

在费城半导体指数步入持续性回调、海外 AI 硬件板块持续低迷的背景下，月之暗面发布 2.8 万亿参数开源大模型 Kimi K3，该模型登顶全球前端代码盲测榜单，单项能力追平甚至超越 GPT-5.6 Sol、Claude Fable 5 等顶级闭源模型，被称为中国又一“DeepSeek 时刻”，引发全球对国产大模型的高度关注；K3 代码性能从前代第 18 名跃升至代码赛道全球前列、综合实力跻身全球第一梯队，同时定价创下国产模型新高，标志着国产大模型正式摆脱低价内卷，进入“性能定价”新阶段；公司当前新一轮融资投前估值达 315 亿美元，半年内涨幅超 7 倍，完成后将启动上市前最后一轮融资，最快 6 个月赴港挂牌，有望带动港股 AI 板块迎来情绪修复与估值重构，可重点关注云服务、算力服务器与 AI 应用厂商等港股 AI 投资主线。

“Kimi 时刻”刷屏华尔街！Kimi K3 开源重估全球 AI 叙事

近期美股半导体板块持续回调，费城半导体指数步入技术性熊市，海外 AI 硬件叙事阶段性退潮。在此背景下，国产大模型领域迎来重磅突破，月之暗面于 7 月 17 日正式发布 2.8 万亿参数开源权重模型 Kimi K3，凭借全球顶级的代码能力与综合性能迅速刷屏华尔街投资圈，被海外市场称为继 DeepSeek 之后的又一个“中国 AI 惊喜时刻”，彻底引爆全球对国产大模型的关注度。

本次发布的 Kimi K3 刷新了开源模型参数规模纪录，同时支持 100 万 Token 上下文窗口与原生多模态能力；其在 Arena.ai 前端代码排行榜中以 1678 分登顶榜首，综合性能可与 Anthropic Fable 5、OpenAI GPT-5.6 Sol 等顶级闭源模型媲美，且调用成本较美国顶级模型低约 40%，打破了国产模型低价内卷的行业格局。在定价策略端，每百万 Token 混合收费约 2.3 美元，显著高于其他国产旗舰但较海外顶级模型低约 40%，K3 兼具性能与性价比优势，也标志着国产大模型正式摆脱低价内卷，迈入“性能定价”的全新发展阶段。认为随着中美前沿 AI 差距大幅收窄，中国 AI 开源模型正接近全球普及的智能临界点，预计 2026 年下半年高端编程领域竞争将进一步加剧，模型参数量将向 2 万亿至 5 万亿区间扩容。

资本端，月之暗面的上市进程正加速推进，公司当前新一轮融资投前估值达 315 亿美元，7 个月内估值涨幅超 7 倍；本轮融资完成后将立即启动上市前最后一轮融资，目标投前估值 500 亿美元，最快 6 个月内赴港挂牌，有望成为港股 AI 大模型板块第三家核心标的，带动整个赛道估值体系重构。在此背景下，港股 AI 板块有望凭借 Kimi K3 验证的国产大模型技术跃迁能力，叠加上市预期催化，迎来情绪修复与估值重估。



图一：全球大模型前端代码盲测榜单

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

资料来源：Arena.ai

➤ 产品、定价、技术三重突破

*性能：从第 18 名到全球前三，实现跨越式跃升

Kimi K3 是月之暗面技术迭代速度的集中体现，其性能跃升幅度远超市场预期。在权威的 Arena.ai 前端代码盲测榜单中，K3 以 1678 的 Elo 分数位居第一，领先第二名 Claude Fable 5 (1634 分) 44 分，领先第三名 GPT-5.6 Sol (1630 分) 48 分。该排名基于 98 个模型、约 48.4 万次两两对决的盲测投票得出，结果具备极强的市场公信力。值得注意的是，Kimi 前代模型 K2.6 在同榜单中仅排名第 18 位，短短数月内实现 17 个位次的跃升，验证了公司极强的技术迭代效率。在细分维度上，7 个前端评测领域中 K3 拿下 6 项第一，仅游戏开发领域小幅落后于 Fable 5。综合评测维度上，K3 在 Artificial Analysis 评分中位居全球第三，仅次于两款海外顶级闭源模型，大幅领先其他国产大模型。功能层面，K3 定位为面向软件工程、深度推理与知识工作的旗舰模型，原生支持视觉理解，采用自研的 Delta Attention 混合线性注意力机制与 Attention Residuals 结构，实现百万上下文下最高 6.3 倍的解码加速，训练效率提升约 25%，在长文档处理、代码库全量分析等场景具备显著优势。



