

AI专题·从FOMO CapEx到ROI CapEx

AI模型迭代聚焦工程能力，AI应用落地锚定高ROI场景

西南证券研究院
海外研究团队
2026年1月

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

核心观点

- ❑ **海外AI投入面临现金流压力，AI投资从FOMO CapEx转向ROI CapEx。** 2024-2025年，海外科技大厂资本开支高增，AI初创企业投入力度加大，且未来开支预期进一步上调，当前行业普遍面临现金流压力，从而促使海外科技厂商寻求多种数据中心建设方式和融资手段缓解压力，AI独角兽IPO进程也有望提速。2024年，AI大基建之初，部分海外云厂商表示“投资不足的风险远远大于投资过度的风险”，AI投资伴随着FOMO情绪；2025年，海外云厂商均强调“云服务供不应求”、“根据需求信号扩展数据中心”、同时愈发关注AI应用的商业化变现，AI投资逐步从FOMO CapEx向ROI CapEx转变。
- ❑ **数据中心面临电力容量限制，算力集群日益强调每瓦特Tokens产出效率。** 通常，海外云厂商在规划数据中心之初，需率先确定能够提供给数据中心的电力容量，再针对数据中心内部的IT设备进行配置。因此，在电力容量限制下，海外云厂商均强调最大化每瓦特下的Tokens产出效率，对芯片/存储/通信等硬件环节，以及软件栈和系统架构进行全面优化，以提高数据中心的算力利用率。同时，海外云厂商日益注重数据中心建设的通用性和灵活性，以适应不同代际的GPU、电气部件和冷却设备，支持多种模型的计算，能够在训练工作和推理负载之间灵活切换，而非为单一技术或单一客户过度投资，协同设计(Co-Design)和通用性(Fungible)成为关键。
- ❑ **大模型工程化能力持续提升，AI产品商业化诉求增强。** 未来，AI大模型将继续围绕长文本/多模态/逻辑推理/工具使用能力迭代升级，当前Transformer架构仍有较大工程优化空间。随着Scaling Law从预训练延伸至强化学习和持续学习，模型对数据集的要求也将发生转变，训练数据将带来模型能力分化，推理阶段有望进一步放大其差异化特征，从而催生不同的AI用例、与不同领域进行深层次融合，在工程化能力提升和锚定高ROI应用场景下，AI商业化进程有望提速，订阅/API/广告/Agent等多种业务模式逐步探索、成形。
- ❑ **云业务增速依赖产能上线节奏，AI云服务有望伴随容量释放实现加速增长。** 2025年，海外AI云服务进入“超大订单+长期基建”模式，云订单金额从十亿级到千亿级不等，同时海外云厂商剩余履约合同金额呈现高速增长，AI云服务有望迈向加速扩张阶段。从海外云厂商表态来看，各公司均表示当前算力产能供不应求，已锁定数吉瓦级别电力容量，由于云业务收入增速将高度依赖产能上线节奏，预计随着2026年算力容量逐步释放，云业务有望迎来加速增长。
- ❑ **相关标的：**1) 芯片：英伟达(NVDA.O)、博通(AVGO.O)；2) 存储：美光(MU.O)、闪迪(SNDK.O)；3) 光通信：Lumentum(LITE.O)、Credo Technology (CRDO.O)；4) 互联网大厂：谷歌(GOOG.L)、Meta(META.O)；5) 云厂商：亚马逊(AMZN.O)、谷歌(GOOG.L)、微软(MSFT.O)。
- ❑ **风险提示：**AI投资带来现金流压力风险；AI云业务增长不及预期风险；AI应用商业化变现不及预期风险等。

目录

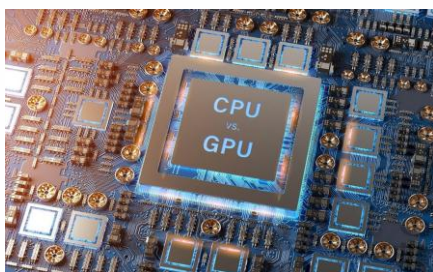
- ◆ **一、资本开支节奏：资本开支预期上调，未来现金流压力加大**
- ◆ **二、数据中心建设：电力容量面临限制，追求每瓦特产出效率**
- ◆ **三、模型能力演进：研究端产品端并进，工程化能力继续增强**
- ◆ **四、AI云业务增速：算力容量加速释放，AI云服务迈向扩张期**

1.1 AI基建→资本开支：AI基建范式转移，资本开支扩张持续

- ❑ **AI范式由“基础设施替代+工作负载升级”转向“AI从推理模型走向智能体”**。1) **计算范式迁移**：过去，数据中心的绝大多数工作负载跑在CPU上，而现在已经向GPU加速计算逐步迁移，部分基础设施已经在加速计算的升级中被逐步替换、扩容。2) **工作负载升级**：在数据中心升级换代背景下，海外科技厂商正在重构核心应用，将搜索、推荐引擎、社交媒体等工作负载迁移至AI原生架构，例如搜索从SEO向GEO迭代升级，该部分需求已逐步构成GPU算力需求的基本盘。3) **AI从“推理模型”走向“智能体”**：当前，AI大模型逐步从简单对话发展至强大的推理模型，未来有望向智能体持续演进，同时，AI产品的商业化进程加速推进，订阅、API、广告等多种业务模式正在成形。而AI基建之初，部分海外云厂商表示“投资不足的风险远远大于投资过度的风险”；如今，海外云厂商均强调“根据需求信号来扩展数据中心”、“云服务供不应求”、“追求每瓦特产出效率”，表明当前数据中心投资建设正在从FOMO CapEx向ROI CapEx转变。

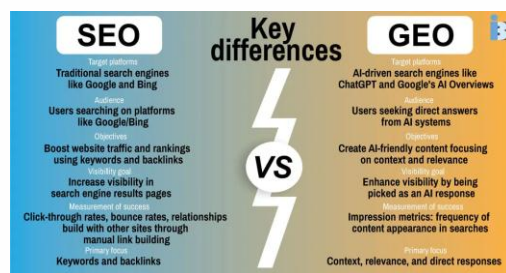
海外AI基建的三重范式转移

计算范式迁移



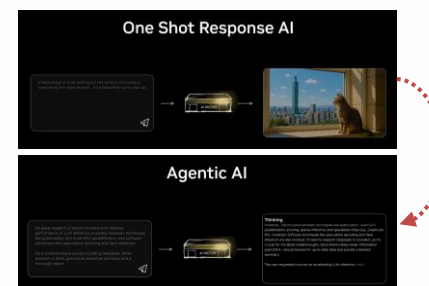
- 当前一部分数据中心基础设施的替换和扩容将用于完成从CPU向GPU的过渡，而非数据中心的简单新建。

工作负载升级



- 搜索、推荐引擎、社交媒体等，正加速迁移至GPU与AI原生架构上，当前这一部分需求已逐步成为GPU需求的“基本盘”。

从“推理模型”走向“智能体”



- 模型从“简单生成一个答复”，走向“长链条推理+自主规划+工具调用”智能体形态，对训练和推理算力形成持续拉动

1.1 AI基建→资本开支：AI基建范式转移，资本开支扩张持续

□ 海外AI云订单签约不断，算力集群持续扩张。2025年，海外科技公司AI云相关合同订单签约明显提速，全球前沿模型的训练与推理需求逐步进入工程化、规模化部署阶段，海外头部AI实验室与云厂商密集开展合作，签订长期算力合作订单，加速锁定IT电力资源，整体合同金额、期限及所对应的算力体量均有抬升。随着算力需求的快速释放，海外数据中心建设进入新一轮扩张周期，2026年吉瓦级算力集群有望成为xAI、Meta、OpenAI、Anthropic等头部AI实验室的基础设施标配。与此同时，数据中心的规模化建设带来资本开支的高强度投入，市场关注点除资本开支外，也逐步将目光转向AI模型应用的落地和商业化变现，构筑AI未来叙事的关键变量。

