.9 RC1, the latest submitted 8 December 2025, for edge and telco-edge hardware.
無法據以確立. Release candidates only; no accepted Edge-MHS specification, telco adoption or 3GPP/O-RAN linkage is shown.
引用於: 第節 第6.5節

S-381 OCP Foundation wiki — OCP-AI-Native-Wireless-Acess - OpenCompute wiki
標準組織 · Global · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.opencompute.org/w/index.php?title=OCP-AI-Native-Wireless-Acess
“Collaborate and partner with Open Air Interface (OAI), 3GPP, TIP, and Open RAN (O-RAN).”
支持的研判. OCP's AI-Native Wireless Access wiki marks its charter and workstreams 'TBA' and aims to ' collaborate and partner with' OpenAirInterface, 3GPP, TIP and 'Open RAN (O-RAN)'.
無法據以確立. An intention on a wiki page; it records no liaison, agreement or cross-citation with 3GPP or O-RAN, and does not name the O-RAN ALLIANCE.
引用於: 第節 第6.5節

S-382 NVIDIA — NVIDIA and Nokia to Pioneer the AI Platform for 6G - Powering America's Return to Telecommunications Leadership
企業新聞稿 · U.S. · 英文 · 28 October 2025 · 閱讀 26 September 2026
https://nvidianews.nvidia.com/news/nvidia-nokia-ai-telecommunications
“NVIDIA will also invest $1 billion in Nokia at a subscription price of $6.01 per share.”
支持的研判. NVIDIA and Nokia announced (28 October 2025) NVIDIA-powered AI-RAN products for Nokia's RAN portfolio, with NVIDIA to invest US$1 billion in Nokia at US$6.01 per share.
無法據以確立. Partnership announcement; operator field trials were to begin in 2026 and the investment was subject to closing conditions; no deployment volumes.
引用於: 第節 第6.5節

S-383 Nokia — Nokia accelerates AI-RAN adoption as global operators embrace AI-native network evolution on NVIDIA platforms
企業新聞稿 · Finland · 英文 · 16 September 2026 · 閱讀 26 September 2026
https://www.nokia.com/newsroom/nokia-accelerates-ai-ran-adoption-as-global-operators-embrace-ai-native-network-evolution-on-nvidia-platforms
"Nokia's AI-RAN platform has already delivered more than 20% improvements in spectral efficiency"
支持的研判. Nokia said (16 September 2026) eight operators, including Chunghwa Telecom and TPG Telecom, run AI-RAN proofs of concept and live trials on NVIDIA platforms, with over 20% spectral-efficiency gains.
無法據以確立. Vendor claims from trials; the spectral-efficiency gain is Nokia's figure, and trials are not commercial deployments.
引用於: 第節 第6.5節

S-384 3GPP — Release 20
標準組織 · Global · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.3gpp.org/specifications-technologies/releases/release-20
"the final '5G-Advanced only' release"
"Stage-3 (protocol details) is targeted for March 2027, followed by the final ASN.1/OpenAPI freeze in June 2027."
支持的研判. 3GPP's Release 20 page calls it the final '5G-Advanced only' release, targets Stage 3 for March 2027, and lists AI/ML among its 5G-Advanced work areas.
無法據以確立. Targets, not fixed dates; the AI/ML item is a listed work area without detail, and the page says nothing about data-centre hardware.
引用於: 第節 第6.5節

S-385 3GPP — Timeline for Release 21
標準組織 · Global · 英文 · 10 June 2026 · 閱讀 27 September 2026
https://www.3gpp.org/news-events/3gpp-news/rel21-timeline
"3GPP has today set a date for the approval of the first 6G standards from the project."
"Stage-3 freeze: Dec. 2028"
支持的研判. 3GPP's 10 June 2026 news item sets the timeline for Release 21, the first 6G standards, with the Stage-3 freeze in December 2028.
無法據以確立. Planned freeze dates set by 3GPP plenaries; they can slip and do not show product or deployment timing.
引用於: 第節 第6.5節

S-386 O-RAN ALLIANCE — Road to Open, Intelligent 6G: O-RAN's First 6G Workshop Toward a Structured and Collaborative 6G Journey
標準組織 · Global · 英文 · 8 December 2025 · 閱讀 27 September 2026
https://www.o-ran.org/blog/road-to-open-intelligent-6g-o-rans-first-6g-workshop-toward-a-structured-and-collaborative-6g-journey
"O-RAN-R006 will cover studies, aligning with 3GPP Release 20. O-RAN-R007 will address normative work items, aligning with 3GPP Release 21."
支持的研判. O-RAN ALLIANCE's report on its first 6G workshop says O-RAN-R006 will cover studies aligned with 3GPP Release 20 and O-RAN-R007 normative work items aligned with Release 21.
無法據以確立. A blog account of workshop consensus and plans ('will cover'), not a formal release decision document; it shows no 6G specification content.
引用於: 第節 第6.5節

S-387 Open Web Foundation (hosted by OCP) — Open Web Foundation Contributor License Agreement 0.9
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://146a55aca6f00848c565-a7635525d40ac1c70300198708936b4e.ssl.cf1.rackcdn.com/images/ed0befaf86bee2568ad720ff4a9a554d1f4260f7.pdf
"No express or implied patent, trademark, or other property rights are granted under this Agreement"
支持的研判. The OWF Contributor License Agreement 0.9 grants a copyright licence and states that no express or implied patent, trademark or other property rights are granted under it.
無法據以確立. The CLA alone carries no patent commitment; OCP's royalty-free patent terms come from the OWFa 0.9 deemed accepted at Final Specification release under OCP's policy.
引用於: 第節 第6.6節; 表 12

S-388 Open Web Foundation — Open Web Foundation Agreement 0.9
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.openwebfoundation.org/the-agreements/the-owf-0-9-agreements-necessary-claims/open-web-foundation-agreement-0-9
"I agree to grant to you a no charge, royalty free license to my Necessary Claims on reasonable and non-discriminatory terms solely for your Implementation."
"including those claims acquired after the Date below, that are necessary to implement the required portions (including the required elements of optional portions) of the Specification"
支持的研判. OWFa 0.9 lets a signatory grant a no-charge, royalty-free RAND licence to its Necessary Claims, defined to include later-acquired claims necessary to the Specification's required portions.
無法據以確立. Agreement text only; the non-assertion option and the termination-on-suit clause sit in other sections. Says nothing on who signs it.
引用於: 第節 第6.6節, 第6.7節, 第C.7節; 表 12

S-389 Open Compute Project Foundation — Templates & Agreements » Open Compute Project
標準組織 · Global · 英文 · 14 August 2026 · 閱讀 27 September 2026
https://www.opencompute.org/contributions/templates-agreements
"Hardware License OCPHL-R Option (for use only with the OCP CLA)"
支持的研判. OCP's Templates & Agreements page lists the OCPHL-R and OCPHL-P hardware licences under 'Archived / Deprecated', each marked 'for use only with the OCP CLA'.
無法據以確立. Page as read in September 2026; it does not say whether archived licences still govern older contributions.
引用於: 第節 第6.6節

S-390 SONiC project (GitHub sonic-net) — sonic-net/sonic-buildimage LICENSE
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://raw.githubusercontent.com/sonic-net/sonic-buildimage/master/LICENSE
"Licensed under the Apache License, Version 2.0"
支持的研判. The SONiC build-image repository's LICENSE file states that it is licensed under the Apache License, Version 2.0.
無法據以確立. Covers the sonic-buildimage repository; it does not describe SONiC's functions, governance or deployment.
引用於: 第節 第6.6節

S-391 JEDEC Solid State Technology Association — JEDEC Patent Policy
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.jedec.org/about-jedec/patent-policy
"All companies, as a condition of committee membership or participation, agree to license their Essential Patent Claims on RAND terms and conditions."
"No company has a duty to conduct a search for Potentially Essential Patents."
支持的研判. JEDEC's patent policy makes RAND licensing of Essential Patent Claims a condition of committee membership or participation, and imposes no duty to search for Potentially Essential Patents.
無法據以確立. Policy summary page; the disclosure duty for known patents is in the full policy (JM21), not these excerpts. Sets no rate.
引用於: 第節 第6.6節, 第C.7節; 表 12

S-392 IEEE SA — IEEE SA Standards Board Bylaws - Clause 6 Patents
標準組織 · Global · 英文 · 3 March 2025 · 閱讀 26 September 2026
https://standards.ieee.org/about/policies/bylaws/sect6-7
"excluding the value, if any, resulting from the inclusion of that Essential Patent Claim's technology in the IEEE Standard"
"shall neither seek nor seek to enforce a Prohibitive Order based on such Essential Patent Claim(s) in a jurisdiction against an implementer who is willing to negotiate in good faith"
支持的研判. IEEE SA bylaws: an LoA licence is free or at Reasonable Rates excluding the value of inclusion in the standard; no prohibitive orders against implementers willing to negotiate.
無法據以確立. Policy text as of the 3 March 2025 page; it sets no rate and says nothing about specific Ethernet patents or holders.
引用於: 第節 第6.6節; 表 12

S-393 Joint Development Foundation (Linux Foundation) / Ultra Ethernet Consortium — UEC Membership Agreement / Project Charters (JDF 5.0.1 RAND)
標準組織 · U.S. · 英文 · 21 October 2024 · 閱讀 26 September 2026
https://cdn.platform.linuxfoundation.org/agreements/uepf.pdf
"offer to all applicants a nonexclusive, worldwide, non-sublicensable, patent license to its Necessary Claims on fair, reasonable, and non-discriminatory terms"
支持的研判. UEC's JDF charter (RAND mode) requires each working-group participant to offer applicants a nonexclusive, worldwide patent licence to its Necessary Claims on fair, reasonable and non-discriminatory terms.
無法據以確立. Charter terms only; they set no royalty level and do not show how any claim reads on the specification or on ESUN.
引用於: 第節 第6.6節; 表 12

S-394 DMTF — DMTF IPR Policy Statement
標準組織 · U.S. · 英文 · 18 September 2025 · 閱讀 27 September 2026
https://www.dmtf.org/sites/default/files/DMTF_IPR_Policy_Statement_09-18-2025.pdf
"Through this policy, DMTF seeks to ensure that its standards can be implemented on fair, reasonable, and non-discriminatory (RAND) terms, and that the rights of both contributors and implementers are respected."
"Contributors to DMTF standards are expected to disclose any patents of which they are aware that may be essential to the implementation of their submissions."
支持的研判. DMTF's 2025 IPR Policy Statement says DMTF seeks to ensure its standards can be implemented on RAND terms and expects contributors to disclose patents they know may be essential.
無法據以確立. A statement of aim and expectation; it sets no licensing commitment itself, which rests on DMTF's 2001 patent policy (S-400).
引用於: 第節 第6.6節; 表 12

S-395 SNIA — Storage Networking Industry Association (SNIA) Intellectual Property Policy v4.0
標準組織 · U.S. · 英文 · 18 September 2015 · 閱讀 26 September 2026
https://www.snia.org/sites/default/files/corp-docs/SNIA_IP_Policy_v4_0_9_18_2015-final.pdf
"must be on reasonable and nondiscriminatory terms and conditions; (c) may be granted either for or without compensation"
支持的研判. SNIA's IP Policy v4.0 (2015) requires essential-claim licences to be on reasonable and nondiscriminatory terms and allows them to be granted either for or without compensation.
無法據以確立. A 2015 policy; 'nonexclusive, worldwide' is elsewhere in the clause, not this excerpt. It sets no rate and names no SNIA–OCP deliverable.
引用於: 第節 第6.6節; 表 12

S-396 Ultra Accelerator Link Consortium, Inc. — Join Us - UALink Consortium
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://ualinkconsortium.org/join-ualink
“Implement UALink technology within the IP protections as outlined in the Participation Agreement”
支持的研判. The UALink Consortium's join page offers members the right to 'Implement UALink technology within the IP protections as outlined in the Participation Agreement'.
無法據以確立. The page does not state the IP policy's licensing mode (RAND or royalty-free); that policy sits in documents not read here.
引用於: 第節 第6.6節; 表 12

S-397 Compute Express Link Consortium, Inc. — Join CXL
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://computeexpresslink.org/join
“Implement with the IP protections as outlined in the Agreements”
支持的研判. The CXL join page lists, as a member benefit, the right to 'Implement with the IP protections as outlined in the Agreements', without stating patent terms.
無法據以確立. The page does not state the IP policy's licensing mode; the policy text sits in member agreements not read here.
引用於: 第節 第6.6節; 表 12

S-398 Universal Chiplet Interconnect Express, Inc. — Join UCIe
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.uciexpress.org/join
“Review and submit a signed Participation Agreement”
支持的研判. UCIe's join page asks applicants to review and submit a signed Participation Agreement and lists an IPR Policy among the governing documents, without stating its patent terms.
無法據以確立. Membership page only; the IPR Policy's licensing mode is in governing documents not read here.
引用於: 第節 第6.6節

S-399 CXL Consortium — CXL Specification (Evaluation Copy Agreement)
標準組織 · U.S. · 英文 · 12 February 2026 · 閱讀 26 September 2026
https://computeexpresslink.org/cxl-specification
“nothing in this Agreement shall be deemed as granting (either expressly, implied or by estoppel) to the End User any kind of license to implement the CXL Specification”
支持的研判. The CXL Consortium's public evaluation-copy agreement (12 February 2026) states that it grants the end user no licence of any kind to implement the CXL Specification.
無法據以確立. Covers CXL's evaluation copy only, not UALink's or UCIe's documents, and not the patent terms in CXL's member IP policy.
引用於: 第節 第6.6節, 第A.2節; 表 12

S-400 DMTF — DMTF Patent Policy; DMTF IPR Policy Statement (09-18-2025)
標準組織 · U.S. · 英文 · 17 October 2001 · 閱讀 26 September 2026
https://www.dmtf.org/sites/default/files/patent-10-18-01.pdf
“non-discriminatory basis either royalty-free or at a fair and reasonable royalty rate”
支持的研判. DMTF's 2001 patent policy asks holders of essential patents to license on a non-discriminatory basis either royalty-free or at a fair and reasonable royalty rate.
無法據以確立. A 2001 policy, still listed by DMTF; it sets no rate, and the 2025 statement (S-394) frames licensing as RAND.
引用於: 第節 第6.6節; 表 12

S-401 ParTec AG via EQS News — ParTec and BF exaQC Take Note of Largely Favorable but Dismissive First-Instance Ruling. A Pyrrhic Victory for NVIDIA?!
企業新聞稿 · Germany · 英文 · 11 March 2026 · 閱讀 26 September 2026
https://www.eqs-news.com/de/news/corporate/partec-and-bf-exaqc-take-note-of-largely-favorable-but-dismissive-first-instance-ruling-a-pyrrhic-victory-for-nvidia/5d907811-53a1-4ae9-9f1d-e0892cb64975_en
“dismissed the patent infringement action brought by ParTec AG”
支持的研判. ParTec says the UPC Munich Local Division dismissed at first instance its infringement action against NVIDIA on EP 3 743 812 on 11 March 2026, and that it will appeal.
無法據以確立. A party's own account of the ruling; not the court's decision text, and it names no OCP specification.
引用於: 第節 第6.6節

S-402 Netlist, Inc. (PRNewswire) — U.S. International Trade Commission Votes to Institute Investigation Into Micron Technology, Inc.
企業新聞稿 · U.S. · 英文 · 24 September 2026 · 閱讀 27 September 2026
https://investors.netlist.com/news/us-international-trade-commission-votes-to-institute-investigation-into-micron-technology-inc/270dd13b-f73f-4be8-a85a-13b5b56ecba8
“The ITC will investigate whether the Respondents infringe Netlist's U.S. Patent Nos. 10,025,731, 10,217,523, 12,373,366 and 12,675,407.”
“These patents read on DDR5 RDIMMs and/or MRDIMMs.”
支持的研判. Netlist says the ITC instituted an investigation (September 2026) into whether Micron and other respondents infringe four Netlist patents, which Netlist says read on DDR5 RDIMMs and/or MRDIMMs.
無法據以確立. The complainant's announcement; institution is not a finding of infringement, and the release is not the ITC's own notice.
引用於: 第節 第6.6節

S-403 Netlist, Inc. (PRNewswire) — Netlist Announces Strategic Alliance with Samsung For Advanced Memory Technology
企業新聞稿 · U.S. · 英文 · 5 August 2026 · 閱讀 26 September 2026
https://investors.netlist.com/news/netlist-announces-strategic-alliance-with-samsung-for-advanced-memory-technology/1c8e5ae6-bde4-434c-afd8-7d80370824af
“the parties agree to settle and mutually release all pending legal actions”
支持的研判. Netlist announced on 5 August 2026 five-year agreements with Samsung, including a patent cross-licence, under which the parties 'settle and mutually release all pending legal actions'.
無法據以確立. Netlist's announcement; it does not list which cases are resolved or disclose payment terms.
引用於: 第節 第6.6節

S-404 U.S. Department of Justice — DAAG Dina Kallay Delivers Keynote at Concurrences Dinner in New York
政府 · U.S. · 英文 · 19 September 2025 · 閱讀 26 September 2026
https://www.justice.gov/opa/speech/daag-dina-kallay-delivers-keynote-concurrences-dinner-new-york
“there are private consortia that impose mandatory, royalty-free cross-licensing obligations on their members.”
支持的研判. A DOJ Antitrust Division deputy assistant attorney general criticised, on 19 September 2025, private consortia that impose mandatory, royalty-free cross-licensing obligations on their members, naming none.
無法據以確立. A speech, not an enforcement action or agency guidance; it names no consortium and does not address OCP.
引用於: 第節 第6.6節

S-405 CHIPS Alliance — CHIPS Alliance FAQ
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.chipsalliance.org/about/faq
“Typically, CHIPS Alliance projects are covered by the Apache License 2.0 for code, and OWFa 0.9 for specs.”
支持的研判. The CHIPS Alliance FAQ says its projects are typically covered by the Apache License 2.0 for code and OWFa 0.9 for specifications.
無法據以確立. 'Typically' is a general statement, not Caliptra's own licence record; Caliptra specifications are published through OCP.
引用於: 表 12

S-406 Open Compute Project — Optical Circuit Switching (OCP community project page)
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/community/optical-circuit-switching
“This subproject is dedicated to developing guiding principles, open standards, and a fully open network technology stack for OCS.”
支持的研判. OCP says its optical circuit switching subproject is dedicated to developing guiding principles, open standards and a fully open network technology stack for OCS.
無法據以確立. An undated project page stating aims; it lists no released OCS specification and says nothing of Google's in-house deployment (S-293).
引用於: 第節 第6.7節; 圖 26

S-407 Hasan Al Maruf, Hao Wang, Abhishek Dhanotia, Johannes Weiner, Niket Agarwal, Pallab Bhattacharya, Chris Petersen, Mosharaf Chowdhury, Shobhit Kanaujia, Prakash Chauhan — TPP: Transparent Page Placement for CXL-Enabled Tiered-Memory
研究 · U.S. · 英文 · 6 June 2022 · 閱讀 27 September 2026
https://arxiv.org/abs/2206.02878
“We propose a novel OS-level application-transparent page placement mechanism (TPP) for CXL-enabled memory.”
“TPP makes a tiered memory system performant as an ideal baseline (<1% gap) that has all the memory in the local tier.”
支持的研判. TPP, published in 2022, is an OS-level transparent page-placement mechanism for CXL-enabled tiered memory, reported within 1% of an ideal baseline and 18% better than Linux.
無法據以確立. Shows the technique was public in 2022; it is not a patent or freedom-to-operate analysis.
引用於: 第節 第6.7節

S-408 Open Compute Project Foundation — Server/OCE/FCSA - OpenCompute wiki
標準組織 · Global · 英文 · 21 September 2026 · 閱讀 27 September 2026
https://www.opencompute.org/wiki/Server/OCE/FCSA
取自來源表格之數值（轉錄）： FCSA 1.1.0 RC0 | Document | NA | For review | 2026-09-14
支持的研判. OCP's FCSA wiki lists FCSA 1.0.0 as released on 12 February 2026 and FCSA 1.1.0 at release candidate RC0, dated 14 September 2026 and marked 'For review'.
無法據以確立. Draft status only; it gives no release date for 1.1.0 and says nothing on when contributors' patent commitments attach (S-047).
引用於: 第節 第6.7節

S-409 Federal Register (BIS, U.S. Department of Commerce) — Framework for Artificial Intelligence Diffusion
監管機關 · U.S. · 英文 · 15 January 2025 · 閱讀 27 September 2026
https://www.federalregister.gov/documents/2025/01/15/2025-00636/framework-for-artificial-intelligence-diffusion
“establishes a worldwide license requirement for ECCNs 3A090.a, 4A090.a, and corresponding .z items in new § 742.6(a)(6)(iii)(A)”
“imposing new controls on the model weights of the most advanced AI models”
支持的研判. BIS's interim final rule of 15 January 2025 imposed a worldwide licence requirement for ECCN 3A090.a and 4A090.a items, an exception for 18 destinations, and controls on advanced closed-model weights.
無法據以確立. A rule later announced for rescission (13 May 2025); weight controls applied only to closed models trained above 10^26 operations.
引用於: 第節 第7.1節; 表 15

S-410 Bureau of Industry and Security — Department of Commerce Announces Rescission of Biden-Era Artificial Intelligence Diffusion Rule
監管機關 · U.S. · 英文 · 13 May 2025 · 閱讀 27 September 2026
https://www.bis.gov/press-release/department-commerce-announces-rescission-biden-era-artificial-intelligence-diffusion-rule-strengthens
“Today, the Department of Commerce (DOC) announced a rescission of the Biden Administration's AI Diffusion Rule”
“Issuing guidance alerting industry to the risks of using PRC advanced computing ICs, including specific Huawei Ascend chips.”
支持的研判. Commerce announced on 13 May 2025 the rescission of the AI Diffusion Rule and issued guidance on the risks of using PRC advanced computing ICs, including specific Huawei Ascend chips.
無法據以確立. An announcement of rescission with guidance; formal rule text and any replacement are separate documents not shown here.
引用於: 第節 第7.1節; 表 15

S-411 Global Trade & Sanctions Law (blog) — Reported Draft Rules Signal New Semiconductor Export Controls Framework
法律評論 · U.S. · 英文 · 13 March 2026 · 閱讀 26 September 2026
https://www.globaltradeandsanctionslaw.com/reported-draft-rules-signal-new-semiconductor-export-controls-framework
“a copy of the draft regulations is not publicly available”
支持的研判. Trade-law commentators, summarising March 2026 news reports of a draft BIS successor to the AI Diffusion Rule, note that a copy of the draft regulations is not publicly available.
無法據以確立. Establishes non-publication only as of 13 March 2026; it relies on press reports and does not show the draft's status at the cut-off.
引用於: 第節 第7.1節; 表 15

S-412 Federal Register (BIS) — Revision to License Review Policy for Advanced Computing Commodities
監管機關 · U.S. · 英文 · 15 January 2026 · 閱讀 27 September 2026
https://www.federalregister.gov/documents/2026/01/15/2026-00789/revision-to-license-review-policy-for-advanced-computing-commodities
“Copackaged DRAM ICs include, for example, high bandwidth memory (HBM).”
“There is a case-by-case license review policy for license applications for exports of commodities with a TPP less than 21,000, and a 'total DRAM bandwidth' less than 6,500 GB/s”
“The semiconductors covered by this rule are the Nvidia H200 and its equivalents, as well as less advanced chips”
支持的研判. BIS's rule, effective 15 January 2026, sets case-by-case review for China or Macau exports of chips below TPP 21,000 and 6,500 GB/s DRAM bandwidth, such as H200 or MI325X.
無法據以確立. Licence policy, not approvals; conditions include U.S. supply and foundry-diversion certifications, and it covers no chips above those thresholds.
引用於: 第節 第7.1節, 第8.5節, 第8.6節, 第C.7節; 表 15; 展望 6, 18, 19 · **研判:** A-08

S-413 Federal Register (Executive Office of the President) — Adjusting Imports of Semiconductors, Semiconductor Manufacturing Equipment, and Their Derivative Products Into the United States (Proclamation 11002)
政府 · U.S. · 英文 · 20 January 2026 · 閱讀 27 September 2026
https://www.federalregister.gov/documents/2026/01/20/2026-01052/adjusting-imports-of-semiconductors-semiconductor-manufacturing-equipment-and-their-derivative-
“imports of Covered Products will be subject to a 25 percent ad valorem duty rate”
“not apply to imports of those Covered Products for use in United States data centers”
支持的研判. Proclamation 11002, effective 15 January 2026, imposes a 25% Section 232 duty on certain advanced computing chips and derivatives, exempting imports for U.S. data centres and other listed uses.
無法據以確立. Covers only products in its annex, with broad exemptions; any broader semiconductor tariff was left for later decision.
引用於: 第節 第7.1節; 表 15

S-414 Federal Register (BIS) — One Year Suspension of Expansion of End-User Controls for Affiliates of Certain Listed Entities
監管機關 · U.S. · 英文 · 12 November 2025 · 閱讀 27 September 2026
https://www.federalregister.gov/documents/2025/11/12/2025-19846/one-year-suspension-of-expansion-of-end-user-controls-for-affiliates-of-certain-listed-entities
“The suspension is set to end November 9, 2026, absent a future extension.”
“In the second phase of this final rule, effective November 10, 2026 and extending indefinitely, the changes included in the Affiliates Rule that are removed in the first stage will be added back into the EAR.”
支持的研判. BIS suspended the 30 September 2025 Affiliates rule (50%-owned affiliates of listed entities) from 10 November 2025 to 9 November 2026; absent extension, its changes return from 10 November 2026.
無法據以確立. Sets dates only; it does not say whether BIS will extend the suspension or how affiliates will be identified in practice.
引用於: 第節 第7.1節, 第7.5節, 第9.2節, 第9.6節, 第C.7節; 表 15; 展望 6, 18; 圖 32

S-415 The White House — Fact Sheet: President Donald J. Trump Strikes Deal on Economic and Trade Relations with China
政府 · U.S. · 英文 · 1 November 2025 · 閱讀 27 September 2026
https://www.whitehouse.gov/fact-sheets/2025/11/fact-sheet-president-donald-j-trump-strikes-deal-on-economic-and-trade-relations-with-china
“China will suspend the global implementation of the expansive new export controls on rare earths and related measures that it announced on October 9, 2025.”
“China will issue general licenses valid for exports of rare earths, gallium, germanium, antimony, and graphite for the benefit of U.S. end users and their suppliers around the world.”
支持的研判. The 1 November 2025 White House fact sheet says China will suspend its 9 October 2025 rare-earth controls and issue general licences, while U.S. suspensions run until 10 November 2026.
無法據以確立. U.S. account of commitments; it states no end date for China's rare-earth suspension or general licences, and no implementation record.
引用於: 第節 第7.1節, 第7.5節, 第9.2節, 第9.6節; 表 15; 展望 6, 18; 圖 32

S-416 South China Morning Post — Nvidia ships first H200 chips to China, but forecasts no data-centre computing revenue
媒體 · Hong Kong · 英文 · 27 August 2026 · 閱讀 26 September 2026
https://www.scmp.com/tech/big-tech/article/3365383/nvidia-ships-first-h200s-china-forecasts-no-data-centre-computing-revenue
“restricted purchases as it promoted home-grown alternatives”
支持的研判. SCMP reported in August 2026 that Beijing restricted purchases of NVIDIA's H200 as it promoted home-grown alternatives, while NVIDIA shipped its first H200s to China.
無法據以確立. Press account; the excerpt lacks its subject, and the article also reports first H200 shipments, so the restriction was not a full block.
引用於: 第節 第7.1節

S-417 Federal Register (International Trade Administration) — American AI Exports Program; Call for Proposals for Pre-Set Consortia
政府 · U.S. · 英文 · 10 April 2026 · 閱讀 26 September 2026
https://www.federalregister.gov/documents/2026/04/10/2026-06952/american-ai-exports-program-call-for-proposals-for-pre-set-consortia
“AI-optimized hardware and related infrastructure, including chips, servers, accelerators, data center storage, cloud services, networking”
支持的研判. ITA's 10 April 2026 notice under EO 14320 calls for pre-set consortia proposals for full-stack AI export packages whose hardware layer includes chips, servers, accelerators, storage, cloud services and networking.
無法據以確立. A call for proposals (due 30 June 2026), not designations; packages also cover data, models, security and applications.
引用於: 第節 第7.1節; 表 15

S-418 Simpson Thacher & Bartlett — U.S. Commerce Department Invites Proposals for American AI Exports Program Excluding PRC Companies and Technologies
法律評論 · U.S. · 英文 · 16 April 2026 · 閱讀 26 September 2026
https://www.stblaw.com/about-us/publications/view/2026/04/16/u.s.-commerce-department-invites-proposals-for-american-ai-exports-program-excluding-prc-companies-and-technologies
“The consortium must exclude open-weight models developed by PRC Companies”
支持的研判. Simpson Thacher summarises the American AI Exports Program call: PRC companies barred from consortium roles, PRC open-weight models excluded, and at least 51% U.S. content by value in hardware.
無法據以確立. Law-firm summary of a call for proposals; the ITA notice governs, and the captured excerpt shows only the open-weight exclusion.
引用於: 第節 第7.1節; 表 15

S-419 The White House — Winning the Race: America's AI Action Plan
政府 · U.S. · 英文 · July 2025 · 閱讀 26 September 2026
https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf
“The United States must meet global demand for AI by exporting its full AI technology stack—hardware, models, software, applications, and standards.”
支持的研判. America's AI Action Plan (July 2025) says the United States must meet global AI demand by exporting its full AI technology stack: hardware, models, software, applications and standards.
無法據以確立. A policy plan, not a binding rule; implementation runs through later orders and programmes.
引用於: 第節 第7.1節

S-420 The AI Policy Network — AIPN Urges Conferees to Preserve Strong AI Provisions in FY27 NDAA
法律評論 · U.S. · 英文 · 31 August 2026 · 閱讀 26 September 2026
https://theaipn.org/aipn-fy27-ndaa-conference-letter
“preserve the work you have done to protect America's technological edge”
支持的研判. An AI Policy Network letter of 31 August 2026 urges FY2027 NDAA conferees to keep Senate chip provisions: the AI OVERWATCH Act, the MATCH Act and the Chip Security Act.
無法據以確立. An advocacy group's letter; it describes pending Senate amendments, not enacted law or the conference outcome.
引用於: 第節 第7.1節; 展望 6

S-421 Congressional Research Service (via EveryCRSReport) — FY2027 NDAA: Status of Legislative Activity (CRS Insight IN12704)
政府 · U.S. · 英文 · 1 September 2026 · 閱讀 26 September 2026
https://www.everycrsreport.com/reports/IN12704.html
“On July 14, cloture failed (50-46) on the motion to proceed to S. 4784.”
支持的研判. CRS records Senate cloture on the FY2027 NDAA motion to proceed failing 50–46 on 14 July 2026, House passage 216–212 on 22 July, and no enactment by 1 September.
無法據以確立. Status as of 1 September 2026 only; it does not show the content of any final conference text.
引用於: 第節 第7.1節

S-422 eCFR (Office of the Federal Register) — 15 CFR 734.3 — Items subject to the EAR (with 15 CFR 734.7 ' Published', https://www.ecfr.gov/current/title-15/subtitle-B/chapter-VII/subchapter-C/part-734/section-734.7)
監管機關 · U.S. · 英文 · 22 September 2026 · 閱讀 26 September 2026
https://www.ecfr.gov/current/title-15/subtitle-B/chapter-VII/subchapter-C/part-734/section-734.3
“Are published, as described in § 734.7”
“or are otherwise patent information or are for a standards-related activity as described in § 734.10”
支持的研判. 15 CFR 734.3(b)(3) lists information and software that are published (734.7) or are for a standards-related activity (734.10) as not subject to the EAR.
無法據以確立. The exclusion covers information and software, not hardware built to them, and depends on the conditions in 734.7 and 734.10.
引用於: 第節 第7.1節; 表 15

S-423 eCFR (Office of the Federal Register) — 15 CFR 734.7 - Published
監管機關 · U.S. · 英文 · 22 September 2026 · 閱讀 26 September 2026
https://www.ecfr.gov/current/title-15/subtitle-B/chapter-VII/subchapter-C/part-734/section-734.7
“made available to the public without restrictions upon its further dissemination”
“Published encryption software classified under ECCN 5D002 remains subject to the EAR unless”
支持的研判. 15 CFR 734.7 treats technology or software made available to the public without dissemination restrictions as published, but keeps published ECCN 5D002 encryption software subject unless conditions are met.
無法據以確立. Defines 'published' only; non-public exchanges may still be excluded under 734.10 (standards) or 734.8 (fundamental research).
引用於: 第節 第7.1節; 表 15

S-424 eCFR (Office of the Federal Register) — 15 CFR 734.10 - Patents and standards-related activity
監管機關 · U.S. · 英文 · 18 September 2026 · 閱讀 27 September 2026
https://www.ecfr.gov/current/title-15/subtitle-B/chapter-VII/subchapter-C/part-734/section-734.10
“"technology" or "software" is not subject to the EAR provided it meets at least one condition in both paragraphs (b)(1) and (2) of this section”
“Occurs with the intent that the resulting standard will be 'published.'”
支持的研判. 15 CFR 734.10(b) excludes from the EAR technology or software released for standards work if it is of listed classifications and the standard is published or intended for publication.
無法據以確立. Regulatory text as of the eCFR date; classification of any item, and whether an activity qualifies, need case-by-case review.
引用於: 第節 第7.1節; 表 15

S-425 Federal Register (BIS) — Additions and Revisions of Entities to the Entity List
監管機關 · U.S. · 英文 · 6 March 2023 · 閱讀 26 September 2026
https://www.federalregister.gov/documents/2023/03/06/2023-04558/additions-and-revisions-of-entities-to-the-entity-list
“acquiring and attempting to acquire U.S.-origin items in support of the China's military modernization efforts”
支持的研判. BIS's March 2023 Entity List rule (published 6 March) added entities, including Inspur Group, for acquiring U.S.-origin items in support of China's military modernisation.
無法據以確立. Lists named parties only; it does not by itself reach unlisted affiliates or published technical information.
引用於: 第節 第7.1節

S-426 Federal Register (BIS) — Additions to the Entity List
監管機關 · U.S. · 英文 · 28 March 2025 · 閱讀 26 September 2026
https://www.federalregister.gov/documents/2025/03/28/2025-05427/additions-to-the-entity-list
“for their contributions to Inspur's development of supercomputers for military end use”
支持的研判. BIS added six Inspur subsidiaries, including Inspur Taiwan, to the Entity List in March 2025 for their contributions to Inspur's development of supercomputers for military end use.
無法據以確立. Covers named parties only; it does not by itself reach unlisted affiliates or published specifications.
引用於: 第節 第7.1節

S-427 Federal Register (Executive Office of the President) — Adjusting Imports of Polysilicon and Its Derivatives Into the United States (Proclamation 11052)
政府 · U.S. · 英文 · 11 August 2026 · 閱讀 27 September 2026
https://www.federalregister.gov/documents/2026/08/11/2026-16400/adjusting-imports-of-polysilicon-and-its-derivatives-into-the-united-states
“Effective with respect to goods entered for consumption, or withdrawn from warehouse for consumption, on or after 12:01 a.m. eastern time on December 4, 2026”
“The applicable minimum import prices for imported polysilicon and polysilicon derivatives shall be: (i) $21 per kilogram for polysilicon; (ii) $100 per kilogram for polysilicon ingots and wafers”
支持的研判. Proclamation 11052 imposes a 15% Section 232 duty and minimum import prices on polysilicon and derivatives, including wafers, from 4 December 2026, capping combined rates at 15% for listed partners.
無法據以確立. Tariff terms only; offsets for approved onshoring plans may apply, and it does not show effects on prices or supply.
引用於: 表 15; 第節 第7.1節

S-428 The White House — Fact Sheet: President Donald J. Trump Advances a Fair and Reciprocal Relationship with China While Hosting Historic State Visit
政府 · U.S. · 英文 · 25 September 2026 · 閱讀 27 September 2026
https://www.whitehouse.gov/fact-sheets/2026/09/fact-sheet-president-donald-j-trump-advances-a-fair-and-reciprocal-relationship-with-china-while-hosting-historic-state-visit
“established the U.S.-China Super Intelligence (SI) Dialogue to exchange views on risks and benefits related to SI”
“The next exchange will occur by November 2026.”
支持的研判. The 25 September 2026 fact sheet records a new U.S.–China Super Intelligence Dialogue, next exchange by November 2026, and does not mention semiconductors or extending the 2025 arrangement.
無法據以確立. Silence in a fact sheet does not show that no extension was agreed or will be; it covers one state visit.
引用於: 第節 第7.1節; 展望 6, 18

S-429 The White House — Accelerating Federal Permitting of Data Center Infrastructure
政府 · U.S. · 英文 · 23 July 2025 · 閱讀 27 September 2026
https://www.whitehouse.gov/presidential-actions/2025/07/accelerating-federal-permitting-of-data-center-infrastructure
“a facility that requires greater than 100 megawatts (MW) of new load dedicated to AI inference, training, simulation, or synthetic data generation”
“a Data Center Project or Covered Component Project for which the Project Sponsor has committed at least $500 million in capital expenditures as determined by the Secretary of Commerce”
支持的研判. Executive Order 14318 (July 2025) speeds federal permitting for qualifying data-centre projects: more than 100 MW of new AI load, or at least US$500 million committed capital.
無法據以確立. Directs federal agencies on permitting; it does not change grid interconnection rules or guarantee project approval or timing.
引用於: 第節 第7.2節

S-430 Federal Energy Regulatory Commission — FERC RM26-4 docket page
監管機關 · U.S. · 英文 · 22 July 2026 · 閱讀 26 September 2026
https://www.ferc.gov/rm26-4
“electricity demand greater than 20 megawatts (MW)”
支持的研判. FERC's RM26-4 page records the Energy Secretary's October 2025 direction to consider an advance notice on interconnecting loads above 20 MW, still at that stage on 22 July 2026.
無法據以確立. A docket page updated 22 July 2026; it does not show later action or the content of any final rule.
引用於: 第節 第7.2節, 第C.7節; 展望 8, 17

S-431 Federal Energy Regulatory Commission (FERC) — Summaries | July 2026 Commission Meeting
監管機關 · U.S. · 英文 · 16 July 2026 · 閱讀 27 September 2026
https://ferc.gov/news-events/news/summaries-july-2026-commission-meeting
“The order directs North American Electric Reliability Corporation (NERC), pursuant to section 215(d)(5) of the Federal Power Act, to develop and submit new or modified Reliability Standards addressing the reliability risks associated with the integration of computational loads into the Bulk-Power System.”
“Both The Reliability Standards and Rules of Procedure revisions are due by December 31, 2026.”
支持的研判. FERC's 16 July 2026 order (RD26-7-000) directs NERC to develop Reliability Standards for computational-load integration risks, with standards and Rules of Procedure revisions due by 31 December 2026.
無法據以確立. A meeting summary of a directive; it does not give the standards' content, and FERC approval comes later.
引用於: 第節 第7.2節

S-432 Nicholas Institute for Energy, Environment & Sustainability, Duke University — Rethinking Load Growth: Assessing the Potential for Integration of Large Flexible Loads in US Power Systems
研究 · U.S. · 英文 · 11 February 2025 · 閱讀 26 September 2026
https://nicholasinstitute.duke.edu/publications/rethinking-load-growth
“76GW of new load—equivalent to 10% of the nation's current aggregate peak demand— could be integrated with an average annual load curtailment rate of 0.25%”
支持的研判. Duke's Nicholas Institute (February 2025) estimates 76 GW of new load, about 10% of U.S. aggregate peak demand, could be integrated with average annual load curtailment of 0.25%.
無法據以確立. A modelling estimate that assumes new loads accept curtailment; it is not a utility or operator commitment.
引用於: 第節 第7.2節, 第7.3節

