仅供内部参考，请勿外传

AI 产业景气跟踪框架

证券研究报告
2026 年 08 月 31 日

策略专题

作者

吴开达 分析师
SAC 执业证书编号：S1110524030001
wukaida@tfzq.com

肖峰 分析师
SAC 执业证书编号：S1110524040003
xiaofeng@tfzq.com

相关报告

1《投资策略：A 股策略周报-沃什杰克逊霍尔年会首秀放鹰》 2026-08-30

2《投资策略：历尽天华成此景，人间万事出艰辛-政策与大类资产配置周观察》 2026-08-25

3《投资策略：美债风险怎么看?-港股策略周观察》 2026-08-24

核心结论：五层景气传导仍在延续，系统性转弱信号尚未显现。本报告构建“需求基础—收入兑现与订单储备—投资扩张—资金保障—资本效率”的五层景气跟踪框架，并沿 AI Lab 及大模型厂商、云厂商两条路径交叉验证，从真实需求、经营兑现到资本回报，持续观察 AI 产业景气变化。我们认为，当前五层指标尚未出现共振转弱，AI 产业或仍处于需求深化、商业化提速与基础设施扩张阶段。随着高强度投资延续，产业判断的重点正由“需求是否真实、投资是否继续”转向“资本效率能否维持”。当前 AI 产业景气仍处扩张区间，系统性转弱信号尚未显现。

需求基础：应用渗透持续深化，真实使用需求仍具韧性。Token 消耗、企业 AI 渗透率及用户使用深度同步提升，表明 AI 应用正在由用户数量扩张转向调用频次提升和场景拓展；模型能力升级与调用成本下降进一步改善应用经济性，推动更多高频及长尾需求释放。GPU 租赁价格在前期下行后触底回升，则显示新增算力供给正在被市场吸收。需求已由用户采用逐步传导至模型调用和底层算力消耗，为商业化兑现提供持续支撑。

收入兑现与订单储备：需求正在转化为当期收入，远期合同增强未来可见性。OpenAI 与 Anthropic 的年化收入快速放量，智谱、MiniMax 及商汤的模型相关收入持续增长，表明 API 调用、产品订阅和企业服务正在形成规模化变现；AWS、Microsoft、Google Cloud 及 Oracle 云业务收入同步扩张，反映模型调用和企业部署进一步转化为云资源消费。在当期收入增长的基础上，Microsoft、Alphabet 和 Oracle 披露的 RPO/Backlog 继续扩张，使经营验证由当前收入进一步延伸至未来合同储备。“真实需求—当期收入—远期订单”的传导链条尚未出现明显断点。

投资扩张：模型研发与算力基础设施投入继续增长，投资端尚未出现趋势性收缩。AI Lab 研发费用保持较快增长，头部厂商仍在围绕模型能力、训练算力和产品迭代持续投入；五大云厂商资本开支全面提速，算力基础设施仍处于集中建设阶段，短期继续支撑芯片、服务器和数据中心等上游需求。但部分厂商资本开支增速已经领先云业务收入增速，意味着产业观察重点正由“企业是否继续投资”转向“新增收入能否有效消化快速形成的算力资产”。

资金保障：外部融资仍然活跃，内部现金覆盖由普遍充裕转向分化。美国 AI 创投金额、融资占比和估值提升倍数共同上行，表明外部资本继续为 AI Lab 及初创企业的模型研发、算力采购和商业化扩张提供资金支持。云厂商经营现金流仍具备较强基础，但随着资本开支加速，部分厂商单季资本开支已经超过经营活动现金流，投资后的现金余量有所收窄。当前资金端尚未形成普遍性约束，但内部现金覆盖能力的分化正在提高后续投资扩张对经营兑现和融资条件的要求。

资本效率：ROIC 整体保持韧性，回报端与资金成本端尚未形成共振恶化。从分析框架看，ROIC 与 WACC 的相对关系可以用于观察公司整体资本回报对资金成本的覆盖程度。本文分别跟踪云厂商 ROIC，以及美国 10 年期国债收益率和美国隐含股权风险溢价，对资本回报与资金成本的相对变化进行方向性判断。当前五大云厂商 ROIC 尚未出现同步下行，厂商间表现虽有分化，但整体回报仍具韧性；与此同时，无风险利率上行与隐含 ERP 中枢回落并存，资金成本的核心驱动尚未形成一致上行。后续若 ROIC 普遍走弱，同时美债利率或隐含 ERP 上升，资本效率压力将明显增强。综合需求、收入与订单、投资、资金及资本效率五层指标，当前产业传导仍在延续，系统性转弱信号尚未显现。

风险提示：产业发展进程不及预期；政策出台和落地具备不确定性；宏观经济波动；模型价格为手工统计，仅供参考。

仅供内部参考，请勿外传

内容目录

1. 核心结论：五层景气传导仍在延续，资本效率成为验证重点 3
2. 需求基础：应用渗透深化，算力供需边际收紧 3
3. 收入兑现与订单储备：商业化放量，未来收入可见性增强 7
4. 投资扩张：AI Lab 研发高增，云厂商 CapEx 全面提速 9
5. 资金保障：外部融资充裕，内部现金覆盖有所分化 10
6. 资本效率：ROIC 整体保持韧性，资金成本约束仍需观察 13
7. 风险提示 15

图表目录

图 1：AI 产业景气度的五层跟踪框架 3
图 2：OpenRouter Token 消耗持续攀升，模型调用需求加速释放 4
图 3：美国企业 AI 使用率继续上行，未来六个月采用意愿更强 4
图 4：用户使用时间越长，ChatGPT 功能覆盖与交互深度越高 5
图 5：模型 API 价格由普遍下调转向分化，部分头部模型开始提价 5
图 6：主流模型价格分层加速，低成本 API 供给显著丰富 6
图 7：模型能力与任务成本总体正相关，高性价比模型已形成 6
图 8：四代 GPU 租赁价格同步触底回升，算力供需边际收紧 7
图 9：OpenAI 与 Anthropic ARR 加速攀升，头部 AI Lab 进入规模化变现阶段 8
图 10：国内大模型厂商收入持续高增，商业化兑现提速（单位：亿元） 8
图 11：北美主要云厂商收入维持高增，AI 需求持续向基础设施层传导（单位：亿美元） 9
图 12：三大云厂商 RPO/Backlog 加速增长，未来收入可见性增强（单位：十亿美元） 9
图 13：智谱研发投入持续扩张，2025 年研发费用升至 31.8 亿元 10
图 14：MiniMax 研发投入保持增长，2022—2025 年研发费用扩大近 24 倍 10
图 15：北美五大云厂商 CapEx 全面提速，2026Q2 同比增速均超 60%（单位：十亿美元） 10
图 16：美国 AI 创投金额在 2026H1 跃升，交易数量保持高位（单位：十亿美元、笔） 11
图 17：AI 吸收超八成美国 VC 投资金额，资本进一步向 AI 集中 12
图 18：AI 项目投前估值随轮次显著抬升，D+轮中位估值突破 40 亿美元（单位：百万美元） 12
图 19：AI 融资估值提升倍数再度上行，平均值与中位数分化扩大 13
图 20：资本开支加速压缩内部现金余量，三家公司单季转负（单位：十亿美元） 13
图 21：五大云厂商 ROIC 整体保持韧性，厂商间走势分化 14
图 22：美债利率上行、隐含 ERP 中枢回落，资金成本约束仍存 14

仅供内部参考，请勿外传

1. 核心结论：五层景气传导仍在延续，资本效率成为验证重点

本报告构建“需求基础—收入兑现与订单储备—投资扩张—资金保障—资本效率”的五层景气跟踪框架，并沿 AI Lab 及大模型厂商、云厂商两条路径交叉验证，从真实需求、经营兑现到资本回报，持续观察 AI 产业景气变化。我们认为，当前五层指标尚未出现共振转弱，AI 产业或仍处于需求深化、商业化提速与基础设施扩张阶段。随着高强度投资延续，产业判断的重点正由“需求是否真实、投资是否继续”转向“资本效率能否维持”。当前 AI 产业景气仍处扩张区间，系统性转弱信号尚未显现。

图 1：AI 产业景气度的五层跟踪框架



资料来源：天风证券研究所

2. 需求基础：应用渗透深化，算力供需边际收紧

我们认为，AI 需求是否真实可沿“采用范围扩大—使用强度加深—应用经济性改善—算力需求兑现”的链条逐级验证。B 端 AI 渗透率衡量企业采用广度，C 端功能覆盖和消息活跃度反映用户黏性，Token 消耗刻画模型实际调用强度；API 价格与模型性价比决定需求规模化扩散的成本门槛，GPU 租赁价格则检验应用增长能否最终转化为底层算力消耗。当前各环节呈现同向改善，表明 AI 需求正在由尝试性使用进入高频调用和规模部署阶段，为后续收入兑现奠定基础。

