下载日志已记录，仅供内部参考

CDSE 华龙证券

从模型突围看AI变局

投资评级：推荐（维持）

---计算机行业策略报告

华龙证券研究所 计算机行业
分析师：孙伯文
SAC执业证书编号：S0230523080004
邮箱：sunbw@hlzq.com
分析师：朱凌萱
SAC执业证书编号：S0230526030001
邮箱：zhulx@hlzq.com

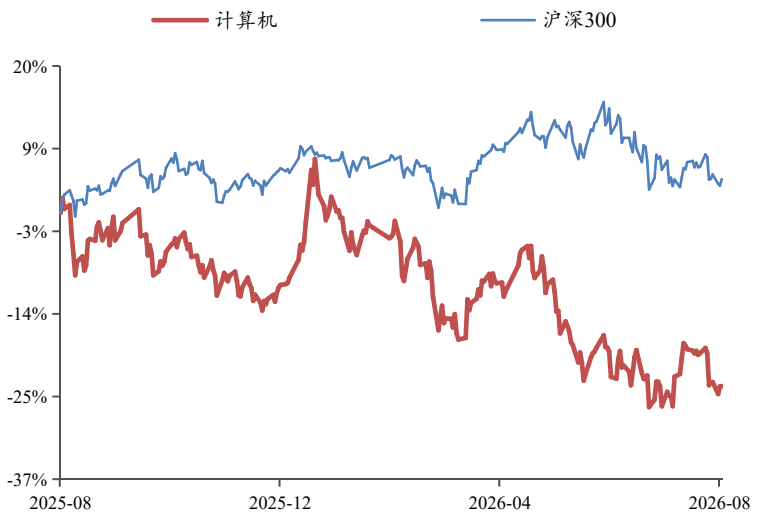
2026年08月27日

证券研究报告

请认真阅读文后免责条款

下载日志已记录，仅供内部参考

最近一年市场走势



相对沪深300表现（2026.08.27）　　（单位：%）

表现	1M	3M	12M
计算机行业	6.06	-12.81	-23.44
沪深300	-0.41	-6.42	3.99

相关报告

《国产模型有望转向高质量竞争，看好AI应用结构性重估机会—计算机行业周报》2026.08.12

《超节点与国产模型共振，看好国产AI投资机会—计算机事件点评》2026.07.28

《华为发布韬定律论文V2版本，国产模型调用量持续领先—计算机行业周报》2026.07.07

请认真阅读文后免责条款

下载日志已记录，仅供内部参考

报告摘要

复盘：截至2026年8月21日，申万计算机指数下跌15.92%，跑输沪深300指数15.68pct，在申万一级31个行业中位列第21位。近十年计算机行业机构持仓占比大致在 28%－37% 区间波动；2026Q1降至27.23%，为近年较低水平。

国产模型：性价比即竞争力。在达到同档智能水平的同时，GLM-5.3以更小的参数规模与更低的调用成本，显著降低了前沿模型能力的使用门槛。Kimi K3 max在BrowseComp 基准中以极低的单任务成本拿下了高分，成本比 Claude Opus 4.8等海外旗舰模型便宜一个数量级。总体而言，两者在性能追平国际顶尖闭源模型的同时，成本仅为对方的十几分之一到几分之一。我们认为，国产头部AI模型在tokens消耗量上已稳居全球第一梯队，部分国产模型凭借高性价比使得增长速率更加陡峭，更预示着持续的扩张潜力。反映出国产AI模型的发展趋势正从技术追赶转向市场引领。我们认为国产模型中，短期以调用量为支撑，中长期能够跑通企业API规模化变现、具备前沿模型迭代能力的标的将更可能获得成长性溢价与估值支撑。

国产算力：以规模换效率。2026年内，国产超节点解决方案密集发布。华为、阿里等超节点产品已进入量产交付环节；中科曙光、新华三等十余家国产厂商发布搭载国产算卡的超节点产品，表明当前国产AI算力的竞争维度已从单卡算力参数延伸至集群互联与系统级算力交付。我们认为超节点从样品走向规模展出释放了国产算力持续突破的明确信号，2027年有望进入国产超节点批量交付元年，国产服务器整机、交换机以及核心上游国产算卡供应商将核心受益。

AI应用：ROI提升逐步可视化。AI编程、AI办公、AI视频生成等AI应用赛道降本增效成果陆续验证，国产AI应用产品力提升，大厂加速争夺AI应用流量入口，商业化落地有望加速。

投资建议：展望未来，我们认为政策催化、国产替代需求以及AI驱动将共同推高计算机行业景气度，行业基本面拐点有望加速确立，建议关注景气赛道中的绩优标的，维持计算机行业**“推荐”**评级。建议关注个股：（1）大模型：智谱（2513.HK） 、MINIMAX-W（0100.HK）；（2）国产算力：寒武纪-U（688256.SH）、海光信息（688041.SH）、中科曙光（603019.SH）、浪潮信息（000977.SZ）、润泽科技（300442.SZ）、宏景科技（301396.SZ）；（3）AI应用：科大讯飞（002230.SZ）、卓易信息（688258.SH）、索辰科技（688507.SH）、能科科技（603859.SH）、万兴科技（300624.SZ）。

风险提示：国产算力建设不及预期；所引用数据资料的误差风险；AI应用落地速度不及预期；技术迭代速度不及预期；重点关注公司业绩不达预期；政策标准出台速度不及预期；科技股资金拥挤度高与波动放大风险；海外流动性超预期收紧风险。

请认真阅读文后免责条款

下载日志已记录，仅供内部参考

报告摘要

表：重点关注公司及盈利预测

股票代码	股票简称	2026/08/21	EPS（元）				PE				投资评级
		股价（元）	2025A	2026E	2027E	2028E	2025A	2026E	2027E	2028E	
000977.SZ	浪潮信息	76.35	1.64	2.43	3.08	3.89	42.7	31.4	24.8	19.6	增持
002230.SZ	科大讯飞	39.19	0.36	0.51	0.67	0.85	139.7	77.2	58.9	45.9	未评级
0100.HK	MINIMAX-W	347.20	-17.23	-1.81	-1.71	-1.56	/	/	/	/	未评级
2513.HK	智谱	1129.00	-12.03	-10.18	-9.72	-4.90	/	/	/	/	未评级
300442.SZ	润泽科技	64.38	3.00	1.98	2.52	3.16	17.6	32.5	25.5	20.4	未评级
300624.SZ	万兴科技	51.24	-0.47	0.12	0.40	0.57	/	420.0	128.7	89.5	未评级
301396.SZ	宏景科技	186.06	0.17	2.22	4.62	9.95	385.5	83.8	40.3	18.7	未评级
603019.SH	中科曙光	84.11	1.49	2.00	2.54	3.26	57.5	42.1	33.1	25.8	未评级
603859.SH	能科科技	38.02	0.93	1.24	1.49	1.62	45.2	30.7	25.5	23.5	增持
688041.SH	海光信息	246.21	1.10	2.00	2.74	4.21	204.0	123.1	89.9	58.4	增持
688256.SH	寒武纪	1035.00	4.93	11.00	17.19	30.83	275.0	94.1	60.2	33.6	增持
688258.SH	卓易信息	97.29	0.66	1.46	2.38	3.43	116.0	66.8	41.0	28.4	未评级
688507.SH	索辰科技	158.46	0.36	0.86	1.14	1.26	278.6	184.9	139.5	125.8	未评级

数据来源：Wind，华龙证券研究所（注：浪潮信息盈利预测来自华龙证券研究所；海光信息、寒武纪、能科科技2026-2027年盈利预测来自华龙证券研究所，2028年盈利预测来自Wind一致预测；其他重点关注公司盈利预测来自Wind一致预测）

