中国AI霸榜全球前十？先分清你说的是哪个榜

文 / 0pm.AI

热度榜、能力榜、开源专榜、含闭源总榜——四种口径，四个故事

前十席，中国模型拿走**9席**。

美国只剩Boson AI一款。

其中**7款**中国开源模型集中发布，英国《自然》将其称为“又一个DeepSeek时刻”。

这组数字是真的。

日期却是**2025年8月**。

把它贴到2026年7月，中国AI能力突破的新闻下面，一条旧热榜就被包装成了新总榜。

“霸榜”没有造假。

被压扁的是时间、参赛名单和评分规则。

九席是真的，新闻已经旧了

2025年8月，智谱GLM-4.5系列、月之暗面Kimi K2、阿里Qwen3系列与Wan-2.2等模型进入抱抱脸开源社群前十。加上已经在榜的腾讯混元与阿里模型，中国模型合计占据**9席**。

这是一张开源专榜。

它主要反映下载、点赞与趋势热度，衡量开发者是否愿意接触、试用和传播一款模型。

到了2026年7月，新的锚点换成了能力突破。

7月16日，Kimi K3在Frontend Code Arena取得**1679分**，超过Claude Fable 5的**1631分**。**7月17日**，彭博又用“又一个DeepSeek时刻”描述市场对中国模型进展的反应。

旧热榜与新能力榜，被一句“7款模型霸榜全球前十”缝在了一起。

两个时间点一旦重叠，去年“谁最热”就会被误读成今年“谁最强”。

“全球第一”至少用了三把尺子

榜单先要回答两件事。

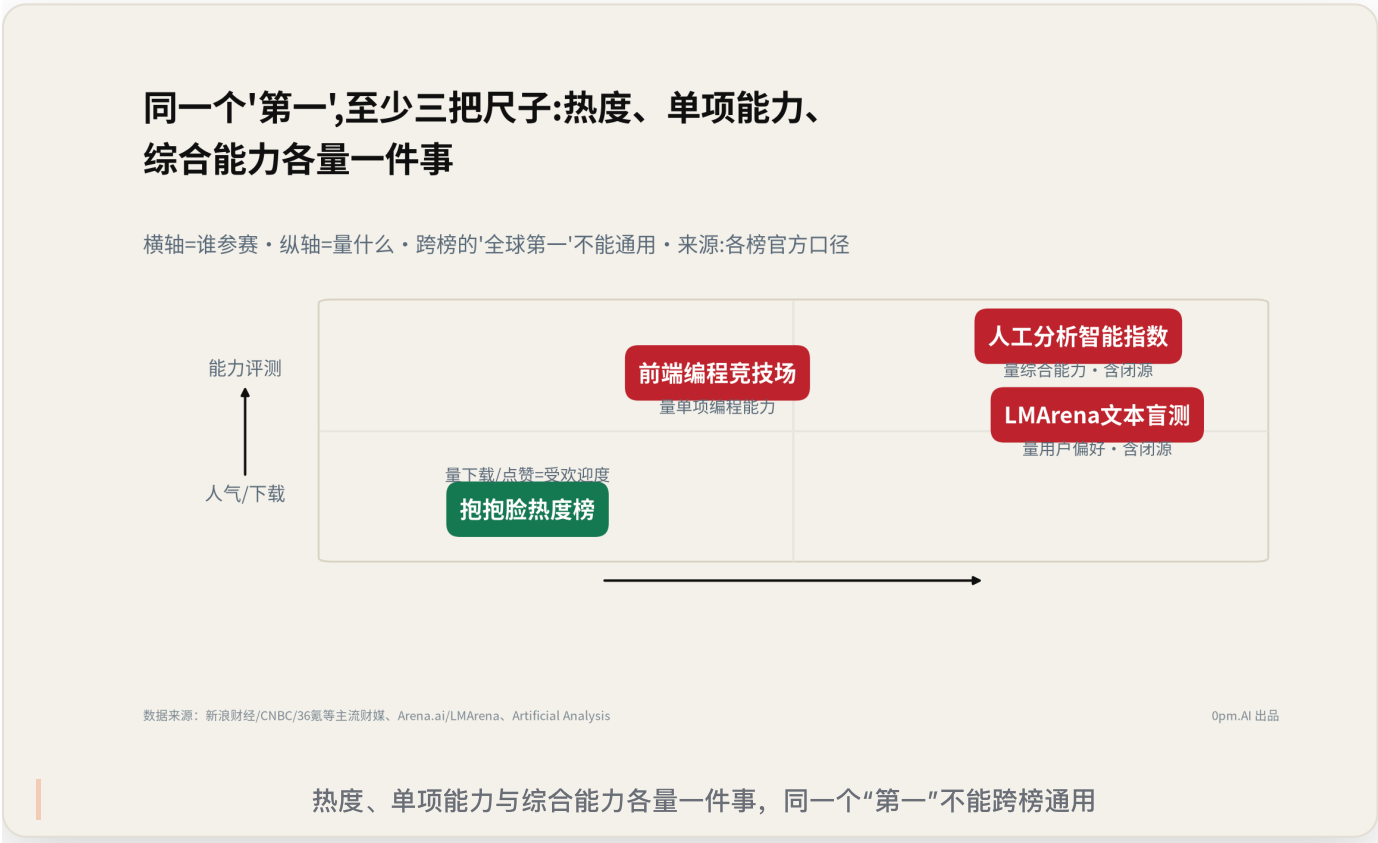
量什么?