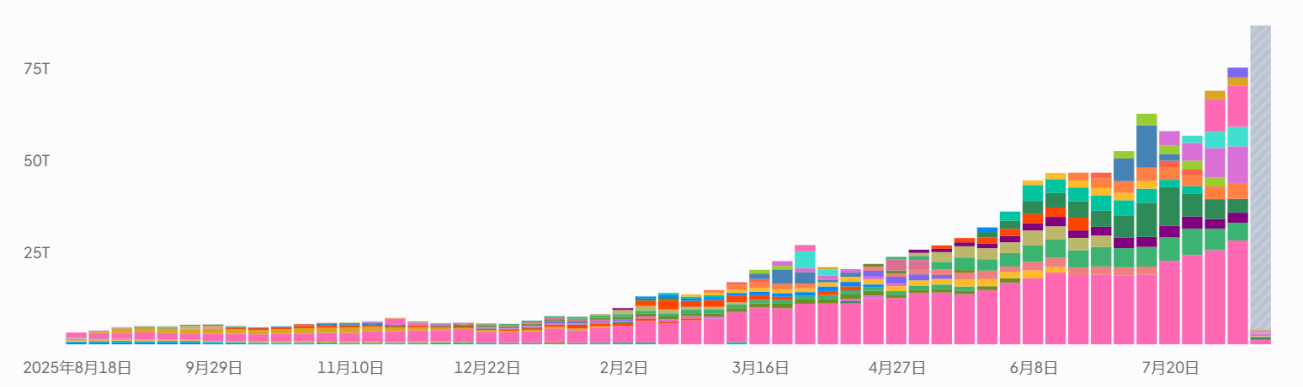
需求侧指标共振向上，AI 产业仍处于渗透率提升、使用强度加深的景气阶段。Token 消耗、B 端渗透率和 C 端使用深度分别反映模型调用规模、企业采用范围和用户黏性。OpenRouter 平台 Token 消耗持续增长，美国企业当前及未来六个月 AI 使用率同步提升（该数据为调查数据），ChatGPT 用户的功能使用广度和消息活跃度随使用时间总体上升。多维指标相互印证，AI 需求正由用户数量扩张进一步转向使用频次提升和应用场景拓展。

模型能力升级与调用成本下降形成正向循环，AI 应用规模化落地的经济性持续改善。模型调价、当前 API 价格以及“智能水平—单任务成本”共同反映 AI 能力的可负担程度。主流厂商通过降价、代际升级和产品分层持续降低调用门槛，成本—能力前沿不断向“更高智能、更低成本”移动，推动更多高频、长尾场景具备商业化基础。我们认为，价格下降并非需求疲弱的表现，而是技术进步和市场竞争推动 AI 需求加速释放的重要条件。

GPU 租赁价格触底回升，算力供需正由前期宽松转向再平衡。2024—2025 年，新增算力集中投放带动 GPU 租赁价格持续下行；2025 年末以来，H100、H200、B200 和 B300 价格相继企稳回升，显示前期新增供给的消化速度加快，底层算力需求仍具支撑。结合 Token 消耗持续增长、B 端渗透率提升及 C 端使用深度加深，AI 需求已由用户采用逐步传导至实际调用和算力消耗，需求侧尚未出现趋势性转弱信号。下一步需要验证的是，这部分使用需求能否持续转化为 AI Lab 和云厂商收入。

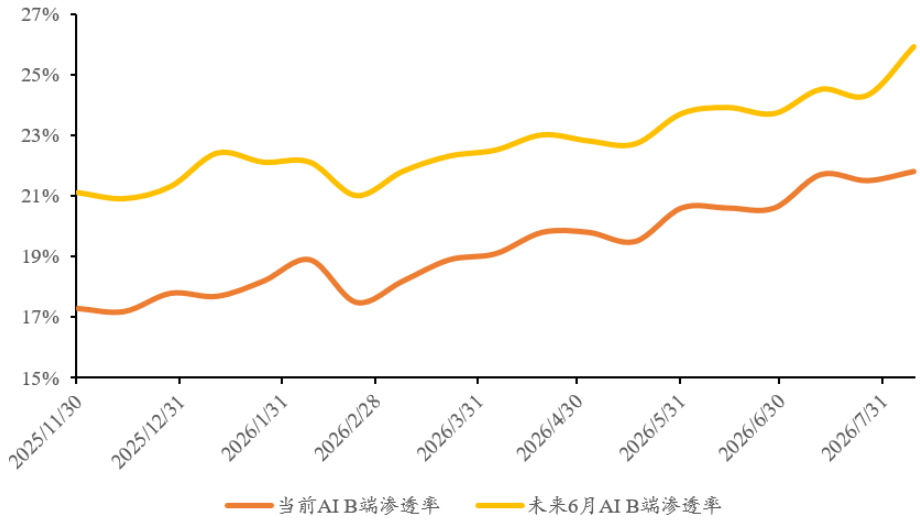
仅供内部参考，请勿外传

图 2：OpenRouter Token 消耗持续攀升，模型调用需求加速释放



资料来源：openrouter，天风证券研究所

图 3：美国企业 AI 使用率继续上行，未来六个月采用意愿更强

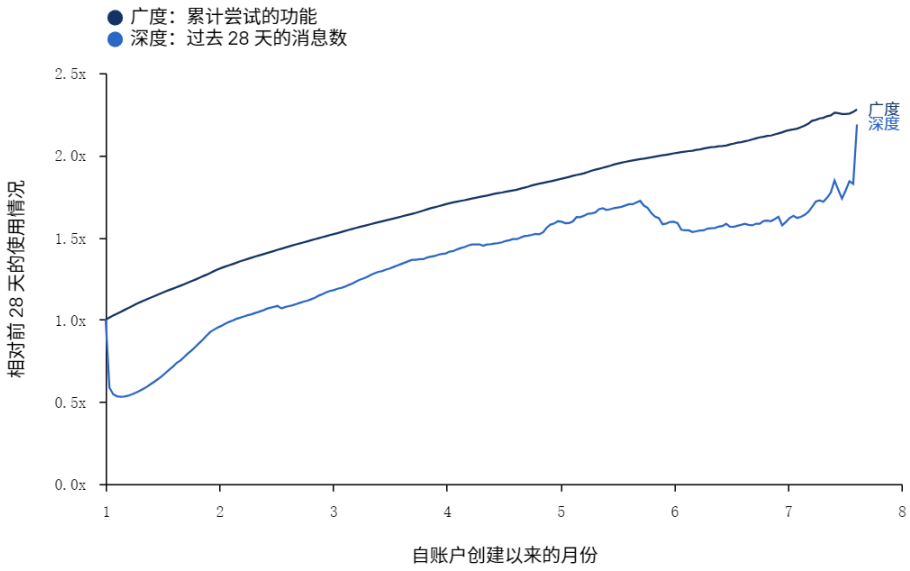


资料来源：U.S. Census Bureau，天风证券研究所

备注：未来 6 月 AI B 端渗透率为调查数据而非预测数据

仅供内部参考，请勿外传

图 4：用户使用时间越长，ChatGPT 功能覆盖与交互深度越高



资料来源：OpenAI，天风证券研究所

图 5：模型 API 价格由普遍下调转向分化，部分头部模型开始提价

日期	厂商/模型	调整前 → 调整后	变动幅度
2025/9/29	DeepSeek V3.2-Exp	官方未逐项披露全部前后价格，但明确表示API新价格立即生效，开发者调用成本下降50%以上	**下降超过50%**
2025/11/6	Kimi K2 Turbo	未缓存输入：¥16→¥8； 缓存输入：¥4→¥1； 输出：¥64→¥58	**-50% / -75% / -9%**
2025/11/24	Claude Opus 4.5	相比Opus 4.1：输入$15→$5；输出$75→$25	均 **-66.7%**
2026/1/27	Qwen3-Max	不超过32K：输入¥3→¥2.5；输出¥12.8→¥10 32K－128K：输入 ¥6.4 → ¥4；输出 ¥25.6 → ¥16 128K－256K：输入 ¥9.6 → ¥7；输出 ¥38.4 → ¥28	约 **-22%**
2026/6/8左右	MiniMax M3	不超过512K：未缓存输入¥4.2→$2.1；输出¥16.8→¥8.4 超过 512K：输入 ¥8.4 → ¥4.2；输出 ¥33.6 → ¥16.8；	均 -50%
2026/6/30发布；8/10永久化	Claude Sonnet 5	相比Sonnet 4.6：未缓存输入$3→$2； 缓存输入$0.30→$0.20；输出$15→$10	均 -33.3%
2026/7/21	Gemini 3.6 Flash	相比Gemini 3.5 Flash：输入$1.50→$1.50；输出$9→$7.50	输入不变；输出 -16.7%
2026/7/30	GPT-5.6 Luna	未缓存输入：$1→$0.20； 缓存输入：$0.10→$0.02； 输出：$6→$1.20	均 -80%
2026/7/30	GPT-5.6 Terra	未缓存输入：$2.50→$2； 缓存输入：$0.25→$0.20； 输出：$15→$12	均 -20%
2026/8/12	Grok 4.6	相比Grok 4.5标准上下文档：未缓存输入$2→$2；缓存输入$0.30→$0.50；输出$6→$6	0% / **+66.7%** / 0%
2026/8/17	DeepSeek V4-Pro-0813	未缓存输入：¥3→峰值¥9/闲时¥4.5；缓存输入：¥0.025→峰值¥0.30/闲时¥0.15；输出：¥6→峰值¥27/闲时¥13.5	峰值：**+200% / +1100% / +350%**； 闲时：**+50% / +500% / +125%**
2026/8/17	DeepSeek V4-Flash-0731	未缓存输入：¥1→峰值¥3/闲时¥1.5；缓存输入：¥0.02→峰值¥0.10/闲时¥0.05；输出：¥2→峰值¥9/闲时¥4.5	峰值：**+200% / +400% / +350%**； 闲时：**+50% / +150% / +125%**

