

亞太區

後來居上的超越？還是永遠差一步？

從伊索寓言到芝諾悖論：人工智能競賽與領先優勢的經濟學

2026 年 10 月

兩隻寓言故事中的烏龜，為我們提供了兩種不同角度來思考人工智能。伊索寓言中的烏龜，最終超越兔子；而在古希臘哲學家芝諾所提出的悖論中，烏龜則始終領先阿基里斯一步：每當阿基里斯追到烏龜的位置，烏龜卻已再向前走了一小段。一個故事的結局是實現了後來居上的超越，另一個則呈現一段看似不斷縮小、卻始終無法完全消除的差距。

中國人工智能實驗室的發展，意外地讓這看似抽象的理論變成了現實的問題。它們在追趕 Anthropic 及 OpenAI 等美國領先企業的過程中各出奇謀，極具創意。先進晶片的供應受到限制，反而促使這些實驗室發展出新的模型設計及訓練方法，務求更有效運用稀缺的運算資源與資本。但問題在於，它們是正在迎頭趕上，還是只是追上領先者過去的水平？

差距正在收窄。SemiAnalysis 於 8 月發表的研究指出，中國領先的開放模型已能媲美 OpenAI 及 Anthropic 僅僅 5 至 6 個月前展現的能力。相比之下，根據該機構 2022 至 2024 年的數據，中國模型當時大約需要 18 個月才能追上。技術發展的前沿仍在向前推進，而領先企業尚未公開的模型可能已走得更遠。然而，競爭對手如今能以更快的速度複製它們的突破，令領先者能夠獨享技術優勢、並將其轉化為收益的時間不斷縮短。¹

對投資者而言，還有另一個問題：實驗室需要領先多少、維持多久，才足以證明為此付出的成本合理？答案不只關乎銷售人工智能模型的企業，也牽涉製造晶片、記憶體、電路板及電源供應器的企業，以及不斷擴大基建的融資投資者與貸款機構。8 月，《華爾街日報》估計，大型科技企業的表外開支承諾，包括未來的數據中心租賃及採購在內，約達 3 萬億美元。問題不只在於金額龐大，更在於時間尺度的不對稱：投資承諾動輒以年計，而技術領先地位卻可能幾個月內消失。²

多家領先的中國實驗室持續發表其工作的詳細技術報告。這種開放態度，讓我們得以審視它們取得進展的方法及成果。

中國部分最值得留意的進展，來自反思模型真正需要甚麼。DeepSeek 成功壓縮對話中較早部分的資訊，只需少量提示便可重新存取相關內容，使 V2 模型的快取記憶體需求較上一代減少逾 90%。V3 模型的架構將類似原則應用於運算方面：雖然模型合共有 6,700 億個參數，但每項單位處理僅需啟用約 370 億個參數。此外，模型亦會選擇性採用較低精確度的數值，從而減少部分運算所需資源，而在誤差容易累積的環節則保留較高精確度。這些創新的理念一脈相承：看似不可避免的運算要求，實際上可能只是系統原有設計方式所造成的高昂成本。^{3,4}

另一類的技術突破則涉及模型的學習方式。「知識蒸餾」是一種模型訓練技術，讓規模較小的學生模型學習較大型的教師模型的推理方式。從經濟學角度看，關鍵的區別其實很簡單：發掘一種新能力，與學會如何複製這種能力，兩者的成本未必相同。追隨者或可受惠於知識轉移，而毋須承擔探索過程中反覆試錯成本。

Anthropic 推出 Fable 5.1，令這類教學材料的價值變得格外明顯。Karo Zieminski 比較有關新版本的英文及中文報導後發現，兩者關注的重點明顯不同：英文內容主要聚焦於定價與安全性；中文內容則更重視防止模型遭蒸餾的措施、推理過程的取用權限，以及中轉服務。Anthropic 亦加強了防護措施，防止其他公司擷取模型的中間推理過程，並利用這些資料訓練另一個模型。同一新版本的相關報導，著眼點亦因應讀者的目的而截然不同：視乎他們是希望使用教師模型，還是了解如何讓另一個模型學習並仿效教師模型的能力。^{5,6}

近期有關多教師同策略蒸餾（MOPD）的研究，進一步延伸這個概念。不同領域的專用教師模型針對不同任務接受訓練，而學生模型則在嘗試自行執行任務時，根據教師模型提供的回饋與指導進行學習。這種方法旨在結合各教師模型的優勢，同時避免在學習不同能力時顧此失彼。6 月發表的論文指出，小米的 MiMo-V2-Flash 已採用這種方法。無論是教導模型的過程，還是接受訓練的模型本身，都體現了相當巧妙的設計。⁷

