下载日志已记录，仅供内部参考

证券研究报告|行业专题报告
互联网
行业评级 强大于市（维持评级）
2026年7月8日

华福证券 HUAFU SECURITIES CO.,LTD.

AGI大时代：从“降本”到“扩大生产”

证券分析师：
杨晓峰 执业证书编号：S0210524020001

下载日志已记录，仅供内部参考

投资要点

- **1、Fable 5推动AGI能力从概念验证进入工程化落地。**Fable 5在对话、推理、代码智能体、知识工作、视觉理解和金融分析等维度均表现突出，尤其在代码智能体任务上较上一代模型和海外主流模型拉开差距。
- **2、AI云厂商加速争夺底层算力入口。**AWS、谷歌云、Meta等厂商均围绕大模型训练与推理需求加码基础设施投入，云厂竞争重心从传统IaaS资源供给延伸至模型绑定、自研芯片、算力租赁和AI服务变现，AI负载成为云增长的重要驱动。
- **3、国产模型依靠性价比和应用能力提升全球存在感。**OpenRouter调用榜和Artificial Analysis测评显示，国产模型在能力差距收窄的同时，凭借更低价格、更快响应和更强多模态产品能力进入全球开发者生态，AI视频等方向已率先形成全球竞争力。
- **投资建议：**建议持续关注全球 AI 产业落地进程。随着模型能力持续迭代以及国产模型渗透持续提升，我们看好全球 AI 产业趋势延续及国内 AI 生态加速崛起。
- **风险提示：**AI竞争激烈，AI应用竞争激烈，AI落地场景变化。

下载日志已记录，仅供内部参考

目录

- 一、AGI再定义：Fable 5的能力跃迁
- 二、AI云发力：大模型厂商加码云端入口
- 三、国产追赶：模型能力与应用生态积极突破
- 投资建议
- 风险提示

下载日志已记录，仅供内部参考

1、Fable5：AGI雏形已逐步形成

- AGI 从概念走向可度量的工程现实，**Fable 5 在通用性维度已显现出人工通用智能的雏形。** AGI（人工通用智能）指在广泛认知任务上比肩乃至超越人类的智能体，区别于仅擅长单一场景的专用模型；OpenAI 以五级路线图（对话 → 推理 → 智能体 → 创新者 → 组织）将目标拆解为可跟踪的能力台阶。Fable 5 系 Anthropic 最新一代前沿模型，在推理、编码、科学与多模态等维度刷新前沿。

图：OpenAI评估框架及Fable 5表现

层级	能力要求	Fable 5 前沿模型现状	距达标差距
L1 对话 Chatbots	自然语言对话	Text Arena 测评中表现第一，共计1509分	无
L2 推理 Reasoners	博士级推理	GPQA ~92%+（评分自同一底层模型的Mythos 5）	极小
L3 智能体 Agents	自主多步骤任务	Stripe 50M行代码库1天完成	中等
L4 创新者 Innovators	推进科学前沿	能力释放存在安全边界： 前沿生物、化学、网络安全能力受控开放	较大
L5 组织 Organizations	替代整个组织	已能压缩特定团队级任务	巨大

资料来源：axios官网，彭博，arena官网， rdworldonline官网，Anthropic官网，华福证券研究所

下载日志已记录，仅供内部参考

2、Fable5：能力代差与前沿学科评测实力全景

- Fable 5 在多项高价值任务上已与主流前沿模型拉开差距，代码智能体优势最为突出，SWE-Bench Pro和FrontierCode Diamond较 Opus 4.8 和 GPT 5.5 提升更明显。
- 本轮评测补充了视觉能力维度，弥补了过去 GPT 系列对比中视觉任务覆盖不足的问题。 Fable 5 在知识工作视觉和空间推理任务中表现靠前。
- 在前沿学科中，Fable 5 在知识工作、分析、工具调用和跨学科推理中也保持领先，能力更贴近企业真实工作流。

