

海外科技

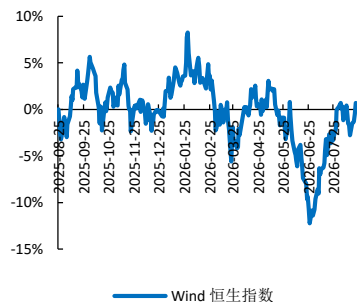
2026 年 08 月 23 日

投资评级：看好（维持）

DeepSeek 补齐视觉能力，Ox Alpha 或成为短期 Fable 级模型市场增量

——行业周报

行业走势图



数据来源：聚源

相关研究报告

《BloomEnergy 受益于 Neocloud 扩产+方案切换共振，现代牙科中报超预期——港股行业点评报告》-2026.8.17

《算力扩容驱动光互联升级，闪迪 HBF 加速 TCB 有望受益——行业周报》-2026.8.16

《功率龙头业绩超预期并上调指引，7 月淘天美护

yoy+11%——行业周报》-2026.8.9

初敏（分析师）

chumin@kysec.cn

证书编号：S0790522080008

杨哲（分析师）

yangzhe@kysec.cn

证书编号：S0790524100001

荀月（分析师）

xunyu@kysec.cn

证书编号：S0790524110001

● DeepSeek：从文本模型向视觉与 Agent 链路补齐

DeepSeek API 文档列出实验性的 V4-Flash-Vision-Exp，文档说明该实验模型在 V4-Flash 基础上加入图片输入，发送给 deepseek-v4-flash-vision-exp 的图片会按其尺寸换算成 token，与文本 token 一并计费。DeepSeek V4 Flash Vision Exp 作为其最新多模态 Agent 实验版本，在 Agent 综合能力上展现出明显迭代效果，部分指标逼近 Opus 4.8 第一梯队水平。DeepSeek 本轮的核心增量是补齐视觉输入与 Agent 开发链路，形成 V4 产品线的多模态执行闭环。

● Ox Alpha：近日受到重视的匿名模型，或对 Fable 级 Coding 市场带来增量

近日，一款名为 Ox Alpha 的模型受到重视：OpenRouter 当前模型目录列出模型 ID 为 stealth/ox-alpha 的 Ox Alpha，公开元数据显示其上下文长度为 1M，支持文本、图片和视频输入，以及工具调用、结构化输出和推理参数；对比页当前显示为免费预览。开发者在 DeepSWE 子集上测试，给出的结果约为 63%，但也有讨论认为该模型已经接近顶级代码水平。我们不对模型提供方做过多猜测，但若该模型确实可对标 SOTA 级 Coding 模型，可以预期该模型或是当前 Fable 级模型市场的一个重要增量。截至 8 月 22 日，Ox Alpha 以 6.54T tokens 排名第 4，标记为新进入榜单；排名高于 DeepSeek V4 Flash 0423、GPT-5.6 Luna、GLM 5.2 等模型。

● 投资建议：

AI：云服务需求与推理算力需求持续增长，云计算高景气延续，国产大模型流量提升带动商业化变现能力提升；AI 智能体产业进入快速增长期，受益标的谷歌、亚马逊、智谱、minimax、Nebius、Coreweave、讯策、德适-B、金山云。

互联网：关注 AI 商业化及应用场景拓展，随多模型能力持续提升、Agent 技术与应用生态快速成熟，token 消耗量有望延续高速增长，推理需求有望带动云计算行业摆脱低价内卷，供给端 DRAM 价格与产能紧约束，CPU、服务器等成本同步上行，共同推动云计算进入以 AI 算力溢价为核心的新阶段；国产 AI 芯片增长迅速；全球化出海有望由粗放扩张转向精细化盈利，推荐阿里巴巴-W、拼多多，百度集团-SW，受益标的腾讯控股。

风险提示：软件及产品推出不及预期，产能及供应链风险，监管政策变动，宏观经济增长放缓，地缘政治风险。

目 录

1、大模型：DeepSeek 补齐视觉能力，Ox Alpha 或成为短期 Fable 级模型市场增量.....	3
1.1、DeepSeek：从文本模型向视觉与 Agent 链路补齐.....	3
1.2、Ox Alpha：近日受到重视的匿名模型，或对 Fable 级 Coding 市场带来增量	4
2、周度数据更新	5
3、投资建议	9
4、风险提示	9

