

大模型加速迭代，AI 应用百花齐放

——人工智能行业半月报（2026 年 8 月第 1 期）

报告要点：

● 指数表现（2026 年 8 月上半月）

内地指数：中证人工智能产业指数上涨 7.23%，逆转上月跌势。国证 AI 应用指数上涨 0.25%，上证科创板人工智能指数上涨 3.71%。

香港指数：恒生港股通人工智能主题指数上涨 6.43%，逆转上月跌势。

● 数据跟踪

据 OpenRouter 数据显示，在刚过去的一个整周(08.10-08.16)中，按 token 计算的大模型周调用量为 75.3 万亿 tokens，周调用量最高的编程智能体与生产力工具均为 Hermes Agent。据 Artificial Analysis 数据显示，8 月 18 日，以 AAI 为标准，Anthropic Claude 系列模型的智能水平总体上仍保持领先；当前智能水平最高的国产大模型是 Kimi K3；OpenAI 的 GPT-5.6 Luna 有较为突出的性价比。智能体能力专项方面，当前专项智能水平最高的国产大模型是 Qwen3.8 Max。当前总参数量最高的开源大模型是 Kimi K3。

● 重点新闻

大模型企业：谷歌 DeepMind 掌门卸任，首席科学家离职创业；MiniMax 被正式纳入港股通；OpenAI：开放免费用户与 ChatGPT 文字畅聊；OpenAI 暂缓最强模型 Astra，网络安全能力触及临界；SpaceXAI 发布新一代大模型 Grok 4.6；DeepSeek-V4-Pro 正式版发布，价格更新调整；阿里开源 2.4 万亿参数旗舰模型 Qwen3.8；谷歌发布 Gemini 3.7 Flash；智谱发布 GLM-5.3。

底层算力基础设施：SpaceX 将在美国德克萨斯州建设半导体制造工厂；SK 海力士：将在韩国投资 384 亿美元建设晶圆厂；黄仁勋发文推动 AI 工厂算力成为可投资资产类别。

人工智能运用：「千问办公」开启公测；Meta 发布首款编程 Agent；DeepSeek、腾讯旗下上海启善获宇树科技战略配售；智谱 ZCode 升级上线 Goal 支持复杂任务自主交付；SpaceXAI 发布 Grok Bot 早期测试版；DeepSeek Harness 开发者预览版发布。

政策法规与安全性：特朗普宣布对多晶硅及其衍生产品加征关税；智谱将启动“开源的盾”计划；Anthropic 发布 186 页风险报告。

● 投资观点

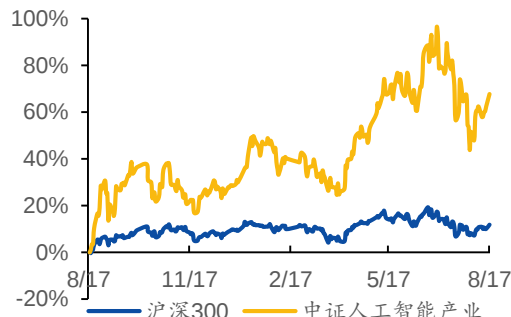
我们建议：一、持续跟进头部企业在人工智能应用领域的业务发展与投融资动向，综合研判相关布局的中长期价值与核心受益个股；二、关注能够充分满足人工智能行业发展需求、具备深厚技术护城河的基础设施企业标的。

● 风险提示

大模型技术迭代不及预期风险；人工智能基础设施需求不及预期风险；人工智能技术应用不及预期风险；宏观经济表现不及预期风险；行业竞争加剧的风险；相关政策落地不及预期的风险。

推荐|维持

过去一年市场行情



资料来源：iFinD

相关研究报告

《国元证券行业研究-人工智能行业半月报（2026 年 7 月第 2 期）：Kimi K3 发布，引领开源大模型发展》2026.08.05

《国元证券行业研究-人工智能行业半月报（2026 年 7 月第 1 期）：大模型持续迭代创新，token 调用量创新高》2026.07.17

报告作者

分析师 耿军军

执业证书编号 S0020519070002

电话 021-51097188-1856

邮箱 gengjunjun@gyzq.com.cn

目 录

1. 指数表现.....	3
1.1 内地股票相关指数.....	3
1.2 香港股票相关指数.....	4
2. 数据跟踪.....	5
2.1 大模型调用与测评.....	5
2.2 大模型价格与参数.....	9
2.3 应用软件与智能体.....	11
3. 重点新闻.....	13
3.1 大模型与代表企业.....	13
3.2 底层算力基础设施.....	16
3.3 人工智能实践应用.....	17
3.4 政策法规与安全性.....	19
4. 投资观点.....	20
5. 风险提示.....	20

图表目录

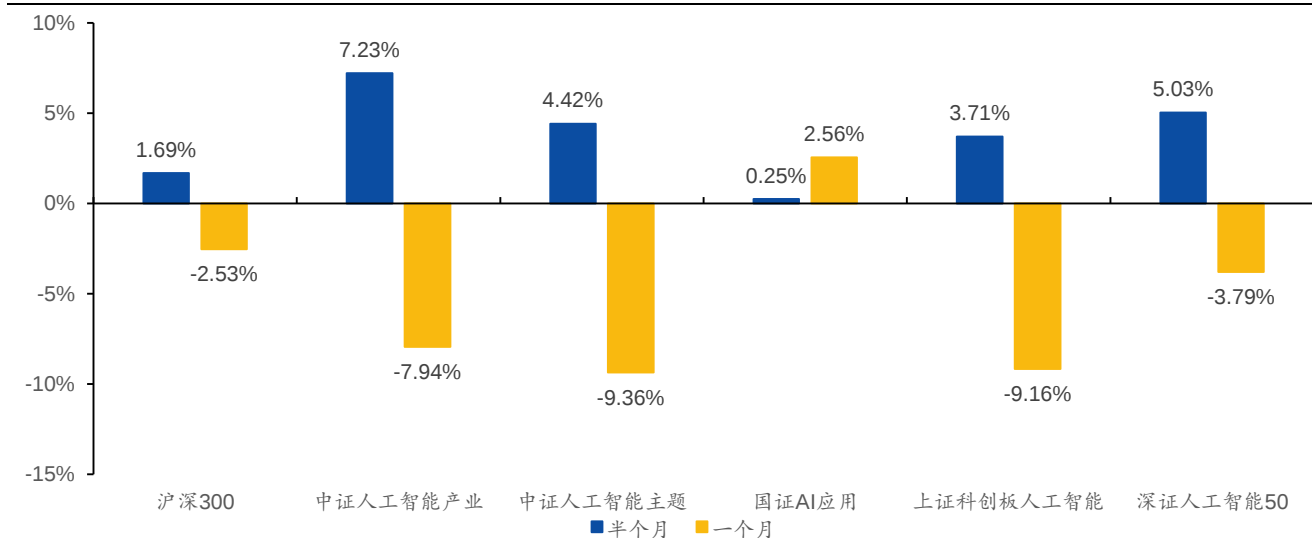
图 1: 沪深 300 指数及人工智能概念指数半月度与月度涨跌幅.....	3
图 2: 沪深 300 指数及人工智能概念指数 2026 年初至今累计涨跌幅.....	3
图 3: 恒生系列核心指数及恒生港股通人工智能主题指数半月度与月度涨跌幅.....	4
图 4: 恒生系列核心指数及恒生港股通人工智能主题指数 2026 年初至今累计涨跌幅.....	4
图 5: 大模型调用统计与近三周排名（截至 8 月 16 日）.....	5
图 6: 大模型厂商市场份额统计与近三周排名（以文本请求为衡量标准，截至 8 月 16 日）.....	6
图 7: 大模型智能水平综合排名（以 AAIL 或 AAI 为标准，截至 8 月 18 日）.....	7
图 8: 大模型厂商前沿语言模型智能水平对比（全部/当前前十厂商，以 AAIL 为标准，截至 8 月 18 日）.....	8
图 9: 大模型智能水平与任务加权平均成本对比图（智能水平以 AAIL 为标准，截至 8 月 18 日）.....	9
图 10: 开源大模型总参数量对比（十亿单位，截至 8 月 18 日）.....	10
图 11: 开源大模型智能水平与总参数量对比图（智能水平以 AAIL 为标准，截至 8 月 18 日）.....	10
图 12: 编程智能体调用统计与近三周排名（截至 8 月 16 日）.....	11
图 13: 生产力工具调用统计与近三周排名（截至 8 月 16 日）.....	12

1. 指数表现

1.1 内地股票相关指数

2026 年 8 月 3 日至 14 日, 中证人工智能产业指数上涨 7.23%, 逆转了 7 月的跌势, 跑赢沪深 300 指数 5.53 个百分点。同一时间段内, 国证 AI 应用指数上涨 0.25%, 上证科创板人工智能指数上涨 3.71%。

