下载日志已记录，仅供内部参考

计算机

报告日期：2026 年 07 月 10 日

Anthropic 或扩大盈利，国产模型驱动融资

——2026.07.04—2026.07.10 周报

投资要点

- **本周大模型行业重点事件集中在 Anthropic 商业化与算力扩张、通用工作智能体升级、国内头部模型厂商再融资三方面。**
- **一是 Anthropic 盈利规模有望继续扩大，Coding 与 API 业务支撑高增长、高毛利的商业模式，并以长约锁定算力供给。**此前《华尔街日报》预计 Anthropic 2026Q2 营业利润或达 5.59 亿美元，如实现将是公司成立以来首次盈利；SemiAnalysis 进一步预测，公司 2026Q3 ARR 或超过 600 亿美元、GAAP EBIT 可能超过 10 亿美元，综合毛利率已由 2024 年的-94%升至 60%以上，其中 API 业务毛利率超过 80%。收入结构方面，超过 65%的 ARR 来自编程场景，75%—85%来自按使用量计费的 API 业务，消费者订阅仅占约 5%；通过 AWS Bedrock、Azure Foundry 等云平台销售的 TaaS 收入占比已由上一季度的 5%—10%升至 15%—20%。算力上，到 2030 年 Anthropic 与 OpenAI 合计算力需求将超过 100 GW，而两家目前合计可用算力刚超过 6 GW；Anthropic 计划在澳大利亚采购至少 1.4GW 数据中心容量，并与 TeraWulf 签署 401MW、20 年期数据中心租约。整体来看，若上述预测兑现，Anthropic 盈利表现将由 Q2 预计首次转正进一步迈向 Q3 规模扩大，Coding 需求与 API 商业模式的高毛利特征持续得到验证，但算力供给仍将是后续增长的核心约束。
- **二是 OpenAI 发布 GPT-5.6 家族并整合产品入口，通用智能体加速从编程工具向跨应用知识工作入口演进。** OpenAI 正式全面开放 GPT-5.6 Sol、Terra 和 Luna，其中 Sol 在 Agents' Last Exam 上达到 53.6 分，领先 Claude Fable 5 达 13.1 分，Terra 与 Luna 以约 1/16 的成本超过 Fable 5。产品侧，ChatGPT Work 内置 Codex 技术并由 GPT-5.6 驱动，可跨应用、文件和浏览器持续执行复杂任务，直接生成表格、幻灯片、文档及 Sites；Codex 周活跃用户已超过 500 万，其中超过 100 万用于软件开发之外的工作。Anthropic 同步将 Claude Cowork 扩展至网页端和移动端，支持任务在云端后台运行及跨设备接续；国内方面，豆包与千问均将在 7 月 15 日下线原有智能体功能， C 端智能体入口收缩。整体来看，前沿模型的代际领先窗口进一步缩短，单位成本性能、知识工作产品以及可跨端、跨应用长时间执行任务的统一工作平台逐渐成为竞争重心。
- **三是智谱与 MiniMax 密集启动大规模再融资，国内模型厂商加速补充算力与研发资本。** 智谱拟配售最多 1,978 万股新 H 股，预计净募资约 313.75 亿港元，主要用于大模型研发与人才培养、算力采购或租赁、算力调度体系建设、业务拓展及战略投资；MiniMax 拟通过配售新股及发行零息可转债合计净募资约 159.57 亿港元，其中 80%用于 AI 基础设施、模型训练与推理研发，10%用于 MiniMax Code 等 Harness 智能体产品的全球商业化。两家公司合计净募资近 470 亿港元，叠加 MiniMax 筹备 2.7 万亿参数开源模型，国内模型厂商的研发与基础设施投入仍在加码。整体来看，国产大模型竞争正由模型能力比拼进一步延伸至融资能力、算力储备与全球商业化，资本实力成为支撑模型持续迭代的重要基础。
- **风险提示**

技术迭代不及预期；盈利改善不及预期；资本开支压力超预期；融资与商业化进展不及预期；监管及安全风险；头部厂商竞争加剧。

行业评级：看好(维持)