這些教師模型可以是實驗室自行開發的模型；而模型蒸餾亦不一定涉及複製競爭對手。同樣地，向另一間實驗室學習，也不代表會永遠落後於對方。研究人員可先透過模型蒸餾掌握既有能力，再透過自身的進一步訓練及研究開發新能力。今天的追隨者，或許就是明天的突破者。

領先者能否持續保持優勢的關鍵，取決於發掘下一項模型能力有多困難。若一家實驗室能夠一次又一次取得新的突破，便有可能維持這段具價值的領先差距。更先進的模型，也可協助研究人員開展相關工作，從而孕育下一代模型。充足的運算能力、客戶的有效回饋，以及強大的研究團隊，或有助支撐並持續推進探索的過程。然而，每一項優勢都需要仔細檢視；不應假定優勢能恆久不變。

然而，在後來居上與無止境追趕之間，還有第三種可能：追隨者或許永遠無法追上領先者，卻仍可削弱持領先地位所帶來的經濟效益。

價格較低的模型，或已足以應付客戶大部分工作，而最先進的模型在最困難的任務上則保持優勢。技術領先仍可延續，但願意支付大幅溢價的客戶群，可能較預期狹窄。實驗室或可持續在模型競賽中勝出，卻難以從勝利中賺取足夠回報。

不過相反情況亦可能出現。只要可靠度略為提升，或足以決定一項高價值任務能否完成。若能力更強的模型可節省昂貴的人手監督時間，即使每單位處理量的價格較高，攤分至每項成功完成的任務，成本仍可能較低。分銷、融入客戶工作流程及服務可靠性亦同樣重要。單憑基準測試分數，無法判斷領先優勢的真正價值。

領先的實驗室亦在成本方面展開競爭。於我們撰寫本文之際，Anthropic 推出 Claude Opus 5.5，OpenAI 則推出 GPT-6 SoL 與 Luna；這些模型在能力及效率據報有所提升之餘，價格亦更低。客戶可選用價格較低的模型，並按任務難度調整模型的推理程度，從而控制開支，而毋須轉用其他供應商。這或有助領先企業既保住最艱難的任務，也留住原本最容易受開放權重模型挑戰的日常工作。財務影響則取決於價格下跌是源於交付成本下降、犧牲利潤率，還是兩者兼有。取得資本的能力亦會影響實驗室在持續投入研究之餘，能維持進取定價多久。因此，必須考慮綜合能力、定價及融資三者。^{8,9}

只要成本下降得更快，或市場規模充分擴大，即使價格下跌亦無礙企業創造可觀回報。效率更高的模型可令智能服務價格低廉至足以刺激整體運算需求：這正是傑文斯悖論（Jevons Paradox）所描述的可能性。然而，單憑使用量增長，並不足以證明每家供應商都有能力履行承諾。若回報遜於預期，融資可能變得困難，開支計劃亦可能改變。長遠而言，運算需求或可錄得龐大增長，但期間的投資週期仍可能充滿波動。

我們的亞洲投資組合參與這個生態系統的多個環節，持股包括記憶體製造商、台積電、矽力杰、聯發科技、光寶科技、健鼎科技及瑞昱半導體。這些公司的業務模式差異顯著，而我們透過投資它們所把握的機會與承擔的風險也各不相同。

台積電為多家晶片設計商生產晶片；我們持有的記憶體企業則供應另一項不可或缺的運算組件。它們的製造專長涉及複雜的實體工序，以及可靠地大規模交付產品的能力。複製這些能力，與學習模型行為是截然不同的問題。然而，技術上的重要性並不能直接決定投資價值。我們仍須評估稀缺產能的溢價、正在形成的新供應，以及在市況轉弱或競爭加劇後，這些投資是否仍能帶來足夠回報。

中國亦正在硬件領域同步追趕。9 月，長鑫存儲宣佈第五代 DRAM 技術平台投入量產，並推出面向智能手機及便攜電子產品的新記憶體產品。這與大規模生產技術要求最高的高頻寬記憶體，同時將次品率控制在具競爭力的水平，是截然不同的成就。然而，實力提升的競爭者即使尚未掌握每項尖端技術，也可先透過增加成熟產品供應而影響回報。若一方面欣賞中國實驗室的創意，另一方面卻假定中國半導體製造商無法另闢蹊徑、突破限制，未免自相矛盾。¹⁰