资料来源：各公司官网，蓝鲸新闻公众号等，天风证券研究所

注：价格数据为手工收集统计,可能存在误差,仅供参考

仅供内部参考，请勿外传

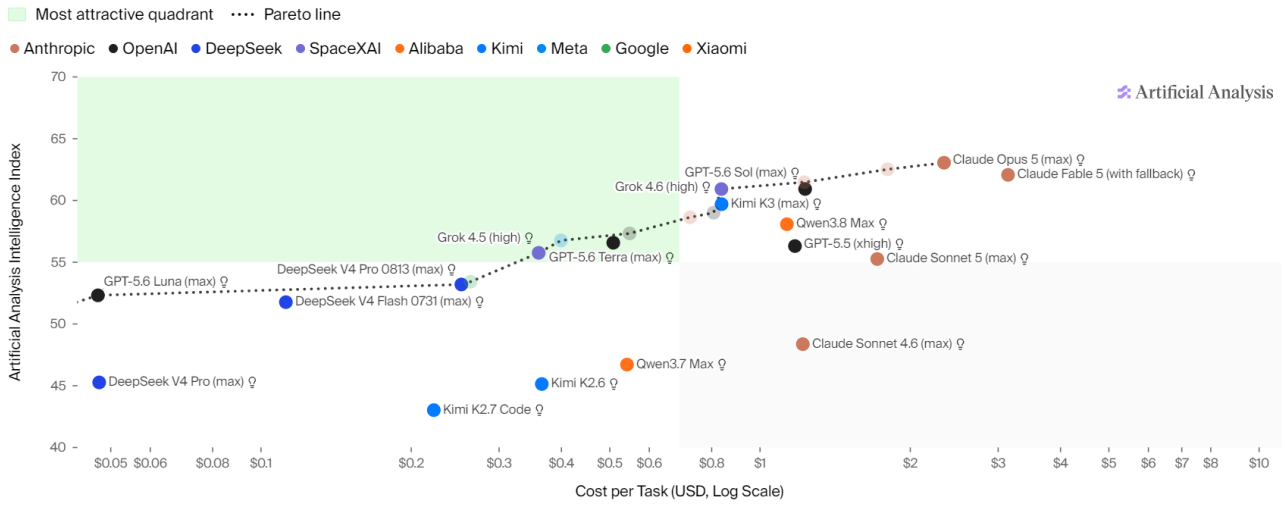
图 6：主流模型价格分层加速，低成本 API 供给显著丰富

厂商	当前代表模型	输入价格	缓存命中输入价格	输出
OpenAI	GPT-5.6 Sol	4美元/百万tokens	0.40美元/百万tokens	20美元/百万tokens
	GPT-5.6 Terra	2美元/百万tokens	0.20美元/百万tokens	12美元/百万tokens
	GPT-5.6 Luna	0.2美元/百万tokens	0.02美元/百万tokens	1.20美元/百万tokens
Anthropic	Claude Fable 5	10美元/百万tokens	1美元/百万tokens	50美元/百万tokens
	Claude Mythos 5	10美元/百万tokens	1美元/百万tokens	50美元/百万tokens
	Claude Opus 5	5美元/百万tokens	0.50美元/百万tokens	25美元/百万tokens
	Claude Sonnet 5	2美元/百万tokens	0.20美元/百万tokens	10美元/百万tokens
xAI/Grok	Grok 4.6	2美元/百万tokens	0.50美元/百万tokens	6美元/百万tokens
	Grok 4.5	2美元/百万tokens	0.30美元/百万tokens	6美元/百万tokens
	Grok Build 0.1	1美元/百万tokens	0.20美元/百万tokens	2美元/百万tokens
Google	Gemini 3.1 Pro Preview	2美元/百万tokens（prompts <= 200k tokens） 4美元/百万tokens（prompts>= 200k tokens）	0.20美元/百万tokens（prompts <= 200k tokens） 0.40美元/百万tokens（prompts>= 200k tokens）	12美元/百万tokens（prompts <= 200k tokens） 18美元/百万tokens（prompts>= 200k tokens）
	Gemini 3.7 Flash	0.75美元/百万tokens（2026.12.31之前） 1.5美元/百万tokens（2027.1.1之后）	0.075美元/百万tokens（2026.12.31之前） 0.15美元/百万tokens（2027.1.1之后）	3.75美元/百万tokens（2026.12.31之前） 7.5美元/百万tokens（2027.1.1之后）
DeepSeek	V4-Pro-0813	峰值：9 元/百万 tokens 闲时：4.5 元/百万 tokens	峰值0.30元/百万tokens； 闲时0.15元/百万tokens	峰值27元/百万tokens； 闲时13.5元/百万tokens
	V4-Flash-0731	高峰：3 元/百万 tokens 闲时：1.5 元/百万 tokens	峰值0.10元/百万tokens； 闲时0.05元/百万tokens	峰值9元/百万tokens； 闲时4.5元/百万tokens
Kimi	Kimi K3	20元/百万tokens	2元/百万tokens	100元/百万tokens
	Kimi K2.6	6.50 元/百万 tokens	1.10元/百万tokens	27元/百万tokens
	Kimi K2.7 Code	6.50 元/百万 tokens	1.30元/百万tokens	27元/百万tokens
智谱	GLM-5.3	8 元/百万 tokens	2元/百万tokens	28元/百万tokens
	GLM-5.3 Flash（限时5折两周）	0.4 元/百万 tokens	0.115元/百万tokens	1.4元/百万tokens
	GLM-5.2	8 元/百万 tokens	2元/百万tokens	28元/百万tokens
MiniMax	MiniMax M3	≤512K：2.1元/百万tokens；>512K：4.2元/百万tokens	≤512K：0.42元/百万tokens；>512K：0.84元/百万tokens	≤512K：8.40元/百万tokens；>512K：16.80元/百万tokens
	MiniMax M2.7	2.1元/百万tokens	0.42元/百万tokens	8.40元/百万tokens
阿里千问	Qwen3.8 Max	12元/百万tokens	-	36元/百万tokens
	Qwen3.7 Max	12元/百万tokens	-	36元/百万tokens
字节豆包	Seed 2.0 Pro	≤32K：3.20 元/百万 tokens 32K－128K：4.80 元/百万 tokens 128K－256K：9.60 元/百万 tokens	-	≤32K：16 元/百万 tokens 32K－128K：24 元/百万 tokens 128K－256K：48 元/百万 tokens
	Seed 2.0 Lite	≤32K：0.60 元/百万 tokens 32K－128K：0.90 元/百万 tokens 128K－256K：1.80 元/百万 tokens	-	≤32K：3.60 元/百万 tokens 32K－128K：5.40 元/百万 tokens 128K－256K：10.80 元/百万 tokens
	Seed 2.0 Mini	≤32K：0.20 元/百万 tokens 32K－128K：0.40 元/百万 tokens 128K－256K：0.80 元/百万 tokens	-	≤32K：2 元/百万 tokens 32K－128K：4 元/百万 tokens 128K－256K：8 元/百万 tokens