S-433 SSRN preprint (listed on Duke Scholars) — Estimating headroom and demand flexibility requirements for fast large-load integration: A PJM case study
研究 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://doi.org/10.2139/ssrn.7416921
“the 2023 PJM system could accommodate up to 32 GW of new flexible loads while requiring curtailment of only 0.1% of their annual energy consumption”
支持的研判. A 2026 preprint estimates the 2023 PJM system could accommodate up to 32 GW of new flexible load with curtailment of only 0.1% of that load's annual energy.
無法據以確立. A modelling study on 2023 system data, not peer reviewed as listed; it assumes loads accept curtailment and is not a PJM finding.
引用於: 第節 第7.2節, 第7.3節

S-434 Commission for Regulation of Utilities (CRU), Ireland — The CRU Publishes its Decision on New Electricity Connection Policy for Data Centres
監管機關 · Ireland · 英文 · 12 December 2025 · 閱讀 27 September 2026
https://www.cru.ie/about-us/news/the-cru-publishes-its-decision-on-new-electricity-connection-policy-for-data-centres/
“New data centres connecting to the electricity network will continue to be required to provide generation and/or storage capacity, either onsite or local in proximity to match the requested data centre maximum import demand capacity.”
“Data centres will be required to meet at least 80% of their annual demand with additional renewable electricity projects generating in the Republic of Ireland.”
支持的研判. Ireland's CRU decided in December 2025 that new data centres must bring local generation or storage matching their import capacity and meet 80% of annual demand from additional Irish renewables.
無法據以確立. A regulatory decision; it gives no connection lead times and does not show which sites have connected under it.
引用於: 第節 第7.2節

S-435 European Commission (DG ENER) — Energy performance of data centres (DG ENER web page)
政府 · EU · 英文 · 未註明日期 · 閱讀 26 September 2026
https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficiency-targets-directive-and-rules/energy-efficiency-directive/energy-performance-data-centres_en
“collects and publishes data, which is relevant for the energy performance and water footprint of data centres with a significant energy consumption”
支持的研判. DG ENER's data-centre page says the European database collects and publishes data relevant to the energy performance and water footprint of data centres with significant energy consumption.
無法據以確立. Describes the reporting scheme; the 500 kW threshold and indicator list come from the delegated regulation (S-436), not this excerpt.
引用於: 第節 第7.2節

S-436 White & Case LLP — Data centres and energy consumption: evolving EU regulatory landscape and outlook for 2026
法律評論 · U.S. · 英文 · 20 October 2025 · 閱讀 26 September 2026
https://www.whitecase.com/insight-alert/data-centres-and-energy-consumption-evolving-eu-regulatory-landscape-and-outlook-2026
“installed information technology power demand of at least 500 kW”
支持的研判. White & Case notes the recast Energy Efficiency Directive's data-centre reporting obligation applies to sites with installed IT power demand of at least 500 kW.
無法據以確立. Law-firm summary (October 2025); the Directive and Delegated Regulation (EU) 2024/1364 govern which energy and water indicators are reported.
引用於: 第節 第7.2節

S-437 European Commission — Making data centres energy efficient thanks to a new EU rating system
政府 · EU · 英文 · 21 September 2026 · 閱讀 26 September 2026
https://commission.europa.eu/news-and-media/news/making-data-centres-energy-efficient-thanks-new-eu-rating-system-2026-09-21_en
“The first sustainability labels for individual data centres are expected to be displayed in 2027.”
“subject to 2-month scrutiny period by the European Parliament and the Council before entering into force”
支持的研判. The Commission (21 September 2026) says its adopted data-centre rating scheme faces two months' scrutiny, expects first labels in 2027 and data centres above 3% of EU electricity by 2030.
無法據以確立. Commission expectations and a delegated act not yet in force; the 3% figure is a Commission projection, not measured use.
引用於: 第節 第7.2節, 第9.2節; 圖 32

S-438 Infocomm Media Development Authority (IMDA) — Launch of second Data Centre - Call for Application
監管機關 · Singapore · 英文 · 1 December 2025 · 閱讀 26 September 2026
https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/factsheets/2025/launch-of-second-data-centre
“1.25 (at 100% IT load) or better”
支持的研判. IMDA's second Data Centre Call for Application, launched 1 December 2025 for at least 200 MW, requires a PUE of 1.25 (at 100% IT load) or better.
無法據以確立. Conditions for applicants to one allocation round; it does not show which sites were awarded or achieved PUE in operation.
引用於: 第節 第7.2節, 第C.7節

S-439 National Development and Reform Commission (NDRC) — 关于印发《数据中心绿色低碳发展专项行动计划》的通知 (attachment PDF)
政府 · China · 簡體中文 · 23 July 2024 · 閱讀 26 September 2026
https://www.ndrc.gov.cn/xxgk/zcfb/tz/202407/P020240723625582947550.pdf
“新建大型和超大型数据中心应优先布局在全国一体化算力网络国家枢纽节点数据中心集群范围内”
APEX譯文: New large and extra-large data centres should be preferentially located within the data-centre clusters of the national hub nodes of the national integrated computing-power network.
“国家枢纽节点数据中心项目电能利用效率不得高于 1.2”
APEX譯文: The power usage effectiveness of national hub-node data-centre projects shall not be higher than 1.2.
支持的研判. China's July 2024 special action plan for green, low-carbon data centres directs new large data centres preferentially to national hub-node clusters and caps PUE at 1.2 for hub-node projects.
無法據以確立. A policy plan jointly issued by four agencies with end-2025 targets; it does not show enforcement or achieved PUE.
引用於: 第節 第7.2節

S-440 Agency for Natural Resources and Energy (METI) — ANRE special feature on data-center energy efficiency (2026)
政府 · Japan · 日文 · 20 May 2026 · 閱讀 26 September 2026
https://www.enecho.meti.go.jp/about/special/johoteikyo/data_center2026.html
“2029年度以降に新設するデータセンターについては、満たすべきエネルギー効率基準（PUE）を1.3以下にすることを求めます”
APEX譯文: For data centres newly established from fiscal year 2029 onward, we require that the energy-efficiency standard (PUE) they must meet be 1.3 or lower.
支持的研判. Japan's Agency for Natural Resources and Energy says data centres newly built from fiscal 2029 will be required to meet a PUE of 1.3 or lower.
無法據以確立. An efficiency standard under energy-conservation law, not a permit or approval condition for building capacity; existing sites face other benchmarks.
引用於: 第節 第7.2節, 第C.7節

S-441 Future of Privacy Forum — South Korea's New AI Framework Act: A Balancing Act Between Innovation and Regulation
法律評論 · U.S. · 英文 · 18 April 2025 · 閱讀 26 September 2026
https://fpf.org/blog/south-koreas-new-ai-framework-act-a-balancing-act-between-innovation-and-regulation
“administrative and financial support to facilitate the construction and operation of data centers”
支持的研判. The Future of Privacy Forum's April 2025 analysis says Korea's AI Framework Act provides administrative and financial support for building and operating data centres.
無法據以確立. Secondary legal commentary written before the Act took effect; it cannot confirm the 22 January 2026 start or implementing decrees.
引用於: 第節 第7.2節

S-442 Lawrence Berkeley National Laboratory (LBNL) — United States Data Center Energy Usage Report: 2025 Update
研究 · U.S. · 英文 · June 2026 · 閱讀 27 September 2026
https://escholarship.org/uc/item/33m6w3x0
直接文件: https://escholarship.org/content/qt33m6w3x0/qt33m6w3x0.pdf
“The Reference Case estimate for 2030 data center electricity use is 649 TWh, representing 11.8% of forecasted 2030 U.S. electricity use as presented in the 2025 North American Electric Reliability Corporation (NERC) Long-Term Reliability Assessment (NERC, 2026).”
“This yields a Compounded Uncertainty Range of 521–843 TWh, equal to 9.5–15.3% of total U.S. electricity use in 2030.”
支持的研判. LBNL estimates U.S. data-centre use at 192 TWh (4.7%) in 2024 and a 2030 Reference Case of 649 TWh (11.8%), within a 521–843 TWh compounded uncertainty range.
無法據以確立. U.S. only and scenario-based; the range compounds scenario bounds, not a statistical interval, and is not comparable with IEA's global series.
引用於: 第節 第7.3節, 第9.3節, 第E.11節; 圖 27, 38; 展望 8

S-443 U.S. Energy Information Administration (EIA) — Short-Term Energy Outlook: Electricity, coal, and renewables
政府 · U.S. · 英文 · 9 September 2026 · 閱讀 26 September 2026
https://www.eia.gov/outlooks/steo/report/elec_coal_renew.php
“We forecast that commercial sector electricity sales will increase by 3.3% in 2026 and 2.7% in 2027”
“Electricity consumption reaches record levels in our forecast, driven by data center development”
支持的研判. EIA's September 2026 STEO forecasts commercial-sector electricity sales growth of 3.3% in 2026 and 2.7% in 2027, with consumption reaching record levels driven by data-centre development.
無法據以確立. A monthly-revised short-term forecast; commercial-sector sales cover more than data centres.
引用於: 第節 第7.3節

S-444 Grid Strategies — Power Demand Forecasts Revised Up for Third Year Running, Led by Data Centers
研究 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://gridstrategiesllc.com/project/load-growth-forecast
“Electricity usage is forecast to grow by an average of 5.7% per year over the next five years, with peak demand growth forecast at 166 GW, a 3.7% annual rate.”
支持的研判. Grid Strategies reports forecast U.S. electricity use growing 5.7% a year over five years, with peak-demand growth forecast at 166 GW, a 3.7% annual rate.
無法據以確立. Aggregates utility planning forecasts, not realised load; the page carries no date (the underlying report is from late 2025).
引用於: 第節 第7.3節

S-445 Grid Strategies — Power Demand Forecasts Revised Up for Third Year Running, Led by Data Centers
研究 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://gridstrategiesllc.com/wp-content/uploads/Grid-Strategies-National-Load-Growth-Report-2025.pdf
“Data centers are the largest driver of demand and energy growth, accounting for about 55% of demand growth”
支持的研判. Grid Strategies finds data centres the largest driver of forecast U.S. demand and energy growth, accounting for about 55% of demand growth.
無法據以確立. A share of forecast growth in utility plans, not of realised load; forecasts may shift with data-centre project attrition.
引用於: 第節 第7.3節

S-446 Meta Platforms, Inc. — Meta Announces Nuclear Energy Projects, Unlocking Up to 6.6 GW to Power American Leadership in AI Innovation
企業新聞稿 · U.S. · 英文 · 9 January 2026 · 閱讀 27 September 2026
https://about.fb.com/news/2026/01/meta-nuclear-energy-projects-power-american-ai-leadership/
“These projects are expected to provide thousands of construction jobs and hundreds of long-term operational jobs, supporting up to 6.6 GW of new and existing clean energy by 2035.”
“Through additional 20-year nuclear energy agreements, we will purchase more than 2.1 GW of energy from two operating Vistra nuclear power plants in Ohio (Perry and Davis-Besse), in addition to the energy from expansions (uprates) at these two Ohio plants and a third Vistra nuclear plant in Pennsylvania (Beaver Valley).”
支持的研判. Meta says nuclear agreements support up to 6.6 GW of new and existing clean energy by 2035: over 2.1 GW from operating Vistra plants, new reactors from 2030 at earliest.
無法據以確立. Agreements and 'as early as' dates, not operating capacity; the 6.6 GW includes existing plants and uprates.
引用於: 第節 第7.3節

S-447 U.S. Department of Commerce — Joint Announcement on the Japan-U.S. Strategic Investment
政府 · U.S. · 英文 · 19 March 2026 · 閱讀 26 September 2026
https://www.commerce.gov/news/press-releases/2026/03/joint-announcement-japan-us-strategic-investment
"The offtake from these projects will include supplying power to co-located data centers."
支持的研判. The U.S.–Japan joint announcement of 19 March 2026, on the second batch of strategic-investment projects, says the projects' offtake will include supplying power to co-located data centres.
無法據以確立. Announced projects and ceilings, not financed or built capacity; the share of output going to data centres is not stated.
引用於: 第節 第7.3節

S-448 UL Solutions — Energy storage system testing and certification
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.ul.com/services/energy-storage-system-testing-and-certification
"UL 9540, the Standard for Energy Storage Systems and Equipment, covers electrical, electrochemical, mechanical and other types of energy storage technologies for systems intended to supply electrical energy."
"UL 9540A, The Test Method for Evaluating Thermal Runaway Fire Propagation in Battery Energy Storage Systems"
支持的研判. UL Solutions' energy-storage service page covers certification to UL 9540, the standard for energy storage systems, and testing to UL 9540A, the thermal-runaway fire-propagation test method.
無法據以確立. The certifier's own service description of its offerings; it does not address rack battery backup units in OCP designs specifically.
引用於: 第節 第7.4節

S-449 European Commission (DG CLIMA) — F-gas legislation (DG CLIMA web page)
政府 · EU · 英文 · 未註明日期 · 閱讀 26 September 2026
https://climate.ec.europa.eu/eu-action/fluorinated-greenhouse-gases/f-gas-legislation_en
"In 2050 HFCs will be phased out in the EU"
支持的研判. The European Commission's F-gas page states that HFCs will be phased out in the EU in 2050 under the F-gas Regulation.
無法據以確立. Covers HFCs; it does not mention data centres, chillers or non-HFC refrigerants, so exposure depends on the refrigerant used.
引用於: 第節 第7.4節

S-450 European Chemicals Agency — ECHA Registry of restriction intentions - PFAS
監管機關 · EU · 英文 · 未註明日期 · 閱讀 27 September 2026
https://echa.europa.eu/restrictions-under-consideration/-/substance-rev/72301/term
"Restriction on the manufacture, placing on the market and use of PFASs."
支持的研判. ECHA's registry of restriction intentions lists a proposed restriction on the manufacture, placing on the market and use of PFASs, whose Annex XV consultation closed on 25 September 2023.
無法據以確立. A proposal still in ECHA's process; the entry names no uses such as heat-transfer fluids and no derogations, and gives no adoption date.
引用於: 第節 第7.4節, 第C.7節

S-451 Freshfields — From reporting to rating: the EU's next steps for data centres
法律評論 · U.K. · 英文 · 24 September 2026 · 閱讀 27 September 2026
https://www.freshfields.com/en/our-thinking/blogs/sustainability/from-reporting-to-rating-the-eus-next-steps-for-data-centres-102o345
"Every reporting data centre gets energy and water A-to-G grades."
"An A grade is a PUE of 1.15 or below and a WUE of 0.1 or below; a G grade is a PUE above 1.9 and a WUE above 1.0."
支持的研判. Freshfields summarises the EU rating scheme: each reporting data centre gets separate energy and water A-to-G grades, with A at PUE ≤1.15 and WUE ≤0.1, not combined into one score.
無法據以確立. Law-firm summary of a Commission act under scrutiny; the published regulation governs, and the summary gives no site's actual grades.
引用於: 第節 第7.4節

S-452 Seoul Economic Daily (citing Counterpoint Research) — Samsung's HBM Market Share Jumps to 33%, Narrowing Gap With SK hynix
媒體 · Korea · 英文 · 3 September 2026 · 閱讀 27 September 2026
https://en.sedaily.com/finance/2026/09/03/samsung-doubles-hbm-market-share-to-33-percent-narrowing
"Samsung held a 33% revenue share of the HBM market in the second quarter, according to market research firm Counterpoint Research on the 3rd."
"The figure still trails SK hynix at 50%, but the gap between the two narrowed to 17 percentage points from 37 percentage points in a single quarter"
支持的研判. Seoul Economic Daily reports Counterpoint Research's 2Q26 HBM revenue shares: SK hynix 50%, Samsung 33% and Micron 18%.
無法據以確立. Secondary report of one market-research firm's estimate; shares are of revenue, not bits, and sum to 101% as published.
引用於: 第節 第7.5節

S-453 Office of Trade Negotiations, Executive Yuan (Taiwan) — 臺美對等貿易協定與投資MOU內容不同，已多次說明將依法併送國會
政府 · Taiwan · 繁體中文 · 15 February 2026 · 閱讀 27 September 2026
https://www.ey.gov.tw/otn/6C43520F4EA0A8D5/acce9636-b048-4504-a1f3-81071d9d71dd
"美方將我方對等關稅調降至15%且不疊加"
APEX譯文: The U.S. side will lower our reciprocal tariff to 15%, without stacking.
"該MOU屬政策與合作架構性文件，內容並未涉及個別廠商之投資金額"
APEX譯文: The MOU is a policy and cooperation framework document; its content does not involve the investment amounts of individual companies.
支持的研判. Taiwan's trade office says the MOU (15 January 2026) and ART (12 February 2026) bring a non-stacked 15% tariff; the MOU names no company amounts; both go to the legislature.
無法據以確立. A 15 February 2026 statement of intent to submit; it does not show that legislative review has begun or its outcome.
引用於: 第節 第7.5節

S-454 U.S. Department of Commerce — First Year Accomplishments under Secretary Howard W. Lutnick
政府 · U.S. · 英文 · 20 January 2026 · 閱讀 26 September 2026
https://www.commerce.gov/news/press-releases/2026/01/first-year-accomplishments-under-secretary-howard-w-lutnick
"$250 Billion Semiconductor fabs, guarantees $250 Billion in full semiconductor supply chain"
支持的研判. Commerce's January 2026 accomplishments list records US$250 billion for semiconductor fabs and US$250 billion in guarantees for the full semiconductor supply chain under the Taiwan trade deal.
無法據以確立. An agency's list of commitments, not disbursed investment; it gives no timetable, company breakdown or guarantee terms.
引用於: 第節 第7.5節

S-455 Executive Yuan (Taiwan) — Taiwan commits major funding to semiconductor facilities, chip design software
政府 · Taiwan · 英文 · 10 September 2026 · 閱讀 26 September 2026
https://english.ey.gov.tw/Page/61BF20C3E89B856/c05dc0da-a607-4418-affd-949a11abd0b7
"The government plans to invest NT$300 billion (approximately US$9.5 billion) in semiconductors through 2033, combining hardware and software upgrades to further reinforce Taiwan's key position in global democratic supply chains."
支持的研判. Taiwan's Executive Yuan says the government plans to invest NT$300 billion (about US$9.5 billion) in semiconductors through 2033.
無法據以確立. A government plan announced 10 September 2026, not an appropriation or spending record; the US$ figure is the Executive Yuan's conversion.
引用於: 第節 第7.5節

S-456 European Commission (Shaping Europe's Digital Future) — Cloud and AI Development Act
政府 · EU · 英文 · 3 June 2026 · 閱讀 26 September 2026
https://digital-strategy.ec.europa.eu/en/policies/cloud-and-ai-development-act
"Level 3: where providers must be owned and controlled from the EU and meet additional criteria, such as personnel citizenship."
支持的研判. The Commission's Cloud and AI Development Act page describes sovereignty Level 3, where providers must be owned and controlled from the EU and meet criteria such as personnel citizenship.
無法據以確立. A June 2026 legislative proposal, not adopted law; that Level 3 is the highest of four levels rests on the page's framework description.
引用於: 第節 第7.5節

S-457 European Commission (Shaping Europe's Digital Future) — European Chips Act
政府 · EU · 英文 · 11 September 2026 · 閱讀 26 September 2026
https://digital-strategy.ec.europa.eu/en/policies/european-chips-act
"The Chips Act 2.0., proposed by the Commission in June 2026, introduces new measures to further boost the chips industry, reduce strategic dependencies"
支持的研判. The Commission's European Chips Act page says Chips Act 2.0, proposed in June 2026, introduces measures to further boost the chips industry and reduce strategic dependencies.
無法據以確立. A proposal described by its author; it is not adopted and the excerpt gives no funding amounts.
引用於: 第節 第7.5節

S-458 IndiaAI (MeitY) — IndiaAI Compute Capacity
政府 · India · 英文 · 未註明日期 · 閱讀 27 September 2026
https://indiaai.gov.in/hub/indiaai-compute-capacity
"10 bidders have been selected to offer diversified AI compute units"
支持的研判. The IndiaAI Compute Capacity page lists '18,000+ GPUs' and says 10 bidders have been selected to offer diversified AI compute units.
無法據以確立. Undated page reflecting the 2024 empanelment (dates shown: August and September 2024); not the programme's current size (38,000 GPUs in February 2026, S-316).
引用於: 第節 第7.5節

S-459 ODCC — 开放数据中心委员会 (ODCC) home page
標準組織 · China · 簡體中文 · 未註明日期 · 閱讀 27 September 2026
https://www.odcc.org.cn
"ODCC2504002 Switchless Scale Up GPU超节点互联系统架构技术规范"
APEX譯文: ODCC2504002 Technical specification for the interconnect system architecture of Switchless Scale Up GPU supernodes
"ODCC2504004 冷板液冷超流体技术规范"
APEX譯文: ODCC2504004 Technical specification for cold-plate liquid-cooling superfluid
支持的研判. ODCC's home page lists projects including a Switchless Scale Up GPU supernode interconnect architecture specification (ODCC2504002) and a cold-plate liquid-cooling 'superfluid' specification (ODCC2504004).
無法據以確立. Project titles only, undated on the extract; they show ODCC's work areas, not adoption, products or any link to OCP.
引用於: 第節 第7.5節

S-460 Open Compute Project Foundation — OCP Regional Project Community: China Mainland
標準組織 · Global · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.opencompute.org/projects/regional-project-community-PRC
"Presentation materials, proposals and meeting minutes are published on the respective project's wiki page and are open to the public in accordance to OCP's Bylaws and IP Policy."
支持的研判. OCP's China Mainland regional community page says presentation materials, proposals and minutes are published openly on project wikis under OCP's Bylaws and IP Policy.
無法據以確立. Describes openness of process, not export-control compliance or whether Chinese firms use OCP specifications.
引用於: 第節 第7.5節; 展望 18

S-461 NIST — Transition to Post-Quantum Cryptography Standards (NIST IR 8547 ipd)
政府 · U.S. · 英文 · 12 November 2024 · 閱讀 27 September 2026
https://nvlpubs.nist.gov/nistpubs/ir/2024/NIST.IR.8547.ipd.pdf
取自來源表格之數值（轉錄）： Table 2: Quantum-vulnerable digital signature algorithms, column 'Transition': ECDSA [FIPS186] and RSA [FIPS186] at 112 bits of security strength, Deprecated after 2030, Disallowed after 2035; ECDSA [FIPS186], EdDSA [FIPS186] and RSA [FIPS186] at ≥ 128 bits of security strength, Disallowed after 2035
取自來源表格之數值（轉錄）： Table 4: Quantum-vulnerable key-establishment schemes, column 'Transition': Finite Field DH and MQV [SP80056A], Elliptic Curve DH and MQC [SP80056A] and RSA [SP80056B] at 112 bits of security strength, Deprecated after 2030, Disallowed after 2035; the same three families at ≥ 128 bits of security strength, Disallowed after 2035
支持的研判. NIST's initial public draft IR 8547 (November 2024) proposes disallowing quantum-vulnerable public-key algorithms after 2035, deprecating those at 112-bit security strength after 2030.
無法據以確立. A draft for comment, not final guidance; timelines vary by algorithm and security strength; not binding on non-federal users.
引用於: 第節 第9.4節, 第9.5節; 圖 32

S-462 Microsoft Corporation (Investor Relations) — Microsoft Fiscal Year 2026 First Quarter Earnings Conference Call
企業新聞稿 · U.S. · 英文 · 29 October 2025 · 閱讀 27 September 2026
https://www.microsoft.com/en-us/investor/events/fy-2026/earnings-fy-2026-q1
"We will increase our total AI capacity by over 80% this year, and roughly double our total datacenter footprint over the next two years, reflecting the demand signals we see."
支持的研判. On its 29 October 2025 earnings call, Microsoft said it will roughly double its total data-centre footprint over the next two years, reflecting the demand signals it sees.
無法據以確立. Management target from October 2025; it does not define how footprint is measured or report progress, which the July 2026 call does (S-075).
引用於: 表 19

S-463 Open Compute Project Foundation — Universal Serial Bus Device Class Specification for OCP Open Boot and Management Framework Interface Consolidation Protocol, v1.0.0
標準組織 · Global · 英文 · 10 April 2026 · 閱讀 26 September 2026
https://www.opencompute.org/documents/obmf-icp-over-usb-1-0-pdf
"requirements and specifications for Universal Serial Bus (USB) Devices that support the Open Boot and Management Framework (OBMF) Interface Consolidation Protocol (OBMF-ICP)"
支持的研判. OCP's USB device-class specification v1.0.0 (10 April 2026) sets requirements for USB devices supporting the Open Boot and Management Framework Interface Consolidation Protocol (OBMF-ICP).
無法據以確立. Specification text only; it does not show implementations, interoperability results or adoption by BMC or SoC vendors.
引用於: 表 21

S-464 Optical Internetworking Forum — OIF Hot Topics (Technical Work)
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://www.oiforum.com/technical-work/hot-topics
"448Gbps Signaling for AI Workshop – April 15-16, 2025"
支持的研判. OIF's Hot Topics page lists a '448Gbps Signaling for AI Workshop' held on 15–16 April 2025, showing exploratory 448 Gb/s electrical work.
無法據以確立. A workshop listing; it does not date the start of 448G framework work (2024) or report the ECOC demonstrations, which rest on S-271.
引用於: 表 22

S-465 DMTF Redfish Forum — Redfish Release 2026.1 Now Available; Redfish Release History
標準組織 · U.S. · 英文 · 20 May 2026 · 閱讀 26 September 2026
https://redfish.dmtf.org/schemas/Redfish_Release_History.pdf
"35 schema updates and new additions to the Message Registry"
支持的研判. DMTF's Redfish Release 2026.1 announcement and release history record Release 2026.1, with Specification v1.24.0 and 35 schema updates and new additions to the Message Registry.
無法據以確立. Recorded URL is the release-history deck (lists 2026.1 under April 2026); the 20 May and 15 September 2026 dates come from DMTF news posts.
引用於: 表 22

S-466 OpenBMC — openbmc/bmcweb (GitHub)
標準組織 · U.S. · 英文 · 未註明日期 · 閱讀 26 September 2026
https://github.com/openbmc/bmcweb
"A do everything Redfish, KVM, GUI, and DBus webserver for OpenBMC"
支持的研判. OpenBMC's bmcweb repository describes itself as 'A do everything Redfish, KVM, GUI, and DBus webserver for OpenBMC', an open-source Redfish implementation.
無法據以確立. Shows one implementation exists; it does not show which Redfish release it supports or any deployment volume.
引用於: 表 22

S-467 Broadcom Inc. — Broadcom Announces Tomahawk® 6 – Davisson, the Industry's First 102.4-Tbps Ethernet Switch with Co-Packaged Optics
企業新聞稿 · U.S. · 英文 · 8 October 2025 · 閱讀 27 September 2026
https://investors.broadcom.com/news-releases/news-release-details/broadcom-announces-tomahawkr-6-davisson-industrys-first-1024
"Broadcom Inc. (NASDAQ: AVGO) today announced that it is now shipping Tomahawk® 6 – Davisson (TH6-Davisson), the company's third-generation Co-Packaged Optics (CPO) Ethernet switch."
"Designed specifically for the accelerating demands of AI networking, TH6-Davisson is the industry's first to deliver an unprecedented 102.4 terabits per second of optically enabled switching capacity."
支持的研判. Broadcom said in October 2025 that it is now shipping Tomahawk 6 – Davisson, its third-generation CPO Ethernet switch, which it says delivers 102.4 Tb/s of optically enabled switching capacity.
無法據以確立. Company announcement; 'industry's first' is Broadcom's claim, and it gives no shipment volumes, customers or production status beyond shipping.
引用於: 表 24

S-468 NVIDIA Newsroom — OpenAI and NVIDIA Announce Strategic Partnership to Deploy 10 Gigawatts of NVIDIA Systems
企業新聞稿 · U.S. · 英文 · 22 September 2025 · 閱讀 26 September 2026
https://nvidianews.nvidia.com/news/openai-and-nvidia-announce-strategic-partnership-to-deploy-10gw-of-nvidia-systems
"The first gigawatt of NVIDIA systems will be deployed in the second half of 2026 on the NVIDIA Vera Rubin platform."
支持的研判. OpenAI and NVIDIA announced (22 September 2025) plans for 10 GW of NVIDIA systems, the first gigawatt to be deployed in 2H 2026 on Vera Rubin.
無法據以確立. A letter of intent with planned timing; not a definitive agreement, delivered capacity or disclosed investment terms.
引用於: 表 24

S-469 Astera Labs (company press release) — Astera Labs Bolsters Leo Smart Memory Controller Family for Agentic AI and General-Purpose Cloud Workloads
企業新聞稿 · U.S. · 英文 · 15 September 2026 · 閱讀 27 September 2026
https://www.asteralabs.com/news/astera-labs-bolsters-leo-smart-memory-controller-family-for-agentic-ai-and-general-purpose-cloud-workloads
"The family is sampling today with hyperscaler customers."
支持的研判. Astera Labs says its Leo CXL smart memory controller family is sampling with hyperscaler customers (September 2026).
無法據以確立. Company release; sampling is not production, and its performance figures and the quoted analyst view are not independent tests.
引用於: 表 25

S-470 Yangtze Memory Technologies Co., Ltd. — YMTC company introduction
企業新聞稿 · China · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.ymtc.com/en/intro.html
"Yangtze Memory Technologies Co., Ltd. (YMTC), established in Wuhan, China in July 2016, is an IDM focusing on the design and manufacturing of 3D NAND flash memory, and is a professional advanced memory solutions provider recognized by global customers."
支持的研判. YMTC describes itself as an integrated device manufacturer (IDM) focused on the design and manufacture of 3D NAND flash memory.
無法據以確立. Undated self-description; gives no market share, capacity, output or DRAM/HBM activity.
引用於: 表 25; 第節 第B.2節

S-471 TrendForce (citing Commercial Times) — ASE's SPIL Breaks Ground on Nearly TWD 100B Plant in Douliu to Boost CoWoS Capacity, Operations Set for 2028
媒體 · Taiwan · 英文 · 12 August 2026 · 閱讀 27 September 2026
https://www.trendforce.com/news/2026/08/12/news-ases-spil-breaks-ground-on-nearly-twd-100b-plant-in-douliu-to-boost-cowos-capacity-operations-set-for-2028
"SPIL broke ground on August 11 on a nearly TWD 100 billion, 6-hectare plant in Douliu, Yunlin"
"which will introduce CoWoS advanced packaging and begin first-phase operations in 2028"
支持的研判. TrendForce News, citing Commercial Times, reports SPIL broke ground on 11 August 2026 on a nearly TWD 100 billion Douliu plant to introduce CoWoS packaging, with first-phase operations in 2028.
無法據以確立. A press-reported plan; it gives no CoWoS capacity figure, and 2028 is a planned start, not confirmed operation.
引用於: 表 26

S-472 Tokyo Electron Limited — Q1 FY2027 Earnings Briefing Q&A
企業新聞稿 · Japan · 英文 · 3 August 2026 · 閱讀 27 September 2026
https://www.tel.com/ir/library/report/index.html
直接文件: https://www.tel.com/ir/library/report/i242su0000000goz-att/FY27Q1_EarningCall_QA_E.pdf
"In advanced packaging, sales of bonders and debonders are also expected to grow significantly. In probers, we are receiving an extraordinary number of inquiries, and revenue growth is anticipated to be high."
支持的研判. Tokyo Electron expects bonder and debonder sales in advanced packaging to grow significantly and prober revenue growth to be high, citing an extraordinary number of prober inquiries.
無法據以確立. Q&A remarks without amounts or dates; inquiries are not orders, and the growth outlook is management's view.
引用於: 表 26

S-473 Semiconductor Manufacturing International Corporation (HKEXnews) — SMIC announcement of unaudited results for the second quarter of 2025
申報文件 · China · 英文 · 7 August 2025 · 閱讀 27 September 2026
https://www.hkexnews.hk/listedco/listconews/sehk/2025/0807/2025080700636.pdf
"Monthly capacity increased to 991,250 standard logic 8-inch equivalent wafers in 2Q25 from 973,250 standard logic 8-inch equivalent wafers in 1Q25."
"the capacity utilization rate increased by 2.9 percentage points sequentially to 92.5%."
支持的研判. SMIC reports monthly capacity of 991,250 8-inch-equivalent wafers in 2Q25, up from 973,250 in 1Q25, with utilisation rising 2.9 points to 92.5%.
無法據以確立. Company results for Q2 2025, a year before the report's cut-off; capacity is in standard-logic equivalents, not leading-edge output.
引用於: 表 26; 第節 第B.2節

S-474 NVIDIA Corporation — NVIDIA Vera Rubin Opens Agentic AI Frontier
企業新聞稿 · U.S. · 英文 · 16 March 2026 · 閱讀 27 September 2026
https://investor.nvidia.com/news/press-release-details/2026/NVIDIA-Vera-Rubin-Opens-Agentic-AI-Frontier/default.aspx
"Global system manufacturers Cisco, Dell Technologies, HPE, Lenovo and Supermicro are expected to deliver a wide range of servers based on Vera Rubin products, as well as Aivres, ASUS, Foxconn, GIGABYTE, Inventec, Pegatron, Quanta Cloud Technology (QCT), Wistron and Wiwynn."
支持的研判. NVIDIA says Cisco, Dell, HPE, Lenovo and Supermicro, as well as Aivres, ASUS, Foxconn, GIGABYTE, Inventec, Pegatron, QCT, Wistron and Wiwynn, are expected to deliver Vera Rubin-based servers.
無法據以確立. NVIDIA's expectation at announcement (March 2026), not a record of shipments or volumes by any named manufacturer.
引用於: 表 27

S-475 Wistron Corporation — Wistron Celebrates Grand Opening of First U.S. Smart Factory Marking Milestone in Global Smart Manufacturing Strategy
企業新聞稿 · Taiwan · 英文 · 21 July 2026 · 閱讀 27 September 2026
https://www.wistron.com/en/Newsroom/2026-07-22
"Wistron Corporation ("Wistron") celebrated the grand opening of its D1 AI smart facility in Fort Worth, Texas, the site where the first NVIDIA GB300 Grace Blackwell Ultra Superchip was built and mass-produced in the United States."
支持的研判. Wistron says it opened its D1 AI smart facility in Fort Worth, Texas, where it says the first U.S.-built NVIDIA GB300 superchip was built and mass-produced.
無法據以確立. Company release; 'first' is Wistron's claim, and it gives no output volumes, capacity or customers beyond NVIDIA.
引用於: 表 27; 第節 第B.2節

S-476 Delta Electronics (Americas) — Delta's Groundbreaking 800 VDC Power Solutions Showcased at OCP Global Summit 2025 to Enable Sustainable AI Factories
企業新聞稿 · Taiwan · 英文 · 14 October 2025 · 閱讀 26 September 2026
https://www.delta-americas.com/en-us/news/deltas-groundbreaking-800-vdc-power-solutions-showcased-at-ocp-global-summit-2025-to-enable-sustainable-ai-factories
"enables 800 V to 12 V outputs at maximum 98.5% efficiency"
支持的研判. Delta showed 800 VDC power solutions at the 2025 OCP Global Summit, including 800 V-to-12 V conversion at up to 98.5% maximum efficiency.
無法據以確立. Vendor exhibit claims for component efficiency; not end-to-end system efficiency, availability dates or shipments.
引用於: 表 28

S-477 Wolfspeed, Inc. — Wolfspeed and LITEON Partner to Support Hyperscale AI Data Center Deployments with 800 VDC Power Solutions
企業新聞稿 · U.S. · 英文 · 6 August 2026 · 閱讀 27 September 2026
https://www.wolfspeed.com/company/news-events/news/wolfspeed-and-liteon-partner-to-support-hyperscale-ai-data-center-deployments-with-800-vdc-power-solutions/
"The successful qualification of Wolfspeed's silicon carbide technology for deployment within LITEON's 800VDC power sidecar and compute rack PSU platforms."
支持的研判. Wolfspeed says its silicon carbide technology has been qualified for deployment in LITEON's 800 VDC power sidecar and compute-rack PSU platforms.
無法據以確立. A qualification announcement; it gives no availability, production or shipment dates for LITEON's 800 VDC products.
引用於: 表 28

S-478 GlobeNewswire / STMicroelectronics — STMicroelectronics expands 800 VDC AI datacenter power conversion portfolio with new 12V and 6V architectures in collaboration with NVIDIA
企業新聞稿 · Switzerland · 英文 · 17 March 2026 · 閱讀 27 September 2026
https://www.globenewswire.com/news-release/2026/03/17/3256868/0/en/STMicroelectronics-expands-800-VDC-AI-datacenter-power-conversion-portfolio-with-new-12V-and-6V-architectures-in-collaboration-with-NVIDIA.html
"today announced the expansion of its 800 VDC power conversion portfolio with two new advanced architectures: 800 VDC to 12V and 800 VDC to 6V."
支持的研判. STMicroelectronics announced on 17 March 2026 an expanded 800 VDC power-conversion portfolio with 800 VDC-to-12 V and 800 VDC-to-6 V architectures, developed in collaboration with NVIDIA.
無法據以確立. Company announcement; product availability, customers, volumes and deployment are not established.
引用於: 表 28

S-479 Mitsubishi Heavy Industries, Ltd. — Q&A Summary — 2024 Medium-Term Business Plan Progress Briefing
企業新聞稿 · Japan · 英文 · 27 May 2026 · 閱讀 27 September 2026
https://www.mhi.com/finance/library/plan/pdf/260527presentation_qa.pdf
"we have set a goal of increasing the number of GTCC units shipped in FY2028 by 30% or more versus current capacity, and from there we are also looking at increasing the number of units shipped by another 50% or higher, i.e., around two times the FY2024 level."
支持的研判. MHI has set a goal of FY2028 GTCC unit shipments 30% or more above current capacity and is looking at a further 50% or more, about twice FY2024.
無法據以確立. A management goal and an aspiration ('looking at'), not guidance; the Q&A does not tie the units to data-centre demand.
引用於: 表 29; 第節 第B.2節

S-480 Mitsubishi Heavy Industries — MHI Advances AI Infrastructure Commercialization with U.S. Deployment of 10MW-Class Chiller and MCP Development
企業新聞稿 · Japan · 英文 · 16 July 2026 · 閱讀 27 September 2026
https://www.mhi.com/news/260716.html
"MHI has shipped its 10MW-class centrifugal chiller test unit to the United States, supporting next-generation high-density AI workloads with arrival scheduled at the Port of Brunswick, Georgia around July."
支持的研判. MHI says it shipped a 10 MW-class centrifugal chiller test unit to the United States for high-density AI workloads, arriving around July 2026, and is developing a Modular Chiller Plant.
無法據以確立. A test unit with no named customer; not a commercial deployment, and the Modular Chiller Plant is still undergoing U.S. safety certification.
引用於: 表 29

S-481 Lumentum Holdings Inc. — Lumentum Announces Fourth Quarter and Full Fiscal Year 2026 Results
企業新聞稿 · U.S. · 英文 · 11 August 2026 · 閱讀 27 September 2026
https://investor.lumentum.com/financial-news-releases/news-details/2026/Lumentum-Announces-Fourth-Quarter-and-Full-Fiscal-Year-2026-Results/default.aspx
""Increasing demand for ultra-high-power CPO lasers, an initial order for ELS modules, as well as our breadth of NPO engagements are the first signs that optics are starting to penetrate in-rack connectivity, significantly upping our optical TAM," said President and CEO Michael Hurlston."
"While our Q4 results demonstrate broad-based traction, key growth drivers such as OCS solutions and our cloud module business, where we are advancing 1.6T adoption, are beginning to layer in."
支持的研判. Lumentum cites increasing demand for ultra-high-power CPO lasers and an initial order for ELS modules, and says growth drivers including OCS solutions are beginning to layer in.
無法據以確立. Company results commentary; it does not quantify CPO, ELS or OCS revenue, and the TAM view is management's.
引用於: 表 30

S-482 imec — Interfacing silicon photonics for high-density co-packaged optics
研究 · Belgium · 英文 · 未註明日期 · 閱讀 27 September 2026
https://www.imec-int.com/en/articles/interfacing-silicon-photonics-high-density-co-packaged-optics
"the tight alignment tolerance because of the small spot size at the optical interface"
"While this enables the integration of V-grooves for passive assembly of SMFs, it places a constraint on the optical I/O density."
支持的研判. An imec article (26 November 2024) names tight alignment tolerance at the optical interface and a constraint on optical I/O density among limits in interfacing silicon photonics for CPO.
無法據以確立. A research-centre article on its own approaches; it does not rank barriers or treat test as a named limit.
引用於: 表 30

