

在費城半導體指數步入持續性回調、海外 AI 硬體板塊持續低迷的背景下，月之暗面發佈 2.8 萬億參數開源大模型 Kimi K3，該模型登頂全球前端代碼盲測榜單，單項能力追平甚至超越 GPT-5.6 Sol、Claude Fable 5 等頂級閉源模型，被稱為中國又一“DeepSeek 時刻”，引發全球對國產大模型的高度關注；K3 代碼性能從前代第 18 名躍升至代碼賽道全球前列、綜合實力躋身全球第一梯隊，同時定價創下國產模型新高，標誌著國產大模型正式擺脫低價內卷，進入“性能定價”新階段；公司當前新一輪融資投前估值達 315 億美元，半年內漲幅超 7 倍，完成後將啟動上市前最後一輪融資，最快 6 個月赴港掛牌，有望帶動港股 AI 板塊迎來情緒修復與估值重構，可重點關注雲服務、算力伺服器與 AI 應用廠商等港股 AI 投資主線。

“Kimi 時刻” 刷屏華爾街！Kimi K3 開源重估全球 AI 敘事

近期美股半導體板塊持續回調，費城半導體指數步入技術性熊市，海外 AI 硬體敘事階段性退潮。在此背景下，國產大模型領域迎來重磅突破，月之暗面於 7 月 17 日正式發佈 2.8 萬億參數開源權重模型 Kimi K3，憑藉全球頂級的代碼能力與綜合性能迅速刷屏華爾街投資圈，被海外市場稱為繼 DeepSeek 之後的又一個“中國 AI 驚喜時刻”，徹底引爆全球對國產大模型的關注度。

本次發佈的 Kimi K3 刷新了開源模型參數規模紀錄，同時支持 100 萬 Token 上下文窗口與原生多模態能力；其在 Arena.ai 前端代碼排行榜中以 1678 分登頂榜首，綜合性能可與 Anthropic Fable 5、OpenAI GPT-5.6 Sol 等頂級閉源模型媲美，且調用成本較美國頂級模型低約 40%，打破了國產模型低價內卷的行業格局。在定價策略端，每百萬 Token 混合收費約 2.3 美元，顯著高於其他國產旗艦但較海外頂級模型低約 40%，K3 兼具性能與性價比優勢，也標誌著國產大模型正式擺脫低價內卷，邁入“性能定價”的全新發展階段。認為隨著中美前沿 AI 差距大幅收窄，中國 AI 開源模型正接近全球普及的智能臨界點，預計 2026 年下半年高端編程領域競爭將進一步加劇，模型參數量將向 2 萬億至 5 萬億區間擴容。

資本端，月之暗面的上市進程正加速推進，公司當前新一輪融資投前估值達 315 億美元，7 個月內估值漲幅超 7 倍；本輪融資完成後將立即啟動上市前最後一輪融資，目標投前估值 500 億美元，最快 6 個月內赴港掛牌，有望成為港股 AI 大模型板塊第三家核心標的，帶動整個賽道估值體系重構。在此背景下，港股 AI 板塊有望憑藉 Kimi K3 驗證的國產大模型技術躍遷能力，疊加上市預期催化，迎來情緒修復與估值重估。



圖一：全球大模型前端代碼盲測榜單

Rank ↕	Rank Spread ↕	Model ↕	Score ↓	Votes ↕	Price \$/M ↕	Context ↕
1	1 → 1	 kimi-k3 Moonshot - Proprietary	1678 +17/-17 	1,828	\$3 / \$15	1M
2	2 → 3	 claude-fable-5 Anthropic - Proprietary	1634 +12/-12	3,021	\$10 / \$50	1M
3	2 → 3	 gpt-5.6-sol-xhigh (codex-harness) OpenAI - Proprietary	1630 +11/-11	4,063	\$5 / \$30	1.1M
4	4 → 4	 glm-5.2 (max) Zai - MIT	1592 +10/-10	5,362	\$140 / \$4.40	1M
5	5 → 8	 claude-opus-4-8-thinking Anthropic - Proprietary	1565 +8/-8	7,978	\$5 / \$25	1M
6	5 → 9	 claude-opus-4-7-thinking Anthropic - Proprietary	1559 +7/-7	11,161	\$5 / \$25	1M
7	5 → 12	 grok-4.5 SpaceXAI - Proprietary	1557 +12/-12	2,887	\$2 / \$6	500K
8	5 → 11	 claude-opus-4-7 Anthropic - Proprietary	1556 +7/-7	10,596	\$5 / \$25	1M
9	6 → 17	 claude-sonnet-5-high Anthropic - Proprietary	1544 +11/-11	3,554	\$2 / \$10	1M
10	8 → 16	 claude-opus-4-6-thinking Anthropic - Proprietary	1542 +6/-6	13,522	\$5 / \$25	1M