海外AI实验室计划截至2026年底上线的约1GW算力集群

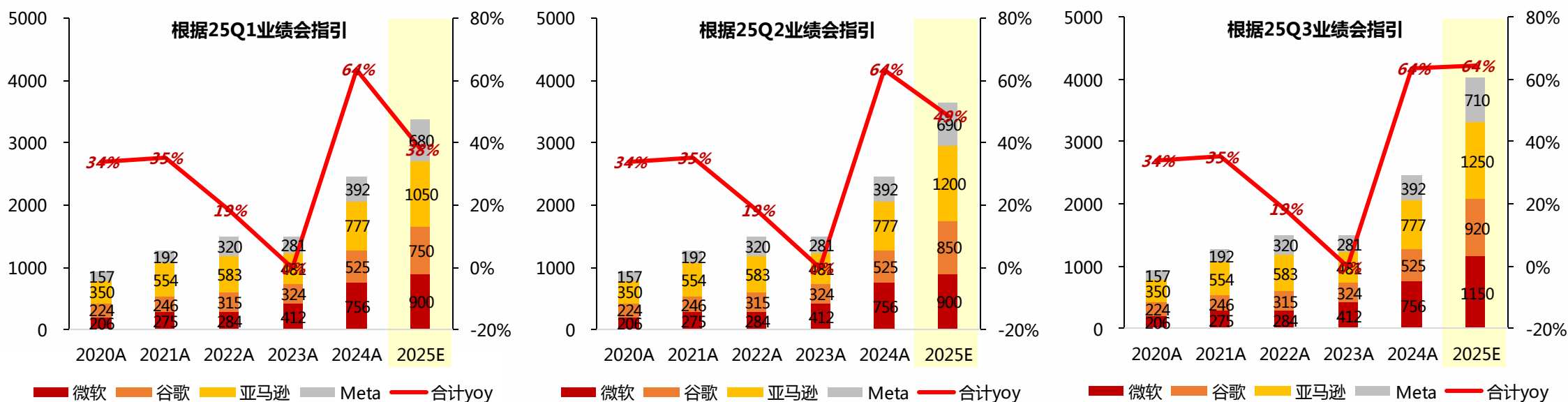
集群名称	Colossus 2	Prometheus	Stargate	Project Rainier
集群需求方与提供商	xAI	Meta	OpenAI&Oracle	Anthropic&AWS
芯片类型	GB200/300	GB200/300	GB200/300	Trainium 2
计划截至2026年IT电力容量 (MW)	1200	1020	880	780
芯片数量 (万颗)	55	50	40	80
每秒浮点运算次数 (EFLOPS)	3,488	3,171	2,470	1,040
图示				

资料来源：SemiAnalysis，State of AI，亚马逊官网，西南证券整理

1.1 AI基建→资本开支：AI基建范式转移，资本开支扩张持续

- ❑ **资本开支2025年指引：多家上调全年资本开支指引，同比增速有望突破60%。** 1) **微软**：资本开支FY2026增长率将高于FY2025；2) **谷歌**：将全年资本开支指引提高至910亿~930亿美元之间；3) **亚马逊**：25Q3业绩会给出2025全年现金资本支出约1250亿美元的指引；4) **Meta**：第三次上调2025年全年资本开支指引，预计全年资本开支（含融资租赁）将在700亿~720亿美元之间。
- ❑ **资本开支2026年指引：投资预期提升，资本开支表态相较前两个季度更加积极。** 1) **微软**：25Q1及25Q2业绩会上表示FY2026资本开支的增长率将低于FY2025，而25Q3业绩会上表示FY2026增长率将高于FY2025；2) **谷歌**：预计2026年资本开支将“显著上升”，此前表述为“继续增加”；3) **亚马逊**：预计2026年资本开支将继续增加；4) **Meta**：预计2026年资本支出增长将显著高于2025年。

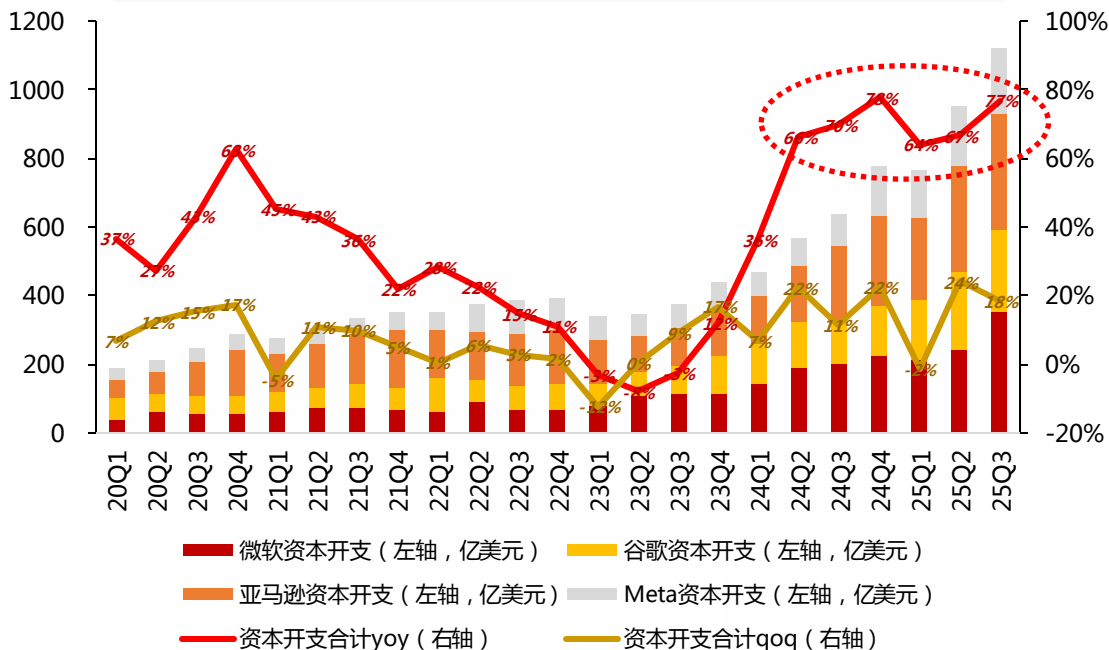
2020-2024年海外四大科技厂商资本开支情况及2025年资本开支预期变化（亿美元）



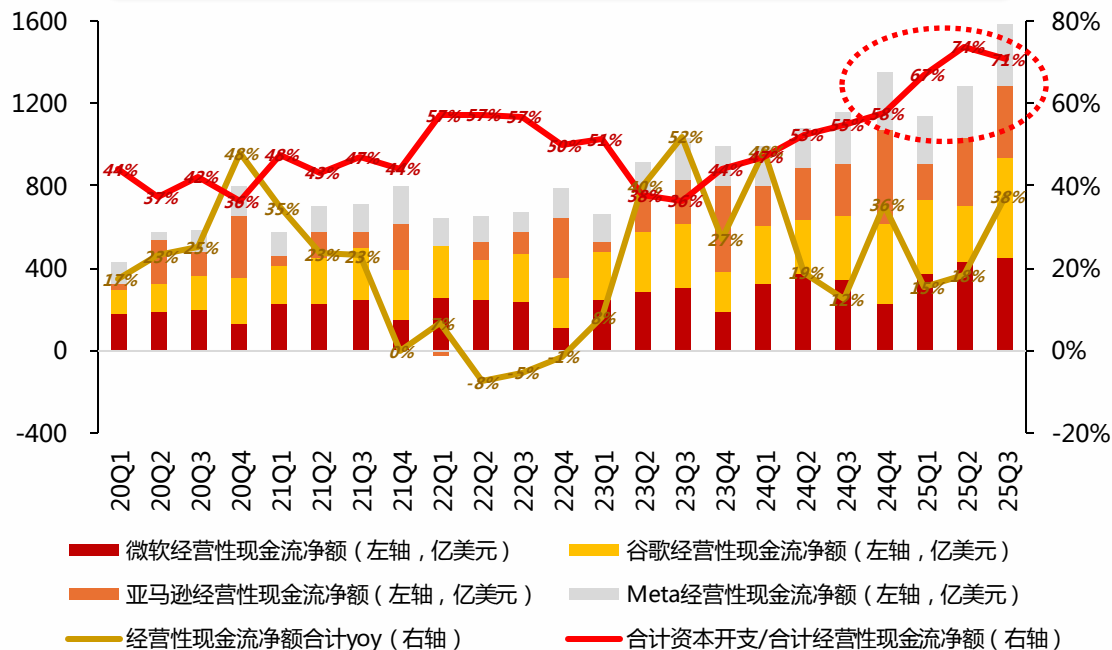
1.2 资本开支→现金流量表：资本开支高速增长，自由现金流压力加大

- **海外四大科技公司：资本开支占经营性现金流净额比例位于高位，业务强劲造血能力强劲支撑自由现金流为正。** 24Q2至25Q3，海外四大科技公司合计资本开支同比增速连续六个季度位于60%以上，超过2020至2023年历史最高同比增速；与此同时，资本开支占经营性现金流净额比例呈现提升态势，24Q4至25Q3，海外四大科技公司合计资本开支占经营性现金流净额的比例连续四个季度超过2020年以来的历史最高水平，各公司自由现金流压力持续加大。但由于海外科技大厂核心业务造血能力强劲，各公司经营性现金流较为充沛，在一定程度上能够支撑高强度投资，在资本开支高速扩张下仍然能够保持自由现金流为正。