资料来源：各公司官网，天风证券研究所

注：统计时间截至 2026.8.30，模型价格数据为手工收集统计，可能存在误差，仅供参考

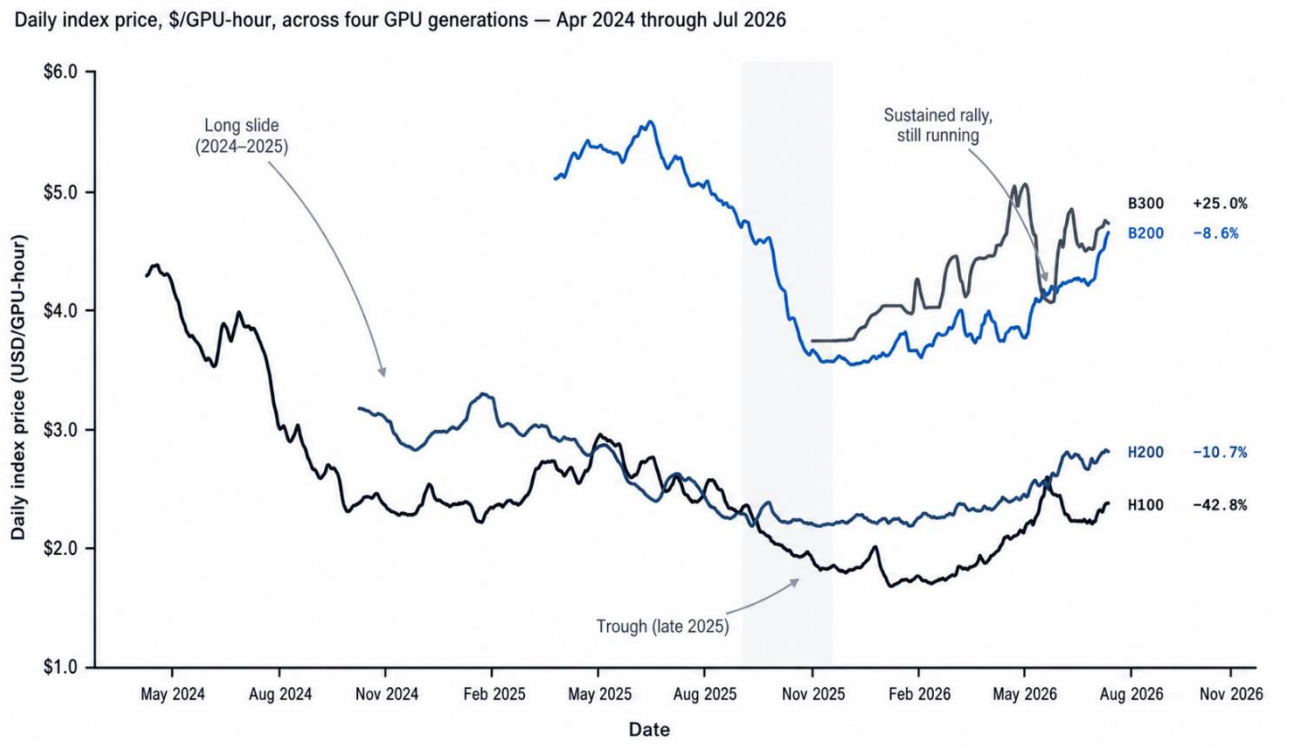
图 7：模型能力与任务成本总体正相关，高性价比模型已形成



资料来源：Artificial Analysis，天风证券研究所

仅供内部参考，请勿外传

图 8：四代 GPU 租赁价格同步触底回升，算力供需边际收紧



资料来源：Architect，天风证券研究所

3. 收入兑现与订单储备：商业化放量，未来收入可见性增强

需求真实只是产业验证的第一步，收入与订单共同决定商业化能否持续。AI Lab 及大模型厂商更接近模型与应用层，其收入变化直接反映模型能力能否通过 API 调用、产品订阅和企业服务实现变现；北美云厂商位于算力基础设施层，云业务收入用于检验模型调用和企业部署能否进一步转化为云资源消费；RPO/Backlog 则代表已经签订但尚未确认收入的合同储备，用于观察未来收入的可见性。三类指标分别对应模型层直接变现、基础设施层收入承接和远期订单支撑，共同构成从“需求真实”到“经营兑现”的完整验证。

AI Lab 收入加速放量，模型层商业闭环已由早期验证迈向规模扩张。按各披露节点的 ARR 或年化收入运行率口径，OpenAI 由 2023 年 8 月超过 10 亿美元增至 2026 年 2 月超过 250 亿美元，不足三年扩大约 25 倍；Anthropic 由 2024 年末约 10 亿美元增至 2026 年 4 月超过 300 亿美元，约 16 个月扩大约 30 倍，且 2025 年下半年以来增长斜率明显抬升。国内模型厂商同样实现持续增长，2023—2025 年，智谱收入由 1.25 亿元增至 7.24 亿元，MiniMax 由 0.25 亿元增至 5.56 亿元，商汤生成式 AI 业务收入由 11.84 亿元增至 36.30 亿元。海外头部厂商年化收入快速放量与国内模型厂商收入增长相互印证，表明模型调用正在通过 API、订阅和企业服务形成真实付费闭环。

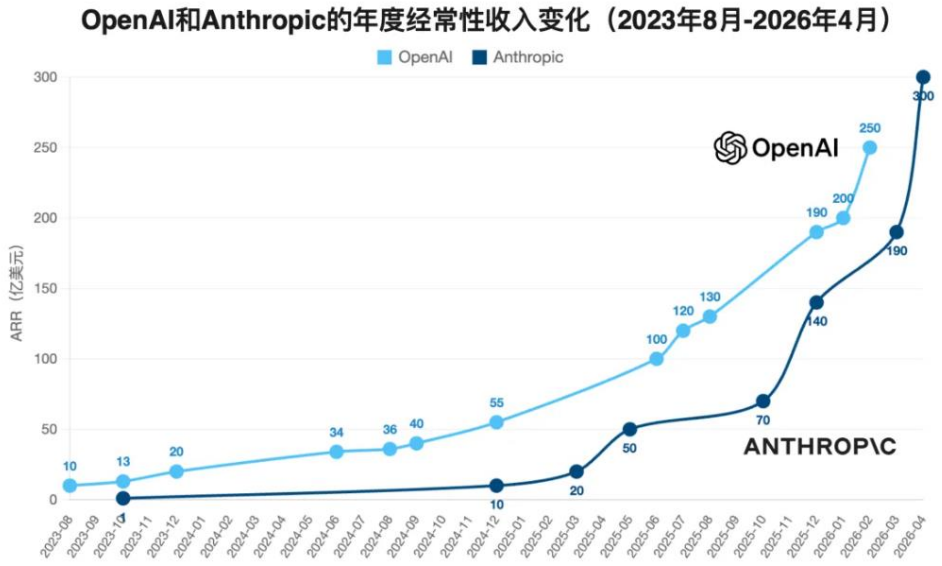
云业务收入维持高增，模型调用正在向基础设施层形成报表兑现。2026Q2，AWS、Microsoft Intelligent Cloud、Google Cloud 和 Oracle OCI/IaaS 单季收入分别达到 422.32 亿、393.06 亿、247.68 亿和 58.00 亿美元，同比分别增长 36.8%、31.6%、81.8%和 93.3%。AWS 与 Microsoft 在接近或超过 400 亿美元的单季收入体量上仍保持较快增长，Google Cloud 与 Oracle OCI/IaaS 则处于更快放量阶段。尽管各家公司披露范围并不完全一致，但云业务收入的共同增长表明，AI 需求并未停留在 Token 消耗和用户活跃层面，而是正在转化为企业部署、模型推理及相关云资源消费。

合同储备同步扩张，进一步将收入验证由当期延伸至未来。2026Q2，Microsoft Commercial RPO、Alphabet Revenue Backlog 和 Oracle RPO 分别达到 6,780 亿、5,195 亿和 6,380 亿美元，同比分别增长 84.2%、380.1%和 363.0%。需要注意的是，三项指标的定义和业务范围并不完全一致，并非同口径的纯 AI 或纯云订单，也不等同于已经确认的收入和现金流；但其

仅供内部参考，请勿外传

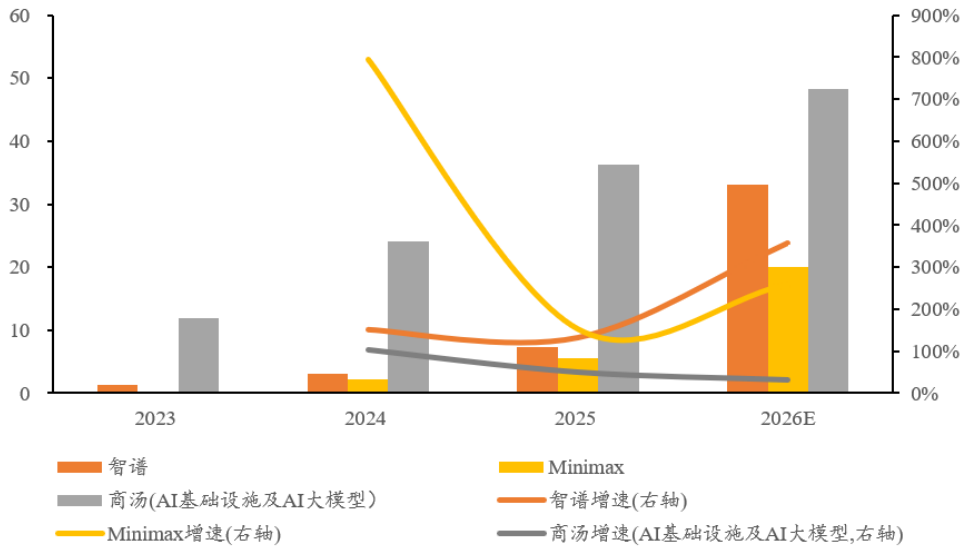
共同增长表明，云厂商未来收入仍具备较强的合同储备支撑。综合来看，AI 需求不仅正在转化为当期模型及云业务收入，持续增长的远期订单也增强了未来收入可见性，“需求基础—当期收入—订单储备”的经营传导尚未出现明显断点，并为企业继续扩大模型研发和算力基础设施投资提供了需求基础。