图表目录

图 1：DeepSeek V4-Flash-Vision-Exp 多模态能力与 Opus4.8 齐平.....	3
图 2：若 Ox Alpha 评分在 63%，则处于 DeepSeek V4 Pro 与 GPT-5.5 之间	4
图 3：Ox Alpha 在 OpenRouter 上排名靠前.....	5
图 4：本周恒生指数（2026.08.17-2026.08.21）在全球主要市场中涨幅居前（单位：%）	5
图 5：本周（2026.08.17-2026.08.21）恒生科技涨幅靠后（%）	6
图 6：本周（2026.08.17-2026.08.21）航运、地产、生物科技类涨幅居前（%）	6
图 7：2026 年 6 月以来港股通当日买入成交净额有所波动（亿元）	6
图 8：恒生沪港通 AH 溢价指数或已触底.....	7
图 9：OpenRouter 的 token 调用量持续提升	8
图 10：国产 AI 在 ZenMux 的 Token 市场份额占比接近 60%	9
表 1：本周（2026.08.17-2026.08.21）港股通头部活跃个股资金流向大模型、互联网等标的（仅计算前十大活跃个股明细）	7

1、大模型：DeepSeek 补齐视觉能力，Ox Alpha 或成为短期

Fable 级模型市场增量

1.1、DeepSeek：从文本模型向视觉与 Agent 链路补齐

DeepSeek API 文档列出实验性的 V4-Flash-Vision-Exp，文档说明该实验模型在 V4-Flash 基础上加入图片输入，发送给 deepseek-v4-flash-vision-exp 的图片会按其尺寸换算成 token，与文本 token 一并计费。

DeepSeek V4-Flash-Vision-Exp 作为其最新多模态 Agent 实验版本，在 Agent 综合能力上展现出明显迭代效果，部分指标逼近 Opus-4.8 第一梯队水平：

在文本 Agent 赛道：该版本相较基线 V4-Flash-0731 实现全面抬升，Terminal Bench 2.1 达到 83.9、Toolathlon-Verified 来到 75.9，代码工具调用、环境执行能力提升显著；NL2Repo、DSBench-Hard 等复杂任务同样取得可观进步，反映模型在工具规划、复杂指令拆解上的优化。

转向多模态 Agent 场景是本次版本核心看点：ApexBench 取得 36.5 分，较旧版本 26.2 大幅跃升；Chartography 图表理解、ZeroBench 多模态任务上，得分已经追平甚至小幅超越 Opus-4.8，证明视觉信息与 Agent 行动链路的融合取得突破。

但模型仍存在短板，AutomationBench 自动化任务得分依旧偏低，复杂现实自动化执行仍是瓶颈。综合来看，V4-Flash-Vision-Exp 验证了多模态 Agent 路线的可行性，兼顾速度与 Agent 能力，是偏向工程落地的迭代版本，但距离顶级模型仍有局部差距。

图1：DeepSeek V4-Flash-Vision-Exp 多模态能力与 Opus4.8 齐平

Benchmark		DeepSeek V4-Flash-Vision-Exp	DeepSeek V4-Flash- 0731	Opus-4.8
文本 Agent 能力评测*	Terminal Bench 2.1	83.9	82.7	85.0
	NL2Repo	57.7	54.2	69.7
	Cybergym	75.3	76.7	78.3
	DeepSWE	59.3	54.4	58.0
	Toolathlon-Verified	75.9	70.3	76.2
	DSBench-Hard	63.6	59.6	71.7
	AutomationBench (Public)	25.7	25.1	27.2
多模态 Agent 能力评测	ApexBench (Pass@1)	36.5	26.2**	39.4
	Agents' Last Exam	27.3	25.2**	25.7
	Chartography	64.3	-	65.0
	ZeroBench (Pass@5)	35.0	-	34.0

