SAMPLE · 样例节选

AI 模型真实支出份额观察
中国厂商 37 天内上升 3.7 个百分点

数据区间 2026-09-21 — 11-01 · 37 个数据日 · 每 8 小时采集一次 · 基于「有效运用值」口径

本报告不统计下载量、不谈跑分，只看一件事：**谁在真金白银地调用模型。** 按各家挂牌价折算的周窗口支出，37 天内从 **70.8 百万美元** 增至 **78.9 百万美元。** 在这块持续变大的盘子里，中国厂商的份额从 **25.0%** 走到 **28.7%**，美国厂商从 **74.0%** 退到 **70.4%。**

+3.7pp	-3.6pp	78.9 M
中国厂商份额变化	美国厂商份额变化	末日周窗口支出（美元）

一、份额走势：不是线性的，是三个台阶

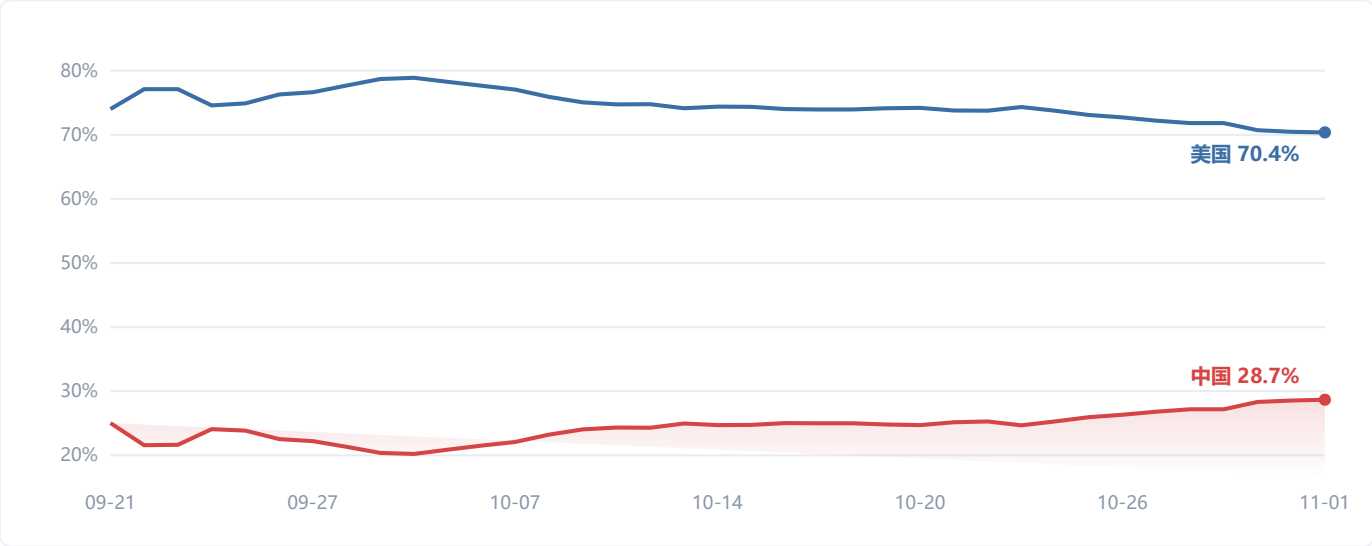


图 1　中国 / 美国厂商在周窗口支出中的份额。灰线为 10% 刻度。

二、谁在推动这条曲线

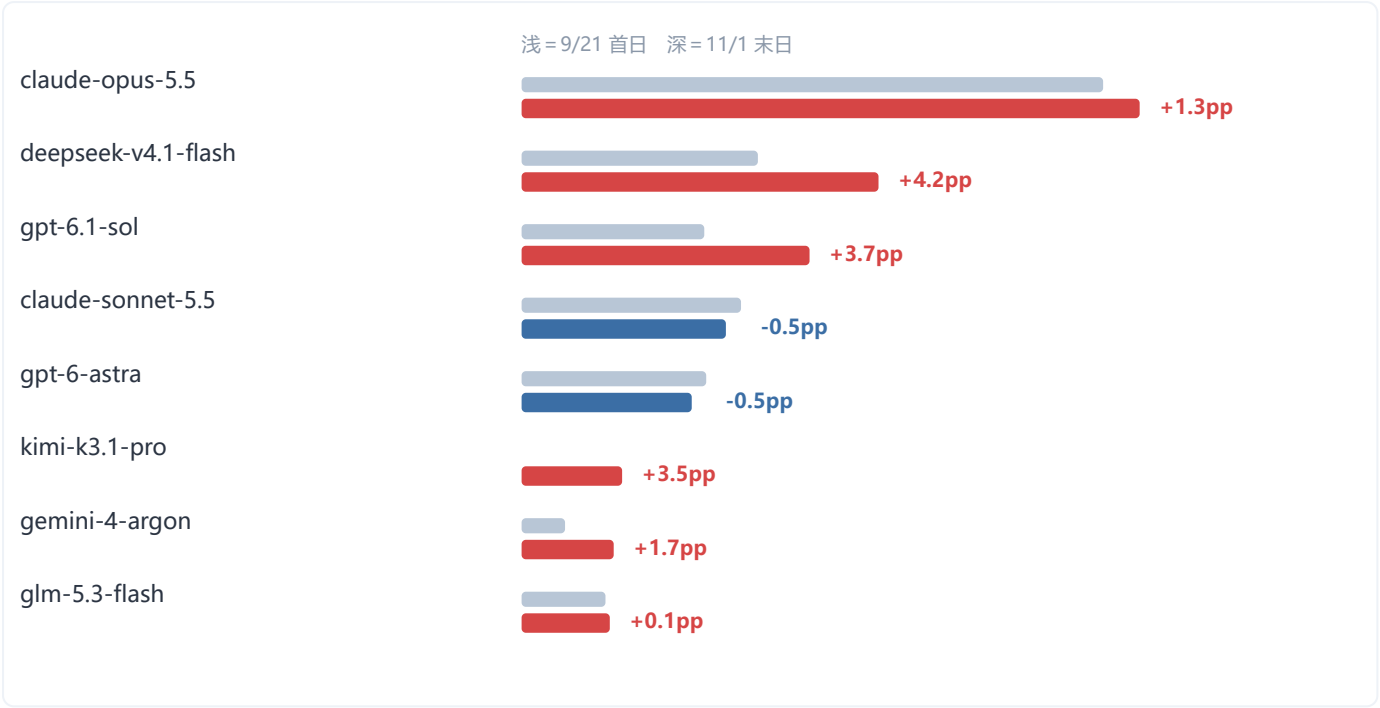


图 2　Top 8 模型族的份额变化（末日 vs 首日）。完整版含全部 295 个族。

三、核心结论

1. **位移来自两个具体模型，而且内部在剧烈换血。**DeepSeek V4.1 Flash 从 8.3% 升至 12.5%（+4.2pp），Kimi K3.1 Pro 从 0 增至 3.5%。两者合计 +7.8pp，是中国厂商净增量（+3.7pp）的两倍——反推可知，**其余中国厂商同期合计失血约 4.1 个百分点。**「中国厂商份额上升」是表象，真实发生的是内部排序重排。