另一方面，矽力杰生產用於電子系統的模擬及電源管理晶片；光寶科技則供應電源解決方案，包括數據中心的高壓電源機架。聯發科技的業務涵蓋定製晶片設計，讓客戶可按需要調整運算方案。瑞昱半導體設計通訊及連接晶片，而健鼎科技則生產印刷電路板。上述多家企業亦服務人工智能以外的龐大市場。這些企業的投資價值，並不單純取決於模型實驗室之間的競賽結果。^{11,12,13,14}

我們刻意將持倉分散於不同環節，但這並不能消除共同風險。供應鏈不同位置的供應商，可能同樣依賴相同客戶的開支。效率提升可擴大需求，卻亦可能減少每項任務所需的特定組件數量。我們的工作正正是要逐一審視這些因素對每家企業的影響，再評估其對整體投資組合的影響。

健鼎科技正好展現了我們所重視的投資判斷。該企業拒絕向 NVIDIA（英偉達）供貨，原因是嚴苛的規格要求及臨時設計變更削弱了可取得的回報。面對與知名客戶合作的機會，亦必須與業務的經濟效益權衡。健鼎科技選擇不追逐最先進技術，因為公司判斷稍為落後一步反而能帶來更吸引的回報。¹⁵

這種勇於說「不」的態度至關重要。我們所投資企業作出的資本配置決策，亦構成我們本身投資風險的一部分。在市場異常熾熱的時期，我們希望管理團隊審慎評估增長需要付出甚麼，包括所需投資、對客戶作出的承諾，以及一旦預期過於樂觀所帶來的後果。只有在滿足這些要求後仍能為股東保留價值，機遇才稱得上優質。

我們亦以相同紀律決定持倉規模。我們在多家大型企業的持倉往往較為適中，並低於它們各自在基準中的權重。持有一家企業，並不代表必須跟隨市場賦予其同等的重要性。我們希望限制集中風險，包括基準本身所包含的集中風險。

當大市升勢持續由少數大型企業帶動時，這意味着我們或需承受相對表現落後；而這正是投資取向上的實際取舍。我們採取絕對回報方針，首要考慮是避免資本永久損失的風險，而非擔心投資表現是否偏離基準指數。指數權重並不能告訴我們，一項投資是否提供充足的安全邊際。

伊索寓言與芝諾悖論之所以值得借鑑，正因它們提醒我們大量問題尚未有定論。我們可以肯定中國實驗室的創新能力，但不應因此假設它們會令最前沿的研究失去經濟效益。也可以肯定現時領先者的優勢，但不應假設其領先地位總能持續帶來吸引的回報。我們應保持謙遜，並將這份態度體現於我們願意支付的價格、持倉規模，以及對管理層提出的期望之中。

我們將持續研究相關技術，並隨着新的證據不斷修正判斷。投資組合既應讓我們分享其發展成果，也要在某項關鍵假設被證明錯誤時，將損失控制在可承受範圍內。無論這場競賽最終更像哪一個故事，我們的首要責任始終是守護客戶資本。

- 1 SemiAnalysis, 《Are Open Models Catching Up》, 2026 年 8 月 21 日。歷史比較涵蓋更廣泛的開放模型，包括美國模型；近期例子則為中國模型。各基準在規模擴展、推理及代理式人工智能時代有所不同。[閱讀原文](#)
 - 2 Peter Rudegeair 及 Peter Santilli, 《Why Big Tech's AI Spending Is \$3 Trillion Higher Than It Seems》, 《華爾街日報》, 2026 年 8 月。估算涉及大型科技企業的承諾，並非指模型實驗室單獨欠下 3 萬億美元債務。[閱讀原文](#)
 - 3 DeepSeek AI, 《DeepSeek V2 技術報告》, 2024 年 5 月。報告所述 93.3% 的減幅，是指相對 DeepSeek 67B 的 KV 快取，而非模型總記憶體。[閱讀原文](#)
 - 4 DeepSeek AI, 《DeepSeek V3 技術報告》, 2024 年 12 月。混合專家模型及選擇性 FP8 訓練。[閱讀原文](#)
 - 5 Karo Zieminski, 《Claude Fable 5.1 What Chinese Coverage Says and English Doesn't》, 2026 年 9 月 3 日。比較涵蓋 136 項中文查詢及一次英文搜尋所呈現的報導，並非具代表性的全國民意調查。[閱讀原文](#)
 - 6 Anthropic, 《Preserved thinking changing how the Messages API handles thinking blocks to protect against distillation》, 2026 年 9 月。推理提取保護措施的範圍及目的。[閱讀原文](#)
 - 7 Ma et al, 《MOPD Multi Teacher On Policy Distillation for Capability Integration in LLM Post Training》, 首次提交日期為 2026 年 6 月 29 日。[閱讀原文](#)
 - 8 Anthropic, 《Introducing Claude Opus 5.5》, 2026 年 9 月 22 日。公司公布的能力、效率及定價變動。[閱讀原文](#)
 - 9 OpenAI, 《Introducing GPT-6 Sol and Luna》, 2026 年 9 月 22 日。公司公布的能力提升及 API 減價。[閱讀原文](#)
 - 10 長鑫存儲 (CXMT), 《CXMT Announces Mass Production of 5th Generation DRAM Technology Platform》, 2026 年 9 月 20 日。公司就 DRAM 及 LPDDR5X 發出的公告；並不足以證明其在領先 HBM 領域具有商業競爭力。[閱讀原文](#)
 - 11 矽力杰, 公司簡介。模擬、電源及訊號鏈產品與應用。[閱讀原文](#)
 - 12 光寶科技, NVIDIA GTC 2026 公告, 2026 年 3 月 17 日。數據中心電源解決方案。[閱讀原文](#)
 - 13 聯發科技, 數據中心解決方案。定製晶片設計及相關工程服務。[閱讀原文](#)
 - 14 瑞昱半導體, 通訊網絡 IC。產品組合。[閱讀原文](#)
 - 15 首域盈信有關健鼎科技的研究及簡報材料, 2026 年 9 月。NVIDIA 例子反映投資團隊對公司決策的說明。
- * 本資料對具體證券的提述（如有）僅供說明用途，不應視成建議買賣有關證券。本資料提述的所有證券可能或可能不會在某時期成為基金組合的部份持倉，而持倉可能作出變動。

