



计算机行业研究

买入（维持评级）
行业周报

证券研究报告

计算机组

 分析师：刘高畅（执业 S1130525120005）
 liugaochang@gjzq.com.cn

 分析师：陈芷婧（执业 S1130525120008）
 chenzhijing@gjzq.com.cn

 分析师：鲍淑娴（执业 S1130526020002）
 baoshuxian@gjzq.com.cn

北美 AI 应用公司显现增长驱动

本周观点

■ **北美应用业绩普遍超预期, AI 已从产品功能传导至经营指标。**近两周, 北美应用厂商批量发布业绩, 包括 Palantir、AppLovin、Cloudflare、Atlassian、Unity、Twilio、HubSpot、Duolingo、Doximity 等, 覆盖企业应用、广告营销、网络安全及 Agent 基础设施、通信、教育、医疗等多个方向, 多家公司收入、客户、订单及使用量保持较快增长。更值得关注的是, AI 正在由新产品、新功能进一步传导至经营指标: Palantir 表示 AIP 持续驱动美国商业业务增长; Atlassian MCP 月活突破 100 万, 采用客户 ARR 增速约为未采用客户的 2 倍; Unity 明确表示 Unity Ads Network 增长由 Unity Vector 驱动; Doximity AI Search 查询量环比增长超 25%; Twilio 的 AI 客户已开始组合使用 Conversation Orchestrator、Memory 和 Agent Connect 等产品。整体来看, AI 应用正加速进入企业核心业务流程, 并逐步体现为收入增长、客户扩张、使用量提升及商业化能力增强。

■ **开源模式释放部署需求, 企业级 AI 加速落地。**一方面, 开放权重模型显著降低高质量模型的获取和部署成本; DeepSeek 等模型持续优化价格性能比, 也使更多高频任务具备商业化落地的经济性。另一方面, 相较于闭源 API+RAG 主要围绕知识库、提示词和模型调用开展应用, 开放权重模型能够支持企业结合自身数据、业务流程和部署环境进行更深入、更丰富的开发, 使 AI 更容易进入财务、研发、工业、营销和企业管理等核心业务流程。随着模型使用成本下降、开发自由度提升, 企业 AI 有望由少量试点走向规模化应用, 模型调用、适配、部署及持续运营需求同步增长, 进一步扩大模型服务、云平台 and 算力基础设施的需求空间。

■ **云与应用商业闭环加速形成, 中国软件企业焕发新活力。**1) 云侧, 企业级 AI 部署需求释放后, 云计算是最直接的受益方向。近期微软、Google 和 AWS 最新财报继续验证, AI 需求正在由基础设施逐步传导至平台服务、企业软件和持续付费, 云厂商 AI 商业闭环进一步形成。微软 FY26Q4 Azure 及其他云服务收入同比增长 43%, Alphabet 2026Q2 Google Cloud 收入同比增长 82%, Amazon 2026Q2 AWS 收入同比增长 36.7%, 三大云厂商均显示 AI 基础设施及企业 AI 需求维持高景气; 2) 应用侧, 基础模型能力提升和使用成本下降后, 企业 AI 落地的关键逐步由“是否拥有模型”转向“能否结合行业数据和业务流程做好二次开发”。国内软件公司长期深耕工业、财税、ERP、营销、办公等垂直场景, 拥有客户入口、行业知识、复杂 workflows 和项目交付经验, 有望将通用模型进一步转化为可收费、可持续优化的行业 Agent 和工作流产品。北美 AI 应用公司显现出强劲增长趋势, 国内我们继续看好云、AI Infra 以及具备行业数据、复杂 workflows 和模型二次开发能力的软件公司。

相关标的

- 1) 大模型: 智谱、Minimax 等;
- 2) 部署&二次开发: 迈富时、范式智能、金蝶国际、汉得信息、用友网络、软通动力等;
- 3) 云: 阿里巴巴、金山云、优刻得、首都在线等;
- 4) AI Infra: 达梦数据、深信服、卓易信息、绿盟科技、合合信息、星环科技、广立微、海天瑞声等;
- 5) 高景气&高壁垒&信创: 德才股份、金山办公、中控技术、税友股份、美年健康、康众医疗、奔图科技、福昕软件、中国软件、顶点软件等。

风险提示

行业竞争加剧的风险; 技术研发进度不及预期的风险; 特定行业下游资本开支周期性波动的风险。



内容目录

一、北美 AI 应用公司显现增长驱动	3
1.1 Palantir: AIP 推动美国商业业务加速, 由试点走向规模化部署	3
1.2 Atlassian (TEAM): MCP 快速放量, AI 使用开始向席位和 ARR 增长传导	3
1.3 HubSpot: Agent 进入 GTM workflow, AI 价值开始接入席位与 Credits 变现	4
1.4 AppLovin: 收入和利润延续高增长, 算法持续优化广告投放效率	4
1.5 Duolingo: AI 扩展内容与交互供给, 持续改善学习体验	4
1.6 Cloudflare: 收入与 RPO 保持较快增长, 布局 Agentic Internet 基础设施	5
1.7 Unity: Vector 推动广告业务增长, AI 驱动开始体现到收入	5
1.8 Doximity: AI Search 使用量提升, AI 进一步进入医生 workflow	6
1.9 Twilio: AI 客户互动带动 Agent 产品组合使用	6
二、开源模式释放部署需求, 企业级 AI 加速落地	6
2.1 NVIDIA 持续倡导开放权重模型	6
2.2 Kimi K3 模型权重、技术报告正式发布	7
2.3 DeepSeek-V4-Flash 正式版 API 上线公测	8
2.4 OpenAI 下调 GPT-5.6 Terra 与 Luna 价格, 进一步提升模型性价比	9
三、云与应用商业闭环加速形成, 中国软件企业焕发新活力	10
3.1 云计算直接受益, 海外三大云厂商财报形成共振	10
3.2 基模上的二次开发尤为重要, 中国软件企业焕发新活力	10
四、相关标的	12
风险提示	12