S-483 Wiwynn Corporation (reviewed consolidated financial statements; auditor KPMG, Taipei) — WIWYNN CORPORATION AND SUBSIDIARIES CONSOLIDATED FINANCIAL STATEMENTS With Independent Auditors' Review Report for the Six Months Ended June 30, 2026 and 2025
申報文件 · Taiwan · 英文 · 7 August 2026 · 閱讀 27 September 2026
https://www.wiwynn.com/hubfs/Investors/Financial_Report/Wiwynn_2026Q2_Financial.pdf
"For certain customer transactions, the Group assists in procuring materials from suppliers designated by the customers and provides related assembly and delivery services. As the Group does not obtain control of such materials before transfer to the customers, the Group is considered an agent in these transactions."
"These consolidated financial statements have been prepared in accordance with the Regulations Governing the Preparation of Financial Reports by Securities Issuers (hereinafter referred to as "the Regulations") and IAS 34 "Interim Financial Reporting" which are endorsed and issued into effect by the Financial Supervisory Commission (hereinafter referred to as FSC)."
取自來源表格之數值（轉錄）: KPMG | Taipei, Taiwan (Republic of China) | August 7, 2026
支持的研判. Wiwynn's KPMG-reviewed interim statements for the six months to 30 June 2026 say that for certain customer transactions it procures customer-designated materials without obtaining control of them and is therefore an agent.
無法據以確立. The note names neither memory nor amounts; the English text is the company's translation, and its Chinese version prevails.
引用於: 第一頁; 第節 第1.5節, 第5.1節, 第5.2節, 第5.3節, 第C.7節; 表 8, 39 · **研判:** A-14

S-484 European Commission (Official Journal of the European Union, via EUR-Lex) — COMMISSION REGULATION (EU) 2017/1987 of 31 October 2017 amending Regulation (EC) No 1126/2008 adopting certain international accounting standards in accordance with Regulation (EC) No 1606/2002 of the European Parliament and of the Council as regards International Financial Reporting Standard 15
政府 · EU · 英文 · 9 November 2017 · 閱讀 27 September 2026
https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32017R1987
"When (or as) an entity that is an agent satisfies a performance obligation, the entity recognises revenue in the amount of any fee or commission to which it expects to be entitled in exchange for arranging for the specified goods or services to be provided by the other party."
"An entity is a principal if it controls the specified good or service before that good or service is transferred to a customer."
支持的研判. The EU-endorsed clarifications to IFRS 15 state that an entity is a principal if it controls the good before transfer, and that an agent recognises revenue in the amount of its fee or commission.
無法據以確立. The EU's adoption text; Taiwan applies IFRS as endorsed by its Financial Supervisory Commission, which contains the same principal-agent guidance.
引用於: 第一頁; 第節 第5.2節, 第C.7節; 表 8

S-485 Advanced Micro Devices, Inc. — AMD Reports Second Quarter 2025 Financial Results
企業新聞稿 · U.S. · 英文 · 5 August 2025 · 閱讀 27 September 2026
https://ir.amd.com/news-events/press-releases/detail/1257/amd-reports-second-quarter-2025-financial-results
"Our current outlook does not include any revenue from AMD Instinct MI308 shipments to China as our license applications are currently under review by the U.S. Government."
"For the quarter, these restrictions led to approximately $800 million in inventory and related charges."
支持的研判. AMD's second-quarter 2025 release says its outlook includes no revenue from MI308 shipments to China because its licence applications are under U.S. review, after about US$800 million of related inventory charges.
無法據以確立. An outlook statement of 5 August 2025; it does not estimate the Chinese demand left out.
引用於: 第節 第1.3節, 第7.1節; 表 8, 14 · **研判:** A-08

S-486 Advanced Micro Devices, Inc. — AMD Reports Fourth Quarter and Full Year 2025 Financial Results
企業新聞稿 · U.S. · 英文 · 3 February 2026 · 閱讀 27 September 2026
https://ir.amd.com/news-events/press-releases/detail/1276/amd-reports-fourth-quarter-and-full-year-2025-financial-results
"For the first quarter of 2026, AMD expects revenue to be approximately $9.8 billion, plus or minus $300 million, including approximately $100 million of AMD Instinct MI308 sales to China."
"Fourth quarter AMD Instinct MI308 revenue to China was approximately $390 million."
支持的研判. AMD guides first-quarter 2026 revenue of about US$9.8 billion including about US$100 million of MI308 sales to China, after about US$390 million of such revenue in the fourth quarter of 2025.
無法據以確立. Company guidance (G) at 3 February 2026; licence conditions may change the China amount in later quarters.
引用於: 第節 第1.3節, 第7.1節; 表 8, 14, 20 · **研判:** A-08

S-487 KLA Corporation — Letter to Shareholders Q1 Fiscal 2026
企業新聞稿 · U.S. · 英文 · 29 October 2025 · 閱讀 27 September 2026
https://d1io3yog0oux5.cloudfront.net/_e8e7c5c005c53adbff4d2646a9bf361d/klatencor/db/1117/10596/letter_to_shareholders/KLA+Corporation+Letter+to+Shareholders+Q1+FY26.pdf
"This outlook is inclusive of the revenue impact related to additional market access loss related to certain customers in China resulting from extended export controls from the US government."
"We estimate the revenue impact on the December quarter and calendar 2026 to be approximately $300 to $350 million for KLA."
支持的研判. KLA says its outlook includes the revenue effect of lost market access to certain Chinese customers under extended U.S. export controls, which it estimates at about US$300-350 million for the December quarter and calendar 2026.
無法據以確立. A company estimate at 29 October 2025, before the suspension of the U.S. Affiliates rule on 10 November 2025.
引用於: 第節 第1.3節, 第7.1節; 表 8, 14, 20 · **研判:** A-08, A-14

S-488 DeepSeek (DeepSeek API Docs, news page) — DeepSeek-V4.1-Flash: Smarter, Faster, More Efficient
企業新聞稿 · China · 英文 · 10 September 2026 · 閱讀 27 September 2026
https://api-docs.deepseek.com/news/news260910
"Cache-hit charges often account for a large share of agent costs. Compressing the cache cuts those costs significantly."
"Planning a large-scale deployment with 2,000 GPUs + a storage cluster? Let's talk."
取自來源表格之數值（轉錄）: Compared with the previous generation, V4.1-Flash's KV cache needs just: | 1/4 the HBM | 1/8 the SSD storage
支持的研判. DeepSeek's product page says V4.1-Flash's KV cache needs a quarter of the HBM and an eighth of the SSD storage of the previous generation, that cache-hit charges are a large share of agent costs, and invites 2,000-GPU deployments.
無法據以確立. A promotional page; the ratios are the developer's own and compare with its previous generation only.
引用於: 第節 第1.3節, 第8.2節, 第8.4節; 表 16

S-489 Moonshot AI (Hugging Face model card) — moonshotai/Kimi-K3
企業新聞稿 · China · 英文 · July 2026 · 閱讀 27 September 2026
https://huggingface.co/moonshotai/Kimi-K3
直接文件: https://huggingface.co/moonshotai/Kimi-K3/raw/main/config.json
"Kimi K3 is an open-weight, native multimodal agentic model and our most capable model to date. It is a 2.8T-parameter model built on Kimi Delta Attention (KDA) and Attention Residuals (AttnRes), with native vision capabilities and a 1-million-token context window."
"Kimi K3 applies quantization-aware training from the SFT stage onward, using MXFP4 weights with MXFP8 activations for broad hardware compatibility."
取自來源表格之數值（轉錄）: Total Parameters | 2.8T | Activated Parameters | 104B | Number of Layers | 93 | Number of Dense Layers | 1 | Attention-Layer Composition | 69 KDA + 24 Gated MLA | Attention Hidden Dimension | 7168 | Number of Attention Heads | 96
取自來源表格之數值（轉錄）: num_hidden_layers | 93 | num_attention_heads | 96 | kv_lora_rank | 512 | qk_rope_head_dim | 64 | max_position_embeddings | 1048576 | format | mxfp4-pack-quantized | num_bits | 4 | group_size | 32 | kv_cache_scheme | null
支持的研判. Moonshot AI's model card describes Kimi K3 as an open-weight 2.8T-parameter model with 104B activated, 69 Kimi Delta Attention and 24 gated MLA layers, a 1M-token context, and MXFP4 weights with MXFP8 activations.
無法據以確立. The developer's description and configuration; the report's weight and cache sizes are its own calculations from them.
引用於: 第節 第1.3節, 第8.1節, 第8.2節, 第8.3節, 第C.7節; 表 16, 17 · **研判:** A-13

S-490 Alibaba Qwen team (Hugging Face model card) — Qwen/Qwen3-235B-A22B-Instruct-2507
企業新聞稿 · China · 英文 · 21 July 2025 · 閱讀 27 September 2026
https://huggingface.co/Qwen/Qwen3-235B-A22B-Instruct-2507
直接文件: https://huggingface.co/Qwen/Qwen3-235B-A22B-Instruct-2507/raw/main/config.json
"To effectively process a 1 million token context, users will require approximately 1000 GB of total GPU memory. This accounts for model weights, KV-cache storage, and peak activation memory demands."
"Context Length: 262,144 natively and extendable up to 1,010,000 tokens"
取自來源表格之數值（轉錄）: num_hidden_layers | 94 | num_attention_heads | 64 | num_key_value_heads | 4 | head_dim | 128 | max_position_embeddings | 262144 | torch_dtype | bfloat16
取自來源表格之數值（轉錄）: license | apache-2.0
支持的研判. Qwen's model card gives Qwen3-235B-A22B-Instruct-2507 a 262,144-token native context extendable to about 1M tokens and 94 layers with 4 KV heads of 128 dimensions, and says a 1M context needs about 1000 GB of GPU memory.
無法據以確立. Developer guidance for its own reference set-up; the report's cache figures are its own calculations at 16 bits.
引用於: 第節 第1.3節, 第8.3節, 第C.7節; 表 16, 17

S-491 Google (The Keyword blog; Sundar Pichai, CEO of Google and Alphabet) — I/O 2026: Welcome to the agentic Gemini era
企業新聞稿 · U.S. · 英文 · 19 May 2026 · 閱讀 27 September 2026
https://blog.google/innovation-and-ai/sundar-pichai-io-2026/
"Fast forward to today, that number jumped 7x to over 3.2 quadrillion per month."
"Last year at I/O, that grew to roughly 480 trillion tokens."
支持的研判. Google says it now processes over 3.2 quadrillion tokens a month across its surfaces, seven times the roughly 480 trillion of a year earlier, against 9.7 trillion two years earlier.
無法據以確立. The company's own count across all its surfaces, without a published definition of tokens processed.
引用於: 第節 第1.3節, 第8.4節, 第8.6節, 第C.7節; 圖 31; 表 18, 20; 展望 7 · **研判:** A-13

S-492 Alphabet Inc. (Form 10-Q filed with the U.S. SEC) — Form 10-Q, Alphabet Inc., quarterly period ended June 30, 2026
申報文件 · U.S. · 英文 · 23 July 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/1652044/000165204426000071/goog-20260630.htm
直接文件: https://s206.q4cdn.com/479360582/files/doc_financials/2026/q2/GOOG-10-Q-Q2-2026.pdf
"As of June 30, 2026, we have entered into leases, primarily related to data centers, that have not yet commenced with future lease payments of $85.2 billion that are not yet recorded."
"These leases will commence between 2026 and 2031 with non-cancelable lease terms between one and 26 years."
支持的研判. Alphabet reports leases, primarily for data centres, not yet commenced with US$85.2 billion of future payments at 30 June 2026, commencing between 2026 and 2031.
無法據以確立. Lease commitments, not capital expenditure or capacity; the megawatts they represent are not stated.
引用於: 第節 第2.2節

S-493 Meta Platforms, Inc. (Form 10-Q filed with the U.S. SEC) — Form 10-Q, Meta Platforms, Inc., quarterly period ended June 30, 2026
申報文件 · U.S. · 英文 · 30 July 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/1326801/000162828026050705/meta-20260630.htm
"As of June 30, 2026, we had $349.31 billion of non-cancelable contractual commitments, comprising both short-term and long-term arrangements."
"These commitments mostly relate to third-party cloud capacity arrangements and investments in servers and network infrastructure, data centers, and consumer hardware products in Reality Labs, with approximately $53.52 billion and $81.65 billion due in 2026 and 2027, respectively."
支持的研判. Meta reports US$349.31 billion of non-cancelable contractual commitments at 30 June 2026, mostly for third-party cloud capacity and investments in servers, network infrastructure, data centres and Reality Labs hardware.
無法據以確立. Commitments include non-AI items and are not capital expenditure; these passages do not split them by type.
引用於: 第節 第2.2節

S-494 Calix, Inc. (Exhibit 99.2 to Form 8-K filed with the U.S. SEC) — Exhibit 99.2: stockholder letter dated July 20, 2026 (no other printed title)
申報文件 · U.S. · 英文 · 20 July 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/1406666/000140666626000028/ex992stockholderletter26q2.htm
"The sequential and year-over-year decreases reflect an industry-wide, rapid rise in memory component costs due to the demand from the building of AI data centers."
"What this means is that over time the incremental memory costs will be neutral to gross profit, however, they will still be a headwind to gross margin."
"The goal of our memory surcharge program is to partner with our customers for cost recovery only."
支持的研判. Calix says an industry-wide rise in memory costs driven by AI data-centre construction cut its gross margin, and that its memory surcharges recover cost only, so over time the costs are neutral to gross profit.
無法據以確立. A stockholder letter describing its own pricing; surcharge amounts and total memory costs are not given.
引用於: 第節 第5.3節; 表 8 · **研判:** A-14

S-495 KLA Corporation — Letter to Shareholders Q2 Fiscal 2026
企業新聞稿 · U.S. · 英文 · 29 January 2026 · 閱讀 27 September 2026
https://d1io3yog0oux5.cloudfront.net/_42d21a5ab5ca507394aee7b9106cc480/klatencor/db/1117/10612/letter_to_shareholders/KLA+Letter+to+Shareholders+-+Q2+FY26.pdf
"This guidance also includes the incremental impact of the rapidly escalating cost of DRAM chips used in the company's image processing computers that ship with our systems, creating a headwind to our gross margin expectations."
"While we expect this pricing environment to be transitory, current modeling suggests that this situation will persist through calendar 2026 and have a roughly 75bps to 100bps negative impact on gross margins for the calendar year."
支持的研判. KLA's guidance for the March 2026 quarter includes the rapidly rising cost of DRAM in the image-processing computers shipped with its systems, a gross-margin headwind of roughly 75-100 basis points for calendar 2026.
無法據以確立. Company guidance at 29 January 2026; KLA models a return to normal memory pricing in its longer-term forecasts.
引用於: 第節 第5.3節; 表 8

S-496 Super Micro Computer, Inc. (press release furnished on Form 8-K, Exhibit 99.1) — Supermicro Announces Fourth Quarter and Full Fiscal Year 2026 Financial Results
申報文件 · U.S. · 英文 · 11 August 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/1375365/000137536526000021/exhibit991_20260630.htm
"Net sales for the full fiscal year ended June 30, 2026, were $39.1 billion versus $22.0 billion for the fiscal year ended June 30, 2025."
"Gross margin for fiscal year 2026 was 10.8% versus 11.1% for the fiscal year ended June 30, 2025."
支持的研判. Supermicro reports fiscal 2026 net sales of US$39.1 billion against US$22.0 billion a year earlier, with gross margin of 10.8% against 11.1%.
無法據以確立. Results-release figures; these passages do not attribute the margin change to any cause.
引用於: 第節 第5.3節; 表 8

S-497 Super Micro Computer, Inc. (Form 10-K filed with the U.S. SEC) — Form 10-K, Super Micro Computer, Inc., fiscal year ended June 30, 2026
申報文件 · U.S. · 英文 · 31 August 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/1375365/000137536526000022/smci-20260630.htm
"GPUs have represented an increasing portion of our cost of sales."
"Cost of sales primarily consists of the costs of materials, contract manufacturing, in-bound shipping, personnel and related expenses including stock-based compensation, tariffs, equipment and facility expenses, warranty costs and write down adjustments for lower of cost or net realizable value and excess and obsolete inventory."
取自來源表格之數值（轉錄）: Net sales (including related party sales of $29,781, $42,259, and $69,791 in fiscal years 2026, 2025, and 2024, respectively) | $ 39,063,072 | $ 21,972,042 | $ 14,989,251
取自來源表格之數值（轉錄）: Gross profit | 4,227,251 | 2,429,922 | 2,061,410
支持的研判. Super Micro's 10-K says GPUs are an increasing share of its cost of sales, which includes materials and tariffs, and reports fiscal 2026 net sales of US$39.1 billion against US$22.0 billion.
無法據以確立. Annual-report disclosures; the GPU share of cost of sales and rebate amounts are not given in these passages.
引用於: 第節 第5.3節; 表 8

S-498 Celestica Inc. (release furnished to the U.S. SEC as Exhibit 99.1) — CELESTICA ANNOUNCES FOURTH QUARTER AND FY 2025 FINANCIAL RESULTS
申報文件 · Canada · 英文 · 28 January 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/1030894/000103089426000006/cls-20251231prcondensedfs.htm
"Substantially all tariffs paid by Celestica are expected to continue to be recovered from our customers, and are not expected to materially impact our adjusted operating earnings (adjusted EBIAT) (non-GAAP)* or adjusted net earnings (non-GAAP)* dollars."
"Our 2026 Annual Outlook Update assumes no material changes to tariffs or trade restrictions compared to what are in effect as of January 28, 2026 and no material changes from current macroeconomic trends and uncertainties."
支持的研判. Celestica's 2026 outlook of 28 January 2026 assumes no material changes to tariffs or trade restrictions and expects substantially all tariffs it pays to continue to be recovered from customers.
無法據以確立. An outlook assumption; it does not quantify tariffs paid or recovered.
引用於: 第節 第5.3節, 第7.1節; 表 8

S-499 Celestica Inc. — Celestica Announces Second Quarter 2026 Financial Results
企業新聞稿 · Canada · 英文 · 27 July 2026 · 閱讀 27 September 2026
https://corporate.celestica.com/news-releases/news-release-details/celestica-announces-second-quarter-2026-financial-results
"no material changes to tariffs, trade restrictions, customs administration, or related refund or recovery processes, including developments relating to duties imposed under the International Emergency Economic Powers Act of 1977 and any replacement, continuing or retaliatory tariff measures, compared to what are in effect as of July 27, 2026"
"that our customers will retain liability for and we will continue to be able to recover substantially all costs from customers relating to product/component tariffs and countermeasures"
支持的研判. Celestica's July 2026 forward-looking statements assume no material changes to tariffs, trade restrictions or refund processes from those in effect at 27 July 2026, and that customers keep liability for tariff costs.
無法據以確立. Assumptions in a forward-looking-statements section, not a quantified tariff outlook.
引用於: 第節 第5.3節; 表 8

S-500 Generac Holdings Inc. (Exhibit 99.1 to Form 8-K filed with the U.S. SEC) — Generac Reports Second Quarter 2026 Results
申報文件 · U.S. · 英文 · 29 July 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/1474735/000143774926024731/ex_994826.htm
"In total, our backlog for products serving the data center market has now increased to approximately $1.6 billion as of today, which does not include any committed volumes from the second hyperscale customer."
"On June 24th, the Company secured a global supply agreement with a second hyperscale customer and is currently negotiating final product specific terms for 2027 and 2028 volumes."
支持的研判. Generac says its data-centre product backlog reached about US$1.6 billion, excluding committed volumes from a second hyperscale customer with which it signed a global supply agreement on 24 June 2026.
無法據以確立. A backlog figure in a results release; the excluded volumes and 2027-28 terms were still being negotiated.
引用於: 第節 第5.3節; 表 8 · **研判:** A-14

S-501 Oracle Corporation (press release furnished on Form 8-K, Exhibit 99.1) — Oracle Announces Record Q4 and FY 2026 Results Driven by Cloud Infrastructure & Cloud Applications
申報文件 · U.S. · 英文 · 10 June 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/1341439/000119312526265848/orcl-ex99_1.htm
"The prepaid and customer supplied hardware portions of our large AI contracts now total $75 billion."
"Most of the RPO increase in both Q3 and Q4 were large scale AI contracts where the customer prepaid Oracle for the purchase of the GPUs, or the customer bought and supplied the GPUs to Oracle."
支持的研判. Oracle says most of its RPO increase in the third and fourth quarters came from large AI contracts in which customers prepaid for GPUs or supplied them, portions that now total US$75 billion.
無法據以確立. A results-release statement; it gives neither the GPU count nor the timing of deployment.
引用於: 第節 第5.3節; 表 8

S-502 Advanced Micro Devices, Inc. — AMD Reports Third Quarter 2025 Financial Results
企業新聞稿 · U.S. · 英文 · 4 November 2025 · 閱讀 27 September 2026
https://ir.amd.com/news-events/press-releases/detail/1265/amd-reports-third-quarter-2025-financial-results
"Our current outlook does not include any revenue from AMD Instinct MI308 shipments to China."
"Our third quarter results did not include any revenue from shipments of AMD Instinct™ MI308 GPU products to China."
支持的研判. AMD's third-quarter 2025 release says its outlook does not include any revenue from MI308 shipments to China, and that third-quarter results included none.
無法據以確立. The release gives no reason for the exclusion and no estimate of the demand left out.
引用於: 第節 第1.3節, 第7.1節; 表 8, 14

S-503 NVIDIA Corporation (Form 10-Q filed with the U.S. SEC) — Form 10-Q, NVIDIA Corporation, quarterly period ended July 26, 2026
申報文件 · U.S. · 英文 · 26 August 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/1045810/000104581026000075/nvda-20260726.htm
直接文件: https://s201.q4cdn.com/141608511/files/doc_financials/2027/NVDA-2027-Q2-10Q-Final-including-exhibits.pdf
"The U.S. government, or USG, granted licenses that would allow us to ship small amounts of H200 products to specific China-based customers, but such sales were restricted by the PRC government, and we have been unable to sell all the products for which we have licenses."
"As a result, any H200s shipped under the new licensing program are subject to a 25% tariff upon importation into the United States."
支持的研判. NVIDIA's 10-Q says U.S. licences allow small H200 shipments to specific China-based customers but the PRC government restricted the sales; licensed H200s bear a 25% U.S. import tariff, and it took a US$0.4 billion H200 charge.
無法據以確立. Quarterly risk and inventory disclosures; they do not quantify Chinese demand or future licensed volumes.
引用於: 表 8, 14; 第節 第5.3節, 第7.1節 · **研判:** A-08

S-504 Applied Materials, Inc. (Form 8-K filed with the U.S. SEC) — Form 8-K (Current Report), date of report September 29, 2025
申報文件 · U.S. · 英文 · 2 October 2025 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/6951/000162828025043753/amat-20250929.htm
"Applied currently estimates that the impact of the BIS Affiliates Rule will reduce its net revenue for the fourth quarter of fiscal 2025 by approximately $110 million."
"In addition, Applied currently expects that the impact of the BIS Affiliates Rule will reduce its net revenue for fiscal 2026 by approximately $600 million."
支持的研判. Applied Materials estimates that the BIS Affiliates rule will reduce its net revenue by about US$110 million in the fourth quarter of fiscal 2025 and about US$600 million in fiscal 2026.
無法據以確立. Company estimates made before the rule's suspension from 10 November 2025; later filings may revise them.
引用於: 第節 第7.1節; 表 14 · **研判:** A-08

S-505 U.S. Department of Commerce, Bureau of Industry and Security (BIS), Federal Register — One Year Suspension of Expansion of End-User Controls for Affiliates of Certain Listed Entities
監管機關 · U.S. · 英文 · 12 November 2025 · 閱讀 27 September 2026
https://www.govinfo.gov/content/pkg/FR-2025-11-12/html/2025-19846.htm
“Effective November 10, 2025, the amendments to 15 CFR parts 732, 734, 736, 744, and 748 in the interim final rule published at 90 FR 47201, on September 30, 2025, are stayed until November 9, 2026.”
“The suspension is set to end November 9, 2026, absent a future extension.”
支持的研判. BIS stays the September 2025 Affiliates rule from 10 November 2025 until 9 November 2026, the suspension ending then absent a future extension.
無法據以確立. A suspension notice; it does not say what follows the end of the stay.
引用於: 第節 第7.1節, 第8.5節; 表 14; 展望 19

S-506 ASML Holding N.V. — ASML reports €7.5 billion total net sales and €2.1 billion net income in Q3 2025
企業新聞稿 · Netherlands · 英文 · 15 October 2025 · 閱讀 27 September 2026
https://www.asml.com/en/news/press-releases/2025/q3-2025-financial-results
“On the other hand, we expect China customer demand, and therefore our China total net sales in 2026, to decline significantly compared to our very strong business there in 2024 and 2025.”
“We do not expect 2026 total net sales to be below 2025.”
支持的研判. ASML expects China customer demand, and so its China net sales in 2026, to decline significantly from very strong 2024 and 2025 levels, while it does not expect total 2026 net sales below 2025.
無法據以確立. A qualitative outlook at 15 October 2025; no China amount is given.
引用於: 第節 第7.1節; 表 14

S-507 ASML Holding N.V. — ASML reports €8.8 billion total net sales and €2.8 billion net income in Q1 2026
企業新聞稿 · Netherlands · 英文 · 15 April 2026 · 閱讀 27 September 2026
https://www.asml.com/en/news/press-releases/2026/q1-2026-financial-results
“We expect that the bandwidth in our 2026 guidance accommodates potential outcomes of ongoing discussions around export controls”
“we now expect total net sales for 2026 to be between €36 billion and €40 billion, with a gross margin between 51% and 53%.”
支持的研判. ASML guides 2026 total net sales of EUR 36-40 billion and says the range accommodates potential outcomes of ongoing discussions around export controls.
無法據以確立. The range does not separate the export-control effect or give a China amount.
引用於: 第節 第7.1節; 表 14

S-508 Synopsys, Inc. (Form 8-K filed with the U.S. SEC) — Form 8-K (Current Report), date of report May 29, 2025
申報文件 · U.S. · 英文 · 29 May 2025 · 閱讀 27 September 2026
https://sec.gov/Archives/edgar/data/883241/000119312525130725/d14316d8k.htm
“On May 29, 2025, following the announcement of Synopsys' financial results of its second fiscal quarter ended April 30, 2025, Synopsys received a letter from the Bureau of Industry and Security of the U.S. Department of Commerce informing Synopsys of new export restrictions related to China”
“has suspended its financial guidance for the third quarter of fiscal year 2025 and full fiscal year 2025.”
支持的研判. Synopsys reports that on 29 May 2025 BIS informed it by letter of new export restrictions related to China, and that it suspended its fiscal 2025 guidance while assessing the impact.
無法據以確立. A current report at the time of the letter; it gives no amount, and later changes to the restrictions are outside it.
引用於: 第節 第7.1節; 表 14

S-509 Cadence Design Systems, Inc. (Form 8-K filed with the U.S. SEC) — Form 8-K (Current Report), date of report May 23, 2025
申報文件 · U.S. · 英文 · 29 May 2025 · 閱讀 27 September 2026
https://sec.gov/Archives/edgar/data/813672/000081367225000079/cdns-20250523.htm
“a license is now required for the export, re-export or in-country transfer of electronic design automation software and technology classified under Export Control Classification Numbers (ECCNs) 3D991 and 3E991 on the Commerce Control List, when a party to the transaction is located in China”
“The new requirements are complex, and we are engaging with BIS to obtain further clarification, as we assess the impact on our business and financial results.”
支持的研判. Cadence reports that a licence is now required to export EDA software and technology classified under ECCNs 3D991 and 3E991 when a party to the transaction is located in China, and that it is seeking clarification from BIS.
無法據以確立. A current report with no amount; later changes to the requirement are outside it.
引用於: 第節 第7.1節; 表 14

S-510 U.S. Department of Commerce, Bureau of Industry and Security (BIS), Federal Register — Revocation of Validated End-User Authorizations in the People's Republic of China
監管機關 · U.S. · 英文 · 2 September 2025 · 閱讀 27 September 2026
https://www.govinfo.gov/content/pkg/FR-2025-09-02/html/2025-16735.htm
“In this final rule, the Bureau of Industry and Security (BIS) amends the Export Administration Regulations (EAR) to revise the existing Validated End-User (VEU) Authorizations list for the People's Republic of China (PRC) by removing Intel Semiconductor (Dalian) Ltd; Samsung China Semiconductor Co. Ltd; and SK hynix Semiconductor (China) Ltd.”
“BIS estimates that these new removal of these entities from the VEU program under the EAR will result in the submission of an additional 1,000 license applications annually, which would be an increase of 495 burden hours.”
支持的研判. BIS's final rule removes Intel Semiconductor (Dalian), Samsung China Semiconductor and SK hynix Semiconductor (China) from the China VEU list, effective 31 December 2025, and expects about 1,000 more licence applications a year.
無法據以確立. The rule's own burden estimate; licensing outcomes after the removal are not shown.
引用於: 第節 第7.1節, 第C.7節; 表 14

S-511 Micron Technology, Inc. (Form 10-K filed with the U.S. SEC) — Form 10-K for the fiscal year ended August 28, 2025
申報文件 · U.S. · 英文 · 3 October 2025 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/723125/000072312525000028/mu-20250828.htm
“Following the May 2023 decision of its cybersecurity review of our products sold in China, the CAC determined that critical information infrastructure operators in China may not purchase Micron products, impacting our revenue with companies headquartered in mainland China and Hong Kong, including direct sales as well as indirect sales through distributors.”
“Further actions by the Chinese government, through CAC action or other means, could impact revenue inside or outside China, or our operations in China, or our ability to ship products to our customers, any of which could have a material adverse effect on our business, results of operations, or financial condition.”
支持的研判. Micron's 10-K says China's CAC determined after a May 2023 review that critical information infrastructure operators in China may not buy Micron products, and warns further Chinese actions could affect revenue.
無法據以確立. A risk-factor disclosure; these passages do not quantify the revenue affected.
引用於: 第節 第7.1節; 表 14

S-512 中科寒武紀科技股份有限公司 (Cambricon Technologies Corporation Limited) — 中科寒武紀科技股份有限公司2025年年度报告
申報文件 · China · 簡體中文 · 13 March 2026 · 閱讀 27 September 2026
https://dataclouds.cninfo.com.cn/shgonggao/hsomarket/2026/20260312/05ca784762a7401b9ed371d917e436dc.PDF
“由于集成电路整个行业链是专业化分工且技术门槛较高，加之公司及部分子公司已被列入“实体清单”，将对公司供应链的稳定造成一定风险。”
APEX譯文: Because the integrated-circuit industry chain is organised by specialised division of labour with high technical thresholds, and the company and some of its subsidiaries have been placed on the 'Entity List', there is a certain risk to the stability of the company's supply chain.
“切换新供应商将产生一定成本，将可能对公司经营业绩产生不利影响。”
APEX譯文: Switching to new suppliers would incur certain costs and could adversely affect the company's operating results.
支持的研判. Cambricon's 2025 annual report says the listing of the company and some subsidiaries on the Entity List poses a risk to supply-chain stability, and that switching suppliers would add costs and could hurt results.
無法據以確立. A risk-factor statement; it names no affected supplier and quantifies no effect.
引用於: 第節 第7.1節; 表 14

S-513 NVIDIA Corporation — CFO Commentary on Third Quarter Fiscal 2026 Results
企業新聞稿 · U.S. · 英文 · 19 November 2025 · 閱讀 27 September 2026
https://s201.q4cdn.com/141608511/files/doc_financials/2026/Q326/Q3FY26-CFO-Commentary.pdf
“H20 sales were insignificant in the third quarter.”
“Revenue is expected to be $65.0 billion, plus or minus 2%.”
支持的研判. NVIDIA's CFO commentary for the third quarter of fiscal 2026 says H20 sales were insignificant and guides fourth-quarter revenue of US$65.0 billion, plus or minus 2%.
無法據以確立. The outlook states neither an exclusion nor an inclusion of China; the report's reading of that silence rests on the whole document.
引用於: 表 14

S-514 KLA Corporation (Form 10-K filed with the U.S. SEC) — Form 10-K, KLA Corporation, fiscal year ended June 30, 2026
申報文件 · U.S. · 英文 · 6 August 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/319201/000031920126000027/klac-20260630.htm
“Failure to obtain export licenses has harmed and could continue to harm our backlog, requiring us to return substantial deposits received from customers in China for purchase orders, and/or further limiting our ability to meet our contractual obligations and sell our products or provide services to our customers in China.”
“For geographical revenue reporting, revenues are attributed to the geographic location in which the customer is located.”
取自來源表格之數值（轉錄）: China | $4,048,358 | 29.8 % | $4,042,567 | 33.3 % | $4,196,727 | 42.8 %
取自來源表格之數值（轉錄）: Total | $13,579,476 | 100.0 % | $12,156,162 | 100.0 % | $9,812,247 | 100.0 %
支持的研判. KLA's 10-K attributes revenue to the customer's location, reports China at 29.8% of fiscal 2026 revenue against 42.8% in fiscal 2024, and says failure to obtain export licences has harmed its backlog.
無法據以確立. Revenue by customer location; the filing does not split the fall in China's share between rules and fab siting.
引用於: 表 14; 第節 第7.1節

S-515 Marvell Technology, Inc. (Form 10-Q filed with the U.S. SEC) — Form 10-Q, Marvell Technology, Inc., quarterly period ended August 1, 2026
申報文件 · U.S. · 英文 · 28 August 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/1835632/000183563226000025/mrvl-20260801.htm
直接文件: https://investor.marvell.com/sec-filings/all-sec-filings/content/0001835632-26-000025/0001835632-26-000025.pdf
“For example, a substantial majority of the shipments made to China relate to sales to non-China based customers that have factories or contract manufacturing operations located within China.”
“These destinations of shipment are not necessarily indicative of the geographic location of the Company's end customers or the country in which the Company's end customers sell devices containing the Company's products.”
取自來源表格之數值（轉錄）： China | $ 1,161.5 | 42 % | $ 583.4 | 29 % | $ 2,219.4 | 43 % | $ 1,292.3 | 33 %
支持的研判. Marvell's 10-Q reports 42% of revenue for the quarter to 1 August 2026 as shipped to China and says a substantial majority relates to non-China-based customers' factories and contract manufacturers there.
無法據以確立. Revenue by shipment destination; the filing does not give the split by end customer.
引用於: 表 14; 第節 第7.1節

S-516 Lam Research Corporation (Form 10-K filed with the U.S. SEC) — Form 10-K, Lam Research Corporation, fiscal year ended June 28, 2026
申報文件 · U.S. · 英文 · 7 August 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/707549/000070754926000037/lrcx-20260628.htm
“For geographical reporting, revenue is attributed to the geographic location in which the customers' facilities are located”
“We present our revenues disaggregated by geographic region based on the location of customers' facilities to which products were shipped and services were rendered.”
取自來源表格之數值（轉錄）： China | 34 % | 34 %
取自來源表格之數值（轉錄）： China | $7,859,811 | $6,205,062 | $6,293,990
支持的研判. Lam Research's 10-K attributes revenue to the location of customers' facilities and reports China at 34% of fiscal 2026 revenue (US$7.9 billion of US$23.2 billion).
無法據以確立. Revenue by facility location, not by the customer's nationality or end use.
引用於: 第節 第7.1節

S-517 ASML Holding N.V. (Form 20-F filed with the U.S. SEC) — Form 20-F, ASML Holding N.V., fiscal year ended December 31, 2025
申報文件 · Netherlands · 英文 · 25 February 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/937966/000162828026011378/asml-20251231.htm
取自來源表格之數值（轉錄）： China | 7,251.8 | 48.6 | 10,195.1 | 72.7 | 9,519.7 | 70.9
取自來源表格之數值（轉錄）： Total | 27,558.5 | 5,799.8 | 28,262.9 | 7,234.0 | 32,667.3 | 8,234.8
支持的研判. ASML's 20-F reports net sales to China of EUR 9.5 billion of EUR 32.7 billion in 2025, against EUR 10.2 billion of EUR 28.3 billion in 2024.
無法據以確立. Rows of the 20-F's geographic table as extracted; the report uses the net-sales columns only.
引用於: 第節 第7.1節

S-518 DeepSeek-AI (Hugging Face model card) — DeepSeek-V4: Towards Highly Efficient Million-Token Context Intelligence
企業新聞稿 · China · 英文 · 24 April 2026 · 閱讀 27 September 2026
https://huggingface.co/deepseek-ai/DeepSeek-V4-Pro
直接文件: https://huggingface.co/deepseek-ai/DeepSeek-V4-Pro/raw/main/config.json
“We present a preview version of DeepSeek-V4 series, including two strong Mixture-of-Experts (MoE) language models — DeepSeek-V4-Pro with 1.6T parameters (49B activated) and DeepSeek-V4-Flash with 284B parameters (13B activated) — both supporting a context length of one million tokens.”
“In the 1M-token context setting, DeepSeek-V4-Pro requires only 27% of single-token inference FLOPs and 10% of KV cache compared with DeepSeek-V3.2.”
取自來源表格之數值（轉錄）： DeepSeek-V4-Pro | 1.6T | 49B | 1M | FP4 + FP8 Mixed*
取自來源表格之數值（轉錄）： DeepSeek-V4-Flash | 284B | 13B | 1M | FP4 + FP8 Mixed*
支持的研判. DeepSeek's V4 preview lists V4-Pro (1.6T parameters, 49B activated) and V4-Flash (284B, 13B), both with a 1M-token context and FP4 experts; at 1M tokens V4-Pro needs 27% of V3.2's single-token FLOPs and 10% of its KV cache.
無法據以確立. A preview model card with the developer's own comparisons; the cache ratio is against DeepSeek-V3.2 at one million tokens only.
引用於: 第節 第8.1節, 第8.2節, 第8.4節, 第8.5節, 第8.6節; 表 16 · **研判:** A-13

S-519 Alibaba Qwen team (Hugging Face model card) — Qwen/Qwen3.8-2.4T-A95B
企業新聞稿 · China · 英文 · 12 August 2026 · 閱讀 27 September 2026
https://huggingface.co/Qwen/Qwen3.8-2.4T-A95B
直接文件: https://huggingface.co/Qwen/Qwen3.8-2.4T-A95B/raw/main/config.json
“Number of Parameters: 2.4T in total and 95B activated”
“Hidden Layout: 23 × (3 × (Gated DeltaNet → MoE) → 1 × (Gated Attention → MoE))”
取自來源表格之數值（轉錄）： full_attention_interval | 4 | head_dim | 256 | num_attention_heads | 64 | num_key_value_heads | 4 | num_hidden_layers | 92 | linear_num_key_heads | 16 | linear_num_value_heads | 128 | max_position_embeddings | 262144 | dtype | bfloat16
取自來源表格之數值（轉錄）： license | other | license_name | qwen3.8-max
支持的研判. Qwen's model card gives Qwen3.8-2.4T-A95B 2.4T parameters with 95B activated, 23 blocks of three Gated DeltaNet layers and one gated-attention layer, 4 KV heads of 256 dimensions, and a custom licence.
無法據以確立. The developer's description and configuration; the report's weight and cache sizes are its own calculations.
引用於: 第節 第8.1節, 第8.2節, 第8.3節, 第C.7節; 表 16, 17

S-520 Z.ai (Zhipu AI) (GitHub repository README) — GLM-5.3 & GLM-5.2 & GLM-5.1 & GLM-5
企業新聞稿 · China · 英文 · 18 August 2026 · 閱讀 27 September 2026
https://github.com/zai-org/GLM-5
直接文件: https://huggingface.co/zai-org/GLM-5.3/raw/main/config.
json
“Compared to GLM-4.5, GLM-5 scales from 355B parameters (32B active) to 744B parameters (40B active), and increases pre-training data from 23T to 28.5T tokens.”
“GLM-5.3 uses the same base model as GLM-5.2 — every gain comes from post-training.”
取自來源表格之數值（轉錄）： num_hidden_layers | 78 | num_attention_heads | 64 | kv_lora_rank | 512 | qk_rope_head_dim | 64 | index_topk | 2048 | index_topk_freq | 4 | index_n_heads | 32 | index_head_dim | 128 | max_position_embeddings | 1048576 | quant_method | fp8 | fmt | e4m3
取自來源表格之數值（轉錄）： GLM-5.3 | 744B-A40B | FP8 | GLM-5.3-BF16 | 744B-A40B | BF16 | GLM-5.3-Flash | 320B-A18B | FP8
支持的研判. Z.ai's repository says GLM-5 scaled to 744B parameters with 40B active and that GLM-5.3 shares GLM-5.2's base model; the configuration gives a 512-dimension latent, 64 positional dimensions and 1M positions.
無法據以確立. The developer's description and configuration; the report's cache figures are its own calculations from them.
引用於: 第節 第8.1節, 第8.2節, 第C.7節; 表 16