请认真阅读文后免责条款

下载日志已记录，仅供内部参考

目录

1 复盘：2026年内行情回顾

2 国产模型：性价比即竞争力

3 国产算力：以规模换效率

4 AI应用：ROI提升逐步可视化

5 投资建议

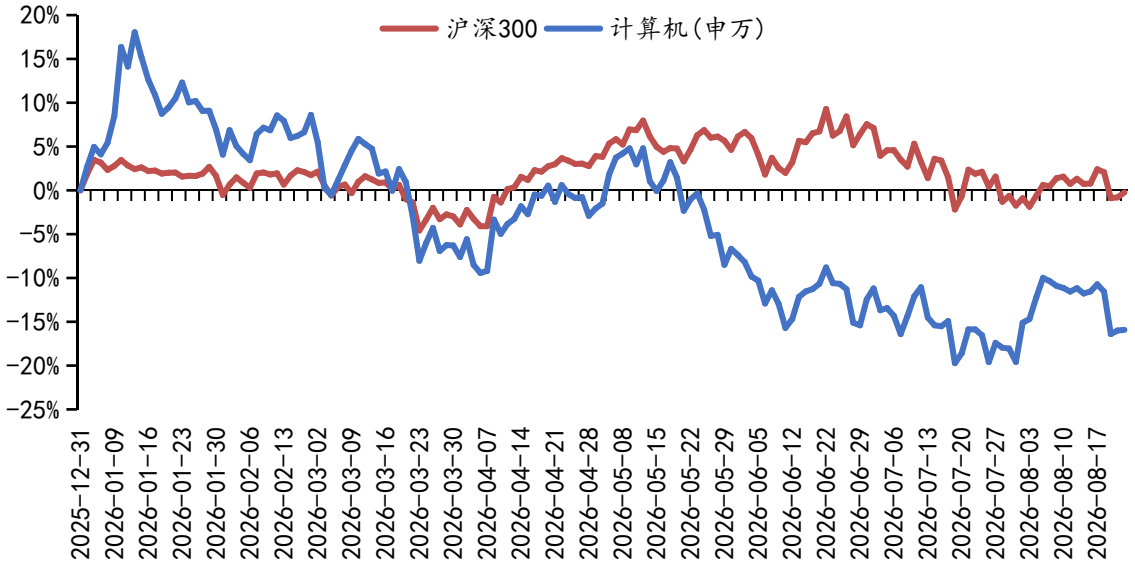
6 风险提示

下载日志已记录，仅供内部参考

01 复盘：2026年内行情回顾

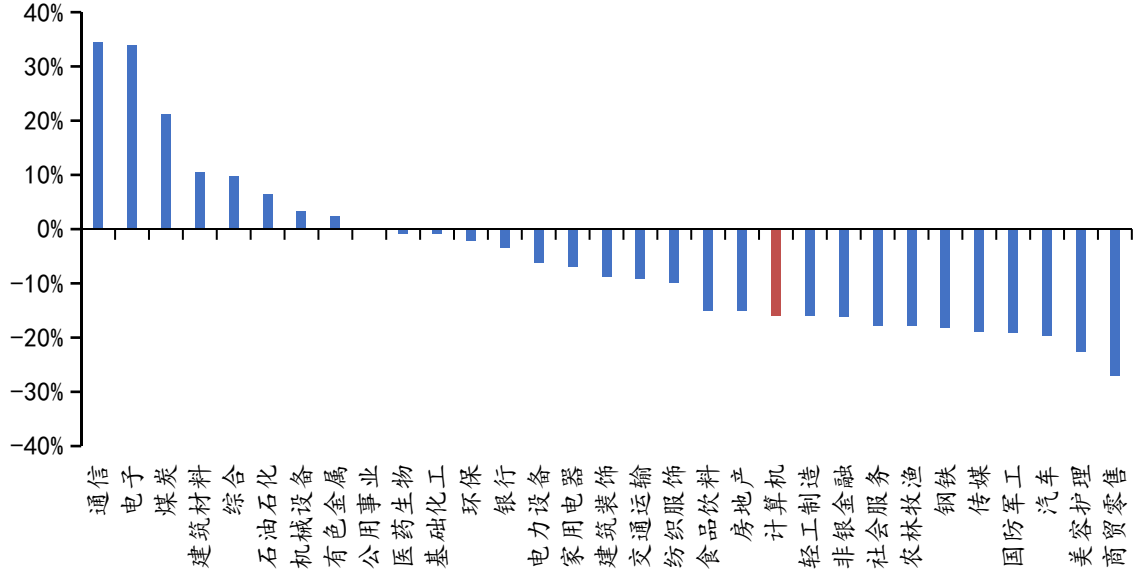
- 截至2026年8月21日，申万计算机指数下跌15.92%，跑输沪深300指数15.68pct，在申万一级31个行业中位列第21位。年初，“龙虾”爆火，计算机指数在云计算、AI应用板块拉动下上涨。后因3月份美伊冲突升级、主题投资情绪回落以及临近财报季市场避险情绪抬升，行业回调明显。年报与一季报披露后，AI硬件业绩普遍超预期，行情启动；同时软件股受“AI吞噬论”影响有所回调。6月中下旬，全球AI资产拥挤度过高，市场风险偏好快速下行，AI资产被挤兑，计算机行业持续调整至7月底，并于8月初反弹。

图1：2026年内计算机行业指数与沪深300走势对比（截至2026.8.21）



数据来源：Wind，华龙证券研究所

图2:2026年内计算机行业涨跌幅排名（截至2026.8.21）



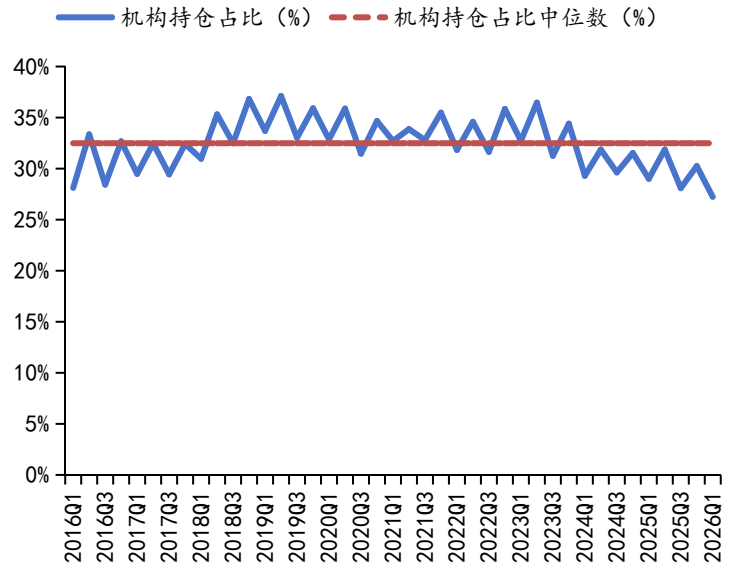
数据来源：Wind，华龙证券研究所

下载日志已记录，仅供内部参考

01 复盘：2026年内行情回顾

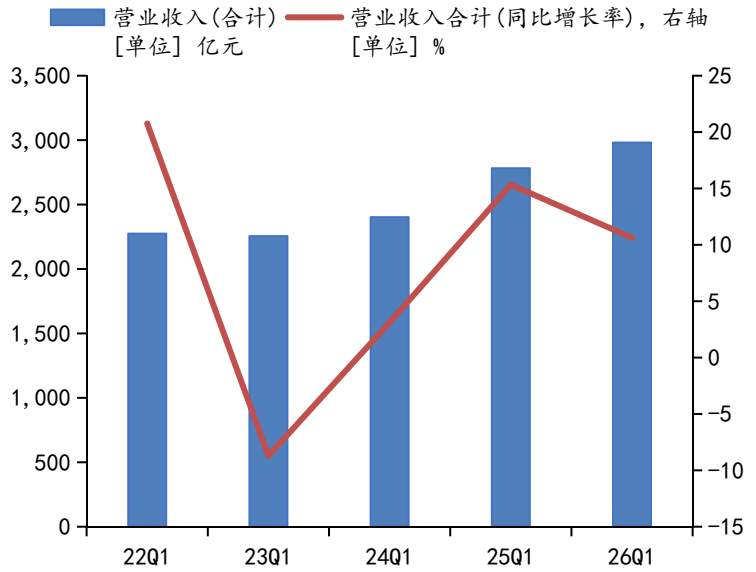
- 近十年计算机行业机构持仓占比大致在 28%－37% 区间波动，高点出现在 2019Q2（37.11%）和2023Q2（36.46%）；2024年以来整体中枢有所下移，多数季度在 29%－32% 之间；2026Q1 降至 27.23%，为近年较低水平。
- 单季度净利润高增长，行业复苏拐点进一步确认。计算机行业上市公司2026Q1单季总营收达2981.21亿元，同比增长10.62%。2026Q1单季归母净利合计值达87.61亿元，同比增长194.87%，单季度利润端实现连续两年同比高增长，呈现明显修复趋势。