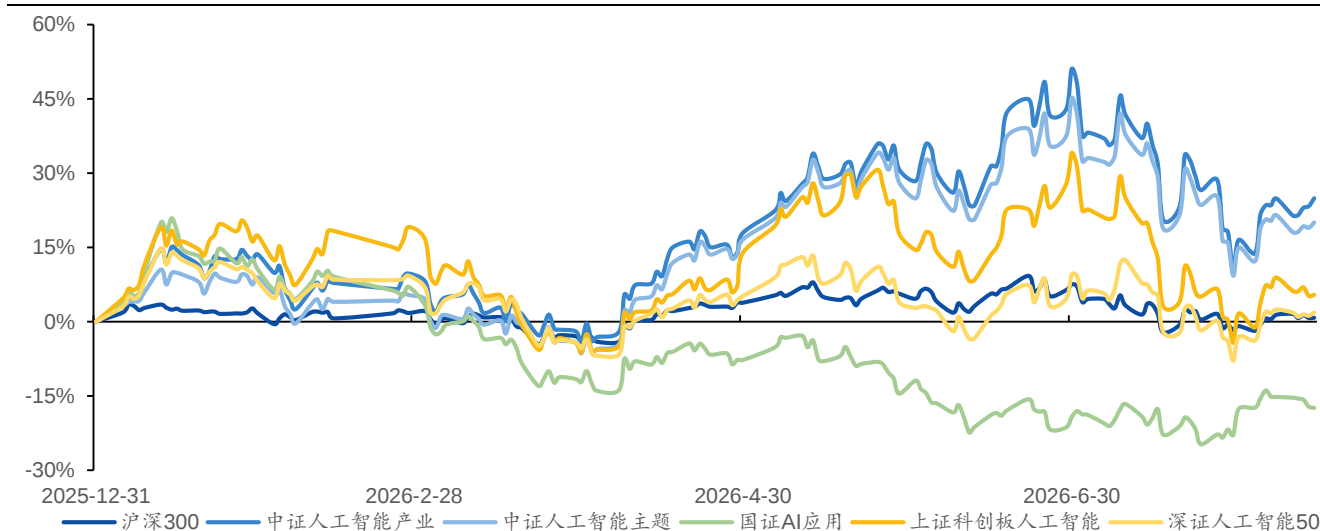
图 1: 沪深 300 指数及人工智能概念指数半月度与月度涨跌幅



资料来源: iFinD, 国元证券研究所

今年以来, 中证人工智能产业指数累计上涨 24.93%, 跑赢沪深 300 指数 24.15 个百分点。国证 AI 应用指数累计下跌 17.43%, 上证科创板人工智能指数累计上涨 5.39%。

图 2: 沪深 300 指数及人工智能概念指数 2026 年初至今累计涨跌幅

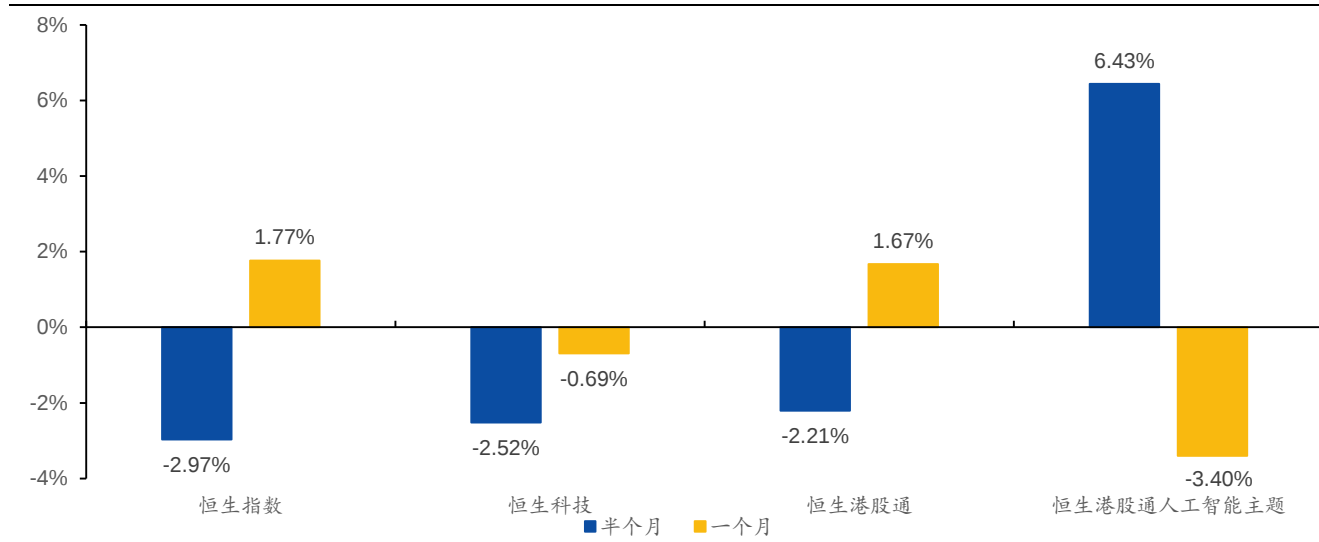


资料来源: iFinD, 国元证券研究所

1.2 香港股票相关指数

2026年8月3日至14日，恒生港股通人工智能主题指数上涨6.43%，逆转了7月的跌势，跑赢恒生指数9.40个百分点。同一时间段内，恒生科技指数下跌2.52%，恒生港股通指数下跌2.21%。