S-521 MiniMax (Hugging Face model card) — MiniMaxAI/MiniMax-M3
企業新聞稿 · China · 英文 · 1 June 2026 · 閱讀 27 September 2026
https://huggingface.co/MiniMaxAI/MiniMax-M3
直接文件: https://huggingface.co/MiniMaxAI/MiniMax-M3/raw/main/config.json
“MiniMax-M3 is a native multimodal model with 1M context. It has ~428B parameters and ~23B activated parameters.”
“Compared with GQA, MSA dramatically reduces the attention compute and memory footprint while preserving model quality.”
取自來源表格之數值（轉錄）： num_hidden_layers | 60 | num_attention_heads | 64 | num_key_value_heads | 4 | head_dim | 128 | max_position_embeddings | 1048576
取自來源表格之數值（轉錄）： license | other | license_name | minimax-community
支持的研判. MiniMax describes MiniMax-M3 as a native multimodal model with a 1M context, about 428B parameters and 23B activated, whose MiniMax Sparse Attention cuts attention compute and memory against GQA.
無法據以確立. The developer's description; its speed-ups against M2 are its own measurements at 1M context.
引用於: 第節 第8.1節; 表 16

S-522 Alibaba Qwen team (Hugging Face model card) — Qwen/Qwen3.5-397B-A17B
企業新聞稿 · China · 英文 · 16 February 2026 · 閱讀 27 September 2026
https://huggingface.co/Qwen/Qwen3.5-397B-A17B
直接文件: https://huggingface.co/Qwen/Qwen3.5-397B-A17B/raw/main/config.json
“Hidden Layout: 15 * (3 * (Gated DeltaNet -> MoE) -> 1 * (Gated Attention -> MoE))”
“Context Length: 262,144 natively and extensible up to 1,010,000 tokens.”
取自來源表格之數值（轉錄）： full_attention_interval | 4 | head_dim | 256 | num_attention_heads | 32 | num_key_value_heads | 2 | num_hidden_layers | 60 | linear_num_key_heads | 16 | linear_num_value_heads | 64 | max_position_embeddings | 262144 | dtype | bfloat16
取自來源表格之數值（轉錄）： license | apache-2.0
支持的研判. Qwen's model card gives Qwen3.5-397B-A17B a layout of fifteen blocks of three Gated DeltaNet layers and one gated-attention layer, a 262,144-token native context extensible to about 1M, and an Apache 2.0 licence.
無法據以確立. The developer's description and configuration; the report's cache figures are its own calculations from them.
引用於: 第節 第8.1節, 第8.2節, 第C.7節; 表 16, 17

S-523 Alibaba Qwen team (Hugging Face model card) — Qwen/Qwen3.8-Flash-Next
企業新聞稿 · China · 英文 · 26 August 2026 · 閱讀 27 September 2026
https://huggingface.co/Qwen/Qwen3.8-Flash-Next
直接文件: https://huggingface.co/Qwen/Qwen3.8-Flash-Next/raw/main/README.md
“Hidden Layout: 12 × (3 × (Gated DeltaNet → MoE) → 1 × (Qwen Sparse Attention → MoE))”
“Context Length: 262,144 natively and extensible up to 1,000,000 tokens.”
取自來源表格之數值（轉錄）： license | other | license_name | qwen-community-1.0 | license_link | LICENSE
支持的研判. Qwen's model card gives Qwen3.8-Flash-Next a layout of three Gated DeltaNet layers to one Qwen Sparse Attention layer, a 262,144-token native context extensible to 1 million, and a custom community licence.
無法據以確立. The developer's description; it states no hardware set-up or memory requirement.
引用於: 第節 第8.1節; 表 16

S-524 Moonshot AI (Hugging Face model card) — moonshotai/Kimi-K2-Instruct-0905
企業新聞稿 · China · 英文 · September 2025 · 閱讀 27 September 2026
https://huggingface.co/moonshotai/Kimi-K2-Instruct-0905
直接文件: https://huggingface.co/moonshotai/Kimi-K2-Instruct-0905/raw/main/config.json
“Both the code repository and the model weights are released under the Modified MIT License.”
“Our model checkpoints are stored in the block-fp8 format, you can find it on Huggingface.”
取自來源表格之數值（轉錄）： num_hidden_layers | 61 | num_attention_heads | 64 | num_key_value_heads | 64 | kv_lora_rank | 512 | q_lora_rank | 1536 | qk_nope_head_dim | 128 | qk_rope_head_dim | 64 | v_head_dim | 128 | max_position_embeddings | 262144 | quant_method | fp8 | fmt | e4m3
取自來源表格之數值（轉錄）： Number of Experts | 384 | Selected Experts per Token | 8 | Number of Shared Experts | 1 | Vocabulary Size | 160K | Context Length | 256K | Attention Mechanism | MLA | Activation Function | SwiGLU
支持的研判. Moonshot AI's model card lists Kimi K2-Instruct-0905 as a 1T-parameter MoE with 32B activated, MLA over 61 layers with a 512-dimension latent, a 256K context, block-FP8 checkpoints and a Modified MIT licence.
無法據以確立. The developer's description and configuration; the report's cache figures are its own calculations.
引用於: 第節 第8.1節, 第8.2節, 第C.7節; 表 16, 17

S-525 MiniMax (Hugging Face model card) — MiniMaxAI/MiniMax-M1-80k
企業新聞稿 · China · 英文 · 16 June 2025 · 閱讀 27 September 2026
https://huggingface.co/MiniMaxAI/MiniMax-M1-80k
直接文件: https://huggingface.co/MiniMaxAI/MiniMax-M1-80k/raw/main/config.json
“We introduce MiniMax-M1, the world's first open-weight, large-scale hybrid-attention reasoning model.”
“Furthermore, the lightning attention mechanism in MiniMax-M1 enables efficient scaling of test-time compute – For example, compared to DeepSeek R1, M1 consumes 25% of the FLOPs at a generation length of 100K tokens.”
取自來源表格之數值（轉錄）： license | apache-2.0
取自來源表格之數值（轉錄）： num_hidden_layers | 80 | num_attention_heads | 64 | num_key_value_heads | 8 | head_dim | 128 | max_position_embeddings | 10240000 | num_local_experts | 32 | num_experts_per_tok | 2
支持的研判. MiniMax presents MiniMax-M1 as the first open-weight, large-scale hybrid-attention reasoning model, whose lightning attention needs 25% of DeepSeek R1's FLOPs at 100K generated tokens; it is Apache 2.0-licensed.
無法據以確立. The developer's own claims of priority and efficiency; the FLOPs comparison is its own, at one generation length.
引用於: 第節 第8.1節, 第8.2節

S-526 DeepSeek-AI (Hugging Face model card) — deepseek-ai/DeepSeek-V3.1
企業新聞稿 · China · 英文 · 21 August 2025 · 閱讀 27 September 2026
https://huggingface.co/deepseek-ai/DeepSeek-V3.1
直接文件: https://huggingface.co/deepseek-ai/DeepSeek-V3.1/raw/main/config.json
“Additionally, DeepSeek-V3.1 is trained using the UE8M0 FP8 scale data format on both model weights and activations to ensure compatibility with microscaling data formats.”
“This repository and the model weights are licensed under the MIT License.”
取自來源表格之數值（轉錄）： num_hidden_layers | 61 | kv_lora_rank | 512 | qk_rope_head_dim | 64 | quant_method | fp8 | fmt | e4m3 | scale_fmt | ue8m0
取自來源表格之數值（轉錄）： DeepSeek-V3.1 | 671B | 37B | 128K
支持的研判. DeepSeek's model card says V3.1 was trained with the UE8M0 FP8 scale format on weights and activations for compatibility with microscaling formats, and that the repository and weights are MIT-licensed.
無法據以確立. The developer's statement of format and licence; it does not say which hardware the format targets.
引用於: 第節 第8.1節, 第8.2節, 第8.5節, 第C.7節

S-527 Z.ai (Zhipu AI) (Hugging Face model card) — zai-org/GLM-5.3-Flash
企業新聞稿 · China · 英文 · 26 August 2026 · 閱讀 27 September 2026
https://huggingface.co/zai-org/GLM-5.3-Flash
直接文件: https://huggingface.co/zai-org/GLM-5.3-Flash/raw/main/config.json
“With 320B total parameters and just 18B active parameters, it outperforms GLM-5.2 across benchmarks and real-world workloads at one-tenth the price, while approaching Claude Opus 4.8 on coding and agentic benchmarks.”
取自來源表格之數值（轉錄）： license | mit
取自來源表格之數值（轉錄）： num_hidden_layers | 45 | num_attention_heads | 64 | kv_lora_rank | 512 | q_lora_rank | 1536 | index_topk | 2048 | max_position_embeddings | 1048576 | n_routed_experts | 288 | num_experts_per_tok | 8
支持的研判. Z.ai's model card lists GLM-5.3-Flash under the MIT licence with 320B total and 18B active parameters, and its configuration gives 45 layers, a 512-dimension latent and 1M positions.
無法據以確立. The developer's card; its benchmark and price comparisons are its own.
引用於: 第節 第8.1節; 表 16

S-528 DeepSeek-AI (Hugging Face model card) — deepseek-ai/DeepSeek-V3
企業新聞稿 · China · 英文 · 26 December 2024 · 閱讀 27 September 2026
https://huggingface.co/deepseek-ai/DeepSeek-V3
直接文件: https://huggingface.co/deepseek-ai/DeepSeek-V3/raw/main/config.json
“We present DeepSeek-V3, a strong Mixture-of-Experts (MoE) language model with 671B total parameters with 37B activated for each token.”
“This code repository is licensed under the MIT License. The use of DeepSeek-V3 Base/Chat models is subject to the Model License.”
取自來源表格之數值（轉錄）： num_hidden_layers | 61 | num_attention_heads | 128 | num_key_value_heads | 128 | kv_lora_rank | 512 | q_lora_rank | 1536 | qk_nope_head_dim | 128 | qk_rope_head_dim | 64 | v_head_dim | 128
取自來源表格之數值（轉錄）： DeepSeek-V3 | 671B | 37B | 128K
支持的研判. DeepSeek's model card describes DeepSeek-V3 as a mixture-of-experts model with 671B parameters, 37B activated per token and a 128K context under its Model License; the configuration gives 61 MLA layers with a 512-dimension latent and 64 positional dimensions.
無法據以確立. The developer's description and configuration; the cache sizes in the report are its own calculations from them.
引用於: 表 16, 17; 第節 第8.2節, 第8.3節, 第C.7節

S-529 DeepSeek-AI (Hugging Face model card) — DeepSeek-V4.1-Flash: Pushing the Limits of KV Cache Compression
企業新聞稿 · China · 英文 · 10 September 2026 · 閱讀 27 September 2026
https://huggingface.co/deepseek-ai/DeepSeek-V4.1-Flash
直接文件: https://huggingface.co/deepseek-ai/DeepSeek-V4.1-Flash/raw/main/config.json
“Combined with FP4 main KV caching (E2M1 format, one E4M3 scale per 16 channels), these designs reduce the global KV cache footprint to 890 bytes per token — roughly 1/4 of DeepSeek-V4-Flash.”
“We introduce DeepSeek-V4.1-Flash, a multimodal Mixture-of-Experts (MoE) model with 552B backbone parameters and support for contexts of up to one million tokens.”
取自來源表格之數值（轉錄）： num_hidden_layers | 40 | num_attention_heads | 64 | num_key_value_heads | 1 | head_dim | 512 | qk_rope_head_dim | 64 | index_topk | 512 | max_position_embeddings | 1048576 | sliding_window | 128
取自來源表格之數值（轉錄）： quant_method | fp8 | scale_fmt | ue8m0 | expert_dtype | fp4
支持的研判. DeepSeek says V4.1-Flash, a multimodal MoE with a 552B backbone and a 1M-token context, cuts its global KV cache to 890 bytes per token with FP4 caching, about a quarter of V4-Flash.
無法據以確立. The developer's own figures; the 890 bytes cover the global cache only, and persistent and sliding-window states are described separately.
引用於: 表 16, 17, 18, 20; 第節 第8.2節, 第8.3節, 第8.4節, 第8.6節, 第C.7節; 展望 7 · **研判:** A-13

S-530 Alibaba Qwen team (GitHub repository README) — Qwen3.8-Flash-Next
企業新聞稿 · China · 英文 · 26 August 2026 · 閱讀 27 September 2026
https://github.com/QwenLM/Qwen3.8-Flash-Next
直接文件: https://raw.githubusercontent.com/QwenLM/Qwen3.8-Flash-Next/main/README.md
“N-gram Embedding looks up a table using the local context to scale model capacity with very little extra computation; the embedding table can be offloaded to host memory and overlapped with model computation through asynchronous prefetching.”
“Qwen3.8-Flash-Next features a 125B-parameter main model, supplemented by an additional 51B N-gram embeddings, with 6B parameters activated per token.”
支持的研判. Qwen's repository describes Qwen3.8-Flash-Next as a 125B-parameter main model with 51B of N-gram embeddings and 6B activated per token, whose embedding table can be offloaded to host memory.
無法據以確立. The developer's description; host-memory offload is an option it describes, not a measured deployment.
引用於: 表 16; 第節 第8.2節

S-531 Z.ai (Zhipu AI) (Hugging Face model card) — zai-org/GLM-4.5
企業新聞稿 · China · 英文 · 28 July 2025 · 閱讀 27 September 2026
https://huggingface.co/zai-org/GLM-4.5
直接文件: https://huggingface.co/zai-org/GLM-4.5/raw/main/config.json
“GLM-4.5 has 355 billion total parameters with 32 billion active parameters, while GLM-4.5-Air adopts a more compact design with 106 billion total parameters and 12 billion active parameters.”
“All are executed on devices that natively support FP8 inference, ensuring both weights and cache are in FP8 format.”
取自來源表格之數值（轉錄）： num_hidden_layers | 92 | num_attention_heads | 96 | num_key_value_heads | 8 | head_dim | 128 | max_position_embeddings | 131072 | torch_dtype | bfloat16
取自來源表格之數值（轉錄）： GLM-4.5 | BF16 | H100 x 16 / H200 x 8 | sglang | GLM-4.5 | FP8 | H100 x 8 / H200 x 4 | sglang
支持的研判. Z.ai gives GLM-4.5 355B parameters with 32B active and 92 layers with 8 KV heads of 128 dimensions, and describes FP8 set-ups in which weights and cache are both FP8; the models are MIT-licensed.
無法據以確立. The developer's description and deployment guidance; the report's cache figure is its own calculation.
引用於: 表 16, 17

S-532 Z.ai (Zhipu AI) (Z.AI Developer Document) — GLM-5.3-Flash/FlashX
企業新聞稿 · China · 英文 · 26 August 2026 · 閱讀 27 September 2026
https://docs.z.ai/guides/vlm/glm-5.3-flash
“As the first open-source frontier model to combine sparse and linear attention, it significantly cuts computation and serving costs while preserving long-context quality — reducing attention computation and KV cache by 3.01× and 4.44× versus GLM-5.3.”
“GLM-5.3-Flash has 320B total parameters with 18B activated.”
支持的研判. Z.ai says GLM-5.3-Flash, with 320B parameters and 18B activated, combines sparse and linear attention to cut attention computation and KV cache by 3.01 and 4.44 times against GLM-5.3.
無法據以確立. The developer's own ratios against its own flagship; not independently measured.
引用於: 表 16; 第節 第8.2節

S-533 MiniMax (Hugging Face model card) — MiniMaxAI/MiniMax-M2
企業新聞稿 · China · 英文 · 27 October 2025 · 閱讀 27 September 2026
https://huggingface.co/MiniMaxAI/MiniMax-M2
直接文件: https://huggingface.co/MiniMaxAI/MiniMax-M2/raw/main/config.json
"Simpler capacity planning with smaller per-request memory and steadier tail latency."
"MiniMax-M2 redefines efficiency for agents. It's a compact, fast, and cost-effective MoE model (230 billion total parameters with 10 billion active parameters) built for elite performance in coding and agentic tasks, all while maintaining powerful general intelligence."
取自來源表格之數值（轉錄）： num_hidden_layers | 62 | num_attention_heads | 48 | num_key_value_heads | 8 | head_dim | 128 | max_position_embeddings | 196608 | quant_method | fp8 | fmt | float8_e4m3fn
支持的研判. MiniMax presents MiniMax-M2 as a compact MoE with 230B parameters and 10B active, promising simpler capacity planning with smaller per-request memory; the configuration gives 62 layers, 8 KV heads and FP8 weights.
無法據以確立. The developer's own positioning; per-request memory depends on context length and the serving set-up.
引用於: 表 16, 17; 第節 第8.2節

S-534 OpenAI (Hugging Face model card) — openai/gpt-oss-120b
企業新聞稿 · U.S. · 英文 · 5 August 2025 · 閱讀 27 September 2026
https://huggingface.co/openai/gpt-oss-120b
直接文件: https://huggingface.co/openai/gpt-oss-120b/raw/main/config.json
"The models were post-trained with MXFP4 quantization of the MoE weights, making gpt-oss-120b run on a single 80GB GPU (like NVIDIA H100 or AMD MI300X) and the gpt-oss-20b model run within 16GB of memory."
"gpt-oss-120b — for production, general purpose, high reasoning use cases that fit into a single 80GB GPU (like NVIDIA H100 or AMD MI300X) (117B parameters with 5.1B active parameters)"
取自來源表格之數值（轉錄）： num_hidden_layers | 36 | num_attention_heads | 64 | num_key_value_heads | 8 | head_dim | 64 | sliding_window | 128 | max_position_embeddings | 131072 | num_local_experts | 128 | experts_per_token | 4 | quant_method | mxfp4
支持的研判. OpenAI says gpt-oss-120b (117B parameters, 5.1B active) was post-trained with MXFP4 MoE weights so that it runs on a single 80 GB GPU; the configuration gives 36 layers and a 128-token sliding window.
無法據以確立. The developer's statement; the alternating layer pattern is read from the configuration and the report's cache figure is its own calculation.
引用於: 表 16, 17; 第節 第8.2節, 第8.5節

S-535 Meta (llama-models GitHub repository model card) — llama-models/models/llama3_3/MODEL_CARD.md
企業新聞稿 · U.S. · 英文 · 6 December 2024 · 閱讀 27 September 2026
https://github.com/meta-llama/llama-models/blob/main/models/llama3_3/MODEL_CARD.md
直接文件: https://github.com/meta-llama/llama-models/blob/main/models/sku_list.py
"All model versions use Grouped-Query Attention (GQA) for improved inference scalability."
"A custom commercial license, the Llama 3.3 Community License Agreement, is available at: https://github.com/meta-llama/llama-models/blob/main/models/llama3_3/LICENSE"
取自來源表格之數值（轉錄）： dim | 8192 | n_layers | 80 | n_heads | 64 | n_kv_heads | 8
取自來源表格之數值（轉錄）： Llama 3.3 (text only) | A new mix of publicly available online data. | 70B | Multilingual Text | Multilingual Text and code | 128k | Yes | 15T+ | December 2023
支持的研判. Meta's model card says Llama 3.3 70B uses grouped-query attention with a 128k context under the Llama 3.3 Community License; its parameters give 80 layers, 64 heads and 8 KV heads at dimension 8,192.
無法據以確立. The developer's card and parameters; the report's cache figure is its own calculation at 16 bits.
引用於: 表 16, 17; 第節 第8.3節

S-536 Malika Aubakirova et al. (OpenRouter Inc.; Andreessen Horowitz), arXiv — State of AI: An Empirical 100 Trillion Token Study with OpenRouter
研究 · U.S. · 英文 · 15 January 2026 · 閱讀 27 September 2026
https://arxiv.org/html/2601.10088v1
"Starting from a negligible base in late 2024 (weekly share as low as 1.2%), Chinese OSS models steadily gained traction, reaching nearly 30% of total usage among all models in some weeks."
"Over the one-year window, they averaged approximately 13.0% of weekly token volume, with strong growth concentrated in the second half of 2025."
支持的研判. The OpenRouter study of over 100 trillion tokens finds Chinese open-source models rose from weekly shares as low as 1.2% in late 2024 to nearly 30% in some weeks, averaging about 13.0% over the year.
無法據以確立. Platform-level observations, not ecosystem-wide conclusions, as the authors say; 'Chinese' includes developers in Taiwan and Hong Kong.
引用於: 第節 第8.1節

S-537 OpenRouter Inc. (company website) — AI Model Rankings
企業新聞稿 · U.S. · 英文 · 27 September 2026 · 閱讀 27 September 2026
https://openrouter.ai/rankings
"Share of text requests made on OpenRouter in the week beginning Sep 14, 2026, with the change in each author's requests against the week before it."
"Models ranked by tokens processed on OpenRouter over the trailing thirty days"
取自來源表格之數值（轉錄）： 1 | deepseek | 25.4% | +9%
取自來源表格之數值（轉錄）： 1 | GLM 5.3 Flash | z-ai | 58.1T | >999%
支持的研判. OpenRouter's rankings for the week beginning 14 September 2026 give DeepSeek 25.4% of text requests, the largest author share, and rank Z.ai's GLM 5.3 Flash first by tokens over the trailing thirty days.
無法據以確立. One platform's routed traffic, which changes weekly; it excludes direct APIs, private deployments and other gateways.
引用於: 第節 第8.1節, 第8.3節

S-538 Vercel Inc. (company blog; Amelia Charles, Eric Dodds) — Open-weight models take 56% of token volume, Astra doubles Fable 5.1 spend
企業新聞稿 · U.S. · 英文 · 17 September 2026 · 閱讀 27 September 2026
https://vercel.com/blog/ai-gateway-production-index-september-2026
"In August, open-weight models ran 56% of all tokens on AI Gateway, marking the first month they took the majority of volume."
"Open-weight models processed 56% of August's gateway tokens and accounted for 14% of spend."
支持的研判. Vercel says open-weight models ran 56% of AI Gateway tokens in August 2026, the first month with a majority, on 14% of spend; in December 2025 they ran fewer than one in ten.
無法據以確立. One gateway's customer traffic; 'open-weight' covers all origins, and token and spend shares depend on its customer mix.
引用於: 第節 第8.1節, 第8.3節; 圖 31

S-539 Menlo Ventures (Tim Tully, Joff Redfern, Deedy Das, Derek Xiao) — 2025: The State of Generative AI in the Enterprise
市場研究 · U.S. · 英文 · 9 December 2025 · 閱讀 27 September 2026
https://menlovc.com/perspective/2025-the-state-of-generative-ai-in-the-enterprise/
"But the model's stagnation—including no new major releases since the April release of Llama 4—has contributed to a decline in overall enterprise open-source share from 19% last year to 11% today."
"Collectively, they account for just 1% of total LLM API usage (roughly 10% of enterprise open-source)."
支持的研判. Menlo Ventures' survey of 495 U.S. enterprise decision-makers finds enterprise open-source share fell from 19% to 11%, with Chinese open-source models about 1% of total LLM API usage.
無法據以確立. A November 2025 survey of U.S. enterprises by an investor; shares are respondents' reported usage.
引用於: 第節 第8.1節

S-540 Hugging Face (company blog; Adina Yakefu, Apolinário, Irene Solaiman) — State of Open Models: Summer 2026 Observations
企業新聞稿 · U.S. · 英文 · 14 August 2026 · 閱讀 27 September 2026
https://huggingface.co/blog/state-of-open-models-summer-2026
"Qwen-based models now account for 151,448 derivatives on the Hub, 2.6× Meta's total footprint and 4.7× the Llama repositories specifically."
"Downloads indicate usage within the Hub ecosystem, but they do not capture API usage, private deployments, or models distributed through other channels."
支持的研判. Hugging Face counts 151,448 derivative repositories built on Qwen, 2.6 times Meta's footprint, and notes that downloads do not capture API usage, private deployments or other channels.
無法據以確立. Counts of derivatives and downloads on one hub, which the authors say do not measure usage elsewhere.
引用於: 第節 第8.1節

S-541 National Institute of Standards and Technology (NIST), Center for AI Standards and Innovation (CAISI), U.S. Department of Commerce — CAISI Evaluation of DeepSeek V4 Pro
政府 · U.S. · 英文 · 1 May 2026 · 閱讀 27 September 2026
https://www.nist.gov/news-events/news/2026/05/caisi-evaluation-deepseek-v4-pro
"CAISI evaluations indicate that DeepSeek V4's capabilities lag behind the frontier by about 8 months (Figure 1)."
"DeepSeek V4 is more cost efficient than other models of similar capability."
支持的研判. CAISI estimates that DeepSeek V4's capabilities lag the frontier by about eight months, and finds it more cost efficient than other models of similar capability and than GPT-5.4 mini on 5 of 7 benchmarks.
無法據以確立. A U.S. government evaluation with its own benchmark and cost method; results apply to the version tested.
引用於: 第節 第8.5節

S-542 Amazon Web Services (What's New) — Kimi K3 by Moonshot AI is now generally available on Amazon Bedrock
企業新聞稿 · U.S. · 英文 · 18 September 2026 · 閱讀 27 September 2026
https://aws.amazon.com/about-aws/whats-new/2026/09/moonshot-ai-kimi-k3-on-amazon-bedrock/
"Kimi K3 is the first open weight model on Amazon Bedrock to support explicit prompt caching, helping reduce latency and input costs when reusing context across model calls."
"Today, Kimi K3 from Moonshot AI is generally available on Amazon Bedrock, giving you a powerful new option for coding and knowledge work."
支持的研判. AWS says Kimi K3 is generally available on Amazon Bedrock and is the first open-weight model there to support explicit prompt caching, reducing latency and input costs when context is reused.
無法據以確立. A provider announcement; it shows availability and features, not usage volumes.
引用於: 第節 第8.1節, 第8.4節, 第C.7節

S-543 Microsoft (Microsoft Foundry Blog; Nick Brady) — What's new in Microsoft Foundry | May 2026
企業新聞稿 · U.S. · 英文 · 31 May 2026 · 閱讀 27 September 2026
https://devblogs.microsoft.com/foundry/whats-new-in-microsoft-foundry-may-2026/
"DeepSeek V4 variants are now available in Microsoft Foundry."
取自來源表格之數值（轉錄）： DeepSeek V4 Pro | High-precision open-model reasoning via Fireworks | Complex reasoning, coding, agent workflows
取自來源表格之數值（轉錄）： Kimi 2.6 | Long-horizon reasoning and agentic workflows via Fireworks | Extended reasoning chains, coding, multi-step agents
支持的研判. Microsoft says DeepSeek V4 variants are available in Microsoft Foundry, listing DeepSeek V4 Pro and Kimi 2.6 as served via Fireworks.
無法據以確立. A release-notes listing of availability; no usage, pricing or hardware information.
引用於: 第節 第8.1節

S-544 Google Cloud (documentation) — Vertex AI release notes
企業新聞稿 · U.S. · 英文 · 10 February 2026 · 閱讀 27 September 2026
https://docs.cloud.google.com/vertex-ai/generative-ai/docs/release-notes
“GLM 5 is available as a managed API in Model Garden.”
“DeepSeek-V3.2 is available as a managed API in Model Garden.”
支持的研判. Google's Vertex AI release notes list GLM 5, DeepSeek-V3.2, Kimi K2 Thinking, MiniMax M2 and Qwen3 models as managed APIs or model-as-a-service offerings in Model Garden.
無法據以確立. Release notes of availability; no usage or hardware information.
引用於: 第節 第8.1節

S-545 Alibaba Qwen team (Hugging Face model card) — Qwen/Qwen3-Next-80B-A3B-Instruct
企業新聞稿 · China · 英文 · 11 September 2025 · 閱讀 27 September 2026
https://huggingface.co/Qwen/Qwen3-Next-80B-A3B-Instruct
直接文件: https://huggingface.co/Qwen/Qwen3-Next-80B-A3B-Instruct/raw/main/config.json
“Hybrid Attention: Replaces standard attention with the combination of Gated DeltaNet and Gated Attention, enabling efficient context modeling for ultra-long context length.”
“Hybrid Layout: 12 * (3 * (Gated DeltaNet -> MoE) -> 1 * (Gated Attention -> MoE))”
取自來源表格之數值（轉錄）: full_attention_interval | 4 | head_dim | 256 | num_attention_heads | 16 | num_key_value_heads | 2 | num_hidden_layers | 48 | linear_num_key_heads | 16 | linear_num_value_heads | 32 | linear_key_head_dim | 128 | linear_value_head_dim | 128 | max_position_embeddings | 262144 | torch_dtype | bfloat16
支持的研判. Qwen's model card says Qwen3-Next replaces standard attention with Gated DeltaNet and gated attention, in a layout of three DeltaNet layers to one attention layer repeated twelve times.
無法據以確立. The developer's description of its architecture; the efficiency claims in the card are its own benchmarks.
引用於: 第節 第8.2節

S-546 Moonshot AI (Hugging Face model card) — moonshotai/Kimi-Linear-48B-A3B-Instruct
企業新聞稿 · China · 英文 · 30 October 2025 · 閱讀 27 September 2026
https://huggingface.co/moonshotai/Kimi-Linear-48B-A3B-Instruct
直接文件: https://huggingface.co/moonshotai/Kimi-Linear-48B-A3B-Instruct/raw/main/config.json
“Hybrid Architecture: A 3:1 KDA-to-global MLA ratio reduces memory usage while maintaining or surpassing the quality of full attention.”
“It reduces the need for large KV caches by up to 75% and boosts decoding throughput by up to $6\times$ for contexts as long as 1M tokens.”
取自來源表格之數值（轉錄）: num_hidden_layers | 27 | full_attn_layers | 4, 8, 12, 16, 20, 24, 27 | kv_lora_rank | 512 | qk_rope_head_dim | 64 | num_attention_heads | 32 | model_max_length | 1048576 | dtype | bfloat16
取自來源表格之數值（轉錄）: Kimi-Linear-Instruct | 48B | 3B | 1M
支持的研判. Moonshot AI says Kimi Linear's 3:1 ratio of Kimi Delta Attention to global MLA layers cuts KV-cache needs by up to 75% and raises decoding throughput up to sixfold at 1M tokens.
無法據以確立. The developer's figures for a 48B research model; 'up to' values at long contexts, not averages.
引用於: 第節 第8.2節

S-547 Moonshot AI (Hugging Face model card) — moonshotai/Kimi-K2-Thinking
企業新聞稿 · China · 英文 · November 2025 · 閱讀 27 September 2026
https://huggingface.co/moonshotai/Kimi-K2-Thinking
直接文件: https://huggingface.co/moonshotai/Kimi-K2-Thinking/raw/main/config.json
“K2 Thinking is a native INT4 quantization model with 256k context window, achieving lossless reductions in inference latency and GPU memory usage.”
“It allows K2 Thinking to support native INT4 inference with a roughly 2x generation speed improvement while achieving state-of-the-art performance.”
取自來源表格之數值（轉錄）: num_hidden_layers | 61 | kv_lora_rank | 512 | qk_rope_head_dim | 64 | max_position_embeddings | 262144 | quant_method | compressed-tensors | format | pack-quantized | num_bits | 4 | group_size | 32 | type | int | kv_cache_scheme | null
支持的研判. Moonshot AI says Kimi K2 Thinking is a native INT4 model, trained with quantization-aware training on its MoE weights, with roughly twice the generation speed at state-of-the-art performance.
無法據以確立. The developer's own performance claims; speed-up and quality are not independently measured.
引用於: 第節 第8.2節

S-548 Moonshot AI (Kimi blog) — Kimi K3: Open Frontier Intelligence
企業新聞稿 · China · 英文 · July 2026 · 閱讀 27 September 2026
https://www.kimi.ai/blog/kimi-k3
“Powered by Mooncake's disaggregated inference architecture, the official Kimi API achieves a cache hit rate above 90% in coding workloads.”
“Since inference efficiency likewise benefits from larger high-bandwidth communication domains, we recommend deploying Kimi K3 on supernode configurations with 64 or more accelerators.”
支持的研判. Moonshot AI says Mooncake's disaggregated serving gives the Kimi API a cache-hit rate above 90% in coding workloads, and recommends deploying Kimi K3 on supernodes of 64 or more accelerators.
無法據以確立. The developer's own serving statistic and recommendation; the hit rate applies to coding workloads on its own API.
引用於: 第節 第8.2節, 第8.3節, 第8.4節

S-549 TrendForce — Stronger North American Demand to Keep Enterprise SSD Prices on an Upward Trend in 4Q26 (TrendForce press release)
市場研究 · Taiwan · 英文 · 21 September 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260921-13246.html
“In China, AI architectures led by DeepSeek are promoting the use of high-capacity QLC enterprise SSDs for KV cache offloading.”
“By selectively offloading cached data from memory to high-speed, high-capacity QLC enterprise SSDs, these architectures can improve model inference efficiency at a lower cost.”
支持的研判. TrendForce reports that DeepSeek-led AI architectures in China are promoting high-capacity QLC enterprise SSDs for KV-cache offloading, and that enterprise SSD orders in 4Q26 could exceed 3Q26's elevated level.
無法據以確立. A market research press release; it does not quantify SSD volumes used for KV caches.
引用於: 第節 第8.2節, 第8.5節, 第8.6節; 表 18; 展望 7

S-550 NVIDIA (NVIDIA Technical Blog) — Inside NVIDIA Rubin GPU Architecture: Powering the Era of Agentic AI
企業新聞稿 · U.S. · 英文 · 21 July 2026 · 閱讀 27 September 2026
https://developer.nvidia.com/blog/inside-nvidia-rubin-gpu-architecture-powering-the-era-of-agentic-ai/
“The decode, or generation, phase of inference is fundamentally memory subsystem bound.”
“Rubin also provides up to 288 GB of HBM4 per GPU, increasing available on-package capacity relative to Blackwell.”
支持的研判. NVIDIA's technical blog says decode is fundamentally memory-subsystem bound, and that Rubin provides up to 288 GB of HBM4 per GPU with up to 22 TB/s of bandwidth, 2.8 times Blackwell.
無法據以確立. A vendor's architecture blog of July 2026; ' up to' specifications may differ from product pages and shipping systems.
引用於: 第節 第8.3節; 展望 15

S-551 NVIDIA Corporation (nvidia.com product page) — Rack-Scale Agentic AI Supercomputer | NVIDIA Vera Rubin NVL72
企業新聞稿 · U.S. · 英文 · 21 September 2026 · 閱讀 27 September 2026
https://www.nvidia.com/en-us/data-center/vera-rubin-nvl72/
“Agentic systems consume up to 15x more tokens than traditional AI applications.”
“AI factories must deliver on token volume and massive context windows with low latency and efficient economics.”
取自來源表格之數值（轉錄）: GPU Memory | Bandwidth | 20.7 TB HBM4 | 1,400 TB/s | 576 GB HBM4 | 38.5 TB/s | 288 GB HBM4 | 19.2 TB/s
取自來源表格之數值（轉錄）: CPU Memory | Up to 54 TB LPDDR5X | Up to 1.5 TB LPDDR5X
支持的研判. NVIDIA's product page gives a Vera Rubin NVL72 rack 20.7 TB of HBM4 at 1,400 TB/s (288 GB and 19.2 TB/s per GPU) and up to 54 TB of LPDDR5X, and says agentic systems consume up to 15x more tokens.
無法據以確立. A product page; specifications may change before shipment, and the token multiple is NVIDIA's claim without a stated basis.
引用於: 第節 第8.3節, 第8.4節; 圖 29; 表 18; 展望 15

S-552 TrendForce — Tight DRAM Supply Gives Suppliers Greater Pricing Power in HBM, with HBM Contract Prices Expected to Surge Multiples Higher in 2027
市場研究 · Taiwan · 英文 · 2 June 2026 · 閱讀 27 September 2026
https://www.trendforce.com/presscenter/news/20260602-13074.html
“In 2027, NVIDIA’s Rubin Ultra platform is expected to further increase HBM capacity per GPU to 384GB.”
“In 2026, HBM demand growth will primarily be driven by AI ASIC capacity upgrades, with HBM capacity per AI chip increasing significantly from 96GB/192GB to 216GB/288GB.”
支持的研判. TrendForce expects HBM per AI chip to rise from 96-192 GB to 216-288 GB in 2026, and NVIDIA's Rubin Ultra platform to raise HBM per GPU to 384GB in 2027.
無法據以確立. Market research expectations (F) of June 2026; TrendForce later reported lower Rubin Ultra configurations under evaluation.
引用於: 第節 第8.3節; 圖 29

S-553 Google (The Keyword blog; Sundar Pichai, CEO of Google and Alphabet) — Q2 2026 earnings call: Remarks from our CEO
企業新聞稿 · U.S. · 英文 · 22 July 2026 · 閱讀 27 September 2026
https://blog.google/company-news/inside-google/message-ceo/alphabet-earnings-q2-2026/
“Our model APIs are now processing approximately 22 billion tokens per minute.”
“That's up from 16 billion just a quarter ago.”
支持的研判. Alphabet's chief executive says its model APIs process about 22 billion tokens a minute, up from 16 billion a quarter earlier.
無法據以確立. Earnings remarks on API traffic only; not comparable with Google's monthly figure for all its surfaces.
引用於: 第節 第8.4節

S-554 National Data Administration of the People's Republic of China (国家数据局) — 刘烈宏出席中国发展高层论坛2026年年会并演讲（全文）
政府 · China · 簡體中文 · 23 March 2026 · 閱讀 27 September 2026
https://www.nda.gov.cn/sjj/jgsz/jld/llh/llhldhd/0323/20260323202046805537 21_pc.html
“2024年年初，中国日均Token调用量为1000亿；至2025年底，跃升至100万亿；今年3月，已突破140万亿，两年增长超千倍。”
APEX譯文: At the start of 2024, China's average daily token calls were 100 billion; by the end of 2025 they had jumped to 100 trillion; in March this year they exceeded 140 trillion, growth of more than a thousandfold in two years.
“截至2025年底，中国智算总规模达159万PFlops，其中，八大国家枢纽（含十大集群）已建成智算规模达138.8万PFlops，占全国比重超过80%。”
APEX譯文: By the end of 2025, China's total intelligent computing capacity reached 1.59 million PFlops, of which the eight national hub nodes (including ten clusters) had built 1.388 million PFlops, over 80% of the national total.
支持的研判. The head of China's National Data Administration says average daily token calls in China rose from 100 billion in early 2024 to 100 trillion at the end of 2025 and over 140 trillion in March 2026.