图3：近十年计算机行业机构持仓占比



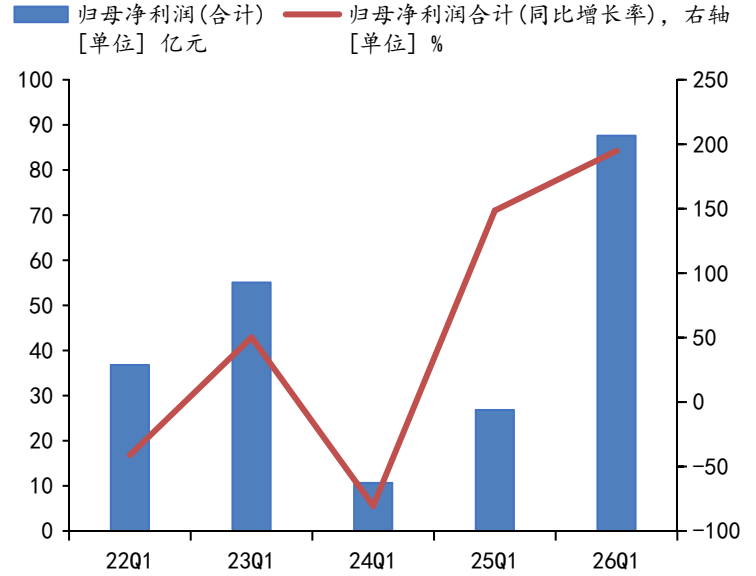
数据来源：Wind，华龙证券研究所

图4:计算机行业营业收入情况



数据来源：Wind，华龙证券研究所

图5:计算机行业归母净利润情况



数据来源：Wind，华龙证券研究所

下载日志已记录，仅供内部参考

目录

1 复盘：2026年内行情回顾

2 国产模型：性价比即竞争力

3 国产算力：以规模换效率

4 AI应用：ROI提升逐步可视化

5 投资建议

6 风险提示

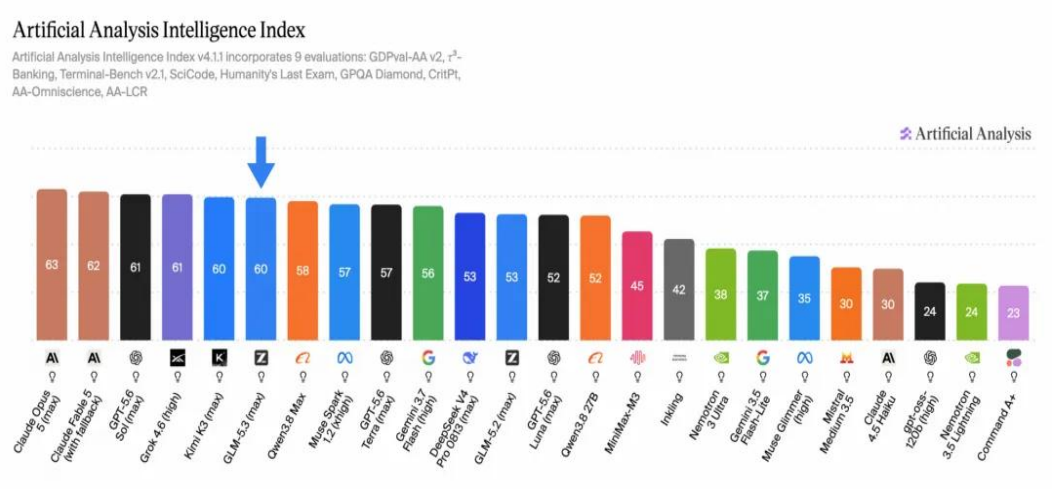
下载日志已记录，仅供内部参考

02 国产模型：性价比即竞争力

智谱、KIMI领跑国产模型，多项测评成绩赶超海外SOTA

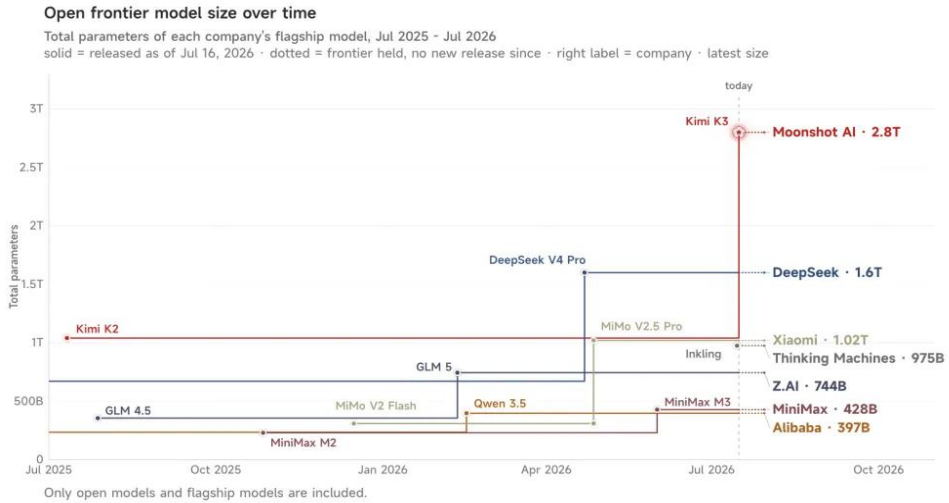
- 2026年8月14日，智谱推出GLM5.3。GLM-5.3是编程能力最强的开源模型，在内部自建体感评测中较GLM-5.2提升50%，在多项公开基准测试中取得开源第一，编程与智能体能力接近Claude Fable 5。
- 2026年7月17日，月之暗面发布新一代大模型Kimi K3，总参数2.8万亿，为目前全球参数最大的开源模型。性能层面，K3在官方测评中整体性能表现仅次于闭源模型Claude Fable 5 和GPT-5.6 Sol。
- 我们认为今年以来国产大模型进步显著，智谱GLM-5.3与Kimi K3在编程等核心场景的表现已紧追海外顶尖SOTA模型。标志着国产模型在技术实力和参数规模上正快速缩小与国际领先水平的差距，展现出强劲的竞争力。

图6：GLM5.3在AA综合智能指数榜单中进入全球前沿模型能力区间



资料来源：智谱，华龙证券研究所

图7：Kimi K3 是首个达到 2.8 万亿参数规模的开源模型



资料来源：Kimi，华龙证券研究所

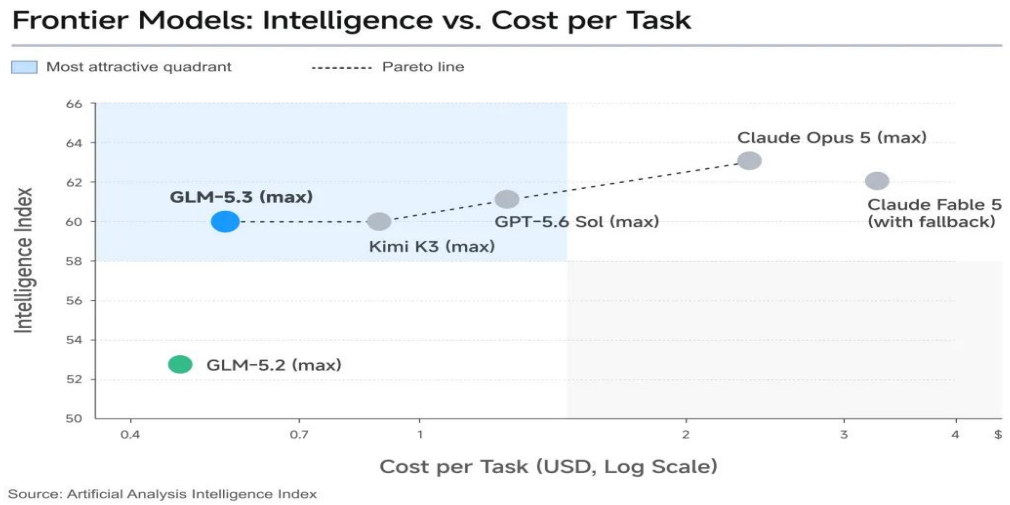
下载日志已记录，仅供内部参考

02 国产模型：性价比即竞争力

从低价策略向高质量竞争的切换

- 中国模型在性能接近甚至部分超越海外主流产品的同时，定价具备压倒性优势。在达到同档智能水平的同时，GLM-5.3以更小的参数规模与更低的调用成本，显著降低了前沿模型能力的使用门槛。Kimi K3 max在 BrowseComp 基准中以极低的单任务成本拿下了高分，成本比 Claude Opus 4.8等海外旗舰模型便宜一个数量级。总体而言，两者在性能追平国际顶尖闭源模型的同时，成本仅为对方的十几分之一到几分之一。
- 年内，国内模型厂商已陆续出现提价表现。其中，智谱2026年一季度API调用定价提升83%，调用量仍增长400%；8月6日，DeepSeek预告整体大幅上调API定价，出现明确的定价策略反转信号。根据OpenRouter数据，自8月3日至8月9日，DeepSeek-V4-Flash-0731周调用量位居平台全球第一，周调用量达8.83万亿Token，环比增长570%。V4-Flash已凭借出色的性价比实现模型调用量爆发式增长。高竞争力与强市场反馈使其具备提价底气。

图8：国产模型GLM-5.3与海外旗舰模型性价比对比



资料来源：智谱，华龙证券研究所

图9：国产模型Kimi K3与海外旗舰模型性价比对比



资料来源：Kimi，华龙证券研究所

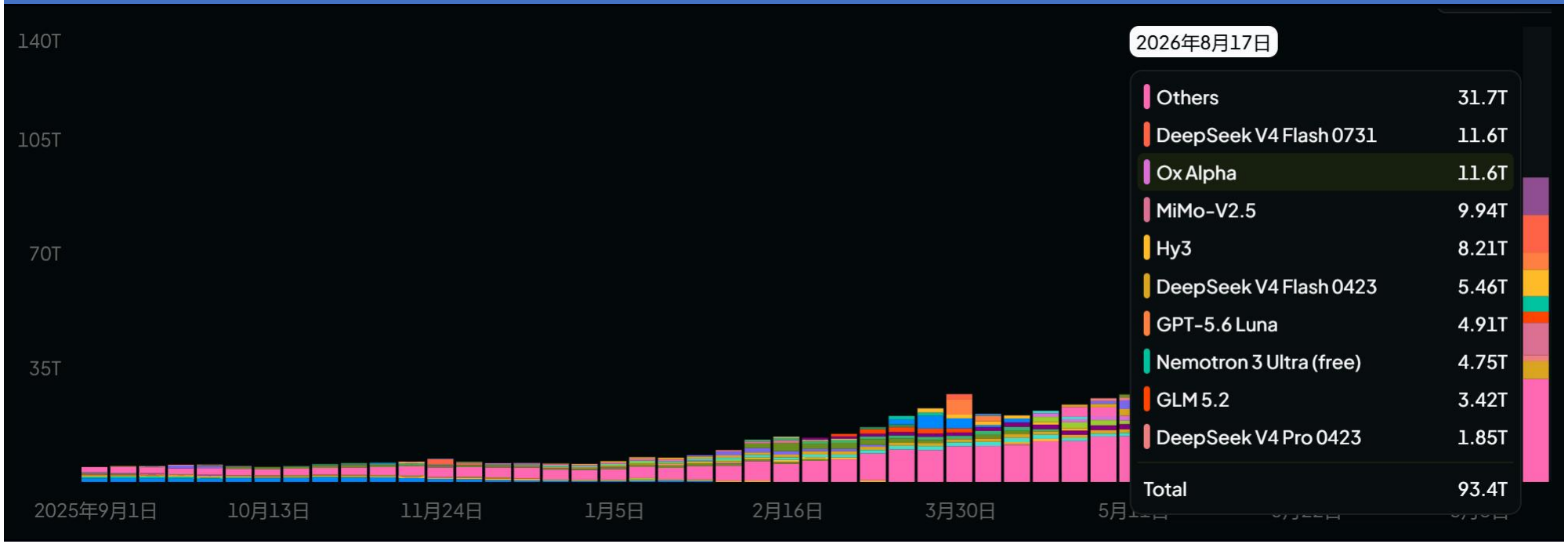
下载日志已记录，仅供内部参考

02 国产模型：性价比即竞争力

国产模型国际认可度持续提升

- 根据OpenRouter的数据，2026年2月，中国AI大模型的全球周调用量（Token消耗量）首次超越美国，标志着中国AI产业在应用落地层面取得里程碑式突破。最新数据显示（2026年8月17日-23日），在全球模型周调用量前九大模型中，国产模型占据七席，这七款模型的合计调用量为52.08T tokens，占总调用量的56.58%，反映出国产大模型在全球开发者实际使用中正获得越来越高的认可度和采用率。
- 国产头部AI模型在tokens消耗量上已跻身全球第一梯队，部分模型凭借突出的性价比实现了更快的增长斜率，也预示着它们仍有持续的扩张空间。折射出国产AI模型的发展正从技术追赶迈向市场引领，国际认可度显著提升。我们认为国产模型中，短期以调用量为支撑，中长期能够跑通企业API规模化变现、具备前沿模型迭代能力的标的将更可能获得成长性溢价与估值支撑。