20Q1-25Q3海外四大科技公司资本开支及增速



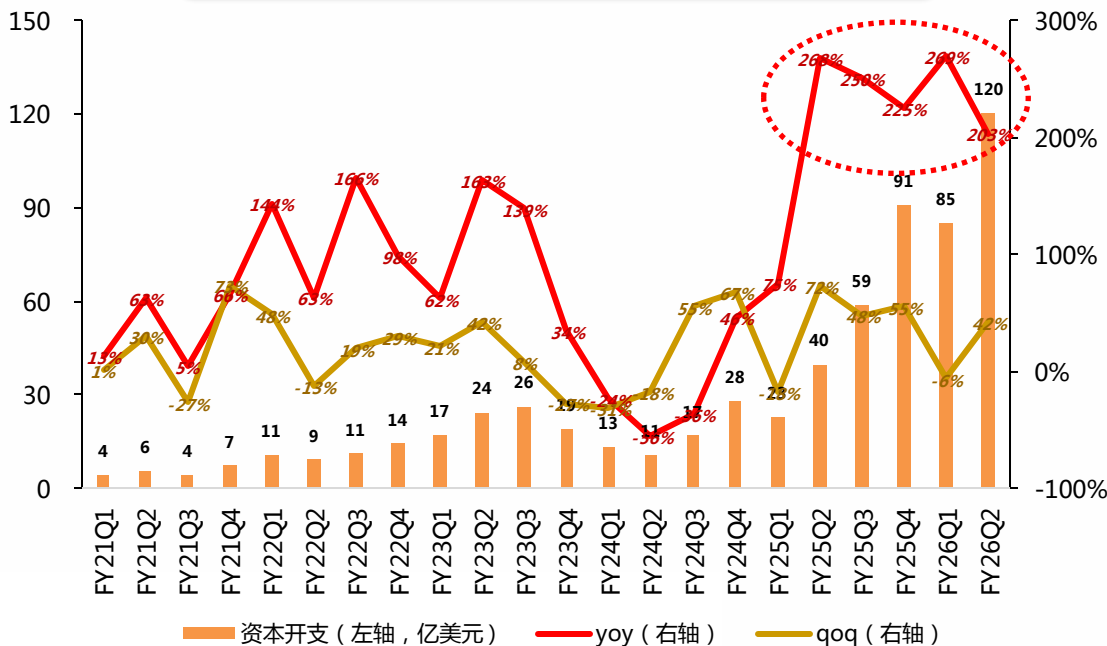
20Q1-25Q3海外四大科技公司经营性现金流净额



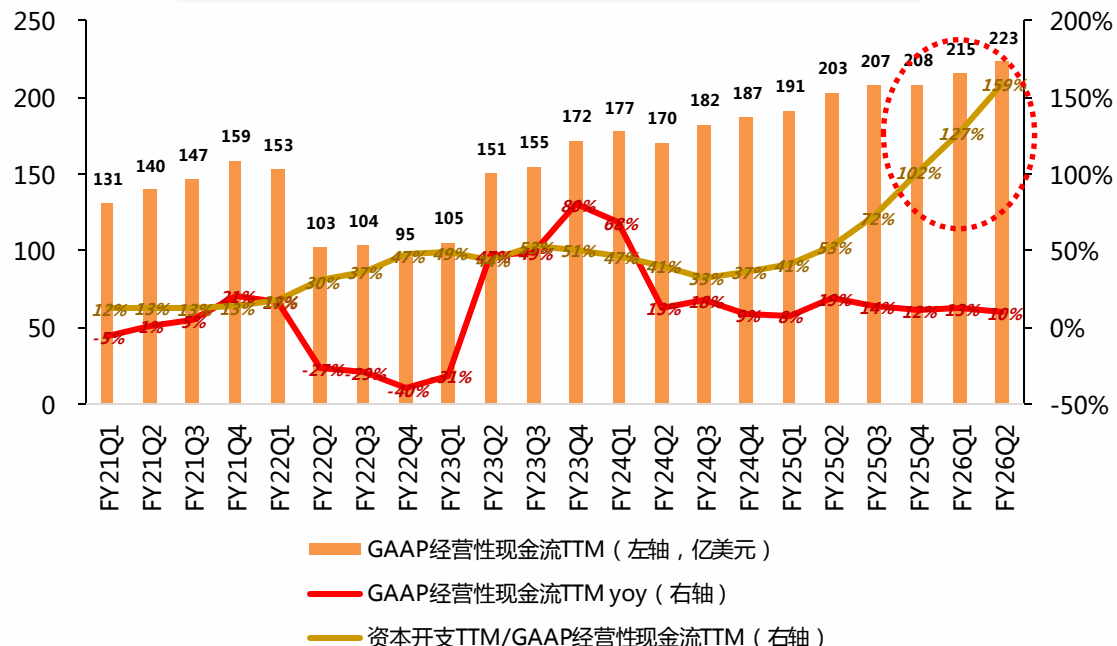
1.2 资本开支→现金流量表：资本开支高速增长，自由现金流压力加大

□ **甲骨文：资本开支同比翻倍增长，自由现金流(TTM)连续三季度为负。** FY25Q2至FY26Q2，甲骨文资本开支连续五个季度呈现2~3倍同比增长，并于FY26Q2业绩会将FY2026资本支出指引从350亿美元上调至500亿美元，而维持全年总收入670亿美元指引不变，收入增速预期暂时难以匹配未来资本开支节奏，市场对公司现金流的担忧进一步加大。FY25Q4至FY26Q2，甲骨文资本开支(TTM)占GAAP经营性现金流(TTM)的比例已连续三个季度超过100%，从而导致公司自由现金流由正转负，FY25Q4至FY26Q2公司自由现金流(TTM)分别为-4亿/-59亿/-132亿美元。在增长资金来源方面，公司于FY26Q2业绩会表示，可通过公共债券、银行融资及私募债务市场等多种渠道获取资金，或通过客户自带芯片安装至数据中心、供应商以租赁方式提供产能等方式缓解现金流压力。