資料來源：Arena.ai

➤ 產品、定價、技術三重突破

*性能：從第 18 名到全球前三，實現跨越式躍升

Kimi K3 是月之暗面技術迭代速度的集中體現，其性能躍升幅度遠超市場預期。在權威的 Arena.ai 前端代碼盲測榜單中，K3 以 1678 的 Elo 分數位居第一，領先第二名 Claude Fable 5 (1634 分) 44 分，領先第三名 GPT-5.6 Sol (1630 分) 48 分。該排名基於 98 個模型、約 48.4 萬次兩兩對決的盲測投票得出，結果具備極強的市場公信力。值得注意的是，Kimi 前代模型 K2.6 在同榜單中僅排名第 18 位，短短數月內實現 17 個位次的躍升，驗證了公司極強的技術迭代效率。在細分維度上，7 個前端評測領域中 K3 拿下 6 項第一，僅遊戲開發領域小幅落後於 Fable 5。綜合評測維度上，K3 在 Artificial Analysis 評分中位居全球第三，僅次於兩款海外頂級閉源模型，大幅領先其他國產大模型。功能層面，K3 定位為面向軟體工程、深度推理與知識工作的旗艦模型，原生支持視覺理解，採用自研的 Delta Attention 混合線性注意力機制與 Attention Residuals 結構，實現百萬上下文下最高 6.3 倍的解碼加速，訓練效率提升約 25%，在長文檔處理、代碼庫全量分析等場景具備顯著優勢。



***定價：打破國產低價內卷，進入性能定價時代**

K3 的定價體系同樣具備行業標誌性意義，創下國產大模型定價新高，徹底脫離了國產模型“低價內卷”的傳統賽道。按照官方 API 定價，K3 人民幣計價為輸入 20 元/百萬 Token、輸出 100 元/百萬 Token，折算混合收費約每百萬 Token 2.3 美元，顯著高於國內同行：Qwen 3.7 Max 約 1.4 美元、智譜 GLM 5.2 約 0.9 美元、MiniMax M3 約 0.22 美元、DeepSeek V4 Pro 約 0.18 美元。

但對比海外頂級模型，K3 仍具備極強的性價比優勢：GPT-5.6 Sol 定價為輸入 5 美元、輸出 30 美元/百萬 Token，Claude Fable 5 定價更高達輸入 10 美元、輸出 50 美元，K3 綜合使用成本較海外頂級模型低約 40%。這種“高於國產、低於海外”的定價策略，本質是公司對自身技術實力的信心體現——模型能力進入全球第一梯隊，定價自然脫離國產低價區間，標誌著國產大模型行業正式從“價格戰”轉向“價值戰”。從商業化回饋看，高定價並未阻礙市場需求，K3 發佈後單日銷售額增至發佈前的 6 倍以上，年化收入 (ARR) 創下歷史最大單日增幅，驗證了高端大模型市場的付費意願與產品認可度。

圖二：Kimi 會員定價體系

Andante 日常使用	Moderato 效率升級	Allegretto 專業首選	Allegro 全能尊享
¥39 / 月 ¥49 按年計費 ¥468 / 年	¥79 / 月 ¥99 按年計費 ¥948 / 年	¥159 / 月 ¥199 按年計費 ¥1,908 / 年	¥559 / 月 ¥699 按年計費 ¥6,708 / 年
預約訂閱	預約訂閱	預約訂閱	預約訂閱
✓ 更多 Agent 額度 - Office 文件處理 - 深度研究 - 網站部署 ✓ Kimi Code 可調用	✓ 2 倍 Agent 額度 - Office 文件處理 - 深度研究 - 網站部署 ✓ Kimi Code 4 倍額度	✓ 4 倍 Agent 額度 - Office 文件處理 - 深度研究 - 網站部署 ✓ Agent 多任務并行 ✓ Kimi Code 20 倍額度	✓ 10 倍 Agent 額度 - Office 文件處理 - 深度研究 - 網站部署 ✓ Agent 多任務并行 ✓ Kimi Code 60 倍額度

資料來源：公司官網

***開源價值：重構開源與閉源的邊界**

K3 作為開放權重模型，其行業影響力遠不止於單一產品。過往全球頂級 AI 能力幾乎全部掌握在 OpenAI、Anthropic 等閉源廠商手中，開源模型與頂級閉源模型存在代際差距。而 K3 首次實現了開源模型在核心專項能力上追平閉源旗艦，大幅抬高了全球開源模型的能力天花板。7 月 27 日 K3 全量權重正式開放後，全球開發者與企業均可基於該模型進行二次開發、本地化部署與行業微調，將直接帶動全球 AI 開發生態的整體升級，加速 AI 技術在各行業的滲透落地。對於國內產業而言，頂級開源模型的出現，將大幅降低國內企業使用高端 AI 能力的門檻與數據安全風險，推動產業智能化進程提速。