***定价：打破国产低价内卷，进入性能定价时代**

K3 的定价体系同样具备行业标志性意义，创下国产大模型定价新高，彻底脱离了国产模型“低价内卷”的传统赛道。按照官方 API 定价，K3 人民币计价每百万 Token 输入 20 元、输出 100 元，折合混合收费约每百万 Token 2.3 美元，显著高于国内同行：Qwen 3.7 Max 约 1.4 美元、智谱 GLM 5.2 约 0.9 美元、MiniMax M3 约 0.22 美元、DeepSeek V4 Pro 约 0.18 美元。

但对比海外顶级模型，K3 仍具备极强的性价比优势：GPT-5.6 Sol 定价为输入 5 美元、输出 30 美元/百万 Token，Claude Fable 5 定价更高达输入 10 美元、输出 50 美元，K3 综合使用成本较海外顶级模型低约 40%。这种“高于国产、低于海外”的定价策略，本质是公司对自身技术实力的信心体现——模型能力进入全球第一梯队，定价自然脱离国产低价区间，标志着国产大模型行业正式从“价格战”转向“价值战”。从商业化反馈看，高定价并未阻碍市场需求，K3 发布后单日销售额增至发布前的 6 倍以上，年化收入 (ARR) 创下历史最大单日增幅，验证了高端大模型市场的付费意愿与产品认可度。

图二：Kimi 会员定价体系

Andante	Moderato	Allegretto	Allegro
日常使用	效率升级	专业优选	全能尊享
¥39 / 月 ¥49	¥79 / 月 ¥99	¥159 / 月 ¥199	¥559 / 月 ¥699
按年计费 ¥468 / 年	按年计费 ¥948 / 年	按年计费 ¥1,908 / 年	按年计费 ¥6,708 / 年
预约订阅	预约订阅	预约订阅	预约订阅
✓ 更多 Agent 额度 - Office 文件处理 - 深度研究 - 网站部署 ✓ Kimi Code 可调	✓ 2 倍 Agent 额度 - Office 文件处理 - 深度研究 - 网站部署 ✓ Kimi Code 4 倍额度	✓ 4 倍 Agent 额度 - Office 文件处理 - 深度研究 - 网站部署 ✓ Agent 多任务并行 ✓ Kimi Code 20 倍额度	✓ 10 倍 Agent 额度 - Office 文件处理 - 深度研究 - 网站部署 ✓ Agent 多任务并行 ✓ Kimi Code 60 倍额度

资料来源：公司官网

***开源价值：重构开源与闭源的边界**

K3 作为开放权重模型，其行业影响力远不止于单一产品。过往全球顶级 AI 能力几乎全部掌握在 OpenAI、Anthropic 等闭源厂商手中，开源模型与顶级闭源模型存在代际差距。而 K3 首次实现了开源模型在核心专项能力上追平闭源旗舰，大幅抬高了全球开源模型的能力天花板。7 月 27 日 K3 全量权重正式开放后，全球开发者与企业均可基于该模型进行二次开发、本地化部署与行业微调，将直接带动全球 AI 开发生态的整体升级，加速 AI 技术在各行业的渗透落地。对于国内产业而言，顶级开源模型的出现，将大幅降低国内企业使用高端 AI 能力的门槛与数据安全风险，推动产业智能化进程提速。