FY21Q1-FY26Q2甲骨文资本开支及增速



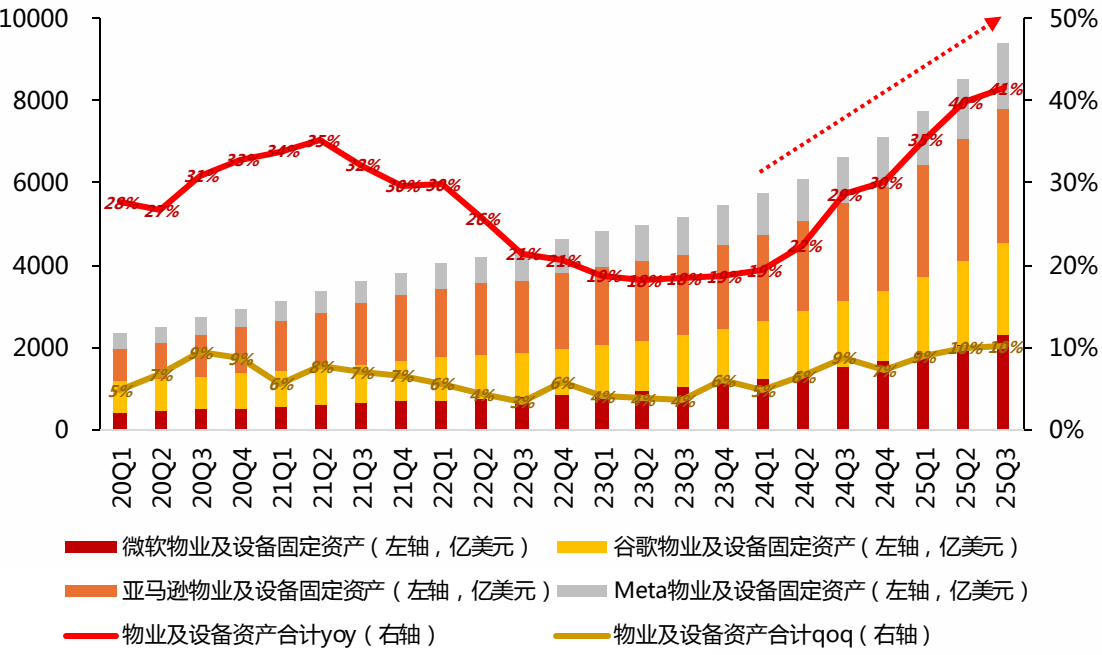
FY21Q1-FY26Q2甲骨文经营性现金流



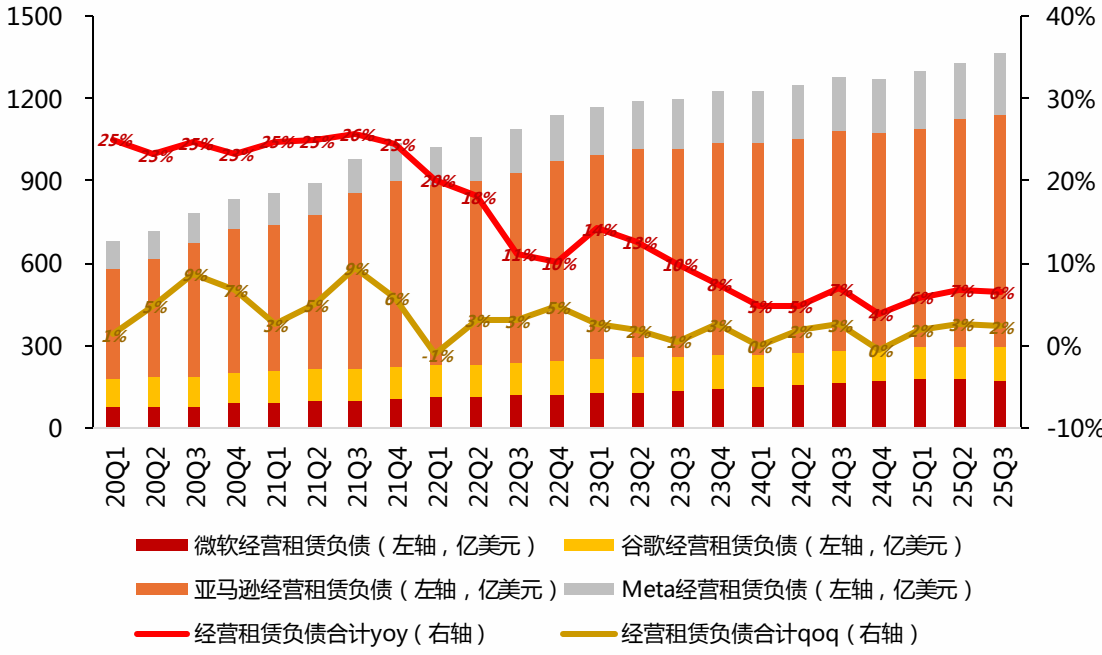
1.3 资本开支→资产负债表：固定资产规模持续攀升，经营租赁负债小幅提升

□ **海外四大科技公司：固定资产规模稳步扩张，经营租赁负债小幅增长。**随着资本开支和数据中心建设持续扩张，24Q1至25Q3，海外四大科技公司的物业及设备固定资产合计规模从5734亿美元增长至9394亿美元，合计同比增速持续攀升；经营租赁资产合计规模从1180亿美元增长至1401亿美元，经营租赁负债合计规模从1227亿美元提升至1364亿美元，同比保持小幅增长。截至25Q3，海外科技大厂的物业及设备固定资产增速相较于经营租赁资产或负债的增速更快，经营租赁资产与物业及设备固定资产的比值约为1:7，表明海外科技大厂近两年主要通过自建模式扩展数据中心，经营性租赁占比相对较小。

20Q1-25Q3海外四大科技公司物业设备资产及增速



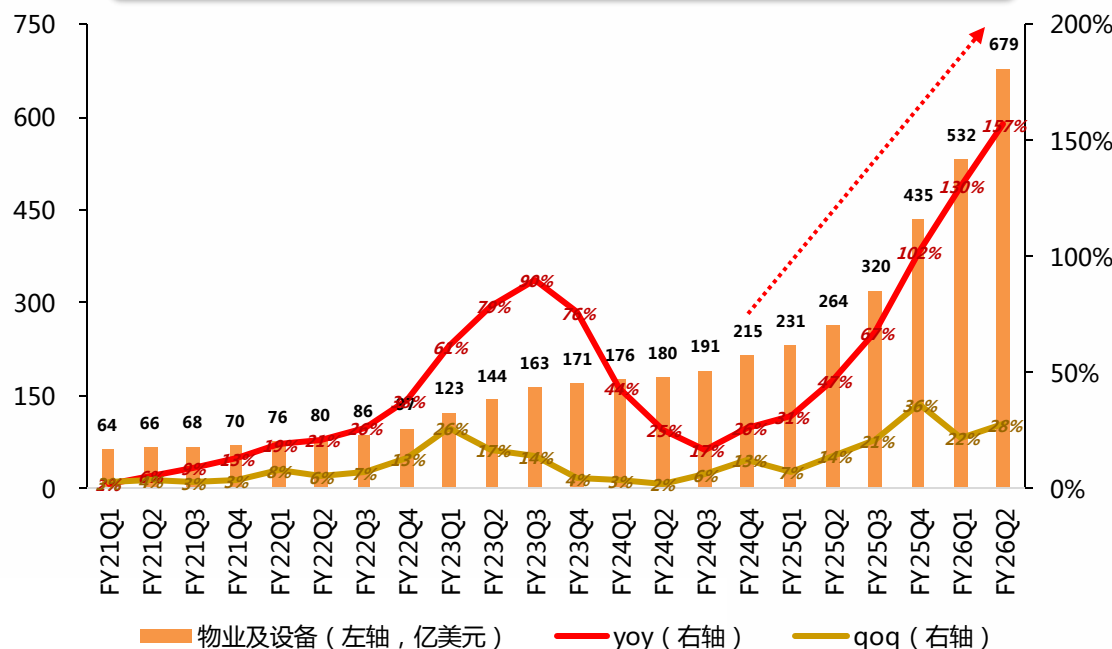
20Q1-25Q3海外四大科技公司经营租赁负债及增速



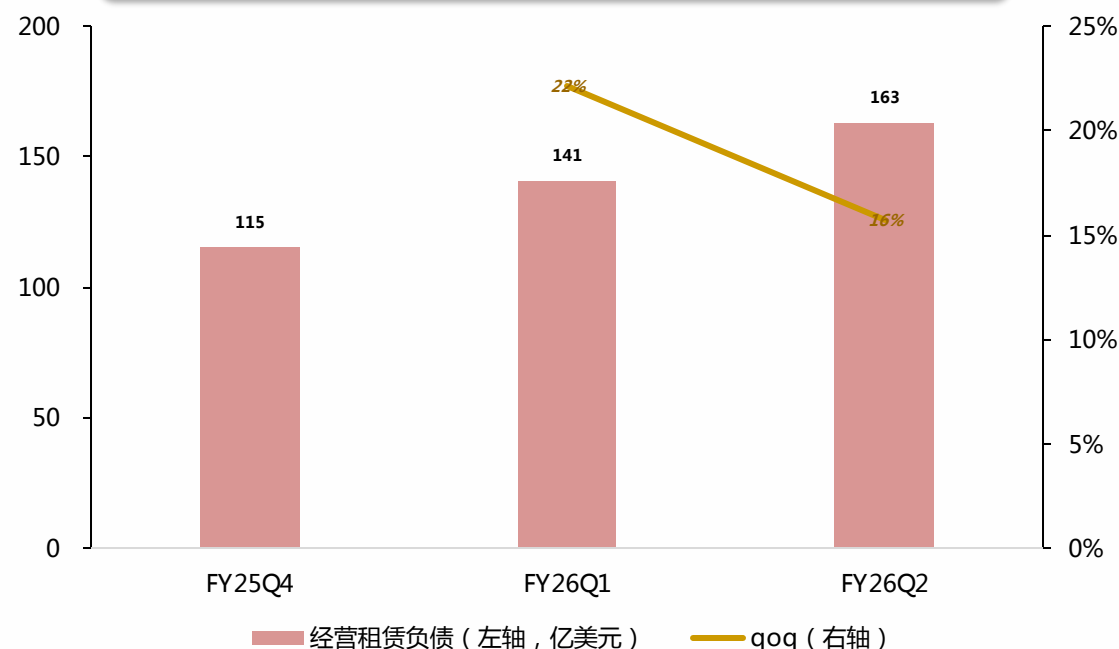
1.3 资本开支→资产负债表：固定资产规模持续攀升，经营租赁负债小幅提升

- **甲骨文：固定资产同比增速连续七个季度持续提升，经营租赁负债已连续披露三个季度。** FY24Q4至FY26Q2，甲骨文物业及设备固定资产合计规模从215亿美元增长至679亿美元，同比增速从26%攀升至157%；同时，甲骨文自FY25Q4开始，已连续三个季度披露经营租赁负债，其经营租赁负债规模从FY25Q4的115亿美元增长至FY26Q2的163亿美元。相较于海外四大科技厂商，甲骨文更快的资本开支节奏带来更快的物业及设备固定资产规模扩张，经营租赁负债环比增速相较海外科技厂商也略有提高，表明甲骨文在投入硬件资产建设数据中心的同时还通过经营租赁负债同步推进。