图 9：OpenAI 与 Anthropic ARR 加速攀升，头部 AI Lab 进入规模化变现阶段



资料来源：财经 AI 湃公众号，天风证券研究所

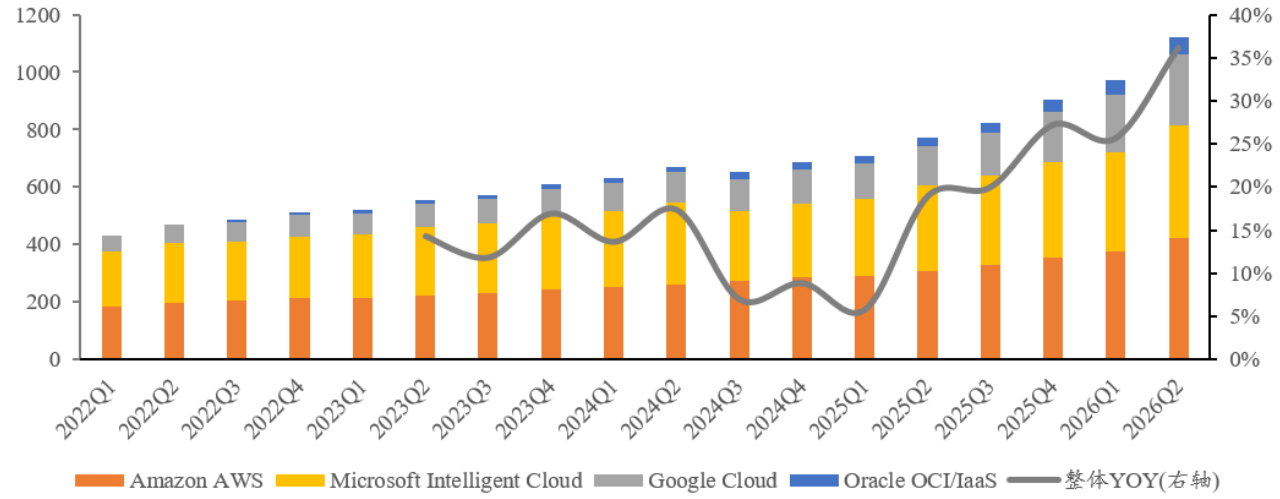
图 10：国内大模型厂商收入持续高增，商业化兑现提速（单位：亿元）



资料来源：Wind，天风证券研究所

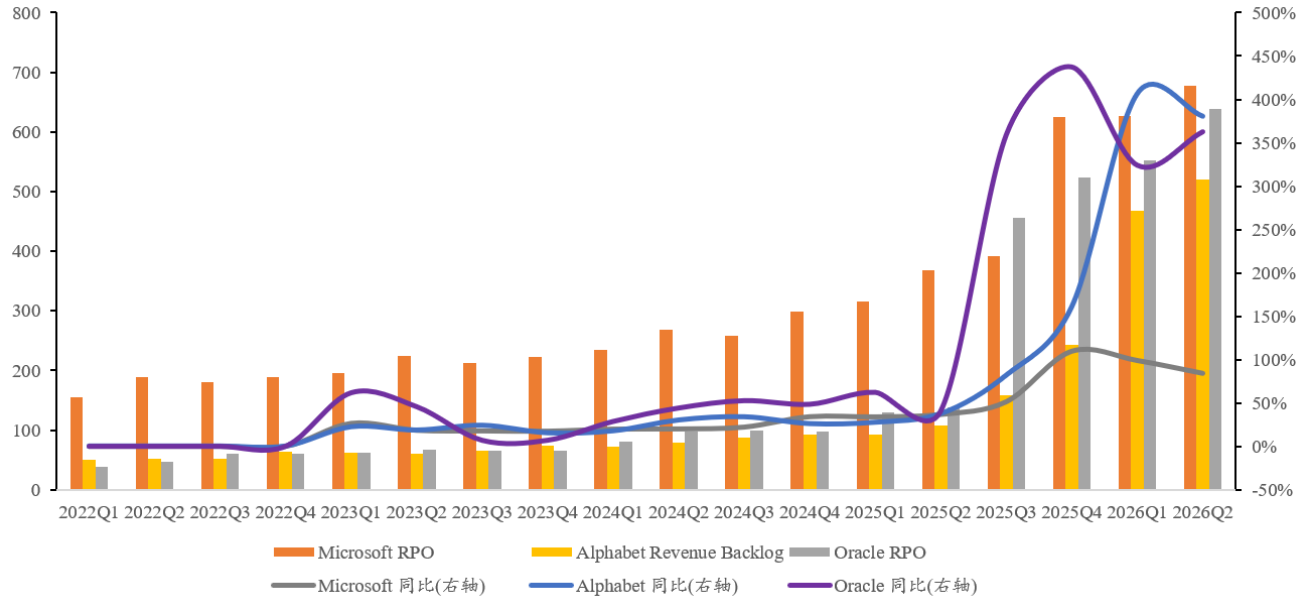
仅供内部参考，请勿外传

图 11：北美主要云厂商收入维持高增，AI 需求持续向基础设施层传导（单位：亿美元）



资料来源：Wind，天风证券研究所

图 12：三大云厂商 RPO/Backlog 加速增长，未来收入可见性增强（单位：十亿美元）



资料来源：各公司财报，天风证券研究所

4. 投资扩张：AI Lab 研发高增，云厂商 CapEx 全面提速

收入持续增长与远期订单扩张，为企业延续模型研发和算力基础设施投入提供经营侧支撑，当前投资扩张仍在延续。AI Lab 研发费用反映企业围绕模型能力、训练算力和产品迭代的投入意愿，云厂商资本开支则衡量数据中心、服务器及网络基础设施的扩张强度；进一步比较 CapEx 同比与云收入同比，可以观察投资扩张速度能否与商业化兑现相匹配。当前研发费用与资本开支仍在同步增长，投资端尚未出现趋势性收缩。

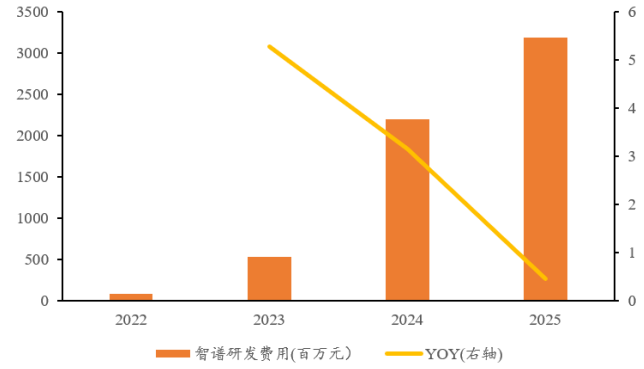
AI Lab 研发投入持续增长，模型侧仍处于能力竞争和产品迭代阶段。2022—2025 年，智谱研发费用由 0.84 亿元增至 31.80 亿元，累计扩大约 37.7 倍；MiniMax 研发费用由 1,056 万美元增至 2.53 亿美元，累计扩大约 23.9 倍。尽管高基数下增速有所回落，2025 年两家公司研发费用仍分别同比增长 44.9%和 33.8%。头部 AI Lab 继续加大模型研发与训练投入，表明模型侧尚未由能力扩张转入主动收缩阶段。

仅供内部参考，请勿外传

云厂商 CapEx 全面提速，算力基础设施仍处于集中建设期。2026Q2，Amazon、Microsoft、Alphabet、Oracle 和 Meta 资本开支分别达到 530.8 亿、358.0 亿、449.2 亿、164.9 亿和 301.2 亿美元，同比分别增长 64.9%、109.6%、100.1%、81.6%和 82.1%。其中，Amazon、Microsoft 和 Alphabet 的 CapEx 增速分别高于对应云业务收入增速 28.2、78.0 和 18.3 个百分点，Oracle 的 CapEx 增速则低于云收入增速 11.7 个百分点。云厂商资本开支继续保持高增长，短期仍将支撑芯片、服务器和数据中心等上游环节的需求。

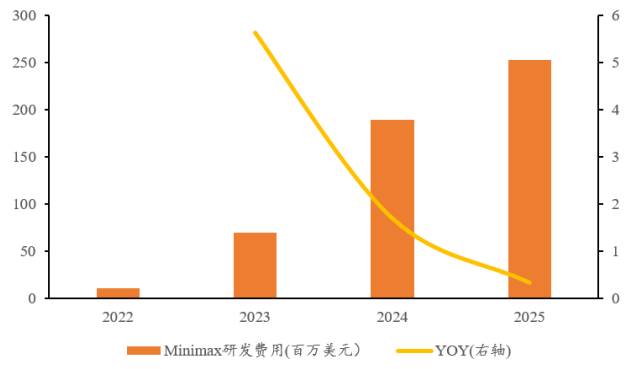
投资扩张尚未出现拐点，但部分平台投入增速已经明显领先收入增速。随着算力资产集中形成，产业判断的重点将由“企业是否继续投资”进一步转向“新增收入能否消化资本开支、资金来源能否支持投资延续”。因此，下一部分将从外部融资、资本开支后的现金余量及 ROIC-WACC 三个维度，进一步检验 AI 投资周期的资金保障能力和资本效率。

