三周，四款开源大模型：中国AI的七月到底发生了什么

文 / 0pm.AI

四款开源模型密集登场，能力真在逼近——但“逼近”两个字要较真

7月17日，周五。

彭博打出标题：「又一个DeepSeek时刻」。

费城半导体指数盘中一度跌约**5.7%**。英伟达盘中跌近**5%**，午后收复大半失地，收盘跌**1.97%**。

市场已经没有复刻2025年1月27日的恐慌。那一天，英伟达收跌约**17%**，单日蒸发约**5890亿美元**市值。

但市场听懂了这次警报。

中国模型带来的压力，已经从“能不能用”，逼近“能不能和最强模型摆在同一张榜上比较”。

Motley Fool事后评价：「最初的反应似乎过度了。」

价格反应可以过度，能力变化却不能轻轻带过。

“三周四款”，先把日历校准

七月的密集发布，从腾讯混元Hy3开始。

7月6日，混元Hy3开放权重；DeepSeek V4采用官方口径，只能写“**7月中旬**”，不能把自媒体流传的7月15日当成确定发布日期。

7月16日，月之暗面在上海世界人工智能大会发布Kimi K3。

GLM-5.2则更早一步，已于**6月17日**开源。

严格按发布日期计算，四款模型并非全部在七月三周内首发。“三周”更准确地指向发布、评测和市场关注集中爆发的窗口。

这处日历误差不影响真正的变化：不同公司的模型，开始在同一阶段冲击前沿能力榜。

这张时间线最值得看的，是密度。

单个模型冲上来，可能是一次技术突进；不同团队连续交卷，说明供给开始成群出现。

四款模型从不同发布台推向同一张能力榜，零散突进开始形成集群

军团的成色，要看你能不能真正拿走

参数规模最抓眼球。

Kimi K3的主流口径为**2.8万亿总参数**，另有**2.7万亿**说法，激活参数尚未公布；混元Hy3为**2950亿总参数、每个token激活210亿参数**。

数字大，不自动等于模型强。

真正决定开源价值的，还有上下文、权重状态、许可证和使用成本。

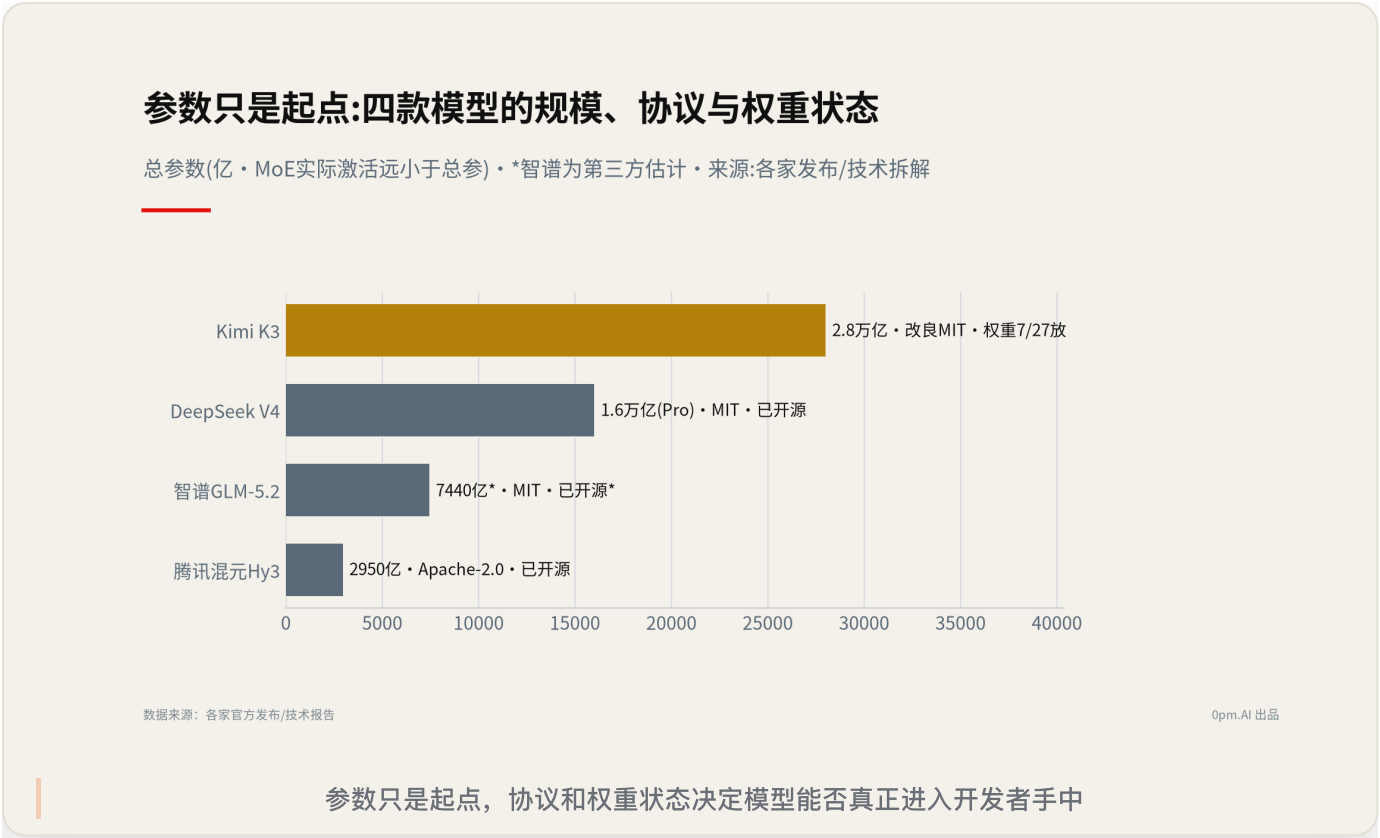
模型	发布或开源节点	参数与上下文	开源协议与状态	已有成本口径
混元Hy3	7月6日	2950亿总参数、210亿激活参数；256K上下文	Apache-2.0，可免费商用	—
DeepSeek V4	官方称7月中旬	—	MIT	—
GLM-5.2	6月17日已开源	官方未披露参数；第三方估计为7440亿总参数、400亿激活参数	MIT	—
Kimi K3	7月16日发布；权重计划7月27日放出	主流口径2.8万亿总参数，存在2.7万亿口径；100万token上下文	Modified MIT，带MAU商用条款	每百万输入token 3美元、输出token 15美元