图表目录

图表 1: Palantir 美国商业业务的客户、订单与合同价值同步加速	3
图表 2: HubSpot Q2 业绩概览	4
图表 3: Duolingo Q2 业绩概览	5
图表 4: Twilio Q2 业绩概览	6
图表 5: 黄仁勋在 X 首帖发布开源模型公开信	7
图表 6: Kimi K3 官方模型测评结果	8
图表 7: DeepSeek-V4-Flash 模型细节	9
图表 8: 主流基础模型任务成本与综合能力对比	10
图表 9: 星环科技新一代 AI 基础设施	11
图表 10: 在各项评测中, Seedance 2.0 的综合表现达到行业领先水平	12



一、北美 AI 应用公司显现增长驱动

近两周北美 AI 应用公司密集发布财报，多家公司收入、客户、订单等指标在 AI 驱动下保持增长。Palantir、AppLovin、Cloudflare、Atlassian、Unity、Twilio、HubSpot、Duolingo 和 Doximity 最新季度收入分别同比增长 93%、53%、36%、28%、24%、22%、20%、18%和 7%。除财务增长外，本轮财报中 AI 与经营指标之间的联系进一步增加：Palantir 表示 AIP 持续驱动美国商业业务动能，美国商业收入、客户及合同指标同步增长；Atlassian MCP 及 Teamwork Graph CLI 月活突破 100 万，MCP 采用客户的付费席位扩张更快、ARR 增速约为未采用客户的 2 倍；Unity 表示 Unity Ads Network 增长由 Unity Vector 驱动；Doximity 工作流活跃处方医生同比增长超过 30%，AI Search 查询量环比增长超过 25%；Twilio 披露的 AI 客户案例则显示，AI 客户互动已开始带动 Conversation Orchestrator、Memory 和 Agent Connect 等产品组合使用。Cloudflare、HubSpot 和 Duolingo 也分别围绕 Agentic Internet、GTM Agent 和 AI 会话练习持续拓展 AI 应用。整体来看，不同公司的 AI 商业化阶段仍有差异，但 AI 正在由新增功能进一步进入核心产品、客户工作流和经营环节。

1.1 Palantir：AIP 推动美国商业业务加速，由试点走向规模化部署

2026Q2 Palantir 总收入同比增长 93%至 19.4 亿美元，其中美国商业收入同比增长 149%、环比增长 28%至 7.64 亿美元，美国商业客户数同比增长 35%、环比增长 6%至 653 家。订单指标同步保持较快增长，美国商业剩余交易价值(RDV)同比增长 124%至 62.4 亿美元，单季美国商业 TCV 达到 21.3 亿美元、同比增长 153%；公司当季签约 220 笔金额不低于 100 万美元的合同，其中 98 笔不低于 500 万美元、73 笔不低于 1,000 万美元。公司同时将 2026 年美国商业收入指引上调至不低于 34.24 亿美元，同比增长至少 134%。

公司在 Q2 Business Update 中明确表示，AIP 持续驱动美国商业业务动能。从实际应用看，AIP 已经进入客户的募资、铀浓缩扩产、SAP 迁移、施工管理和客户服务等生产场景，将企业内部数据、业务AI模型接入具体业务流程。随着 AIP 在客户生产环境中的部署持续扩大，公司美国商业客户数量、合同规模及收入均保持较快增长。

图表1：Palantir 美国商业业务的客户、订单与合同价值同步加速

Q2 2026 Highlights

- US revenue grew +115% Y/Y and +23% Q/Q to \$1.57 billion
- US commercial revenue grew +149% Y/Y and +28% Q/Q to \$764 million
- US government revenue grew +90% Y/Y and +18% Q/Q to \$809 million
- Revenue grew +93% Y/Y and +19% Q/Q to \$1.94 billion
- Rule of 40 score of 155%
- Closed 220 deals of at least \$1 million, 98 deals of at least \$5 million, and 73 deals of at least \$10 million
- Adjusted free cash flow of \$1.22 billion; 63% margin
- Adjusted operating income of \$1.19 billion; 62% margin
- US commercial remaining deal value ("RDV") grew +124% Y/Y and +27% Q/Q to \$6.24 billion
- Highest ever quarter of US commercial total contract value ("TCV") of \$2.13 billion; +153% Y/Y
- Overall TCV grew +49% Y/Y to \$3.37 billion
- Adjusted EPS of \$0.41; GAAP EPS of \$0.41

来源：Palantir 《Q2 2026 Business Update》，国金证券研究所

1.2 Atlassian (TEAM)：MCP 快速放量，AI 使用开始向席位和 ARR 增长传导

Atlassian FY26Q4 收入同比增长 28%至 17.66 亿美元，其中云收入同比增长 31%至 12.13 亿美元；订阅 ARR 同比增长 23%至 66.06 亿美元，RPO 同比增长 44%。公司 AI 相关产品使用量快速提升，MCP Server 和 Teamwork Graph CLI 月活当季翻倍至超过 100 万，MCP 调用量增长超过 400%，通过 MCP 生成的 Jira 工单和 Confluence 页面环比接近 4 倍。公司同时披露，MCP 采用客户的产品黏性更高、付费席位扩张更快，ARR 增速约为未采用客户的 2 倍。

产品层面，Atlassian 通过 Teamwork Graph 连接 Jira、Confluence 等产品中的研发协作数据、企业知识、组织关系和权限体系，并通过 MCP Server 向 Claude、Cursor、ChatGPT 等外部 AI 工具开放这些业务上下文。在权限及审计机制下，外部 Agent 可以读取企业信息，并进一步创建 Jira 工单、Confluence 页面及执行相关任务。随着 MCP 使用量增长，公司 AI 能力已经由内置功能逐步扩展至外部 Agent 与 Atlassian 工作流之间的连接，并开始体现到客户席位和 ARR 等经营指标。



1.3 HubSpot: Agent 进入 GTM 工作流, AI 价值开始接入席位与 Credits 变现

HubSpot Q2 收入同比增长 20%至 9.12 亿美元, 客户数同比增长 14%至 30.64 万家, 单客户订阅收入同比增长 4%至 1.18 万美元, 客户数量和单客户收入均保持增长。AI 方面, 公司表示 Agent 已经开始为 GTM 团队带来可衡量的业务结果, 并持续围绕 AI 调整产品、定价和 GTM 策略。Investor Presentation 中, 公司将“提供清晰的 AI 价值”列为增长抓手, 并通过核心席位以及 Credits 对 AI 使用进行商业化。

