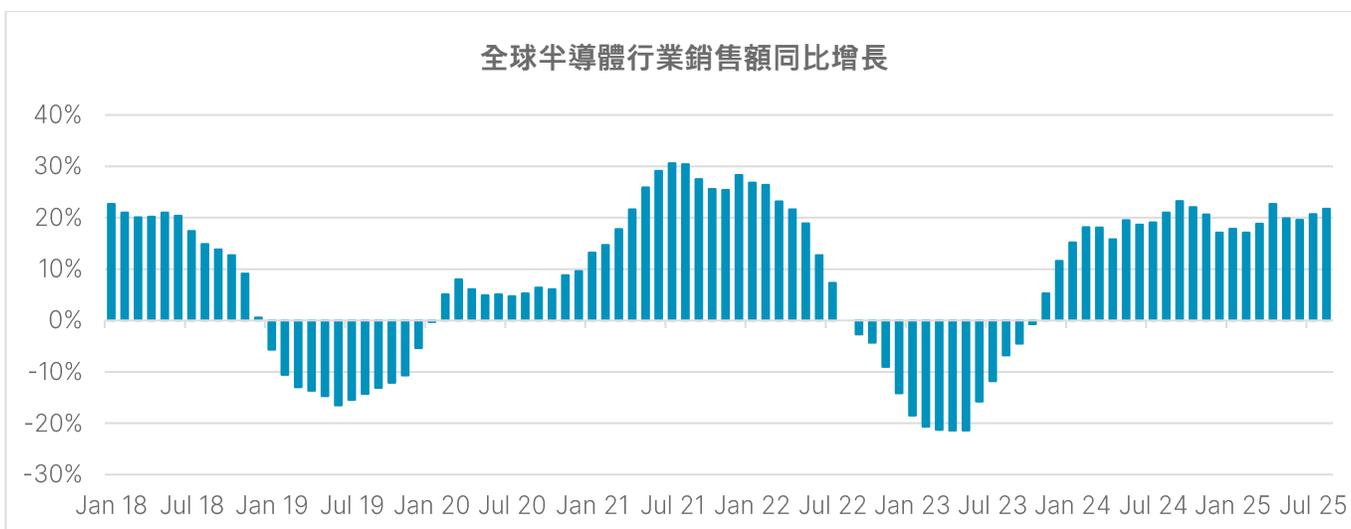


# 半導體如何塑造新一代 AI 創新格局

亞太區指數研究部主管蔡朗賢 David Tsoi, CFA, CAIA, FRM, CESGA, CAMS

人工智能 (AI) 依然是當今最具革命性的技術，並由一眾半導體公司主導和引領突破性進展。在邏輯和記憶晶片需求的帶動下，全球半導體市場預計今年將延續 2024 年的強勁復甦，全年增長 18%，總收入達到 8,000 億美元，其中美洲及亞太區預期將會錄得最高增長。數據中心的擴張繼續帶動顯著增長，專注於 AI 和半導體創新的公司尤其受惠。



資料來源：美國半導體行業協會。截至 2025 年 10 月 3 日。

## AI 維持強勁增長

儘管面臨關稅上升及經濟阻力增加等挑戰，隨著 AI 驅動的變現機會逐步浮現，超大規模雲端服務商 (hyperscalers) 的資本支出繼續維持上升趨勢。2025 年第二季度，全球數據中心資本支出按年飆升 43%。微軟、亞馬遜和 Google 均表示，AI 工作負載需求持續超出現有基礎設施的承載能力，預計全年將繼續擴大新增容量以應對需求增長。OpenAI 擬斥資約 4,000 億美元，聯同甲骨文及軟銀集團，在美國興建五個全新數據中心據點。這些新據點最終將具備 7 吉瓦 (gigawatts) 的電力容量，顯著提升 OpenAI 的運算能力，以支援其各項服務。Meta 亦正在建設多個多吉瓦級數據中心集群，以加速實現其 AI 目標。其中首個數據中心預計將於明年啓用，屆時其 2026 年的資本支出將進一步上升。Meta 將 AI 作為其廣告策略重心，並計劃在明年年底前讓品牌客戶能夠全面使用 AI 工具來設計和制定具針對性的廣告活動。根據客戶的預算，這些新工具將可生成整個廣告，包括圖片、影片和文字，並發送給目標受眾。