FY21Q1-FY26Q2甲骨文物业设备资产及增速



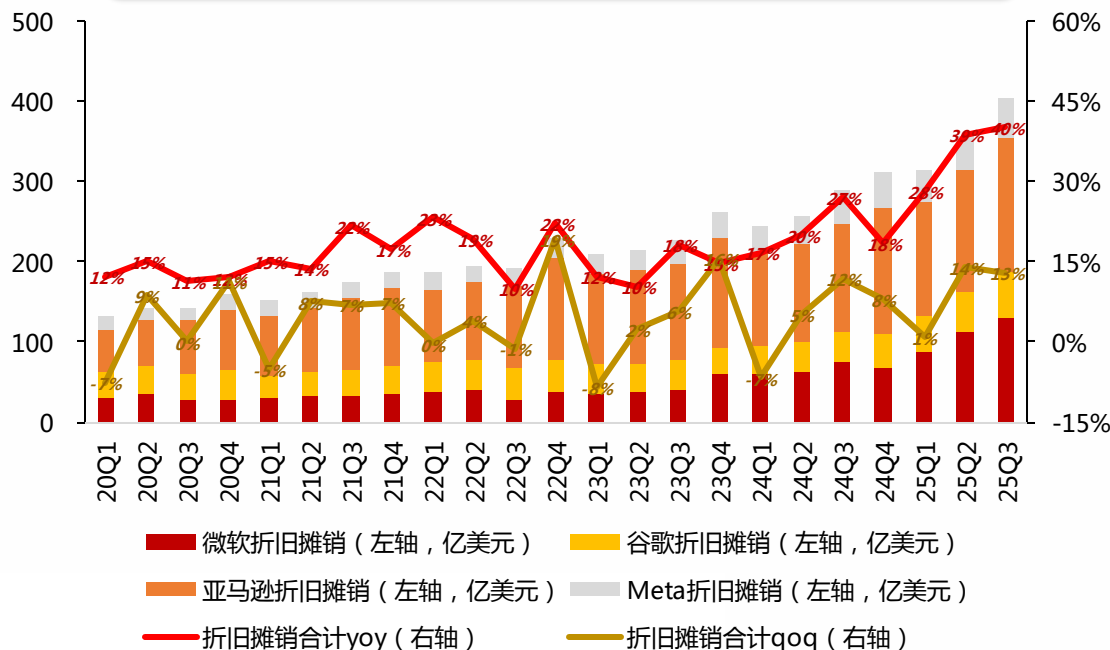
FY25Q4-FY26Q2甲骨文经营租赁负债及增速



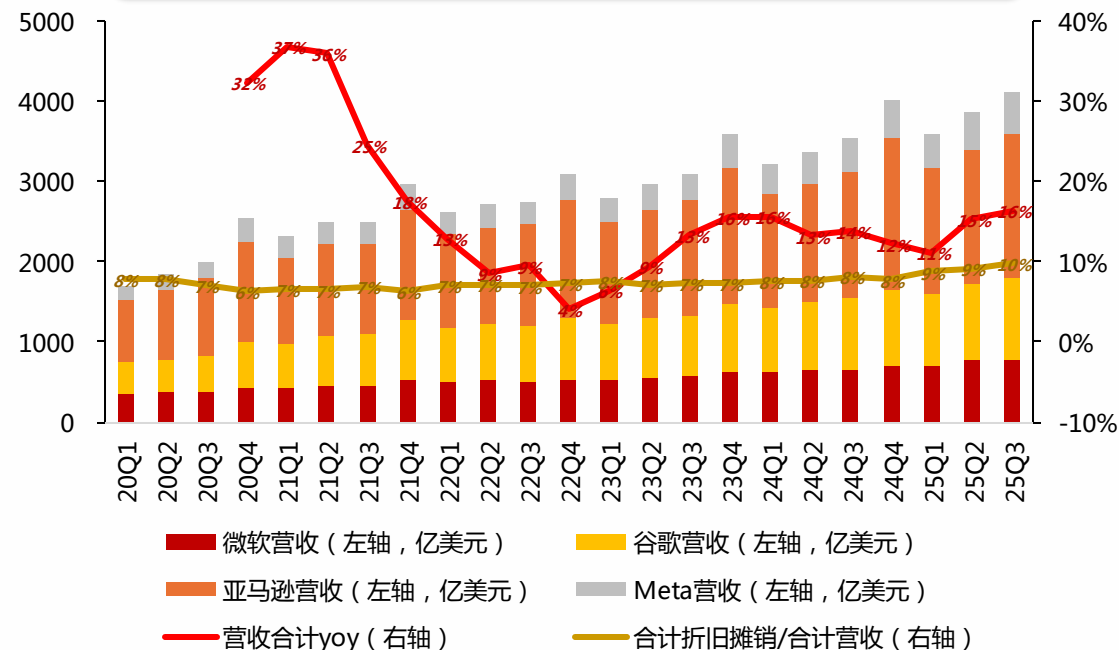
1.4 资本开支→利润表：折旧摊销增长加速，利润端压力日益显现

□ **海外四大科技公司：折旧摊销金额增长加速，其占收入比例略有抬升。** 25Q3海外四大科技公司合计折旧摊销金额约404亿美元，yoy+40%，qoq+13%，合计折旧摊销金额占合计收入的比例约10%，该比例为20Q1以来单季度最高水平，表明基础设施建设带来的折旧成本压力日益显现。其中，谷歌预计25Q4折旧增速有望进一步加快；亚马逊认为随着数据中心资产的投入使用，相关折旧会对利润率产生一定影响；Meta预计2026年总支出同比增长更快，部分将由基础设施折旧推动。由于折旧摊销对利润端的压制有滞后效应，因此，随着折旧摊销加速，未来海外科技大厂在利润端的压力将进一步显现。