图10：OpenRouter全球AI大模型排行榜



资料来源：OpenRouter，华龙证券研究所

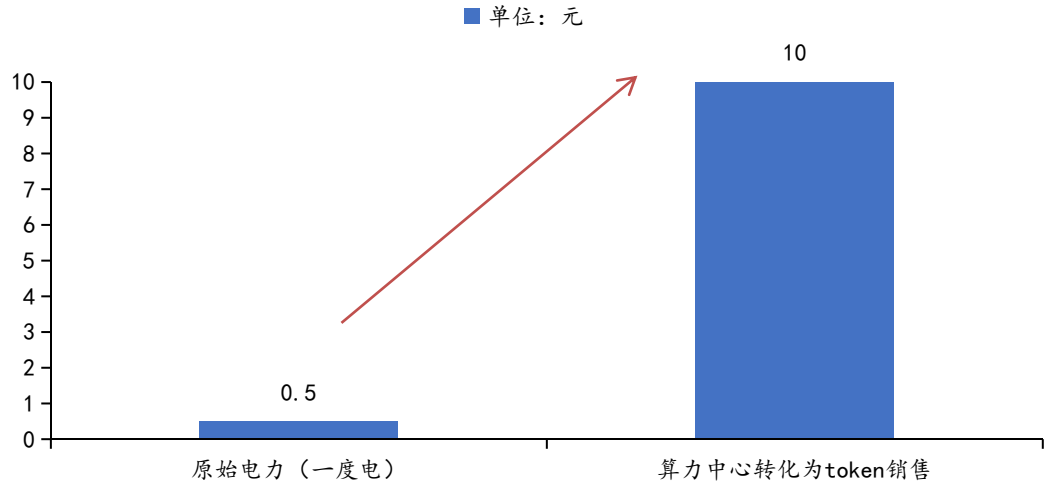
下载日志已记录，仅供内部参考

02 国产模型：性价比即竞争力

高性价比模型+电价优势，看好token经济开启正循环

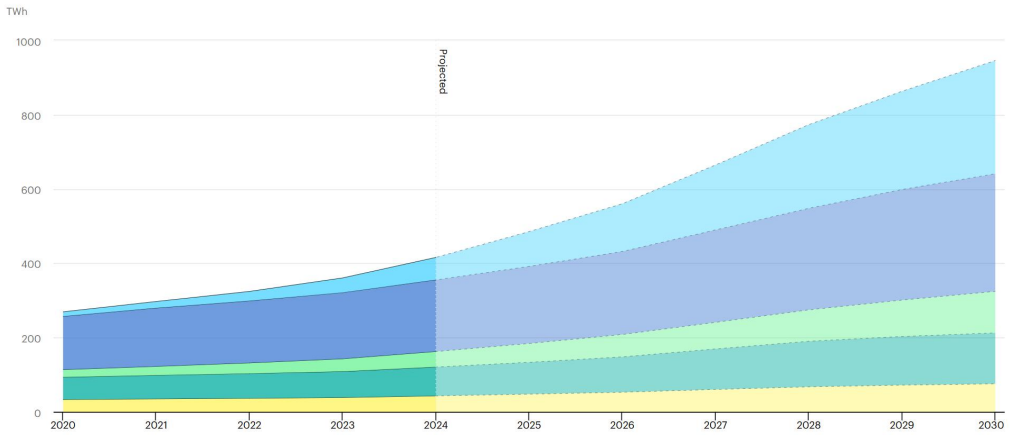
- 人工智能模型的训练和部署主要在数据中心进行。传统数据中心使用 10 兆瓦（MW）到 25 兆瓦（MW）的电力，而超大规模人工智能中心的需求可能超过 100 MW——相当于 10 万户家庭的年电力消耗量。2024 年，数据中心占全球电力需求的 1.5%。到 2030 年，根据国际能源署的基本情景预测，这一比例将上升到约 3%，全球数据中心的电力需求将超过两倍，达到约 945 太瓦时（TWh）。
- 国产绿电成本有优势，能够为tokens提供定价上的支持。与北美电力进行对比。由于AI训练和推理都是耗电巨兽，电力及算力折旧成本可占数据中心运营总成本的70%左右，一个国家的电网稳定性、电力成本（特别是绿色电力成本）决定了其Token的生产成本竞争力。从电费角度看，中国工业用电0.4-0.6元/度，美国0.8-1.2元/度，叠加国内电力输送和响应速度优势，国产token具备显著的电价优势支撑。根据专业人士测算，在国内新能源发电场景中，0.5元/度的原始电力通过算力中心转化为token销售，可以获得20倍的价值放大。

图11：电力-token转化



资料来源：央视财经，华龙证券研究所

图12：2020-2030年全球数据中心耗电量预测（最后更新时间：2025年4月）



资料来源：IEA，华龙证券研究所

下载日志已记录，仅供内部参考

02 国产模型：性价比即竞争力

Token经济中的涨价传导

- **云服务：**2026年一季度已经进入涨价实证期。其中，海外云先涨、国内中小云跟涨、随后头部云厂正式调价，模型厂商同步提价。本轮涨价涉及GPU云服务、存储、数据传输、模型调用以及安全类产品，主要是因为AI硬件成本上行叠加需求快速增加。建议关注涨价传导对云服务厂商收入弹性带来的放大效应。
- **算力租赁：**大模型从训练走向推理，算力需求将进入7x24小时持续消耗。高端机型供需仍处于不平衡状态，预计未来1-2年仍有较强价格支撑，建议关注算力租赁龙头厂商的毛利率及业绩兑现情况。

表1：云服务/模型涨价情况梳理

服务类型	涨价区间	涨价厂商和幅度
AI算力/GPU云服务器	普遍上调5%—30%，部分紧缺算力涨幅更高	金山云GPU云服务器（GPU高效计算型P6V）上调15-50%；阿里云平头哥真武810E等算力卡产品价格上调5%至34%；百度智能云AI算力相关产品服务价格上调5%至30%；腾讯云GPU云服务器价格上调5%；AWS EC2机器学习容量块价格上调15%；
大模型API/套餐订阅	部分模型API涨价	腾讯云混元HY2.0 Instruct输入价由0.0008元/千Tokens涨至0.004505元/千Tokens，涨幅约463%；智谱GLM Coding Plan订阅套餐价格整体上调30%起
存储/高性能文件系统	AI相关存储产品上调约30%	金山云KPFS文件存储部分上调30%-50%；阿里云CPFS智算版上调30%；百度智能云并行文件存储等上调30%；网宿科技宣布对象存储产品上调40%
数据传输/CDN	部分地区翻倍	谷歌云北美数据传输由0.04美元/GiB涨至0.08美元/GiB，涨幅100%；欧洲涨60%、亚洲涨约42%；网宿科技对部分CDN产品上调35%—40%
DDoS/安全网络产品	安全防御类产品同步上调	阿里云DDoS高防弹性95价格由100元/Mbps/月上调至150元/Mbps/月

资料来源：金山云，网宿科技，财联社，中国工信网，证券日报网，澎湃新闻，华龙证券研究所

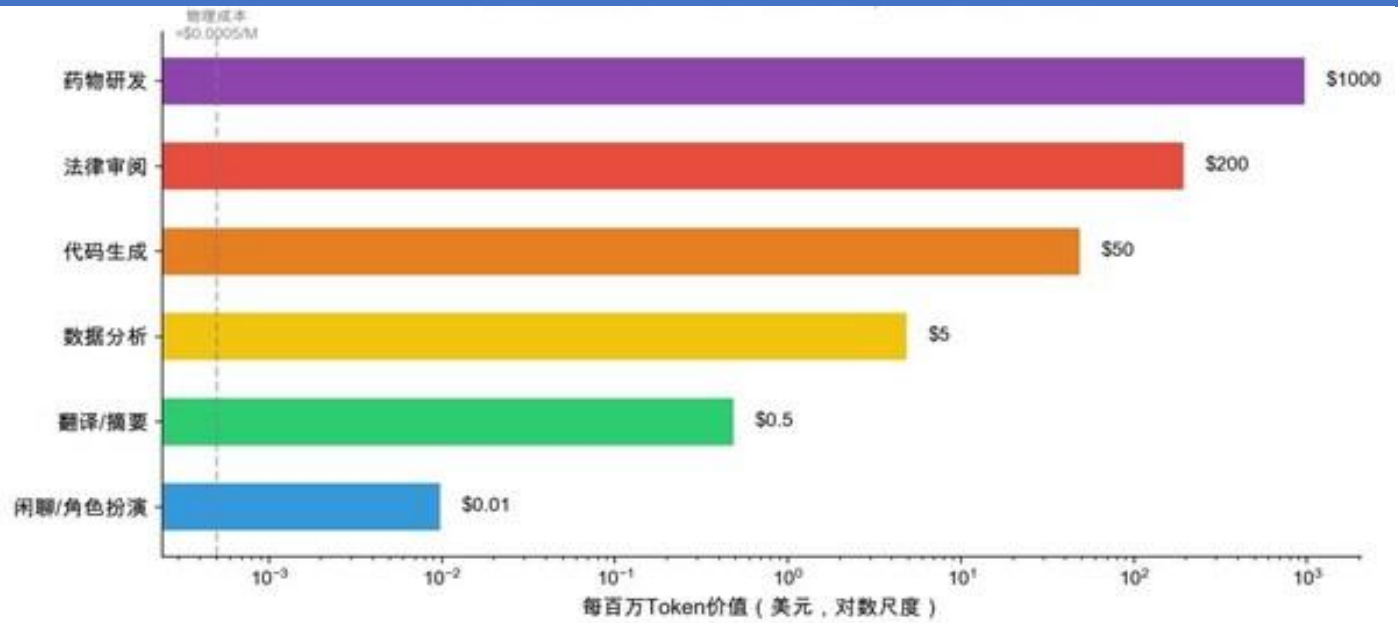
下载日志已记录，仅供内部参考

02 国产模型：性价比即竞争力

从传统业态到token工厂

- “Token工厂”可以理解为将电力、GPU、冷却、网络和推理优化技术进行打包，规模化、工业化地生产大模型可消费 Token 的新型智算基础设施。产业层面上，弘信电子已在无锡落地基于昇腾架构的超节点Token工厂，直接将算力集群与推理优化打包交付并按Token计费。
- 商业模式上，Token工厂突破了传统算力租赁按机柜或时长收费的资源售卖逻辑，转向以Token产量为计量单位的服务收费模式，收入来源包括基础算力服务费、按Token调用量计费、API接口服务费及基于应用效果的收益分成。其运营核心在于通过提升单卡日均有效Token产能、缓存命中率及推理框架效率，最大化每度电的Token产出比，从而降低单位成本并提升利润空间，有望成为token时代的价值转化核心载体。