HubSpot 当前主要将 AI 嵌入 CRM 既有的营销、销售和服务流程, 由 Agent 调用 Smart CRM 中的客户数据、历史互动及业务上下文完成相关任务, 并通过核心席位和 Credits 形成收费。相较于单独销售一个 AI 产品, 公司当前的商业化方式更多是将 AI 能力融入原有 CRM 及 GTM 产品体系, 提升现有产品使用价值并扩展按使用量收费的空间。

图表2: HubSpot Q2 业绩概览

Q2 2026 Financial Summary

Q2 2026 Actuals	Quarterly Results	Increase (Decrease) Y/Y ²	Currency Tailwind (Headwind)	Constant Currency Y/Y ³
Total Revenue	\$912m	19.8%	2.3 pts	17.5%
Subscription Revenue	\$894m	20.1%	-	-
Professional Services & Other Revenue	\$18m	8.5%	-	-
Gross Profit¹	\$766m	17.7%	-	-
Gross Profit Margin ¹	84.0%	(1.5) pts	-	-
Operating Profit¹	\$185m	43.5%	-	-
Operating Profit Margin ¹	20.3%	3.4 pts	0.5 pts	2.9 pts
Calculated Billings	\$930m	14.2%	(2.4) pts	16.6%
Free Cash Flow¹	\$168m	44.4%	-	-

来源: HubSpot《Q2'26 Investor Presentation》, 国金证券研究所

1.4 AppLovin: 收入和利润延续高增长, 算法持续优化广告投放效率

AppLovin Q2 收入同比增长 53%至 19.24 亿美元, 调整后 EBITDA 同比增长 58%至 16.14 亿美元, 公司指引 Q3 收入达到 20.55—20.85 亿美元、调整后 EBITDA 率达到 83%, 收入和利润继续保持较快增长。公司营销平台主要通过算法持续优化广告供需匹配和投放效果, 在提升广告主 ROI 的基础上吸引广告预算, 并扩大平台交易规模。

AI 及算法能力是 AppLovin 广告平台的重要产品基础, 公司持续利用相关技术提升用户匹配、广告投放及变现效率。不过, 本季度官方业绩材料并未单独量化 AXON 对收入增长的具体贡献, 因此现阶段更适合将 AI 和算法能力视为支撑广告平台效率提升的重要驱动因素, 而不对本季度收入增长进行单独的 AI 贡献拆分。

1.5 Duolingo: AI 扩展内容与交互供给, 持续改善学习体验

Duolingo Q2 DAU 同比增长 23%至 5,870 万, 付费订阅用户同比增长 17%至 1,270 万, 收入同比增长 18%至 2.99 亿美元, 用户规模和订阅业务继续增长。公司表示, 本季度 DAU 增长加速主要来自产品改善、营销以及一次性用户召回活动等多方面因素。

AI 方面, 公司持续扩大 AI 会话练习 Video Call 范围, 多数新的 Super 订阅用户已经可以使用该功能。生成式 AI 使 Duolingo 能够规模化提供个性化会话和互动式学习体验, 并提高教学内容供给效率。公司表示, AI 正在帮助其以更低的边际成本提供更高质量的学习体验, 因此目前 AI 主要通过丰富教学内容和交互方式参与产品迭代, 并进一步服务用户增长和订阅转化。


图表3: Duolingo Q2 业绩概览

Q2 Highlights

User Metrics	Q2 2025	Q2 2026
Daily Active Users	47.7M	58.7M 23% YoY
Paid Subscribers at period end	10.9M	12.7M 17% YoY

Financial Metrics	Q2 2025	Q2 2026
Revenue	\$252.3M	\$298.5M 18% YoY
Total Bookings	\$268.0M	\$289.1M 8% YoY
Net Income	\$44.8M 17.8% margin	\$33.2M 11.1% margin
Adjusted EBITDA	\$78.7M 31.2% margin	\$77.3M 25.9% margin
Net Cash Provided by Operating Activities	\$90.7M 35.9% margin	\$88.3M 29.6% margin
Free Cash Flow	\$86.3M 34.2% margin	\$78.6M 26.3% margin

来源: Duolingo 《Q2 2026 Shareholder Letter》, 国金证券研究所

1.6 Cloudflare: 收入与 RPO 保持较快增长, 布局 Agentic Internet 基础设施

Cloudflare Q2 收入同比增长 36%至 6.961 亿美元, Current RPO 同比增长 35%; 公司预计 Q3 收入为 7.36—7.37 亿美元, 并将全年收入指引上调至 28.64—28.70 亿美元。管理层表示, 随着互联网向 AI 答案引擎和 Agent 驱动的商业模式转变, 机器间流量正在推动互联网流量结构发生变化, 公司将自身定位为 Agentic Internet 所需基础设施、控制能力、开发者工具及支付轨道的提供者。

AI 战略方面, Cloudflare 重点围绕 Agent 产生和访问互联网流量所需的网络、控制及开发者能力进行布局。管理层认为, AI 答案引擎和 Agent-driven commerce 的发展将持续增加 machine-to-machine traffic, 公司希望依托现有全球网络和开发平台承接相关需求。

1.7 Unity: Vector 推动广告业务增长, AI 驱动开始体现到收入

Unity Q2 收入同比增长 24%至 5.465 亿美元, 战略收入同比增长 38%至 4.864 亿美元, 其中 Grow Solutions 收入同比增长 35%至 3.89 亿美元。公司表示, Grow Solutions 增长主要来自 Unity Ads Network, 而 Unity Ads Network 的增长由 Unity Vector 驱动。公司预计 Q3 战略收入为 5.40—5.50 亿美元, 同比增长 44%—47%。管理层同时表示, Unity Vector AI 的持续成功正在为创作者、玩家和股东创造价值。