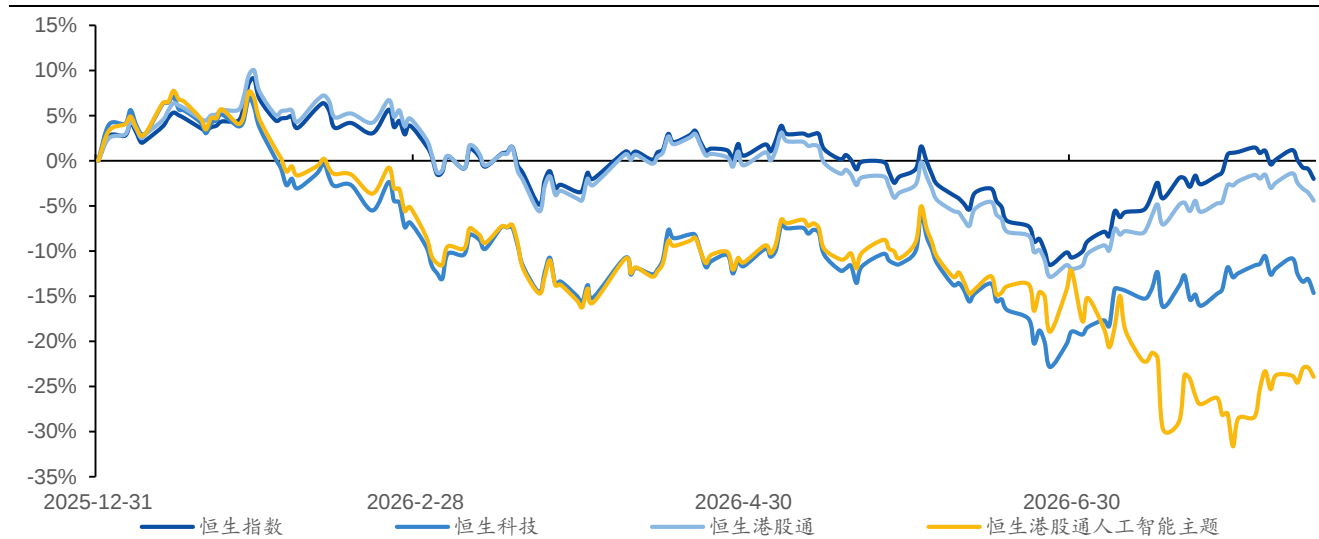
图3：恒生系列核心指数及恒生港股通人工智能主题指数半月度与月度涨跌幅



资料来源：iFinD，国元证券研究所

今年以来，恒生港股通人工智能主题指数累计下跌23.94%，跑输恒生指数21.93个百分点。恒生科技指数累计下跌14.65%，恒生港股通指数累计下跌4.43%。

图4：恒生系列核心指数及恒生港股通人工智能主题指数2026年初至今累计涨跌幅



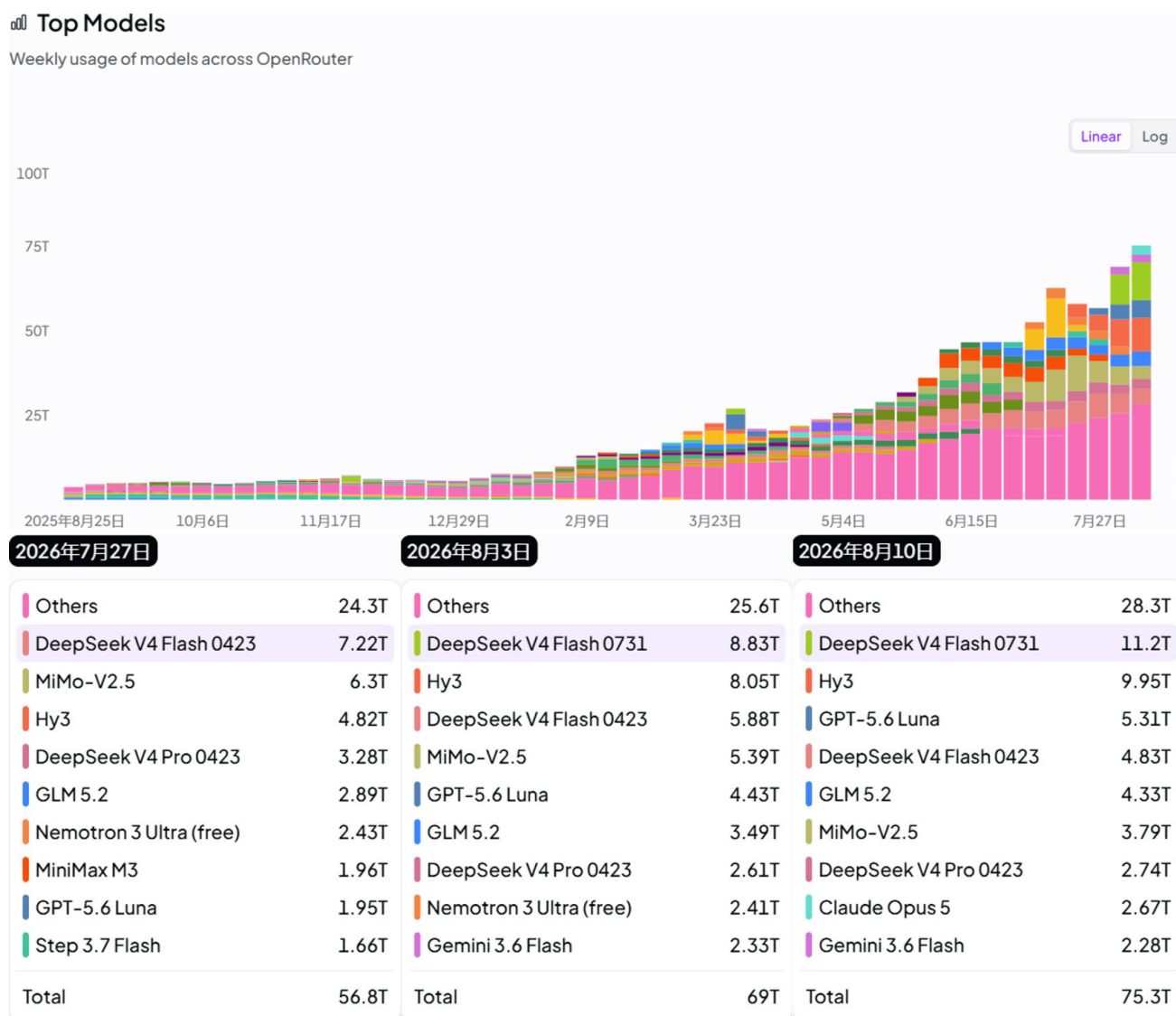
资料来源：iFinD，国元证券研究所

2. 数据跟踪

2.1 大模型调用与测评

据 OpenRouter 统计数据显示，按 token（词元）计算的大模型周调用量在刚过去的一个整周(2026.08.10-2026.08.16)为 75.3 万亿 tokens，较上一个整周(2026.08.03-2026.08.09)的 69 万亿 tokens 有所上升，周调用量再创新高。其中，该周调用排名前五的大模型分别为 DeepSeek 的 DeepSeek V4 Flash 0731、腾讯的 Hy3、OpenAI 的 GPT-5.6 Luna、DeepSeek 的 DeepSeek V4 Flash 0423 以及智谱的 GLM 5.2。

图 5：大模型调用统计与近三周排名（截至 8 月 16 日）

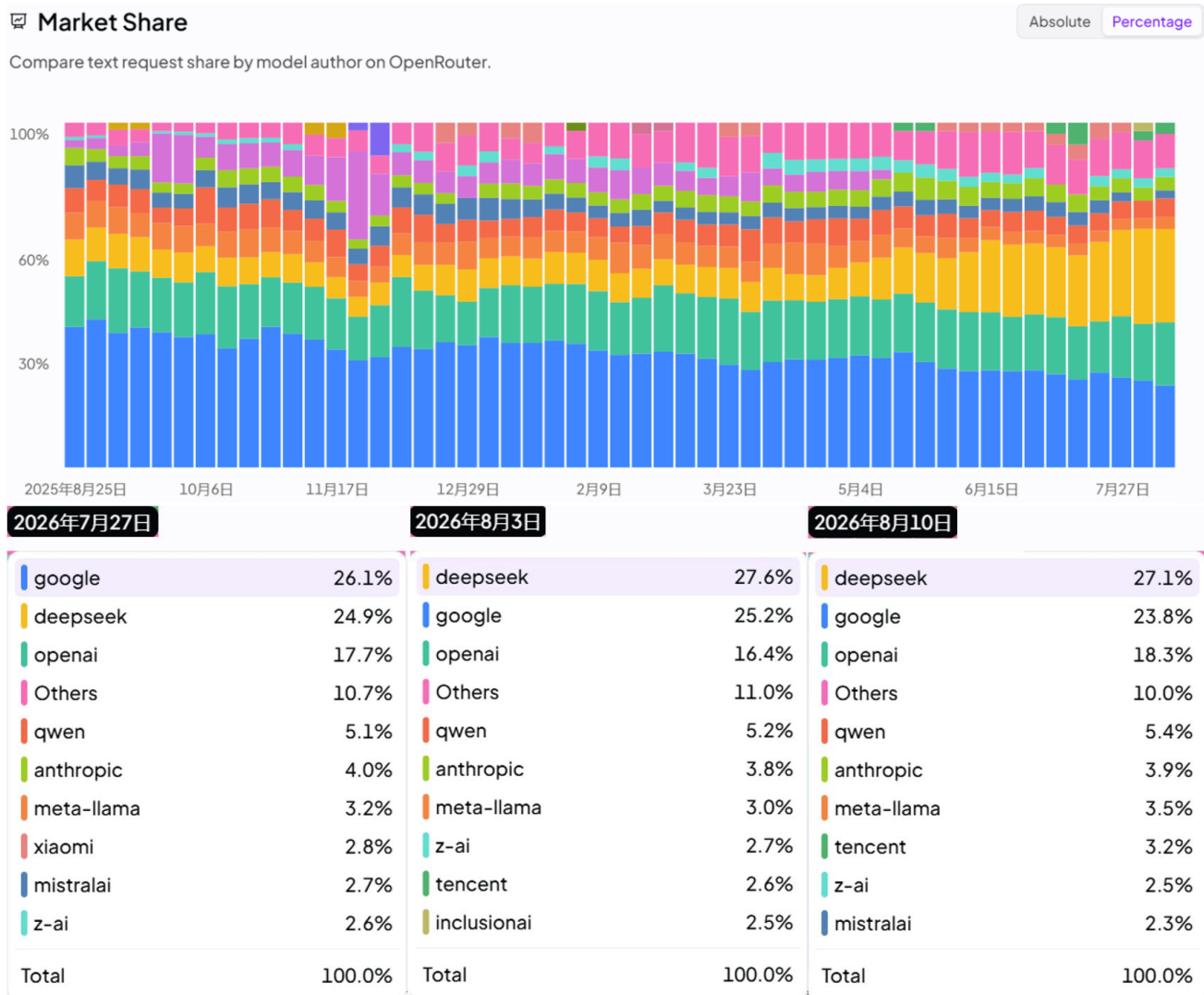


资料来源：OpenRouter，国元证券研究所

以文本请求为衡量标准，在最近过去的一个整周(2026.08.10-2026.08.16)，市场份额排名首位的是 DeepSeek(27.1%)，排名第二的是谷歌(23.8%)。

DeepSeek 在文本请求方面的市场份额已连续两周超过谷歌。OpenAI、阿里千问和 Anthropic 则继续在相应市场份额上排名三到五位。

图 6：大模型厂商市场份额统计与近三周排名（以文本请求为衡量标准，截至 8 月 16 日）



资料来源：OpenRouter，国元证券研究所

当地时间 2026 年 8 月 6 日，Artificial Analysis 小幅更新了 AAIL 的计算方法（现为 v4.1.1 版），连带受影响的是 AAAI 的计算方法。当前 Artificial Analysis 平台上显示的 AAIL、AAAI 及相关数据，均采用这种最新计算方法计算得出。

8 月 18 日，以 AAIL 为标准，Anthropic Claude 系列模型的智能水平总体上仍保持领先；OpenAI GPT 系列模型的智能水平优势下降；当前智能水平最高的国产大模型是 Kimi K3。智能体能力专项方面，8 月 18 日，以 AAAI 为标准，Claude Opus 5 (max) 和 Grok 4.6 (high) 的专项智能水平排名前两位，当前智能水平最高的国产大模型是 Qwen3.8 Max。