分析师：刘雯蜀
执业证书号：S1230523020002
liuwenshu03@stocke.com.cn

分析师：郑毅
执业证书号：S1230524070002
zhengyi@stocke.com.cn

相关报告

1 《中国大模型获海外认可》 2026.07.04

2 《AI 越火，安全越贵》 2026.07.02

3 《CPO 开启光路通信新格局》 2026.06.28

下载日志已记录，仅供内部参考

正文目录

1 国内厂商动态 4
1.1 字节跳动 4
1.2 通义 4
1.3 智谱 5
1.4 MiniMax 5
1.5 Kimi 6
2 海外厂商动态 6
2.1 Anthropic 6
2.2 OpenAI 7
2.3 Gemini 8
3 风险提示 8

下载日志已记录，仅供内部参考

图表目录

图 1： Seedream 5.0 Pro 直接用于高信息密度的内容生产 4
图 2： Seedream 5.0 Pro 能深刻解析画面中的位置信息 4
图 3： Fun-ASR-Flash 在 Artificial Analysis 错字率全球最低 4
图 4： MiniMax 再融资所得款项用途及分配计划 5
图 5： SemiAnalysis 对 Anthropic 2026Q1—Q2 经营数据的预测与实际值对比 6
图 6： GPT-5.6 在 Agents' Last Exam 中实现更高性能与更低成本 7
图 7： GPT-5.6 Sol 在编程智能体评测以更低成本成为 SOTA 7
图 8： GPT-Live 提供快速、自然的响应，同时由 GPT-5.5 在后台处理搜索任务 8

下载日志已记录，仅供内部参考

1 国内厂商动态

1.1 字节跳动

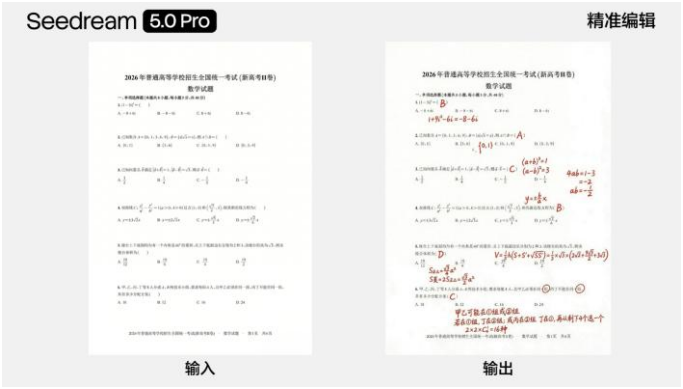
- **豆包智能体功能将于 7 月 15 日下线，引导用户转向猫箱 App。** 智能体信息及历史对话数据将保留一段时间，但 2026 年 10 月 15 日后将无法在豆包内查看或恢复；字节跳动旗下的猫箱 App 可创建新的智能体、开启对话服务，如有需要，后续可在猫箱 App 开启智能体新体验。
- **7 月 8 日正式发布多模态图像创作模型 Seedream 5.0 Pro。** 重点提升图文匹配、结构合理性、文字渲染和画面美感，并强化复杂信息可视化、精准编辑、真实影像质感及多语种生成；模型已上线火山方舟，豆包、即梦将陆续接入。

图1： Seedream 5.0 Pro 直接用于高信息密度的内容生产



资料来源：公司官网，浙商证券研究所

图2： Seedream 5.0 Pro 能深刻解析画面中的位置信息

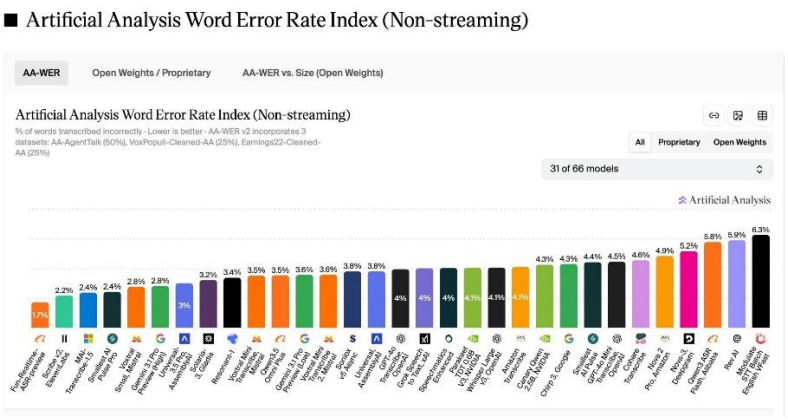


资料来源：公司官网，浙商证券研究所

1.2 通义

- **升级实时语音识别模型 Fun-ASR-Realtime。** 首字延迟为百毫秒级，识别准确率接近离线模型，支持 16 种方言和 30 种语言，适用于直播字幕、会议转写及客服等低延迟场景；在影视飓风「重返荒岛」100 小时直播中，实时字幕由 Fun-ASR-Realtime 全程提供，累计识别字幕 6 万余条、共计 132 万字。**离线模型 Fun-ASR-Flash 近期也已上线**，在评测平台 Artificial Analysis 上，Fun-ASR-Flash（榜单名为“Fun-realtime-ASR-preview”）错字率为 1.7%，位列全球第一。