表中的空白不拿猜测补齐。

混元Hy3采用Apache-2.0，在这轮对照中商用边界最宽松；GLM-5.2与DeepSeek采用MIT；Kimi K3虽然走开放权重路线，Modified MIT仍带有MAU商用条款。

Kimi K3还要多加一条时间注脚：截至7月16日发布时，完整权重尚未实际放出，Artificial Analysis仍将其标为Proprietary。它计划到7月27日才完成开放权重这一步。

开源的价值不在发布会，而在权重能否拿走、许可是否真能用、部署成本是否承受得起。



这张表把“开源军团”的成色拉开了：四款模型方向相同，开放程度并不相同。

1679分，把“逼近”写进了榜单

Kimi K3最硬的一仗，发生在前端编程盲测。

Frontend Code Arena中，Kimi K3拿到**1679分**，超过Claude Fable 5的**1631分**，排名第一。

七个前端领域里，它拿下六个第一，仅游戏领域排名第二。相比前代K2.6的第18名，这不是小步挪动。

它在一个明确、可重复比较的能力切面上，越过了领先闭源模型。

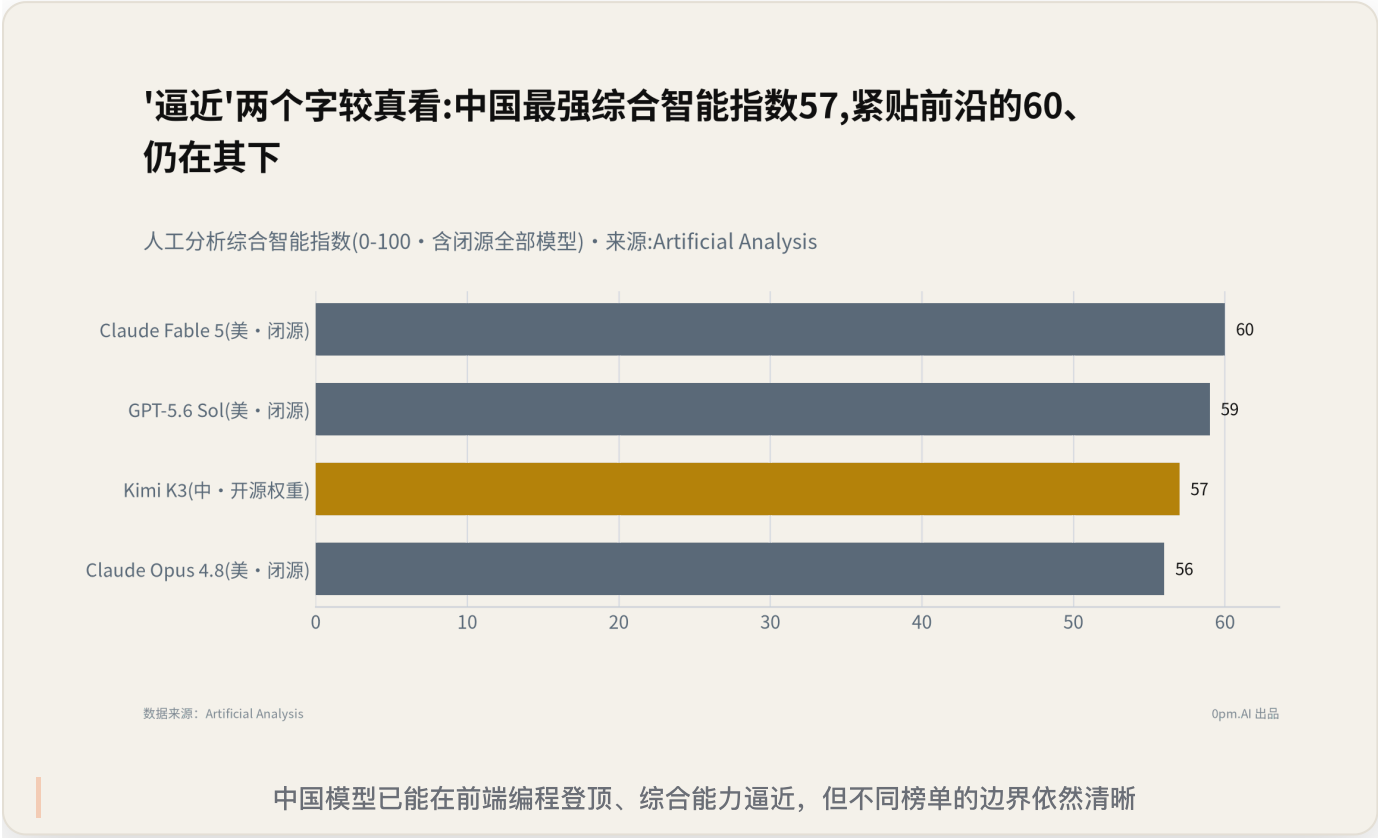
GLM-5.2提供了另一块证据。

富瑞转述的Artificial Analysis榜单截面中，GLM-5.2综合智能指数排到全球第三，编程能力全球第四，智能体能力全球第二。富瑞把它称为「中国AI里程碑」。

过去谈中国开源，常见优势是便宜、开放、部署方便。

七月出现的新东西，是能力坐标。

中国模型第一次密集站进前沿比较区，评价标准也从“国产模型里表现如何”，切换成“与全球头部模型还差多少”。



这张图要看的不是一个孤立冠军，而是两种逼近同时发生：细项开始穿透，综合榜开始贴近。

榜首要认，边界也得认

“霸榜前十”的故事早已流传过一次。

那是**2025年8月**的旧事，依据Hugging Face热度榜，衡量的是开源模型的热度与下载表现。它不能移到2026年7月，更不能拿来证明综合能力已经全面领先。

今年七月的证据更硬，也更窄。