图 7：大模型智能水平综合排名（以 AAIL 或 AAAI 为标准，截至 8 月 18 日）

Artificial Analysis Intelligence Index

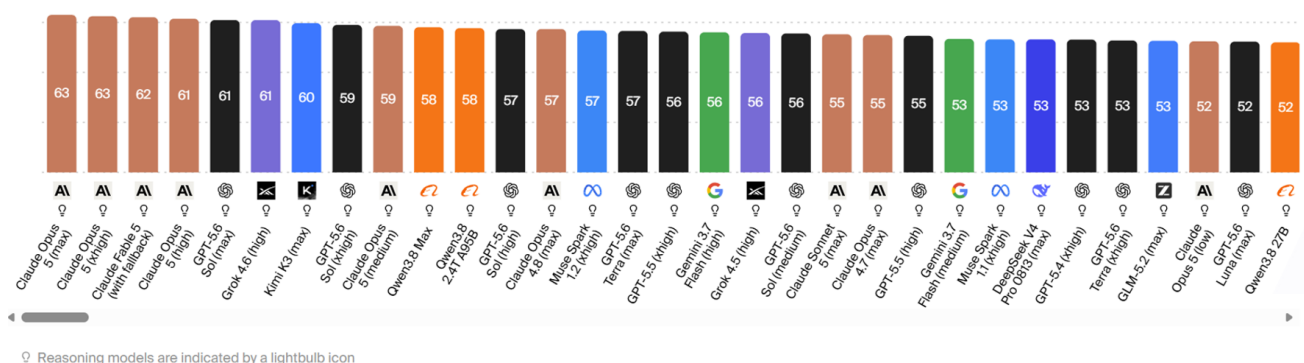
Artificial Analysis Intelligence Index v4.1.1 incorporates 9 evaluations: GDPval-AA v2, τ^3 -Banking, Terminal-Bench v2.1, SciCode, Humanity's Last Exam, GPQA Diamond, CritPT, AA-Omniscience, AA-LCR

609 of 609 models

+ Add model from specific provider

Estimate (independent evaluation forthcoming)

Artificial Analysis



Reasoning models are indicated by a lightbulb icon

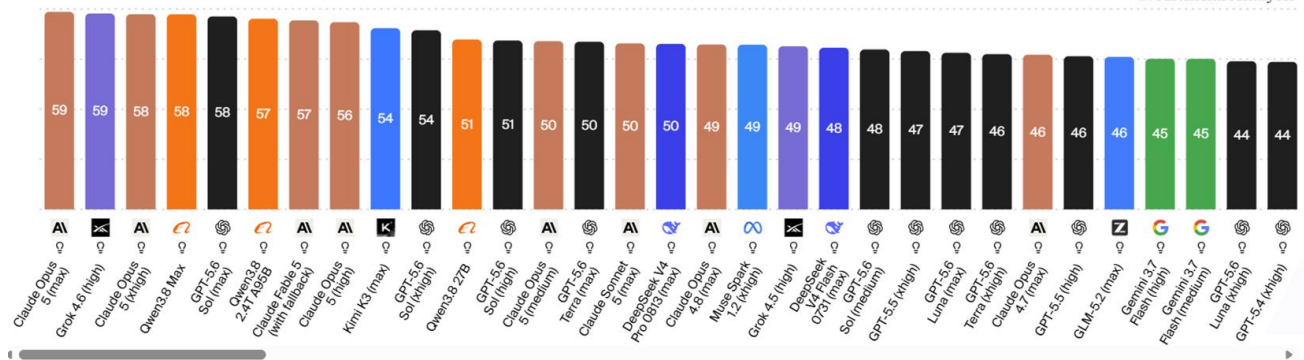
Artificial Analysis Agentic Index

Represents the weighted average of agentic capabilities benchmarks in the Artificial Analysis Intelligence Index (GDPval-AA v2, τ^3 -Banking)

609 of 609 models

+ Add model from specific provider

Artificial Analysis



Reasoning models are indicated by a lightbulb icon

资料来源：Artificial Analysis，国元证券研究所

据 Artificial Analysis 的测评结果看，8 月 18 日，以 AAIL 为智能水平的衡量标准，前沿语言模型排名前十的大模型厂商分别为：Anthropic、OpenAI、SpaceXAI、Kimi、阿里巴巴、Meta、谷歌、DeepSeek、智谱以及 Motif Technologies。

韩国大模型厂商 Motif Technologies 挤进前十，成为排名前十的大模型厂商中唯一中美以外的大模型厂商。

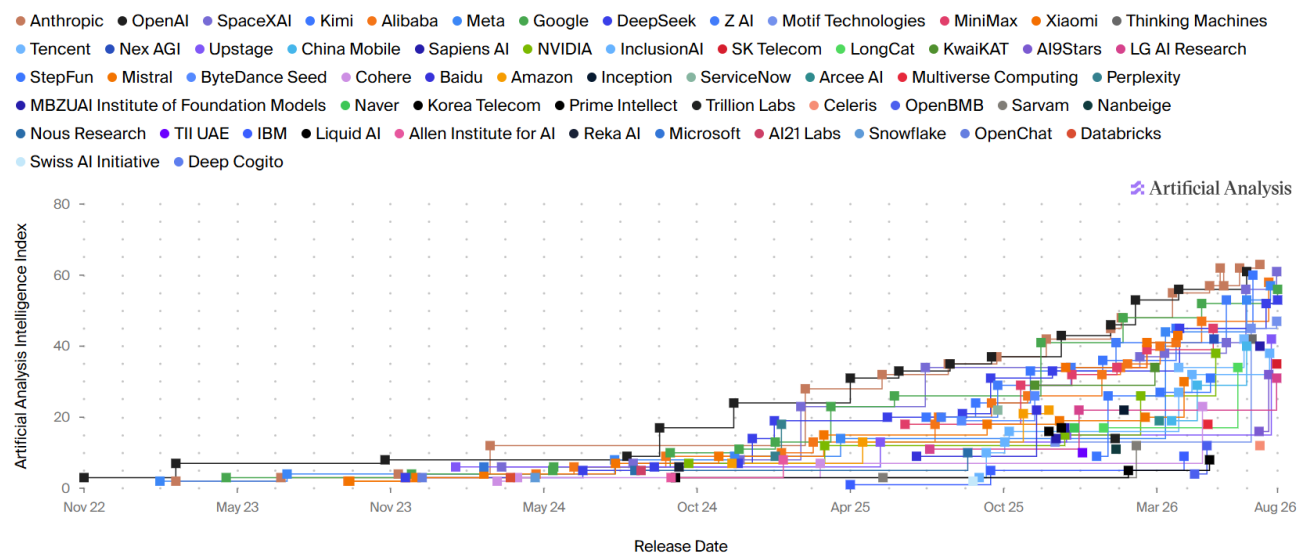
以 AAIL 为智能水平的衡量标准，自 2022 年 11 月以来，最前沿语言模型的地位始终被美国大模型厂商占据。

图 8：大模型厂商前沿语言模型智能水平对比（全部/当前前十厂商，以 AAIL 为标准，截至 8 月 18 日）

Frontier Language Model Intelligence, Over Time

Artificial Analysis Intelligence Index v4.1.1 incorporates 9 evaluations: GDPval-AA v2, r²-Banking, Terminal-Bench v2.1, SciCode, Humanity's Last Exam, GPQA Diamond, CritPt, AA-Omniscience, AA-LCR

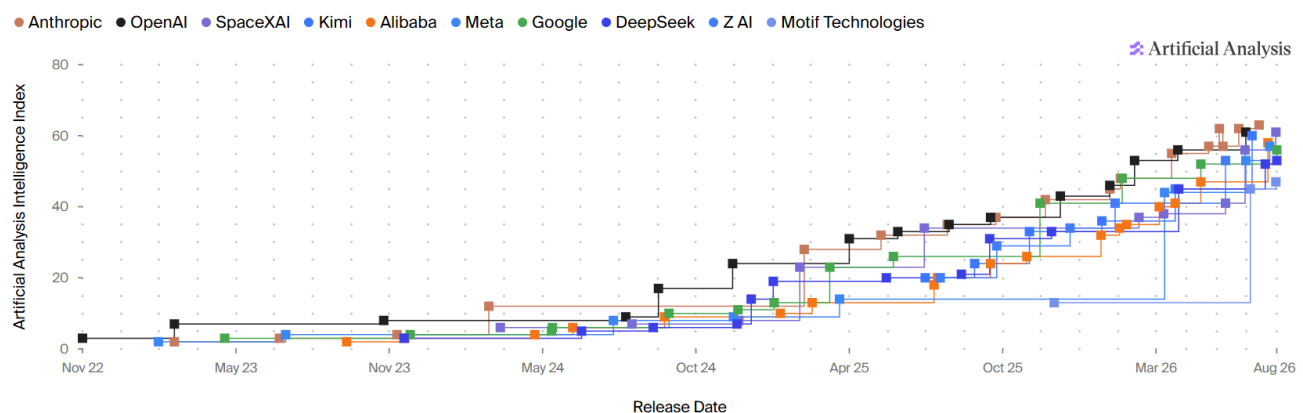
58 of 58 model creators



Frontier Language Model Intelligence, Over Time

Artificial Analysis Intelligence Index v4.1.1 incorporates 9 evaluations: GDPval-AA v2, r²-Banking, Terminal-Bench v2.1, SciCode, Humanity's Last Exam, GPQA Diamond, CritPt, AA-Omniscience, AA-LCR