图13:各垂类Token的价值光谱



资料来源：CXO UNION，Litowitz et al. (2026)；Bergemann et al. (2025)，华龙证券研究所

下载日志已记录，仅供内部参考

目录

1 复盘：2026年内行情回顾

2 国产模型：性价比即竞争力

3 国产算力：以规模换效率

4 AI应用：ROI提升逐步可视化

5 投资建议

6 风险提示

下载日志已记录，仅供内部参考

03 国产算力：以规模换效率

全球CSP资本开支继续增加，腾讯、阿里资本开支提速

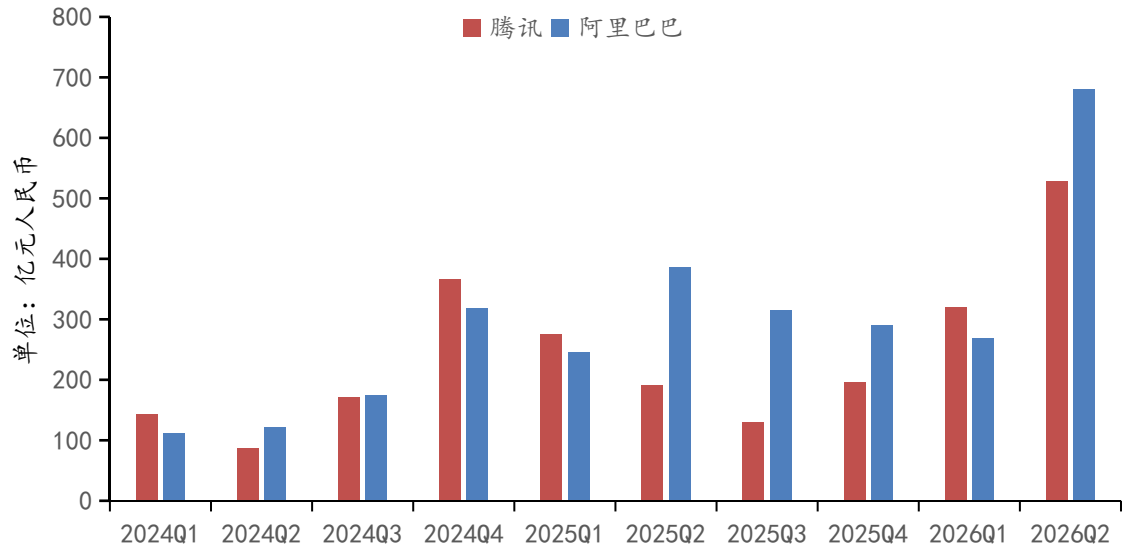
- 2026年7月，北美四大云厂最新2026年资本开支指引整体维持高位且多数上调，其中亚马逊、谷歌和Meta分别将指引上调至2200亿美元、1950-2050亿美元和1300-1450亿美元，仅微软因会计处理调整将指引下调至1750亿美元，但实际投资计划未变。海外云厂投资主要集中于人工智能芯片、数据中心、服务器和网络基础设施，显示出对AI等前沿技术持续扩张的坚定信心。
- 与此同时，中国主要的科技企业同样在加大资本投入。其中，腾讯2026年第二季度财报披露，当季资本开支达527.8亿元，同比增长176%，环比增长65%。阿里2026Q2资本开支单季度高达676.78亿人民币，同比增长75%。国内科技企业持续的投入态势与北美云厂商相呼应，共同指向全球科技行业在云计算和人工智能基础设施领域正在进行一场高强度的“军备竞赛”。
- 我们认为，当前全球科技巨头的资本开支正处于一个高强度、高增长的周期，针对AI等核心技术的投资热潮在未来几年仍将持续。

表2：北美四大云厂最新资本支出及指引

公司	日期	资本开支指引
微软	2026年7月29日	2026年资本支出指引从年初的1900亿美元预期下调至1750亿美元（但实际投资计划不变，下调资本开支数字是由于会计处理的调整）
Meta	2026年7月30日	2026财年资本支出预期区间下限上调50亿美元，上限维持不变，预计为1300亿至1450亿美元，此前预计为1250亿至1450亿美元
谷歌	2026年7月22日	2026年全年资本支出指引从1800-1900亿美元上调至1950-2050亿美元，并明确 2027年资本开支将显著高于2026年
亚马逊	2026年7月30日	2026年资本支出指引从2000亿美元上调至2200亿美元

数据来源：各公司业绩说明会，观点网，智通财经，21世纪经济网，华龙证券研究所

图14:阿里、腾讯资本开支情况



数据来源：各公司财报，Wind，华龙证券研究所

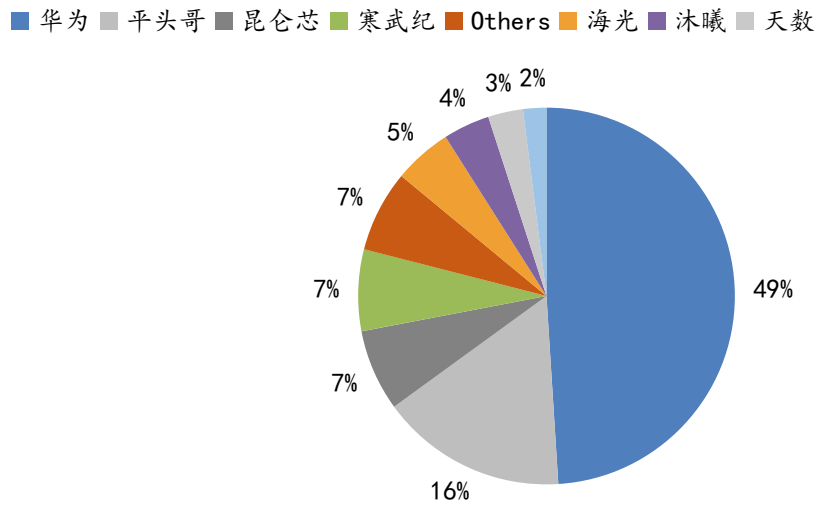
下载日志已记录，仅供内部参考

03 国产算力：以规模换效率

国产替代推进，中国芯如何“分蛋糕”

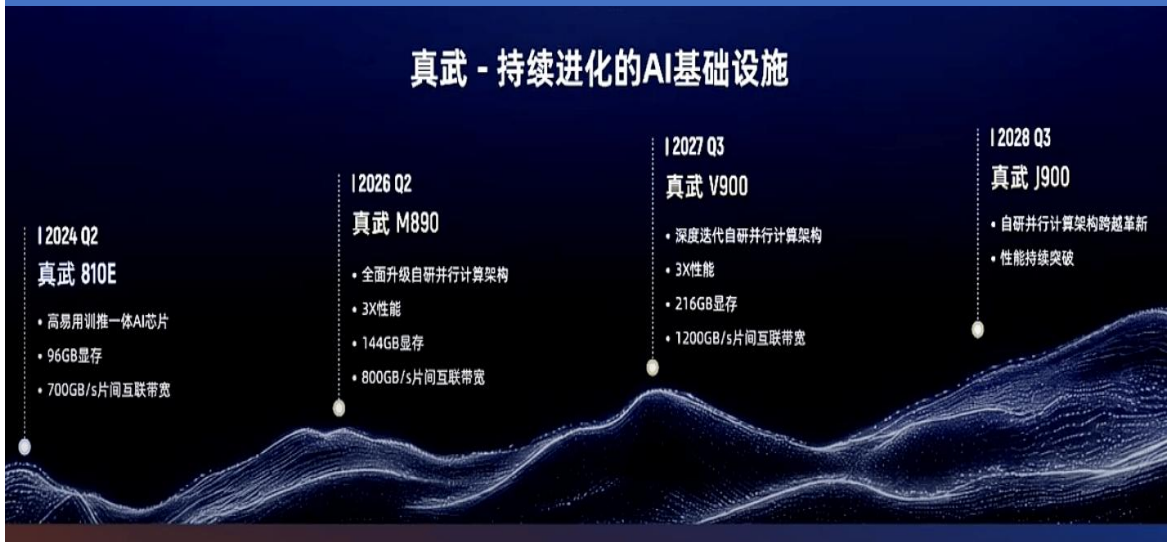
- **中国CSP资本开支与国产替代同步推进。**据IDC数据，2025年中国市场AI加速卡市场中，中国本土芯片厂商合计出货约165万张，整体市场份额提升至41%，国产替代进入加速通道。国产份额由华为昇腾、寒武纪、海光等通用AI加速芯片，以及阿里平头哥、百度昆仑、腾讯自研ASIC两条路线共同构成，互联网大厂依靠自研芯片显著提升内部算力自给比例。以阿里平头哥真武技术路线图为例，2026-2027年，真武系列芯片将保持每年性能提升三倍，显存与片间互联带宽同步提升。
- 在美方出口管制持续、国内政府采购清单倾斜、HBM等配套产业链成熟的共同推动下，中国AI芯片市场正在以远超预期的速度完成从"英伟达主导"到"国产并行"再到"国产主导"的格局重构。据TrendForce预测，受美国芯片出口限制与中国本土芯片产能扩张双重影响，AMD和英伟达等在中国高端AI芯片的市场份额2026年将降至10%，国产芯片则占据90%，同时预计中芯国际和上海华虹制造的芯片未来几年出货量年复合增长率可达50%。

