戴着镣铐跳舞：中国为什么集体押注开源

文 / 0pm.AI

开源不是理想主义，是被算力管制逼出的最优解——是护城河也是天花板

7月6日，腾讯把混元Hy3的权重挂上网，Apache-2.0协议，最宽松的那一档，随便商用。7月16日，月之暗面在上海世界人工智能大会上放出Kimi K3，**2.8万亿参数**（另有2.7万亿口径），号称将成为全球最大的公开可下载模型——完整权重要到7月27日才放出，发布当天仍只是API可用，改良MIT协议。更早在6月17日，智谱已把GLM-5.2开源，采用MIT。

再往前数，DeepSeek——MIT。

把这个月的中国头部AI摆成一排，你会看到一件反常的事：混元、DeepSeek、GLM、Kimi，几乎清一色开源。而大洋对面，OpenAI和Anthropic把权重锁在保险柜里，一行都不给。

按常理，落后的一方应该攥紧手里每一张牌。为什么反过来了？为什么追赶的这一方，反而把家底摊在桌上？

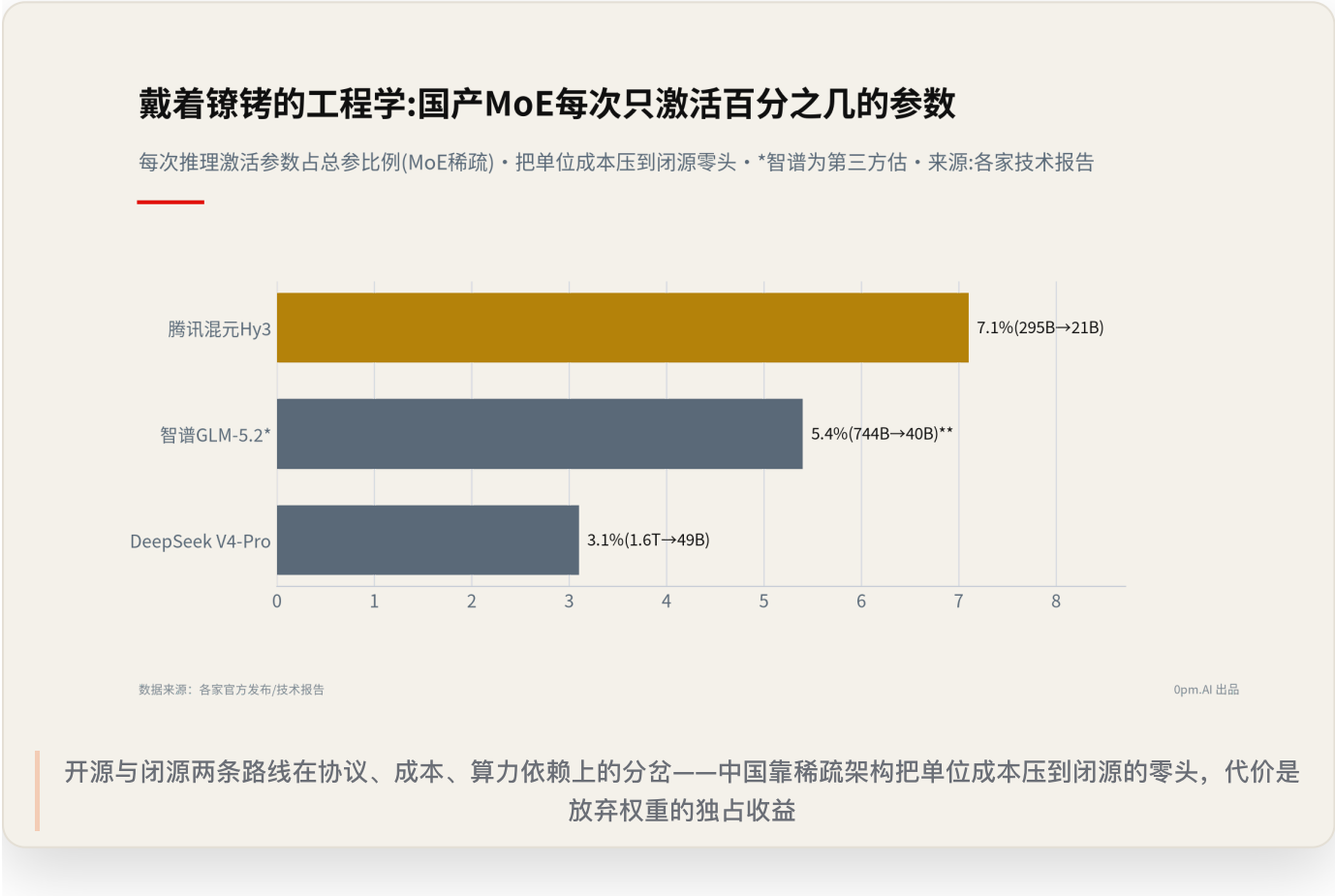
我的判断很直接：这不是理想主义发作，是被算力管制逼出来的一条活路。这条活路既是护城河，也是天花板。