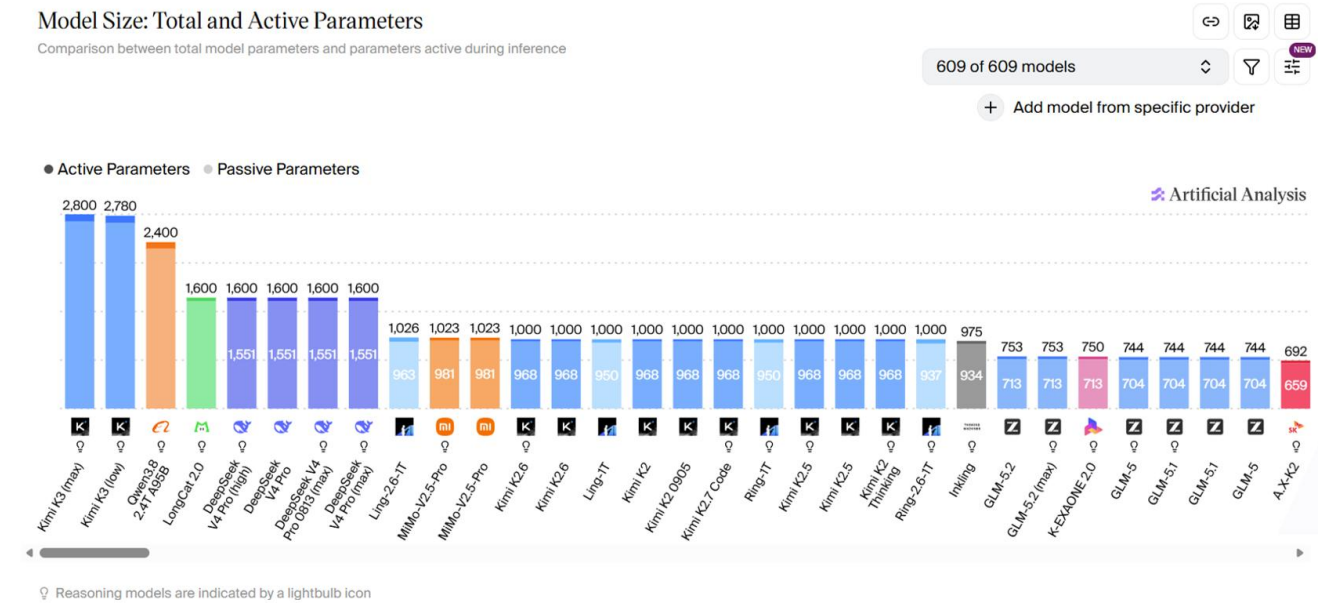
10 of 58 model creators



资料来源：Artificial Analysis，国元证券研究所

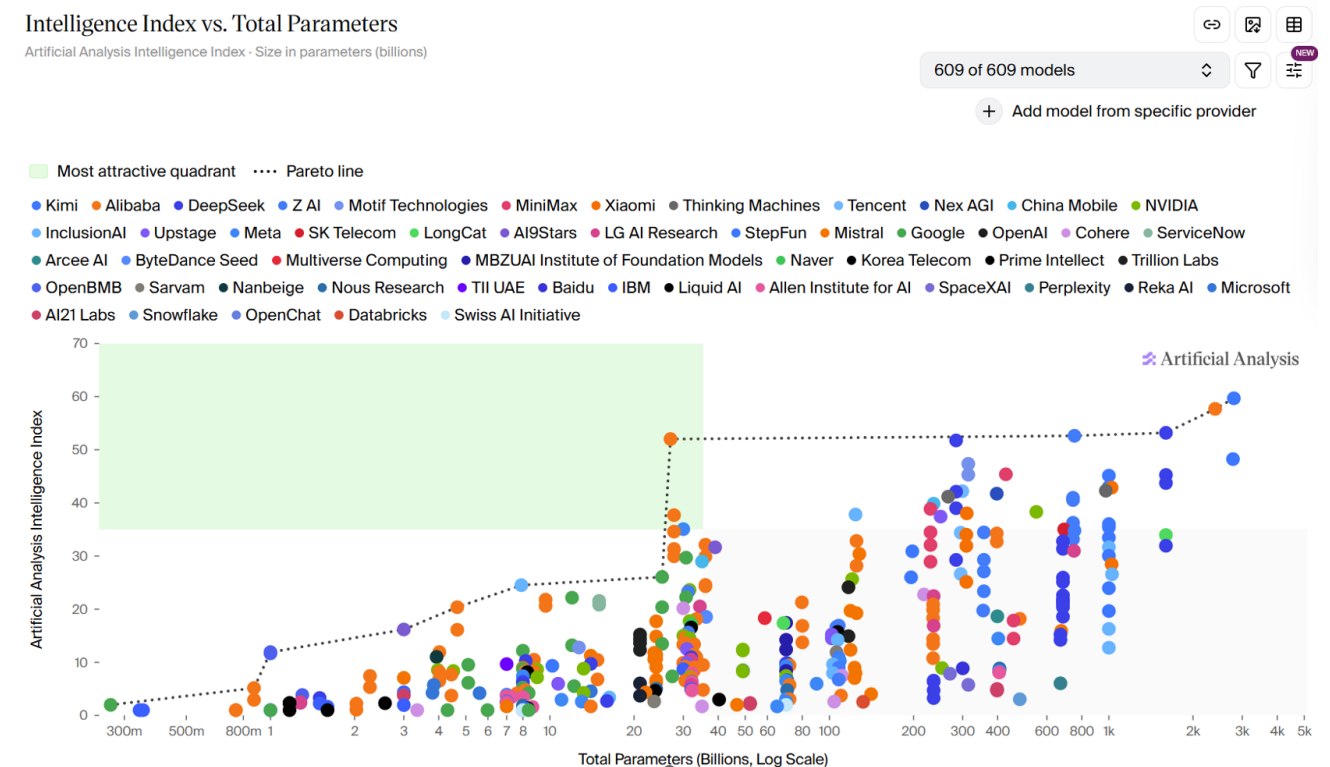
据 Artificial Analysis 数据，开源大模型中，当前总参数量最高的大模型为 Kimi K3，其在 max 和 low 思考强度模式下的总参数量分别为 2.8 万亿和 2.78 万亿。