免責聲明

本資料所載只屬一般性資料，並不包括亦不構成投資或投資產品建議。有關資料來自首源集團認為可靠的來源，而且在發表本資料時為準確的資料，首源集團並無就有關資料的中肯、準確、完整或正確作出明確或隱含的聲明或保證。在法律允許的範圍內，首源集團、其任何聯營公司、任何董事、主管人員或僱員概不就任何因使用本資料而直接或間接引致的損失承擔任何責任。本資料所載資料僅供參考，並不構成投資意見及 / 或建議，且不應作為任何投資決定的基礎。本資料所述部分基金未獲授權在若干司法管轄區向公眾人士要約 / 發售。

投資價值及收益可跌亦可升，閣下也許未能取回投資本金。過往表現未必可作日後業績的標準。投資者欲知有關詳情，請參閱基金銷售文件，包括風險因素。本資料對具體證券的提述（如有）僅供說明用途，不應視成建議買賣有關證券。本資料提述的所有證券可能或可能不會在某時期成為基金組合的部份持倉，而持倉可能作出變動。

本資料僅供一般參考，並不可視為提供全面的資料或特別建議。本資料所述意見為發件人在發表報告時的觀點，並不一定為首源集團的觀點，並可隨時予以修訂。本資料並非銷售文件，且不構成投資建議。未取得具體的專業意見前，任何人士不應依賴本資料內容，及 / 或有關資料所載的任何事宜為行事基礎。未經首源集團預先同意，不得複製或傳閱本資料的全部或部分資料。本資料只可於符合相關司法管轄區的適用法例下使用及 / 或收取。

於香港，此資料由首源投資（香港）有限公司發行，並未經香港證監會審閱。首源集團、首源投資、首域盈信資產管理、信衛資產管理、RQI Investors 及 Igneo Infrastructure Partners 是首源投資（香港）有限公司的商業名稱。於新加坡，此資料則由首源投資（新加坡）發行（公司註冊號碼：196900420D）。此廣告或資料並未經新加坡金融管理局審閱。首源集團（註冊編號 53507290B）、首源投資（註冊編號 53236800B）、首域盈信資產管理（註冊編號 53314080C）、信衛資產管理（註冊編號 53310114W）、RQI Investors（註冊編號 53472532E）及 Igneo Infrastructure Partners（註冊編號 53447928J）是首源投資（新加坡）的商業名稱。首源投資（香港）有限公司及首源投資（新加坡）是首源集團投資管理業務的一部分，該公司最終由全球金融集團三菱日聯金融集團（“MUFG”）擁有。首源集團在不同司法管轄區內有多間公司。

在法律允許的範圍內，MUFG 及其附屬公司對本資料包含的任何聲明或信息概不負責。MUFG 及其任何附屬公司均不保證本資料中提及的任何投資或公司的業績或履行償還資本。本資料中提及的任何投資並不是對 MUFG 或其附屬公司的存款或其他負債，而投資存在投資風險，包括收入和資本損失。