過往，AI 的需求主要集中於訓練工作負載，尤其是前沿 AI 模型。在繼續投入資源建設更龐大 AI 模型的同時，科技巨頭亦將轉為重新分配更多的投資於推理（inference）領域。推理是指已完成訓練的 AI 模型處理新數據，從而產生灼見、進行預測或支援決策的階段。訓練模型基本上是一次性支出，對模型進行指示（推理）生成詞元（token），而每個詞元皆附帶成本。在 Google I/O 2025 大會主題演講中，Alphabet 行政總裁桑德爾·皮查伊透露，2025 年 4 月，Alphabet 所有產品和應用程序編程接口（API）處理高達 480 萬億個詞元，較去年同期激增 50 倍。詞元量快速攀升反映 AI 模型的應用日益普及，預示對運算能力將有更大需求，並帶動對晶片的需求增長。

## AI 推理時代

隨著新推理模型面世，投資也加速轉向推理領域。傳統的 AI 模型雖然能快速回應，並擅於識別模式，卻往往無法理解更廣泛的背景，且難以處理複雜的推理任務。推理模型旨在將複雜的問題分解為更小、更易處理的步驟，再透過明確的邏輯推理解決問題。此類模型能夠展示其推理過程，並遵循結構更加清晰的思考流程，因而在處理用戶查詢時需要更長的運算時間。這些模型在推理過程中需要使用明顯較多的運算資源進行推理，以處理複雜的問題。從基本模式識別演進至結構化推理對 AI 的發展至關重要，可釋放 AI 有效應對各種現實世界複雜挑戰的潛能。隨著 AI 應用範圍快速擴大，對推理能力的需求亦將相應激增。

## AI 代理的崛起浪潮

AI 代理旨在為機構的運作模式帶來革命性的轉變，令生產力和營運效率實現突破性增長。這些智能系統在設計上能夠透過理解目標、制定決策並採取行動來獨立執行任務，從而達成預設目標。雖然人類負責定義期望結果，但 AI 代理能自主選擇達成有關目標所需的最佳行動。AI 代理的應用範圍廣泛，從支援學術研究、簡化線上購物流程，以至規劃度假行程皆可。客戶服務、銷售與營銷以及資訊科技及網路安全是未來六個月內最常部署或計劃引入 AI 代理的三大業務領域。隨著企業逐步將 AI 代理整合至各種營運場景，對運算基礎設施的需求正在急劇上升。

## 客製化 AI 晶片的崛起

超大規模雲端服務商日益專注於特殊應用晶片（ASIC）基礎設施，以滿足激增的 AI 需求。ASIC 專為特定工作負載打造，不但在執行相關任務時的效率遠高於圖形處理器（GPU），更可大幅降低成本。儘管開發 ASIC 初期涉及龐大的投資金額，但一旦前期成本被吸收，預計應用這些晶片運行生成式 AI 工作負載的長期成本將會下降。例如，Google 於 2025 年 4 月推出專為推理工作負載而設的第七代 Tensor Processing Unit（TPU）「Ironwood」。過去 Google 的自家 TPU 僅限於內部使用，如今公司正擴大對外開放，以加快雲端業務的增長。邁威爾科技（Marvell Technology）預計客製化運算裝置市場規模將於 2028 年增長至 554 億美元，規模超過 2023 年的八倍。

## AI 帶動高頻寬記憶體（HBM）技術的強勁需求

HBM 為尖端記憶體技術，旨在提升數據存取速度，同時降低耗電量，對 AI 運算效能至關重要。預計 HBM 在動態隨機存取記憶體（DRAM）領域所佔市場份額，將從 2024 年的 18% 躍升至 2030 年超過 50%。從下一代 HBM4 開始，基底晶片（base die）將採用邏輯製程（logic processes）生產，減低耗電量之餘，亦可根據客戶需求客製化。受 AI 訓練和推理工作負載令運算需求日益增長所帶動，HBM 市場前景依然強勁。作為輝達主要 HBM 供應商，SK 海力士於 2025 年第二季佔全球 HBM 出貨量的 62%，預計全球 HBM 市場將以每年 30% 的增速持續增長至 2030 年。