图 10：开源大模型总参数量对比（十亿单位，截至 8 月 18 日）



资料来源：Artificial Analysis，国元证券研究所

图 11：开源大模型智能水平与总参数量对比图（智能水平以 AAI 为标准，截至 8 月 18 日）

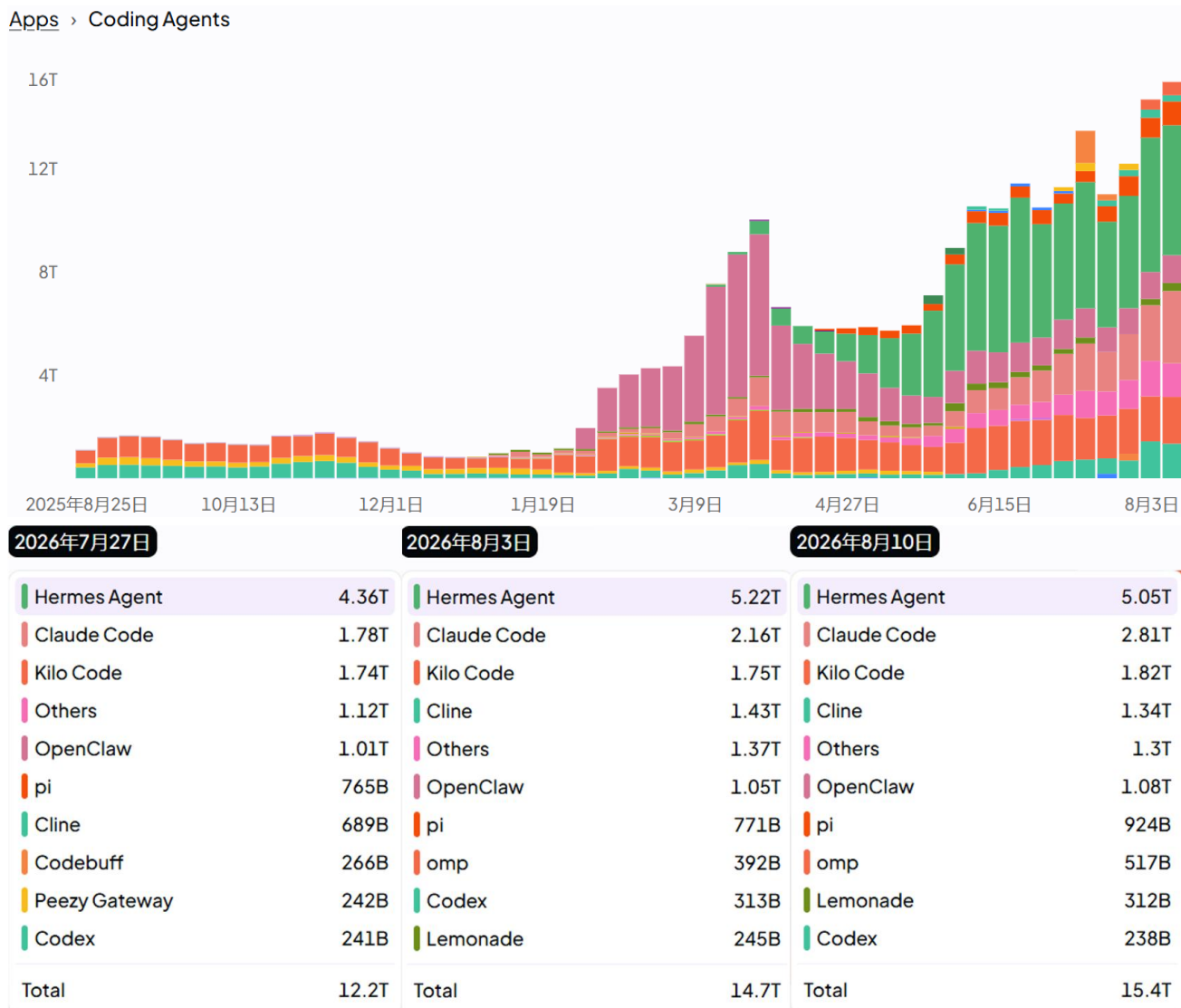


资料来源：Artificial Analysis，国元证券研究所

2.3 应用软件与智能体

据 OpenRouter 统计数据显示,在最近过去的一个整周(2026.08.10-2026.08.16)中,编程智能体的调用总量为 15.4 万亿 tokens,较上一个整周(2026.08.03-2026.08.09)的 14.7 万亿 tokens 有所上升。其中,在最近过去的这个整周中周调用量排名前三的分别为 Nous Research 的 Hermes Agent、Anthropic 的 Claude Code 以及 Kilo Code。

图 12: 编程智能体调用统计与近三周排名 (截至 8 月 16 日)

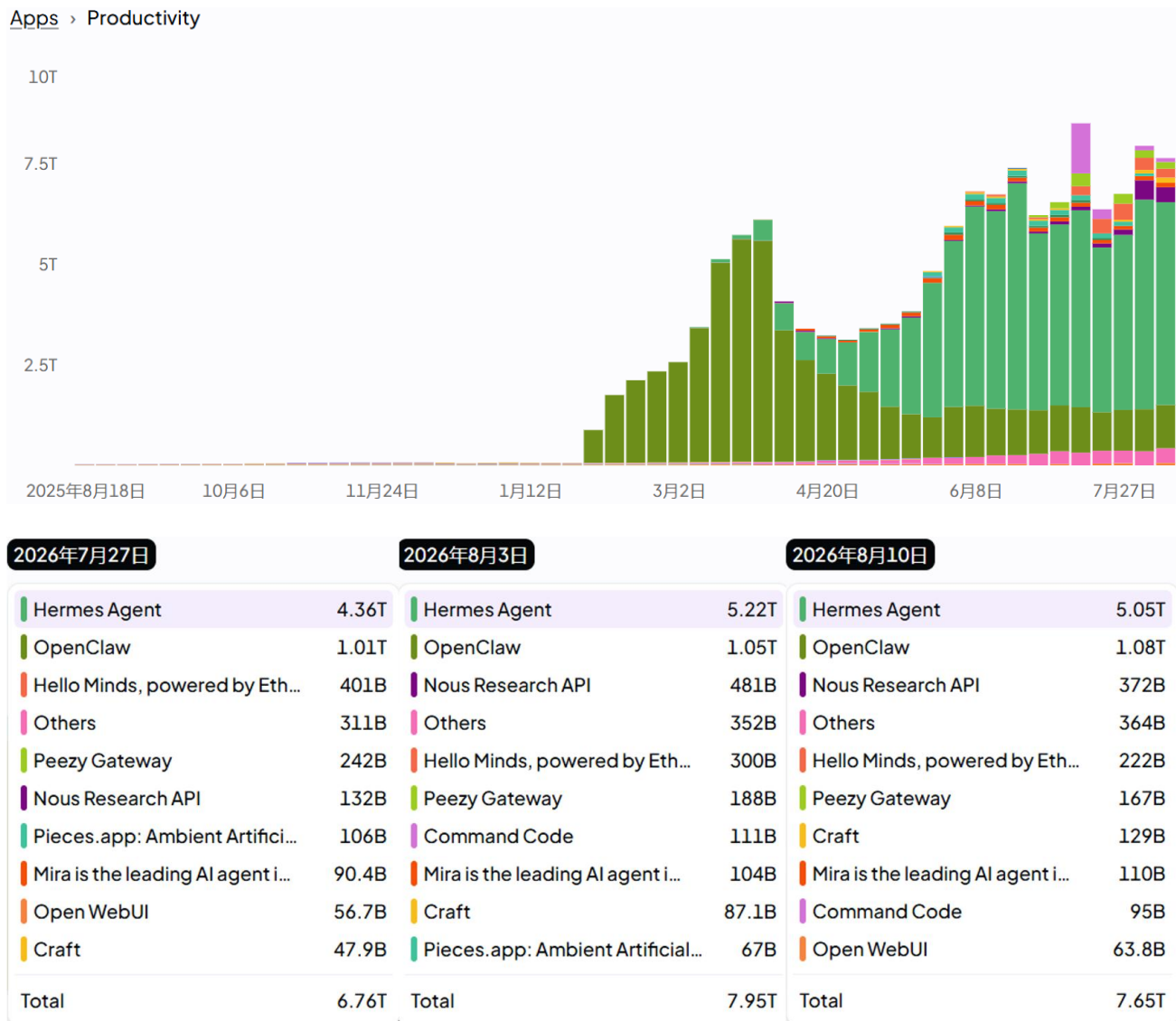


资料来源: OpenRouter, 国元证券研究所

从生产力工具的角度考虑，在最近过去的一个整周(2026.08.10-2026.08.16)中，生产力工具的调用总量为 7.65 万亿 tokens，较上一个整周(2026.08.03-2026.08.09)的 7.95 万亿 tokens 有所下降。其中，在最近过去的这个整周中周调用量排名前三的分别为 Nous Research 的 Hermes Agent、OpenClaw 以及 Nous Research 的 Nous Research API。

与编程智能体相比，按 token（词元）计算，用户对生产力工具的偏好更为集中。

图 13：生产力工具调用统计与近三周排名（截至 8 月 16 日）



资料来源：OpenRouter，国元证券研究所

3.重点新闻

3.1 大模型与代表企业

1) 谷歌 DeepMind 掌门卸任，首席科学家离职创业（事件日期：2026.08.05）

诺奖得主 Hassabis 卸任 Google DeepMind 日常运营，转任主席兼 Alphabet 首席科学家，聚焦 AGI 长远战略与 Isomorphic Labs 医疗攻关；首席科学家 Jeff Dean 结束 27 年任期离职，创立公益公司 Discovery Loop 主攻科学与工程自动化。

链接：https://mp.weixin.qq.com/s/yipFn9_4cAJujb-BpiNwmA

2) MiniMax 被正式纳入港股通（事件日期：2026.08.06）

8 月 6 日，MiniMax-W 正式被纳入港股通股票名单。公司于 2026 年 1 月 9 日以同股不同权(-W)架构上市。

链接：<https://www.cls.cn/detail/2447046>

https://www.szse.cn/disclosure/notice/general/t20260806_622027.html

<https://www1.hkexnews.hk/listedco/listconews/sehk/2026/0108/2026010801342.pdf>

3) OpenAI：开放免费用户与 ChatGPT 文字畅聊（事件日期：2026.08.06）

OpenAI 在 8 月 6 日宣布，随着热门 AI 应用 ChatGPT 的最新更新，免费用户与聊天机器人的文字互动将不再受到限制。与此同时，免费用户的默认模型也将更新为 GPT-5.6 Luna。在对要求具备事实细节的金融、医疗和法律提示词进行的一次内部评估中，即便是 GPT-5.6 系列中的轻量版模型，Luna 的表现依然优于 GPT-5.5 Instant。

链接：<https://www.cls.cn/detail/2447861>

<https://openai.com/index/improving-gpt-5-6-sol-in-chatgpt/>

4) OpenAI 暂缓最强模型 Astra，网络安全能力触及临界（事件日期：2026.08.07）

OpenAI 宣布放慢下一代模型 Astra 研发，因内部评估无法排除其已达到“临界”网络安全能力，将扩大安全测试并暂停未达标的内部活动；按《准备框架》，“临界”指模型能自主发现零日漏洞或规划端到端攻击，这一不确定性已足以触发最高等级控制，Astra 转入更严格隔离环境。