1 前沿物理研究

Physical Superintelligence实测：36小时完成GPT-5.5需4天才能达到的前沿物理推理层高

2 生命科学/蛋白质设计

与Fable 5同一底层模型的Mythos 5蛋白质设计：14靶标中9个产出强候选药物较人类科学家提速约10x，全程自主无人工干预

3 基础科学综合推理

Humanity’s Last Exam 59%，全球唯一超50%模型，覆盖数学、物理、化学、生命科学、人文多维科学推理

图：Anthropic模型能力评测表

	Claude Mythos 5 / Fable 5	Claude Mythos Preview	Claude Opus 4.8	GPT 5.5	Gemini 3.1 Pro
Agentic coding SWE-Bench Pro	**80.3%**	77.8%	69.2%	58.6%	54.2%
Agentic coding FrontierCode (Diamond)	**29.3%** xhigh	—	13.4% xhigh	5.7% xhigh	—
Knowledge work GDPval-AA	**1932**	—	1890	1769	1314
Knowledge work vision GDPpdf	**29.8%** no tools	—	22.5% no tools	24.9% no tools	16.7% no tools
Spatial reasoning Blueprint-Bench 2	**38.6%**	—	14.5%	36.2%	26.5%
Tool use AutomationBench	**17.4%**	—	15.5%	12.9%	9.6%
Computer use OSWorld-Verified	85.0%	**85.4%**	83.4%	78.7%	76.2%
Legal Legal Agent Benchmark	**13.3%**	—	10.4%	2.1%	0.0%
Multidisciplinary reasoning Humanity's Last Exam	**59.0%*** no tools	56.8% no tools	49.8% no tools	41.4% no tools	44.4% no tools
	64.5%* with tools	**64.7%** with tools	57.9% with tools	52.2% with tools	51.4% with tools
Biology BioMysteryBench	**46.1%*** hard	29.6% hard	40.0% hard	—	—
	83.9%* human solved	82.6% human solved	80.4% human solved	—	—
Agentic coding Terminal-Bench 2.1	**88.0%***	—	82.7%	83.4% Codex CLI	70.7% Gemini CLI
Cybersecurity ExploitBench (Cap%)	**78.0%***	69.0%	40.0%	34.0%	—
Health HealthBench Professional	**66.0%***	64.7%	56.9%	51.8%	—

Methodology: Reported scores are within a 1-3 percentage point difference for Claude Mythos 5 and Claude Fable 5. This table shows the higher score of the two. Starred (*) benchmarks show a larger difference due to our blocking safeguards for cybersecurity and biology-related questions. For these benchmarks, Claude Fable 5 performs closer to Claude Opus 4.8 due to fallbacks. See the system card for details.

资料来源：Anthropic官网，DBR官网，Arxiv官网，华福证券研究所

下载日志已记录，仅供内部参考

目录

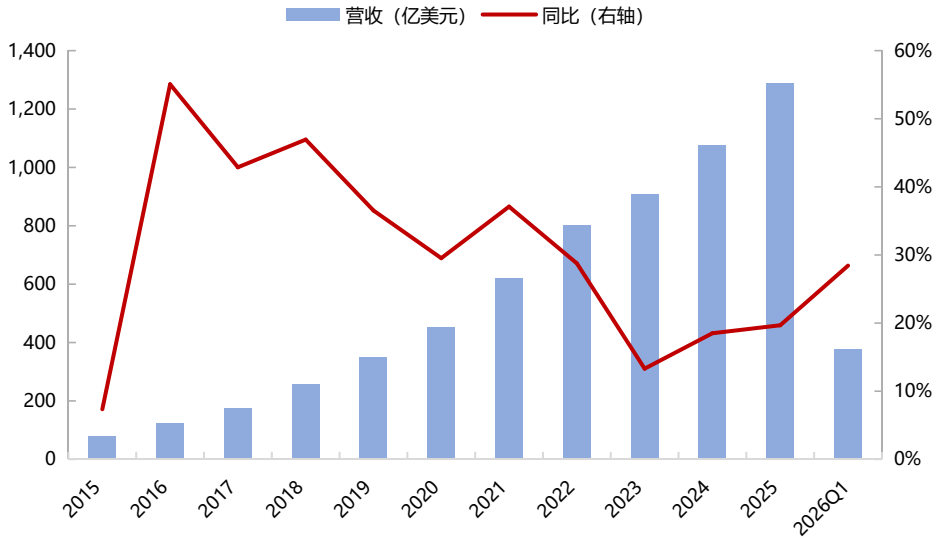
- 一、AGI再定义：Fable 5的能力跃迁
- 二、AI云发力：大模型厂商加码云端入口
- 三、国产追赶：模型能力与应用生态积极突破
- 投资建议
- 风险提示