無法據以確立. An official's speech; the counting basis for token calls is not defined, and the figures are not independently verified.
引用於: 第節 第8.4節, 第8.6節; 圖 31

S-555 財联社 (Cailian Press), via 中国金融信息网 (cnfin.com) — 在2026火山引擎force原动力大会上，火山引擎总裁谭待表示，截至2026年6月，豆包大模型日均tokens调用量达180万亿，较发布时增长超1500倍，过去一年增长超10倍。
媒體 · China · 簡體中文 · 23 June 2026 · 閱讀 27 September 2026
https://www.cnfin.com/kx/detail/20260623/4430319_1.html
“在2026火山引擎force原动力大会上，火山引擎总裁谭待表示，截至2026年6月，豆包大模型日均tokens调用量达180万亿，较发布时增长超1500倍，过去一年增长超10倍。”
APEX譯文: At the 2026 Volcano Engine FORCE conference, Volcano Engine president Tan Dai said that as of June 2026 the Doubao models' average daily token calls reached 180 trillion, more than 1,500 times the level at launch and more than tenfold growth over the past year.
支持的研判. Cailian Press reports Volcano Engine's president saying that the Doubao models averaged 180 trillion token calls a day as of June 2026, more than ten times a year earlier.
無法據以確立. A media report of a company statement; the counting basis is not defined.
引用於: 第節 第8.4節

S-556 Vercel Inc. (company blog; Amelia Charles, Harpreet Arora) — Open-weight models surge to 29% of volume, price per token flattens
企業新聞稿 · U.S. · 英文 · 13 July 2026 · 閱讀 27 September 2026
https://vercel.com/blog/ai-gateway-production-index-july-2026
“Open-weight models ran 29% of gateway tokens, up from 11% in April, on under 4% of spend.”
“This analysis is based on anonymized, aggregate routing data from the Vercel AI Gateway through June 2026.”
支持的研判. Vercel says open-weight models ran 29% of AI Gateway tokens, up from 11% in April, on under 4% of spend, based on routing data through June 2026.
無法據以確立. One gateway's anonymised, aggregated customer traffic; not an industry-wide measure.
引用於: 圖 31

S-557 OpenRouter Inc. (with Andreessen Horowitz co-authors) — State of AI 2025: 100T Token LLM Usage Study
研究 · U.S. · 英文 · December 2025 · 閱讀 27 September 2026
https://openrouter.ai/state-of-ai
“Average prompt tokens per request have increased roughly fourfold from around 1.5K to over 6K while completions have nearly tripled from about 150 to 400 tokens.”
“What was effectively a negligible slice of usage in early Q1 now exceeds fifty percent.”
支持的研判. OpenRouter's study reports that reasoning-optimised models rose to over half of its tokens in 2025, and that average prompt tokens per request rose about fourfold to over 6,000 while completions nearly tripled.
無法據以確立. One routing platform's traffic in 2025; its customer mix differs from enterprise and consumer use.
引用於: 第節 第8.4節; 展望 15

S-558 Anthropic (Engineering blog) — How we built our multi-agent research system
企業新聞稿 · U.S. · 英文 · 13 June 2025 · 閱讀 27 September 2026
https://www.anthropic.com/engineering/built-multi-agent-research-system
“In our data, agents typically use about 4× more tokens than chat interactions, and multi-agent systems use about 15× more tokens than chats.”
“There is a downside: in practice, these architectures burn through tokens fast.”
支持的研判. Anthropic reports that in its data agents typically use about four times the tokens of chat interactions and multi-agent systems about fifteen times.
無法據以確立. Figures from the developer's own research system and data; not an industry-wide measure.
引用於: 第節 第8.4節; 展望 15

S-559 DeepSeek-AI (Hugging Face model card) — deepseek-ai/DeepSeek-R1-0528
企業新聞稿 · China · 英文 · 28 May 2025 · 閱讀 27 September 2026
https://huggingface.co/deepseek-ai/DeepSeek-R1-0528
直接文件: https://huggingface.co/deepseek-ai/DeepSeek-R1-0528/raw/main/config.json
“This advancement stems from enhanced thinking depth during the reasoning process: in the AIME test set, the previous model used an average of 12K tokens per question, whereas the new version averages 23K tokens per question.”
“The DeepSeek R1 model has undergone a minor version upgrade, with the current version being DeepSeek-R1-0528.”
取自來源表格之數值（轉錄）: num_hidden_layers | 61 | num_attention_heads | 128 | kv_lora_rank | 512 | qk_rope_head_dim | 64 | max_position_embeddings | 163840 | quant_method | fp8
支持的研判. DeepSeek says R1-0528 used an average of 23K tokens per AIME question against 12K for the previous version, reflecting deeper reasoning.
無法據以確立. One benchmark's average, stated by the developer; it does not measure tokens per task in production workloads.
引用於: 第節 第8.4節

S-560 OpenAI — Introducing GPT-5
企業新聞稿 · U.S. · 英文 · 7 August 2025 · 閱讀 27 September 2026
https://openai.com/index/introducing-gpt-5/
“In our evaluations, GPT-5 (with thinking) performs better than OpenAI o3 with 50-80% less output tokens across capabilities, including visual reasoning, agentic coding, and graduate-level scientific problem solving.”
“GPT-5 gets more value out of less thinking time.”
支持的研判. OpenAI says GPT-5 with thinking outperforms o3 with 50-80% fewer output tokens across capabilities in its evaluations.
無法據以確立. The developer's own evaluations; output-token savings vary by task and are not independently measured.
引用於: 第節 第8.4節

S-561 DeepSeek (DeepSeek API Docs, Chinese news page) — DeepSeek-V3.1 发布
企業新聞稿 · China · 簡體中文 · 21 August 2025 · 閱讀 27 September 2026
https://api-docs.deepseek.com/zh-cn/news/news250821
“我们的测试结果显示，经过思维链压缩训练后，V3.1-Think 在输出 token 数减少 20%-50% 的情况下，各项任务的平均表现与 R1-0528 持平。”
APEX譯文: Our test results show that after chain-of-thought compression training, V3.1-Think matches R1-0528 in average performance across tasks while its output tokens fall by 20%-50%.
“需要注意的是，DeepSeek-V3.1 使用了 UE8M0 FP8 Scale 的参数精度。”
APEX譯文: Note that DeepSeek-V3.1 uses UE8M0 FP8 Scale parameter precision.
支持的研判. DeepSeek's Chinese release says V3.1-Think matches R1-0528's average performance with 20-50% fewer output tokens after chain-of-thought compression training, and that V3.1 uses UE8M0 FP8 scale precision.
無法據以確立. The developer's own test results, averaged over its chosen tasks; not an independent measurement.
引用於: 第節 第8.4節

S-562 Alibaba Group (Alizila news site) — Alibaba Cloud's 2026 Apsara Conference: Full-Stack AI Roadmap Along with Global Market Expansion Plan
企業新聞稿 · China · 英文 · 24 September 2026 · 閱讀 27 September 2026
https://www.alizila.com/alibaba-clouds-2026-apsara-conference-full-stack-ai-roadmap-along-with-global-market-expansion-plan/
“Agent Context connects a company's documents, business systems, chat records and multimodal data into a single foundation, cutting token usage by up to 67%.”
“The Zhenwu V900 delivers three times the performance of its predecessor, the Zhenwu M890, featuring 216 GB of GPU memory and 1,200 GB/s of inter-chip bandwidth with native support for FP8 and FP4.”
支持的研判. Alibaba Cloud says its Agent Context cuts token usage by up to 67%, and that its Zhenwu V900 chip, with 216 GB of GPU memory and native FP8 and FP4 support, enters mass production in the first quarter of 2027.
無法據以確立. Company announcements; 'up to 67%' is a best case, and the chip's volumes and memory source are not stated.
引用於: 第節 第8.4節

S-563 Microsoft Corporation (Investor Relations) — Microsoft Fiscal Year 2025 Second Quarter Earnings Conference Call
企業新聞稿 · U.S. · 英文 · 29 January 2025 · 閱讀 27 September 2026
https://www.microsoft.com/en-us/investor/events/fy-2025/earnings-fy-2025-q2
“On inference, we have typically seen more than 2X price-performance gain for every hardware generation, and more than 10X for every model generation due to software optimizations.”
“You don’t want to buy too much of anything at one time because of the Moore’s Law every year is going to give you 2x. Your optimization is going to give you 10x.”
支持的研判. Microsoft's chief executive said inference typically gains more than 2X in price-performance per hardware generation and more than 10X per model generation through software, and cautioned against buying too much at one time.
無法據以確立. Earnings-call remarks of 29 January 2025; the multiples are typical values the company observed, not measurements.
引用於: 第節 第8.4節

S-564 DeepSeek (DeepSeek API Docs) — Models & Pricing
企業新聞稿 · China · 英文 · 10 September 2026 · 閱讀 27 September 2026
https://api-docs.deepseek.com/quick_start/pricing/
“The prices listed below are in units of per 1M tokens.”
“Off-peak rates are half of the peak rates.”
取自來源表格之數值（轉錄）: 1M INPUT TOKENS (CACHE HIT) | OFF-PEAK | $0.003 | $0.022 | PEAK | $0.006 | $0.044
取自來源表格之數值（轉錄）: 1M INPUT TOKENS (CACHE MISS) | OFF-PEAK | $0.15 | $0.66 | PEAK | $0.3 | $1.32
支持的研判. DeepSeek's pricing page lists V4.1-Flash at US$0.006 per million input tokens on a cache hit and US$0.30 on a miss at peak hours, with off-peak rates half of peak.
無法據以確立. List prices at the date read; they change often and do not show actual costs or discounts.
引用於: 第節 第8.4節, 第C.7節

S-565 SK hynix Inc. (SK hynix Newsroom) — The Next-Generation Memory Architecture in the AI Era? SK hynix Charts the Direction at 'FMS 2026'
企業新聞稿 · Korea · 英文 · 7 August 2026 · 閱讀 27 September 2026
https://news.skhynix.com/en/fms-2026/
“Through this, SK hynix demonstrated that transferring or reusing the KV cache on the basis of CXL Pooled Memory can significantly improve the performance of an AI Agent Serving system compared with legacy systems that rely on network communication.”
“It also proved that storing the entire KV cache on SSD can deliver fast read performance and large-capacity storage at the same time.”
支持的研判. SK hynix says it demonstrated at FMS 2026 that transferring or reusing KV cache over CXL pooled memory improves AI-agent serving against network-based systems, and that SSDs can hold the entire KV cache.
無法據以確立. A company account of its own demonstrations; no deployment, customer or volume figures.
引用於: 第節 第8.4節; 表 18; 展望 15

S-566 KIOXIA Holdings Corporation — Flash Memory Scales AI Inference / Investor Day
企業新聞稿 · Japan · 英文 · 2 June 2026 · 閱讀 27 September 2026
https://www.kioxia-holdings.com/en-jp/ir/library/event.html
直接文件: https://www.kioxia-holdings.com/content/dam/kioxia-hd/en-jp/ir/library/event/asset/Kioxia_Investor_Day_2026_en_script.pdf
“Within this segment, inference AI applications will serve as a primary growth catalyst, and are likely to achieve an estimated CAGR of 86%.”
“Notably, datacenter demand is expected to grow significantly, by 46%, with increasing AI server capacity driving market expansion.”
支持的研判. Kioxia expects data-centre NAND demand to grow by 46% with AI server capacity, and inference AI applications, the main catalyst, to achieve an estimated 86% compound annual growth.
無法據以確立. Investor-day projections (F) of bit demand through 2028; growth bases and segment definitions are the company's own.
引用於: 第節 第8.4節; 展望 15

S-567 TrendForce — HBM Market Bulletin - Sep. 22, 2026
市場研究 · Taiwan · 英文 · 22 September 2026 · 閱讀 27 September 2026
https://www.trendforce.com/research/download/RP260922FJ3
“Amid tight supply and richer HBM4 mix, TrendForce sharply lifts its 2027 HBM price outlook as 8-Hi leads shipments.”
“2027 HBM forecast raised on tight supply, richer HBM4 mix.”
支持的研判. TrendForce's HBM bulletin of 22 September 2026 says it sharply lifted its 2027 HBM price outlook amid tight supply and a richer HBM4 mix, with 8-high stacks leading 2027 shipments.
無法據以確立. A bulletin summary; the revised price figures are not given in these passages.
引用於: 第節 第8.4節

S-568 Micron Technology, Inc. — Micron Technology, Inc. Fiscal Q1 2026 Earnings Call Prepared Remarks
企業新聞稿 · U.S. · 英文 · 17 December 2025 · 閱讀 27 September 2026
https://investors.micron.com/static-files/088991c5-a249-4f66-a0a6-258d9b66f3f9
“We forecast an HBM TAM CAGR of approximately 40% through calendar 2028, from approximately $35 billion in 2025 to around $100 billion in 2028.”
“This $100 billion HBM TAM milestone is now projected to arrive two years earlier than in our prior outlook.”
支持的研判. Micron forecasts the HBM market to grow about 40% a year, from about US$35 billion in 2025 to around US$100 billion in 2028, two years earlier than in its prior outlook.
無法據以確立. A company forecast (F) of market value, which depends on prices as well as bits and may be revised.
引用於: 第節 第8.4節

S-569 U.S. Department of Commerce, Bureau of Industry and Security (BIS), Federal Register — Foreign-Produced Direct Product Rule Additions, and Refinements to Controls for Advanced Computing and Semiconductor Manufacturing Items
監管機關 · U.S. · 英文 · 5 December 2024 · 閱讀 27 September 2026
https://www.federalregister.gov/documents/2024/12/05/2024-28270/foreign-produced-direct-product-rule-additions-and-refinements-to-controls-for-advanced-computing
直接文件: https://www.govinfo.gov/content/pkg/FR-2024-12-05/pdf/2024-28270.pdf
“All HBM stacks currently in production exceed this threshold.”
“this IFR will control HBM having a 'memory bandwidth density' greater than 2 GB per second per square millimeter (mm).”
支持的研判. BIS's December 2024 rule controls HBM with a memory bandwidth density above 2 GB per second per square millimetre, noting that all HBM stacks then in production exceed that threshold.
無法據以確立. An interim final rule as published; later amendments and licence exceptions change how it applies.
引用於: 第節 第8.5節; 展望 19

S-570 U.S. Government, Electronic Code of Federal Regulations (eCFR), Title 15 (BIS Export Administration Regulations) — § 740.25 License Exception High Bandwidth Memory (HBM)
監管機關 · U.S. · 英文 · 18 September 2026 · 閱讀 27 September 2026
https://www.ecfr.gov/current/title-15/subtitle-B/chapter-VII/subchapter-C/part-740/section-740.25
“The following exports, reexports, or transfers (in-country) are only authorized under this License Exception HBM if the 3A090.c item has a memory bandwidth density less than 3.3 GB/s/mm2 and both of the following conditions apply:”
“The 3A090.c items exported, reexported, or transferred to or within Macau or a destination specified in Country Group D:5 must be directly purchased by the designer of the co-packaged commodity not otherwise prohibited from receipt of the item; and”
支持的研判. License Exception HBM authorises exports of 3A090.c HBM to Macau or D:5 destinations only below 3.3 GB/s/mm2, when bought directly by the co-packaged chip's designer and shipped directly to the packaging site.
無法據以確立. Regulatory text with further conditions, including exporter headquarters and excluded advanced-node facilities.
引用於: 第節 第8.5節

S-571 The White House (Presidential Proclamation and Annex) — ADJUSTING IMPORTS OF SEMICONDUCTORS, SEMICONDUCTOR MANUFACTURING EQUIPMENT, AND THEIR DERIVATIVE PRODUCTS INTO THE UNITED STATES
政府 · U.S. · 英文 · 14 January 2026 · 閱讀 27 September 2026
https://www.whitehouse.gov/presidential-actions/2026/01/adjusting-imports-of-semiconductors-semiconductor-manufacturing-equipment-and-their-derivative-products-into-the-united-states/
直接文件: https://www.whitehouse.gov/wp-content/uploads/2026/01/2026SemiConductor.prc_.rel-ANNEX.pdf
“Except as otherwise provided in this proclamation, imports of Covered Products will be subject to a 25 percent ad valorem duty rate.”
“(1) a total processing performance greater than 14,000 and less than 17,500, and a total DRAM bandwidth greater than 4,500 GB/s and less than 5,000 GB/s”
支持的研判. The January 2026 proclamation imposes a 25% duty on covered semiconductor products, defined as logic ICs within set bands of total processing performance and total DRAM bandwidth.
無法據以確立. The proclamation's definitions and exceptions are longer than these passages; the duty's scope depends on its annex.
引用於: 第節 第8.5節

S-572 Huawei Technologies Co., Ltd. (华为) — 昇腾950超节点真机亮相2026世界人工智能大会
企業新聞稿 · China · 簡體中文 · 17 July 2026 · 閱讀 27 September 2026
https://www.huawei.com/cn/news/2026/7/atlas-950-superpod
“本次WAIC，华为首次线下展出昇腾950超节点（Atlas 950 SuperPoD），该产品基于灵衢互联协议和超节点架构，实现业界最大1024卡规模，提供1 EFLOPS FP8 、2 EFLOPS FP4澎湃算力，”
APEX譯文: At this WAIC, Huawei showed the Ascend 950 SuperPoD (Atlas 950 SuperPoD) offline for the first time; based on the Lingqu interconnect protocol and supernode architecture, it reaches the industry's largest scale of 1,024 cards and provides 1 EFLOPS of FP8 and 2 EFLOPS of FP4 computing power,
“拥有256TB全局统一内存编址空间，依托TB级NPU互联超大带宽与3μs 超低RTT时延，突破集群内计算、存储等各类资源的通信瓶颈。”
APEX譯文: It has a 256 TB globally unified memory addressing space and, relying on TB-class NPU interconnect bandwidth and ultra-low 3 μs RTT latency, breaks through the communication bottlenecks between computing, storage and other resources within the cluster.
支持的研判. Huawei says it showed a physical Atlas 950 SuperPoD at WAIC 2026, with 1,024 cards, 1 EFLOPS FP8 and 2 EFLOPS FP4, and a 256 TB globally unified memory address space.
無法據以確立. A company showcase article; unified memory addressing is address space, not HBM capacity, and no delivery volumes are given.
引用於: 第節 第8.5節

S-573 Huawei Technologies Co., Ltd. — Advancing the Agentic World, Building a Solid Silicon Foundation
企業新聞稿 · China · 英文 · 17 September 2026 · 閱讀 27 September 2026
https://www.huawei.com/en/news/2026/9/hc-wang-keynote
“Ascend 960DT will be available in Q1 2027, three quarters ahead of the company's original roadmap.”
“A single Atlas 960E SuperPoD can scale up to 4,096 NPUs, delivering 8 EFLOPS of FP8 compute performance, with up to 1 petabyte of HBM capacity.”
支持的研判. Huawei says the Ascend 960DT will be available in the first quarter of 2027, three quarters ahead of its original roadmap, and that an Atlas 960E SuperPoD scales to 4,096 NPUs with up to 1 PB of HBM.
無法據以確立. A keynote's roadmap (plans, not deliveries); the HBM's source and volumes are not stated.
引用於: 第節 第8.5節, 第8.6節; 展望 19

S-574 CXMT Corporation (长鑫科技集团股份有限公司), Shanghai Stock Exchange STAR Market (688825) — 长鑫科技集团股份有限公司 2026 年半年度报告
申報文件 · China · 簡體中文 · 29 August 2026 · 閱讀 27 September 2026
https://static.cninfo.com.cn/finalpage/2026-08-29/1225529543.PDF
“主要得益于全球算力需求的快速增长和全球主要厂商产能调配等因素影响，全球 DRAM 产品供不应求，价格呈现大幅上涨趋势，带动公司主要 DRAM 产品销售毛利大幅增长所致。”
APEX譯文: mainly owing to factors including the rapid growth of global computing-power demand and capacity reallocation by the world's major manufacturers, global DRAM products were in short supply and prices rose sharply, driving a substantial increase in the gross profit on sales of the company's main DRAM products.
“当前人工智能产业正处于基础设施集中投资建设期，但其商业化应用仍处于持续拓展过程中，公司报告期内相关领域内的收入占比相对较低。”
APEX譯文: The AI industry is currently in a period of concentrated infrastructure investment, but its commercial applications are still expanding, and the share of the company's revenue from related fields during the reporting period was relatively low.
取自來源表格之數值（轉錄）: 营业收入 | 150,310,076,758.00 | 15,437,924,611.51 | 873.64
取自來源表格之數值（轉錄）: 归属于上市公司股东的净利润 | 77,605,227,520.00 | -2,332,058,185.65 | 不适用
支持的研判. CXMT's half-year report gives H1 2026 revenue of RMB150.3 billion, up 873.64%, and net profit of RMB77.6 billion, which it attributes to a global DRAM shortage with sharply rising prices; AI-related revenue remained a relatively low share.
無法據以確立. Interim statements; the link between results and AI demand is the company's own explanation, and HBM is not mentioned in the sections read.
引用於: 第節 第8.5節; 展望 19

S-575 Alibaba Group Holding Limited (Form 6-K, Exhibit 99.1, SEC EDGAR) — Alibaba Group Announces June Quarter 2026 Results
申報文件 · China · 英文 · 20 August 2026 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/1577552/000110465926099220/tm2623667d1_ex99-1.htm
“The significant year-over-year increase was due to several reasons, including fluctuations in procurement cycles, increase in CPU-compute capacity driven by anticipated growing customer adoption of AI agents, and higher pricing of a broad range of chip components.”
“During the quarter ended June 30, 2026, capital expenditures were RMB67,678 million (US$9,975 million), an increase of 75% compared to RMB38,676 million in the same quarter of 2025, reflecting our continued investments in AI infrastructure to meet strong and growing customer demand.”
支持的研判. Alibaba's capital expenditure rose 75% to RMB67.7 billion in the June 2026 quarter; it attributes the rise partly to more CPU-compute capacity for expected AI-agent adoption and higher prices of a broad range of chip components.
無法據以確立. Group-wide capex, not split by servers, memory or facilities; the component-price effect is not quantified and memory is not named.
引用於: 第節 第8.5節

S-576 Tencent Holdings Limited (Investor Relations) — TENCENT ANNOUNCES 2026 SECOND QUARTER RESULTS
企業新聞稿 · China · 英文 · 12 August 2026 · 閱讀 27 September 2026
https://www.tencent.com/wp-content/uploads/2026/08/Tencent-Announces-2026-Second-Quarter-Results.pdf
“Capital expenditure was RMB52.8 billion, up 176% YoY.”
“At the infrastructure level, we substantially stepped up our procurement of compute, which will enable us to convert usage of our applications and models into revenue going forward.”
支持的研判. Tencent reports second-quarter 2026 capital expenditure of RMB52.8 billion, up 176% year on year, after substantially stepping up its procurement of compute to convert usage of its models and applications into revenue.
無法據以確立. Capex and procurement are company-wide and not split by chips, memory or facilities.
引用於: 第節 第8.5節

S-577 Alibaba Group (Alizila news site) — AliViews: Eddie Wu Shares Alibaba's Strategic Full-Stack AI Roadmap at the 2026 Apsara Conference
企業新聞稿 · China · 英文 · 22 September 2026 · 閱讀 27 September 2026
https://www.alizila.com/aliviews-eddie-wu-shares-alibabas-strategic-full-stack-ai-roadmap-at-the-2026-apsara-conference/
“Global shortages across the AI data center supply chain are currently limiting the speed at which we can scale our compute infrastructure.”
“However, we recognize that the industry's mid-to-long-term demand far outpaces our supply capabilities.”
支持的研判. Alibaba's chief executive says global shortages across the AI data-centre supply chain limit how fast it can scale its compute infrastructure, and that mid- to long-term demand far outpaces its supply capabilities.
無法據以確立. A conference speech; it names no component in shortage and quantifies neither demand nor supply.
引用於: 第節 第8.5節

S-578 National Business Daily (每日经济新闻, nbd.com.cn) — DeepSeek官微留言：UE8M0 FP8是针对即将发布的下一代国产芯片设计
媒體 · China · 簡體中文 · 21 August 2025 · 閱讀 27 September 2026
https://www.nbd.com.cn/articles/2025-08-21/4021882.html
“DeepSeek官微在置顶留言里说，UE8M0FP8是针对即将发布的下一代国产芯片设计。”
APEX譯文: In a pinned comment, DeepSeek's official account said that UE8M0 FP8 is designed for the next generation of domestic chips to be released soon.
支持的研判. National Business Daily reports a pinned comment by DeepSeek's official account saying UE8M0 FP8 was designed for next-generation domestic chips about to be released.
無法據以確立. A media report of a social-media comment; it names no chip or supplier.
引用於: 第節 第8.5節; 展望 19

S-579 Cailian Press (财联社) — DeepSeek-V4发布！迈入百万上下文时代 华为昇腾超节点全面支持
媒體 · China · 簡體中文 · 24 April 2026 · 閱讀 27 September 2026
https://m.cls.cn/detail/2354690
“华为昇腾超节点全系列产品支持DeepSeek V4系列模型，实现DeepSeek V4-Pro 20ms和DeepSeek V4-Flash 10ms低时延推理。”
APEX譯文: Huawei Ascend's full range of SuperPoD products supports the DeepSeek V4 series models, achieving low-latency inference of 20 ms for DeepSeek V4-Pro and 10 ms for DeepSeek V4-Flash.
“DeepSeek方面称，受限于高端算力，目前V4-Pro的服务吞吐仍有限，预计下半年昇腾950超节点批量上市后，Pro价格会大幅下调。”
APEX譯文: DeepSeek said that, constrained by high-end computing power, V4-Pro's service throughput is still limited for now, and it expects Pro prices to be cut sharply once Ascend 950 SuperPoDs are launched in volume in the second half of the year.
支持的研判. Cailian Press reports that Huawei's full range of Ascend SuperPoDs supports DeepSeek V4, and that DeepSeek expects to cut V4-Pro prices sharply once Ascend 950 SuperPoDs ship in volume in the second half of 2026.
無法據以確立. A media report of company statements at launch; the latency figures are vendor claims.
引用於: 第節 第8.5節; 展望 19

S-580 National Institute of Standards and Technology (NIST), Center for AI Standards and Innovation (CAISI), U.S. Department of Commerce — CAISI Evaluation of DeepSeek AI Models Finds Shortcomings and Risks
政府 · U.S. · 英文 · 30 September 2025 · 閱讀 27 September 2026
https://www.nist.gov/news-events/news/2025/09/caisi-evaluation-deepseek-ai-models-finds-shortcomings-and-risks
“The best U.S. model outperforms the best DeepSeek model (DeepSeek V3.1) across almost every benchmark.”
“Agents based on DeepSeek's most secure model (R1-0528) were, on average, 12 times more likely than evaluated U.S. frontier models to follow malicious instructions designed to derail them from user tasks.”
支持的研判. NIST's CAISI found that the best U.S. model outperformed the best DeepSeek model on almost every benchmark, and that DeepSeek-based agents were on average 12 times more likely to follow malicious instructions.
無法據以確立. A U.S. government evaluation of selected models and benchmarks at one date; newer models may differ.
引用於: 第節 第8.5節

S-581 U.S. Congress (National Defense Authorization Act for Fiscal Year 2026, Public Law 119-60, section 1532), statutory text as reproduced by the Emerging Technology Observatory AGORA database (Georgetown CSET) — National Defense Authorization Act for Fiscal Year 2026, Sec. 1532. Guidance and prohibition on use of certain artificial intelligence
政府 · U.S. · 英文 · 18 December 2025 · 閱讀 27 September 2026
https://agora.eto.tech/instrument/2694
直接文件: https://www.govinfo.gov/content/pkg/PLAW-119publ60/html/PLAW-119publ60.htm
“Except as provided in subsection (b), not later than 30 days after the date of the enactment of this Act, the Secretary of Defense shall require the exclusion and removal of covered artificial intelligence from the systems and devices of the Department of Defense.”
“Except as provided in subsection (b), not later than 30 days after the date of enactment of this Act, no contractor may, during the period of performance of such contractor under a contract with the Department of Defense, use covered artificial intelligence with respect to the performance of a contract with the Department.”
支持的研判. Section 1532 of the fiscal 2026 NDAA requires the Defense Secretary to remove covered AI, including DeepSeek's, from Department systems within 30 days, and bars contractors from using it in contract performance.
無法據以確立. A statute with exceptions in subsection (b); it applies to the Department of Defense and its contractors only.
引用於: 第節 第8.5節

S-582 State Council of the PRC (gov.cn) — 国务院关于深入实施“人工智能+”行动的意见 (国发〔2025〕11号)
政府 · China · 簡體中文 · 26 August 2025 · 閱讀 27 September 2026
https://www.gov.cn/zhengce/content/202508/content_7037861.htm
“支持人工智能开源社区建设，促进模型、工具、数据集等汇聚开放，培育优质开源项目。”
APEX譯文: Support the building of AI open-source communities, promote the pooling and opening of models, tools, datasets and the like, and cultivate high-quality open-source projects.
“到2027年，率先实现人工智能与6大重点领域广泛深度融合，新一代智能终端、智能体等应用普及率超70%，智能经济核心产业规模快速增长，人工智能在公共治理中的作用明显增强，人工智能开放合作体系不断完善。”
APEX譯文: By 2027, take the lead in achieving broad and deep integration of AI with six key fields; the penetration rate of applications such as new-generation intelligent terminals and agents will exceed 70%, the scale of the core industries of the intelligent economy will grow rapidly, the role of AI in public governance will be markedly strengthened, and the AI open-cooperation system will keep improving.
支持的研判. China's State Council opinion on the 'AI+' action supports AI open-source communities and sets targets for new-generation intelligent terminals, agents and similar applications to exceed 70% penetration by 2027 and 90% by 2030.
無法據以確立. A policy opinion with targets; it does not define how penetration is measured or bind specific firms.
引用於: 第節 第8.5節

S-583 International Energy Agency (IEA) — Energy demand from AI – Energy and AI
研究 · Global · 英文 · 10 April 2025 · 閱讀 27 September 2026
https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai
“It sees global electricity demand from data centres in 2035 that is around 45% higher than in the Base Case, exceeding the 1 700 TWh mark and reaching around 4.4% of global electricity demand.”
“This leads to a plateau in energy demand at around 700 TWh, limiting the growth of the data centre share of global electricity demand to less than 2% in 2035.”
支持的研判. The IEA's cases put 2035 data-centre electricity demand above 1,700 TWh in its Lift-Off case, at a plateau of about 700 TWh in its Headwinds case and around 970 TWh with high efficiency.
無法據以確立. Scenario cases (S) under IEA assumptions, not forecasts; the Base Case figure is not in these passages.
引用於: 第節 第8.6節

S-584 NVIDIA Corporation (NVIDIA Technical Blog) — How to Reduce KV Cache Bottlenecks with NVIDIA Dynamo
企業新聞稿 · U.S. · 英文 · 18 September 2025 · 閱讀 27 September 2026
https://developer.nvidia.com/blog/how-to-reduce-kv-cache-bottlenecks-with-nvidia-dynamo/
“It directly offloads KV Cache from GPU memory to more scalable and affordable storage systems like CPU RAM, local SSDs, or remote network storage.”
“The KV Cache grows linearly with prompt length and must reside in GPU memory during the generation process for fast access.”
支持的研判. NVIDIA says the KV cache grows linearly with prompt length and must reside in GPU memory during generation, and that Dynamo offloads it to CPU memory, local SSDs or network storage.
無法據以確立. A vendor's technical blog on its own software; it shows capability, not how widely offloading is deployed.
引用於: 展望 15

S-585 CXL Consortium (Compute Express Link blog) — Increasing Tokens-per-Dollar with CXL: FMS 2026 Recap
標準組織 · Global · 英文 · 21 August 2026 · 閱讀 27 September 2026
https://computeexpresslink.org/blog/increasing-tokens-per-dollar-with-cxl-fms-2026-recap-4788/
"The panel explored how CXL memory pooling and sharing can provide additional memory capacity closer to compute, along with memory tiering and KV Cache offloading strategies that reduce reliance on slower storage."
"As AI models and context windows continue to grow, so does the amount of KV Cache data that needs to be stored and accessed during inference."
支持的研判. The CXL Consortium's FMS 2026 recap says a panel explored CXL memory pooling and sharing, memory tiering and KV-cache offloading to add capacity near compute and reduce reliance on slower storage.
無法據以確立. A consortium blog summarising a panel; it cites no specification for moving KV caches and no deployments.
引用於: 展望 15

S-586 KLA Corporation — Letter to Shareholders Q3 Fiscal 2026
企業新聞稿 · U.S. · 英文 · 29 April 2026 · 閱讀 27 September 2026
https://d1io3yog0oux5.cloudfront.net/_42d21a5ab5ca507394aee7b9106cc480/klatencor/db/1117/10655/letter_to_shareholders/KLA+Shareholder+Letter+-+Q3+FY26.pdf
"Our view of elevated memory pricing persisting through at least calendar 2026 is unchanged and we continue to see a roughly 100bps negative impact on gross margins over the next several quarters."
"As discussed last quarter, the guidance also includes the persistent impact of elevated DRAM chip costs for the company's image-processing computers that ship with our systems, creating a headwind to the company's gross margins."
支持的研判. KLA's April 2026 letter says its guidance still includes elevated DRAM chip costs for its image-processing computers and that it sees a roughly 100 bps gross-margin impact over the next several quarters.
無法據以確立. An input cost for KLA's own computers; no dollar amount, supplier or quarterly split.
引用於: 第節 第5.3節; 表 8

S-587 KLA Corporation — Letter to Shareholders Q4 Fiscal 2026
企業新聞稿 · U.S. · 英文 · 28 July 2026 · 閱讀 27 September 2026
https://d1io3yog0oux5.cloudfront.net/_42d21a5ab5ca507394aee7b9106cc480/klatencor/db/1117/10668/letter_to_shareholders/KLA+Earnings+Shareholder+Letter+-+Q4+FY26.pdf
"In descending order, the quarterly regional revenue split was Taiwan at 31%, China at 26%, North America at 18%, Korea at 10%, Japan at 6%, Europe at 5%, and Rest of Asia at 4% of total revenue."
"This was also at the upper end of our guidance range, driven by a more favorable services mix than modeled and manufacturing scale that served as positive offsets to the challenging memory pricing environment and tariff headwinds."
支持的研判. KLA's July 2026 letter puts China at 26% of revenue for the June 2026 quarter and names the memory pricing environment and tariffs as headwinds offset by services mix and scale.
無法據以確立. The letter, as read, gives no export-control revenue estimate; an excerpt cannot show an absence.
引用於: 表 8, 14, 20; 第節 第7.1節; 展望 6

S-588 Advanced Micro Devices, Inc. — AMD Reports First Quarter 2026 Financial Results
企業新聞稿 · U.S. · 英文 · 5 May 2026 · 閱讀 27 September 2026
https://ir.amd.com/news-events/press-releases/detail/1284/amd-reports-first-quarter-2026-financial-results
"For the second quarter of 2026, AMD expects revenue to be approximately $11.2 billion, plus or minus $300 million."
支持的研判. AMD guides second-quarter 2026 revenue of about US$11.2 billion, plus or minus US$300 million, in an outlook sentence with no China or MI308 qualifier.
無法據以確立. The release, read twice, mentions neither China nor MI308; an excerpt cannot show an absence.
引用於: 第節 第7.1節; 表 14

S-589 Synopsys, Inc. (Form 8-K filed with the U.S. SEC) — Form 8-K (Current Report), date of report July 2, 2025
申報文件 · U.S. · 英文 · 2 July 2025 · 閱讀 27 September 2026
https://sec.gov/Archives/edgar/data/883241/000119312525155294/d80081d8k.htm
"On July 2, 2025, Synopsys received a letter from the Bureau of Industry and Security of the U.S. Department of Commerce informing Synopsys that the export restrictions related to China, pursuant to a letter received on May 29, 2025, have now been rescinded, effective immediately."
"Synopsys is working to restore access to the recently restricted products in China."
支持的研判. Synopsys reports that on 2 July 2025 BIS informed it that the China export restrictions of its 29 May 2025 letter were rescinded with immediate effect, and that it is restoring access to the restricted products.
無法據以確立. A current report; it gives no guidance, amount or product list.
引用於: 第節 第7.1節; 表 14

S-590 Cadence Design Systems, Inc. (Form 8-K filed with the U.S. SEC) — Form 8-K (Current Report), date of report July 2, 2025
申報文件 · U.S. · 英文 · 2 July 2025 · 閱讀 27 September 2026
https://www.sec.gov/Archives/edgar/data/813672/000081367225000093/cdns-20250702.htm
"On July 2, 2025, BIS informed Cadence that the license requirements set forth in the May 23, 2025 letter from BIS were rescinded effective immediately."
支持的研判. Cadence reports that on 2 July 2025 BIS informed it that the licence requirements of its 23 May 2025 letter were rescinded with immediate effect.
無法據以確立. A current report; it gives no guidance and no amount.
引用於: 第節 第7.1節; 表 14

S-591 TrendForce — Liquid Cooling to Scale in AI Data Centers, Penetration to Surpass 30% in 2025, Says TrendForce
市場研究 · Taiwan · 英文 · 21 August 2025 · 閱讀 28 September 2026
https://www.trendforce.com/presscenter/news/20250821-12682.html
"Penetration in AI data centers is projected to surge from 14% in 2024 to 33% in 2025, with continued growth in the following years."
"TrendForce's latest research on the liquid cooling industry reveals that the rollout of NVIDIA's GB200 NVL72 rack servers in 2025 will accelerate AI data center upgrades, driving liquid cooling adoption from early pilot projects to large-scale deployment."
支持的研判. TrendForce projected in August 2025 that liquid-cooling penetration in AI data centres would rise from 14% in 2024 to 33% in 2025.
無法據以確立. The release gives the basis as AI data centres, not AI chips, and does not say whether it counts chips, servers, racks or capacity; it gives no values for 2026 or 2027.
引用於: 第一頁; 圖 18; 第節 第E.2節

S-592 IDC — Servers Market Insights
市場研究 · Global · 英文 · 28 July 2026 · 閱讀 28 September 2026
https://www.idc.com/promo/servers/
"The worldwide server market delivered 30.7% spending growth in the first quarter of 2026 while unit growth was modest at 3.3% year over year."
"IDC's baseline assumes continuation of hyperscalers capex at announced levels, ongoing rack-scale GPU deployment with redistribution from ODM custom-rack to OEM rack-scale and standard-rack configurations, and a constrained memory/NAND supply environment with elevated pricing through at least the first half of 2027."
取自來源表格之數值（轉錄）: Servers Market Insights annual table, untitled (Value (US$M); Annual Growth %) | 2025, Non-x86, Value (US$M): $154,971 | 2025, Non-x86, Annual Growth %: 209.3% | 2025, x86, Value (US$M): $298,560 | 2025, x86, Annual Growth %: 46.1% | 2025, Total, Value (US$M): $453,531 | 2025, Total, Annual Growth %: 78.2% | 2026, Non-x86, Value (US$M): $311,072 | 2026, Non-x86, Annual Growth %: 100.7% | 2026, x86, Value (US$M): $335,925 | 2026, x86, Annual Growth %: 12.5% | 2026, Total, Value (US$M): $646,998 | 2026, Total, Annual Growth %: 42.7% | 2027, Non-x86, Value (US$M): $444,640 | 2027, Non-x86, Annual Growth %: 42.9% | 2027, x86, Value (US$M): $485,922 | 2027, x86, Annual Growth %: 44.7% | 2027, Total, Value (US$M): $930,562 | 2027, Total, Annual Growth %: 43.8%
支持的研判. IDC's server page (updated 28 July 2026) reports worldwide server spending up 30.7% and units up 3.3% in the first quarter of 2026, and tabulates server spending by processor type for 2025 to 2027 with its baseline assumptions.
無法據以確立. The page does not define the market or mark 2026 and 2027 as forecasts; its split is by processor architecture (x86 and non-x86), not AI against other servers.
引用於: 第一頁; 第節 第2.3節; 圖 6; 表 36, 37, 41

S-593 International Monetary Fund (IMF) — World Economic Outlook Update, July 2026: Global Economy in Crosscurrents of War and Technology
研究 · Global · 英文 · 8 July 2026 · 閱讀 28 September 2026
https://www.imf.org/en/publications/weo/issues/2026/07/08/world-economic-outlook-update-july-2026
直接文件: https://www.imf.org/-/media/files/publications/weo/2026/update/july/english/text.pdf
"In the United States, growth is projected at 2.3 percent in 2026 and 2.2 percent in 2027, virtually unchanged from April. Activity is supported by fiscal policy, accommodative financial conditions, and continued technology-related business investment and productivity strength."
"Global growth is projected at 3.0 percent for 2026 and 3.4 percent for 2027, broadly unchanged cumulatively from the April 2026 World Economic Outlook."
支持的研判. The IMF projects world growth of 3.0% in 2026 and 3.4% in 2027 and U.S. growth of 2.3% and 2.2%, citing technology-related business investment among the supports for U.S. activity.
無法據以確立. The update does not quantify AI-related investment or its contribution to growth; world output is a PPP-weighted aggregate, and the figures are July 2026 projections.
引用於: 第一頁; 第節 第E.1節, 第E.6節

S-594 Dell'Oro Group — Data Center Physical Infrastructure Market Forecast to Reach $120 Billion by 2030, According to Dell'Oro Group
市場研究 · U.S. · 英文 · 19 August 2026 · 閱讀 28 September 2026
https://www.delloro.com/news/data-center-physical-infrastructure-market-forecast-to-reach-120-billion-by-2030/
"Thermal Management remains the fastest-growing DCPI segment, with liquid cooling the fastest-growing technology as rack densification moves the technology from an option to a precondition."
"What has changed in this forecast is where the risk sits. Demand is no longer the open question—delivery is. Equipment lead times, construction labor, grid interconnection, and community consent all remain constrained, especially with the first statewide data center moratorium now in force."
"According to a recently published report from Dell'Oro Group, the trusted source for market information about the telecommunications, security, networks, and data center industries, worldwide Data Center Physical Infrastructure (DCPI) manufacturer revenue is projected to grow at a 22 percent compound annual growth rate (CAGR) from 2025 to 2030, reaching $120 billion by the end of the period."
支持的研判. Dell'Oro Group forecasts data-centre physical infrastructure revenue growing 22% a year from 2025 to 2030, to US$120 billion, with nearly 200 GW of capacity added and thermal management and liquid cooling growing fastest.
無法據以確立. A market-research forecast of manufacturers' revenue for power, thermal and rack infrastructure; it prints no 2025 value, and the 200 GW appears in a subheading without a start year or geography.
引用於: 第一頁; 第節 第4.3節, 第4.4節

S-595 Board of Governors of the Federal Reserve System, FEDS Notes (author: Geng Li) — Do Major Technology Advancements Lead to Overinvestment?
研究 · U.S. · 英文 · 6 July 2026 · 閱讀 28 September 2026
https://www.federalreserve.gov/econres/notes/feds-notes/do-major-technology-advancements-lead-to-overinvestment-20260706.html
直接文件: https://www.federalreserve.gov/econres/notes/feds-notes/do-major-technology-advancements-lead-to-overinvestment-accessible-20260706.htm

“After fluctuating within a relatively narrow range over the previous ten years, this share turned sharply higher in 2025 and, as of the first quarter of 2026, was about 0.8 percentage points higher than the level of 2024, only a touch lower than the year 2000 peak.”