图3： Fun-ASR-Flash 在 Artificial Analysis 错字率全球最低



资料来源：阿里云公众号，浙商证券研究所

下载日志已记录，仅供内部参考

- **阿里内部全面禁用 Claude Code，推荐 Qoder 作为替代方案。** 因近期 Claude Code 被曝存在植入后门的安全风险，阿里经综合评估后已将其列入高风险软件名单；自 7 月 10 日起，阿里将全面禁止内部员工在办公环境下使用 Claude Code。
- **千问向用户推送“千问智能体服务下线提醒”。** 千问智能体功能与服务将于 2026 年 7 月 15 日正式下线，下线后用户将无法继续访问相关智能体配置及历史对话记录。
- **通义千问对接中国邮政 EMS 系统，提供录取通知书物流实时提醒。** 配套志愿日历、个性化志愿报告工具，截至 7 月 8 日已有 2300 万份免费志愿报告被用户领取。

1.3 智谱

- **拟配售最多 1,978 万股新 H 股，预计净募资约 313.75 亿港元。** 以每股 1,588 港元向不少于 6 名独立专业、机构及其他投资者配售，预计募集资金总额约 314.11 亿港元；配售价较 7 月 8 日收盘价折让 12.99%，预计于 7 月 13 日完成。募集资金计划于 2027 年底前使用，主要投向大模型研发和人才培养、算力采购或租赁、相关前置贷款偿还、技术服务及算力调度体系建设，同时用于业务拓展、战略投资与并购，以及补充运营资金。

1.4 MiniMax

- **据路透社报道，计划推出 2.7 万亿参数规模的新一代大模型（M3 总参数 428B，激活参数 23B）。** 若发布将是中国企业开发的规模最大的开源模型，也可能成为全球最大的此类模型。
- **创始人兼 CEO 闫俊杰发布内部全员信，回应近期市场波动，并强调公司长期方向不变。** 宣布从即日起到公司实现 AGI 的那一天，他将不再从公司领取任何薪酬；未来四年，他将拿出个人名下相当于公司总股本 4% 的股份，用于激励团队成员；同时拿出 1% 的股份设立专项基金，持续支持相关开源社区的发展。
- **拟在上市约半年后再次融资：** 拟以每股 268 港元配售 3,560 万股新 A 类普通股，预计净募资约 94.91 亿港元；同时 65 亿港元同时发行 2027 年到期的零息有担保可转债，初始转股价为每股 335 港元，预计净募资约 64.66 亿港元；两项交易合计净募资约 159.57 亿港元，资金计划用途见下表。

图4： MiniMax 再融资所得款项用途及分配计划

所得款項淨額擬定用途	佔所得款項淨額總額的概約百分比	所得款項淨額分配（百萬港元）	動用所得款項淨額的預期時間表
繼續加強AI基礎設施及模型研究與開發	80.00%	12,765.78	2027年12月31日或之前
加快Harness產品的全球商業化及開發	10.00%	1,595.72	2029年6月30日或之前
營運資金及一般企業用途	10.00%	1,595.72	2029年6月30日或之前
總計	**100.00%**	**15,957.23**	