Unity Vector 主要应用于广告业务, 公司持续推进 Vector 在广告匹配和投放环节的优化。本季度管理层已经将 Unity Ads Network 增长明确归因于 Vector, 因此 Vector 的产品进展开始直接体现到 Grow Solutions 及整体战略业务的增



长。

1.8 Doximity: AI Search 使用量提升, AI 进一步进入医生 workflow

Doximity FY27Q1 收入同比增长 7%至 1.566 亿美元, 工作流活跃处方医生同比增长超过 30%, AI Search 查询量环比增长超过 25%。公司预计 FY27Q2 收入为 1.70—1.71 亿美元, 并将全年收入指引更新至 6.71—6.81 亿美元。AI 方面, 公司表示临床 AI 助手 Doximity Ask 在 NOHARM 基准测试中成为表现最好的美国模型, 本季度同时实现 record engagement。

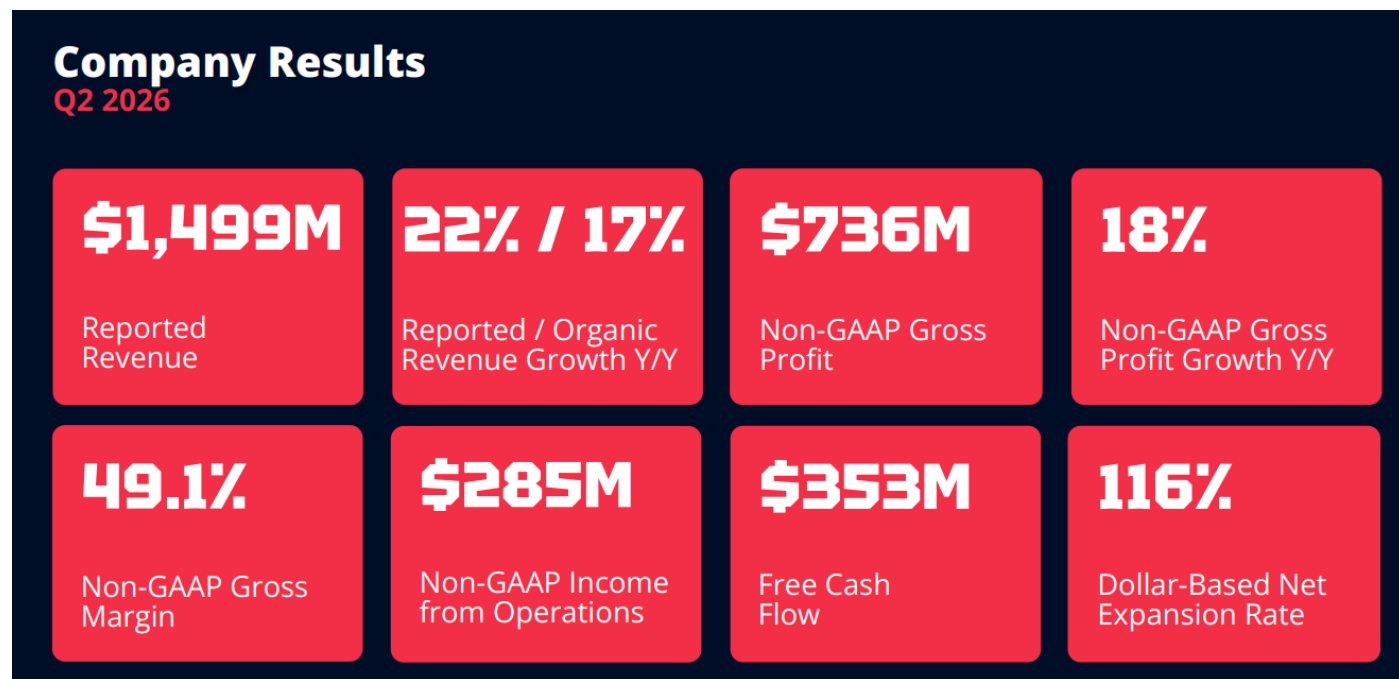
产品层面, 公司持续围绕 Doximity Ask 和 AI Search 将 AI 能力嵌入医疗专业人士的日常使用场景。工作流活跃处方医生和 AI Search 查询量的增长显示, AI 相关功能的使用仍在扩大, 公司继续通过临床问答、搜索等场景深化医生 workflow 中的 AI 应用。

1.9 Twilio: AI 客户互动带动 Agent 产品组合使用

Twilio Q2 收入同比增长 22%至 14.99 亿美元, 有机收入同比增长 17%, Dollar-Based Net Expansion Rate 为 116%, 自由现金流达到 3.53 亿美元。公司预计 FY27Q2 收入为 15.05—15.15 亿美元, 并将全年收入增长指引上调至 18%—18.5%。AI 应用方面, 公司披露 Finance 247 的 AI 助手 Carla 已处理约 28.8 万次客户对话; Vozzi 则组合使用 Conversation Orchestrator、Memory 和 Twilio Agent Connect 构建 Agentic 客户互动。

从客户案例看, Twilio 的 AI 应用已经从单一对话能力延伸至包含对话编排、上下文记忆和 Agent 连接的产品组合。公司正在利用原有通信能力, 为 AI Agent 提供客户交互及上下文相关能力, 并持续扩展 AI 原生客户的应用场景。

图表4: Twilio Q2 业绩概览



来源: Twilio Q2 2026 Earnings Results, 国金证券研究所

二、开源模式释放部署需求, 企业级 AI 加速落地

我们在前两篇报告《开源模型跃迁会带来什么影响》《开源模型: 云与应用闭环加速形成》中提出, 开放权重模型产业意义不仅在于模型能力提升, 更在于降低企业获取、开发和部署高质量模型的门槛, 推动企业 AI 由调用模型逐步走向开发、部署和运营模型。本周产业变化进一步强化了上述趋势。