图15：2025年中国芯片厂商AI加速卡在国内数据中心适配数量占比



数据来源：IDC，半导体产业观察，华龙证券研究所

图16:阿里平头哥技术路线图



资料来源：南方都市报，华龙证券研究所

下载日志已记录，仅供内部参考

03 国产算力：以规模换效率

国产算卡+国产大模型适配加速，国内AI产业闭环正在形成

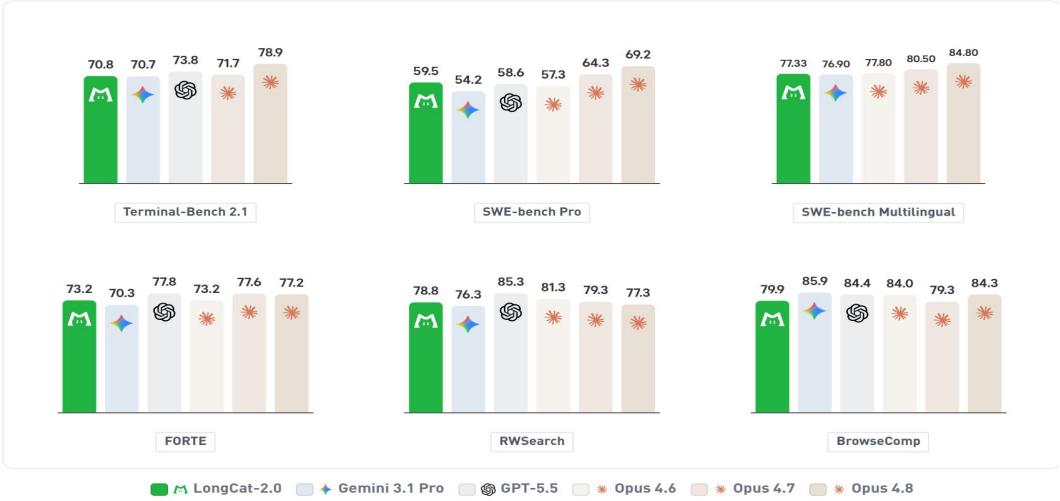
- **推理端快速适配，算力与模型互利共赢。**DeepSeek于今年4月宣布，受限于高端算力，目前DeepSeek-v4-Pro的服务吞吐十分有限，预计下半年随着昇腾950超节点批量上市后，Pro的价格会大幅下调。海光信息于7-8月分别宣布海光DCU完成国产模型Kimi K3、Minimax H3适配。
- **不止步于推理端，国产算力迈向训练端。**美团于今年6月底发布并开源新一代万亿参数大模型 LongCat-2.0，其完整训练流程与大规模部署均全部使用国产算力集群，验证国产算力支持模型训练、推理闭环的可行性。我们认为随着国产算力在训练端的拓展，国产卡的覆盖范围将加速扩大。

图17：昇腾950超节点实现DeepSeek-V4高吞吐、低时延推理



资料来源：华为，华龙证券研究所

图18：基于国产算力集群训练和大规模部署的LongCat-2.0测评得分



资料来源：Longcat，华龙证券研究所

下载日志已记录，仅供内部参考

03 国产算力：以规模换效率

国产超节点：以集群补足单卡代差

- 国产算力芯片单卡性能相较英伟达 H100、B 系列等高端产品存在明显差距，叠加国内晶圆制程（中芯国际为主）相较海外 3/5nm 先进制程存在代差。在此背景下，国内发展超节点技术是以集群形态补单芯片性能不足的短板，通过系统级能力实现弯道超车。

- 以华为超节点为例，截至 2025 年，华为CloudMatrix 384 已进入工程化交付阶段；CloudMatrix 384 在算力总量、HBM 总容量和 Scale-Up 域规模上采取了更大规模的组织方式。相较于NVIDIA 的优势主要体现在单卡密度、工具链成熟度和标准化交付能力，华为更偏向以规模换效率。华为于今年上线的Atlas 950超节点，相比英伟达同样将在明年下半年上市的NVL144，Atlas 950超节点卡的规模是其56.8倍，总算力是其6.7倍，内存容量是其15倍，达到1152TB；互联带宽是其62倍，达到16.3PB/s。

图19：昇腾950超节点（Atlas 950 SuperPoD）真机



资料来源：华为，华龙证券研究所

表3:华为CloudMatrix 384与NVIDIA GB200 NVL72对比

维度	华为 CloudMatrix 384	NVIDIA GB200 NVL72
计算芯片数量	384 个昇腾 910C NPU	72 个 Blackwell GPU
系统算力	约 300 PFLOPs（BF16/FP16）	180 PFLOPs（FP16/BF16）
Scale-Up 域规模	384 NPU	72 GPU
系统 HBM 容量	49.2 TB	13.8 TB
系统 HBM 带宽	1229 TB/s	576 TB/s
Scale-Up 带宽	269 TB/s（双向）	130 TB/s（双向）
物理形态	16 柜超节点	单柜 NVL72
系统总功耗	约 559 kW	约 145 kW
软件栈	CANN / HCCL / MindIE / MindSpeed	CUDA / NCCL / TensorRT

资料来源：《超节点技术体系白皮书》，华龙证券研究所

下载日志已记录，仅供内部参考

03 国产算力：以规模换效率

国产超节点：以集群补足单卡代差

- **2026年内，国产超节点解决方案密集发布。**华为、阿里等超节点产品已进入量产交付环节；沐曦、新华三等十余家国产厂商发布搭载国产算卡的超节点产品，表明当前国产AI算力的竞争维度已从单卡算力参数延伸至集群互联与系统级算力交付。我们认为超节点从样品走向规模展出释放了国产算力持续突破的明确信号，2027年有望进入国产超节点批量交付元年，国产服务器整机、交换机以及核心上游国产算卡供应商将核心受益。

表4：国产超节点产品梳理

厂商	超节点产品	芯片基础	大致落地进度	厂商	超节点产品	芯片基础	大致落地进度
华为昇腾	昇腾384超节点/ 昇腾950超节点	自研昇腾卡	昇腾384已商用落地 **750多套**，是国内唯一训练出SOTA模型的超节点，覆盖互联网/运营商/金融/政务等行业；昇腾950（1024卡）于2026年WAIC完成真机首秀，预计**2026 Q4上市**	沐曦	曦景S600超节点	自研沐曦GPU	2026年7月17日，沐曦股份正式发布新一代AI超节点产品——曦景S600。在系统交付方面，曦景S600联合产业生态伙伴完成整机及系统级适配，实现芯片、整机、软件平台的协同优化
阿里云	灵骏真武M890超节点	自研平头哥真武M890芯片	2026年8月12日，阿里云宣布灵骏真武M890超节点实例正式上线，并在内蒙古乌兰察布地域开售。该超节点是国内首个成功运行超2万亿参数大模型的超节点形态算力，KimiK3和Qwen3.8-Max已通过该实例提供服务。	浪潮信息	元脑SD200超节点	国产GPU	2025年发布，实现64路本土GPU芯片高速互连，已实现商用
百度智能云	天池超节点（P800）	自研昆仑芯P800	天池256与天池512超节点即将上市，未来，还将基于昆仑芯M系列，推出千卡、四千卡超节点	新华三	H3C UniPoD S80000系列超节点	兼容多厂商	2026年5月发布，实现从32卡到1024卡灵活扩展，最大可扩展至16384卡互联规模，全面适配智算中心、企业专属算力、行业大模型平台等多样场景。
中兴通讯	OEX 超节点	多国产芯片异构	中兴通讯联合曦智科技、壁仞科技、沐曦股份、燧原科技、天数智芯等合作伙伴，凭借“基于OEX+dOCS架构的国产高性能Matrix超节点”项目斩获2026世界人工智能大会SAIL之星奖项	摩尔线程	MTT C256 超节点	/	实现并柜扩展可达 256 卡极速互联，将卡间通信延迟压缩至亚微秒级。