资料来源：公司公告，浙商证券研究所

下载日志已记录，仅供内部参考

1.5 Kimi

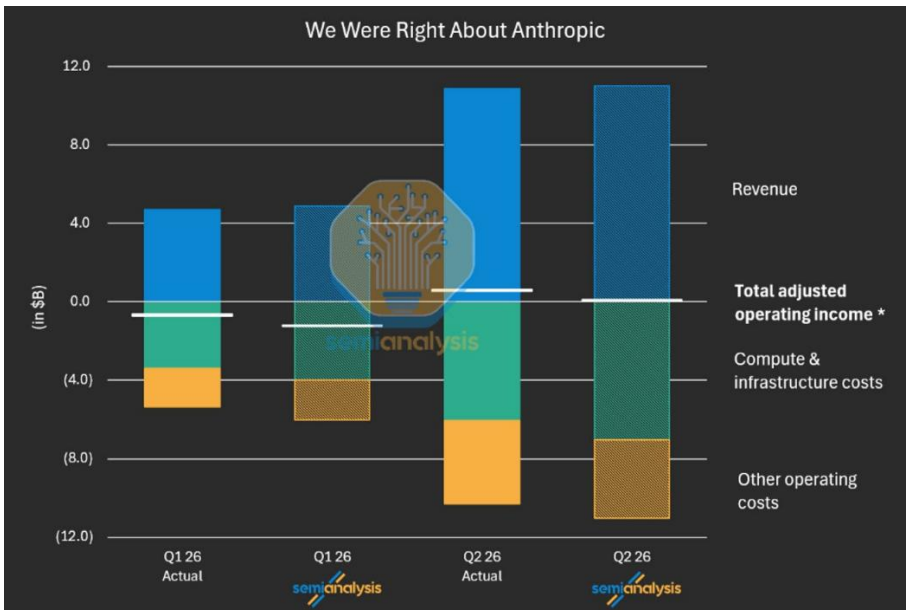
- **本周暂无重要更新。**

2 海外厂商动态

2.1 Anthropic

- **据 SemiAnalysis，26Q3 利润可能超过 10 亿美元，ARR 增长至 600 亿美元以上：**
  - ✓ 若按此速度持续（即每月净新增 ARR 约 150 亿美元），到 2027 年底，Anthropic ARR 有望达 3000 亿美元；
  - ✓ 目前超过 65% 的 ARR 来自与编程相关的应用场景，其中 Cursor、Cognition、Loveable 和 Replit 等编程工具类创业公司合计贡献了约 60 亿美元 ARR；
  - ✓ 约 75%—85% 的 ARR 来自按使用量计费的 API 业务，消费者订阅仅占 ARR 总额的 5%；作为对比，OpenAI 在 2026Q1 超过 65% 的收入来自订阅模式，消费者 ARR 占比约为 40%；
  - ✓ 目前综合毛利率已升至 60% 以上（2024 年综合毛利率为 -94%），其中 API 业务的毛利率超过 80%；
  - ✓ 分销方面，通过 AWS Bedrock、Azure Foundry 等超大规模云平台进行间接销售的“代币即服务”（TaaS）模式正快速增长：该渠道目前占 ARR 的 15%—20%，一个季度前仅为 5%—10%（需向云平台支付 20%—30% 的收入分成）；
  - ✓ 到 2030 年，Anthropic 与 OpenAI 两家合计算力需求将超过 100 GW，而 2025 年和 2026 年的净新增算力分别仅为 2.5 GW 和 5 GW，两家公司目前合计可用算力刚超过 6 GW。