20Q1-25Q3海外四大科技公司折旧摊销及增速



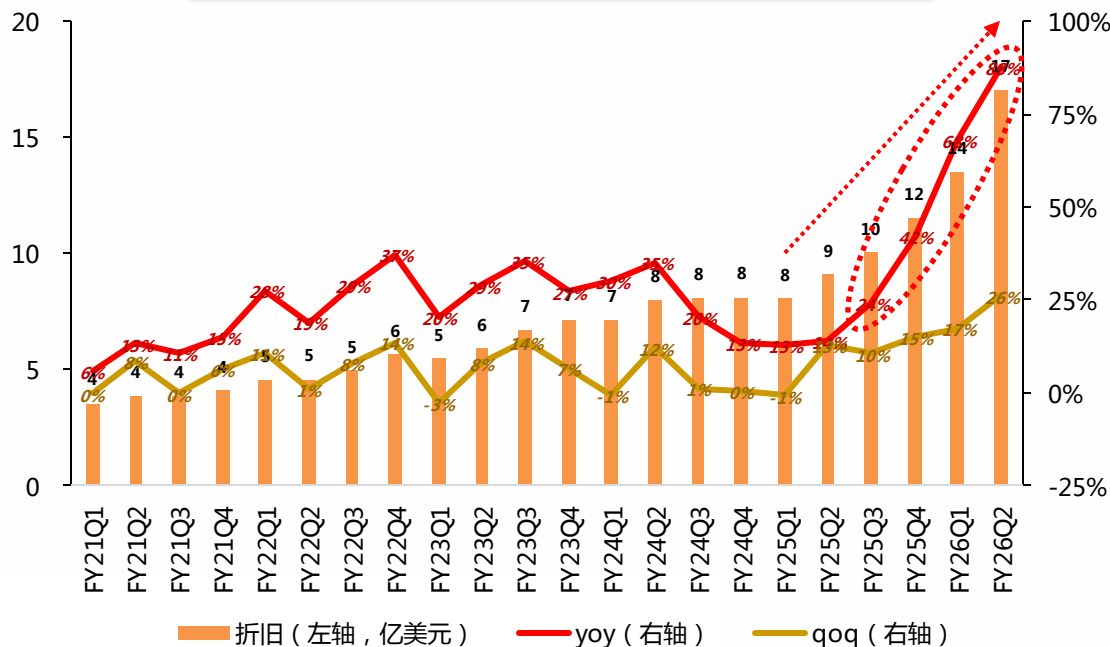
20Q1-25Q3海外四大科技公司折旧摊销占收入比例



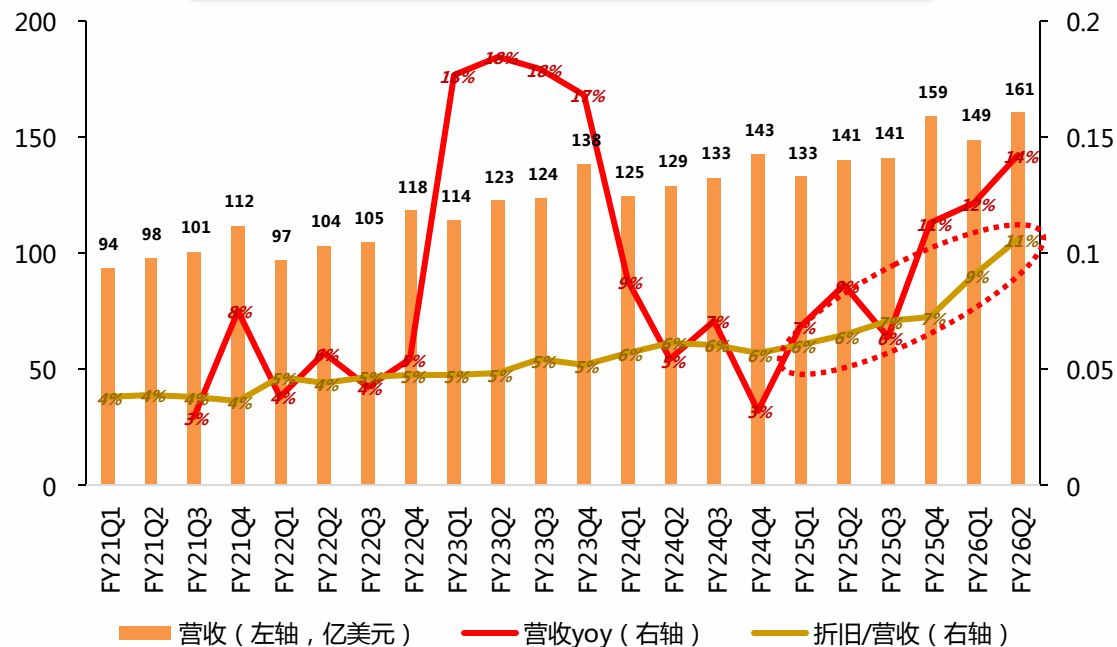
1.4 资本开支→利润表：折旧摊销增长加速，利润端压力日益显现

- **甲骨文：资本开支同比增速领先折旧约三个季度，折旧占收入比例连续三个季度明显提升。** FY25Q1至FY26Q2，甲骨文折旧金额从8亿美元提升至17亿美元，资本开支同比增速大约领先折旧同比增速3个季度（FY24Q4资本开支出现同比大幅提升，对应折旧摊销同比增速从FY25Q3开始呈现明显上行）。FY25Q1至FY26Q2，公司折旧金额占收入的比例从6%提升至11%，提升幅度相较海外科技厂商更加明显，对利润端的挤压将进一步凸显。

FY21Q1-FY26Q2甲骨文折旧及增速



FY21Q1-FY26Q2甲骨文折旧占收入比例



1.5 资本开支→融资需求：AI投入带来现金压力，科技企业融资需求加剧

□ **海外科技厂商：逐步从现金投资转向债务融资，未来现金流能力将成为关键。** 2024年至2025年，在AI数据中心建设过程中，大部分资本投资主要通过企业自身留存收益和经营性现金流实现，尽管大多数科技企业现金流仍然充沛，但近两年的巨额资本开支对现金流的压力持续加大，部分海外科技企业开始加快发债节奏，债务融资现象逐步显现。根据各公司公告及标普全球评级公告，2025年9月以来，谷歌/亚马逊/Meta/甲骨文/CoreWeave分别新发债务250亿/150亿/300亿/180亿/22.5亿美元，合计发行债务规模达900亿美元，表明海外科技公司逐步从现金投资转向债务融资。未来，海外科技企业的偿债能力和流动性变化将对企业发展产生重要影响，并将成为市场评估AI基建叙事能否延续的关键。

截至2025年底海外科技厂商最新发债情况

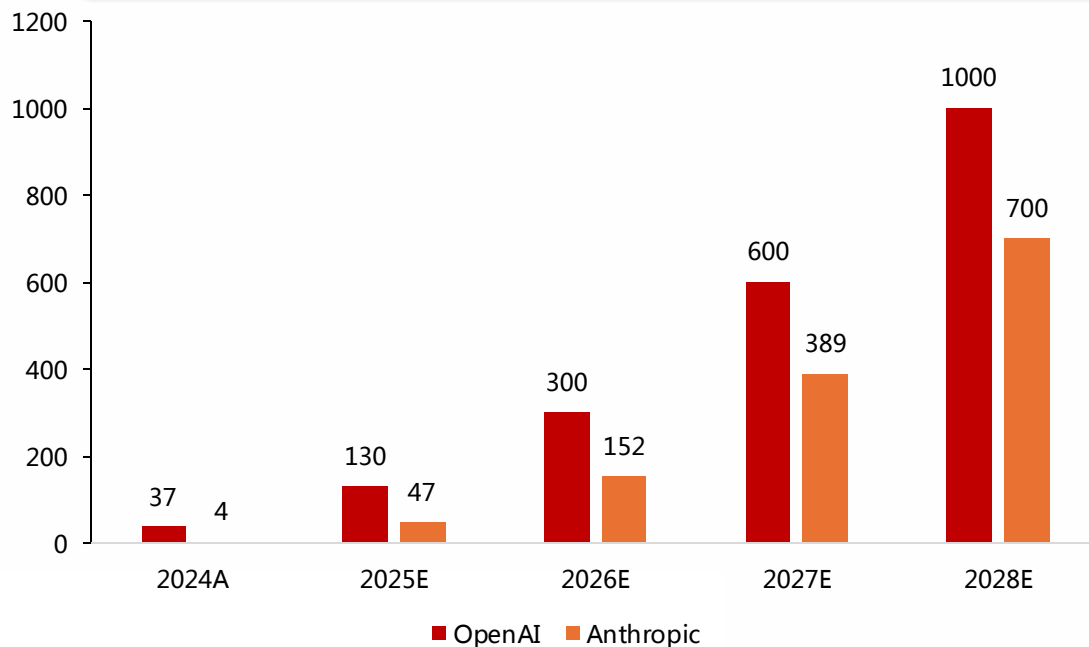
公司名称	发行人信用评级	发行人评级展望	最新债项宣布时间	债项评级	债项计划情况
微软	AAA	Stable/稳定	/	/	/
谷歌	AA+	Stable/稳定	2025年11	AA+	计划在美国筹集175亿美元债务，在欧洲筹集65亿欧元（约75亿美元）债务，用于一般企业用途。截至2025年9月30日，现金等价物及可交易证券为985亿美元，债务为220亿美元，公司还有能力将净债务增加至超过2000亿美元，而仍未触及设定的按标普全球评级口径调整后的1倍净杠杆下调阈值。
亚马逊	AA	Stable/稳定	2025年11月	AA	计划通过债券发行筹集150亿美元，用途包括但不限于偿还债务/并购/投资/营运资金/对子公司投资/资本性支出/股票回购。截至2025年9月30日，标普全球评级口径调整后的杠杆率为0.5倍；随着此次发债，S&P Global预计该指标将小幅上升，但距离对设定的1.5倍下调触发阈值仍有充足缓冲空间。
Meta	AA-	Stable/稳定	2025年10月	AA-	计划发行300亿美元债券，发行所得资金用于进一步巩固其卓越的流动性状况，并用于一般公司用途，包括AI基础设施扩展等。S&P Global预计Meta按标普全球评级口径调整后的杠杆率仍将显著低于当前评级1.0倍的下调触发阈值。
甲骨文	BBB	Negative/负面	2025年9月	BBB	计划通过债券发行筹集约180亿美元，用于支持AI基础设施建设。S&P Global将评级下调触发阈值从3.5倍放宽至4倍，以表明对甲骨文业务转型采取更长期的视角，即在收获收入之前需要前期投入。
Coreweave	B+	Stable/稳定	2025年12月	B	计划发行22.5亿美元可转换债券，用于支持新数据中心建设。S&P Global预计公司到2027年，其经营资金与债务比率将超过12%，经营现金流与债务比率将达到10%~15%，这一水平适合当前评级。