“A large part of the recent investment boom has been AI-driven, and the pace of IP and equipment investment acceleration relative to GDP growth has been comparable to that seen during the 1990s, but this time starting at a much higher base level.”

“Solow (1959) introduced the seminal idea of embodied technology—without new capital being invested, new technology cannot increase productivity by itself.”

“From the perspective of the model, the risk of overinvestment has been taken into account in the formulation of optimal investment. In this sense, overinvestment is not a mistake caused by irrational exuberance over the new technology.”

取自來源表格之數值（轉錄）： Figure 2. IP and Equipment Investment Share in GDP (Share, decimal fraction of GDP; annual data through 2025 and quarterly for 2026:Q1) | 1995: 0.098812781 | 1996: 0.101398471 | 1997: 0.104796213 | 1998: 0.107924703 | 1999: 0.112000581 | 2000: 0.114857087 | 2001: 0.106455362 | 2002: 0.097519466 | 2023: 0.105323337 | 2024: 0.105406512 | 2025: 0.109140793 | 2026Q1: 0.113328514

支持的研判. A Federal Reserve Board staff note charts U.S. IP and equipment investment as a share of GDP (11.49% in 2000, 11.33% in 2026:Q1) and models overinvestment after general purpose technologies.

無法據以確立. The note does not say whether the series is private or total, nominal or real, and does not measure AI investment; its overinvestment findings come from a theoretical model, and the views are the author's.

引用於: 第一頁; 表 36, 37, 38; 第節 第E.3節, 第E.5節, 第E.7節; 圖 34

S-596 IDC, Resource Center blog (Juan Seminara, Research Director, Worldwide Infrastructure Trackers) — AI Infrastructure Spending Caps Historic Year at ~$90 Billion in Q4 2025; 2029 Spending to Eclipse $1 Trillion
市場研究 · Global · 英文 · 16 April 2026 · 閱讀 28 September 2026
https://www.idc.com/resource-center/blog/ai-infrastructure-spending-caps-historic-year-at-90-billion-in-q4-2025-2029-spending-to-eclipse-1-trillion/

“Full-year 2025 AI infrastructure spending totaled $318 billion, more than double the $153 billion recorded in 2024.”

“IDC projects AI infrastructure spending will reach $487 billion in 2026, representing approximately 53% year-over-year growth.”

支持的研判. IDC puts AI-infrastructure spending at US$153 billion in 2024 and US$318 billion in 2025, projects US$487 billion in 2026 and forecasts more than US$1 trillion by 2029.

無法據以確立. The post does not define AI infrastructure beyond referring to IDC's tracker taxonomy; the growth drivers it lists are upside factors, not its base case.

引用於: 第一頁; 第節 第2.4節; 圖 10

S-597 TrendForce — Combined CapEx of Top Eight CSPs to Exceed $710 Billion in 2026; Google Leads ASIC Deployment with TPUs, Says TrendForce
市場研究 · Taiwan · 英文 · 25 February 2026 · 閱讀 28 September 2026
https://www.trendforce.com/presscenter/news/20260225-12934.html

“Combined capital expenditures by the world's eight leading CSPs—Google, AWS, Meta, Microsoft, Oracle, Tencent, Alibaba, and Baidu— are projected to exceed $710 billion in 2026, representing approximately 61% YoY growth.”

“TrendForce estimates that Alphabet, Google's parent company, will see its 2026 capital expenditure surpass $178.3 billion, up 95% YoY.”

支持的研判. TrendForce projected in February 2026 that the combined 2026 capex of eight leading cloud providers would exceed $710 billion, about 61% more than in 2025.

無法據以確立. The set of eight excludes ByteDance, so the figure is not comparable with TrendForce's later nine-company totals; the company figures are TrendForce estimates, not guidance.

引用於: 第節 第2.1節, 第E.2節; 圖 33

S-598 TrendForce — North American AI Data Center Expansion Drives 2026 CapEx of Top Nine CSPs to US$830 Billion, Says TrendForce
市場研究 · Taiwan · 英文 · 6 May 2026 · 閱讀 28 September 2026
https://www.trendforce.com/presscenter/news/20260506-13033.html

“As a result, TrendForce has revised its forecast for the combined CapEx of the world's top nine CSPs—Google, AWS, Meta, Microsoft, Oracle, ByteDance, Tencent, Alibaba, and Baidu—up to approximately US$830 billion in 2026, with the annual growth rate raised from 61% to 79%.”

“Microsoft has increased its CapEx outlook to $190 billion, implying approximately 130% YoY growth.”

支持的研判. TrendForce raised its forecast of the nine largest cloud providers' 2026 capex to about US$830 billion in May 2026, the growth rate from 61% to 79%, and relays higher guidance from Microsoft, Google and Meta.

無法據以確立. The 61% it revised was published for eight companies without ByteDance; the release gives no individual figures for most of the nine and no 2025 base.

引用於: 第節 第2.1節, 第E.2節; 圖 33

S-599 Gartner — Gartner Says Worldwide AI Spending Will Total $2.5 Trillion in 2026
市場研究 · Global · 英文 · 15 January 2026 · 閱讀 28 September 2026
https://www.gartner.com/en/newsroom/press-releases/2026-1-15-gartner-says-worldwide-ai-spending-will-total-2-point-5-trillion-dollars-in-2026

“Worldwide spending on AI is forecast to total $2.52 trillion in 2026, a 44% increase year-over-year, according to Gartner, Inc. a business and technology insights company.”

取自來源表格之數值（轉錄）： Table 1: Worldwide AI Spending by Market, 2025-2027 (Millions of U.S. Dollars) | AI Services, 2025: 439,438 | AI Services, 2026: 588,645 | AI Services, 2027: 761,042 | AI Cybersecurity, 2025: 25,920 | AI Cybersecurity, 2026: 51,347 | AI Cybersecurity, 2027: 85,997 | AI Software, 2025: 283,136 | AI Software, 2026: 452,458 | AI Software, 2027: 636,146 | AI Models, 2025: 14,416 | AI Models, 2026: 26,380 | AI Models, 2027: 43,449 | AI Platforms for Data Science and Machine Learning, 2025: 21,868 | AI Platforms for Data Science and Machine Learning, 2026: 31,120 | AI Platforms for Data Science and Machine Learning, 2027: 44,482 | AI Application Development Platforms, 2025: 6,587 | AI Application Development Platforms, 2026: 8,416 | AI Application Development Platforms, 2027: 10,922 | AI Data, 2025: 827 | AI Data, 2026: 3,119 | AI Data, 2027: 6,440 | AI Infrastructure, 2025: 964,960 | AI Infrastructure, 2026: 1,366,360 | AI Infrastructure, 2027: 1,748,212 | Total AI Spending, 2025: 1,757,152 | Total AI Spending, 2026: 2,527,845 | Total AI Spending, 2027: 3,336,690 | Source: Gartner (January 2026)

支持的研判. Gartner's January 2026 forecast puts 2026 AI spending at US$2.52 trillion and AI infrastructure at US$1,366 billion (Table 1, 2025–2027).

無法據以確立. A market-research forecast; the release does not define AI infrastructure, and Gartner changed its market structure in September 2026, so this total is not on the later basis.

引用於: 第節 第2.1節, 第E.2節; 圖 33

S-600 Gartner — Gartner Forecasts Worldwide AI Spending to Grow 47% in 2026
市場研究 · Global · 英文 · 19 May 2026 · 閱讀 28 September 2026
https://www.gartner.com/en/newsroom/press-releases/2026-05-19-gartner-forecasts-worldwide-ai-spending-to-grow-47-percent-in-2026

“"Through the next several years, the need for capacity will make AI infrastructure, including AI-optimized IaaS, AI-optimized servers, AI network fabric, AI processing semiconductors and devices, the largest segment of the market, accounting for over 45% of spending, which will be driven by vendors," said John-David Lovelock, Distinguished VP Analyst at Gartner.”

取自來源表格之數值（轉錄）： Table 1: Worldwide AI Spending by Market, 2025-2027 (Millions of U.S. Dollars) | AI Services, 2025: 436,351 | AI Services, 2026: 585,527 | AI Services, 2027: 759,418 | AI Cybersecurity, 2025: 25,920 | AI Cybersecurity, 2026: 51,347 | AI Cybersecurity, 2027: 85,997 | AI Software, 2025: 282,897 | AI Software, 2026: 453,209 | AI Software, 2027: 638,431 | AI Models, 2025: 15,494 | AI Models, 2026: 32,604 | AI Models, 2027: 59,161 | AI Platforms for Data Science and Machine Learning, 2025: 21,292 | AI Platforms for Data Science and Machine Learning, 2026: 29,928 | AI Platforms for Data Science and Machine Learning, 2027: 42,639 | AI Application Development Platforms, 2025: 6,587 | AI Application Development Platforms, 2026: 8,416 | AI Application Development Platforms, 2027: 10,922 | AI Data, 2025: 826 | AI Data, 2026: 3,126 | AI Data, 2027: 6,480 | AI Infrastructure, 2025: 975,581 | AI Infrastructure, 2026: 1,431,509 | AI Infrastructure, 2027: 1,890,310 | Total AI Spending, 2025: 1,764,947 | Total AI Spending, 2026: 2,595,667 | Total AI Spending, 2027: 3,493,358 | Source: Gartner (May 2026)

支持的研判. Gartner's May 2026 forecast puts 2026 AI spending at US$2.59 trillion and AI infrastructure, defined as AI-optimized IaaS, servers, network fabric and processing semiconductors and devices, at US$1,432 billion.

無法據以確立. A market-research forecast with its own taxonomy, not comparable in level with other firms' AI-infrastructure figures.

引用於: 第節 第2.1節, 第E.2節; 圖 33

S-601 Gartner — Gartner Forecasts Worldwide AI Spending to Grow 49.5% in 2026
市場研究 · Global · 英文 · 16 September 2026 · 閱讀 28 September 2026
https://www.gartner.com/en/newsroom/press-releases/2026-09-16-gartner-forecasts-worldwide-ai-spending-to-grow-49-point-5-percent-in-2026

“"Demand for AI infrastructure (including AI-optimized IaaS, AI-optimized servers, AI network fabric, AI processing semiconductors and devices) to support anticipated future workloads remains strong and inelastic to pressures from memory-related pricing increases," said John-David Lovelock, Distinguished VP Analyst at Gartner.”

“In the current forecast, Gartner has separated cross-functional agents and assistants from AI software and added consumer agents and assistants into the AI spending forecast to better show the emerging opportunity in this area.”

取自來源表格之數值（轉錄）： Table 1: Worldwide AI Spending by Market, 2025-2027 (Millions of U.S. Dollars) | AI Services, 2025: 434,046 | AI Services, 2026: 576,481 | AI Services, 2027: 745,655 | AI Cybersecurity, 2025: 25,920 | AI Cybersecurity, 2026: 51,347 | AI Cybersecurity, 2027: 85,997 | AI Software, 2025: 288,168 | AI Software, 2026: 461,637 | AI Software, 2027: 656,353 | AI Agents and Assistants, 2025: 16,481 | AI Agents and Assistants, 2026: 29,219 | AI Agents and Assistants, 2027: 65,472 | Generative AI Models, 2025: 13,021 | Generative AI Models, 2026: 28,266 | Generative AI Models, 2027: 51,620 | AI Platforms for Data Science and Machine Learning, 2025: 19,405 | AI Platforms for Data Science and Machine Learning, 2026: 26,445 | AI Platforms for Data Science and Machine Learning, 2027: 35,552 | AI Application Development Platforms, 2025: 6,885 | AI Application Development Platforms, 2026: 9,541 | AI Application Development Platforms, 2027: 12,478 | AI Data, 2025: 826 | AI Data, 2026: 3,126 | AI Data, 2027: 6,480 | AI Infrastructure, 2025: 981,920 | AI Infrastructure, 2026: 1,484,397 | AI Infrastructure, 2027: 1,977,685 | Total AI Spending, 2025: 1,786,671 | Total AI Spending, 2026: 2,670,460 | Total AI Spending, 2027: 3,637,292 | Source: Gartner (September 2026)

支持的研判. Gartner's September 2026 forecast puts AI spending at US$2.7 trillion in 2026 (+49.5%) and AI infrastructure at US$982 billion (2025), US$1,484 billion (2026) and US$1,978 billion (2027).

無法據以確立. The September forecast separated agents and assistants from AI software, so its total is not on the basis of the January and May releases; its taxonomy is not comparable in level with IDC's.
引用於: 第節 第2.1節, 第2.4節, 第E.2節; 圖 10, 33

S-602 SEMI — Global Semiconductor Equipment Sales Forecast to Reach a Record $229 Billion in 2028, SEMI Reports
市場研究 · Global · 英文 · 14 July 2026 · 閱讀 28 September 2026
https://www.semi.org/en/semi-press-release/global-semiconductor-equipment-sales-forecast-to-reach-a-record-229-billion-dollars-in-2028-semi-reports
"DRAM equipment sales are projected to rise 39.0% to $38.8 billion in 2026, followed by 27.4% growth in 2027 and 15.0% growth in 2028, reaching $56.9 billion. NAND equipment sales are expected to grow 30.7% to $13.9 billion in 2026, then increase 31.1% in 2027 and 14.5% in 2028, reaching $20.8 billion, driven by 3D NAND layer migration and investment in higher-density architectures."
"Global sales of total semiconductor manufacturing equipment by original equipment manufacturers (OEMs) are forecast to reach a record high of $165.9 billion in 2026, increasing 23.2% year-on-year, SEMI announced today in its Mid-Year Total Semiconductor Equipment Forecast – OEM Perspective. Growth momentum is expected to continue through 2028, with total equipment sales forecast to reach a record $229.5 billion, marking five consecutive years of growth as AI-driven demand reshapes semiconductor manufacturing investment."
支持的研判. SEMI forecasts semiconductor equipment sales of US$165.9 billion in 2026 and US$229.5 billion in 2028, with DRAM equipment up 39.0% to US$38.8 billion in 2026 and at US$56.9 billion in 2028.
無法據以確立. Equipment makers' sales, not chip revenue or capacity; for most segments the release prints growth rates but no 2025 or 2027 dollar values.
引用於: 第節 第3.2節; 圖 16

S-603 SEMI — SEMI Projects 300mm Memory Equipment Investment to Surpass $50 Billion in 2026
市場研究 · Global · 英文 · 29 June 2026 · 閱讀 28 September 2026
https://www.semi.org/en/semi-press-release/semi-projects-300mm-memory-equipment-investment-to-surpass-50-billion-dollars-in-2026
"Worldwide 300mm fab equipment investment in the memory sector is projected to surpass $50 billion for the first time in 2026, rising 29% to $52 billion before increasing another 11% to $57 billion in 2027, SEMI reported today in its latest 300mm Fab Outlook."
"Worldwide 300mm memory capacity is also projected to increase, reaching 4.1 million wafers per month in 2026 and 4.2 million wafers per month in 2027."
支持的研判. SEMI projects 300 mm memory fab-equipment investment of US$52 billion in 2026 and US$57 billion in 2027, and 300 mm memory capacity of 4.1 and 4.2 million wafers a month.
無法據以確立. The second-quarter 2026 edition; the third-quarter update prints higher growth rates. The 2029 value of about US$80 billion appears only in the subtitle, and spending on fab equipment is a different measure from equipment makers' sales.
引用於: 第節 第3.2節; 圖 16

S-604 SEMI — 300mm Fab Outlook (2024 to 2029 and extended to 2031) - 3Q 2026 Update
市場研究 · Global · 英文 · 2026 · 閱讀 28 September 2026
https://www.semi.org/en/products-services/market-data/300mm-fab-outlook
"Global 300mm front-end fab equipment spending is projected to reach a record US$149 billion in 2026, representing 31% year-over-year growth."
"Memory equipment spending is projected to grow by 32% in 2026, 15% in 2027, 12% in 2028, and 21% in 2029, driven by strong AI-related memory demand."
支持的研判. SEMI's 300 mm Fab Outlook (third-quarter 2026 update) projects fab-equipment spending of US$149 billion in 2026, memory equipment up 32% in 2026 and 15% in 2027, and a 6% decline in 2030.
無法據以確立. The product page prints no edition date and no dollar values beyond 2026; the full report is sold to subscribers.
引用於: 第節 第3.2節; 圖 16

S-605 TechInsights — McClean Report April 2026: Memory (DRAM, NAND, NOR) Market Update
市場研究 · Global · 英文 · 24 April 2026 · 閱讀 28 September 2026
https://www.techinsights.com/blog/mcclean-report-april-2026-memory-dram-nand-nor-market-update
"Driven by an AI "supercycle" and relentless demand for HBM, memory IC sales are projected to grow a record-breaking 248% in 2026, reaching $706 billion."
"Consequently, DRAM and NAND flash will dominate 99% of market revenue. This reveals a major structural shift: while AI components have low unit volume, their soaring prices account for the majority of total revenue."
支持的研判. TechInsights' McClean Report projects memory IC sales up 248% in 2026 to US$706 billion, led by HBM and enterprise QLC SSDs, with DRAM and NAND at 99% of market revenue.
無法據以確立. A summary of subscription analysis; it prints no 2025 base, and its memory definition is not the same as other firms' DRAM or semiconductor revenue series.
引用於: 第節 第3.2節

S-606 Uptime Institute (Uptime Intelligence) — 2026 Cooling Systems Survey [Results and Crosstab files]
市場研究 · Global · 英文 · 12 August 2026 · 閱讀 28 September 2026
https://intelligence.uptimeinstitute.com/resource/2026-cooling-systems-survey-results-and-crosstab-files
"Results from Uptime Institute's 2026 Cooling Systems Survey (n=1,034) focus on the usage of data center cooling systems across the industry, zeroing in on the continued adoption of direct liquid cooling."
"The attached data files below provide full results of the survey, including data cuts based on organization type, region, and data center size, as well as comparison to prior-year results where applicable."
支持的研判. The Uptime Institute's 2026 Cooling Systems Survey drew 1,034 respondents and focuses on the adoption of direct liquid cooling.
無法據以確立. The public page gives the sample and scope but no result shares; the results are in files for subscribers.
引用於: 第節 第4.3節

S-607 Dell'Oro Group — AI Back-End Networks Switch Sales Surpass Front-End Networks for the First Time in 2Q 2026, According to Dell'Oro Group
市場研究 · U.S. · 英文 · 3 September 2026 · 閱讀 28 September 2026
https://www.delloro.com/news/ai-back-end-networks-switch-sales-surpass-front-end-networks-for-the-first-time-in-2q2026/
"According to a recently published report from Dell'Oro Group, the trusted source for market information about the telecommunications, security, networks, and data center industries, 2Q 2026 marked a major inflection point as switch sales in AI back-end networks surpassed those in front-end networks for the first time."
"Supply availability will also play an increasingly important role in shaping the competitive landscape and overall market growth. We expect the market to remain supply-constrained, rather than demand-constrained, for at least the next one-to-two years."
支持的研判. Dell'Oro Group reports that in 2Q 2026 switch sales in AI back-end networks surpassed those in front-end networks for the first time, and expects the market to stay supply-constrained.
無法據以確立. The release prints no dollar values or growth rates, only vendor ranks.
引用於: 第節 第4.6節

S-608 Dell'Oro Group — AI Back-end Switch Sales to Approach $1 Trillion over the Next Five Years, According to Dell'Oro Group
市場研究 · U.S. · 英文 · 28 July 2026 · 閱讀 28 September 2026
https://www.delloro.com/news/ai-back-end-switch-sales-to-approach-1-trillion-over-the-next-five-years/
"According to a recently published report from Dell'Oro Group, the trusted source for market information about the telecommunications, security, networks, and data center industries, cumulative spending on data center switches deployed in AI back-end networks to support next generation AI infrastructure is forecast to approach $1 trillion over the next five years (2026-2030). Scale-up is projected to account for more than half of the market by 2030."
"We have substantially raised our revenue and port-shipment forecasts since our January 2026 outlook to reflect both stronger demand for accelerated compute and higher infrastructure costs. Despite this significant upward revision, we believe the market will remain supply-constrained rather than demand-constrained over our five-year forecast horizon."
支持的研判. Dell'Oro Group forecasts cumulative spending on AI back-end network switches approaching US$1 trillion over 2026–2030, with scale-up above half the market by 2030, after substantially raising its forecasts.
無法據以確立. A market-research forecast; it prints no annual values.
引用於: 第節 第4.6節

S-609 Dell'Oro Group — AI Back-End Switch Market Will Push Past $100 Billion by 2030, According to Dell'Oro Group
市場研究 · U.S. · 英文 · 4 February 2026 · 閱讀 28 September 2026
https://www.delloro.com/news/ai-back-end-switch-market-will-push-past-100-billion-by-2030/
""Scale-up compute fabrics have been historically dominated by proprietary fabrics such as NVLink but just as Ethernet prevailed over InfiniBand in large-scale scale-out environments, we are now seeing alternative technologies such as UALink and Ethernet gain momentum in scale-up compute fabrics. While we predict a strong adoption of UALink, we expect Ethernet to emerge as the long-term winner across both scale-up and scale-out architectures," continued Boujelbene."
"According to the Data Center Switch – AI Back-end Networks Report by Dell'Oro Group, the trusted source for market information about the telecommunications, security, networks, and data center industries, spending on data center switches deployed in AI back-end networks is forecast to surpass $100 billion by 2030, driven by expanding deployments across scale-up, scale-out, and scale-across domains."
"The majority of switch ports deployed in AI back-end networks have shifted to 800 Gbps, and are expected to be 1600 Gbps by 2027 and 3200 Gbps by 2030."
支持的研判. Dell'Oro Group forecast in February 2026 that spending on AI back-end network switches would surpass US$100 billion by 2030, with ports moving to 1,600 Gb/s by 2027 and 3,200 Gb/s by 2030 and Ethernet the long-term winner.
無法據以確立. Raised by Dell'Oro's July 2026 forecast; the release does not say whether the US$100 billion is annual.
引用於: 第節 第4.6節

S-610 LightCounting — AI creates a new wave in demand for optical transceivers and accelerates CPO adoption
市場研究 · Global · 英文 · January 2026 · 閱讀 28 September 2026
https://www.lightcounting.com/newsletter/en/january-2026-optics-for-ai-clusters-366
"We estimate that the market reached $16.5 billion in 2025 and it will reach $26 billion in 2026."
"In our current assumptions, sales of optics to the TOP 5 Cloud companies will account for 3.1% of their capex in 2026 (up from 2.7% in 2025) and increase to 4.1% by 2031."
支持的研判. LightCounting estimates AI-cluster Ethernet optical transceiver and CPO sales at US$16.5 billion in 2025 and US$26 billion in 2026, 60% growth in each year, and optics at 2.7% and 3.1% of the top five clouds' capex.
無法據以確立. The newsletter's own market definition; the capex shares are stated as current assumptions, and the forecast beyond 2026 is in a report sold to subscribers.
引用於: 第節 第4.7節; 圖 21

S-611 LightCounting — $100 Billion Market for AI Cluster Optics by 2030?
市場研究 · Global · 英文 · March 2026 · 閱讀 28 September 2026
https://www.lightcounting.com/newsletter/en/march-2026-ethernet-optics-382
"There is a reasonable chance for annual sales of optical interconnects used in AI clusters to reach $100 billion by 2030."
"Sales of Ethernet optical transceivers doubled in 2024 and increased by another 70% in 2025, exceeding our expectations."
支持的研判. LightCounting says Ethernet optical transceiver sales doubled in 2024 and rose 70% in 2025, expects about 60% in 2026, and sees a reasonable chance of US$100 billion of AI-cluster optics sales by 2030.
無法據以確立. The US$100 billion is a possibility, not the base forecast ('Soft Landing'), and the 70% covers all Ethernet transceivers, a different scope from the January series.
引用於: 第節 第4.7節; 圖 21

S-612 Synergy Research Group — Hyperscale Operators to Account for 67% of all Data Center Capacity by 2031
市場研究 · U.S. · 英文 · 7 April 2026 · 閱讀 28 September 2026
https://www.srgresearch.com/articles/hyperscale-operators-to-account-for-67-of-all-data-center-capacity-by-2031
"The number of large data centers operated by hyperscale companies reached 1,360 at the end of the fourth quarter of 2025, and new data from Synergy Research Group shows that they now account for 48% of the worldwide capacity of all data centers. Almost 60% of that hyperscale capacity is now in own-built, owned data centers with the balance being in leased facilities."
"With non-hyperscale colocation capacity accounting for another 20% of total capacity, that leaves enterprise on-premise data centers with just 32% of the total. This is in stark contrast to 2018, when 56% of data center capacity was in on-premise facilities. Looking ahead to 2031, hyperscale operators will account for 67% of all capacity, while on-premise will drop to just 19%."
支持的研判. Synergy Research Group reports hyperscale operators at 48% of worldwide data-centre capacity at the end of 2025 (on-premise 32%, colocation 20%) and forecasts 67% by 2031.
無法據以確立. Shares of physical capacity, not spending; no capacity figure in MW or GW and no 2031 colocation share are printed.
引用於: 第節 第5.6節; 圖 23

S-613 Synergy Research Group — Hyperscale Spending Spree is Driving Dramatic Growth in Data Center Capacity
市場研究 · U.S. · 英文 · 19 December 2025 · 閱讀 28 September 2026
https://www.srgresearch.com/articles/hyperscale-spending-spree-is-driving-dramatic-growth-in-data-center-capacity
"Over the last three years, quarterly hyperscale operator capex has ballooned, growing by almost 180% and reaching $142 billion in Q3 of this year. This has enabled a comparable 170% increase in the amount of operational capacity added each quarter."
"The companies with the broadest data center footprint are the leading cloud providers – Amazon, Microsoft, and Google. In addition to a huge data center footprint in their home US market, each also has multiple data centers in many other countries around the world. In aggregate, the three now account for 58% of all hyperscale data center capacity."
支持的研判. Synergy Research Group reports 1,297 hyperscale data centres in Q3 2025, quarterly hyperscale capex up almost 180% in three years to US$142 billion, and Amazon, Microsoft and Google at 58% of hyperscale capacity.
無法據以確立. Based on 21 large cloud and internet firms; it prints no capacity figure in MW or GW.
引用於: 第節 第5.6節

S-614 Gartner — Gartner Forecasts Worldwide AI-Optimized IaaS Spending to Grow 96% Through 2026
市場研究 · Global · 英文 · 10 August 2026 · 閱讀 28 September 2026
https://www.gartner.com/en/newsroom/press-releases/2026-08-10-gartner-forecasts-worldwide-artificial-intelligence-optimized-iaas-spending-to-grow-96-percent-in-2026
"In 2026, global spending on inference ($23.3 billion) will surpass that of training ($19 billion)."
取自來源表格之數值（轉錄）: Table 1: Spending on Infrastructure as a Service, Worldwide, 2025-2027 (Millions of Dollars) | Total AI-optimized IaaS, 2025 Spending: 21,529 | Total AI-optimized IaaS, 2025 Growth (%): 180.0 | Total AI-optimized IaaS, 2026 Spending: 42,276 | Total AI-optimized IaaS, 2026 Growth (%): 96.4 | Total AI-optimized IaaS, 2027 Spending: 66,143 | Total AI-optimized IaaS, 2027 Growth (%): 56.5 | Total IaaS, 2025 Spending: 222,170 | Total IaaS, 2025 Growth (%): 25.3 | Total IaaS, 2026 Spending: 287,347 | Total IaaS, 2026 Growth (%): 29.3 | Total IaaS, 2027 Spending: 359,899 | Total IaaS, 2027 Growth (%): 25.2 | Source: Gartner (August 2026)
支持的研判. Gartner forecasts AI-optimized IaaS spending of US$42.3 billion in 2026 (+96%) and US$66.1 billion in 2027, with inference (US$23.3 billion) surpassing training (US$19 billion) in 2026.
無法據以確立. One Gartner category of cloud capacity, not all AI spending; the release does not define AI-optimized IaaS or give 2027 inference dollars.
引用於: 第節 第8.4節; 圖 30

S-615 Federal Reserve Bank of St. Louis, On the Economy blog (Hannah Rubinton, Economist; Bontu Ankit Patro, Research Associate) — Tracking AI's Contribution to GDP Growth
研究 · U.S. · 英文 · 12 January 2026 · 閱讀 28 September 2026
https://www.stlouisfed.org/on-the-economy/2026/jan/tracking-ai-contribution-gdp-growth
"Together, the AI categories contributed 0.97 percentage points to real GDP growth in the first three quarters of 2025 (0.90 percentage points, excluding data centers, for which data were not available in 2000), higher than the 0.81 percentage points in 2000."
取自來源表格之數值（轉錄）: Dot-com vs. AI: Contribution of IT- or AI-Related Investment Categories to Real GDP Growth in 2000 and 2025 (real GDP growth in percent; contributions in annualized percentage points) | 2000, Real GDP Growth: 2.94% | 2000, Information Processing Equipment: 0.58 | 2000, Software: 0.11 | 2000, R&D: 0.12 | 2000, Data Centers: — | First 9 Months of 2025, Real GDP Growth: 2.51% | First 9 Months of 2025, Information Processing Equipment: 0.42 | First 9 Months of 2025, Software: 0.35 | First 9 Months of 2025, R&D: 0.13 | First 9 Months of 2025, Data Centers: 0.07
支持的研判. St. Louis Fed economists estimate that information-processing equipment, software, R&D and data centres added 0.97 percentage points to U.S. real GDP growth in the first three quarters of 2025 (39% of growth), against 0.81 points (28%) in 2000.
無法據以確立. The categories are proxies that include non-AI spending; the figures are averages of annualised quarterly contributions as of January 2026, and the shares are of growth, not of GDP.
引用於: 表 36, 37, 38, 40; 第節 第E.3節, 第E.9節; 圖 35

S-616 U.S. Bureau of Economic Analysis, via FRED (Federal Reserve Bank of St. Louis) — Private fixed investment: Private fixed investment in information processing equipment and software
政府 · U.S. · 英文 · 9 April 2026 · 閱讀 28 September 2026
https://fred.stlouisfed.org/series/A679RC1A027NBEA
直接文件: https://fred.stlouisfed.org/data/A679RC1A027NBEA.txt
取自來源表格之數值（轉錄）: A679RC1A027NBEA (Billions of Dollars, annual) | 1995: 253.845 | 1996: 279.190 | 1997: 316.549 | 1998: 349.334 | 1999: 396.403 | 2000: 450.601 | 2001: 423.583 | 2002: 389.257 | 2021: 975.002 | 2022: 1079.382 | 2023: 1111.800 | 2024: 1194.137 | 2025: 1381.380
取自來源表格之數值（轉錄）: FRED data file header, A679RC1A027NBEA | Source: U.S. Bureau of Economic Analysis | Release: Gross Domestic Product | Seasonal Adjustment: Not Seasonally Adjusted | Frequency: Annual | Units: Billions of Dollars | Date Range: 1947-01-01 to 2025-01-01 | Last Updated: 2026-04-09 7:56 AM CDT
支持的研判. BEA's annual series for private investment in information-processing equipment and software, in billions of current dollars (US$450.6 billion in 2000, US$1,381.4 billion in 2025).
無法據以確立. Not deflated and not AI-specific; excludes R&D and data-centre buildings; subject to BEA revision.
引用於: 表 36; 第節 第E.3節; 圖 36

S-617 U.S. Bureau of Economic Analysis, via FRED (Federal Reserve Bank of St. Louis) — Real Private Fixed Investment: Private Fixed Investment in Information Processing Equipment and Software
政府 · U.S. · 英文 · 13 March 2026 · 閱讀 28 September 2026
https://fred.stlouisfed.org/series/A679RL1A225NBEA
直接文件: https://fred.stlouisfed.org/data/A679RL1A225NBEA.txt
取自來源表格之數值（轉錄）: A679RL1A225NBEA (Percent Change from Preceding Period, annual) | 1995: 17.4 | 1996: 19.8 | 1997: 23.4 | 1998: 22.9 | 1999: 22.9 | 2000: 17.9 | 2001: -0.4 | 2002: -2.9 | 2021: 13.0 | 2022: 11.1 | 2023: 2.3 | 2024: 6.2 | 2025: 15.7
取自來源表格之數值（轉錄）: FRED data file header, A679RL1A225NBEA | Source: U.S. Bureau of Economic Analysis | Release: Gross Domestic Product | Seasonal Adjustment: Not Seasonally Adjusted | Frequency: Annual | Units: Percent Change from Preceding Period | Date Range: 1948-01-01 to 2025-01-01 | Last Updated: 2026-03-13 7:52 AM CDT
支持的研判. BEA's annual real growth of private investment in information-processing equipment and software (17.4% to 23.4% a year in 1995–2000; 15.7% in 2025).
無法據以確立. Annual-average growth, not fourth quarter over fourth quarter; not AI-specific; subject to BEA revision.
引用於: 表 36, 38; 第節 第E.3節; 圖 36

S-618 TrendForce — AI Server Industry Analysis—2Q26
市場研究 · Taiwan · 英文 · 3 July 2026 · 閱讀 28 September 2026
https://www.trendforce.com/research/download/RP260703GL3
"Market Scale: AI server shipments approach 2.8M units in 2026, with YoY growth near 30.5%; total server output value exceeds 624B USD, with AI servers accounting for over 74%."
"Liquid Cooling: High-end AI chip liquid cooling penetration rises from 41% in 2025 to 65%; overall AI chip rate reaches 53%, driven by NVIDIA and Google ASIC platforms."
支持的研判. TrendForce's 2Q26 AI-server report page gives 2026 AI-server shipments approaching 2.8 million units, with growth near 30.5%.
無法據以確立. The abstract of a report sold to subscribers; its figures are approximations, and TrendForce's August release raised the growth forecast.
引用於: 第節 第E.2節; 圖 33

S-619 TrendForce — TrendForce 2024: Riding the Wave of Revolutionary Tech Trends
市場研究 · Taiwan · 英文 · 17 October 2023 · 閱讀 28 September 2026
https://www.trendforce.com/presscenter/news/20231017-11886.html
"TrendForce estimates that shipments of AI servers (including those equipped with GPUs, FPGAs, and ASICs) will exceed 1.2 million units in 2023—a YoY increase of 37.7%—making up nearly 9% of total server shipments."
"This number is expected to grow 38% in 2024 with AI occupying a share of 12%."
支持的研判. TrendForce estimated in October 2023 that AI-server shipments would exceed 1.2 million units in 2023, up 37.7%, and forecast 38% growth for 2024.
無法據以確立. Published before the year closed, in an annual outlook covering many sectors; the page carries no later revision.
引用於: 第節 第E.2節

S-620 TrendForce — AI Server Growth in 2025 Faces Uncertainties, DeepSeek Effect to Boost AI Inference Share, Says TrendForce
市場研究 · Taiwan · 英文 · 12 February 2025 · 閱讀 28 September 2026
https://www.trendforce.com/presscenter/news/20250212-12469.html
"TrendForce's latest investigations reveal that global AI server shipments grew by 46% in 2024, driven by strong demand from CSPs and OEMs."
"Under this scenario, global shipments of AI servers in 2025 are projected to grow by nearly 28% year-over-year."
支持的研判. TrendForce estimated that AI-server shipments grew 46% in 2024 and, in its base case, forecast nearly 28% growth for 2025.
無法據以確立. Growth rates only, with no unit count or base; the 2025 figure is one of three scenarios.
引用於: 第節 第E.2節

S-621 TrendForce — Geopolitical Tensions and US Export Restrictions Slightly Temper AI Server Shipment Growth in 2025, Says TrendForce
市場研究 · Taiwan · 英文 · 2 July 2025 · 閱讀 28 September 2026
https://www.trendforce.com/presscenter/news/20250702-12630.html
"However, geopolitical tensions and US export restrictions impacting the Chinese market have led TrendForce to slightly revise its forecast, projecting a 24.3% YoY increase in global AI server shipments for the year."
"With sustained demand from North American CSPs and OEM customers, global AI server shipments are projected to maintain double-digit growth in 2025."
支持的研判. TrendForce revised its forecast of 2025 AI-server shipment growth to 24.3%, citing geopolitical tensions and U.S. export restrictions.
無法據以確立. A mid-year forecast, not a completed-year figure; the release gives no unit volume.
引用於: 第節 第E.2節

S-622 International Monetary Fund (IMF), The 2026 IMF Annual Report: Navigating a Precarious World (In Focus section) — AI: Deployment and Disruption
研究 · U.S.; global · 英文 · 23 September 2026 · 閱讀 28 September 2026
https://www.imf.org/annual-report/2026/in-focus/ai-deployment-and-disruption/
"In 2025, technology investments related to AI added an estimated 0.5 percentage point to US GDP growth."
"Private-sector-driven investment in AI could top $2 trillion globally in 2026, according to some external estimates, making it among the fastest-growing drivers of growth in recent years."
"In addition, productivity growth in the US has accelerated over the past several years, which may in part reflect early impacts from the adoption of AI."
"From a business perspective, there's a risk that the payoff from expensive investments in AI, increasingly debt financed, could prove illusory. This could lead to a sharp reversal in equity valuations, wealth destruction, and layoffs."
支持的研判. The IMF estimates that AI-related technology investment added 0.5 percentage point to U.S. GDP growth in 2025, cites external estimates of more than US$2 trillion of AI investment globally in 2026, and warns of debt and circular-financing risks.
無法據以確立. A short narrative with no method or source for either figure; the page carries no date.
引用於: 第節 第E.3節, 第E.6節, 第E.9節, 第E.10節; 表 37, 38