资料来源：各公司官网，网通社，财经网科技，科技快报，华龙证券研究所

下载日志已记录，仅供内部参考

目录

1 复盘：2026年内行情回顾

2 国产模型：性价比即竞争力

3 国产算力：以规模换效率

4 AI应用：ROI提升逐步可视化

5 投资建议

6 风险提示

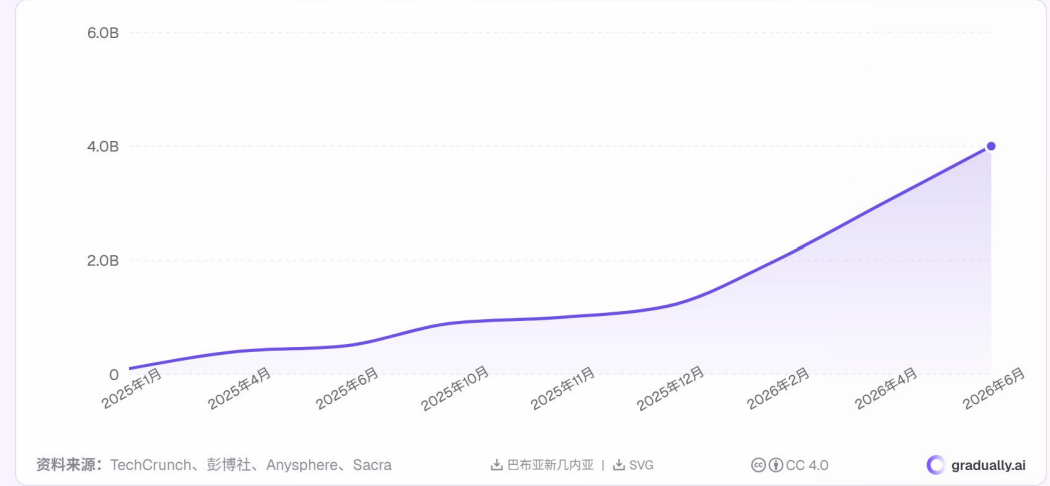
下载日志已记录，仅供内部参考

04 AI应用：ROI提升逐步可视化

AI编程是这一轮大模型浪潮中最先跑出清晰ROI、最先实现规模化付费的应用层赛道

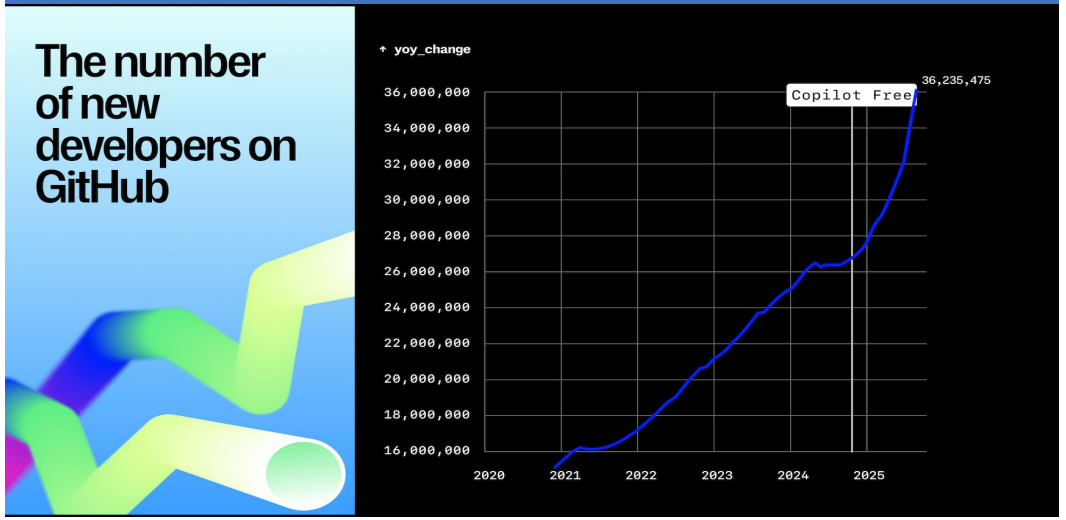
- **海外头部AI编程公司ARR快速爬坡。**海外龙头CursorARR创人类SaaS史上最快的增长曲线；2026年Q2，GitHub Copilot拥有约5000万用户，收入环比加速增长超60%。GitHub对《财富》500强企业覆盖率超90%。头部企业商业化数据成为这条赛道最有说服力的投资锚点。
- **AI编程辅助下，企业ROI清晰。**AI编程的效率转化清晰可量化，是大型企业提高AI编程覆盖率的动力。在Cursor的客户案例中，一个从零开始构建的应用交付周期，可从原定的4个月压缩至3周。
- **国内AI编程市场有望快速扩张。**据2026年7月IDC数据披露，2025年中国AI编程市场规模3.99亿元，预计2026年底将增长至11.73亿元（一年近3倍）。头部互联网企业腾讯已实现研发体系中50%新增代码由AI辅助生成。我们认为AI在编程中的渗透率曲线仍将陡峭。

图20：Cursor ARR 创SaaS史上最快的增长曲线



资料来源：gradually.ai，华龙证券研究所

图21：AI编程工具Copilot Free的推出加速Github开发者数量增长



资料来源：Github，华龙证券研究所

下载日志已记录，仅供内部参考

04 AI应用：ROI提升逐步可视化

大厂争夺AI办公流量入口

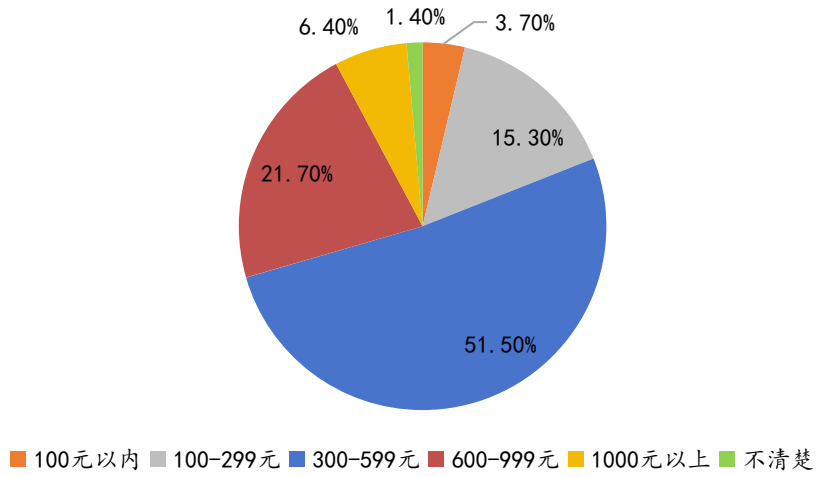
- **企业级Agent快速渗透。**IDC预计，活跃Agent的数量将从2025年的约2860万，快速攀升至2030年的22.16亿。这意味着五年后，能够帮助企业执行任务的数字劳动力数量将是今天的近80倍，年复合增长率139%，换言之，平均每年活跃Agent数量都将以超过一倍的速度增长。Agent真正干活的频率增长得更快，年执行任务数将从2025年的440亿次暴涨至2030年的415万亿次，年复合增长率高达524%。
- **数据飞轮效应下，办公是AI商业化的超级入口。**办公场景的高频调用能持续获取企业私域数据，用以微调模型，形成"用户使用越多、模型越懂企业、用户越离不开"的正向循环。艾媒咨询数据显示，使用AI办公智能体的企业中有近六成是100-499人的中型企业；51.5%的企业可接受单员工年度使用智能体的成本上限是300-599元。企业对员工使用智能体付费接受度较好。

图22：全球企业活跃Agent数量与年度执行量预测（统计时间：2026年1月）



资料来源：IDC，华龙证券研究所

图23：企业接受单员工年度使用智能体成本上限



资料来源：艾媒咨询，华龙证券研究所

下载日志已记录，仅供内部参考

04 AI应用：ROI提升逐步可视化

大厂争夺AI办公流量入口

- **2026年7-8月，国内AI办公赛道出现"集中兵力"现象，腾讯Workbuddy领跑。**2026 年 7 月AI办公赛道桌面(PC)客户端国内总月活用3000 万，其中腾讯 WokrBuddy 月活用户 1115.23 万，AI办公智能体总榜第一，增速 304.4%；百度搭子月活用户 674.3 万，增速1063.79%，AI办公智能体增速榜第一；字节 TraeWork 月活 265.99 万，增速 51.08%。8月，阿里千问办公开启公测，初步打通钉钉IM；随后，字节发布豆包工作，与飞书深度打通。至此，国内互联网三大巨头全部入局，AI办公赛道有望成为下一个AI应用热门竞技场。

图24：腾讯Workbuddy企业版及办公智能体套件全景图



资料来源：Workbuddy，华龙证券研究所

表5：国内大厂争夺AI办公入口

厂商	产品	关键进展
腾讯	WorkBuddy	2026年3月9日，腾讯自研的桌面智能体 WorkBuddy正式上线。已接入混元 Hy3 preview、DeepSeek-V4、智谱GLM-5.1、Kimi-K2.6等最新主流大模型；上线skills（技能）、专家模式、自动化定时任务，支持微信客服号、ClawBot插件等多种方式连接。2026年7月MAU超千万。
阿里	千问办公	2026年8月3日，千问办公开启公测，已初步打通钉钉IM
字节	豆包工作	2026年8月25日，字节发布豆包工作——面向生产力场景的Agent产品，与飞书深度打通。

资料来源：各公司官微，华龙证券研究所

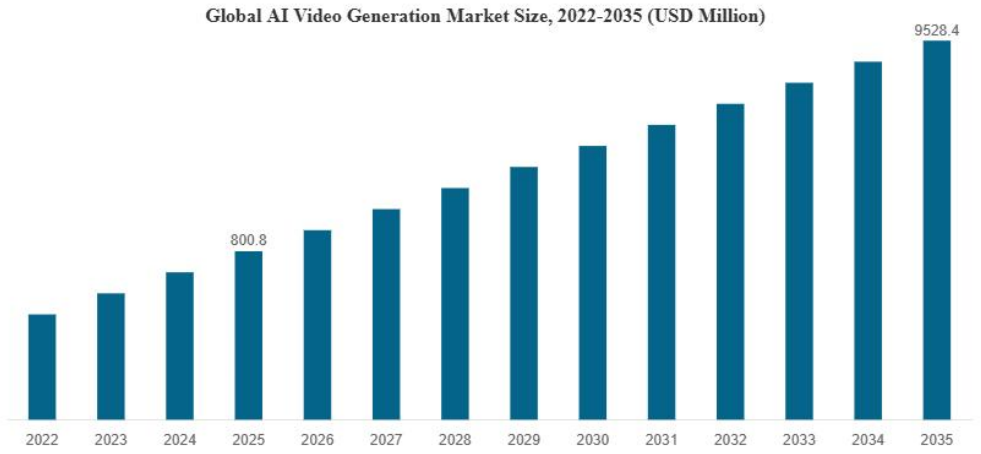
下载日志已记录，仅供内部参考

AI应用：ROI提升逐步可视化

AI视频生成降本效果显著

- **AI视频生成降本增效成果显著，视频生成迈向生产力工具。**借助字节的Seedance 2.0模型，内容制作方能将从剧本创作到剪辑的全流程提效80%-90%，对比传统工艺降本70%-90%；据字节跳动最新发布的Seedance 2.5模型定价估算，其输出5秒模型的价格更是低至3.36元/个。随着模型能力提升，AI视频生成的高效低成本优势将进一步显现。
- **商业化路径清晰，AI视频生成市场规模有望快速增长。**据Astute Analytica，在2026年至2035年的预测期内，全球人工智能视频生成市场规模复合年增长率将达到28.1%。今年二季度，可灵AI营收超8.5亿元，同比增长超过200%。截至6月，可灵AI全球用户突破1亿，较2025年底增长约67%；企业客户近5万家，同样较2025年底增长约67%。标志着国产AI视频生成模型已从"追赶"转为"引领"。