---

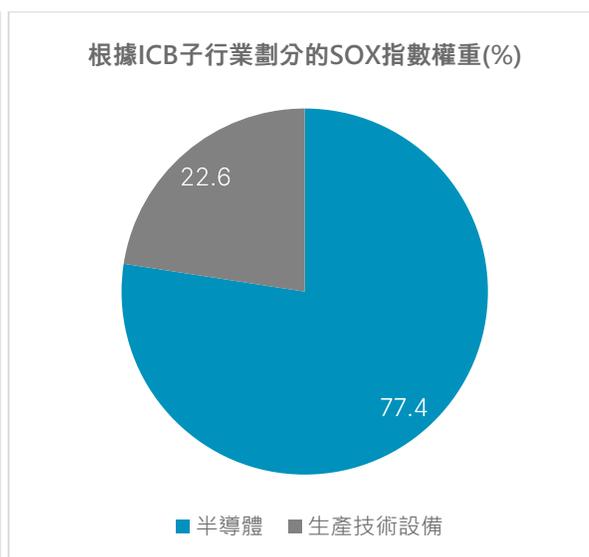
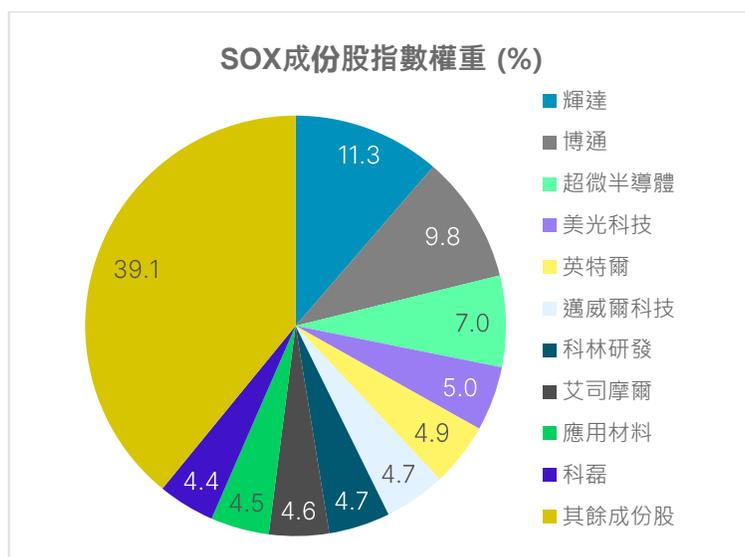
## 費城半導體指數 (SOX™) — 半導體行業的領先指標

納斯達克的費城半導體指數 (SOX) 涵蓋前 30 大主要從事半導體設計、分銷、製造和銷售的美國上市公司股票及美國存託憑證 (ADR)。過去三年，SOX 的總回報率高達 185%，較紐約證券交易所半導體指數高出 20 個百分點，也是標普半導體精選行業指數回報的 1.6 倍。