Kimi K3 发布完整模型权重, 增加了高能力开放模型供给; DeepSeek V4-Flash 开放 API 公测, 进一步完善模型服务和 Agent 工具适配, 且在定价方面再次体现出极致性价比; NVIDIA 持续倡导开放权重模型, 并围绕模型、数据、训练、部署和开发组件建设完整生态; OpenAI 下调 Terra 和 Luna 价格, 降低企业调用闭源模型的成本。我们认为, 开放权重模型持续完善、模型价格性价比不断提升以及开发生态持续成熟, 共同推动高质量基础模型由“可用”进一步走向“易用”, 释放企业级 AI 部署需求, 加快商业化落地。

2.1 NVIDIA 持续倡导开放权重模型

黄仁勋在 X 账号首帖中分享了英伟达参与联署的开源模型公开信, 并表示, AI 将改造每个行业、赋能每家公司、由每个国家建设; 开源模型能够强化安全与网络安全、加速创新与扩散并支持主权化, 世界同时需要前沿闭源模型与前沿



开源模型。NVIDIA 在 7 月 14 日发布的 Nemotron Labs 文章中指出，开放模型允许企业围绕自有领域知识与工作流进行定制、检查和改进，并可与前沿闭源模型共同组成多模型系统；NeMo 套件覆盖模型定制、评测、Agent 优化和治理，开放模型、权重、数据集与工具共同构成开发基础。

图表5：黄仁勋在 X 首帖发布开源模型公开信



Jensen Huang  
@JensenHuang

For my first post, I'm sharing a letter @NVIDIA signed on why open models matter.

AI will transform every industry, power every company, and be built by every country.

Open models strengthen safety and cybersecurity, accelerate innovation and diffusion, and enable sovereignty.

The world needs both frontier closed models and frontier open models.

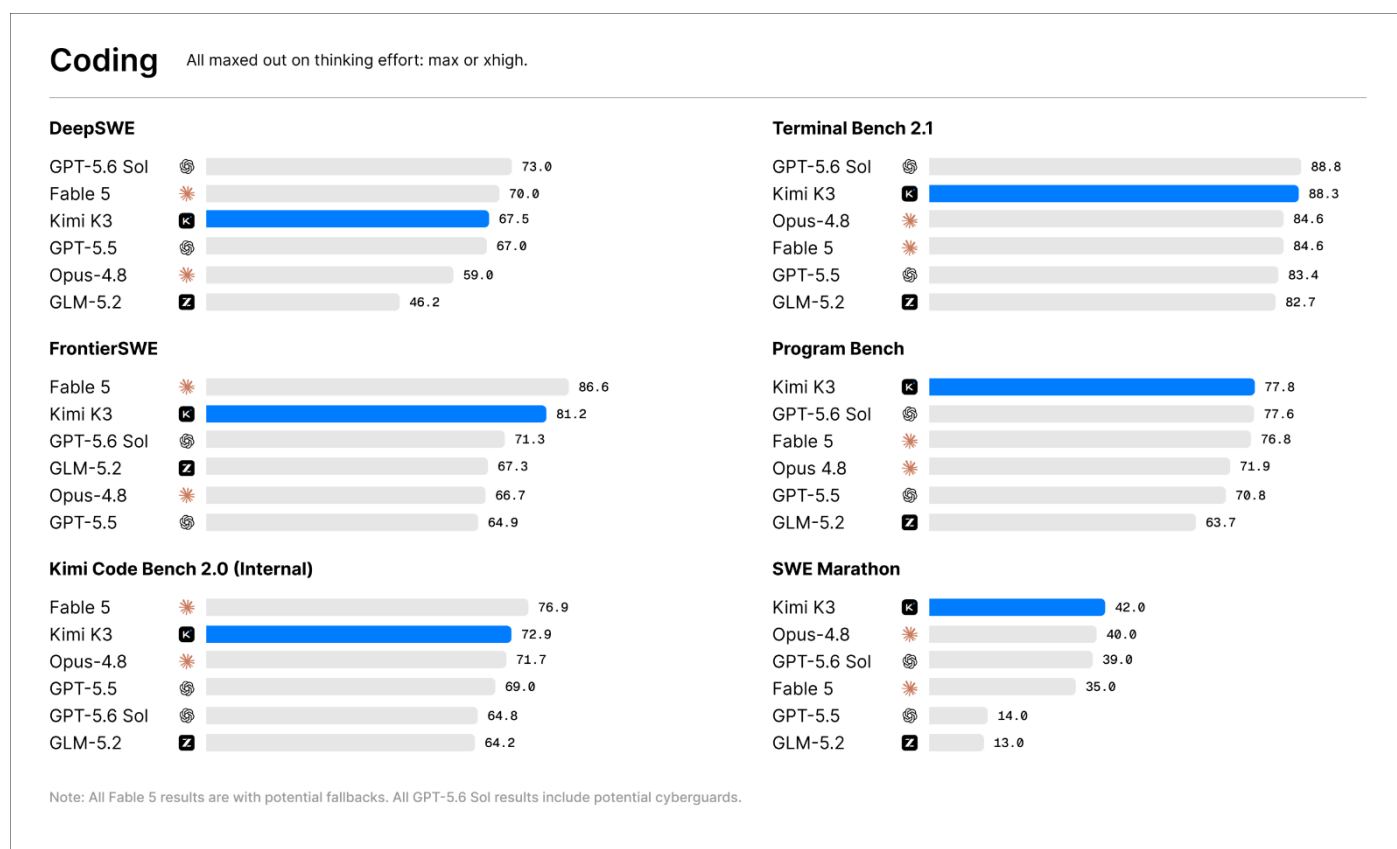
来源：环球时报，国金证券研究所

2.2 Kimi K3 模型权重、技术报告正式发布

月之暗面正式开放 Kimi K3 模型权重。在 7 月 27 日 Kimi K3 开放日，月之暗面发布 Kimi K3 的模型权重、技术报告，并开源支撑 Kimi K3 模型训练的关键 Infra 技术：MoonEP、FlashKDA 和 AgentEnv。Kimi 团队于 7 月发布 Kimi K3，技术报告显示其为 2.8T 总参数、104B 激活参数的 MoE 模型，支持原生视觉与 100 万 Token 上下文，并公开完整模型权重。其官方评测中，Kimi K3 在 Terminal-Bench 2.1 取得 88.3 分，在 SWE-Marathon 取得 42.0 分，显示国产开放模型在长链路编码和 Agent 任务上的可用性持续提升。