让谁参赛?

更准确地说，热度榜、能力榜、开源榜、含闭源总榜并非四条平行跑道。它们是两条轴：一条决定测量对象，一条决定参赛范围。

传播把这个二维坐标压成了“全球榜”三个字。

榜单或平台	主要测量对象	参赛范围	第一名能够证明什么
抱抱脸开源社群榜	下载、点赞、趋势热度	开源模型	开发者关注度与生态渗透
Frontend Code Arena	前端编程盲测表现	参评模型	特定编程任务能力
Artificial Analysis	综合智能指数	包含闭源模型	多项基准汇总后的综合位置
LMArena文本榜	用户盲测偏好Elo	包含闭源模型	用户对回答质量的相对偏好



热度第一说明有人用。

编程第一说明一项能力强。

综合榜靠前，才说明模型在更宽的任务面上逼近前沿。

三者都值钱，价格完全不同。

同一款模型站在热度、单项和综合三把尺子前，刻度不同，名次自然不同

中国模型在不同尺子下都有真成绩。

OpenRouter官方博客与具名分析显示，开放权重模型能力前五中，中国模型占**4席**：GLM-5、Qwen3.5、Kimi K2.5和DeepSeek V4。美国模型仅NVIDIA Nemotron 3 Ultra进入这一组。