Kimi K3在前端Code Arena以1679分登顶，证明的是前端编程能力；它在Artificial Analysis全部模型综合榜中约为**57分、排名第四**。

LMArena文本盲测里，Kimi K3为**1486±11分、排名第九**。

在这两张包含闭源模型的总榜前15名中，中国模型都只有Kimi K3一款。

GLM-5.2的“全球可用模型第一”，同样要保留“可用模型”这个限定。闭源的Claude Fable 5仍排在它前面。

榜单可以证明逼近，不能替任何模型颁发“全面第一”。

可也别因此把突破说轻了。

榜首只有一个，逼近却是一种产业状态：从无法比较，走到每个细项都必须较真。

这个七月真正值得认账的，并非中国模型把海外闭源模型挤出了总榜。中国开源已经从“便宜替代”，进入前沿模型的能力参照系。

细项能赢，综合榜能贴近，多个团队能连续交卷。

“全面追平”仍然没有证据。

继续把中国开源理解成低配版本，同样站不住了。

当开源模型已经能在细项登顶、综合榜逼近前沿，你还该用“开源等于低配”理解下一轮竞争吗?

来源：彭博、雅虎财经、Motley Fool、Artificial Analysis、LMArena、Frontend Code Arena、富瑞研报及新浪财经转述、量子位、Tom's Hardware、Decrypt、MarkTechPost、DataLearnerAI、Simon Willison。

风险提示：本文不构成任何投资建议，历史表现不代表未来，市场有风险，投资需谨慎。

0pm.AI 出品

三周四款开源模型·彭博喊fear·但'霸榜'至少混了三种口径

真图驱动　一手权威源　多源交叉核验

数据与口径

数据来源

各家官方发布/技术报告(智谱开放文档/月之暗面/DeepSeek/腾讯混元)、Artificial Analysis人工分析、Frontend Code Arena/LMArena、Bloomberg彭博、Tom's Hardware/VentureBeat/CNBC、智谱华章招股书/港交所/年报、新浪财经/财联社/36氪/量子位、斯坦福AI指数/CAISI(美方口径)

访问官网

扫码直达 0pm.ai

官网入口
0pm.AI