图表6: Kimi K3 官方模型测评结果



来源: Kimi K3 技术报告, 国金证券研究所

2.3 DeepSeek-V4-Flash 正式版 API 上线公测

DeepSeek V4 能力与性价比同步升级。DeepSeek V4 于 4 月 24 日首次发布, 4 月 31 日开放 V4-Flash-0731 API 公测。官方更新显示, 本次版本保持与 Preview 相同架构和规模, 主要进行后训练升级; 在 Terminal Bench 2.1、DeepSWE、Toolathon verified 等任务上分别取得 82.7、54.4、70.3 分, 并原生支持 Responses API、针对 Codex 完成适配。同时, V4-Flash 延续极致性价比, 百万 Token 输入价格为缓存命中 0.02 元、缓存未命中 1 元, 输出价格仅 2 元。



图表7: DeepSeek-V4-Flash 模型细节

模型		deepseek-v4-flash	deepseek-v4-pro
BASE URL (OpenAI 格式)		https://api.deepseek.com	
BASE URL (Anthropic 格式)		https://api.deepseek.com/anthropic	
模型版本		DeepSeek-V4-Flash-0731	DeepSeek-V4-Pro
思考模式		支持非思考与思考模式 (默认) 切换方式详见 思考模式	
上下文长度		1M	
输出长度		最大 384K	
功能	Json Output	支持	支持
	Tool Calls	支持	支持
	Responses API⁽¹⁾	支持	暂不支持
	Anthropic API	支持	支持
	对话前缀续写 (Beta)	支持	支持
	FIM 补全 (Beta)	仅非思考模式支持	仅非思考模式支持
价格 ⁽²⁾	百万tokens输入 (缓存命中)	0.02元	0.025元
	百万tokens输入 (缓存未命中)	1元	3元
	百万tokens输出	2元	6元
并发限制 ⁽³⁾		2500	500

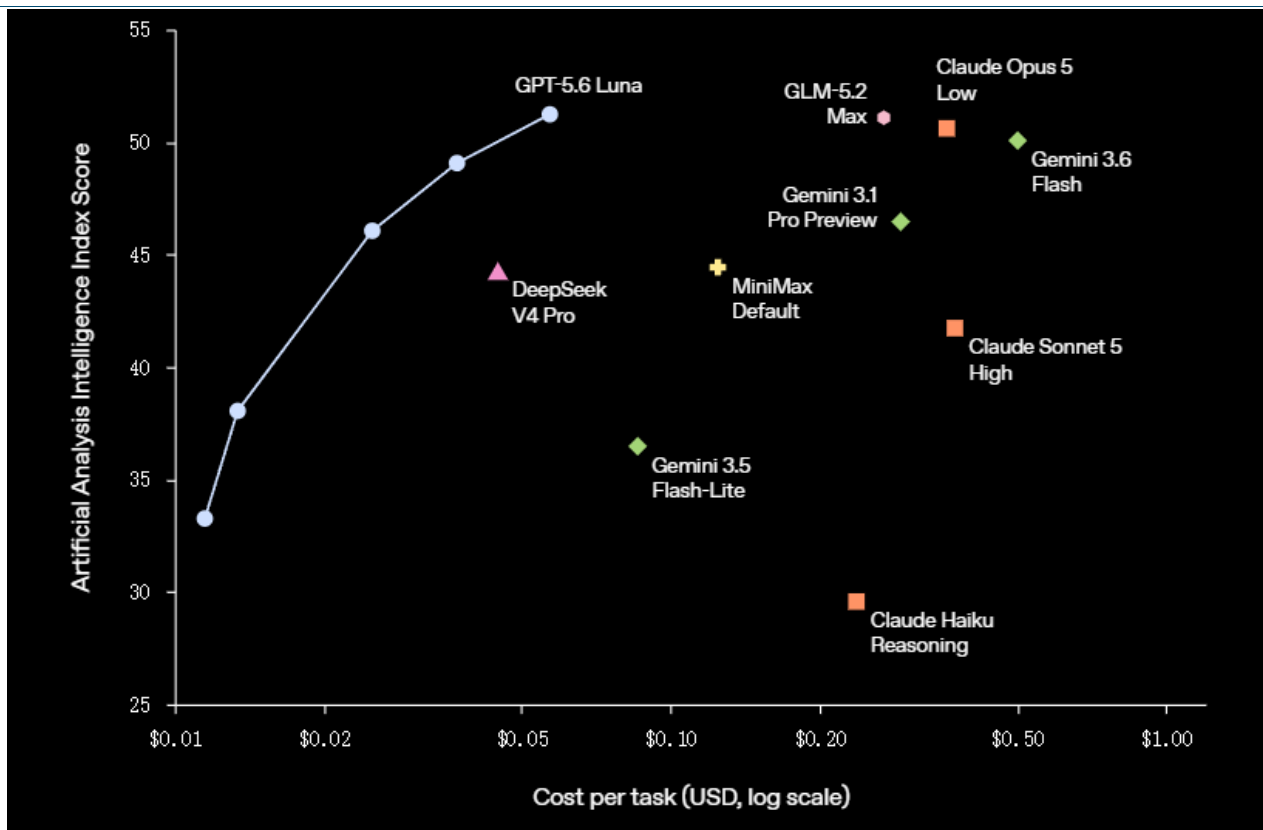
来源: DeepSeek API 文档, 国金证券研究所

2.4 OpenAI 下调 GPT-5.6 Terra 与 Luna 价格, 进一步提升模型性价比

7月30日, OpenAI 下调 GPT-5.6 Terra 与 Luna 的 API 价格: Terra 每百万 Token 输入、输出价格分别由 2.5 美元、15 美元降至 2 美元、12 美元, 降幅均为 20%; Luna 每百万 Token 输入、输出价格分别由 1 美元、6 美元降至 0.2 美元、1.2 美元, 降幅均为 80%。其中, Terra 定位于日常企业任务的能力与成本平衡, Luna 则面向高频、规模化任务。OpenAI 表示, Luna 能够调用工具并完成多步骤 workflow, 降价后可使更多 Agent 和批量任务具备规模化部署的经济性; Terra 和 Luna 在 ChatGPT Work 及 Codex 中的使用额度消耗也同步下降。



