

淺談「中美AI發展的近況與差異」

曹豫

一、背景

自 2023 年 OpenAI 發明 ChatGPT 之後，中美率先進入智算時代，生成式人工智能的強勢崛起對算力產生了巨大的需求。

中國國務院於 2020 年明確提出「算力服務國家治理、經濟發展和民生改善」的戰略目標。

筆者根據自身工作經歷與行業背景，總結自 2023 年以來，中國數據中心從「分佈式」發展快速往「集羣式」數據中心發展。

其主要原因係在人工智能快速發展的背景下（平均每 6-7 個月更新一次算法），「集羣式」數據中心在財務、運維管理、能耗效率、市場適應性維度等都具有優越性。

「集羣式」多機房的協同佈局與管理將顯著降低數據中心對電力、冷卻等獨立配套設施的需求和運營成本，此與國內“雙碳”目標高度契合。

更重要的是 AI 的大模型訓練存在物理距離限制。採用異地聯合訓練服務器時，「當機」概率將呈現指數級攀升。這中間涉及訊號傳輸與系統穩定性等問題（因牽涉過多技術背景，不在此贅述）。

因此在 AI 大模型訓練時代，數據中心往「集群式」發展為必然之趨勢。

二、中美 AI 基礎設施投入比較

在 2024 年，美國幾大科技巨頭微軟、Google、Tesla、Meta、亞馬遜總投入至少約 3000 億美金，採購包含芯片和數據中心等基礎設施。

中國科技巨頭阿里、字節、百度、騰訊、華為保守估計投入約計 2000-3000 億人民幣，雙方差異七倍以上。

2025 年下半年開始，Nvidia、OpenAI 開始加速投資，每年的金額都在 1000 億美金級別。對比中國，最引人注意的即是阿里巴巴宣布未來五年總投入 3800 億人民幣。

但以筆者親身經驗，中國 AI 投資真正落地進展仍相較緩慢。

以用來衡量數據中心損益平衡的標準：每千瓦租賃金額

從 2020 年高點 1000 人民幣到目前約 300 人民幣。數據中心建設、折舊、運維等費用需確保每千瓦租賃價格至少在 260 人民幣以上。

截至 2025 年 Q3，全國各地機房租賃價格未見起色。

更令人擔憂的是，中國新建算力中心的空置率保守估計在 30-40% 以上，對比美國新建算力中心空置率 3-10%，雙方機房利用率存在巨大差距。

三、AI 投入差異可能原因

1. 中國 GPU 芯片不足

截至 2025 年，AI 大模型訓練當中，有 80% 需倚靠 Nvidia 的算力芯片。由於供給嚴重不足，影響國內 AI 大模型的投資進展。筆者判斷，這現象應該會在 2025 年七月後逐步緩解。主要係觀

察到中芯國際已突破 7nm 製程，以寒武紀和華為 910C 系列為首的 GPU 芯片應可以放量生產。

2. 算法架構差異

Deepseek 於 2024 年底發布依據 MOE 架構的算法模型，此模型極大的改變世界 AI 模型基礎，也體現出中美 AI 發展的側重點不同。

由於算力不足，且中國缺乏巨大開放的資本市場支持，因此中國式模型在效率上追求極致，優先關注成本節省，這間接影響 AI 基礎設施支出。

（註：MixtureofExperts 專家混合模型，是一種模組化深度學習架構，旨在解決大規模模型所面臨的巨大計算和儲存壓力）

3. AI 理念不同

美國追求「通用型 AI」，所謂「通用型 AI」是指一種能自我思考和進化的智能體。

目前中國追求的是更高效的智能工具。

這兩種思維差異導致雙方在算力投入上有著極大的出入。

以筆者觀點來看，雙方選擇並無對錯。中國雖受到算力資源限制，卻擁有完整工業體系，數據，充足的能源和電網，讓 AI 以工具化方式快速落地並滲透至各行各業是最優解。

美國嘗試選擇通用智能追求。

何謂通用智能追求？

我們先問自己一個問題，什麼是理解？

三歲小孩可以自動的認識自己的父母。但 AI 辨認人臉，需要將所有參數重新計算。一個人類大腦在工作時耗電功率大約在 20W。一個基礎 AI 算力集群大概是 200MW，兩者相差一千萬倍。

也就是說，現階段 AI 大模型基於 LLM 架構，以統計學的方式並不一定能通往真正的智能體。

（註：LargeLanguageModel 大型語言模型，一種利用大量數據進行預訓練的進階人工智慧模型，能夠理解、生成和處理人類語言）

小孩是否認識父母，可能只需要記得 30% 的臉部長相或是藉由聲音、氣味、行為模式等來進行分辨。

也就是說，我們對事物的理解是「分散式」架構，可以按照不同區塊的關鍵信息進行多重確認，不用對整體信息進行全量確認。這大大降低算力的消耗。

達到認知還有一個關鍵要素，即是「聯想能力」。

當 AI 回答一個未知的問題時，需要調用全量算力。它無法通過類似經驗來回答問題。

舉例：小明「打」了小美的手，小美痛得哭出來了。

當我們把動詞改為「敲打」時，AI 無法通過上下文，例如「手、痛、和哭」，來推測「敲打」的含義。這代表 AI 並沒有真正理解這段文字的物理意義。

當然，美國的另一方案「Token 座標化」或許可以解決部分問題，這可以留到下一次進行分享。

另一個關鍵爭論點，當前的 LLM 語言可能只是一種「縮小可能性」的算法架構。它追求在一個受限範圍內的最優解。

真正的智能體必須具備犯錯的自由，才能解決尚未有答案的問題，也就是自我進化能力。

這也代表，AI 在發展過程中，必須產生自我的喜好、給自己設定問題甚至拒絕回答問題，這是目前 LLM 模型追求最優解的架構下無法達成。

四、結論

筆者大膽預測，當前基於 LLM 語言所開發的 AI 大模型，可能與未來「真正」智能體發展路線並不相符。

筆者也研判，美國已經開始關注 LLM 語言的局限性，並著手進行下一輪大規模投入。目前中美雙方所有已投入的 AI 基礎建設並非無意義，因為此會促進人類對於下階段 AI 發展的思考。對下一個大語言模型的發展必有所助益。

曹豫

一個服務於著名互聯網企業的台灣青年

2025/10/09