开源在这里不是价值观，是唯一还能打的牌

先看成本这一面，因为镣铐的印子就刻在这里。

拿不到最先进制程、拿不到最高端的算力，中国团队只能在架构和工程上抠。混元Hy3是**2950亿参数**的稀疏模型，可每次前向只点亮约**210亿**；官方给的部署建议是8卡GPU（H20-3e那一级）就能跑，还配了三档推理力度，能省则省。这不是炫技，是没得选——每一分算力都得掰成两半花。

抠出来的结果，是价格。Kimi K3每完成一个任务的成本压到约**0.94美元**，输出token比上一代还少了两成。便宜到什么程度？在OpenRouter这类开发者平台上，最常用的前十个模型里，有六个是中国货；美国公司自己调用中国模型的token占比，最近一周超过**30%**，峰值逼近**46%**。



开源与闭源两条路线在协议、成本、算力依赖上的分岔——中国靠稀疏架构把单位成本压到闭源的零头，代价是放弃权重的独占收益

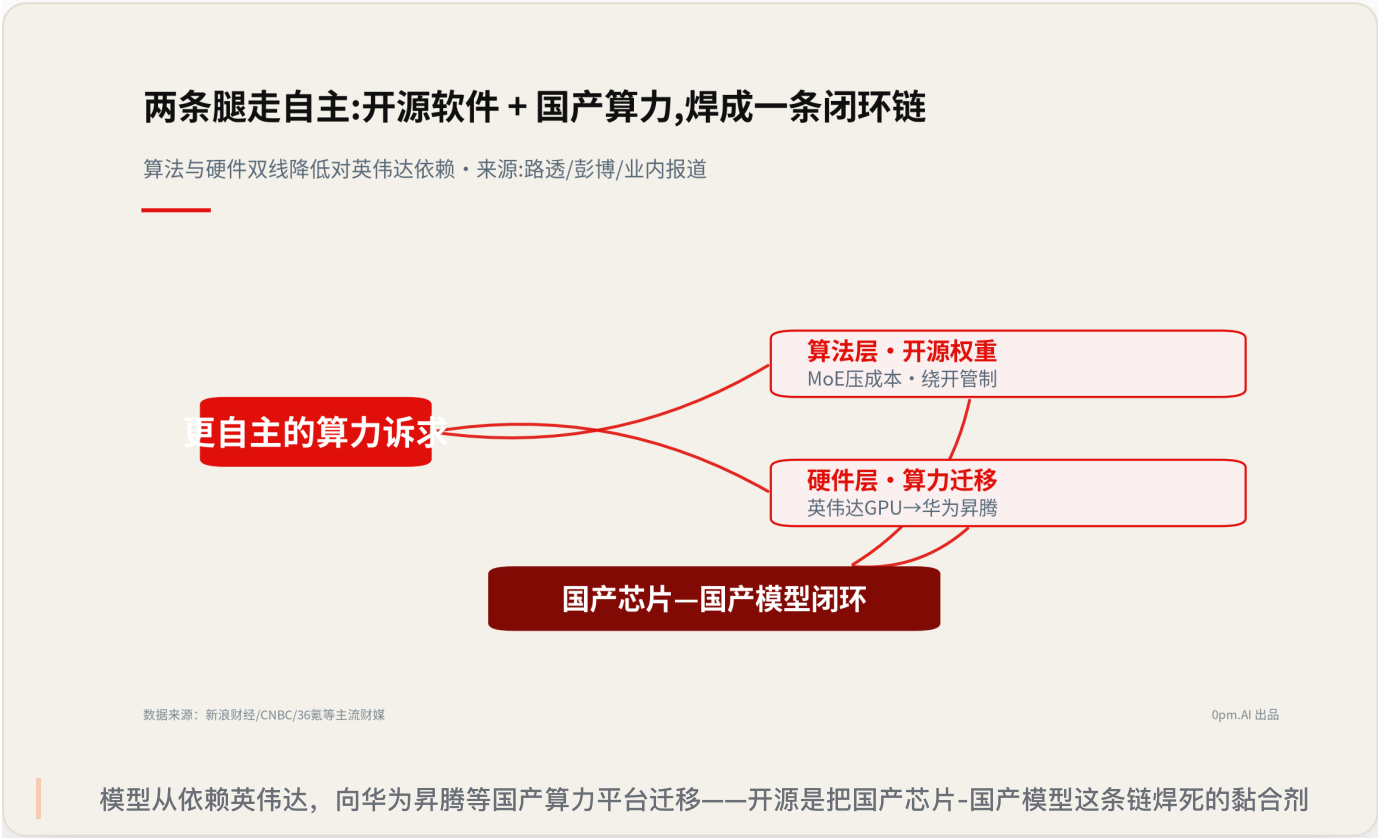
这里藏着第一个逻辑差：便宜从来不是慈善。它是被管制训练出来的工程本能。名义上叫"开源普惠"，翻译过来是——闭源那张高价牌，中国团队根本没有筹码去打，只能把开源、把低价，做成自己的武器。

把权重送出去，是为了换回一整个生态