图表8：主流基础模型任务成本与综合能力对比



来源：OpenAI 官网，国金证券研究所

三、云与应用商业闭环加速形成，中国软件企业焕发新活力

3.1 云计算直接受益，海外三大云厂商财报形成共振

我们认为，企业级 AI 部署需求释放后，云计算是最直接的受益方向。近期微软、Google 和 AWS 最新财报继续验证，AI 需求正在由基础设施逐步传导至平台服务、企业软件和持续付费，云厂商 AI 商业闭环进一步形成。

微软 FY26Q4 Microsoft Cloud 收入 593 亿美元，同比增长 27%；Azure 及其他云服务收入同比增长 43%，FY26 Azure 全年收入超过 1000 亿美元。Microsoft 365 Copilot 付费席位超过 3000 万，商业剩余履约义务达到 6780 亿美元、同比增长 84%。公司指引 FY27Q1 Azure 固定汇率增速约 45%，并表示需求仍高于可用供给。

Alphabet 2026Q2 Google Cloud 收入 248 亿美元，同比增长 82%；云业务营业利润 88 亿美元，营业利润率 35.6%；云业务积压订单 5140 亿美元，环比增加约 500 亿美元。公司称增长由 AI 基础设施和企业 AI 解决方案共同推动，GCP 增速快于 Cloud 整体。Alphabet 2026Q2 资本开支 449 亿美元，约 60%用于服务器、40%用于数据中心和网络设备；公司将 2026 年全年资本开支指引提高至 1950 亿—2050 亿美元，并表示需求仍快于投资扩张。

Amazon 2026Q2 AWS 收入 422 亿美元，同比增长 36.7%，为 18 个季度以来最快增速；AWS 营业利润 166 亿美元。公司披露 AI 业务年化收入运行率超过 250 亿美元，自研芯片业务年化运行率亦超过 250 亿美元，二者均保持三位数增长。截至 2026Q2 的过去 12 个月自由现金流为-76 亿美元，主要受物业及设备投入同比增加 661 亿美元影响，增量主要投向 AI。

3.2 基模上的二次开发尤为重要，中国软件企业焕发新活力

基础模型能力提升和使用成本下降后，企业 AI 落地的关键逐步由“是否拥有模型”转向“能否结合行业数据和业务流程做好二次开发”。国内软件公司长期深耕工业、财税、ERP、营销、办公等垂直场景，拥有客户入口、行业知识、复杂工作流和项目交付经验，有望将通用模型进一步转化为可收费、可持续优化的行业 Agent 和工作流产品。

AI Infra：美股持续验证的产业趋势

复杂 Agent 推动数据库、数据治理、模型管理、可观测性、安全及混合云编排需求持续提升，美股数据库、安全及云软件公司业绩持续验证 AI Infra 产业趋势。

AI 基础设施软件可以协助企业将原始数据转化为高质量数据资产，并降低企业 AI 能力的开发难度。AI 基础设施软件整合了数据清洗、生成、标注、存储等功能，实现了对企业内部异构数据源的统一访问与集中管理。通过提升整体数据的可用性，显著减少数据孤岛与冗余，为构建结构合理、语义一致的数据资产体系奠定基础。此外，AI 基础设施软



件可以提供模型设计、测试与迭代的框架，降低构建 AI 应用的复杂性，助力企业更高效地开发和部署人工智能。例如，AI 模型开发软件可自动执行重复任务，优化数据识别、分类及管理流程，从而高效处理大规模数据集，加快 AI 模型训练及部署，减少企业从头搭建模型开发和管理体系的成本。

数据底座构成 AI 基础设施软件中的关键部分。在大数据场景下，以集中式架构关系型数据库为代表的传统数据管理软件不能很好适应数据的“4V”特性，面临较多技术挑战。随着信息技术的发展，尤其是生成式 AI 的出现，数据处理需求的变化推动了数据管理软件技术的变革，从技术发展方向来看，数据管理软件技术依次按照“关系型、集中式”向“非关系型、分布式”并进一步向“多模型、云原生”这三个阶段进行演变。随着技术不断成熟，分布式架构将逐渐成为主流。自底向上，传统的集中式资源管理调度逐渐向基于云原生技术的分布式统一资源管理平台发展；数据管理软件技术架构也会因为计算模式的转变发生重大变革，传统的集中式数据库逐渐向分布式、多模型数据库发展；传统数据分析软件逐渐向新型的分布式数据开发和智能分析软件发展，智能分析软件需要提供机器学习和大模型的统一运营视角。

我们认为，随着 AI 系统复杂度提升，数据库/可观测性/安全/恢复/混合云/工作负载编排等相关的 AI Infra 工具越发成为刚需，因此复杂 Agent 的应用，将进一步打开 AI Infra 板块的 TAM 扩张边界。