资料来源：DeepSeek 官网

DeepSeek 本轮的核心增量是补齐视觉输入与 Agent 开发链路，形成 V4 产品线
的多模态执行闭环。视觉输入使模型可以处理截图、图表、界面和文档页面；Harness
则把模型调用、工具执行和会话状态放进同一开发者工作流。两者叠加后，Agent
的工作路径从“读文本—生成答案”进一步扩展为“观察多模态信息—调用工具—执行
任务—返回结果”。这是产品能力边界的补齐，短期意义大于再增加一个纯文本模型
名称。

1.2、Ox Alpha: 近日受到重视的匿名模型，或对 Fable 级 Coding 市场带来增量

近日，一款名为 Ox Alpha 的模型受到重视：OpenRouter 当前模型目录列出模型
ID 为 stealth/ox-alpha 的 Ox Alpha，公开元数据显示其上下文长度为 1M，支持文本、
图片和视频输入，以及工具调用、结构化输出和推理参数；对比页当前显示为免费
预览。开发者在 DeepSWE 子集上测试，给出的结果约为 63%，但也有讨论认为该
模型已经接近顶级代码水平。我们不对模型提供方做过多猜测，但若该模型确实可
对标 SOTA 级 Coding 模型，可以预期该模型或是当前 Fable 级模型市场的一个重要
增量。

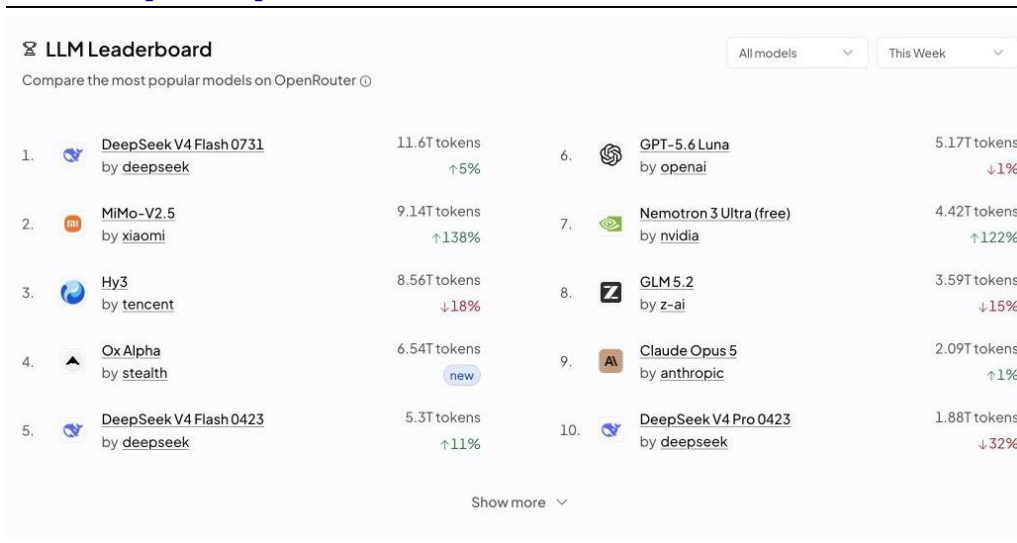
图2：若 Ox Alpha 评分在 63%，则处于 DeepSeek V4 Pro 与 GPT-5.5 之间

MODEL	PASS@1	AVG COST	OUT TOK	STEPS
AI claude-opus-5 [max]	74% ±4%	\$11.84	118k	99
gpt-5.6-sol [max]	73% ±3%	\$8.39	60k	61
AI claude-fable-5 [max]	70% ±4%	\$21.63	119k	88
glm-5.3 [max]	69% ±3%	\$3.99	80k	124
kimi-k3 [max]	69% ±5%	\$4.65	81k	98
gpt-5.6-luna [max]	67% ±4%	\$0.61	73k	102
gpt-5.5 [xhigh]	67% ±6%	\$7.23	46k	82
x grok-4.6 [xhigh]	67% ±2%	\$5.50	71k	87
gemini-3.7-flash [high]	65% ±2%	\$2.18	107k	125
deepseek-v4-pro [max]	63% ±6%	\$0.24	106k	155
AI claude-opus-4.8 [max]	59% ±2%	\$13.22	135k	120
qwen3.8-max [xhigh]	57% ±3%	\$3.73	95k	111
muse-spark-1.2 [xhigh]	55% ±2%	\$3.70	99k	101