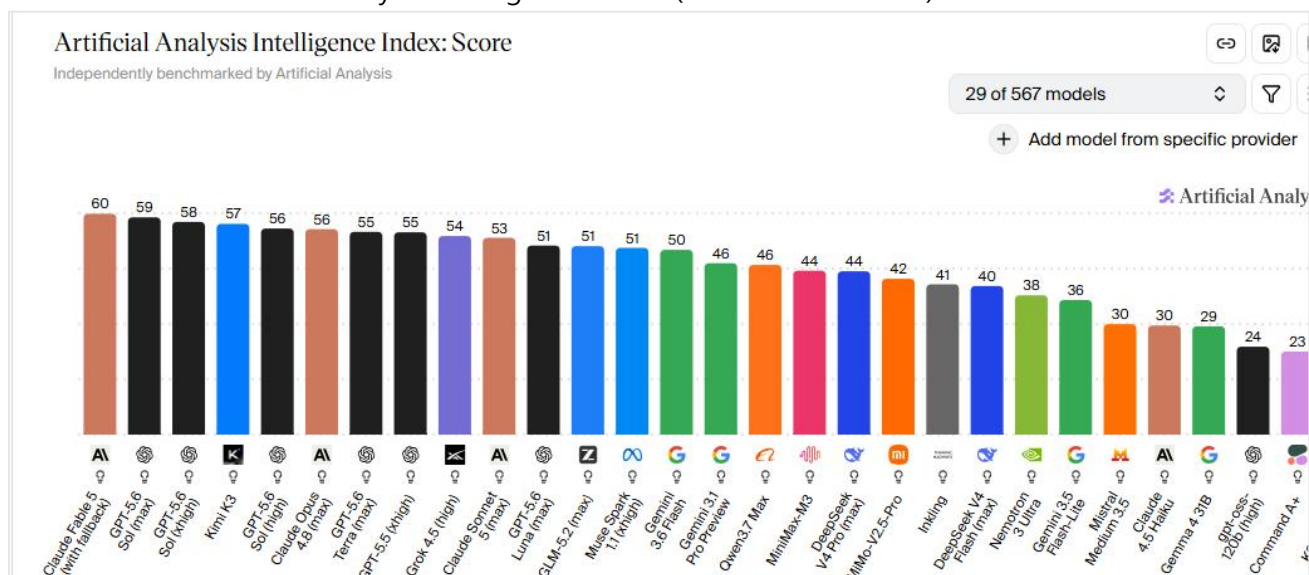


➤ 行業影響：重塑全球 AI 敘事，兩大格局生變

*中美 AI 技術差距快速收窄，國產模型全球話語權提升

Kimi K3 的發佈最核心的產業意義，在於直接打破了“中國 AI 大幅落後美國”的刻板認知。僅一年前，國產大模型還普遍被認為落後海外頂尖水準 1-2 代；而當前，國產模型已在代碼、長上下文等核心賽道實現反超，綜合能力躋身全球第一梯隊。這種技術躍遷並非個例，DeepSeek、月之暗面、智譜等多家廠商持續實現技術突破，形成了群體式崛起的態勢。傑富瑞等海外機構已明確修正對中國 AI 的判斷，認為中美前沿 AI 的差距已大幅縮小，中國大模型廠商具備全球競爭力。長期來看，國產大模型將從“跟隨者”逐步轉向“並行競爭者”，在全球 AI 產業規則制定、技術標準、市場份額中擁有更多話語權。

圖三：Artificial Analysis Intelligence Index (人工分析智能指數) 大模型綜合得分排名



資料來源：Artificial Analysis

*行業定價體系重構，AI 應用落地進程提速

此前國內大模型行業陷入普遍的價格戰，廠商持續降價爭奪市場份額，行業盈利模式模糊。K3 的高價策略與市場正向回饋，為行業探索出了新路徑——只要技術能力達到相應層級，市場願意支付對應的溢價。這將推動國內大模型行業形成清晰的價值分層：頭部廠商憑藉頂級能力瞄準高端企業級市場，以高定價覆蓋研發成本；中腰部廠商聚焦垂類場景與性價比路線，差異化競爭。行業整體將從“拼低價、沖調用量”的粗放競爭，轉向“拼性能、拼落地、拼盈利品質”的高質量競爭，有利於頭部優質廠商的長期發展。