成本是守，开源本身是攻。

闭源模型是一门生意，开源是一场圈地。你把权重免费放出去，全世界的开发者替你调试、替你适配、替你写教程，用着用着，标准就成了你的。这套打法绕开了管制最要命的地方：算力可以卡，但拦不住代码在GitHub上流动。

更狠的一步藏在算力迁移里。GLM-5.2开源当天，就适配了华为昇腾在内的**8大国产算力平台**——这意味着模型可以跑在非英伟达的芯片上。DeepSeek更进一步，据路透7月7至8日的报道，它正在自研推理芯片，项目约一年前启动，还处早期；同时从年初开始在内蒙古乌兰察布招数据中心的运维和交付岗，从轻资产研发往自建算力集群转。



> 模型从依赖英伟达，向华为昇腾等国产算力平台迁移——开源是把国产芯片-国产模型这条链焊死的黏合剂

一条国产芯片托着国产模型、模型又反过来喂养国产芯片的闭环，正在被开源焊起来。

软实力也一并收进口袋。GLM-5.2在前端编程盲测里拿下官方口径的「全球可用模型第一」，富瑞（Jefferies）给智谱的评价是「中国AI里程碑」。Kimi K3在Frontend Code Arena上得**1679分**登顶，压过Claude Fable 5的**1631分**，七个前端领域里六个第一，比上一代从第18名一口气蹿到榜首。