资料来源：S&P Global，各公司公告，西南证券整理

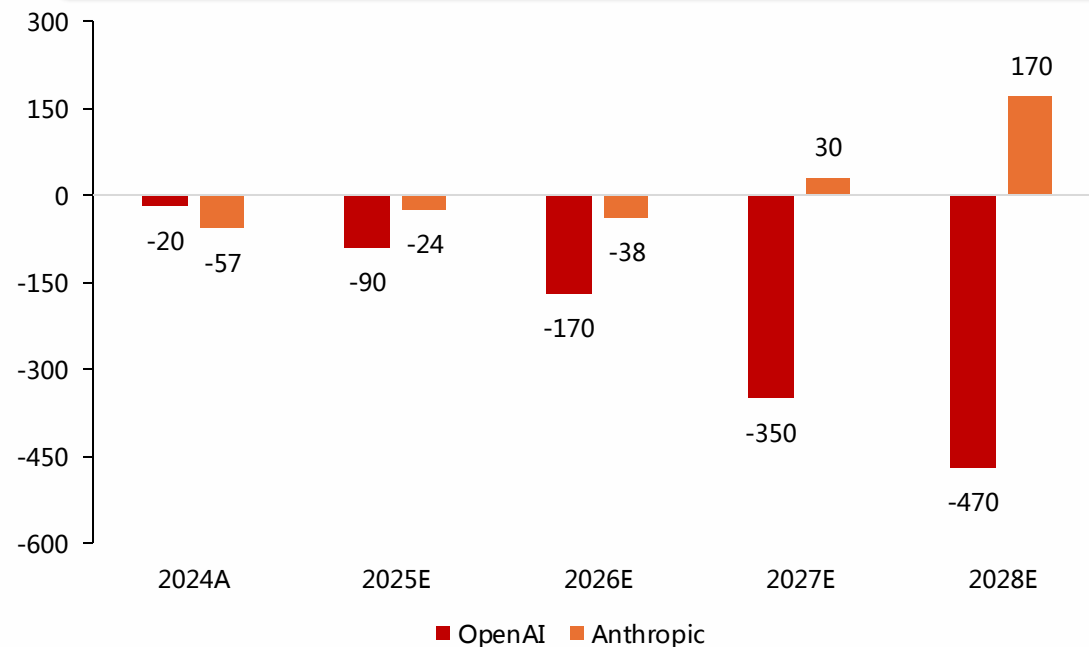
1.5 资本开支→融资需求：AI投入带来现金压力，科技企业融资需求加剧

- **海外科技创企：AI投入带来大额现金需求，企业融资及产品变现诉求加剧。**2025年11月，OpenAI CEO Sam Altman公开表示，未来8年内预计会约1.4万亿美元与基础设施建设和算力相关的支出承诺，大额投入带来更多现金需求，从而推动AI创企加速推进外部融资、旗下AI产品商业化变现诉求加剧。根据The Information预测，到2028年，OpenAI有望创造1000亿美元收入，Anthropic有望实现700亿美元收入；但在高额AI投入下，预计2028年，OpenAI自由现金流仍然为负，而Anthropic可实现170亿美元自由现金流。

2024A-2028E OpenAI和Anthropic收入（亿美元）



2024A-2028E OpenAI和Anthropic自由现金流（亿美元）



1.5 资本开支→融资需求：AI投入带来现金压力，科技企业融资需求加剧

❑ **海外科技创企—OpenAI：融资金额持续增加，资本密度加速推进IPO。** 1) **2016-2022年：**OpenAI公司成立初期，启动资源主要来自于多种形式的捐赠，包括1.37亿美元的现金捐赠，以及来自亚马逊AWS（180万美元）、微软Azure（至少5000万美元）和谷歌GCP（至少5000万美元）的计算资源；2019年获得微软等战略投资；2021年进行Pre-A轮融资，微软再次追加20亿美元战略投资。**2) 2023年-2025年：**OpenAI凭借AI大模型上的领先优势，不断获得科技大厂和金融机构的战略投资、风险投资或债务融资，融资金额和融资频率相较公司发展前期明显提升，推动公司在资金支持和商业合作上实现双重受益。**3) 2026-2027年：**在1.4万亿美元投资承诺背景下，OpenAI的现金需求将更加迫切，除一级市场融资外，公司有望在二级市场进行融资，未来IPO进程有望加速。

截至2025年OpenAI融资情况

时间	融资类型	融资额（亿美元）	投资方数量	领投方
2025年12月11日	Corporate Round	10	1	Disney
2025年10月2日	二级市场	66	5	T. Rowe Price/Thrive Capital/SoftBank/MGX/Dragoneer
2025年3月31日	风险融资	400	16	SoftBank
2024年10月3日	债务融资	40	9	Goldman Sachs/Morgan Stanley/JP Morgan
2024年10月2日	风险融资	66	15	Thrive Capital
2024年2月16日	二级市场	95	5	Thrive Capital
2023年4月28日	二级市场	3	10	Andreessen Horowitz/K2 Global/Sequoi/Thrive Capital
2023年1月23日	Corporate Round	100	1	Microsoft
2021年1月1日	二级市场		4	Tiger Global Management
2019年7月23日	Corporate Round	10	2	Microsoft
2019年3月11日	种子轮		2	Khosla Ventures, Reid Hoffman Foundation
2016年8月22日	预种子轮	0.0012	1	Y Combinator

1.5 资本开支→融资需求：AI投入带来现金压力，科技企业融资需求加剧

❑ 海外科技创企—Anthropic：融资密度加大，亚马逊与谷歌追加多轮投资。1) Anthropic&亚马逊：2023年9月，Anthropic与亚马逊达成战略合作，亚马逊直接下注40亿美元并获得少量股权；2024年，亚马逊继续追加投资40亿，一跃成为Anthropic最大投资者，同时明确将亚马逊AWS作为其主要云服务提供商，持续在AWS上训练和部署Claude等系列模型，双方合作程度持续加深。2) Anthropic&谷歌：2023年2月及9月，谷歌分别投资3亿和12.5亿美元，成为Anthropic重要股东之一，并在后续Anthropic的多轮融资中持续追加，总投资额已达数十亿美元，同时，Anthropic逐步与谷歌展开云合作，并计划采用TPU承担部分AI工作负载。

截至2025年Anthropic融资情况

时间	融资类型	融资额（亿美元）	投资方数量	领投方
2025年9月2日	F轮融资	130.0	24	Fidelity/ICONIQ Capital/Lightspeed Venture Partners
2025年5月16日	债务融资	25.0	7	/
2025年3月3日	E轮融资	35.0	12	Lightspeed Venture Partners
2025年1月22日	Corporate Round	10.0	1	Google
2024年11月22日	Corporate Round	40.0	1	Amazon
2024年6月11日	风险融资	/	1	/
2024年5月31日	二级市场	4.5	15	G Squared/MVP
2024年3月22日	二级市场	8.8	14	Advanced Technology Investment Company/MVP
2024年1月11日	D轮融资	7.5	4	Menlo Ventures
2023年10月27日	Corporate Round	20.0	1	Google
2023年9月	Corporate Round	12.5	1	Amazon
2023年5月	C轮融资	4.5	/	Spark Capital/Google/Salesforce Ventures/Zoom
2023年2月	Corporate Round	3.0	1	Google
2022年4月	B轮融资	5.8	/	Sam Bankman-Fried
2021年5月	A轮融资	1.2	/	Dustin Moskowitz/Eric Schmidt/Jaan Tallinn

目录

- ◆ 一、资本开支节奏：资本开支预期上调，未来现金流压力加大
- ◆ 二、数据中心建设：电力容量面临限制，追求每瓦特产出效率
- ◆ 三、模型能力演进：研究端产品端并进，工程化能力继续增强
- ◆ 四、AI云业务增速：算力容量加速释放，AI云服务迈向扩张期