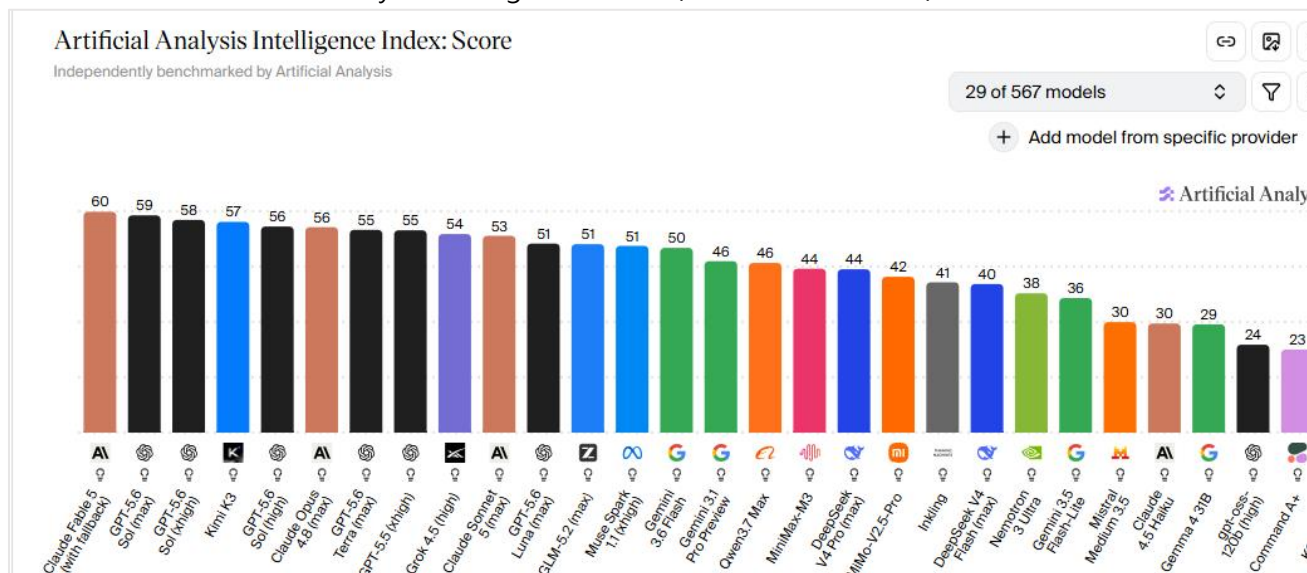


➤ 行业影响：重塑全球 AI 叙事，两大格局生变

*中美 AI 技术差距快速收窄，国产模型全球话语权提升

Kimi K3 的发布最核心的产业意义，在于直接打破了“中国 AI 大幅落后美国”的刻板认知。仅一年前，国产大模型还普遍被认为落后海外顶尖水平 1-2 代；而当前，国产模型已在代码、长上下文等核心赛道实现反超，综合能力跻身全球第一梯队。这种技术跃迁并非个例，DeepSeek、月之暗面、智谱等多家厂商持续实现技术突破，形成了群体式崛起的态势。杰富瑞等海外机构已明确修正对中国 AI 的判断，认为中美前沿 AI 的差距已大幅缩小，中国大模型厂商具备全球竞争力。长期来看，国产大模型将从“跟随者”逐步转向“并行竞争者”，在全球 AI 产业规则制定、技术标准、市场份额中拥有更多话语权。

图三：Artificial Analysis Intelligence Index (人工分析智能指数) 大模型综合得分排名



资料来源：Artificial Analysis

*行业定价体系重构，AI 应用落地进程提速

此前国内大模型行业陷入普遍的价格战，厂商持续降价争夺市场份额，行业盈利模式模糊。K3 的高价策略与市场正向反馈，为行业探索出了新路径——只要技术能力达到相应层级，市场愿意支付对应的溢价。这将推动国内大模型行业形成清晰的价值分层：头部厂商凭借顶级能力瞄准高端企业级市场，以高定价覆盖研发成本；中腰部厂商聚焦垂类场景与性价比路线，差异化竞争。行业整体将从“拼低价、冲调用量”的粗放竞争，转向“拼性能、拼落地、拼盈利质量”的高质量竞争，有利于头部优质厂商的长期发展。