2. **份额可以一个月内被同厂新品吃掉。**Claude Opus 5 从 3.8% 掉到 1.6%（-2.2pp），同期同厂 Opus 5.5 稳在 20% 以上——这不是市场丢失，是代际替换。模型迭代的最短替换周期已压缩到数周量级。

3. **总量在涨，而且更向头部集中。**周窗口支出 37 天增长 11%；头部 3 个模型族的合计份额从 35.1% 升至 44.3%（+9.2pp）。钱变多的同时，集中度也在上升——这对做模型选型的团队意味着：可选清单在收窄，而不是变宽。

口径与方法　支出（有效运用值，美元）= (min(输入 token, 请求数 × 64K) + 超出部分 × 0.10) × 输入挂牌价 + 输出 token × 输出挂牌价。超额部分按缓存读取价折算，以反映主流厂商价目表的 0.1 倍档。

样本与覆盖　37 个数据日的周窗口快照（每 8 小时采集，同日取最后一期）；剔除 Preview 型号以保持可比。可定价记录覆盖 **91.3%** 的 token 量（13834 / 23675 条记录），未覆盖部分为嵌入类与长尾小模型，不影响头部结论。

局限　聚合基于单一公开聚合平台的调用口径，反映「经该平台结算的真实支出」，不等同于厂商总收入；价格按公开挂牌价折算，不含企业折扣。