S-623 Bank for International Settlements (BIS Bulletin No 120) — Financing the AI boom: from cash flows to debt
研究 · U.S. · 英文 · 7 January 2026 · 閱讀 28 September 2026
https://www.bis.org/publications/bulletin-120-financing-ai-boom-cash-flows-debt
直接文件: https://www.bis.org/publications/bulletin-120-financing-ai-boom-cash-flows-debt.pdf
"For example, at around 1% of US GDP, it is similar in size to the US shale boom of the mid-2010s and half as large as the rise in IT investment during the dot-com boom of the 1990s."
"Leading firms in the information technology (IT) sector have historically financed much of their investments internally out of operating cash flows. However, the scale of current and anticipated AI-related investment needs is now so vast that firms are seeking external sources of funding."
"Indeed, free cash flows have recently lagged capital expenditures in absolute amounts."
"The size of anticipated investment needs will require firms to shift the source of financing from operating cash flows to debt, with private credit playing a rapidly increasing role."
支持的研判. The Bank for International Settlements finds AI-related investment surging, leading IT firms shifting from cash-flow funding to debt with private credit rising fast, and the rise in AI-related investment at around 1% of U.S. GDP.
無法據以確立. Private-credit figures rest on PitchBook data for broad AI-related sectors; the US$300–600 billion by 2030 is an extrapolation, not a forecast of borrowing.
引用於: 第節 第E.3節, 第E.6節, 第E.8節, 第E.10節, 第E.11節; 表 37, 38, 39

S-624 U.S. Bureau of Economic Analysis, via FRED (Federal Reserve Bank of St. Louis) — Private fixed investment: Nonresidential: Equipment: Information processing equipment: Computers and peripheral equipment
政府 · U.S. · 英文 · 9 April 2026 · 閱讀 28 September 2026
https://fred.stlouisfed.org/series/B935RC1A027NBEA
直接文件: https://fred.stlouisfed.org/data/B935RC1A027NBEA.txt
取自來源表格之數值（轉錄）: B935RC1A027NBEA (Billions of Dollars, annual) | 1995: 66.110 | 1996: 72.793 | 1997: 81.375 | 1998: 87.887 | 1999: 97.231 | 2000: 103.178 | 2001: 87.556 | 2002: 79.725 | 2021: 147.162 | 2022: 160.258 | 2023: 149.064 | 2024: 178.731 | 2025: 270.729
取自來源表格之數值（轉錄）: FRED data file header, B935RC1A027NBEA | Source: U.S. Bureau of Economic Analysis | Release: Gross Domestic Product | Seasonal Adjustment: Not Seasonally Adjusted | Frequency: Annual | Units: Billions of Dollars | Date Range: 1959-01-01 to 2025-01-01 | Last Updated: 2026-04-09 7:55 AM CDT
支持的研判. BEA's annual series for private investment in computers and peripheral equipment, in billions of current dollars (US$103.2 billion in 2000, US$270.7 billion in 2025).
無法據以確立. Not deflated and not AI-specific; one component of information-processing equipment; subject to BEA revision.
引用於: 第節 第E.3節; 圖 36

S-625 Board of Governors of the Federal Reserve System (Finance and Economics Discussion Series 2000-11) — Explaining the Investment Boom of the 1990s
研究 · U.S. · 英文 · 7 March 2000 · 閱讀 28 September 2026
https://www.federalreserve.gov/econres/feds/explaining-the-investment-boom-of-the-1990s.htm
直接文件: https://www.federalreserve.gov/pubs/feds/2000/200011/200011pap.pdf
"Computer investment is very sensitive to the cost of capital, far more so than investment in non-computing equipment."
"First, the pace at which firms replace depreciated capital has increased. Second, investment has been more sensitive to the cost of capital. We document that these two features stem from the special behavior of investment in computers and therefore propose a disaggregated approach. This produces an econometric model that successfully explains the 1990s equipment investment boom."
"Prices Matter: Many previous studies have found limited roles for price variables, stressing the ability of an accelerator model to explain the cyclical behavior of investment. In contrast, we find an important role for equipment prices. Specifically, falling computer prices played a crucial role in the investment boom of the 1990s."
支持的研判. A Federal Reserve Board working paper of 2000 explains the 1990s equipment boom by faster replacement of depreciating computers and computer investment's sensitivity to the cost of capital, with falling computer prices crucial.
無法據以確立. Econometric estimates for the boom up to the late 1990s; the paper does not cover the 2000–2001 bust, and the views are the authors'.
引用於: 第節 第E.3節, 第E.5節, 第E.7節

S-626 U.S. Bureau of Economic Analysis (working paper WP2002-2) — Information Processing Equipment and Software in the National Accounts
研究 · U.S. · 英文 · April 2002 · 閱讀 28 September 2026
https://www.bea.gov/research/papers/2002/information-processing-equipment-and-software-national-accounts
直接文件: https://www.bea.gov/sites/default/files/papers/WP2002-2.pdf
"In the U.S. national income and product accounts (NIPA's), most of the types of goods in the investment category "information processing (IP) equipment and software" have experienced rapidly changing technology and are thus candidates for inclusion in the new economy."
"The NIPA price indexes for computers and peripheral equipment, computer software, and communication equipment all, at least in part, include quality adjustments based on hedonic studies."
支持的研判. BEA economists describe rapid technical change in information-processing equipment and software and the hedonic quality adjustment of its price indexes.
無法據以確立. A 2002 staff paper whose views are the authors' own, not a statement of current BEA practice.
引用於: 第節 第E.3節

S-627 Board of Governors of the Federal Reserve System (speech by Governor Lisa D. Cook) — The Opportunities and Risks AI Presents for the Economy and Financial System
政府 · U.S. · 英文 · 27 May 2026 · 閱讀 28 September 2026
https://www.federalreserve.gov/newsevents/speech/cook20260527a.htm
"To date, companies have announced more than $1.5 trillion in data-center plans, only a small portion of which have been realized."
"Many of the hyperscaler firms have executed large investment-grade bond deals in recent months to fund AI capital expenditures."
"Another emerging trend that may have implications for financial stability relates to the fact that firms are increasingly tapping debt markets to finance the capital investments relating to AI infrastructure."
支持的研判. Federal Reserve Governor Cook notes firms increasingly tapping debt markets for AI infrastructure, hyperscalers' large investment-grade bond deals and smaller developers' private and asset-backed debt.
無法據以確立. A speech with no amounts; its financing statements rest on press reports cited in its footnotes.
引用於: 表 37, 38, 39, 41; 第節 第E.6節

S-628 National Bureau of Economic Research (NBER Working Paper 25148) — The Productivity J-Curve: How Intangibles Complement General Purpose Technologies
研究 · U.S. · 英文 · October 2018 · 閱讀 28 September 2026
https://www.nber.org/papers/w25148
"General purpose technologies (GPTs) such as AI enable and require significant complementary investments, including co-invention of new processes, products, business models and human capital. These complementary investments are often intangible and poorly measured in the national accounts, even when they create valuable assets for the firm."
"We develop a model that shows how this leads to an underestimation of productivity growth in the early years of a new GPT, and how later, when the benefits of intangible investments are harvested, productivity growth will be overestimated."
支持的研判. Brynjolfsson, Rock and Syverson model a productivity J-curve in which intangible complementary investment makes measured productivity understate a general purpose technology's early gains, and find AI-related effects small but growing.
無法據以確立. A model result, read from the working paper's abstract; not a forecast.
引用於: 表 37, 38; 第節 第E.7節, 第E.9節

S-629 Board of Governors of the Federal Reserve System (Finance and Economics Discussion Series 2000-20) — The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?
研究 · U.S. · 英文 · April 2000 · 閱讀 28 September 2026
https://www.federalreserve.gov/econres/feds/the-resurgence-of-growth-in-the-late-1990s-is-information-technology-the-story.htm
直接文件: https://www.federalreserve.gov/pubs/feds/2000/200020/200020pap.pdf
"All in all, we estimate that the use of information technology and the production of computers accounted for about two-thirds of the 1 percentage point step-up in productivity growth between the first and second halves of the decade."
"Thus, to answer the question posed in the title of this paper, information technology largely is the story."
支持的研判. Federal Reserve Board economists estimate that IT use and computer production accounted for about two-thirds of the 1 percentage point step-up in U.S. productivity growth in the late 1990s.
無法據以確立. A growth-accounting estimate read from the abstract; it says nothing about the later infrastructure retrenchment.
引用於: 第節 第E.5節; 表 38

S-630 Federal Reserve Bank of Richmond (Economic Quarterly, Volume 89/4, Fall 2003) — Boom and Bust in Telecommunications
研究 · U.S. · 英文 · 2003 · 閱讀 28 September 2026
https://www.richmondfed.org/publications/research/economic_quarterly/2003/fall/couperhejkalwolman
直接文件: https://www.richmondfed.org/-/media/RichmondFedOrg/publications/research/economic_quarterly/2003/fall/pdf/wolman.pdf
"Since the final quarter of 2000, year-over-year communications investment growth was negative for seven straight quarters. In terms of investment levels, the low point came in quarter four of 2001, at under $93 billion—only 69 percent of the same figure one year earlier."
"From the first quarter of 1996 to the fourth quarter of 2000, investment in communications equipment grew from approximately $62 billion per year to over $135 billion per year in constant 1996 dollars (Figure 2). This represents average annual growth of nearly 18 percent."
支持的研判. Richmond Fed economists record U.S. communications-equipment investment rising from about US$62 billion a year (1996 Q1) to over US$135 billion (2000 Q4) in constant 1996 dollars, then falling to under US$93 billion by 2001 Q4.
無法據以確立. BEA data as available in 2003, with approximate levels; the views are the authors', and the article makes no comparison with AI.
引用於: 第節 第E.5節; 圖 37; 表 38

S-631 Federal Reserve Bank of New York (Current Issues in Economics and Finance, Volume 10, Number 6) — What Investment Patterns across Equipment and Industries Tell Us about the Recent Investment Boom and Bust
研究 · U.S. · 英文 · May 2004 · 閱讀 28 September 2026
https://www.newyorkfed.org/research/current_issues/ci10-6.html
直接文件: https://www.newyorkfed.org/medialibrary/media/research/current_issues/ci10-6.pdf
"Significantly, investment by these industries in 2000 likely reflected a severe misperception of future returns."
"We find that investment trends in the communications industries indeed had a disproportionate effect—in relation to the industries' share of the overall economy—on both the end of the boom in 2000 and the onset of the bust in 2001."
支持的研判. A New York Fed study finds that communications-industry investment had a disproportionate effect on the end of the 2000 boom and the 2001 bust and likely reflected a misperception of future returns.
無法據以確立. The profit-expectation finding is inferred from a standard investment model under stated assumptions and is hedged ('likely').
引用於: 第節 第E.5節

S-632 Board of Governors of the Federal Reserve System (speech by Vice Chair Philip N. Jefferson) — AI, the Economy, and Financial Stability
政府 · U.S. · 英文 · 21 November 2025 · 閱讀 28 September 2026
https://www.federalreserve.gov/newsevents/speech/jefferson20251121a.htm
"In the late 1990s and early 2000s, many, albeit not all, dot-com firms typically had little to no realized earnings and only speculative revenue prospects even as they obtained external funding and were listed on public exchanges with relative ease amid pronounced investor enthusiasm. In contrast, the firms most closely identified with AI technologies in today's market generally have well-established and growing earnings streams."
"The second difference is that because of their solid earnings base, the price-to-earnings ratios of AI-related firms have thus far remained well below the peak ratios of dot-com firms in the late 1990s and early 2000s."
"However, it is important to acknowledge that the rapid growth of private capital markets—which are not publicly traded—in recent years may obscure the extent of investor enthusiasm for financing AI firms, as a growing number of private firms reportedly now identify as AI-focused or AI-native enterprises."
"Finally, with respect to financial leverage, both dot-com previously and AI firms recently exhibited limited reliance on debt financing for the most part."
支持的研判. Federal Reserve Vice Chair Jefferson contrasts the dot-com and AI eras: established earnings, price-to-earnings ratios well below dot-com peaks, about 50 AI-focused listed firms against more than 1,000 dot-com listings, and limited use of debt in both.
無法據以確立. A speech, with no P/E values or earnings data; the firm counts are not like-for-like, the AI count resting on one ETF's focus.
引用於: 第節 第E.6節, 第E.8節; 表 38

S-633 Bank for International Settlements (BIS Working Papers No 1367) — The AI investment race
研究 · U.S. · 英文 · 14 July 2026 · 閱讀 28 September 2026
https://www.bis.org/publications/working-paper-1367-ai-investment-race
直接文件: https://www.bis.org/publications/working-paper-1367-ai-investment-race.pdf
"We study a dynamic contest in which firms competing for a few dominant positions over-commit resources. The over-investment leaves the sector exposed to revenue disappointment that could turn boom into bust. The larger the boom, the deeper the eventual bust."
"The race to commit early through debt and circular financing also makes a bust more likely. Calibrated to balance sheet and deal data, the model points to over-investment of around 1.5 times the efficient level, rising to around three times where demand is less elastic. A network analysis shows that stress in one firm could cascade to others through chains of financial exposures."
"At $\beta = 0.80$ the model sits closer to the frictionless limit $\beta \to 1$ where the contest externality vanishes, so the headline distortions should be read as floors rather than point estimates."
"The financing structure is also central to the mechanism, with debt and circular stakes both propelling the boom and exacerbating the losses during busts."
支持的研判. A Bank for International Settlements working paper models AI investment as a winner-take-most contest, with over-investment of around 1.5 times the efficient level (about three times where demand is less elastic) and debt deepening the bust.
無法據以確立. A calibrated theoretical model: its ratios are not an observed measure of excess capacity, and the author calls them floors rather than point estimates.
引用於: 第節 第E.6節, 第E.7節, 第E.11節; 表 40

S-634 Organisation for Economic Co-operation and Development (OECD) — OECD Economic Outlook, Interim Report September 2026: Weathering Successive Shocks
研究 · Global · 英文 · 23 September 2026 · 閱讀 28 September 2026
https://www.oecd.org/en/publications/oecd-economic-outlook-interim-report-september-2026_f751d02b-en.html
直接文件: https://www.oecd.org/content/dam/oecd/en/publications/reports/2026/09/oecd-economic-outlook-interim-report-september-2026_8312492f/f751d02b-en.pdf
"Strong AI-related activity has also bolstered investment, production and trade."
"Prolonged disruptions to exports from the Middle East, weather-related supply shocks associated with a very strong El Niño, further increases in long-term sovereign bond yields or AI investment returns that fail to meet expectations are key downside risks that could weaken growth."
支持的研判. The OECD's September 2026 interim outlook states that strong AI-related activity has bolstered investment, production and trade, and names AI investment returns below expectations among key downside risks.
無法據以確立. Only the landing-page abstract could be read; it gives no projections and no quantification of AI-related investment.
引用於: 第節 第E.6節

S-635 American Economic Association (American Economic Journal: Macroeconomics) — The Productivity J-Curve: How Intangibles Complement General Purpose Technologies
研究 · U.S. · 英文 · January 2021 · 閱讀 28 September 2026
https://www.aeaweb.org/articles?id=10.1257/mac.20180386
"We apply our method to US data and find that adjusting for intangibles related to computer hardware and software yields a TFP level that is 15.9 percent higher than official measures by the end of 2017."
"We call this phenomenon the Productivity J-curve."
支持的研判. The published version of the productivity J-curve paper finds U.S. TFP 15.9% above official measures by the end of 2017 after adjusting for hardware and software intangibles.
無法據以確立. Abstract only; a model-based adjustment, not a forecast.
引用於: 第節 第E.9節

附錄E

經濟與財務校準：成長、GDP、資本循環與dot-com時代

美國設備與智慧財產投資近期的熱潮大部分由AI帶動，其占GDP的比重已接近2000年的水準。dot-com時代所提出的問題是：各層級的資本投入速度，是否超出使用情況、生產力與現金流所能吸納的範圍。

適用讀者： 財務、投資、政府與策略讀者，以及任何需要跨來源比較成長率、GDP占比或資本循環者。

本附錄將本報告的產業證據置於經濟尺度之上。它區分四項經常被混用的量——投資水準、成長率、占國內生產毛額（gross domestic product, GDP）的比重，以及對GDP成長的貢獻——並加入支應此輪建置的財務結構。其後，本附錄以聯準會、美國經濟分析局（Bureau of Economic Analysis, BEA）、國際清算銀行（Bank for International Settlements, BIS）與國際貨幣基金（International Monetary Fund, IMF）所公布的指標，將1990年代後期的資訊科技（information technology, IT）投資循環與當前循環加以對照。此一對照旨在校準與檢驗，而非預測。本附錄回答的問題是：哪些可衡量的條件能使今日的資本保持生產性，哪些條件又會使其中某一層級的建置超前於需求？

E.1 經濟分析衡量什麼，不衡量什麼

四項量回答四個不同的問題。 以美元計的投資水準衡量規模：供應商營收與資金需求。其成長率衡量動能，即此輪循環是否正在加速。其占GDP的比重衡量相對於經濟體規模的資本密集度；當分子與分母出自同一套國民所得帳時，這是跨數十年最乾淨的比較。其對實質GDP成長的貢獻以百分點表示，兼具規模與成長兩個面向：某一類別的貢獻大致等於其占GDP的比重乘以其實質成長率，因此唯有在規模龐大且成長迅速時，其貢獻才最大。這四項量絕不繪於同一條數值軸上。第五個問題，即該投資如何取得資金，則須借助企業財務報表，因為營運現金流、債務、租賃、預付款與客戶出資的資產，可能使同一座機櫃或同一處場址出現在報表的不同位置（第E.10節）。

A 全球支出與國民所得帳須使用相稱的分母。 企業與市場研究的數字多為全球數字，而此處用以衡量投資對比產出的國民所得帳則屬美國。以全球資本數字除以美國GDP會混用不同地理範圍，並高估該比率；因此本報告僅以全球成長率作為全球比率的對照基準，例如國際貨幣基金（IMF）以購買力平價權數所作的世界實質成長率預估，2026年為3.0%、2027年為3.4%[S-593]；以名目支出成長率與之對照，僅能校準量級。美國數列僅對比美國GDP。如圖1中17%對比3%這類成長率倍數僅用於校準量級，並非對成長的貢獻[S-086]。

E.2 成長率的數學：CAGR、YoY、QoQ與預測版次

= OCP所公布的22.2%是兩個端點之間的複合成長率，本報告則將其所隱含的路徑繪為橋接值。 IDC為OCP所作的研究，將OCP認可基礎設施的支出列為2025年1,320億美元、2029年2,950億美元，CAGR為22.2%[S-004]。計算公式：CAGR = （期末 ÷ 期初）$^{1/\text{years}} - 1$；此處依已公布的四捨五入端點計算為（295 ÷ 132）$^{1/4} - 1 = 22.27\%$，對比已公布的22.2%。以已公布的22.2%自2025年推算，2026年為1,613億美元、2027年為1,971億美元、2028年為2,409億美元、2029年為2,943億美元，其中2029年仍採已公布的2,950億美元。各中間值並非IDC的預測，第一頁將其繪為不帶標記的細線（圖1）。國際能源總署（IEA）2025年485 TWh與2030年950 TWh的數值[S-005]隱含（950 ÷ 485）$^{1/5} - 1$ = 每年14.4%，並以同一方式作橋接。

= 逐季成長率連鎖相乘即構成全年的變動，這也是第一頁DRAM列載明全部五個季度的原因。 TrendForce對DRAM產業營收的估計值，3Q25成長30.9%、4Q25成長29.4%、1Q26成長81%、2Q26成長59.5%[S-120; S-121; S-035; S-118]。連鎖相乘後，1.309 × 1.294 × 1.81 × 1.595 = 4.89，即1,547.3億美元相對於316.3億美元的倍數[S-117; S-118]：四個季度內+389%。若改採同一期間的複合成長率，則會抹平構成此一證據的2026年初加速走勢。該數列衡量的是營收而非位元：TrendForce將1Q26的跳升歸因於傳統DRAM合約價格在一個季度內上漲約93–98%[S-035]，因此該倍數主要衡量的是稀缺之下所擷取的價值，而非所出貨的記憶體；就3Q25而言，TrendForce亦將部分原因歸於位元出貨量與HBM數量的提升[S-120]。

A 以100%為上限的占比，須以百分點解讀。 TrendForce預期液冷（liquid cooling）將自2025年約占AI晶片的33%，升至2026年的53%，並於2027年接近60%[S-006]：兩年內+27個百分點。若以成長率表示，同一變動約為每年35%；若依該速率持續，至2029年即會超過100%，而任何滲透率均不可能如此。2024年14%的數值亦採另一基礎：TrendForce 2025年8月的發布將其列為AI資料中心的滲透率[S-591]，較其以AI晶片為基礎的數列早一年，因此本報告既不將兩者相連，亦不跨兩者計算成長率。

F TrendForce於六個月內兩度上調其2026年雲端資本支出預測，並一度擴大其企業樣本。 2026年2月，該機構預估八家領先雲端服務供應商（CSP，不含ByteDance）2026年的合計資本支出（capex）將超過7,100億美元，約較2025年高出61%[S-597]。5月，該機構將納入ByteDance後的九家CSP列為約8,300億美元，「with the annual growth rate raised from 61% to 79%」[S-598]。8月，該機構估計這九家將超過8,867億美元，約成長90%，2027年則約為1.3兆美元[S-013]。其對2026年AI伺服器出貨成長率的預測，至4月之前維持在約28%，其後上調：1月為逾28%、4月為約28%、7月接近30.5%、8月近31%[S-008; S-038; S-618; S-013]。

F Gartner對2026年AI基礎設施的預測，自1月至9月上調8.6%。 Gartner 2026年的三次發布，將該年AI基礎設施支出分別列為1月13,660億美元、5月14,320億美元、9月14,840億美元，2027年則分別為17,480億美元、18,900億美元與19,780億美元[S-599; S-600; S-601]。其AI支出總額自2.53兆美元上調至2.67兆美元，惟9月時Gartner「separated cross-functional agents and assistants from AI software and added consumer agents and assistants」，故各項總額並非基於同一基礎；入圖者為AI基礎設施一項，因5月與9月的發布對其採相同定義（1月的發布未給出定義）[S-601; S-600]（圖33）。

A 修正本身即帶有資訊。 兩項支出預測的每一版次均高於前一版次，AI伺服器預測則於4月之後上調；TrendForce 5月的修正係在企業財測指引上調之後提出，其發布文件並轉述了Microsoft、Google與Meta的指引[S-598]。遞減的版次序列，將是在任何企業削減預算之前，產能正常化路徑（第E.11節）的早期徵兆。較早的各版次顯示預測可能朝兩個方向失準：TrendForce於2023年10月預測2024年AI伺服器出貨成長38%，但2025年2月估計為46%；其後又將2025年的預測自2025年2月的近28%下調至7月的24.3%[S-619; S-620; S-621]。本報告採用最新版次並註明其日期，其餘各版次則保留為信念如何變動的紀錄。

E.3 GDP校準：水準、占比與對成長的貢獻

GDP為資本支出提供了共同的尺度。一家公司的資本預算或某項產業預測說明的是支出金額有多少；而將同一數字對比經濟體的規模，則可說明此輪建置的規模是否足以帶動成長，以及其是否與過往的熱潮相似。兩項指標回答不同的問題：某一類別占GDP的比重說明其規模有多大，其對成長的貢獻則說明它在特定期間內增添了多少。本節同時研讀兩者，且僅限美國數列，並維持各類別依其來源所定義的範圍。

F 美國在智慧財產與設備上的投資，已接近其2000年占GDP的比重。 聯準會理事會2026年7月的一份幕僚報告，依美國經濟分析局（BEA）資料繪出美國智慧財產（intellectual property, IP）與設備投資占GDP的比重[S-595]。該比重自1995年的9.88%升至2000年11.49%的高峰，至2002年回落至9.75%；2023年為10.53%、2024年為10.54%、2025年為10.91%，2026年第1季為11.33%[S-595]。依該報告的用語，該比重「turned sharply higher in 2025」，且2026年初較其2024年水準高出約0.8個百分點，「only a touch lower than the year 2000 peak」[S-595]（圖34）。

A 其步調與1990年代後期相似，但起點較高；而該水準所衡量的是規模，而非結果。 該報告指出，近期熱潮有大部分「has been AI-driven」，且其相對於GDP成長的步調「has been comparable to that seen during the 1990s, but this time starting at a much higher base level」[S-595]。11.33%與11.49%之間約0.15個百分點的落差（以未四捨五入的數值計為0.153），說明美國經濟再次將與dot-com建置高峰時期大致相當比例的產出投入設備與智慧財產，但並未說明該等資本將獲得何種報酬。該報告自身的模型，將過度投資視為不確定性下最適投資的一種可能結果，而非「irrational exuberance」的結果[S-595]。

? 未解問題——該報告對其數列的界定，僅止於美國智慧財產與設備投資，以及以美國經濟分析局（BEA）為來源。 它並未說明分子屬民間或全體、名目或實質[S-595]。因此本報告僅將該比率與自身比較，作為在單一基礎上跨兩個期間的規模衡量，而不將其與其他數列合併。

F AI相關投資類別2025年對美國成長的貢獻，高於可比的資訊科技類別於2000年的貢獻。 聖路易聯邦準備銀行的經濟學家Hannah Rubinton與研究助理Bontu Ankit Patro，衡量了美國經濟分析局（BEA）三項投資數列——資訊處理設備、軟體與研究發展（R&D）——以及人口普查局自2014年起的資料中心興建資料，對實質GDP成長的貢獻[S-615]。以2025年前三季平均計，此四項為2.51%的年化實質GDP成長貢獻了0.97個百分點；相對地，2000年可取得的三個類別以該年全部四季平均計，為2.94%的成長貢獻了0.81個百分點；若不計資料中心，2025年的數字為0.90[S-615]。以占成長的比重計即為39%，不計資料中心則為36%，相對於2000年的28%[S-615]（圖35）。

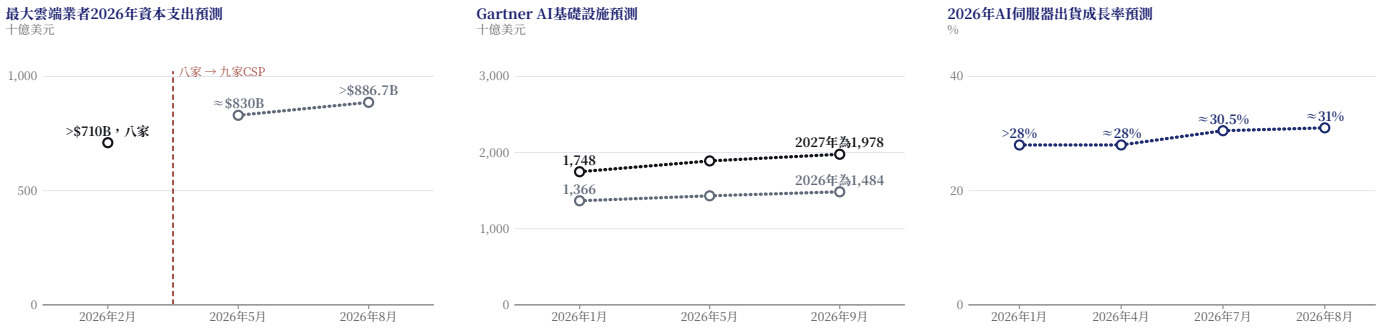
A 組成已自硬體轉向軟體。 資訊處理設備2000年貢獻0.58個百分點，2025年為0.42個百分點；軟體則自0.11個百分點升至0.35個百分點，占2025年成長的14%，相對於2000年的3.7%；研發（R&D）大致持平，分別為0.12與0.13個百分點[S-615]。依此衡量，當前循環較dot-com建置更偏重軟體，研發的貢獻則大致未變——縱使本報告記錄了記憶體、封裝與電力的實體稀缺。各類別為作者所採的代理指標，而非AI支出的衡量[S-615]。該篇文章並未扣除進口淨額：其所計貢獻為

表36。 每一個經濟用語只回答一個問題，用於其他問題即生誤導
比較兩個成長數字之前，須先確認二者屬同一種數字，且涵蓋同一種期間。

用語	衡量內容	適用時機	不適用時機	本報告中的實例
複合年成長率（CAGR）	（期末值 ÷ 期初值）$^{1/years} - 1$	兩端點相隔逾一年，且複合計算具有意義時	某一占比接近飽和，或為逐季的衝擊型數列時	OCP認可支出，年增22.2%，2025–29年[S-004]
年對年（YoY）	相對於一年前同期的變動	已結束的期間；具季節性的數列	端點為不規則的預測日期時	DRAM產業營收，自2Q25至2Q26為+389%[S-117; S-118]
季對季（QoQ）	相對於前一季的變動	短週期的價格與營收	用於長期結構性預測時	DRAM產業營收，1Q26為+81%[S-035]
百分點（pp）	兩個占比之間的差額	滲透率、組成與GDP占比	用於營收成長時	AI晶片的液冷（liquid cooling），自33%升至約60%：+27個百分點[S-006]
成長率倍數	一個成長率除以另一個成長率	校準量級	會因此暗示某一量在水準上較大時	2025年資料中心用電量+17%，對比全部用電量+3%[S-086]
占GDP比重	某一流量除以GDP，且採同一套帳	資本密集度的時序變化	會以全球支出對比單一國家的GDP時	美國智慧財產與設備投資，2026年第1季為GDP的11.33%[S-595]
對成長的貢獻	某一類別所增添的實質GDP成長百分點	規模與成長同時重要時	所指實為占經濟體的比重時	AI相關類別，2025年前九個月為0.97個百分點[S-615]
名目與實質	當期美元；剔除價格變動後的數量	營收與融資採名目值，產出與生產力採實質值	會在同一項比較中混用兩者時	2025年資訊處理設備與軟體的民間投資：13,810億美元，實質+15.7%[S-616; S-617]
價格、數量與組成	營收 ≈ 價格 × 數量 × 組成	記憶體、伺服器與企業營收	來源揭露的內容不足以將三者分離時	1Q26伺服器支出+30.7%，對比出貨量+3.3%[S-592]
推導橋接值	已公布的成長率在已公布端點之間所給出的數值	繪出端點預測所隱含的路徑	該等數值會被解讀為預測時	OCP認可支出與資料中心用電量（圖1）
預測版次	於特定日期所公布的預測	顯示某一觀點如何變動時	會將各版次接合為單一數列時	2026年雲端資本支出（capex）與AI基礎設施預測（圖33）

本報告所使用的經濟語彙、各用語所衡量的內容、其適用與不適用的時機，並附本報告中的一項實例及其來源。

圖33。 2026年的各項預測在年內遭上調：雲端資本支出、AI基礎設施，以及自7月起的AI伺服器成長率



每一個點均為對同一日曆年的預測，依所示日期公布。左圖：最大雲端業者2026年的合計資本支出（capex），以十億美元計，資料來自TrendForce（2月為八家公司，5月起為九家；樣本變動處已標註）[S-597; S-598; S-013]。中圖：Gartner對2026年與2027年AI基礎設施的預測，以十億美元計[S-599; S-600; S-601]。右圖：TrendForce對2026年AI伺服器出貨成長率的預測，以%計[S-008; S-038; S-618; S-013]。

範圍 本圖為預測版次，而非隨時間變化的支出：其中沒有任何一點是已實現值，亦無任何版次與另一版次接合。TrendForce 2月的數字（「exceed $710 billion」）與8月的數字（「will exceed US$886.7 billion」）均為下限；其1月的成長率數字（「more than 28%」）亦為下限。

某一類別占GDP的名目比重乘以其實質成長率，並以全額計入進口設備，而GDP則於淨出口項下扣除進口。

F 國際貨幣基金（IMF）與國際清算銀行（BIS）經濟學家的估計一致認為AI相關投資在總體經濟上具重要性，但對其規模的看法不同。 IMF的2026年年報指出，2025年「technology investments related to AI added an estimated 0.5 percentage point to US GDP growth」[S-622]。由Iñaki Aldasoro、Sebastian Doerr與Daniel Rees撰寫的國際清算銀行（Bank for International Settlements, BIS）公報發現，半導體製造設施與資料中心的支出在2022年以後的三年間，對GDP成長平均貢獻0.4個百分點，而2022年之前的貢獻微不足道；該公報並將AI相關投資的增幅（約為美國GDP的1%）列為「half as large as the rise in IT investment during the dot-com boom of the 1990s」[S-623]。各項估計所涵蓋的類別、期間與方法互異，因此本報告將其並列呈現，不予平均。

F 同一筆投資，名目數列與實質數列講述不同的故事。 美國經濟分析局（BEA）資訊處理設備與軟體民間投資的當期美元數列，自1995年的2,538億美元升至2000年的4,506億美元，2002年回落至3,893億美元，再自2023年的11,118億美元升至2025年的13,814億美元[S-616]。以實質計，同一筆投資在1995年至2000年的每一年均成長17.4%至23.4%，其後2001年下滑0.4%、2002年下滑2.9%；2021年至2025年則分別成長13.0%、11.1%、2.3%、6.2%與15.7%[S-617]。僅就電腦及週邊設備而言，其當期美元投資2025年成長51.5%至2,707億美元，相對於2000年1,032億美元的高峰[S-624]（圖36）。

= 兩輪循環的名目複合成長率看似相近；實質成長率則不然。 依美國經濟分析局（BEA）的當期美元數值，該投資1995年至2000年每年成長12.2%，即（450.6 ÷ 253.8）$^{1/5} - 1$；2023年至2025年每年成長11.5%，即（1,381.4 ÷ 1,111.8）$^{1/2} - 1$[S-616]。將BEA的實質成長率複合計算，1996–2000年為每年21.4%，2024–2025年為每年10.8%[S-617]。此一落差隱含該投資的價格在1990年代後期每年下跌約7.6%——當時經品質調整後的電腦價格下跌，對該輪投資熱潮發揮了關鍵作用[S-625; S-626]——並於2024–2025年每年上漲約0.6%，此與第3.2節所述零組件價格上漲相符，惟不足以構成證明。

E.4 資本循環診斷指標：價格、數量、使用、報酬與資金來源

資本循環的判斷不僅取決於其規模。 同樣的投資水準，可能屬穩健亦可能屬建置過度，視產能被使用的程度、使用是否轉為營收、營收是否足以涵蓋資金成本，以及該建置如何取得資金而定。因此本報告以一組固定重複的指標研讀AI建置，各指標並對應能夠提供該指標的來源類型（表37）。其中數項尚未以可比形式公開，尤以已安裝加速器（accelerator）與已供電容量的使用率，以及2025–2026年建置的報酬為著；此等項目以未解問題列示，而非逕予估計。

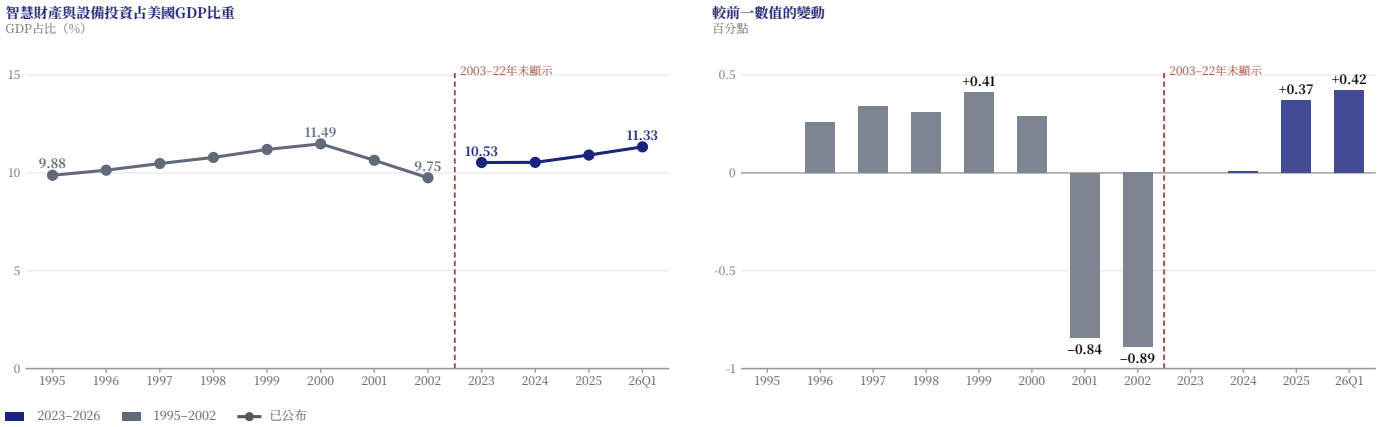
? 未解問題——AI加速器的使用率並無公開且可比的衡量。 所閱讀的來源提供的是新增、已簽約或已預訂的產能，部分並提供已啟用電力對比已簽約電力，但未提供已安裝加速器（accelerator）或已供電容量中投入生產使用的比例。缺少此一指標，即無法僅憑投資水準判讀資本過剩或短缺，這也是本報告改以價格、交期（lead time）與合約研讀兩者的原因。

E.5 dot-com時代的基準：1995–2002年

1990年代後期的熱潮，是由通用技術帶動的資本循環中最接近的先例，也是最常被用來質疑今日支出的例子。惟唯有依其實際經過加以解讀，它才具參考價值：可衡量的生產力確有實質提升，同時伴隨集中於單一層級——通訊設備——的建置過度。本節依當時的聯準會研究及其後的各項分析陳述此一紀錄，使兩輪循環的對照（第E.7節；第E.8節）立基於前一輪循環的實際表現，而非其名聲。

F 1990年代後期的投資熱潮提升了可衡量的生產力。 聯準會理事會的經濟學家於2000年估計，「the use of information technology and the production of computers accounted for about two-thirds of the 1 percentage point step-up in productivity growth between the first and second halves of

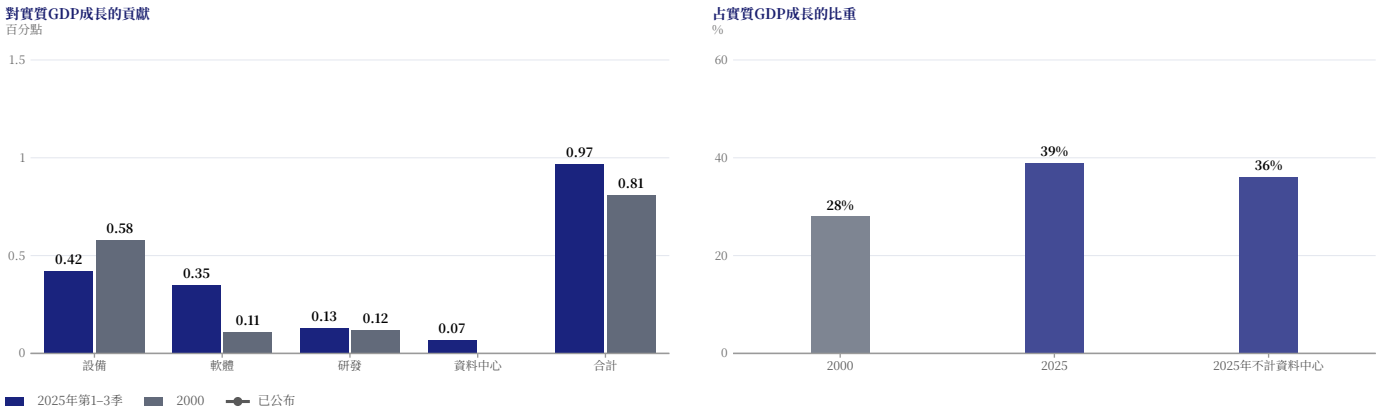
圖34。 美國設備與智慧財產投資占GDP的比重已回升至接近2000年水準，且正以1990年代後期的速度上升



左圖：美國智慧財產與設備投資占GDP的比重，以%計，出自聯準會理事會2026年7月6日的幕僚報告（FEDS Notes，Geng Li），並依美國經濟分析局（BEA）資料繪製；為年度數值，2026年第1季單獨列示[S-595]。右圖：較前一數值的變動，以百分點計，依同一數列推導；2026年第1季的柱狀為單一季度與2025年年度數值的比較。