图 13：智谱研发投入持续扩张，2025 年研发费用升至 31.8 亿元



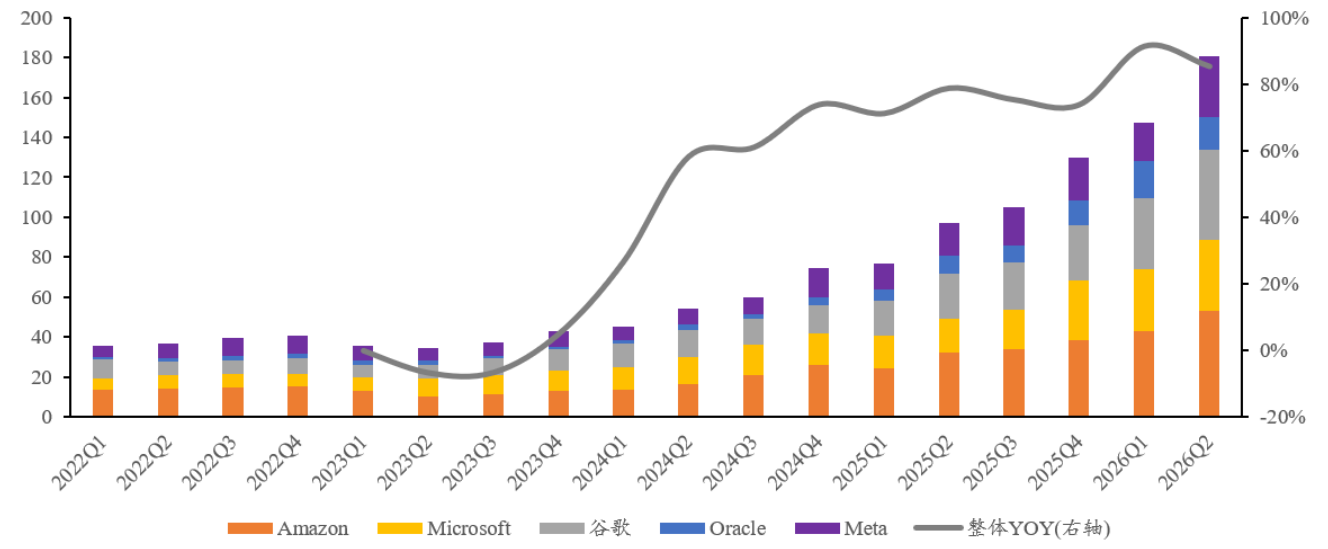
资料来源：Wind，天风证券研究所

图 14：MiniMax 研发投入保持增长，2022—2025 年研发费用扩大近 24 倍



资料来源：Wind，天风证券研究所

图 15：北美五大云厂商 CapEx 全面提速，2026Q2 同比增速均超 60%（单位：十亿美元）



资料来源：Wind，天风证券研究所

5. 资金保障：外部融资充裕，内部现金覆盖有所分化

资金保障是检验 AI 投资扩张能否延续的重要前提。当前 AI Lab 研发投入和云厂商资本开支仍在扩张，投资周期能否延续，首先取决于外部资本能否持续供给，以及经营现金流能否覆盖高强度投入。基于此，我们以美国 AI VC 融资金额、AI 融资占比和估值提升倍数观察 AI Lab 及初创企业的外部资金环境，以“经营活动现金流净额－现金资本开支”衡量云

仅供内部参考，请勿外传

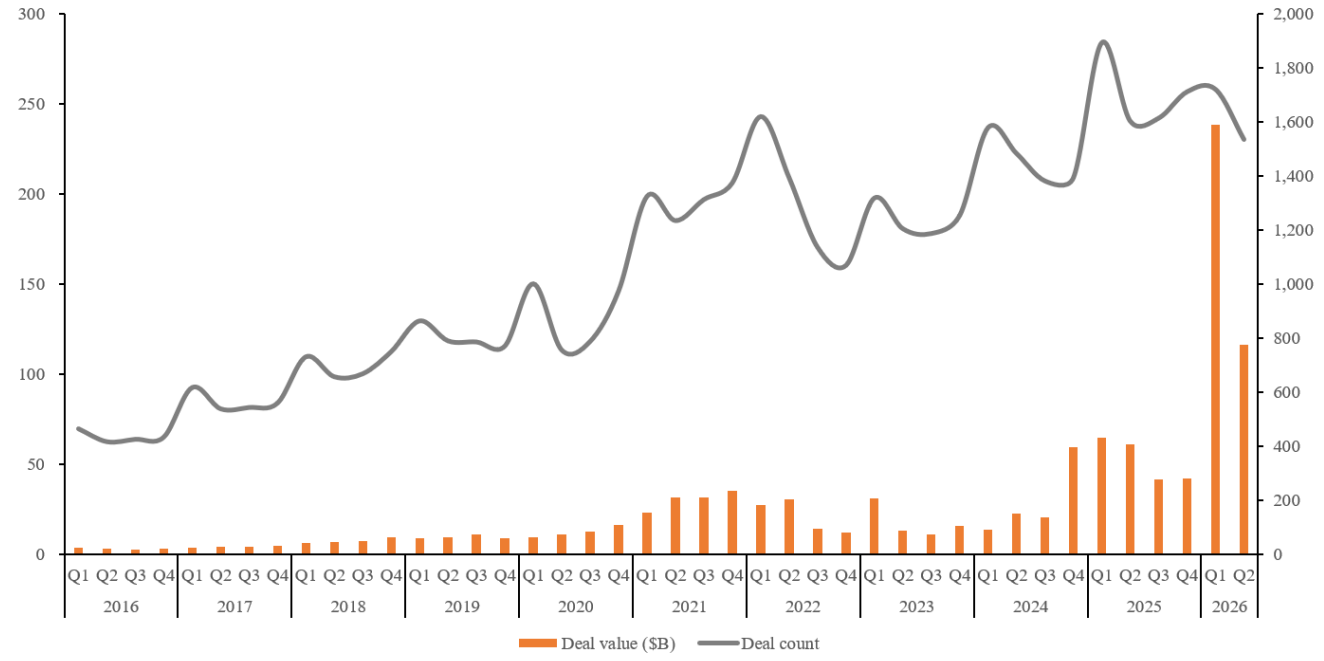
厂商的内部资金余量。

AI 融资规模与估值水平同步上行，外部资本继续加码 AI 产业。2026 年上半年，美国 AI VC 投资金额达到 3,548.8 亿美元；2026Q1 和 Q2，AI 投资金额占全部 VC 投资金额的比例分别达到 88.6%和 81.0%，交易数量占比分别为 44.3%和 42.0%。融资金额占比显著高于交易数量占比，显示 AI 项目的融资规模继续领先市场平均水平。同期，AI 融资估值提升倍数的中位数和平均数分别升至 2.19 倍和 3.23 倍，D+轮项目投前估值中位数达到 42.5 亿美元。融资规模、资本占比与估值水平共同上行，为 AI Lab 的模型研发、算力采购和商业化扩张提供了持续的外部资金支持。

云厂商现金覆盖出现分化，资本开支加速正在提高内部资金占用。2026Q2，Microsoft 和 Meta 的“经营活动现金流净额－现金资本开支”分别为 196.4 亿和 17.4 亿美元，仍保持正向现金余量；Amazon、Alphabet 和 Oracle 则阶段性转为负值，表明其单季基础设施投入已经超过当期经营现金流。整体来看，头部云厂商仍具备较强的经营造血基础，但资本开支快速扩张已经压缩部分厂商投资后的现金余量，内部现金覆盖能力正由普遍充裕转向分化。

综合来看，外部融资仍然活跃，内部现金覆盖虽然出现分化，但尚未形成普遍性的资金约束。 后续若 AI 融资条件明显收紧，并伴随云厂商资本开支后现金余量持续下降，资金端对投资扩张的约束将明显增强。

