

大模型重压之下

五天里，中国大模型的三家代表公司先后遭遇不同维度的压力测试。

9月9日—13日，DeepSeek在新旧版本切换时，最终选择保留DeepSeek V4 Pro API(应用程序编程接口)，并下调DeepSeek V4.1 Flash定价。Kimi(月之暗面)一边启动企业合作伙伴计划，一边就网传的创始人及团队信息紧急辟谣。智谱则在9月13日公告中披露，通过配售新股和发行可转债，完成融资约50亿美元。三家公司曾共享技术神话的红利，如今进入现实考场，产品迭代要回应用户反馈，规模扩张伴随管理复杂度，高估值也需业绩验证。竞赛仍在继续，但重心移向运营。



9月9日—14日三家中国大模型公司关键事件时间线

时间	主体	关键事件
9月9日	Kimi(月之暗面)	面向全球招募7名“Wild Card”关键人才
9月10日	DeepSeek	发布V4.1 Flash并同步下调API价格,最初计划有序下线V4 Pro
9月10日	Kimi	启动企业合作伙伴“登月计划”
9月11日	DeepSeek	改口:9月14日之后继续提供V4 Pro的API调用服务
9月12日	Kimi	辟谣网传创始人及团队信息,称纯属虚构、已向公安机关报案
9月13日	智谱	公告配售新H股并发行零息可换股债券
9月14日	智谱	低开低走收报721港元,较前一交易日跌9.1%

摇摆频繁

一天时间，一家明星公司改变了主意。
9月10日，DeepSeek正式发布DeepSeek V4.1 Flash模型，一款全新模型结构系列中的最小尺寸模型，相应的API同步上线，计划有序下线DeepSeek V4 Pro模型。次日，DeepSeek又称在2026年9月14日之后继续提供DeepSeek V4 Pro的API调用服务。让公司改变主意的是用户需求。

9月14日，北京商报记者登录DeepSeek开放平台官网，DeepSeek V4.1 Flash发布通知、调价通知，以及继续提供DeepSeek V4 Pro的API调用服务的通知均以蓝色提亮。

根据DeepSeek介绍，随着DeepSeek V4.1 Flash上线，API价格同步下调。空闲时段百万Token输入缓存命中价格0.02元、输入缓存未命中价格1元，输出单价4元；高峰时段价格为空闲时段价格的2倍。

DeepSeek的定价像一口钟摆，近几个月摆动得格外频繁。

4月，DeepSeek发布V4系列模型预览版，1.6T参数版本为DeepSeek V4 Pro，284B参数版是DeepSeek V4 Flash。8月正式上线，API价格上调，且采用峰谷定价，空闲时段价格为高峰时段价格的一半。

当时，DeepSeek V4 Flash空闲时段百万Token输入缓存命中价格为0.05元，输入缓存未命中价格为1.5元，输出价格为4.5元。高峰时段，上述价格均上涨1倍，分别是0.1元、3元、9元。

9月10日正式发布DeepSeek V4.1 Flash模型时，DeepSeek依然采用峰谷定价，但下调了V4.1 Flash的定价。

空闲时段DeepSeek V4.1 Flash百万Token输入缓存命中中的价格是DeepSeek V4 Flash的40%，百万Token输入缓存未命中价格是DeepSeek V4 Flash的2/3，输出价格是DeepSeek V4 Flash的8/9。

多线探路

数月间，DeepSeek经历了涨价、分层、换代、降价，产品层面的调整速度，明显快于此

前的节奏。即便如此，这家大模型公司的API价格在国内外的性价比均有口皆碑。

“DeepSeek频繁调价，就是要找中间价，公司现在处在调价周期，还没有摸透价格弹性区间。”文渊智库创始人王超向北京商报记者谈了自己的猜测，“融资后，DeepSeek的风格有了明显的变化。”

定价是一种试探方式，招新是另一种。9月9日，Kimi通过官方账号在全球招聘7名关键人才，一个人可以成为一支队伍。次日，Kimi启动企业合作伙伴计划，与各行各业的IT服务商、集成商伙伴共建前线城市部署工程师队伍，深入客户现场，帮助企业把AI真正部署到核心业务中。有业内人士认为，这是Kimi发力B(企业)端的信号。事实上，Kimi早期尝试C端突破后，B端一直是公司的布局重点。

试探的同时伴随着噪声，同一周内，网传月之暗面创始人及团队相关信息，公司紧急辟谣。对于正处于IPO传闻中的Kimi而言，团队在扩张期的稳定性，与产品迭代同样重要。

全球最大大模型API聚合平台

OpenRouter的数据，提供了一个更直接的侧面。本周Kimi系列模型并没有进入OpenRouter最热门的模型前十，DeepSeek V4 Flash 0731、DeepSeek V4.1 Flash、DeepSeek V4 Flash 0423分别位列第四、第六和第七，智谱的GLM 5.3 Flash位列第三。

资本耐力

产品层的探路之外，是这门生意对资本的持续渴求。

9月13日，智谱发布了一则配售新股、发行可换股债券的公告，以每股714港元的价格配售最多2196.5万股新H股，较9月11日(即紧接配售协议签署日期前的交易日)的收盘价793港元折让约9.96%，较截至2026年9月11日(含该日)最后5个连续交易日约891.9港元的平均收盘价折让19.95%。此次配售所得约157亿港元。

根据公告，智谱于9月12日已同意发行201.4亿元债券，可按初始换股价每股892.5港元(可予调整)转换为新H股。较9月11日的收

盘价溢价约12.55%，较截至9月11日最后5个连续交易日约891.9港元的平均收盘价溢价约0.07%。

9月14日，智谱低开低走，收报721港元，较前一交易日下跌约9.1%。1月上市时，智谱发行价116.2港元，6月股价一度涨至2980港元，至7月股票解禁时，股价超过1500港元。

回到此次融资，配售新股叠加发行可换股债券两项合计融资约50亿美元。

募资用途写得很直接：所得款项净额约60%将用于下一代GLM基础模型及完全自训练(Fully Self Training)体系的研究，以及大规模训练、生产推理、算力资源及相关技术基础设施的部署及升级；约15%将用于支持本公司的业务拓展计划、战略投资及潜在并购；约25%将用于优化资本结构、补充日常运营所需营运资金及其他一般企业用途。

这次融资是北京社科院副研究员王鹏的关注重点，他告诉北京商报记者，“智谱上市后仍快速推进大规模再融资，直接印证了大模型行业重投入、长周期的底层属性。头部玩家已经从算力的使用者，转向算力基础设施的建设者，国内大模型的竞争正式进入拼长期资本耐力的阶段”。

从基本面看，已上市的智谱和MiniMax都需要证明造血能力。2026年上半年智谱实现营收约9.5亿元，同比增长399.7%，经调整净亏损约19.6亿元，同比扩大12.1%。2026年上半年，MiniMax总营收1.2亿美元，同比增长283.1%，经调整净亏损约2.9亿美元，同比扩大111.2%。

产品也好，业绩也罢，有关国产大模型的动态和格局还不到终局，9月14日有报道称，DeepSeek的CFO即将到岗，截至北京商报记者发稿，DeepSeek并未作出官方回应。

北京商报记者 魏蔚
图片来源:视觉中国

四九城里的

玩溜北京城

吃穿玩

您找我

玩

溜

北

京

城

100%原创视频

好玩的 好吃的 有趣的

带你避开旅游中的“坑”

有彩蛋哦~

有彩蛋哦~

扫码关注 玩溜北京城