资料来源：deepSWE 官网

OpenRouter 周榜提供了一个可量化的采用度信号：截至 8 月 22 日，Ox Alpha
以 6.54T tokens 排名第 4，标记为新进入榜单；排名高于 DeepSeek V4 Flash 0423、
GPT-5.6 Luna、GLM 5.2 等模型。

图3: Ox Alpha 在 OpenRouter 上排名靠前



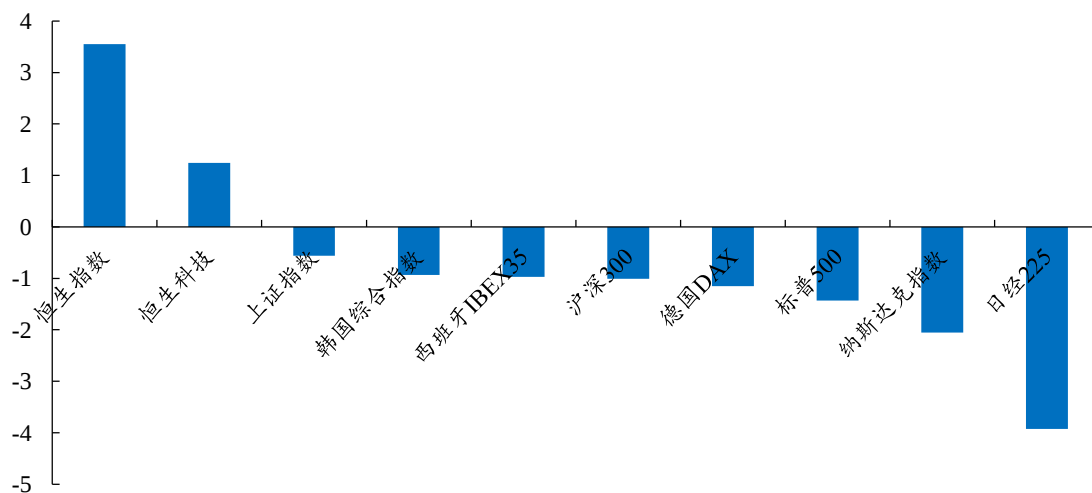
资料来源: OpenRouter 官网

Ox Alpha 的影响在于把测试门槛进一步压低。百万级上下文、图像/视频输入、工具调用与免费预览的组合,会降低开发者测试长文档、代码仓库和复杂 Agent 任务的成本,并抬高其他模型在“上下文长度+工具可用性”上的对比基线。对竞争格局而言,这类匿名模型也会加快能力扩散和价格透明化。

2、周度数据更新

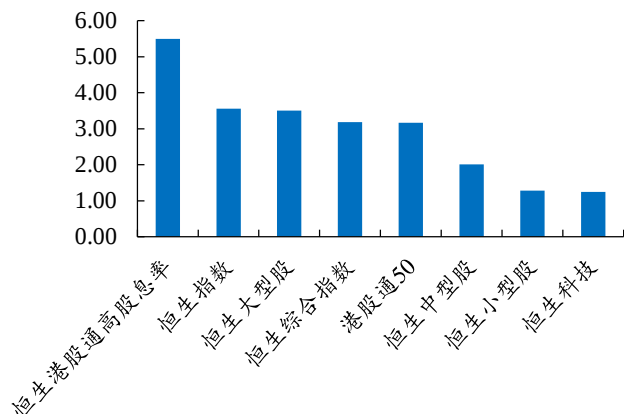
本周恒生指数上涨 3.55%, 在全球主要市场中涨幅居前, 从成分上看, 高股息、航运、地产、医药科技类等涨幅居前, 恒生科技本周上涨 1.24%。

图4: 本周恒生指数 (2026.08.17-2026.08.21) 在全球主要市场中涨幅居前 (单位: %)