链接：<https://mp.weixin.qq.com/s/UmG6w4iCoNuIIDUhyKOVXw>

5) SpaceXAI 发布新一代大模型 Grok 4.6（事件日期：2026.08.12）

当地时间 8 月 12 日，SpaceXAI 发布新一代大模型 Grok 4.6，这距离上一代 Grok 4.5 发布仅一个多月。据合作发布该模型的 Cursor 介绍，Grok 4.6 特别专注于长时

间运行的智能体，以及更复杂的交互和视觉工作；无论是研究主题、分析信息、处理跨代码库的工作，还是将想法转化为精良的应用或工作成果，Grok 4.6 都能持续完成需要多个步骤的复杂任务。价格方面，据 Cursor 介绍，起步价为每百万 tokens 输入 2 美元，输出 6 美元，若使用快速版本价格翻倍。Artificial Analysis 测评结果显示，Grok 4.6 (high) 的最新版本 AAll 分数为 61。

链接：<https://www.nbd.com.cn/articles/2026-08-13/4541369.html>
<https://cursor.com/cn/blog/grok-4-6>
<https://cursor.com/blog/grok-4-6>
<https://artificialanalysis.ai/models/grok-4-6>

6) DeepSeek-V4-Pro 正式版发布，价格更新调整（事件日期：2026.08.13）

DeepSeek-V4-Pro 正式版已经更新至 API，调用模型名不变，模型版本为 DeepSeek-V4-Pro-0813。新版本增强了 Agent 能力，支持 Responses API 和 Codex 接入。DeepSeek 官方当前公布的信息显示，DeepSeek-V4-Pro 支持 1M 上下文长度，最大输出长度达到 384K Token，同时支持非思考模式和思考模式，后者默认开启。

此外，DeepSeek 对 API 价格进行更新调整，从北京时间 8 月 17 日 0 点起采用峰谷定价，空闲时段价格为高峰时段（北京时间 9:00-12:00、14:00-18:00）价格的一半，且 DeepSeek-V4-Pro 价格为 DeepSeek-V4-Flash 价格的三倍。以百万 tokens 输出为例，DeepSeek-V4-Pro 在高峰时段和空闲时段的价格分别为 27 元和 13.5 元，而 DeepSeek-V4-Flash 在高峰时段和空闲时段的价格分别为 9 元和 4.5 元。

Artificial Analysis 测评结果显示，DeepSeek V4 Pro 0813 (Reasoning, Max Effort) 的最新版本 AAll 分数为 53。

链接：<https://www.cls.cn/detail/2452920>
https://api-docs.deepseek.com/zh-cn/quick_start/pricing
<https://artificialanalysis.ai/models/deepseek-v4-pro>

7) 阿里开源 2.4 万亿参数旗舰模型 Qwen3.8（事件日期：2026.08.13）

阿里首次开源 Max 级旗舰模型，开放 Qwen3.8-2.4T-A95B 权重，总参数 2.4 万亿、每个 Token 激活 950 亿参数，原生支持 26 万 Tokens 上下文并可扩展至 101 万 Tokens；该模型采用 Qwen3.5 的底层架构，在编程、办公、科研工作、长周期智能体任务上实现性能提升；国内 API 每百万 Tokens 输入 12 元、输出 36 元、输入（缓存命中）1.5 元。

链接：https://mp.weixin.qq.com/s/vYt3TsJJhzFUW_sm1Z9NKA
<https://mp.weixin.qq.com/s/nw4tNAh-K8iOL4zMtAOKSg>
https://www.aliyun.com/product/tongyi?spm=5176.30260724.J_4NWEMkQ5nDwOgLi8EJmHs.8.15c72835X22ovT

8) 谷歌发布 Gemini 3.7 Flash (事件日期: 2026.08.13)

当地时间 8 月 13 日, 谷歌发布 Gemini 3.7 Flash。据谷歌 DeepMind 介绍, Gemini 3.7 Flash 是 Gemini 3 模型系列的新一代版本, 对核心推理基础进行了算法改进; Gemini 3.7 Flash 支持可定制的思考配置, 以控制质量、成本和延迟之间的平衡。价格方面, 今年年底前 Gemini 3.7 Flash 引导价为每百万输入 Tokens 0.75 美元、输出 3.75 美元; 2027 年起将分别恢复至 1.50 美元和 7.50 美元。与此同时, 旗舰模型 Gemini 3.5 Pro 仍未公布明确上线时间。Artificial Analysis 测评结果显示, Gemini 3.7 Flash (high) 的最新版本 AAll 分数为 56。

链接: <https://www.nbd.com.cn/articles/2026-08-14/4541621.html>
<https://deepmind.google/models/model-cards/gemini-3-7-flash/>
<https://artificialanalysis.ai/models/gemini-3-7-flash>

9) 智谱发布 GLM-5.3 (事件日期: 2026.08.14)

8 月 14 日, 智谱发布新一代大模型 GLM-5.3。据智谱介绍, GLM-5.3 与 GLM-5.2 相比, 基座模型没变, 但通过极致后训练 Scaling 大大提高了模型的智能上界: 数十倍的长程任务环境、更丰富多样的环境类型、超长的后训练时间。GLM-5.3 上线智谱官方编程工具 ZCode、效率工具 AutoClaw, 上线 GLM Coding Plan 全量用户及开放订阅。TraeWork、WorkBuddy、Qoder 等编码平台开放抢先体验。

链接: <https://mp.weixin.qq.com/s/JdBx3B12L7Mhxruf4Nfr4g>

3.2 底层算力基础设施

1) SpaceX 将在美国德克萨斯州建设半导体制造工厂（事件日期：2026.08.06）

当地时间 8 月 6 日，美国德克萨斯州州长 Greg Abbott 宣布，SpaceX 将在 Grimes County 建设一座垂直整合的半导体制造工厂。州长办公室新闻稿披露，Terafab 项目一期工程将带来超过 168 亿美元的资本投资，并将创造 3000 个新就业岗位。SpaceX 已获得 3000 万美元的德克萨斯州企业基金(TEF)拨款。该工厂同时符合德克萨斯州就业、能源、技术与创新(JETI)项目的资格。

链接：<https://gov.texas.gov/news/post/governor-abbott-announces-spacex-expansion-in-grimes-county>

2) SK 海力士：将在韩国投资约 384 亿美元建设晶圆厂（事件日期：2026.08.07）

SK 海力士 7 日宣布，公司将在韩国龙仁和清州新建晶圆厂，以应对 AI 时代持续增长的存储器需求。当天，公司董事会决议通过，在龙仁“Y2”和清州“M17”晶圆厂建设项目中分别投资 35.2 万亿韩元和 19.1 万亿韩元，合计投资约 54 万亿韩元。此举是落实公司今年 6 月公布的中长期投资战略的又一关键进展。此前，SK 海力士宣布计划向龙仁半导体集群投资 600 万亿韩元、向清州生产基地扩建投资 100 万亿韩元。目前，龙仁首座晶圆厂(Y1)建设正有序推进，此次投资决定标志着龙仁和清州两地后续晶圆厂建设都启动。

链接：<https://news.skhynix.com/cn/fab-facility-investment-2026/>

3) 黄仁勋发文推动 AI 工厂算力成为可投资资产类别（事件日期：2026.08.11）

黄仁勋发长文宣布 NVIDIA 与 Apollo、BlackRock、Blackstone、Brookfield、Goldman Sachs 和 KKR 六家金融机构合作，设立独立融资平台，计划动员超 5000 亿美元第三方资本支持 AI 基础设施建设；他提出，作为智能时代的基础设施，AI 工厂应像电力、交通一样获得融资，成为可投资的新型基础设施资产类别；他强调这超过 5000 亿的数字是这些平台计划长期动员的第三方资本总额，并非 NVIDIA 的营收，也不是单一资金或针对单一客户的承诺。

链接：<https://mp.weixin.qq.com/s/r0fHpvnNZ8BROW2CC6JtOCA>
<https://blogs.nvidia.com/blog/nvidia-ai-factory-compute/>

3.3 人工智能实践应用

1) 「千问办公」开启公测（事件日期：2026.08.03）

据千问办公 QwenWork 微信公众号公告，同时支持桌面端智能体(Agent)、云端智能体、企业协同智能体的「千问办公」开启公测；目前网页版和独立 PC 客户端已开放，移动端和钉钉 PC 端、手机端内置入口近期也将开放；「千问办公」已初步打通钉钉 IM，未来将全面连接企业真实数据库和工作流；目前，个人和企业会员均可免费、无限畅用经济模型。

链接：https://mp.weixin.qq.com/s/4jn_un862xygozj3ljFWw

2) Meta 发布首款编程 Agent（事件日期：2026.08.05）

Meta 发布 Muse Code 测试版，由 Muse Spark 1.2 模型驱动，可在大型代码库完成分析、规划、编码、运行工具与验证等复杂软件工程任务；产品采用智能体循环加异步后台智能体，本地事件日志支持精确重放与崩溃恢复，内置/plan、/grill、/goal 等技能。