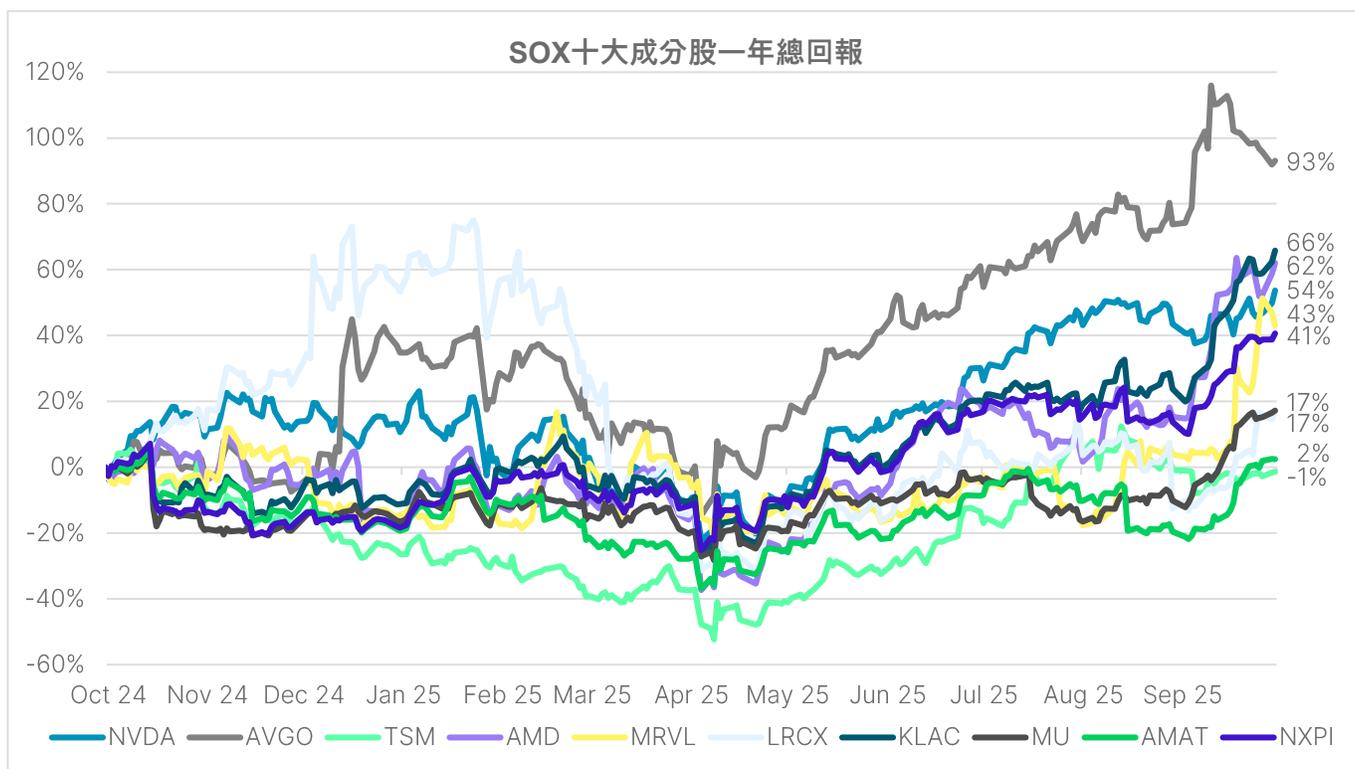
資料來源：納斯達克全球指數、彭博。截至 2025 年 9 月 30 日。

SOX 為修正後的市值加權指數，市值排名最高的三隻成份股的權重上限分別為 12%、10%和 8%，其餘成份股在每季再平衡時各自的權重上限為 4%。有關完整指數編製方案，請瀏覽我們的[網站](#)。截至 2025 年 9 月底，十大成份股佔指數權重的 60.9%。根據行業分類基準 (ICB) 分類系統，指數權重的 77.4%為半導體子行業，其餘則屬於生產技術設備子行業。



資料來源：納斯達克全球指數、慧甚。截至 2025 年 9 月 30 日。

過去 12 個月，十大持股中有九檔錄得正總回報。十大成份股的一年總回報平均達 39%。在同一行業內，個別股份表現存在重大差異。十大成份股中表現最佳與最差的股份（博通和超微半導體）之間的一年總回報差距高達 95 個百分點。這說明了即使是投資於單一產業或主題，分散投資仍然重要。



資料來源：納斯達克全球指數、慧甚。截至 2025 年 9 月 30 日。

### 輝達 ( 權重：11.3% )

作為 SOX 的最大成份股，輝達在十大成份股中表現排名第四，在過去 12 個月錄得 54% 的增幅。繼於 2023 年 5 月躋身「萬億美元市值俱樂部」後，該公司在 2025 年 7 月成為史上首家市值達到 4 萬億美元的上市公司。輝達仍然保持明顯的技術優勢，市場對 AI 推論的需求飆升，加上已達到規模經濟效應，帶動 Blackwell 出貨量加速增長。該公司認為，除 AI 外，機械人技術將會是其具有最大增長潛力的市場，而自動駕駛汽車將為首個主要的商業應用範圍。該公司最近推出的全新輝達 Omniverse 函式庫及輝達 Cosmos 世界基礎模型，讓開發者結合 AI 推論與可擴展、符合物理定律的精準模擬技術，生產新一代的機械人及自動駕駛汽車。輝達計劃向 OpenAI 投資高達 1,000 億美元，加速擴建容量至少達 10 吉瓦的數據中心，以及向英特爾投資 50 億美元，以共同開發新一代數據中心及個人運算產品。