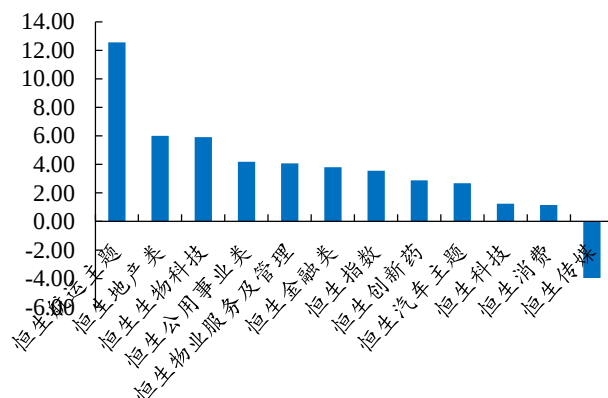
数据来源: Wind、开源证券研究所

图5：本周（2026.08.17-2026.08.21）恒生科技涨幅靠后（%）



数据来源：Wind、开源证券研究所

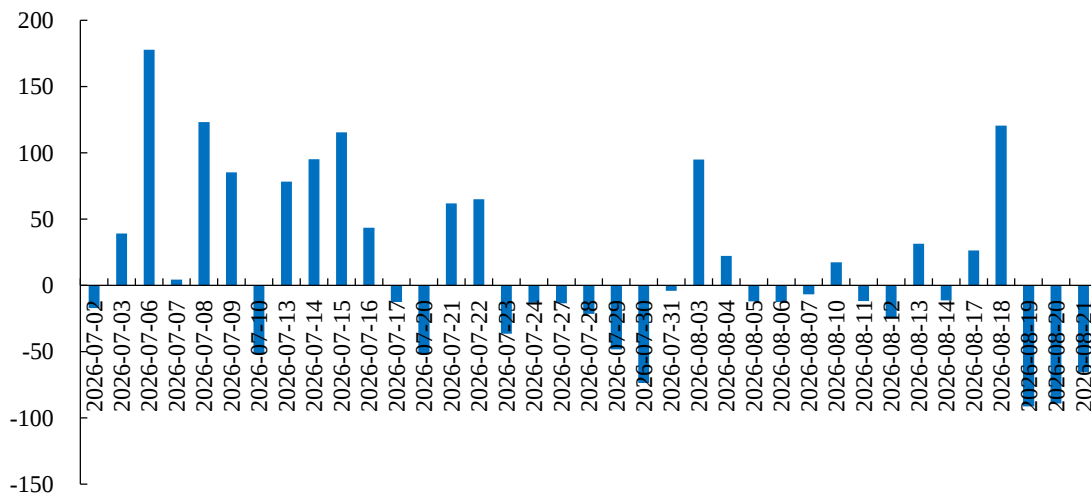
图6：本周（2026.08.17-2026.08.21）航运、地产、生物科技类涨幅居前（%）



数据来源：Wind、开源证券研究所

本周港股通成交净额为流出 98.91 亿元，环比转负，从头部活跃个股交易层面来看，阿里巴巴-W (+37.1 亿港元)、中芯国际 (+19.8 亿港元)、智谱 (+14.5 亿港元)、快手-W (+11.4 亿港元) 资金净流入居前。

图7：2026 年 6 月以来港股通当日买入成交净额有所波动（亿元）



数据来源：Wind、开源证券研究所

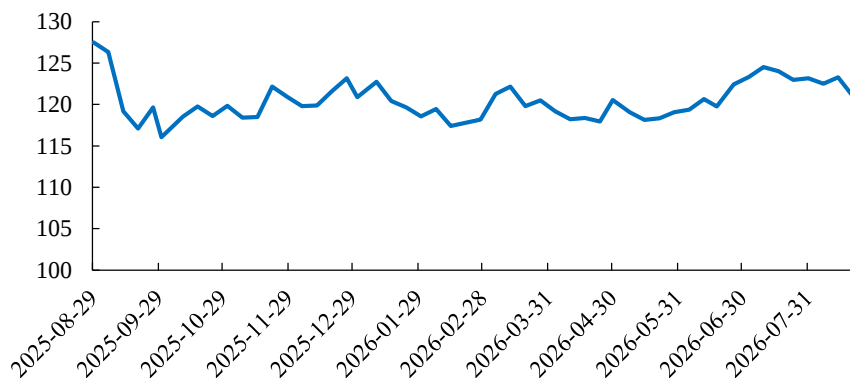
表1：本周（2026.08.17-2026.08.21）港股通头部活跃个股资金流向大模型、互联网等标的（仅计算前十大活跃个股明细）