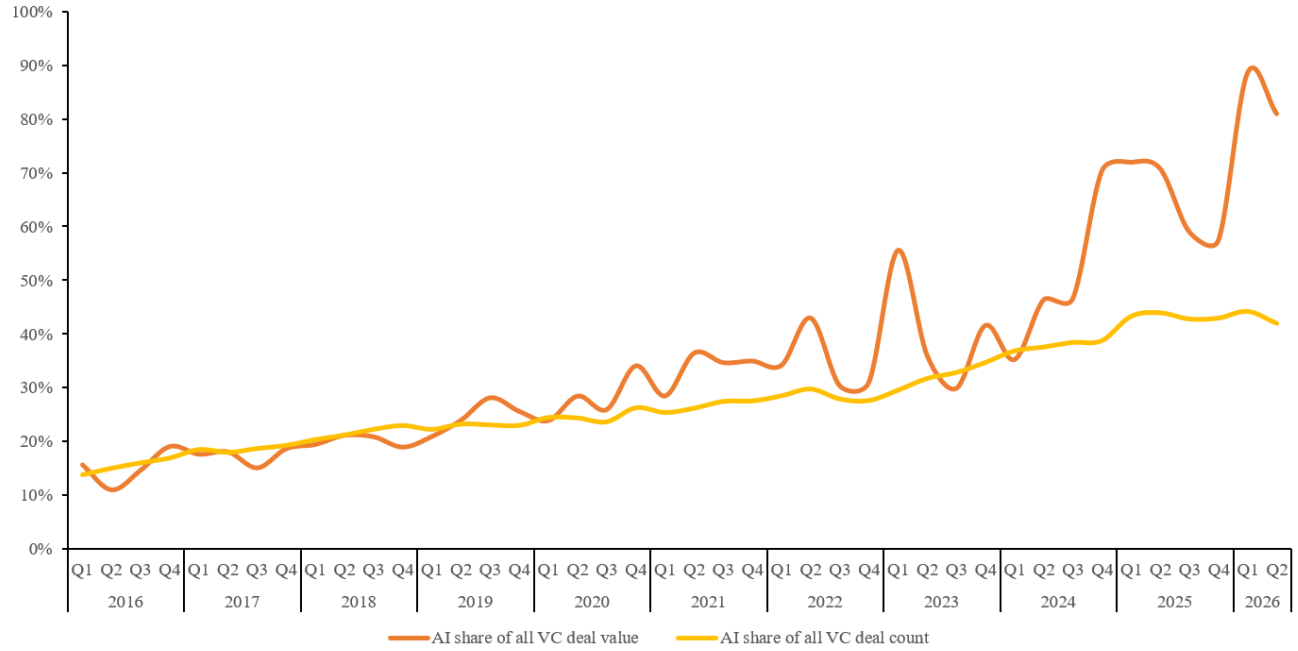
图 16：美国 AI 创投金额在 2026H1 跃升，交易数量保持高位（单位：十亿美元、笔）



资料来源：NVCA，天风证券研究所

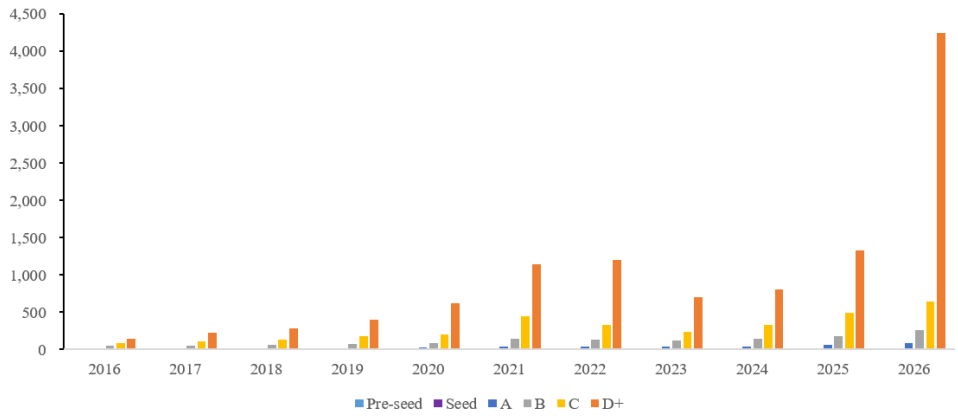
仅供内部参考，请勿外传

图 17：AI 吸收超八成美国 VC 投资金额，资本进一步向 AI 集中



资料来源：NVCA，天风证券研究所

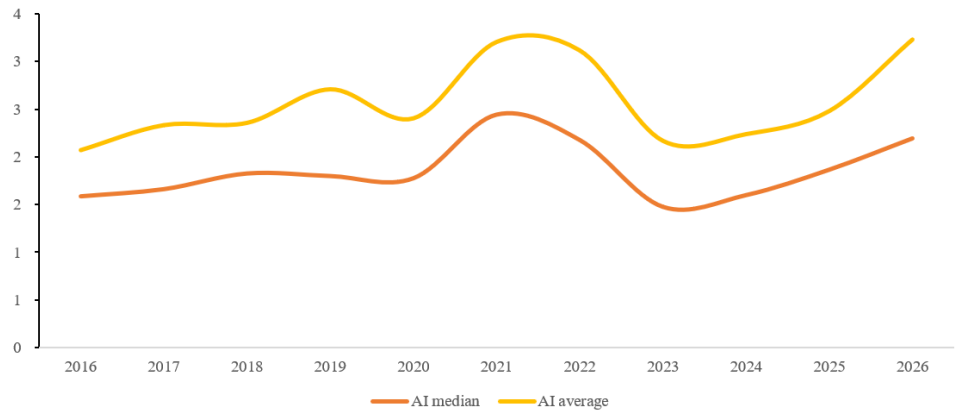
图 18：AI 项目投前估值随轮次显著抬升，D+轮中位估值突破 40 亿美元（单位：百万美元）



资料来源：NVCA，天风证券研究所

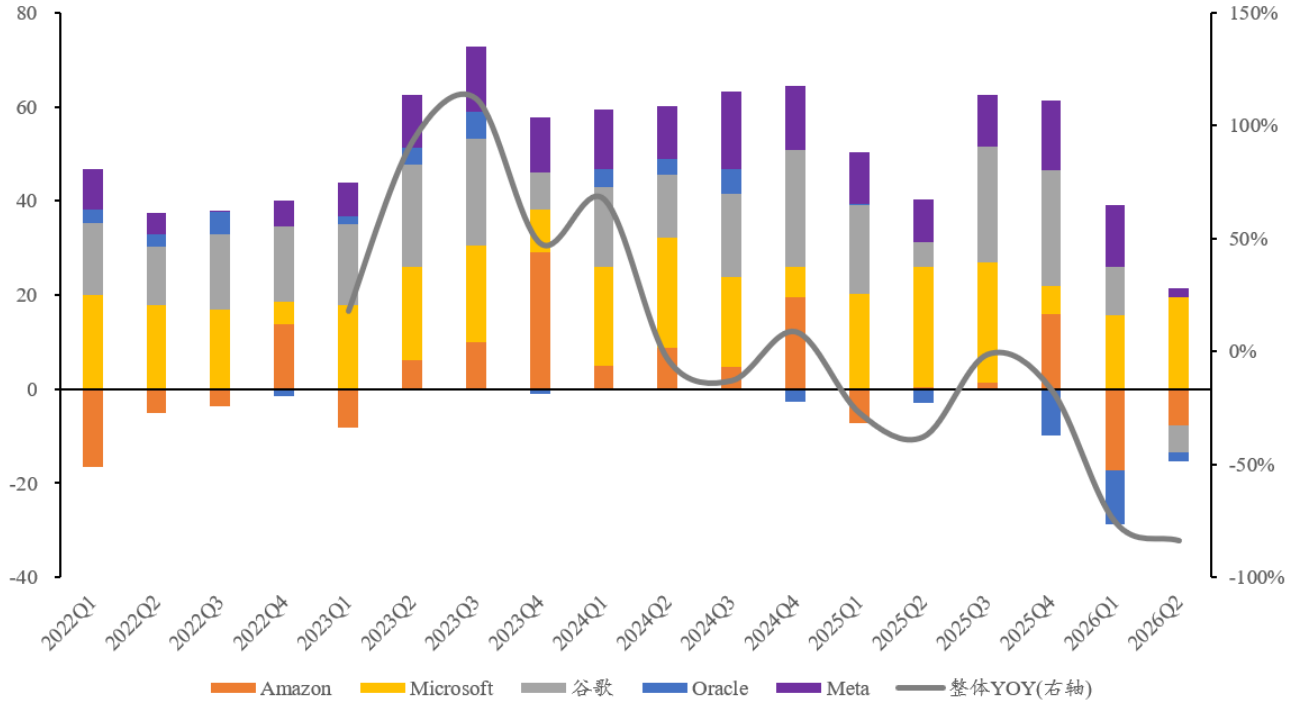
仅供内部参考，请勿外传

图 19：AI 融资估值提升倍数再度上行，平均值与中位数分化扩大



资料来源：NVCA，天风证券研究所

图 20：资本开支加速压缩内部现金余量，三家公司单季转负（单位：十亿美元）



资料来源：Wind，天风证券研究所

6. 资本效率：ROIC 整体保持韧性，资金成本约束仍需观察

资本效率是检验 AI 投资扩张可持续性的关键关口。资金保障决定企业能否维持高强度投入，资本效率则反映公司整体经营回报能否覆盖资金成本，并为后续资本开支扩张提供回报基础。从分析框架看，ROIC-WACC 可以用于观测公司整体资本回报对资金成本的覆盖程度。从趋势判断上，美国 10 年期国债收益率和美国隐含股权风险溢价分别代表无风险利率与市场风险补偿，其变化能够反映 WACC 市场端核心驱动的方向。将上述指标与云厂商 ROIC 走势结合，可以对 ROIC-WACC 安全边际进行定性观察：若 ROIC 走弱且两项资金成本变量上行，资本回报安全边际趋于收窄；若 ROIC 改善或资金成本变量回落，安全边际则有望扩大。