下载日志已记录，仅供内部参考

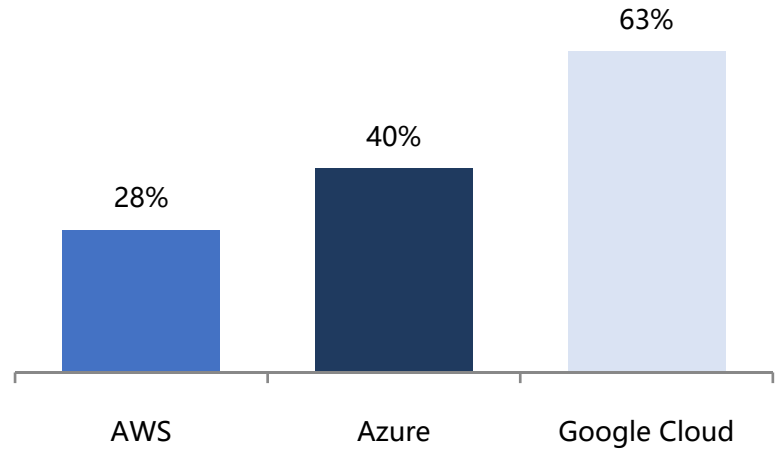
1、AWS：被 Anthropic 拉动，提速增长

- AWS 一季度增速重回加速通道，AI 负载放量叠加大模型客户绑定，是这轮再加速的核心驱动。2026Q1 收入 375.9 亿美元、同比+28%，**创 15 个季度最快，**年化收入约 1500 亿美元；增量主要来自对 Anthropic 的加码，并为深化与 OpenAI 的合作预留接口。**结构层面：**AWS 的 AI 相关收入年化已超 150 亿美元，自研 Trainium 芯片业务年化突破 200 亿美元。