股票代码	股票名称	成交净买入（亿港元）	港股通持股数占总股本比例
9988.HK	阿里巴巴-W	37.1	10.3%
0981.HK	中芯国际	19.8	29.7%
2513.HK	智谱	14.5	11.6%
1024.HK	快手-W	11.4	19.3%
0100.HK	MINIMAX-W	6.2	8.0%
2269.HK	药明生物	5.2	24.8%
3986.HK	兆易创新	2.7	55.5%
1548.HK	金斯瑞生物科技	2.1	24.1%
6166.HK	剑桥科技	1.9	54.5%
9992.HK	泡泡玛特	0.2	25.0%
0700.HK	腾讯控股	-2.1	11.8%
0762.HK	中国联通	-2.7	12.0%
1952.HK	云顶新耀	-3.2	58.2%
3690.HK	美团-W	-3.4	18.8%
6869.HK	长飞光纤光缆	-5.4	43.4%
1888.HK	建滔积层板	-5.6	13.7%
9926.HK	康方生物	-6.3	47.9%
1093.HK	石药集团	-10.9	28.4%
1810.HK	小米集团-W	-22.5	19.3%
1347.HK	华虹宏力	-35.6	20.5%

资料来源：Wind、开源证券研究所

本周恒生沪港通 AH 溢价指数为 120.96，环比有所下降，AH 股溢价或已触底，考虑到该指数成分中金融行业权重超 50%，能源行业权重超 15%，背后或更多反映对红利概念有需求的资金特征，随着后续更多 A 股科技类标的赴港上市，或带动港股板块的“成长性溢价”持续向 A 股靠拢。