另外，顶级开源模型的出现，将极大丰富全球 AI 开发生态。一方面，大量中小企业、开发者无需支付高昂的闭源 API 费用，即可使用接近顶级水平的 AI 能力开展创新，大幅降低 AI 创业与应用开发的门



檻；另一方面，开源模型支持本地化部署，能够满足金融、政务、工业等领域的数据安全与隐私要求，打开 AI 在敏感行业的落地空间。同时，开源生态的繁荣将反向催生出更多模型微调、推理优化、行业解决方案的需求，带动算力、数据、AI 应用等整条产业链的需求扩容，形成“基础模型开源—生态应用繁荣—反哺模型迭代”的正向循环。

*公司资本进程：估值半年涨 7 倍，港股上市进入倒计时

技术突破的同时，月之暗面的资本化进程也在全面加速。据彭博社等多家媒体披露，公司当前正推进新一轮融资，投前估值约 315 亿美元，预计近期完成交割。值得注意的是，2025 年底公司估值仅为 43 亿美元，短短半年时间涨幅超 7 倍，是当前全球估值增长最快的 AI 独角兽之一。本轮融资完成后，公司将立即启动上市前最后一轮私募股权融资，目标投前估值达 500 亿美元，为赴港上市做最终准备。架构层面，公司已启动拆除原红筹 VIE 架构的工作，转为适配港股上市的股权结构，拟聘请中金公司与高盛担任联席保荐机构，最快 6 个月内登陆香港资本市场，最早有望于 2027 年初正式挂牌。若顺利上市，月之暗面将成为继智谱、MiniMax 之后，港股第三家纯大模型龙头企业，也将是当前港股 AI 板块中估值规模、技术实力最顶尖的标的之一，其上市将显著提升港股 AI 赛道的全球吸引力与资金关注度，带动整个板块的估值体系重构。

➤ 港股核心受益标的有哪些？

*算力服务器与整机

大模型开源带动全行业私有化部署需求爆发，服务器整机厂商率先受益于推理算力采购放量。联想集团 (0992.HK) 作为全球 AI 服务器龙头，覆盖从训练到推理的全场景服务器产品，企业端 K3 私有化部署将直接拉动通用推理服务器采购需求，同时公司可提供算力调度与私有化部署全栈服务，具备从硬件到方案的一体化交付能力。中兴通讯 (0763.HK) 实现全栈 ICT 布局，AI 服务器、交换机、光模块、算力操作系统全覆盖，深度参与国产算力生态建设，直接受益于 AI 推理算力的分布式部署浪潮。

* IDC 与算力云服务

大模型推理与私有化部署高度依赖高等级数据中心与云算力资源，IDC 与云服务商迎来需求增量。万国数据-SW (9698.HK) 是国内第三方高等级 IDC 龙头，核心节点资源适配大模型推理与私有化部署需求，是企业落地本地算力的核心载体。金山云 (3896.HK) 作为国内头部云服务商，提供大模型训练、微调、推理全流程算力服务，K3 开源后中小客户私有化部署上云需求快速增长，公司云算力业务有望持续放量。



***应用内容与数据要素**

高质量正版数据是大模型能力迭代的核心基础，K3 开源带动全行业模型训练与微调需求攀升，数据资产价值迎来持续重估。阅文集团 (0772.HK) 拥有国内最大的正版网文 IP 库，海量优质文字内容是大模型核心训练数据资产，同时公司可基于大模型快速衍生 IP 内容，实现数据价值与内容生产效率的双向提升。快手-W (1024.HK) 坐拥海量短视频内容数据资产，同时视频智能理解、AI 内容生成场景落地加速，K3 多模态能力开源将显著降低技术落地门槛，推动公司 AI 应用商业化提速。商汤-W (0020.HK)：计算机视觉 AI 龙头，同时布局大模型与 AIDC 算力基础设施，日日新大模型持续迭代，算力资源稀缺性凸显，直接受益于大模型行业需求扩容。

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。