图：AWS营收及增速



图：海外云2026Q1营收同比增速



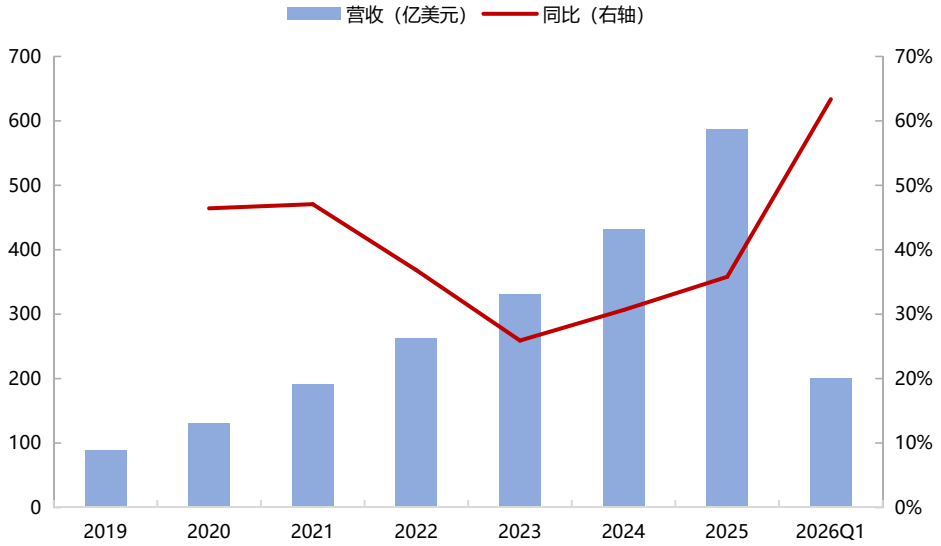
资料来源：wind，cnbc官网，华福证券研究所

下载日志已记录，仅供内部参考

2、Google Cloud：增速领跑海外云

➢ **Google Cloud以 63% 的增速领跑三大云，合同储备近乎翻倍，收入确定性显著抬升。** 2026Q1 收入 200.3 亿美元、同比 +63%，单季首破 200 亿，增速全面超越 AWS（28%）与 Azure（40%）；backlog 环比接近翻倍至 4600 亿美元以上，锁定未来数个季度的收入兑现节奏。企业级 AI 首次成为云增长第一引擎，基于自研生成式模型的产品收入同比增长近 800%。

图：谷歌云服务营收及增速



图：对应算力部署

$460B+

合同储备 backlog · 环比近翻倍

~800%

自研生成式模型产品收入同比

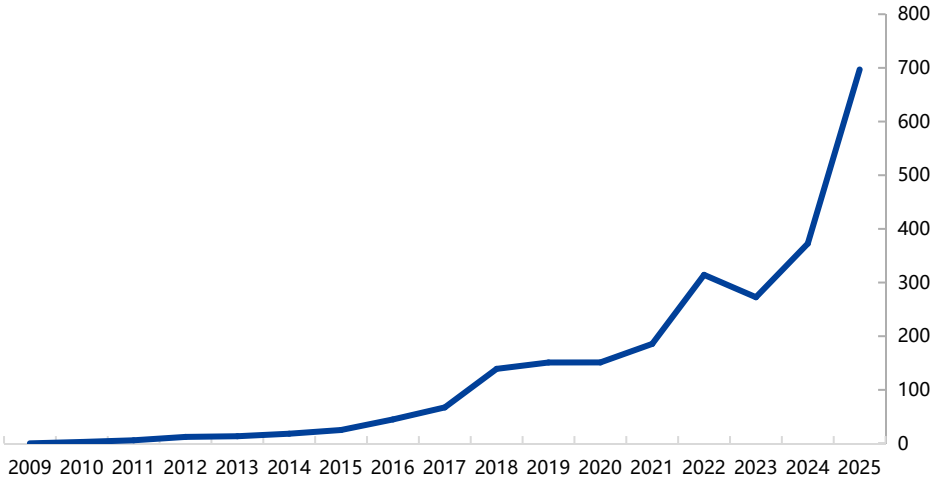
资料来源：wind，Google官网，华福证券研究所

下载日志已记录，仅供内部参考

3、Meta云服务：从自用算力走向算力变现

- 近期Meta云基础设施业务进入规划阶段，拟将过剩 AI 算力与模型能力对外出售，AI 基建投入开始从成本项转向潜在收入项。2026 年 capex 指引已上调至 1250–1450 亿美元，核心投向包括高价组件采购与数据中心扩容；若后续云业务落地，**有望在广告主业之外形成AI 算力租赁的新增变现路径。**目前该计划仍处于早期阶段，战略方向存在调整可能，短期更适合作为 AI capex 回收路径的观察变量。

图：Meta CapEx （亿美元）



图：Meta近期算力部署

$107B

单季新增多年期基建承诺

1GW+

与博通合作自研芯片 · 主打算力效率

资料来源：wind，Meta26Q1官方电话会，华福证券研究所

下载日志已记录，仅供内部参考

4、马斯克算力部署

- **马斯克算力模式是先堆单体规模、再对外租赁变现，走出不同于三度云的差异化路径。**
- **规模端：**Colossus 1 已部署超 22 万块 GPU、功率 300MW+，Colossus 2 向 2GW、55.5 万块 GPU、约 180 亿美元投资推进。
- **变现端：**2026 年 5 月起 Anthropic 整租 Colossus 1，月付约 12.5 亿美元，合同排至 2029 年、累计对价有望超 400 亿美元；6 月起 Colossus 2 部分算力已被租用。大模型推理支出最终回流至底层算力基建，印证"能力—算力—现金流"的传导链条。