一个被管制的国家，用免费的代码，在别人的开发者社区里插旗。这盘棋走得很聪明。

> 一个人戴着沉重的镣铐，却在镣铐允许的半径里跳出了完整的舞步——约束划定了动作，也逼出了动作

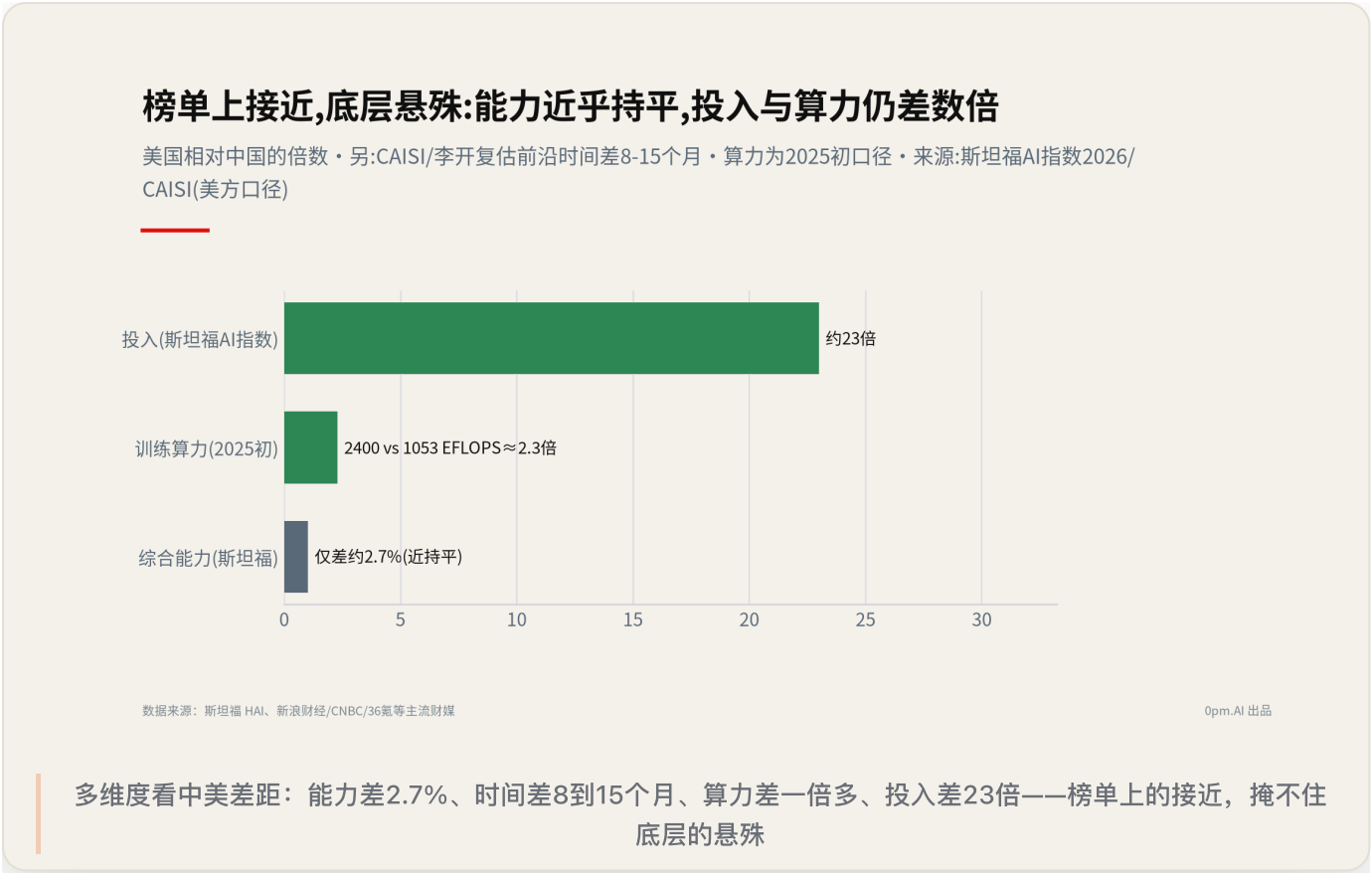
霸榜和登顶，是两件事

聪明归聪明，别被"霸榜"两个字晃了眼。这是最容易被民族情绪带偏的地方，也是本篇最要泼的一盆冷水。

中国开源模型确实在**开源专榜**上领跑。但把闭源模型一起拉进来看**总榜**，画风就变了：Artificial Analysis的综合智能指数前四是Fable 5、两款GPT-5.6 Sol，第四才是Kimi K3；前十五名里，中国模型只剩它一个名字。LMArena的文本盲测同样，Kimi K3排第9，前十五里也只有它一款。