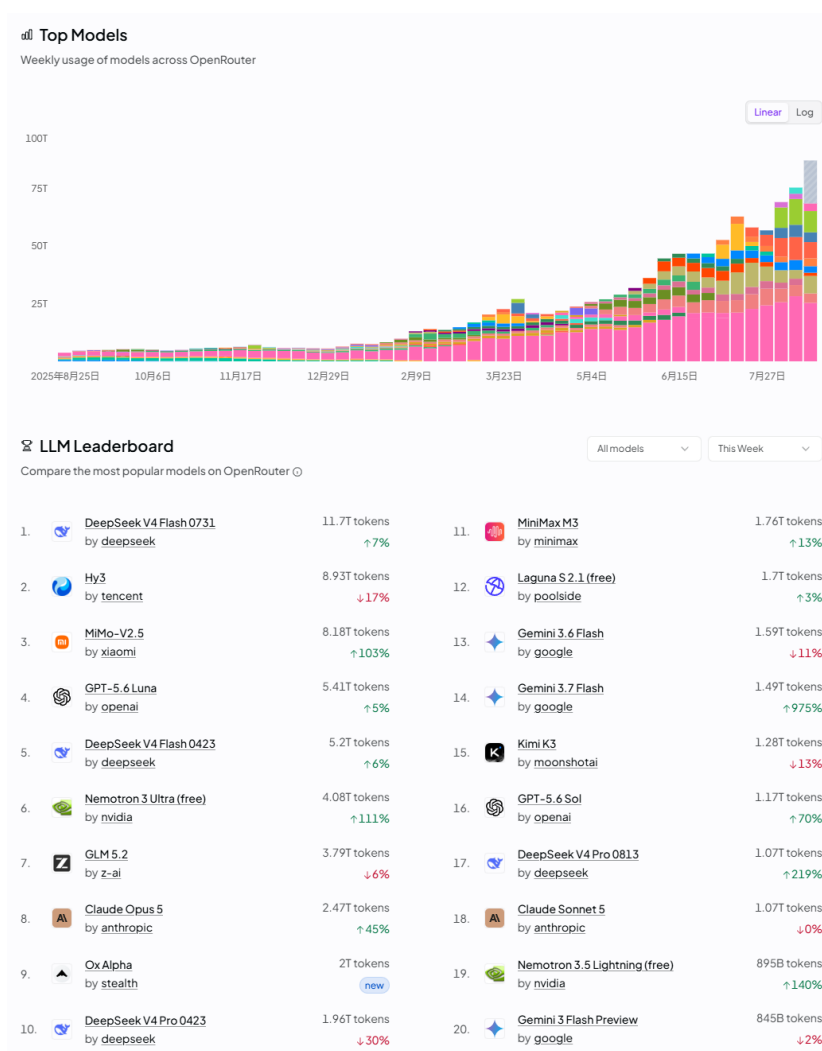
图8：恒生沪港通 AH 溢价指数或已触底



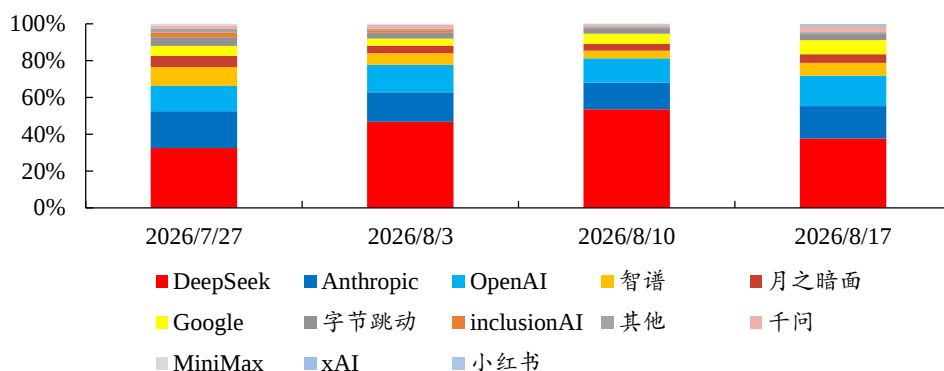
数据来源：Wind、开源证券研究所

本周（2026.08.17-2026.08.21）OpenRouter 的 token 调用量持续提升，其中国产模型 DeepSeek V4 Flash、Hy3、MiMo-V2.5 调用量位居前三。在 ZenMux 统计的 Token 市场份额中，国产 AI 占比接近 60%，DeepSeek 市场份额继续保持第一，占比为 37.8%。依托开源生态及出色的性价比优势，国产模型持续扩大市场影响力。

图9：OpenRouter 的 token 调用量持续提升



数据来源：OpenRouter 官网

图10：国产 AI 在 ZenMux 的 Token 市场份额占比接近 60%


数据来源：ZenMux、开源证券研究所

3、投资建议

AI：云服务需求与推理算力需求持续增长，云计算高景气延续，国产大模型流量提升带动商业化变现能力提升；AI 智能体产业进入快速增长期，受益标的谷歌、亚马逊、智谱、minimax、Nebius、Coreweave、讯策、德适-B、金山云。

互联网：关注 AI 商业化及应用场景拓展，随多模型能力持续提升、Agent 技术与应用生态快速成熟，token 消耗量有望延续高速增长，推理需求有望带动云计算行业摆脱低价内卷，供给端 DRAM 价格与产能紧约束，CPU、服务器等成本同步上行，共同推动云计算进入以 AI 算力溢价为核心的新阶段；国产 AI 芯片增长迅速；全球化出海有望由粗放扩张转向精细化盈利，推荐阿里巴巴-W、拼多多，百度集团-SW，受益标的腾讯控股。

4、风险提示

软件及产品推出不及预期：科技行业软件更新较快，市场竞争激烈，若产品推出不及时或不及预期，或影响企业未来竞争力和创收情况。

产能及供应链风险：若产能爬坡速度较慢，供给端受限，创收弹性无法充分释放，则相关行业产业链业绩或受到影响。

监管政策变动：科技政策及垂类行业相关政策的变动，或将影响公司增长景气度及战略决策，进一步影响公司的增长节奏。

宏观经济增长放缓：若宏观经济增长放缓影响终端消费需求、企业 IT 支出等，将影响产业链相关企业的创收能力。

地缘政治风险：地缘政治变动或伴随国际政策调整，或对科技行业供应链等相关企业带来限制。

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）》已于2017年7月1日起正式实施。根据上述规定，开源证券评定此研报的风险等级为R4（中高风险），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境内专业投资者及风险承受能力为C4、C5的普通投资者。若您并非境内专业投资者及风险承受能力为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。本报告清晰准确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。

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