另外，頂級開源模型的出現，將極大豐富全球 AI 開發生態。一方面，大量中小企業、開發者無需支付高昂的閉源 API 費用，即可使用接近頂級水準的 AI 能力開展創新，大幅降低 AI 創業與應用開發的門



檻；另一方面，開源模型支持本地化部署，能夠滿足金融、政務、工業等領域的數據安全與隱私要求，打開 AI 在敏感行業的落地空間。同時，開源生態的繁榮將反向催生出更多模型微調、推理優化、行業解決方案的需求，帶動算力、數據、AI 應用等整條產業鏈的需求擴容，形成“基礎模型開源—生態應用繁榮—反哺模型迭代”的正向迴圈。

*公司資本進程：估值半年漲 7 倍，港股上市進入倒計時

技術突破的同時，月之暗面的資本化進程也在全面加速。據彭博社等多家媒體披露，公司當前正推進新一輪融資，投前估值約 315 億美元，預計近期完成交割。值得注意的是，2025 年底公司估值僅為 43 億美元，短短半年時間漲幅超 7 倍，是當前全球估值增長最快的 AI 獨角獸之一。本輪融資完成後，公司將立即啟動上市前最後一輪私募股權融資，目標投前估值達 500 億美元，為赴港上市做最終準備。架構層面，公司已啟動拆除原紅籌 VIE 架構的工作，轉為適配港股上市的股權結構，擬聘請中金公司與高盛擔任聯席保薦機構，最快 6 個月內登陸香港資本市場，最早有望於 2027 年初正式掛牌。若順利上市，月之暗面將成為繼智譜、MiniMax 之後，港股第三家純大模型龍頭企業，也將是當前港股 AI 板塊中估值規模、技術實力最頂尖的標的之一，其上市將顯著提升港股 AI 賽道的全球吸引力與資金關注度，帶動整個板塊的估值體系重構。

➤ 港股核心受益標的有哪些？

*算力伺服器與整機

大模型開源帶動全行業私有化部署需求爆發，伺服器整機廠商率先受益於推理算力採購放量。聯想集團 (0992.HK) 作為全球 AI 伺服器龍頭，覆蓋從訓練到推理的全場景伺服器產品，企業端 K3 私有化部署將直接拉動通用推理伺服器採購需求，同時公司可提供算力調度與私有化部署全棧服務，具備從硬體到方案的一體化交付能力。中興通訊 (0763.HK) 實現全棧 ICT 佈局，AI 伺服器、交換機、光模組、算力操作系統全覆蓋，深度參與國產算力生態建設，直接受益於 AI 推理算力的分佈式部署浪潮。

* IDC 與算力雲服務

大模型推理與私有化部署高度依賴高等級數據中心與雲算力資源，IDC 與雲服務商迎來需求增量。萬國數據-SW (9698.HK) 是國內第三方高等級 IDC 龍頭，核心節點資源適配大模型推理與私有化部署需求，是企業落地本地算力的核心載體。金山雲 (3896.HK) 作為國內頭部雲服務商，提供大模型訓練、微調、推理全流程算力服務，K3 開源後中小客戶私有化部署上雲需求快速增長，公司雲算力業務有望持續放量。



***應用內容與數據要素**

高質量正版數據是大模型能力迭代的**核心基礎**，K3 開源帶動全行業模型訓練與微調需求攀升，數據資產價值迎來持續重估。閱文集團 (0772.HK) 擁有國內最大的**正版網文 IP 庫**，海量優質文字內容是大模型**核心訓練數據資產**，同時公司可基於大模型快速衍生 IP 內容，實現數據價值與內容生產效率的雙向提升。快手-W (1024.HK) 坐擁海量**短視頻內容數據資產**，同時**視頻智能理解、AI 內容生成場景**落地加速，K3 多模態能力開源將顯著降低技術落地門檻，推動公司 AI 應用商業化提速。商湯-W (0020.HK)：**電腦視覺 AI 龍頭**，同時佈局大模型與 AIDC 算力基礎設施，日日新大模型持續迭代，**算力資源稀缺性凸顯**，直接受益於大模型行業需求擴容。

免責聲明：本報內容所提供資料所述或與其相關的任何投資或潛在交易，均受限於閣下司法轄區適用的法律及監管規定，而閣下須單獨就遵守該等法律及監管規定負責。本報內容僅供參考，不構成任何投資建議。本公司對所提供的財經資訊已力求準確，但對其中全部或部分內容的準確性、完整性或有效性，不承擔任何責任或提供任何形式保證。如有錯失遺漏，本公司恕不負責。另請注意證券與虛擬資產價格可升可跌，尤其虛擬資產的風險極高，投資者應對有關產品保持審慎及自行承擔投資風險。