图：马斯克算力部署情况

指标	Colossus 1	Colossus 2	远期目标
GPU 规模	22 万+（H100/H200/GB200）	55.5 万（Blackwell 为主）	100 万+
功率	300MW+	5月实现1GW （引据Epoch AI预测）	2 GW （引据马斯克自述）
定位 / 进展	已整租，转推理负载	约 $180 亿，训练主力	自研芯片 TeraFab
算力变现	Anthropic $1.25B/月 至 2029	部分算力租赁	多前沿模型客户算力平台

资料来源：XAI官网，techcrunch官网，tomshardware官网，reuters官网，sherwood官网，华福证券研究所

下载日志已记录，仅供内部参考

目录

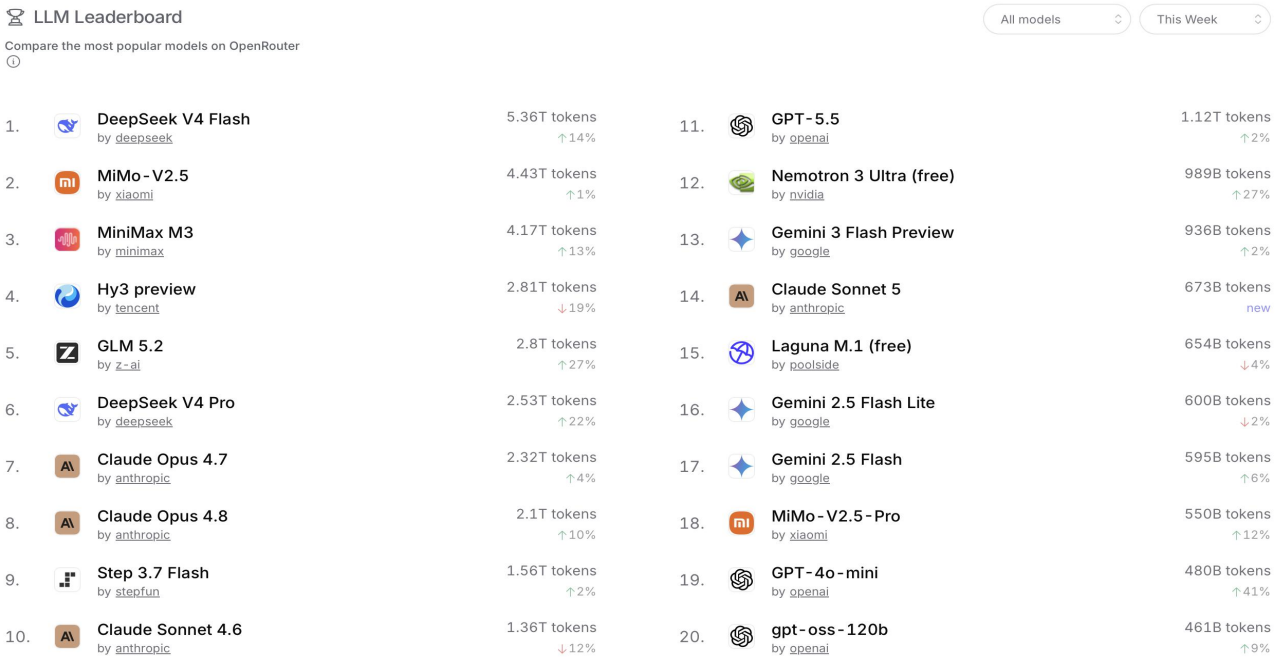
- 一、AGI再定义：Fable 5的能力跃迁
- 二、AI云发力：大模型厂商加码云端入口
- 三、国产追赶：模型能力与应用生态积极突破
- 投资建议
- 风险提示

下载日志已记录，仅供内部参考

1、国产模型积极追赶

- 从 OpenRouter 周度 LLM 调用榜看，国产模型已在全球开发者生态中形成较强存在感。 排名靠前的国产模型普遍具备**够用能力 + 低调用成本**的组合优势。相比海外头部闭源模型，国产模型在推理、代码、长文本和日常 Agent 任务上已经能覆盖大量中高频场景，同时价格更低、响应更快，因此更容易被开发者用于大规模调用。