图25：全球AI视频生成市场规模



资料来源：Astute Analytica，华龙证券研究所

图26：Seedance 2.5 视频价格估算

- 输入不含视频

分辨率	宽高比	输出视频时长（秒）	doubao-seedance-2.5 视频价格	
			（元/个）	（元/s）
480p	16:9	5	3.36	0.67
720p	16:9	5	7.56	1.51
1080p	16:9	5	18.71	3.74

- 输入包含视频

分辨率	宽高比	输入视频时长（秒）	输出视频时长（秒）	doubao-seedance-2.5 视频价格（元/个）
480p	16:9	2~30	5	3.63~14.12 最低价对应输入2~4秒 最高价对应输入30秒
720p	16:9	2~30	5	8.16~31.75 最低价对应输入2~4秒 最高价对应输入30秒
1080p	16:9	2~30	5	20.12~78.25 最低价对应输入2~4秒 最高价对应输入30秒

资料来源：火山引擎，华龙证券研究所

下载日志已记录，仅供内部参考

目录

1 复盘：2026年内行情回顾

2 国产模型：性价比即竞争力

3 国产算力：以规模换效率

4 AI应用：ROI提升逐步可视化

5 投资建议

6 风险提示

下载日志已记录，仅供内部参考

05 投资建议

- **国产模型：性价比即竞争力。**我们认为，国产头部AI模型在tokens消耗量上已稳居全球第一梯队，部分国产模型凭借高性价比使得增长速率更加陡峭，更预示着持续的扩张潜力。反映出国产AI模型的发展趋势正从技术追赶转向市场引领。我们认为国产模型中，短期以调用量为支撑，中长期能够跑通企业API规模化变现、具备前沿模型迭代能力的标的将更可能获得成长性溢价与估值支撑。
- **国产算力：以规模换效率。**2026年内，国产超节点解决方案密集发布，表明当前国产AI算力的竞争维度已从单卡算力参数延伸至集群互联与系统级算力交付。我们认为超节点从样品走向规模展出释放了国产算力持续突破的明确信号，2027年有望进入国产超节点批量交付元年，国产服务器整机、交换机以及核心上游国产算卡供应商将核心受益。
- **AI应用：ROI提升逐步可视化。**AI编程、AI办公、AI视频生成等AI应用赛道降本增效成果陆续验证，国产AI应用产品力提升，大厂加速争夺AI应用流量入口，商业化落地有望加速。
- **投资建议：**展望未来，我们认为政策催化、国产替代需求以及AI驱动将共同推高计算机行业景气度，行业基本面拐点有望加速确立，建议关注景气赛道中的绩优标的，维持计算机行业**“推荐”**评级。建议关注个股：（1）大模型：智谱（2513.HK）、MINIMAX-W（0100.HK）；（2）国产算力：寒武纪-U（688256.SH）、海光信息（688041.SH）、中科曙光（603019.SH）、浪潮信息（000977.SZ）、润泽科技（300442.SZ）、宏景科技（301396.SZ）；（3）AI应用：科大讯飞（002230.SZ）、卓易信息（688258.SH）、索辰科技（688507.SH）、能科科技（603859.SH）、万兴科技（300624.SZ）。

下载日志已记录，仅供内部参考

05 投资建议

表6：重点关注公司及盈利预测

股票代码	股票简称	2026/08/21	EPS（元）				PE				投资评级
		股价（元）	2025A	2026E	2027E	2028E	2025A	2026E	2027E	2028E	
000977.SZ	浪潮信息	76.35	1.64	2.43	3.08	3.89	42.7	31.4	24.8	19.6	增持
002230.SZ	科大讯飞	39.19	0.36	0.51	0.67	0.85	139.7	77.2	58.9	45.9	未评级
0100.HK	MINIMAX-W	347.20	-17.23	-1.81	-1.71	-1.56	/	/	/	/	未评级
2513.HK	智谱	1129.00	-12.03	-10.18	-9.72	-4.90	/	/	/	/	未评级
300442.SZ	润泽科技	64.38	3.00	1.98	2.52	3.16	17.6	32.5	25.5	20.4	未评级
300624.SZ	万兴科技	51.24	-0.47	0.12	0.40	0.57	/	420.0	128.7	89.5	未评级
301396.SZ	宏景科技	186.06	0.17	2.22	4.62	9.95	385.5	83.8	40.3	18.7	未评级
603019.SH	中科曙光	84.11	1.49	2.00	2.54	3.26	57.5	42.1	33.1	25.8	未评级
603859.SH	能科科技	38.02	0.93	1.24	1.49	1.62	45.2	30.7	25.5	23.5	增持
688041.SH	海光信息	246.21	1.10	2.00	2.74	4.21	204.0	123.1	89.9	58.4	增持
688256.SH	寒武纪	1035.00	4.93	11.00	17.19	30.83	275.0	94.1	60.2	33.6	增持
688258.SH	卓易信息	97.29	0.66	1.46	2.38	3.43	116.0	66.8	41.0	28.4	未评级
688507.SH	索辰科技	158.46	0.36	0.86	1.14	1.26	278.6	184.9	139.5	125.8	未评级

数据来源：Wind，华龙证券研究所（注：浪潮信息盈利预测来自华龙证券研究所；海光信息、寒武纪、能科科技2026-2027年盈利预测来自华龙证券研究所，2028年盈利预测来自Wind一致预测；其他重点关注公司盈利预测来自Wind一致预测）

下载日志已记录，仅供内部参考

目录

1 复盘：2026年内行情回顾

2 国产模型：性价比即竞争力

3 国产算力：以规模换效率

4 AI应用：ROI提升逐步可视化

5 投资建议

6 风险提示

下载日志已记录，仅供内部参考

06 风险提示

（1）国产算力建设不及预期。算力是AI应用基石，国产算力建设不达预期将会延缓AI应用的落地速度。

（2）所引用数据资料的误差风险。本报告数据资料来源于公开数据，将可能对分析结果造成影响。

（3）AI应用落地速度不及预期。当前市场上AI应用的定价、商业模式以及市场监管等方面仍处于探索阶段。

（4）技术迭代速度不及预期。技术迭代速度影响商业化进程。

（5）重点关注公司业绩不达预期。重点关注公司业绩会受到各种因素影响，如果业绩不达预期，会使得公司股价受到影响。

（6）政策标准出台速度不及预期。人工智能赛道需政策提供支持和引导。

（7）科技股资金拥挤度高与波动放大风险。资金高度集中于科技赛道，交易拥挤度处于高位，可能加剧波动。

（8）海外流动性超预期收紧风险。全球美元流动性走向对港股影响较大，若海外流动性环境恶化，港股流动性将面临收紧。

下载日志已记录，仅供内部参考

免责及评级说明部分

分析师声明：

本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉尽责的职业态度，独立、客观、公正地出具本报告。不受本公司相关业务部门、证券发行人、上市公司、基金管理公司、资产管理公司等利益相关者的干涉和影响。本报告清晰准确地反映了本人的研究观点。本人在预测证券品种的走势或对投资证券的可行性提出建议时，已按要求进行相应的信息披露，在自己所知情的范围内本公司、本人以及财产上的利害关系人与所评价或推荐的证券不存在利害关系。本人不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。据此入市，风险自担。

投资评级说明：

投资建议的评级标准	类别	评级	说明
报告中投资建议所涉及的评级分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后的6-12个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅。其中：A股市场以沪深300指数为基准。	股票评级	买入	股票价格变动相对沪深 300 指数涨幅在 10%以上
		增持	股票价格变动相对沪深 300 指数涨幅在 5%至 10%之间
		中性	股票价格变动相对沪深 300 指数涨跌幅在-5%至 5%之间
		减持	股票价格变动相对沪深 300 指数跌幅在-10%至-5%之间
		卖出	股票价格变动相对沪深 300 指数跌幅在-10%以上
	行业评级	推荐	基本面向好，行业指数领先沪深 300 指数
		中性	基本面稳定，行业指数跟随沪深 300 指数
		回避	基本面向淡，行业指数落后沪深 300 指数

免责声明：

华龙证券股份有限公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。

本报告的风险等级评定为R4，仅供符合本公司投资者适当性管理要求的客户（C4及以上风险等级）参考使用。本公司不会因为任何机构或个人接收到报告而视其为当然客户。

本报告信息均来源于公开资料，本公司对这些信息的准确性和完整性不作任何保证。本报告所载的意见、评估及预测仅反映报告发布当日的观点和判断。在不同时期，本公司可能会发出与本报告所载意见、评估及预测不一致的研究报告，但本公司没有义务和责任及时更新本报告所涉及的内容并通知客户。同时，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。以往表现并不能指引未来，未来回报并不能得到保证，并存在损失本金的可能。

本报告仅为参考之用，并不构成对具体证券或金融工具在具体价位、具体时点、具体市场表现的投资建议，也不构成对所述金融产品、产品发行或管理人作出任何形式的保证。在任何情况下，本公司仅承诺以勤勉的职业态度，独立、客观地出具本报告以供投资者参考，但不就本报告中的任何内容对任何投资做出任何形式的承诺或担保。据此投资所造成的任何一切后果或损失，本公司及相关研究人员均不承担任何形式的法律责任。

在法律许可的情况下，本公司及所属关联机构可能会持有报告中提及的公司所发行证券的头寸并进行证券交易，也可能为这些公司提供或正在争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。本公司的员工可能担任本报告所提及的公司的董事。客户应充分考虑可能存在的利益冲突，勿将本报告作为投资决策的唯一参考依据。

版权声明：

本报告版权归华龙证券股份有限公司所有，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。

华龙证券研究所

北京
地址：北京市东城区安定门外大街189号天鸿宝景大厦西配楼F4层
邮编：100033

兰州
地址：兰州市城关区东岗西路638号文化大厦21楼
邮编：730030
电话：0931-4635761

上海
地址：上海市浦东新区浦东大道720号11楼
邮编：200000

深圳
地址：深圳市福田区民田路178号华融大厦辅楼2层
邮编：518046