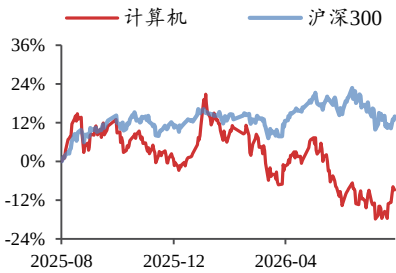


计算机

2026 年 08 月 09 日

投资评级：看好（维持）

行业走势图



数据来源：聚源

相关研究报告

《DeepSeek-V4-Flash 超高性价比重塑 Agent 竞争格局——行业点评报告》
-2026.8.3

《AI 应用商业闭环逐步跑通，垂直场景迎来价值重估——行业点评报告》
-2026.8.2

《新叙事新场景，看好 AI 应用及 AI Infra 投资机会——行业点评报告》
-2026.8.2

DeepSeek API 拟大幅提价，算力景气度高企

——行业点评报告

张越（分析师）

zhangyue1@kysec.cn

证书编号：S0790524090003

叶进新（联系人）

yejinxin@kysec.cn

证书编号：S0790126060053

● DeepSeek 年内多轮调价，国产大模型低价周期或将迎来拐点

截至 2026 年 8 月 7 日，DeepSeek API 定价历经多轮调整。4 月平台下调缓存命中输入价格，DeepSeek-V4-Pro 开启限时 2.5 折优惠；5 月优惠结束后定价降至原价 1/4，输入缓存命中低至 0.025 元/百万 tokens，整体降幅达 75%，带动其他厂商同步降价，国内大模型价格战持续升温。公司预告 V4 正式版上线后将引入峰谷定价机制，工作日高峰时段价格翻倍以调度算力资源；8 月 6 日公告将整体上调 API 定价，预计涨幅较大。伴随智谱、Kimi、豆包等头部厂商先后调整定价或收紧供给，行业逻辑正从价格战抢量转向训练端投入与模型能力比拼。

● 架构缓存优化支撑低价，产业逻辑切换是本次提价的核心动因

极致的技术优化是 DeepSeek 低价策略的核心支撑，低价失效、训练升级或是本轮涨价的根本动因。成本结构上，推理成本为核心支出，V4 系列采用 MoE 稀疏架构与混合注意力机制，V4 Flash 总参数 2840 亿但单 Token 仅激活约 130 亿参数，搭配 98%命中率的缓存折扣，大幅压低单位 Token 计算量，单任务成本较同智能水平海外模型低约 60%，基准测试成本仅为部分旗舰模型的百分之一；训练、运维、空闲算力等构成隐性全链路成本。当前模型调用量激增，V4 Flash 周调用量最高达 7.22 万亿 Token，OpenRouter 平台使用量排名第一，占据约 58%份额，持续高负荷下已出现服务拥堵、容量不足问题；同时大模型赛道赢家通吃，用户核心锚点是产品力，低价换量无法构筑长期壁垒。

● 全球调用量稳居首位，提价或将利好行业商业化生态优化

DeepSeek 当前稳居全球大模型调用量榜首，凭借头部规模优势，本次提价具备行业风向标效应，或将推动行业竞争导向发生转变。竞争格局方面，7 月下旬全球大模型调用量排名前五均为国产模型，其中 DeepSeek-V4-Flash 位列第一，DeepSeek-V4-Pro 升至第四，规模优势显著。提价的影响分长短周期来看：短期来看，价格上调将直接抬升单位 Token 调用成本，对重度依赖低价 API 的中小开发者与 AI 应用团队形成成本压力，也将推动市场侧更重视峰谷调度、缓存优化与多模型选型来对冲成本；长期来看，提价将推动行业走出价格战误区，引导厂商资源向训练端技术攻坚与模型能力升级倾斜，以产品硬实力构筑核心壁垒，推动国产大模型步入技术驱动的良好竞争周期。