图：LLM排行榜（openrouter）



资料来源：Openrouter官网，华福证券研究所

下载日志已记录，仅供内部参考

2、国产模型性能差距收窄

- 从模型能力看，GLM-5.2 的 Intelligence Index 已接近 GPT-5.5、Claude Sonnet 等前沿模型，但价格仅约 0.90 美元 / 百万 tokens，且在输出速度和响应延迟上具备优势。随着开发者对单位 token 成本和任务完成效率的敏感度提升，**国产模型有望凭借性价比高持续提升全球调用份额。**

图：模型能力评分（artificial analysis）

Model	Features: Context Window	Features: Creator	Intelligence: Artificial Analysis Intelligence Index	Price: Blended USD/1M Tokens	Speed: Median Tokens/s	Latency: Latency First Chunk (s)	End-to-End Response Time: Total Response (s)
Claude Fable 5 (with fallback)	1M	Anthropic	60	$7.70	70	149.84	156.94
Claude Opus 4.8 (max)	1M	Anthropic	56	$3.85	66	39.46	47.04
GPT-5.5 (xhigh)	922k	OpenAI	55	$4.35	91	87.99	93.48
Claude Opus 4.7 (max)	1M	Anthropic	54	$3.85	57	16.60	25.43
Claude Sonnet 5 (max)	1M	Anthropic	53	$1.54	79	191.31	197.67
GPT-5.5 (high)	922k	OpenAI	53	$4.35	84	24.88	30.86
GLM-5.2 (max)	1M	Z AI	51	$0.90	215	1.37	13.00
GPT-5.5 (medium)	922k	OpenAI	50	$4.35	90	6.04	11.57
Gemini 3.5 Flash	1M	Google	50	$1.31	198	28.68	31.20
Gemini 3.1 Pro Preview	1M	Google	46	$1.74	158	30.10	33.27

资料来源：Artificial Analysis官网，华福证券研究所

下载日志已记录，仅供内部参考

3、AI 视频：效果与成本双优，国产厂商占据全球高位

- AI 视频成为国产模型最先跑到全球前列的应用方向之一。 从 Artificial Analysis 有声视频生成模型榜单看，前十名中国产团队占据多数，Seedance 2.0 位列第一，国产视频模型已经具备全球可比的产品能力。对于短剧、广告、电商素材、社媒内容等场景，视频模型的价值更接近真实内容生产流程。**价格端同样具备竞争力，**榜单中 Seedance 2.0 API 价格约 9.07 美元/分钟，低于 Google Veo 3.1 的 24 美元/分钟；效果进入第一梯队后，国产模型有望依靠成本优势进一步切入海外创作者和企业内容生产市场。

图：有声视频生成模型测评表现

	Range	Creator	Model	Elo	95% CI	Samples	Released	API Pricing[1]
1	1	ByteDance Seed	Dreamina Seedance 2.0 720p	1,223	−9/9	9,628	Mar 2026	$9.07 /min
2	2	Alibaba-ATH	HappyHorse-1.1	1,153	−11/11	2,939	Jun 2026	$9.90 /min
3	3	Alibaba-ATH	HappyHorse-1.0	1,126	−9/9	6,634	Apr 2026	$13.20 /min
4	4–6	Skywork AI	SkyReels V4	1,109	−10/10	3,794	Mar 2026	$21.00 /min
5	4–6	KlingAI	Kling 3.0 1080p (Pro)	1,109	−8/8	7,224	Feb 2026	$20.16 /min
6	4–9	Alibaba	Wan 2.7	1,102	−10/10	3,302	Apr 2026	$9.00 /min
7	6–10	KlingAI	Kling 3.0 Omni 1080p (Pro)	1,099	−8/8	7,013	Feb 2026	$16.80 /min
8	6–10	KlingAI	Kling 3.0 720p (Standard)	1,099	−8/8	7,184	Feb 2026	$15.12 /min
9	6–11	Google	Veo 3.1	1,095	−8/8	7,187	Jan 2026	$24.00 /min
10	7–12	KlingAI	Kling 3.0 Omni 720p (Standard)	1,092	−8/8	6,935	Feb 2026	$13.44 /min