### 博通 ( 權重：9.8% )

作為 SOX 第二大成份股，博通一躍成為十大指數成份股中表現最佳的一員，截至 2025 年 9 月的一年總回報率高達 93%。該公司繼續主導 AI ASIC 及 AI 網絡半導體市場，為全球七大超大規模雲端服務商提供客製化晶片產品，包括為 Google、Meta 及字節跳動提供加速器。Open AI 亦承諾向博通採購客製化晶片及網路元件，預計於 2026 年下半年開始建造容量達到 10 吉瓦的 AI 數據中心，並預期於 2029 年年底前竣工。此外，博通在 2023 年收購威睿 ( VMware ) 後，擁有龐大的企業軟體業務。其基礎設施軟體部門上一季度的經營利潤率為 77%，高於去年同期的 67%。

## 美光科技 ( 權重 : 5.0% )

作為第四大 SOX 成分股，美光科技 2025 年第二季的 DRAM 營收市佔率為 22%，過去 12 個月的總回報為 62%。該公司預期 11 月季度毛利率將達 51.5%，自 2019 年以來首度突破 50% 的門檻。受 HBM3E 產品量產帶動，美光科技的 HBM 年化營收接近 80 億美元。隨著 HBM4 產品量產及相關客戶群增加至六家，該公司的強勢將延續至明年。在智慧型手機領域，具備 AI 功能的裝置滲透率持續提升仍然是推動行動裝置 DRAM 容量增加的關鍵因素。

## 台積電 ( 權重 : 4.3% )

台積電是全球最大的半導體合約製造商，過去 12 個月的總回報為 63%。由 AI 驅動的需求仍然是台積電的主要增長動力，其市場主導地位亦為其帶來強大的議價能力，該公司來自高性能運算業務的淨收入由去年的 51% 升至今年的 57%。7 奈米或更小尺寸的高階晶片佔該公司 2025 年第三季度晶圓總收入的 74%。儘管美國總統近期宣佈對進口半導體加徵關稅，台灣方面已確認，鑒於台積電在美國擁有龐大的生產投資，故此已獲新徵收關稅的豁免。

## 總結

為支援日趨複雜的 AI 模型架構，市場對進階訓練能力的需求持續上升，同時，AI 推理已儼然成為關鍵的增長動力。詞元數量急升反映 AI 模型的應用正持續擴大。AI 代理的普及亦勢必改變多個行業的格局，同時大幅推高運算的需求。此動態發展突顯了半導體產業在推動下一波 AI 創新浪潮中擔當重要的角色。

納斯達克的費城半導體指數 ( SOX ) 在截至 2025 年 9 月底的三年總回報為 185%。追蹤 SOX 的基金包括國泰美國費城半導體基金 ( 台灣 : 00830 )、Invesco 費城半導體 ETF ( 納斯達克 : SOXQ )、未來資產 TIGER 美國費城半導體納斯達克 ETF ( 韓國 : 381180 )、Global X 半導體 ETF ( 日本 : 2243 ) 及 Yurie 費城半導體指數基金 ( 韓國 : 7D01596 )。未來資產 TIGER 美國費城半導體槓杆 ETF ( 韓國 : 423920 ) 以兩倍槓杆追蹤 SOX。

免責聲明：

Nasdaq®、PHLX Semiconductor™、SOX™ 及 Nasdaq-100® 為 Nasdaq, Inc. 的註冊商標。上文所載資料僅供參考及教育用途，不應詮釋為針對特定證券或整體投資策略的投資建議。Nasdaq, Inc. 及其任何關聯公司概不對買賣任何證券作出任何建議，亦不對任何公司的財務狀況作出任何陳述。有關納斯達克上市公司或納斯達克專有指數的聲明並不保證未來的表現。實際結果可能與所明示或暗示的結果存在重大差異。過去表現並不代表未來的結果。投資者在投資前應自行作出盡職調查並仔細評估公司。強烈建議投資者徵詢證券專業人士的建議。任何由於翻譯造成之差異或分歧均不具約束力及對合規或執法也無法律效力。若對本譯本所提供資訊有任何疑問，請參考英文版本。

© 2025. Nasdaq, Inc. 保留所有權利。

---