● 受益标的

国产算力：寒武纪、海光信息、天数智芯、禾盛新材、沐曦股份、摩尔线程、中国长城、龙芯中科、芯原股份、品高股份、锐捷网络等、浪潮信息、工业富联、紫光股份、中科曙光、华勤技术等。

算力租赁：行云科技、宏景科技、利通电子、杰创智能、协创数据、航锦科技、润泽科技、东阳光、华策影视、亿田智能、平治信息、盛视科技、盈峰环境、熙菱信息等。

● **风险提示：**AI 模型竞争风险、AI 应用竞争风险、AI 模型发展不及预期风险等。

特别声明

《证券期货投资者适当性管理办法》、《证券经营机构投资者适当性管理实施指引（试行）》已正式实施。根据上述规定，开源证券评定此研报的风险等级为R3（中风险），因此通过公共平台推送的研报其适用的投资者类别仅限定为专业投资者及风险承受能力为C3、C4、C5的普通投资者。若您并非专业投资者及风险承受能力为C3、C4、C5的普通投资者，请取消阅读，请勿收藏、接收或使用本研报中的任何信息。因此受限于访问权限的设置，若给您造成不便，烦请见谅！感谢您给予的理解与配合。

分析师声明

本研究报告的署名人员具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告，并对内容和观点负责。本报告清晰地反映了署名人员的研究观点，所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。本报告署名人员保证他们报酬的任何一部分不曾与，不与，也将不会与本报告中的具体的推荐意见或观点有直接或间接的联系。

股票投资评级说明

	评级	说明
证券评级	买入（Buy）	预计相对强于市场表现 20%以上；
	增持（outperform）	预计相对强于市场表现 5%~20%；
	中性（Neutral）	预计相对市场表现在-5%~+5%之间波动；
	减持（underperform）	预计相对弱于市场表现 5%以下。
行业评级	看好（overweight）	预计行业超越整体市场表现；
	中性（Neutral）	预计行业与整体市场表现基本持平；
	看淡（underperform）	预计行业弱于整体市场表现。

备注：评级标准为以报告日后的 6~12 个月内，证券相对于市场基准指数的涨跌幅表现，其中 A 股基准指数为沪深 300 指数（北交所基准指数为北证 50 指数）、港股基准指数为恒生指数、新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）、美股基准指数为标普 500 或纳斯达克综合指数。我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比重建议；投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。

分析、估值方法的局限性说明

本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。

法律声明

开源证券股份有限公司是经中国证监会批准设立的证券经营机构，已具备证券投资咨询业务资格。

本报告仅供开源证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。本报告是发送给开源证券客户的，属于商业秘密材料，只有开源证券客户才能参考或使用。

本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他金融工具的邀请或向人做出邀请。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动，过往的业绩表现不应作为其日后表现的预示。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突，不应视本报告为做出投资决策的唯一因素。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。本公司未确保本报告充分考虑到个别客户特殊的投资目标、财务状况或需要。本公司建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。若本报告的接收人非本公司的客户，应在基于本报告做出任何投资决定或就本报告要求任何解释前咨询独立投资顾问。投资者应自主作出投资决策并自行承担投资风险，任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

开源证券在法律允许的情况下可参与、投资或持有本报告涉及的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。开源证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

本报告的版权归本公司所有。本公司对本报告保留一切权利。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

开源证券研究所

上海

地址：上海市浦东新区世纪大道1788号陆家嘴金控广场1号楼3层
邮编：200120
邮箱：research@kysec.cn

北京

地址：北京市西城区西直门外大街18号金贸大厦C2座9层
邮编：100044
邮箱：research@kysec.cn

深圳

地址：深圳市福田区金田路2030号卓越世纪中心1号楼45层
邮编：518000
邮箱：research@kysec.cn

西安

地址：西安市高新区锦业路1号都市之门B座5层
邮编：710065
邮箱：research@kysec.cn