图表9：星环科技新一代 AI 基础设施



来源：星环科技 25 年年报，国金证券研究所

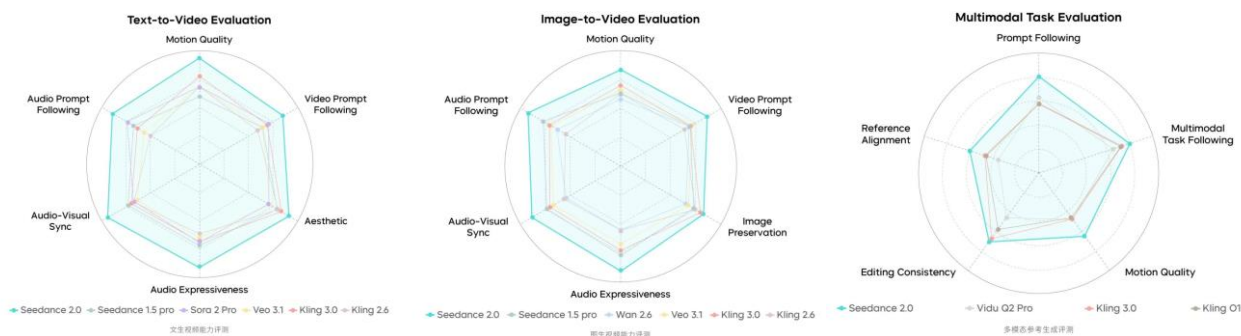
高景气高壁垒：基模上的二次开发价值提升

开放权重模型进一步放大了 workflow 软件的价值量。开放权重模型使软件公司能够结合行业数据、业务流程和客户需求，对通用模型进行二次开发和专属部署，并将 AI 真正嵌入营销、销售、客服、研发等核心业务流程。以近期迈富时业绩为例，根据公司 2026 年半年度盈喜，公司预计 2026 年上半年 AI 应用业务收入达到 10.68 亿—11.80 亿元，同比增长 112%—134%，约占公司总收入的 57.5%；公司表示，增长主要来自场景 Token 计费模式在 GEO、外贸及经营分析等 Agent 产品中加速渗透，以及 AI Agentforce 智能体中台 3.0、KnowForce AI 知识中台和 GenAI OS 等产品完善，带动客户数和客单价同步提升。

高景气方面，2026 年 2 月 10 日，字节跳动旗下即梦 AI 正式发布 Seedance 2.0；该模型在 Artificial Analysis Video Arena 以 Elo 1,269 的成绩登顶，超越 Google Veo 3、OpenAI Sora 2 和 Runway Gen-4.5，成为当前评测体系下综合表现最强的视频生成模型。生成质量的稳定性（抽卡成功率）对 ai 视频成本非常关键，让模型生成一段 15 秒的视频，过去 AI 视频抽卡成功率不到 20%，根据极客公园的测试和多位从业者在 X 上的反馈，Seedance2.0 的可用率高达 90% 以上。模型能力的跃迁直接降低废片率，是漫剧 agent 商业化的关键拐点。我们认为，从抽卡成功率提升的角度看，漫剧 agent 的真实生产成本大幅压缩，这意味着“批量化 AI 内容生产”在 2026 年将首次具备规模化的单位经济模型。



图表10：在各项评测中，Seedance 2.0 的综合表现达到行业领先水平



来源：字节跳动官方公众号，国金证券研究所

信创：业务增速回升预期

信创：临近重要替换节点，政策引导+AI 技术革新引领行业迈入加速期

政策端：一方面，信创临近重要替换节点，党政及行业均加速替换，且进一步迈入替换深水区；另一方面，供应链安全相关的政策引导进一步完善，4月7日，国务院总理李强签署国务院令，公布《国务院关于产业链供应链安全的规定》，这是我国首部针对产业链供应链安全的专项行政法规，为党政、金融、电信、能源、交通、航空航天等关键行业加速推进信创替代提供了坚实依据。在政策驱动下，金融、电信、交通、能源行业信创建设加速，信创进程在深度、广度上均显著提升。

技术端：AI 与信创的融合已从产业趋势升级为发展刚需。4月22日，2026中国数据库技术与产业大会在北京举办，行业各界达成共识：国内数据库产业已完成规模化替代的阶段性目标，正式进入“AI 驱动+市场筛选”的高质量发展阶段，AI 技术正从内核层面重塑数据库架构，向量数据库、AI 原生数据库逐步成为产业主流，打破了传统数据库“功能叠加”的发展模式。达梦数据、人大金仓、OceanBase、PolarDB 等国内头部数据库企业，在大会上集中展示了 AI 融合成果，向量检索、多模数据统一管控、智能自治运维等功能已成为产品核心配置。

四、相关标的

- 1) 大模型：智谱、Minimax 等；
- 2) 部署&二次开发：迈富时、范式智能、金蝶国际、汉得信息、用友网络、软通动力等；
- 3) 云：阿里巴巴、金山云、优刻得、首都在线等；
- 4) AI Infra：达梦数据、深信服、卓易信息、绿盟科技、合合信息、星环科技、广立微、海天瑞声等；
- 5) 高景气&高壁垒&信创：德才股份、金山办公、中控技术、税友股份、美年健康、康众医疗、奔图科技、福昕软件、中国软件、顶点软件等。

风险提示

■ 行业竞争加剧的风险：

在信创等政策持续加码支持计算机行业发展的背景下，众多新兴玩家参与到市场竞争之中，若市场竞争进一步加剧，竞争优势偏弱的企业或面临出清，某些中低端品类的毛利率或受到一定程度影响。

■ 技术研发进度不及预期的风险：

计算机行业技术开发需投入大量资源，如果相关厂商新品研发进程不及预期，表面层面将呈现出投入产出在较长时期的滞后特征。

■ 特定行业下游资本开支周期性波动的风险：

部分计算机公司系顺周期行业，下游资本开支波动与行业周期性相关性较强，或在个别年份对于上游软件厂商的营收表现产生扰动。



行业投资评级的说明：

买入：预期未来 3—6 个月内该行业上涨幅度超过大盘在 15%以上；

增持：预期未来 3—6 个月内该行业上涨幅度超过大盘在 5%—15%；

中性：预期未来 3—6 个月内该行业变动幅度相对大盘在 -5%—5%；

减持：预期未来 3—6 个月内该行业下跌幅度超过大盘在 5%以上。



特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员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。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级（含 C3 级）的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海	北京	深圳
电话：021-80234211	电话：010-85950438	电话：0755-86695353
邮箱：researchsh@gjzq.com.cn	邮箱：researchbj@gjzq.com.cn	邮箱：researchsz@gjzq.com.cn
邮编：201204	邮编：100005	邮编：518000
地址：上海浦东新区芳甸路 1088 号 紫竹国际大厦 5 楼	地址：北京市东城区建国内大街 26 号 新闻大厦 8 层南侧	地址：深圳市福田区金田路 2028 号皇岗商务中心 18 楼 1806



【小程序】
国金证券研究服务



【公众号】
国金证券研究