图5： SemiAnalysis 对 Anthropic 2026Q1—Q2 经营数据的预测与实际值对比



资料来源：SemiAnalysis，浙商证券研究所

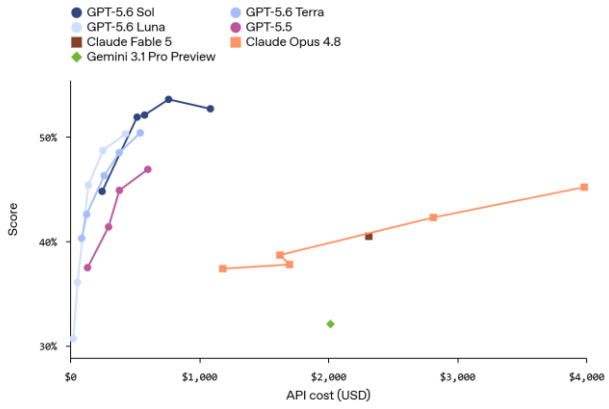
下载日志已记录，仅供内部参考

- 计划在澳大利亚采购至少 1.4 GW 数据中心算力，对应建设成本约 150 亿美元，高于市场此前预期。 公司目标在 2027 年底前启用至少 1 GW 算力，并已在澳大利亚设立办公室推进本地化基础设施布局。
- 与 TeraWulf 签署 20 年 AI 数据中心租约，合同收入约 190 亿美元。Anthropic 将租用位于美国肯塔基州专为 AI 打造的数据中心设施，总关键 IT 负载规模约 401 MW；项目将分阶段建设，预计首批产能于 2027 年下半年投运，并于 2028 年初全面投入运营；TeraWulf 预计该租约在初始租期内将带来约 190 亿美元的合同收入。
- 首席财务官 Krishna Rao 在播客节目中披露，公司净收入留存率（NRR）高达 500%。 意味着在 Q1 贡献了 300 亿美元 ARR 的客户，一年前贡献的收入仅为 20 亿美元。
- Claude Cowork 开始向移动端与网页端开放。 任务可在云端后台运行，并在电脑、网页及手机间跨端接续，即使没有设备联网，预定好的任务也会继续执行；在网页版和桌面版上，Chat 与 Cowork 合二为一。
- Claude Fable 5 免费使用期延长至 7 月 12 日。付费用户最多可使用每周额度的 50%，期满后需转用 usage credits。

2.2 OpenAI

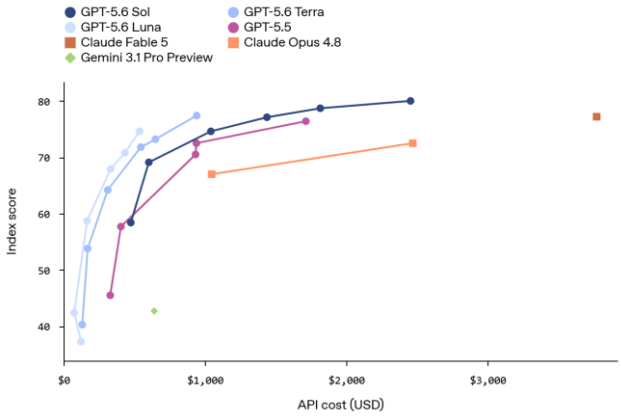
- **GPT-5.6 模型家族于当地时间 7 月 9 日正式全面开放。** 旗舰 Sol、均衡款 Terra、低成本款 Luna 同步推出，主打 token 效率与降本；在 Agents' Last Exam 上，Sol 以 53.6 创下新高、领先 Fable 5 达 13.1 分，中等推理档以约 1/4 成本领先 11.4 分，Terra 与 Luna 以约 1/16 成本反超 Fable 5。

图6： GPT-5.6 在 Agents' Last Exam 中实现更高性能与更低成本



资料来源：公司官网，浙商证券研究所

图7： GPT-5.6 Sol 在编程智能体评测以更低成本成为 SOTA



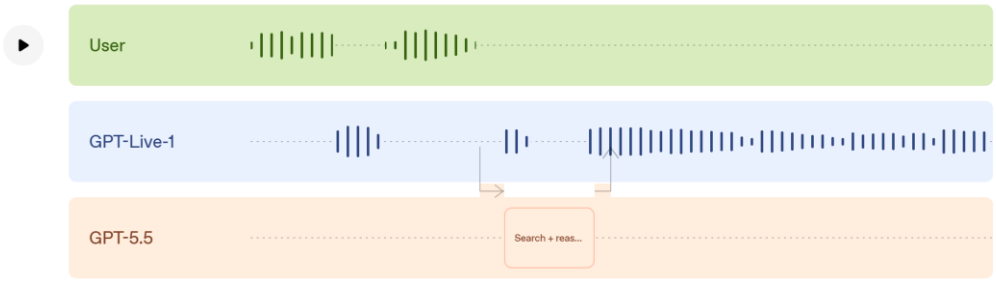
资料来源：公司官网，浙商证券研究所

- **Codex 与 ChatGPT 合并为 ChatGPT Work。** ChatGPT Work 内置 Codex 技术、由 GPT-5.6 驱动，可跨应用与文件自主行动、持续处理复杂项目数小时，产出表格、幻灯片、文档及 Sites；支持定时任务、桌面端内置浏览器与 Computer Use；Codex 应用并入新版 ChatGPT 桌面端，Atlas 将逐步下线。
- **每周超 500 万人使用 Codex，其中超 100 万人将其用于软件开发之外的工作。**

下载日志已记录，仅供内部参考

- **推出新语音模型 GPT-Live。** 采用全双工架构，可同时听与说，复杂任务可交由后台 GPT-5.5 处理；随新前沿模型发布，将持续更新 GPT-Live 所使用的模型。