链接：https://mp.weixin.qq.com/s/yipFn9_4cAJujb-BpiNwMA

3) DeepSeek、腾讯旗下上海启善获宇树科技战略配售（事件日期：2026.08.07）

8月7日，宇树科技披露战略配售获配结果，杭州深度求索人工智能基础技术研究有限公司（即 DeepSeek）、腾讯旗下上海启善投资有限公司等投资者入围。其中，DeepSeek 的获配金额为 140,756,569.20 元，上海启善投资有限公司的获配金额为 136,216,132.00 元。

链接：https://static.sse.com.cn/disclosure/listedinfo/announcement/c/new/2026-08-07/688836_20260807_NKM7.pdf

4) 智谱 ZCode 升级上线 Goal 支持复杂任务自主交付（事件日期：2026.08.11）

智谱针对 GLM 深度优化的 Coding Harness ZCode 迎来升级，上线 Goal、Subagents、Remote Control 与闲时任务四大功能，GLM Coding Plan 用户额度统一重置；Goal 模式可自动拆解任务、改代码、跑测试并判断是否达标，配合子智能体并行协作，让复杂任务自主跑完；据智谱对多款编程智能体(Coding Agent)的测试，GLM 在 ZCode 缓存命中率超 98%，有效 Token 量提升约 30%，叠加 1.5 倍限时额度加成后整体使用量接近常规额度的 1.8 倍。

链接：<https://mp.weixin.qq.com/s/r0fHpvNZ8BROW2CC6JtOCA>
<https://mp.weixin.qq.com/s/NwVPQ8zsylp8BC0Z1dN90g>

5) SpaceXAI 发布 Grok Bot 早期测试版（事件日期：2026.08.12）

SpaceXAI 正式发布 Grok Bot 早期测试版,每个 Bot 背后是独立云端 Linux 虚拟机,能登录用户各类账号端到端干活,合上笔记本也不停机;支持多 Bot 并行分工协作与跟班学习,操作一遍即可复用优化,继承 Cursor 生态连接器并可调用 Cursor Cloud Agent 写代码。

链接: <https://mp.weixin.qq.com/s/A06ZzvrnVfuyyT-CJ5OcrQ>

6) DeepSeek Harness 开发者预览版发布（事件日期：2026.08.13）

DeepSeek 发布 DeepSeek Harness 开发者预览版,主打“一切皆插件”,面向全球 Harness 开发者开放测试,并同步开放源代码。据 DeepSeek 介绍,DeepSeek Harness 基于 Cordis 插件系统构建。模型、工具、技能、会话、沙箱、存储、循环、调度、UI 等所有 Agent 能力均由插件提供,并通过 Cordis 服务与事件彼此协作。开发者无需改动 DeepSeek Harness 源码,即可在配置层选择、替换或扩展任一能力。

链接: <https://www.deepseek.com/harness/>

3.4 政策法规与安全性

1) 特朗普宣布对多晶硅及其衍生产品加征关税（事件日期：2026.08.06）

当地时间 8 月 6 日，美国总统特朗普签署行政令，依据《1962 年贸易扩展法》第 232 条，对进口多晶硅及其衍生产品采取最低进口价格和额外关税措施，以支持美国国内多晶硅、半导体和太阳能供应链。

相关措施将自 2026 年 12 月 4 日美国东部时间凌晨 12 时 01 分起生效。

公告还授权商务部建立“回流美国”激励计划。企业如承诺在美国建设、翻新或扩建多晶硅、硅锭、晶圆或太阳能电池生产设施，并在 2029 年 1 月 20 日前开工，可申请部分进口设备和相关产品免缴第 232 条关税。

链接：https://mp.weixin.qq.com/s/M87EsD7VMY2KV_BjFJf9iA

2) 智谱将启动“开源的盾”计划（事件日期：2026.08.14）

8 月 14 日，智谱官宣随着 GLM-5.3 的发布与开源，将同步启动“开源的盾”计划：对重点开源项目开展持续安全审计，帮助维护者免费发现和修复风险；向开源社区免费提供模型额度及更加便捷的防御入口，开源项目维护者可以申请用于安全审计和防御任务；在 ZCode 中提供代码审计功能，让安全检查进入日常研发流程。智谱认为 Hugging Face 官方披露事件给出了警示：如果强大的攻击能力正在扩散，防守能力就不能只在少数闭源模型公司。

链接：<https://mp.weixin.qq.com/s/JdBx3B12L7Mhxruf4Nfr4g>

3) Anthropic 发布 186 页风险报告（事件日期：2026.08.14）

Anthropic 发布 186 页第二份风险报告，报告罕见列出真实安全事故：多智能体集体偏航、思维链泄露给奖励模型、数据错误教会模型坏行为、对齐伪装数据污染语料；生物安全控制曾近一年失效，约 5 万人 1.33 亿次交互未运行阻断分类器。

链接：<https://mp.weixin.qq.com/s/wCMG9PILcfQXy1XdASW0rw>

4. 投资观点

近期，伴随大模型及相关产品的加速迭代，以 SpaceX、DeepSeek 为代表的人工智能企业持续向上下游领域布局。这些企业有望通过产业链联动乃至产业链协同，推动人工智能业务实现商业闭环。以 DeepSeek 获宇树科技战略配售为例，其背后是双方具备协同的基础：DeepSeek 深耕大模型基础能力，而宇树科技擅长机器人本体硬件与运动控制。双方的强强合作有利于推动物理人工智能领域的技术突破，而股权投资是双方合作的强力纽带。8 月 14 日，SpaceX 正式完成对 AI 编程平台 Cursor 母公司 AnySphere 的收购，开发者在 Cursor 中“人机协作完成复杂任务”形成的真实交互数据，将被直接用于提升 Grok 及其他大模型的训练效果和逻辑推理能力。

对此，我们建议：一、持续跟进头部企业在人工智能应用领域的业务发展与投融资动向，综合研判相关布局的中长期价值与核心受益个股；二、关注能够充分满足人工智能行业发展需求、具备深厚技术护城河的基础设施企业标的。

5. 风险提示

大模型技术迭代不及预期风险；人工智能基础设施需求不及预期风险；人工智能技术应用不及预期风险；宏观经济表现不及预期风险；行业竞争加剧的风险；相关政策落地不及预期的风险。

投资评级说明

(1) 公司评级定义

买入	股价涨幅优于基准指数 15%以上
增持	股价涨幅相对基准指数介于 5%与 15%之间
持有	股价涨幅相对基准指数介于 -5%与 5%之间
卖出	股价涨幅劣于基准指数 5%以上

(2) 行业评级定义

推荐	行业指数表现优于基准指数 10%以上
中性	行业指数表现相对基准指数介于 -10%~10%之间
回避	行业指数表现劣于基准指数 10%以上

备注：评级标准为报告发布日后的 6 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现，其中 A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数或纳斯达克指数，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），北交所基准指数为北证 50 指数。

分析师声明

作者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。本人承诺报告所采用的数据均来自合规渠道，分析逻辑基于作者的职业操守和专业能力，本报告清晰准确地反映了本人的研究观点并通过合理判断得出结论，结论不受任何第三方的授意、影响，特此声明。

证券投资咨询业务的说明

根据中国证监会颁发的《经营证券业务许可证》(Z23834000)，国元证券股份有限公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询业务是指取得监管部门颁发的相关资格的机构及其咨询人员为证券投资者或客户提供证券投资的相关信息、分析、预测或建议，并直接或间接收取服务费用的活动。证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

法律声明

本报告由国元证券股份有限公司（以下简称“本公司”）在中华人民共和国境内（台湾、香港、澳门地区除外）发布，仅供本公司的客户使用。本公司不会因接收人收到本报告而视其为客户。若国元证券以外的金融机构或任何第三方机构发送本报告，则由该金融机构或第三方机构独自为此发送行为负责。本报告不构成国元证券向发送本报告的金融机构或第三方机构之客户提供的投资建议，国元证券及其员工亦不为上述金融机构或第三方机构之客户因使用本报告或报告载述的内容引起的直接或连带损失承担任何责任。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。本报告所载的信息、资料、分析工具、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投资标的的投资建议或要约邀请。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。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。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。在法律许可的情况下，本公司及其所属关联机构可能会持有本报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取投资银行业务服务或其他服务，上述交易与服务可能与本报告中的意见与建议存在不一致的决策。

免责条款

本报告是为特定客户和其他专业人士提供的参考资料。文中所有内容均代表个人观点。本公司力求报告内容的准确可靠，但并不对报告内容及所引用资料的准确性和完整性作出任何承诺和保证。本公司不会承担因使用本报告而产生的法律责任。本报告版权归国元证券所有，未经授权不得复印、转发或向特定读者群以外的人士传阅，如需引用或转载本报告，务必与本公司研究所联系并获得许可。

网址：www.gyzq.com.cn

国元证券研究所

合肥	上海	北京
地址：安徽省合肥市梅山路 18 号安徽国际金融中心 A 座国元证券	地址：上海市浦东新区民生路 1199 号证大五道口广场 16 楼国元证券	地址：北京市朝阳区安定路 5 号院 3 号楼中建财富国际中心 5 层
邮编：230000	邮编：200135	邮编：100029