资料来源：Artificial Analysis官网，华福证券研究所

下载日志已记录，仅供内部参考

目录

- 一、AGI再定义：Fable 5的能力跃迁
- 二、AI云发力：大模型厂商加码云端入口
- 三、国产追赶：模型能力与应用生态积极突破
- 投资建议
- 风险提示

下载日志已记录，仅供内部参考

投资建议

➢ **建议1：看好全球 AI 应用加速落地。**

- 随着模型能力持续迭代，AI 正从工具型应用走向更广泛的产业场景渗透，全球 AI 应用生态有望持续扩容。

➢ **建议2：看好 AI 云与算力基础设施长期景气。**

- 全球云厂商持续加码 AI 相关投入，算力需求、云服务收入和基础设施建设有望形成长期支撑。

➢ **建议3：看好国内 AI 生态加速崛起。**

- 国产模型在能力、成本和应用场景上持续优化，国内 AI 应用、模型服务和出海机会有望进一步打开。

下载日志已记录，仅供内部参考

目录

- 一、AGI再定义：Fable 5的能力跃迁
- 二、AI云发力：大模型厂商加码云端入口
- 三、国产追赶：模型能力与应用生态积极突破
- 投资建议
- 风险提示

下载日志已记录，仅供内部参考

风险提示

- 1、AI竞争激烈。
- 2、AI应用竞争激烈。
- 3、AI落地场景变化。

下载日志已记录，仅供内部参考

分析师声明及一般声明

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，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰准确地反映了本人的研究观点。本人不曾因，不因，也将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。

一般声明

华福证券股份有限公司（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。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。

本报告的信息均来源于本公司认为可信的公开资料，该等公开资料的准确性及完整性由其发布者负责，本公司及其研究人员对该等信息不作任何保证。本报告中的资料、意见及预测仅反映本公司于发布本报告当日的判断，之后可能会随情况的变化而调整。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。本公司不保证本报告所含信息及资料保持在最新状态，对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

在任何情况下，本报告所载的信息或所做出的任何建议、意见及推测并不构成所述证券买卖的出价或询价，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。投资者应自行决策，自担投资风险。

本报告版权归“华福证券股份有限公司”所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。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。未经授权的转载，本公司不承担任何转载责任。

下载日志已记录，仅供内部参考

特别声明及投资声明评级

特别声明

投资者应注意，在法律许可的情况下，本公司及其本公司的关联机构可能会持有本报告中涉及的公司所发行的证券并进行交易，也可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。投资者请勿将本报告视为投资或其他决定的唯一参考依据。

投资评级声明

类别	评级	评级说明
公司评级	买入	未来6个月内，个股相对市场基准指数涨幅在20%以上
	持有	未来6个月内，个股相对市场基准指数涨幅介于10%与20%之间
	中性	未来6个月内，个股相对市场基准指数涨幅介于-10%与10%之间
	回避	未来6个月内，个股相对市场基准指数涨幅介于-20%与-10%之间
	卖出	未来6个月内，个股相对市场基准指数涨幅在-20%以下
行业评级	强于大市	未来6个月内，行业整体回报高于市场基准指数5%以上
	跟随大市	未来6个月内，行业整体回报介于市场基准指数-5%与 5%之间
	弱于大市	未来6个月内，行业整体回报低于市场基准指数-5%以下

备注：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）相对同期基准指数的相对市场表现。其中，A股市场以沪深300指数为基准；香港市场以恒生指数为基准；美股市场以标普500指数或纳斯达克综合指数为基准（另有说明的除外）。

下载日志已记录，仅供内部参考

华福证券
HUAFU SECURITIES CO.,LTD.

诚信专业 发现价值

联系方式

华福证券研究所 上海
公司地址：上海市浦东新区浦明路1436号陆家嘴滨江中心MT座20楼
邮编：200120
邮箱：hfyjs@hfzq.com.cn