图8： GPT-Live 提供快速、自然的响应，同时由 GPT-5.5 在后台处理搜索任务



资料来源：公司官网，浙商证券研究所

- **据《金融时报》报道，已与美国政府展开初步谈判，提议出让 5% 的公司股权。** 奥特曼还建议美国所有主要 AI 公司都拿出同等比例的股权，纳入一个类似阿拉斯加永久基金（Alaska Permanent Fund）的公共投资工具，向美国政府和居民分红；这项提议目前仍处于早期概念阶段。

2.3 Gemini

- **下一场 Made by Google 硬件发布会将于 8 月 12 日在纽约举行。** 可能将推出 Pixel 11 系列。

3 风险提示

技术迭代不及预期。 若模型能力、Coding 与 Agent 等核心方向的迭代节奏放缓，可能影响产品竞争力与商业化进度。

盈利改善不及预期。若推理成本下降慢于价格竞争与调用量增长，或云平台渠道分成比例进一步上升，毛利率与盈利改善的持续性可能不及预期。

资本开支压力超预期。 超长期租约与数据中心投入持续加码，若算力投入回报周期拉长，将加大头部厂商现金流与融资压力。

融资与商业化进展不及预期。 相关公司融资安排能否按计划完成存在不确定性；企业级部署与 Agent 产品的实际付费转化与续费若不及预期，将拖累收入释放节奏。

监管及安全风险。 出口管制、数据与内容合规及政府持股等政策安排存在不确定性；若模型或工具被曝安全漏洞，可能影响相关产品商业化推进。

头部厂商竞争加剧。 模型、价格、生态与人才竞争持续升温，可能压缩利润空间并加大份额波动。

下载日志已记录，仅供内部参考

股票投资评级说明

以报告日后的 6 个月内，证券相对于沪深 300 指数的涨跌幅为标准，定义如下：
1.买　入 ： 相对于沪深 300 指数表现 + 20％以上；
2.增　持 ： 相对于沪深 300 指数表现 + 10％ ～ + 20％；
3.中　性 ： 相对于沪深 300 指数表现 - 10％ ～ + 10％之间波动；
4.减　持 ： 相对于沪深 300 指数表现 - 10％以下。

行业的投资评级：

以报告日后的 6 个月内，行业指数相对于沪深 300 指数的涨跌幅为标准，定义如下：
1.看　好 ： 行业指数相对于沪深 300 指数表现 + 10%以上；
2.中　性 ： 行业指数相对于沪深 300 指数表现 - 10% ～ + 10%以上；
3.看　淡 ： 行业指数相对于沪深 300 指数表现 - 10%以下。
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。投资者不应仅仅依靠投资评级来推断结论。

法律声明及风险提示

本报告由浙商证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格，经营许可证编号为：Z39833000）制作。本报告中的信息均来源于我们认为可靠的已公开资料，但浙商证券股份有限公司及其关联机构（以下统称“本公司”）对这些信息的真实性、准确性及完整性不作任何保证，也不保证所包含的信息和建议不发生任何变更。本公司没有将变更的信息和建议向报告所有接收者进行更新的义务。
本报告仅供本公司的客户作参考之用。本公司不会因接收人收到本报告而视其为本公司的当然客户。
本报告仅反映报告作者的出具日的观点和判断，在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求。对依据或者使用本报告所造成的一切后果，本公司及/或其关联人员均不承担任何法律责任。
本公司的交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。本公司没有将此意见及建议向报告所有接收者进行更新的义务。本公司的资产管理公司、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。
本报告版权均归本公司所有，未经本公司事先书面授权，任何机构或个人不得以任何形式复制、发布、传播本报告的全部或部分内容。经授权刊载、转发本报告或者摘要的，应当注明本报告发布人和发布日期，并提示使用本报告的风险。未经授权或未按要求刊载、转发本报告的，应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。



浙商证券研究所

上海总部地址：杨高南路 729 号陆家嘴世纪金融广场 1 号楼 25 层
北京地址：北京市东城区朝阳门北大街 8 号富华大厦 E 座 4 层
深圳地址：广东省深圳市福田区广电金融中心 33 层
上海总部邮政编码：200127
上海总部电话：(8621) 80108518
上海总部传真：(8621) 80106010
浙商证券研究所：https://www.stocke.com.cn