开源榜第一，和含闭源的总榜第一，中间隔着的不是名次，是整个前沿。

差距有多大，几家不同立场的机构给了不同的尺子。美国CAISI（NIST下属）今年5月的评估认为，DeepSeek落后前沿约**8个月**（该报告含安全指控，作中立记录、不当定论）；李开复的公开估计更长，约**15个月**。斯坦福的AI指数则给了个更扎心的对照——顶尖中美模型的性能差已缩到约**2.7%**，可背后的投入差高达约**23倍**。再往上游看，2025年初的口径下，中美总训练算力约是**2400对1053 EFLOPS**。



> 多维度看中美差距：能力差2.7%、时间差8到15个月、算力差一倍多、投入差23倍——榜单上的接近，掩不住底层的悬殊

用2.7%的性能差去讲"追平"，却对23倍的投入差、一倍多的算力差闭口不谈，这是刷榜叙事最爱藏的账。

代价也开始显形。Kimi K3能力上去了，幻觉率却从上一代的**39%**升到**51%**（AA-Omniscience基准）——省算力、抢速度，是要还的。资本市场那边，富瑞一边喊里程碑，一边提醒：智谱的P/ARR约**94倍**，而Anthropic只有约**18倍**。榜单上的逼近，和估值上的透支，是同时发生的。

同一道墙，挡住了别人，也罩住了自己

现在可以把这道镣铐的两面拼起来看了。

护城河这一面是真的。开源换来了生态、换来了成本优势、换来了在别人平台上30%以上的token占有率，还把国产芯片和国产模型绑成了一条链。这些是闭源对手一时半会儿抢不走的地盘。

天花板那一面也是真的，而且没松动。最先进的制程和算力仍在管制之外，前沿的定义权还攥在别人手里。开源能让你跑得快、铺得广，却没法让你第一个撞线——因为撞线要的那台机器，你根本买不到。

护城河和天花板，本就是同一道墙的两面。它替你挡住了追兵，也把你自己罩在了下面。所谓"约束下的最优解"，重音永远在前半句：解是最优的，约束没有解除。

尾声

回头看"又一个DeepSeek时刻"，最深的一层从来不是谁又追上了谁。彭博7月17日把市场那阵恐慌命名为"Another DeepSeek Moment"，可真正在分岔的，是两种范式——一边是开源普惠、把成本打到地板，另一边是闭源前沿、把最强的东西攥在手里。中国被推上了前一条路，美国守在后一条路。

镣铐既划定了动作，也逼出了动作。问题不在于这支舞跳得好不好看——它已经足够好看了。问题在于：当有一天镣铐真的解开，这套为约束量身长出来的本事，是会长成翅膀，还是会发现，自己早已习惯了只在那个半径里起跳？

风险提示：本文仅作行业逻辑分析，不构成任何投资建议。历史数据不代表未来表现，市场有风险，投资需谨慎。文中涉及榜单、算力、估值等数据均标注了口径与时点，不同来源口径存在差异，请以官方披露为准。

来源说明：本文数据与事实来自Artificial Analysis、Arena.ai（LMArena）官方排行榜，富瑞（Jefferies）研报（经新浪财经转述），美国CAISI/NIST评估，斯坦福HAI《2026 AI Index》，及CNBC、路透社、Tom's Hardware、Decrypt、MarkTechPost、量子位、IT之家等公开报道。

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