範圍 該報告未說明此一數列屬民間或全體、名目或實質；此處將其視為採單一基礎的單一數列，且僅與自身比較。2003年至2022年未予顯示。

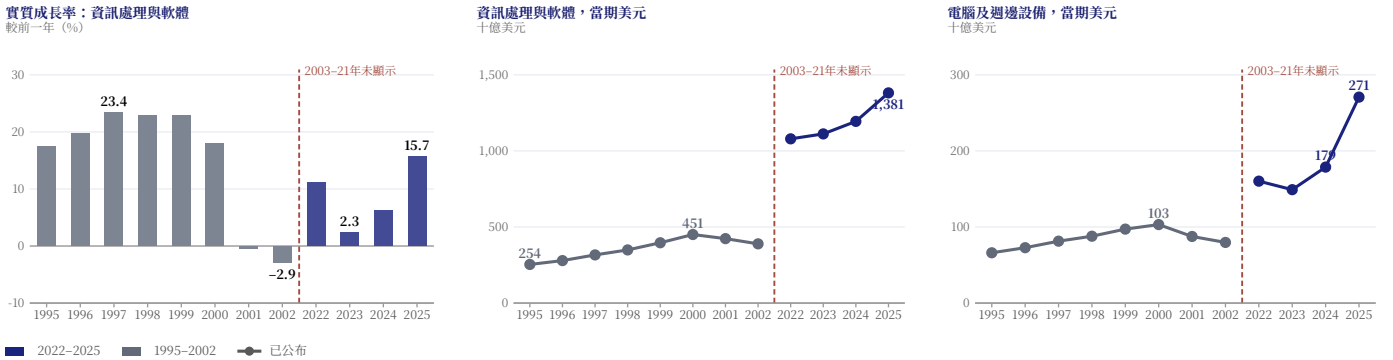
圖35。 AI相關類別2025年對美國成長的增益高於資訊科技類別於2000年的增益；此一上升由軟體帶動



左圖：依投資類別劃分，對美國年化實質GDP成長的貢獻（設備係指資訊處理設備），以百分點計，為2000年四個季度與2025年前三季的平均值，出自聖路易聯邦準備銀行，並依美國經濟分析局（BEA）與人口普查局資料；資料中心無2000年數值，2000年的合計僅涵蓋三個類別[S-615]。右圖：各類別占實質GDP成長的比重，以%計[S-615]。

範圍 各類別為作者就AI相關投資所採的代理指標，而非AI支出的衡量；各項占比係占實質GDP成長，而非占GDP。各項貢獻以全額計入進口設備，因GDP係於帳內其他項目中扣除進口。

圖36。 1990年代的熱潮是一段持續的高實質成長期；當前數列則於2025年急速加速



美國民間固定投資，資料來自美國經濟分析局（BEA），經FRED取得。左圖：資訊處理設備與軟體，較前一年的實質成長率，以%計[S-617]。中圖：同一筆投資的當期美元金額，以十億美元計[S-616]。右圖：電腦及週邊設備，當期美元，以十億美元計[S-624]。2003年至2021年未予顯示。

範圍 本圖涵蓋上述類別的全部民間投資，而非AI專屬支出；資料中心建物與研發（R&D）不計入。當期美元數列混合了價格與數量。各數值為FRED於2026年3月與4月所印數值，可能經美國經濟分析局（BEA）修正。

the decade」，並認為「information technology largely is the story」[S-629]。同年的一份理事會工作論文，以電腦投資的兩項特徵解釋該輪設備熱潮：企業更快汰換已折舊的資本，整體折舊率自1989年的0.13升至1997年的0.16；且電腦投資「very sensitive to the cost of capital」，而電腦價格的下跌發揮了「a crucial role」[S-625]。

同一輪熱潮在單一層級造成建置過度。 里奇蒙聯邦準備銀行《Economic Quarterly》的一篇文章（Couper、Hejkal與Wolman，2003年）記載，以1996年不變美元計，美國通訊設備投資自1996年第1季每年約620億美元，成長至2000年第4季逾1,350億美元，平均每年成長近18%[S-630]。其後年對年成長率連續七個季度為負；水準於2001年第4季觸底，低於930億美元，「only 69 percent of the same figure one year earlier」；通訊設備占民間投資總額的比重亦自2000年的近7%降至2002年底的4.8%[S-630]（圖37）。

此一建置過度很可能源於其後證明過於樂觀的預期。 紐約聯邦準備銀行的一項研究發現，相對於通訊產業占經濟體的比重，該產業的投資對2000年熱潮的終結與2001年衰退的開始均具有「a disproportionate effect」，且其2000年的投資「likely reflected a severe misperception of future returns」[S-631]。依其模型的假設，電話產業投資率的上升隱含每年8.7%至10.1%的預期實質利潤成長率，該研究認為此對一個成熟產業而言極為異常[S-631]。2026年的理事

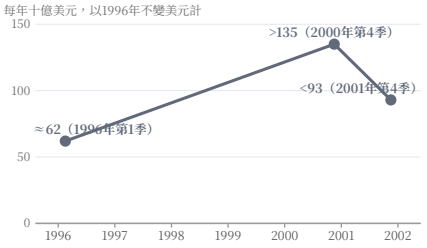
表37。 十四項指標用以研讀資本循環，自總體規模到地緣政治；其中數項尚未以可比形式公開

規模、價格與融資目前已可衡量；而尚未以可比形式公開的使用率，以及僅部分公開的報酬，才決定該等資本是否配置得宜。

構面	指標	重要性何在	本報告研讀之處	目前是否公開？
總體規模	智慧財產與設備投資占GDP的比重	經濟體的資本密集度	聯準會理事會依美國經濟分析局（BEA）資料所作的幕僚報告[S-595]	是：BEA的帳為季資料；該報告提供至2025年的年度數值與2026年第1季數值
對成長的貢獻	實質GDP成長的百分點	規模與成長兼具	聖路易聯邦準備銀行、國際貨幣基金（IMF）、國際清算銀行[S-615; S-622; S-623]	是，但定義互異
實體需求	AI伺服器、加速器（accelerator）、HBM位元數、已送電的百萬瓦數	數量	第2.3節；第3.2節；第2.2節	部分：無任何供應商陳述HBM位元數
價格	DRAM與HBM合約價格	供應商的議價能力（bargaining power）	TrendForce[S-007; S-035]	僅市場研究
組成	HBM與伺服器DRAM占比；加速器（accelerator）組成；伺服器支出對比出貨量	價值移轉	[S-008; S-007; S-592]	僅市場研究
使用率	已安裝加速器（accelerator）與已供電容量的使用情形	已安裝資本是否產生收益	—	未以可比的公開形式提供
變現	AI營收、使用量與已簽約需求	將產能轉化為需求	第2.2節；第8.4節	部分
資金來源	資本支出（capital expenditure）對比營運現金流	內部資金負擔	國際清算銀行[S-623]；申報文件	是，逐一公司
財務槓桿	債務、利息、私募信貸	金融傳導	[S-623; S-627]	部分
前瞻涵蓋度	未出貨訂單（backlog）、剩餘履約義務（RPO）、長期協議（long-term agreement）、預訂	需求可見度	第2.2節；第5.4節	是，但由各公司自行定義
報酬	分部利潤率；在可重建的情況下的資本報酬率	經濟回報	第5.1節	部分
生產力	勞動生產力與總要素生產力	總體經濟回報	[S-622; S-628]	尚在早期
集中度	供應商與客戶集中度	議價能力（bargaining power）與脆弱性	第5.6節；第7.5節	部分
地緣政治	被排除或取決於許可的營收	法律上可服務的需求	第7.1節	是，見於企業展望

本報告用以研讀AI資本循環的各項指標、各指標的重要性何在、本報告於何處研讀該指標，以及截至資料截止日的公開可得情形。

圖37。 美國通訊設備投資於1990年代後期成長逾一倍，其後一年內下滑近三分之一



美國通訊設備的實質民間固定投資，以1996年不變美元計，單位為每年十億美元，取里奇蒙聯邦準備銀行文章所指明的三個時點：1996年第1季約620億美元、2000年第4季逾1,350億美元、2001年第4季低於930億美元[S-630]。

範圍 為2003年可取得的美國經濟分析局（BEA）資料，以1996年不變美元計；該文章的近似用語（「approximately」、「over」、「under」）均保留於標籤之中。各時點之間的逐季路徑未予顯示。

會幕僚報告則引述熱潮期間所鋪設的「dark fiber」，作為熱潮結束後所留下的無生產性資本的一例[S-595]。

A 此一教訓具層級特定性。 1990年代同時產生了可衡量的生產力提升，以及集中於單一層級——通訊網路——的資本過剩，兩者並不相斥[S-629; S-630; S-631; S-595]。因此就AI而言，本報告逐層檢驗建置超前於需求的風險，涵蓋加速器（accelerator）、HBM與DRAM、先進封裝（advanced packaging）、資料中心廠體、發電容量預訂、電網併接、光學連結與專業AI雲端容量；絕不以某一層級判讀全系統過剩，亦不以另一層級判讀全系統稀缺。圖2所列的規劃時鐘即為其原因：依季時鐘下單的層級可能在一年內即建置過度，而依五至十年時鐘者則仍屬稀缺。

E.6 AI時代的循環：2023–2026年

當前循環必須在其仍在進行之際加以衡量。其規模取決於年內即經修正的資本計畫、範圍互異的各項預測，以及超前於實際建置進度的各項公告，因此並無任何單一數字足以描述它。本節將各項估計並列呈現而不予加總，繼而研讀今日的領先投資者與dot-com時代者之間的差異何在：其獲利、其評價，以及其如何為此輪建置取得資金。

F 當前循環的金額規模更大，部分層級的速度更快，且其規模仍在形成之中。 TrendForce估計，九家最大的雲端業者2026年的資本支出（capital expenditure）將超過8,867億美元，約較2025年高出90%，2027年則約為1.3兆美元[S-013]。國際能源總署（IEA）報告指出，五家大型科技公司的資本支出2025年已暴增至逾4,000億美元，並將於2026年再成長75%[S-086]。國際貨幣基金（IMF）引述外部估計，認為民間部門的AI投資「could top $2 trillion globally in 2026」[S-622]；一位聯準會理事並指出，已公告的資料中心計畫逾1.5兆美元，「only a small portion of which have been realized」[S-627]。上述各項估計採不同範圍，不予加總。

F 領先企業帶著既成獲利進入此輪循環。 聯準會副主席Philip Jefferson於2025年11月對照兩個時代時表示，「albeit not all」，但許多dot-com企業「typically had little to no realized earnings」，而與AI關聯最密切的企業則「generally have well-established and growing earnings streams」[S-632]。該等企業的本益比「have thus far remained well below the peak ratios of dot-com firms」；依某一衡量方式，約有50家公開交易企業聚焦AI，相對於1990年代後期高峰時逾1,000家掛牌的dot-com企業[S-632]。他並表示，私募資本市場「may obscure the extent of investor enthusiasm for financing AI firms」[S-632]。

F 融資來源正由現金流轉向債務。 國際清算銀行（Bank for International Settlements, BIS）經濟學家指出，領先的資訊科技企業「have historically financed much of their investments internally out of operating cash flows」，但自由現金流「have recently lagged capital expenditures in absolute amounts」，並預期投資需求的規模將使融資轉向債務，「with private credit playing a rapidly increasing role」[S-623]。對AI相關產業的私募信貸放款已自近乎於零成長至逾2,000億美元，約占未償私募信貸放款的8%；依該公報的外推，至2030年可能達到約3,000億至6,000億美元[S-623]。聯準會理事Lisa Cook於2026年5月指出，許多超大規模雲端業者（hyperscaler）「have executed large investment-grade bond deals in recent months」，而規模較小的開發商則透過私募與資產擔保債務募資[S-627]。

A 兩個時期大體上都少有債務依賴；差異在於此次建置目前所需的外部資金規模，以及這些資金落在何處。 Jefferson指出，「both dot-com previously and AI firms recently exhibited limited reliance on debt financing for the most part」[S-632]。目前正在改變的是建置所需外部資金的規模：來自最大買方的債券、提供給較小開發商的私募信貸、客戶預付款，以及落在企業自身資本支出（capital expenditure）之外的客戶提供硬體與營業租賃（第E.10節）。財務槓桿升高，使各層級的經濟性對使用率與再融資成本更為敏感；在國際清算銀行（BIS）的工作論文中，專用且以債務融資取得的資本，在同步緊縮之際面對的是交易稀薄的轉售市場[S-633]；國際貨幣基金（IMF）則警告，一家企業的問題可能連鎖擴散至其他企業[S-622]。

F 官方展望同時將AI列為支撐因素與風險。 國際貨幣基金（IMF）7月展望預估2026年全球成長率為3.0%、2027年為3.4%，美國則分別為2.3%與2.2%，並指出「continued technology-related business investment and productivity strength」為支撐美國經濟活動的因素之一[S-593]。經濟合作暨發展組織（OECD）9月期中展望表示：「Strong AI-related activity has also bolstered investment, production and trade.」該展望並將「AI investment returns that fail to meet expectations」列為主要下行風險之一[S-634]。國際貨幣基金2026年年度報告警告，超大規模雲端業者（hyperscaler）與晶片製造商之間的循環融資，提高了一家企業的問題連鎖擴散至其他企業的風險[S-622]。

E.7 兩輪循環的共同之處

A 兩輪循環都必須先為通用技術（無論已確立或僅屬候選）投入資本，其後才可能獲得回報。 聯準會理事會幕僚報告將資訊科技與蒸汽機、鐵路及電力同列為通用技術，並以技術須體現於資本作為其模型的立論基礎：「without new capital being invested, new technology cannot increase productivity by itself」

[S-595]。該報告並指出，過去許多通用技術雖提升了生產力，卻伴隨「顯著的過度投資」與資本過剩[S-595]。兩輪循環中投資成長均快於GDP，從占比即可見其勢（圖34）；且兩者皆因單位成本下降而帶動更多使用：1990年代的電腦價格推升了電腦資本的累積[S-625]，本報告模型證據中的每token記憶體用量與每token成本亦然（第8.4節）。

A 兩輪循環在資本投入之際，報酬皆屬未知，且可衡量的報酬可能較晚才出現。 在聯準會理事會幕僚報告的模型中，過度投資的風險是不確定性下最適投資的一部分，因此過度投資「is not a mistake caused by irrational exuberance over the new technology」[S-595]。國際清算銀行（Bank for International Settlements, BIS）的一份工作論文，將當前的競逐模擬為一場「firms competing for a few dominant positions over-commit resources」的競賽[S-633]。而在生產力J曲線模型中，可衡量的生產力低估了通用技術的早期效益，因為它所需的互補性投資「are often intangible and poorly measured in the national accounts」[S-628]（第E.9節）。

E.8 結構上的差異何在

A 四項差異最為明顯，且各自改變了應監測的對象。 第一，領先買方已有既成獲利，且向來多以營運現金流支應其大部分投資，惟自由現金流目前已落後於資本支出，因此首要的檢驗項目是現金創造能力與資本支出的對比，而非評價[S-632; S-623]。第二，瓶頸屬於實體性質，且分散於記憶體、封裝、光學、電力與電網之間（圖3），各自依不同時鐘增添產能（圖2），即使其中一項元件趨緩，仍可能使供應商的議價能力（bargaining power）維持於高位。第三，能源是第一順位的投入要素：2025年資料中心用電量成長17%，而全部電力僅成長3%；國際能源總署（IEA）估計，AI帶動的成長推力可能使2035年全球能源需求水準提高1%至4%，視採用程度而定[S-086; S-005]。第四，出口規則決定哪些需求可以合法滿足（第7.1節）。

E.9 生產力、互補性資本與J曲線

對新技術的大量投資，一般預期會反映於生產力，而這項效益若未出現，往往被解讀為過度投資的證明。兩者的連結較為緩慢且較不直接：企業必須先建立該技術所需的互補性資本，例如新流程、新技能與軟體，而其中有相當部分屬於無形且衡量不佳。本節闡述此一論點、生產力J曲線，以及本報告用以區辨延後出現的回報與不會出現的回報所觀察的各項指標。

F 可衡量的生產力可能落後於通用技術的投資。 Brynjolfsson、Rock與Syverson以模型指出，互補性的無形投資會導致「an underestimation of productivity growth in the early years of a new GPT」，其後則導致高估，「when the benefits of intangible investments are harvested」[S-628]。他們就與電腦硬體及軟體相關的無形資產調整美國資料後發現，截至2017年底，總要素生產力（total factor productivity, TFP）高於官方統計15.9%[S-635]。在工作論文版本（2018年，2020年修訂）中，他們發現AI相關無形資本對可衡量生產力的影響「雖小但正在擴大」[S-628]。

A J曲線是一項關於時點的假說，而非承諾。 它說明了為何以某一年的資本支出對比同一年的生產力並不成熟；它並未顯示回報必將到來，或各層級都能達到其資金成本的報酬水準。本報告以下列驗證指標來研讀J曲線：每小時產出；軟體對成長的貢獻度——聖路易聯邦準備銀行估計2025年前九個月為0.35個百分點，2000年則為0.11個百分點，而研發大致持平[S-615]；企業採用情況；可歸因於已部署AI的營收或成本節省；每一項有效任務的推論成本；以及已安裝運算資源與已供電容量的使用率。國際貨幣基金（IMF）指出，美國生產力成長已加速，並「may in part reflect early impacts from the adoption of AI」[S-622]。

E.10 融資、財務槓桿與資金來源結構

資本支出（capital expenditure）只是同一批產能取得資金的多種方式之一。 現金資本支出除以營運現金流，可衡量企業將其現金創造能力重新投入的比例，但自有資料中心的企業與承租資料中心的企業之間、自行購入記憶體的供應商與以代理人身分為客戶採購的供應商之間，此一比率並不可比。因此本報告依資金管道分類每一筆已揭露金額：營運現金流、債券與股權、私募信貸與資產擔保債務、租賃、客戶預付款與客戶提供硬體，以及代理採購（表39）。

AI基礎設施的報酬應視為一道階梯，而非單一比率。 首先是使用率：已安裝的產能是否被使用？其次是營收轉換：使用是否產生已認列營收？再者是毛利：價格是否高於運算、記憶體與電力的增量成本？接著是現金轉換，即扣除營運資金與支援成本之後。再者是資本回收：在設備過時之前，累計現金是否已回收所投入的資本？最後是經濟報酬：系統成熟之後，報酬是否超過資金成本？本報告不公布跨公司的投入資本報酬率，因為租賃、客戶出資資產與代理採購使各公司的資本基礎並不相同。

IF 若轉向債務的趨勢持續，融資本身將成為一項需求變數。 國際清算銀行（BIS）的一份公報，將對AI企業的私募信貸外推至2030年約3,000億至6,000億美元[S-623]；國際貨幣基金（IMF）則警告，AI投資的回報「increasingly debt financed, could prove illusory」[S-622]。應觀察最大買方的資本支出（capital expenditure）與營運現金流的對比、超大規模雲端業者（hyperscaler）發行債券的信用價差、資料中心開發商的私募信貸條件，以及其租約中的承租人集中度。若最大買方持續以營運現金流支應資本支出的成長，且開發商的債務仍與投資等級承租人掛鉤，則此一研判（reading）將趨於薄弱。

E.11 至2035年的條件式路徑與預測期限

四條經濟路徑描述此輪循環可能如何收束，但不附機率。 這些路徑以經濟語彙重述第9.1節的基準路徑與各項替代路徑，依資本獲得驗證、移轉至另一處瓶頸、趨於正常化，或在某一層級留下過剩的各項條件（conditions）加以區分（表40）。其訊號依不同時鐘到來，從季度合約價格到跨年度的使用率與現金回收皆有（表41）。

F 有一項模型為過剩路徑給出了數字，但其身分是模型，而非觀察結果。 國際清算銀行（Bank for International Settlements, BIS）的工作論文《The AI investment race》，將AI企業陣營置於「贏家囊括多數」的競賽中建模；該模型以資產負債表與交易資料校準，指出「over-investment of around 1.5 times the efficient level, rising to around three times where demand is less elastic」，並認為債務與循環持股既推升了榮景，也加深了隨後的衰退[S-633]。其完全均衡下的過度投資倍數為1.40，作者並表示，這些主要的失真幅度「should be read as floors rather than point estimates」[S-633]。本報告僅以此結果來框定過剩路徑；它並非對當前過剩產能的估計。

F 透明的展望會公布其敏感度，而非僅給出一個中心數字。 勞倫斯柏克萊國家實驗室（Lawrence Berkeley National Laboratory, LBNL）的美國資料中心用電模型，在其參考情境下將2030年用電量定為649 TWh，相當於美國電力的11.8%，並落在521至843 TWh的綜合不確定性區間內[S-442]。單一變動對結果的影響幅度差異甚大：資訊科技設備安裝量減少使其下降10.8%至578 TWh；ASIC增加使其上升2.4%至664 TWh；AI晶片壽命縮短使其下降9.1%；AI伺服器閒置功耗與使用率提高則使其上升20.6%至782 TWh（百分比依LBNL表2所印數值，於四捨五入前計算）[S-442]（圖38）。在LBNL的各項情境中，AI伺服器閒置功耗與使用率上升的假設，對2030年總量的影響大於安裝量縮減的假設。

IF 在稀缺向外移轉且資金供應跟上的情況下，此輪循環仍處於限制條件移轉路徑。 此一研判（reading）成立的前提是：記憶體價格至2027年維持堅挺[S-007]、封裝產能持續處於配額狀態（圖3）、送電啟用仍是新場址的進度關鍵（第7.2節），且最大買方一如過往，持續以營運現金流支應其大部分資本支出（capital expenditure）[S-623]。應觀察美國經濟分析局（BEA）後續發布的智慧財產與設備投資占GDP比重、資本支出與AI基礎設施預測的下一批版次，以及資料中心開發商的債券與私募信貸條件。若預測版次轉為向下，而已簽約產能仍持續上升，或使用率揭露顯示已安裝的加速器大規模閒置，則此一研判不成立。

E.12 圖表資料、來源順位與重建

經濟類來源依固定的優先順位閱讀。 首先是國家統計，即美國經濟分析局（BEA）、人口普查局與勞工統計局；其次是聯準會理事會與各聯邦準備銀行的研究；再者是國際清算銀行（Bank for International Settlements, BIS）、國際貨幣基金（IMF）與經濟合作暨發展組織（OECD）的分析；接著是經查核或經會計師核閱的公司申報文件與財報資料；再者是市場研究，用於公開統計未提供的數量；機制方面則參考學術論文。證券的市場價格，例如半導體指數股票型基金（ETF），僅能顯示投資人如何為此輪循環定價，絕不能顯示實體需求，本報告亦未納入任何此類圖表。幕僚報告與演講依其自身效力閱讀：上述數項來源均載明其觀點僅為作者個人意見。

本附錄中的每一項數值均可依紀錄重建。 每一項入圖數值，包括每一項推導橋接值與每一項百分點變動，均連同其類別、單位與來源編號收錄於本報告的圖表資料集；每一則引述亦收錄於來源查核登錄表（附錄D）中該來源的卡片。推導值於出現處載明其計算公式。GDP占比、貢獻表與美國經濟分析局（BEA）數列，均依其發布機構於2026年所印數值呈現，並可能經修正；國民所得帳的後續版次，既可能改動1990年代後期的數值，也可能改動近期的數值。

本章來源

每一編號均連結至附錄D中該來源的卡片，載有原始網址與該來源的原文。

企業. S-010 NVIDIA (26 Aug 2026); S-044 Oracle (10 Jun 2026); S-045 Microsoft (29 Apr 2026); S-063 Wiwynn (7 Aug 2026); S-075 Microsoft (29 Jul 2026); S-483 Wiwynn (7 Aug 2026)

標準組織與聯盟. S-004 OCP (1 Dec 2025)

政府與監管機關. S-005 IEA (16 Apr 2026); S-086 IEA (16 Apr 2026); S-616 U.S. Bureau of Economic Analysis (9 Apr 2026); S-617 U.S. Bureau of Economic Analysis (13 Mar 2026); S-624 U.S. Bureau of Economic Analysis (9 Apr 2026); S-627 Board of Governors of the Federal Reserve System (27 May 2026); S-632 Board of Governors of the Federal Reserve System (21 Nov 2025)

研究機構. S-442 LBNL (Jun 2026); S-593 IMF (8 Jul 2026); S-595 Board of Governors of the Federal Reserve System (6 Jul 2026); S-615 Federal Reserve Bank of St. Louis (12 Jan 2026); S-622 International Monetary Fund (23 Sep 2026); S-623 Bank for International Settlements (7 Jan 2026); S-625 Board of Governors of the Federal Reserve System (7 Mar 2000); S-626 U.S. Bureau of Economic Analysis (Apr 2002); S-628 National Bureau of Economic Research (Oct 2018); S-629 Board of Governors of the Federal Reserve System (Apr 2000); S-630 Federal Reserve Bank of Richmond (2003); S-631 Federal Reserve Bank of New York (May 2004); S-633 Bank for International Settlements (14 Jul 2026); S-634 OECD (23 Sep 2026); S-635 American Economic Association (Jan 2021)

市場研究. S-006 TrendForce (17 Aug 2026); S-007 TrendForce (25 Aug 2026); S-008 TrendForce (20 Jan 2026); S-013 TrendForce (3 Aug 2026); S-035 TrendForce (1 Jun 2026); S-038 TrendForce (15 Apr 2026); S-117 TrendForce (2 Sep 2025); S-118 TrendForce (7 Sep 2026); S-120 TrendForce (26 Nov 2025); S-121 TrendForce (26 Feb 2026); S-591 TrendForce (21 Aug 2025); S-592 IDC (28 Jul 2026); S-597 TrendForce (25 Feb 2026); S-598 TrendForce (6 May 2026); S-599 Gartner (15 Jan 2026); S-600 Gartner (19 May 2026); S-601 Gartner (16 Sep 2026); S-618 TrendForce (3 Jul 2026); S-619 TrendForce (17 Oct 2023); S-620 TrendForce (12 Feb 2025); S-621 TrendForce (2 Jul 2025)

表38。 兩輪循環在投資占GDP比重上相似，在組成、獲利、融資、能源與地緣政治上則有差異

投資占GDP比重相近，並不代表結果相同；以下各項差異才是應觀察的重點。

構面	1990年代後期資訊科技與網路循環	當前AI基礎設施循環	研判
技術	通用技術：資訊科技[S-595]	候選的通用技術，需要互補性投資[S-628]	兩者在獲得回報前都需要互補性資本
投資強度	智慧財產與設備投資於2000年達到美國GDP的11.49%高峰[S-595]	2026年第1季為11.33%[S-595]	規模已接近2000年的高點
對成長的貢獻	2000年為0.81個百分點，占實質GDP成長的28%[S-615]	2025年前九個月為0.97個百分點，占成長的39%（不計資料中心則為0.90個百分點與36%）[S-615]	至少相當，且更由軟體主導
實質成長：資訊科技設備與軟體	1995年至2000年每一年均介於17.4%與23.4%之間[S-617]	2023年、2024年與2025年分別為2.3%、6.2%與15.7%[S-617]	目前這一波的持續期間仍較短
建置過度的層級	通訊設備：1996年至2000年成長逾一倍，隨後一年內下滑31%（2001年第4季為2000年第4季的69%）[S-630]	記憶體、封裝、電力與電網依不同時鐘擴張；所閱讀的來源在資料截止時均報告各項皆有稀缺情形（圖3）	風險分散於更多層級
生產力	資訊科技使用與電腦生產：約占1990年代後期生產力躍升的三分之二[S-629]	美國生產力成長「may in part reflect early impacts from the adoption of AI」[S-622]	有待檢驗，不應逕予假定
上市企業獲利	許多dot-com企業幾無已實現獲利[S-632]	領先AI企業已有既成且持續成長的獲利[S-632]	起步時的現金基礎較為穩健
上市企業	高峰時有逾1,000家dot-com企業掛牌[S-632]	依某一以ETF為基礎的衡量方式，約50家聚焦AI；私募市場則持續擴大[S-632]	廣度有一部分落在公開市場之外
融資	大體上對債務的依賴有限；在電信業，大量首次公開發行（IPO）隨後轉為大量破產申請[S-632; S-630]	向來以現金支應，目前轉向債券、私募信貸與資產擔保債務[S-623; S-627]	財務槓桿敏感度正在上升
資產年限	網路設備與光纖，部分閒置未用（「暗光纖」，dark fiber）[S-595]	Microsoft截至2026年3月當季的資本支出（capex）約三分之二投入短年限資產，主要為GPU與CPU[S-045]；Microsoft並將其資料中心的使用年限自15年延長至25年[S-075]	殘值風險依層級而異
能源	—	2025年全球資料中心用電量成長17%；國際能源總署（IEA）基準情境下2030年約為950 TWh[S-086; S-005]	能源是第一順位的投入要素
供給集中度	—	HBM製造商僅三家；尖端晶圓代工與封裝高度集中（表25；表26）	產能配額與驗證（qualification）本身即具價值
地緣政治	—	地緣政治限制已進入財測指引：NVIDIA的展望假設來自中國的Data Center運算營收為零[S-010]（第7.1節）	可服務的需求可能與技術上的需求出現落差

以1990年代後期資訊科技與網路循環對照當前AI基礎設施循環，逐一構面比較，並附本報告所採的研判（reading）；來源如各欄所引註。破折號表示本報告所閱讀的1990年代來源未記錄該項。

表39。 六種資金管道使同一批產能出現在帳上的不同位置

當營業租賃或客戶提供硬體承擔了其中一部分時，企業的資本支出（capital expenditure）會低估其實際新增的產能；而當供應商以代理人身分購入零組件時，其營收也會低估實際出貨的規模。

管道	紀錄內容	出現位置	後續觀察重點
營運現金流	領先的資訊科技企業向來多以內部資金支應其大部分投資；自由現金流目前在絕對金額上已落後於資本支出（capex）[S-623]。Oracle 2026財年的自由現金流為負237億美元[S-044]	現金流量表中的資本支出	資本支出與營運現金流的對比
債券與股權	超大規模雲端業者（hyperscaler）的大規模投資等級債券發行[S-627]；Oracle於2026財年募得430億美元債務與50億美元股權，並預期2027財年將再募得約400億美元[S-044]	債務與股權；利息成本	發行量、信用價差、到期結構
私募信貸與資產擔保債務	對AI相關產業的私募信貸放款逾2,000億美元；規模較小的開發商向私募債務基金與資產擔保市場借款[S-623; S-627]	多記於開發商，落在承租人資產負債表之外	貸款條件與承租人集中度
租賃	Microsoft稱：「Finance leases are included in capital expenditures while operating leases are not」[S-075]	使用權資產；僅融資租賃計入資本支出	租賃承諾及其到期結構
客戶預付款與客戶提供硬體	Oracle稱：「The prepaid and customer supplied hardware portions of our large AI contracts now total $75 billion」[S-044]	新聞稿未載明；客戶提供硬體不計入供應者的資本支出，預付款則降低其淨資金需求	預付款與資本支出的對比
代理採購	Wiwynn以代理人身分為特定客戶採購記憶體，因其在移轉前並未取得該等物料的控制權，故該等金額不計入其營收與銷貨成本[S-063; S-483]	落在營收與銷貨成本之外	實體部署規模與帳列營收的對比

AI基礎設施的資金來源，每一管道均附紀錄中的一項實例、該金額在企業帳上出現的位置，以及後續觀察重點。

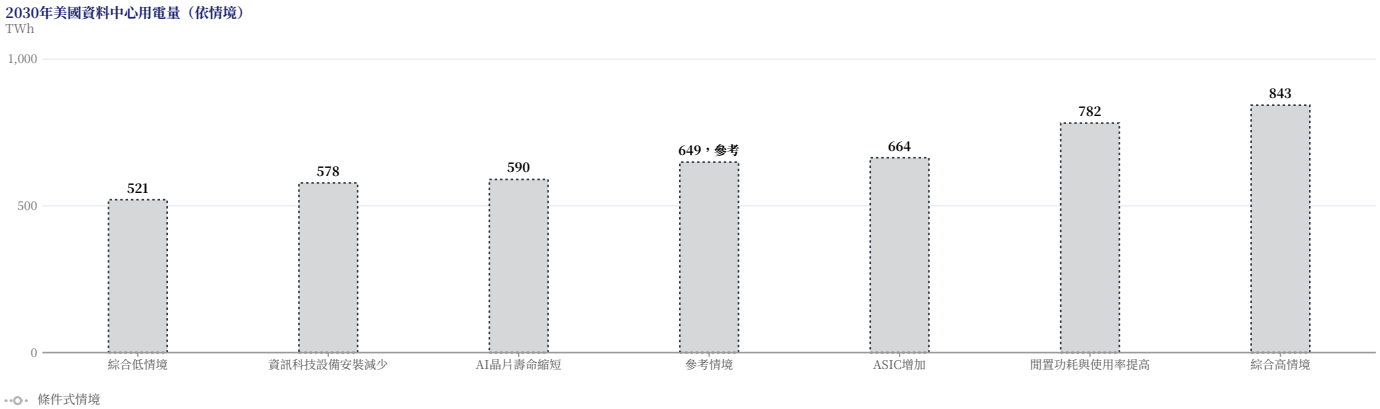
表40。 四條經濟路徑指明此輪建置獲得驗證、移轉、趨於正常化或留下過剩的各項條件

問題不在於AI是否可行，而在於各層級的建置步調是否為使用情況與現金流所能吸納。

路徑	條件	研判	最接近的展望	應觀察的訊號
生產性資本深化	使用率維持高位；AI營收與成本節省隨產能一同上升；每項任務的推論成本下降；生產力效益擴散至科技業以外；債務以營運現金流償付	高額投資由產出與需求予以驗證	基準路徑（第9.1節）	軟體與研發對成長的貢獻度[S-615]
限制條件移轉	運算與記憶體供給改善；瓶頸移往電力、散熱、光學、場址送電啟用或融資	支出維持高位，議價能力（bargaining power）則沿供應層次移動	基準路徑；替代路徑二	渦輪機未出貨訂單與併網時程（第7.2節）
產能正常化	供給追上需求；零組件價格與交期下降；報酬壓縮，但產能仍具生產性	稀缺溢價消失，但技術本身並未失敗	替代路徑一	DRAM合約價格下跌；預測版次向下修正（圖33）
資本過剩	已簽約產能持續超出可變現需求；使用率走弱；取消案例增加；價格下跌快於成本；融資成本上升；專用資產的轉售價值下滑[S-633]	某一層級的建置超前於可回收的需求	替代路徑三	取消案例、租約重新議定與信用價差

本報告的四條條件式經濟路徑、各自所需的條件（conditions）、其研判（reading）、第9章中最接近的展望，以及應觀察的訊號。任何路徑均不附機率。

圖38。 在LBNL的各項情境中，AI伺服器的使用方式對2030年美國資料中心用電量的影響，大於增加ASIC



LBNL依模型情境（scenario）推估的2030年美國資料中心用電量，單位為TWh，取自其2025年更新版（2026年6月）表2；參考情境為649 TWh[S-442]。
範圍 係單一模型針對美國所設的情境，並非附機率的預測；綜合情境結合了數項變動。與國際能源總署（IEA）的全球數列不可比。

表41。 每一預測期限各以不同變數研讀：先看執行與價格，最後看結構
預測期限（horizon）愈長，單一產品藍圖愈不重要，經濟結構則愈為關鍵。

預測期限	研讀依據	訊號
一年	執行與定價	HBM與伺服器DRAM合約價格[S-007]；AI伺服器與一般伺服器的出貨量與支出[S-013; S-592]；預測版次（圖33）；送電啟用延遲（第7.2節）；債券發行量[S-627]與信用價差；企業財測指引中的出口假設（第7.1節）
三年	供給反應與使用率	已完成驗證而非僅經宣布的記憶體與封裝產能（圖16）；已完成送電啟用的電力與電網專案；2025年至2027年建置批次的使用率；稀缺供應商的利潤率；債務償付與租賃負擔
五年	資本報酬與架構	2025年至2027年建置的現金回收；加速器與HBM的汰換週期；記憶體分層與資源池化（第8.3節）；每token記憶體用量（第8.4節）；科技業以外的生產力
2035年	結構，而非點預測	AI工作負載的能源強度；基礎設施的資本強度；運算、記憶體與電力之間的平衡；地理分布；多供應商介面；HBM、尖端邏輯、光學與電力設備上殘留的集中度

本報告在各預測期限（horizon）所研讀的變數，並附本報告中承載這些變數的訊號；來源如引註。

法律免責聲明

本報告係以公開規格、法規與企業申報文件、業績發布、企業與政府出版物、研究，以及市場研究資料為基礎。報告全篇依來源查核紀律標示並載明佐證來源引註，該紀律旨在支持追溯性、獨立審閱，以及對重大事實與關鍵解讀的來源層級查證。凡來源用語對關鍵解讀具重要性者，均一併納入底層紀錄中的相關原文摘錄，使讀者得以直接檢視原始措辭，並對照其來源脈絡評估該解讀。此一證據紀律將所引事實與Apex Standards自身的分析加以區分，使重大結論的依據透明可見，俾使讀者與客戶得以依底層紀錄檢驗關鍵解讀。前瞻性觀點僅於所載明的條件下提出，除經所引來源明確支持外，不給定機率、排名或承諾日期。本報告所載資訊係嚴格依「AS IS」（現狀）原則提供，僅供資訊與教育用途，不構成法律、投資、會計、安全認證或工程方面之建議。對於準確性、完整性、特定用途適用性或不侵權，均不作任何明示或默示之聲明或保證。報告中提及之第三方個人、組織、產品、標準或服務，僅供資訊參考之用，並不表示與Apex Standards存在關聯、獲其贊助，或構成任一方對另一方之背書。在適用法律允許之最大範圍內，Apex Standards對於錯誤、遺漏、對本資訊之信賴，或因使用本資訊所生之後果，概不承擔任何責任或賠償義務，且不負有於本報告發布後對其進行更新、修改或修訂之義務。

Legal disclaimer

This report is based on public specifications, regulatory and corporate filings, results releases, company and government publications, research, and market-research materials. Supporting source references are identified and specified throughout the report under a source-audit discipline designed to support traceability, independent review, and source-level verification of material facts and critical interpretations. Where source language is material to a critical interpretation, relevant verbatim excerpts from the underlying record are included so that readers can review the original wording directly and assess the interpretation against its source context. This evidentiary discipline distinguishes cited facts from Apex Standards' own analysis and makes the basis of material conclusions transparent, enabling readers and clients to test critical interpretations against the underlying record. Forward-looking views are presented only under stated conditions, without assigning probabilities, rankings, or committed dates unless expressly supported by the cited source. The information is provided strictly on an "AS IS" basis for informational and educational purposes and does not constitute legal, investment, accounting, safety-certification, or engineering advice. No representation or warranty, express or implied, is made as to accuracy, completeness, fitness for a particular purpose, or non-infringement. References to third-party individuals, organisations, products, standards, or services are provided solely for informational purposes and do not imply affiliation with, sponsorship by, or endorsement of or by Apex Standards. To the fullest extent permitted by applicable law, Apex Standards assumes no responsibility or liability for errors, omissions, reliance on the information, or consequences arising from its use, and undertakes no obligation to update, amend, or revise this report after publication.

關於Apex Standards

Apex Standards是一家獨立的技術、標準、市場與智慧財產權情報機構，專注於關鍵及新興技術。我們提供先進的分析軟體平台、研究工具、經來源查核的分析，以及策略顧問服務，涵蓋關鍵通訊標準、3GPP、IEEE、IETF、ITU、Open RAN、多媒體編解碼器、JEDEC、半導體、光子學、AI基礎設施、OCP、CXL、HBM，以及相關技術與供應鏈生態系。

我們為政府機關、企業、研究機構、技術長與技術團隊、投資人及企業發展主管，提供具決策品質且可供稽核的情報，用於策略定位、市場與供應鏈分析、投資與併購盡職調查，以及技術評估。我們協助客戶減少策略盲點、揭露被忽略的風險，並更早識別高價值機會。

網站：www.apexstandards.com
電子郵件：support@apexstandards.com

About APEX STANDARDS

Apex Standards is an independent technology, standards, market and IP intelligence organization focused on critical and emerging technologies. We provide advanced analytical software platforms, research tools, source-audited analysis, and strategic advisory across critical communication standards, 3GPP, IEEE, IETF, ITU, Open RAN, multimedia codecs, JEDEC, semiconductors, photonics, AI infrastructure, OCP, CXL, HBM, and related technology and supply-chain ecosystems.

We support governments, corporations, research institutions, CTO and technical teams, investors, and corporate-development leaders with decision-grade, auditable intelligence for strategic positioning, market and supply-chain analysis, investment and M&A diligence, and technology assessment. We help clients reduce strategic blind spots, uncover overlooked risks, and identify high-value opportunities earlier.

Website: www.apexstandards.com
Email: support@apexstandards.com