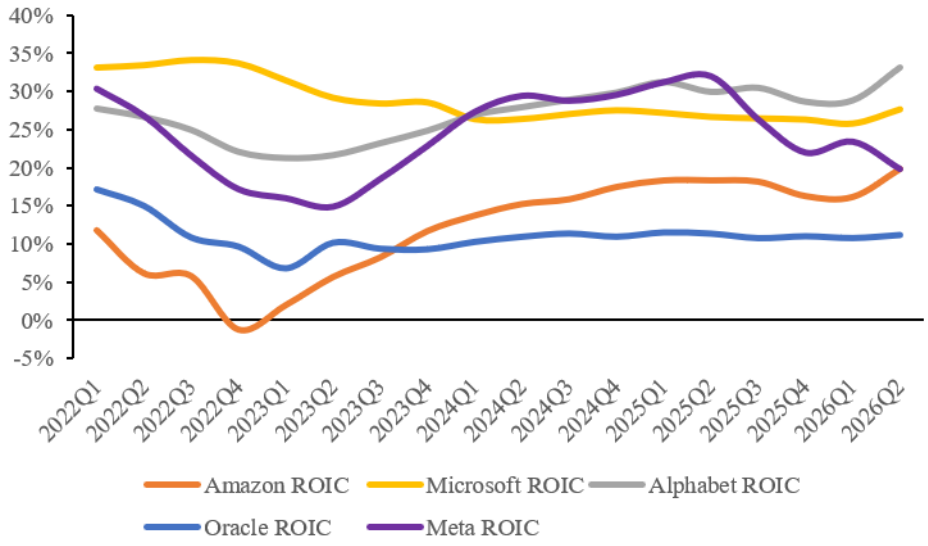
云厂商 ROIC 整体保持韧性，但不同厂商的资本回报走势有所分化。Microsoft 和 Alphabet 的 ROIC 总体处于较高水平，Alphabet 在前期回落后重新改善；Amazon 从早期低位逐步

仅供内部参考，请勿外传

修复，盈利和资本回报能力持续恢复；Oracle 整体波动相对有限，Meta 则在前期快速回升后有所回落。五家云厂商 ROIC 尚未出现同步下行，但不同厂商在业务结构、盈利能力和资本开支阶段上的差异，正在转化为资本回报表现的分化。

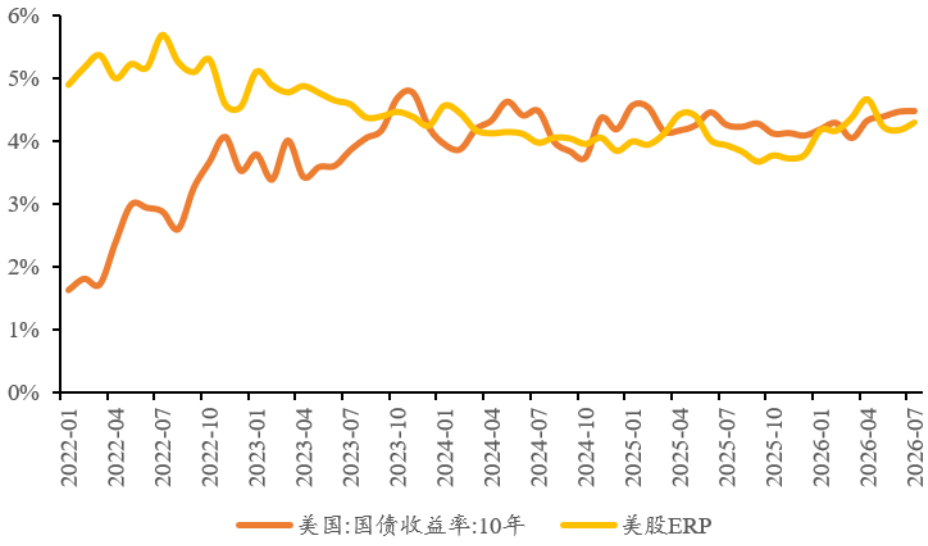
无风险利率上行与隐含股权风险溢价回落并存，资金成本约束仍需持续观察。美国 10 年期国债收益率自低位明显上升并维持高位波动，美国隐含股权风险溢价则总体回落、近期有所企稳，两项指标对权益资本成本形成方向不同的影响。隐含 ERP 下降能够部分缓解风险补偿压力，但无风险利率维持高位意味着企业仍需保持较强的资本回报，才能形成充足的回报安全边际。后续若云厂商 ROIC 普遍回落，同时美债利率或隐含 ERP 上升，资本效率约束将趋于增强；反之，ROIC 保持稳定或改善、资金成本基础变量趋于缓和，则有利于投资周期延续。

图 21：五大云厂商 ROIC 整体保持韧性，厂商间走势分化



资料来源：Wind，天风证券研究所

图 22：美债利率上行、隐含 ERP 中枢回落，资金成本约束仍存



资料来源：Wind，NYU Stern，天风证券研究所

仅供内部参考，请勿外传

7. 风险提示

1）产业发展进程不及预期：人工智能技术迭代期，若算法突破速度放缓或硬件配套供应受限，可能导致产业发展降速。

2）政策出台和落地具备不确定性：人工智能相关政策的出台和已出台政策的落地情况具备不确定性，若政策支持力度下降可能导致相关投资放缓。

3）宏观经济波动：全球利率环境超预期变化或影响科技板块估值。

4）模型价格为手工统计，仅供参考：本文所引用的大模型价格数据主要通过第三方平台和行业资料手工收集整理。不同模型在版本、输入与输出 Token 计费、上下文缓存、批量调用、服务等级、地区及优惠政策等方面存在差异，且厂商可能随时调整定价，因此相关数据可能存在遗漏、滞后或统计误差。

仅供内部参考，请勿外传

分析师声明

本报告署名分析师在此声明：我们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，本报告所表述的所有观点均准确地反映了我们对标的证券和发行人的个人看法。我们所得报酬的任何部分不曾与，不与，也将不会与本报告中的具体投资建议或观点有直接或间接联系。

一般声明

除非另有规定，本报告中的所有材料版权均属天风证券股份有限公司（已获中国证监会许可的证券投资咨询业务资格）及其附属机构（以下统称“天风证券”）。未经天风证券事先书面授权，不得以任何方式修改、发送或者复制本报告及其所包含的材料、内容。所有本报告中使用的商标、服务标识及标记均为天风证券的商标、服务标识及标记。
本报告是机密的，仅供我们的客户使用，天风证券不因收件人收到本报告而视其为天风证券的客户。本报告中的信息均来源于我们认为可靠的已公开资料，但天风证券对这些信息的准确性及完整性不作任何保证。本报告中的信息、意见等均仅供客户参考，不构成所述证券买卖的出价或征价邀请或要约。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。客户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，必要时就法律、商业、财务、税收等方面咨询专家的意见。对依据或者使用本报告所造成的一切后果，天风证券及/或其关联人员均不承担任何法律责任。
本报告所载的意见、评估及预测仅为本报告出具日的观点和判断。该等意见、评估及预测无需通知即可随时更改。过往的表现亦不应作为日后表现的预示和担保。在不同时期，天风证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。
天风证券的销售人员、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。天风证券没有将此意见及建议向报告所有接收者进行更新的义务。天风证券的资产管理部门、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。

特别声明

在法律许可的情况下，天风证券可能会持有本报告中提及公司所发行的证券并进行交易，也可能为这些公司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。因此，投资者应当考虑到天风证券及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突，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	说明	评级	体系
股票投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	买入	预期股价相对收益 20%以上
		增持	预期股价相对收益 10%-20%
		持有	预期股价相对收益-10%-10%
		卖出	预期股价相对收益-10%以下
行业投资评级	自报告日后的 6 个月内，相对同期沪深 300 指数的涨跌幅	强于大市	预期行业指数涨幅 5%以上
		中性	预期行业指数涨幅-5%-5%
		弱于大市	预期行业指数涨幅-5%以下

天风证券研究

北京	海口	上海	深圳
北京市西城区德胜国际中心 B 座 11 层	海南省海口市美兰区国兴大道 3 号互联网金融大厦 A 栋 23 层 2301 房	上海市虹口区东长治路 855 号北外滩金茂中心 T1 栋 19 楼	深圳市福田区益田路 5033 号平安金融中心 71 楼
邮编：100088	邮编：570102	邮编：200082	邮编：518000
邮箱：research@tfzq.com	电话：(0898)-65365390	电话：(8621)-60256006	电话：(86755)-23915663
	邮箱：research@tfzq.com	邮箱：research@tfzq.com	传真：(86755)-82571995
			邮箱：research@tfzq.com