这是能力格局。

2025年那张前十占九席的抱抱脸榜，讲的是生态热度。

把两组数字放在一起，可以确认中国开源模型既获得了使用，也进入了能力前列。把它们说成同一张榜，认知增量反而消失了。

打开闭源模型，九席缩成一席

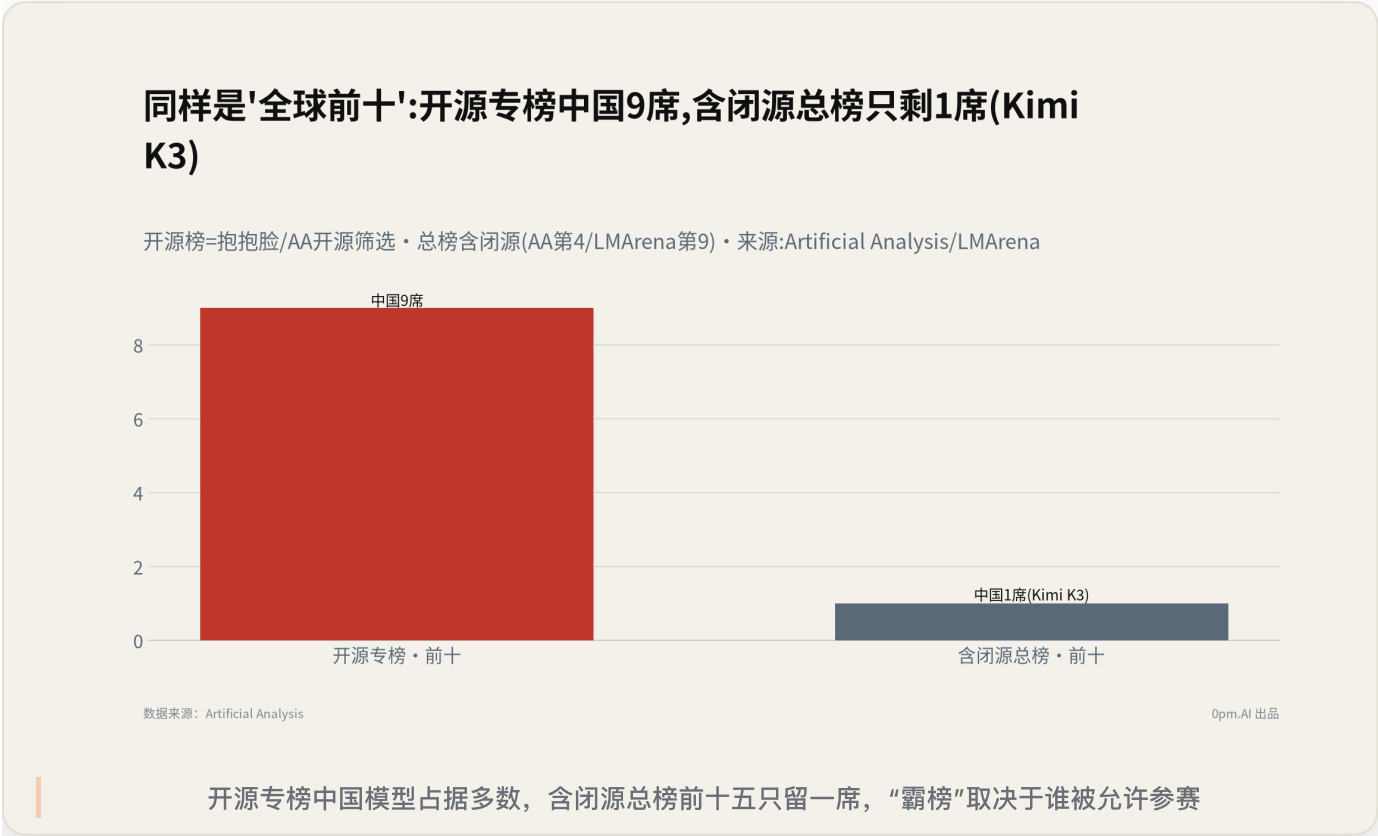
真正改变故事的，是参赛名单。

只看开源权重模型，中国可以占据前五中的**4席**，一度包揽开源榜前六。

把Anthropic、OpenAI、Google等闭源模型放回来，名次立刻重排。

Artificial Analysis给定榜单截点中，Kimi K3位列综合智能指数**第4**，前三名均为美国闭源模型。前十五名里，中国模型只有Kimi K3一款。

LMArena文本盲测中，Claude Fable 5排名第一，Kimi K3位列**第9**；前十五名同样只有这一款中国模型。



9席与**1席**都是真的。

前者证明中国已经成为全球开源生态的主导力量之一，后者提醒你，含闭源的综合前沿仍由美国模型占据多数席位。

榜单还是那块记分牌，参赛池换了。

还有一个容易漏掉的变量：榜单是快照。

富瑞在6月21日转述的榜单截点中，GLM-5.2位列Artificial Analysis综合智能指数全球第三，并称其为“中国AI里程碑”。Kimi K3发布后，新的给定截点又呈现出Kimi K3第四的格局。

不标日期的名次，保质期可能比一条热搜还短。

最后一层冷水，泼给跑分机制

即使日期、指标和参赛名单全部对齐，榜单也不能代替真实使用。

Kimi K3在Frontend Code Arena登顶，却在Artificial Analysis综合榜位列第四，在LMArena文本榜排第九。

同一款模型，单项第一、综合第四、用户偏好第九。

它们没有互相否定。

模型能力本来就不是一根刻度。

Kimi K3还有一个更尖锐的对照：AA-Omniscience基准中，其幻觉率由K2.6的**39%**升至**51%**。能力总分上升，并不保证每个可靠性指标同步改善。

可复现性也有时间差。

Kimi K3于**7月16日**发布，开放权重计划在**7月27日**放出。发布日只能通过API使用，Artificial Analysis因此将其标为Proprietary。许可采用Modified MIT，并附带MAU商用条款。

在完整权重放出前，“开放权重模型”的最终复现链条还没有闭合。

评测体系本身也会受污染。

人大高瓴团队相关研究所涉及的C-Eval基准题泄露争议，暴露了训练语料与测试题重叠的风险。论文《榜单幻觉》则用Meta案例讨论榜单失真；该案例并非针对国产厂商，却说明广泛使用的排行榜同样可能制造领先幻觉。

这些证据不能推出“中国模型靠刷分”。

它们只够推出一个更克制的结论：单次跑分越亮眼，越需要公开权重、独立复测与真实任务把它钉牢。

该打折的是叙事，不是进步

我的立场很明确。

“中国AI全球前十占九席”这句话，离开**2025年8月、抱抱脸、开源热度榜**三个限定词，就不再可靠。

“中国模型没有真突破”同样站不住。

开放权重能力前五占四席、Kimi K3登顶前端编程竞技场、进入含闭源综合榜第四，这些都是能力逼近前沿的硬证据。

中国AI最值钱的进展，并非在所有榜单上同时拿第一。

它已经能用不同模型，在开源生态、编程能力和综合智能三个层面反复进入全球前列。

爽文只需要一个“霸”字。

判断产业位置，需要先看尺子。

下一次再看到“中国AI全球第一”，你先问谁赢了，还是先问这张榜到底量了什么？

来源说明

来源：抱抱脸榜单同源报道（Yahoo财经中国香港、鉅亨网）；OpenRouter官方博客及Understanding AI具名分析；Arena AI旗下LMArena与Frontend Code Arena；Artificial Analysis官方排行榜；富瑞研报相关转述；Tom's Hardware；人大高瓴团队相关研究及36氪报道；论文《The Leaderboard Illusion》。

风险提示：本文不构成投资建议，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