2.1 数据中心建设模式：自建/租赁等多种方式共建，通用性/灵活性成为建设共识

- **第一梯队云厂商主要采取自建模式，第二梯队云厂商带壳部署或租赁占比较多。** 1) **微软**：自建+租赁。微软在25Q3的349亿美元资本支出中，有111亿美元是用于大型数据中心站点的融资租赁。2) **亚马逊**：自建+带壳部署。3) **谷歌**：自建为主。谷歌于25Q3业绩会表示，多数数据中心均由自己建造，公司对其进行优化，确保以最高效的方式完成。4) **甲骨文**：带壳部署、租赁为主。公司将资本开支主要用于设备，而非建筑；建筑由合作伙伴建成后收取租金；公司不拥有土地或建筑物，而是拥有设备，并对设备进行优化。5) **Coreweave**：带壳部署、租赁为主。截至25Q3，公司任何单一数据中心供应商在签约电力组合中的占比都不超过20%，同时，公司已启动自建数据中心计划，以提升运营自主性并加快扩张。6) **Nebius**：定制化/租赁+自建同时存在。Nebius新泽西数据中心委托合作伙伴按照制定的规格和设计建造；同时公司正逐步减少租赁或共址机房的比例，更多转向自建设施。

海外云厂商数据中心主流建设模式对比

建设模式	自建 (Self-Built)	带壳部署 (Powered Shell)	定制化 (Build-to-Suit)	整租机房 (Turn-Key)	托管租赁 (Colocation)
定义	1) 无需第三方数据中心供应商； 2) 云厂商：完全自建/自有/自营；需要经过拿地/电力接入/能评/环评/报建/施工等全流程	1) 第三方数据中心供应商：提供物理外壳/土建/主配电/主冷却系统； 2) 云厂商：需自行部署UPS/电池组/母线槽/机柜等	1) 第三方数据中心供应商：为客户建设定制化数据中心，产权归属于数据中心供应商； 2) 云厂商：决定数据中心结构/供电等设计方案，通常采用10~20年长期租约	1) 第三方数据中心供应商：基础设施（建筑物/电力/UPS/冷却）已全部Ready； 2) 云厂商：只需搬入机柜和服务器等即可运营	1) 第三方数据中心供应商：提供现成数据中心机柜/空间/电力等全部资源； 2) 云厂商：按需直租即可，按机柜/月、按 kW/月计费
适用场景	长期战略、技术定制要求高的企业	控制需求高、但想缩短上线周期	对定制性有要求但不想自建的客户	快速扩容、愿承担全部运营责任	短期扩容、也适合中小/初创AI公司
控制权	□□□□□ 最强	□□□□ 强	□□□ 中	□□ 弱	□ 最弱
部署速度	🐢🐢🐢 最慢 (12-24个月)	🐢 中等 (6-12个月)	🐢🐢 偏慢 (12-18个月)	🐘 偏快 (3-6个月)	🚀🚀🚀 最快 (1-3个月)
前期投入	💰💰💰💰💰 Capex最高	💰💰💰💰 Capex较高	💰💰💰 Capex中等，主要由数据中心供应商承担；云厂商承担定制化部分	💰💰 Capex略低	💰 Capex最低
后期运营成本	💰 长期摊薄、长期TCO低	💰💰 TCO较低	💰💰💰 长期租金锁定，取决于合同结构	💰💰💰 较高，租户承担运营成本	💰💰💰💰 短期便宜，长期昂贵

资料来源：各公司公告，西南证券整理

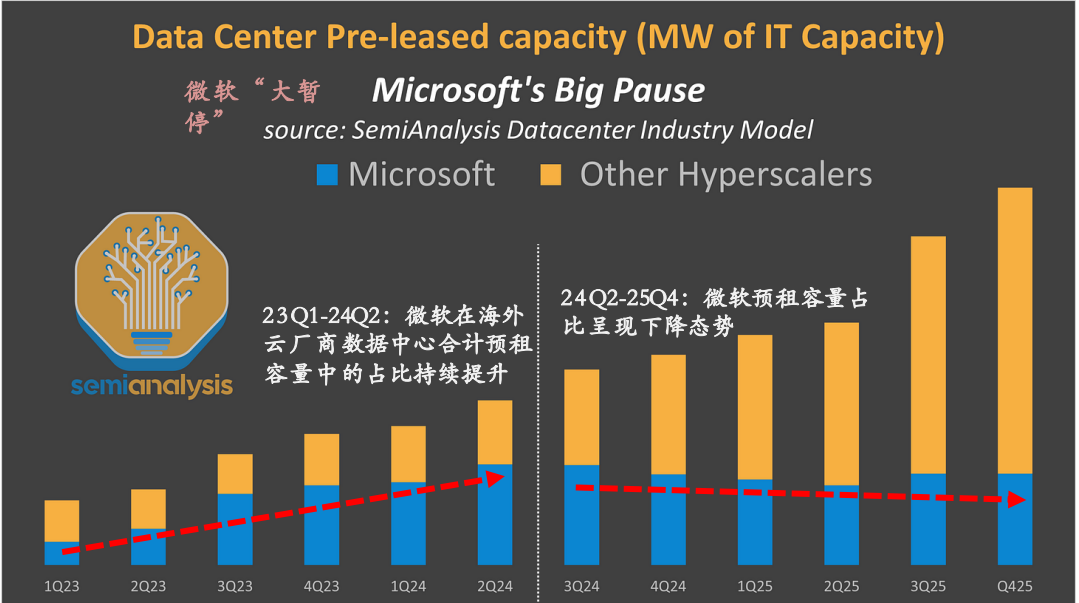
2.1 数据中心建设模式：自建/租赁等多种方式共建，通用性/灵活性成为建设共识

- ❑ **数据中心通用性成为云厂建设共识，以保证在不同计算任务之间灵活切换。**目前，多数海外云厂商均强调，在建设数据中心时，目标是建立一个通用的、能支持多代技术和多种模型的超大规模计算业务，而非为单一技术或单一客户过度投资，以适应不同代际的GPU、满足更灵活的功率和冷却要求，同时能够在训练和推理工作负载之间灵活切换。根据SemiAnalysis信息，2024年中，微软基于风险与回报的考量，实施数据中心建设的“大暂停”，截至24Q2，微软在海外云厂商数据中心预租总量中的占比一度超过60%，而自25Q3开始，微软逐步停止租赁活动，其他超大云厂商的预租规模大幅增加，25Q4微软预租容量占比下降至25%。微软于2024年中开启的数据中心“大暂停”调整，代表着海外云厂商在FOMO capex和ROI capex之间的战略平衡，并强调数据中心是为多类模型、多代硬件设计，而非为单一模型、某一代硬件建设，基础设施需要灵活地跨越多方面要求。

海外云厂商有关“数据中心通用性/灵活性”表述

云厂商	“数据中心通用性/灵活性”相关表述
微软	强调服务器集群的 通用性(fungible) ，追求计算集群（CPU/GPU/存储）的 灵活性 ，可以同时服务于第一方、第三方和商业云客户。
亚马逊	/
谷歌	强调为AI优化的数据中心，通过整合高性能AI专用存储、串联高带宽光学互联网络，动态调整集群结构，以便 在训练与推理之间切换 ，无需中断服务。
甲骨文	公司为所有客户提供的AI云基础设施完全相同，OCI从一开始就选择裸金属虚拟化，以及安全擦除硬件等技术。
CoreWeave	强调基础设施是 可转用的/通用的(fungible) ，可以从一个客户切换到另一个客户；同时基础设施按照最苛刻的规格建造，既可用于训练，也可用于推理，在设计上确保最大程度的 灵活性和可选性 。
Nebius	/

23Q1-25Q4微软在海外云厂数据中心预租容量占比



2.2 数据中心建设目标：电力容量成为核心瓶颈，追求最大化每瓦特产出效率

❑ **电力供应成为核心瓶颈，追求最大化每瓦特产出效率。**当前，海外云厂商普遍表态当前数据中心的建设瓶颈主要在于空间和电力供应，从具体环节上看，电力瓶颈主要在于供电外壳（Powered-shell）层面。通常，云厂商在规划建设某一数据中心时，需要率先确定支持该数据中心的电力容量，然后再对数据中心内部的各类IT设备进行配置。因此，在电力容量限制下，海外云厂商均开始强调最大化现有数据中心的功率使用、最大化每瓦特下的tokens产出效率，数据中心建设将针对芯片、存储、网络等硬件环节，以及软件栈和系统架构进行全面优化，以提高数据中心能效和算力利用率。

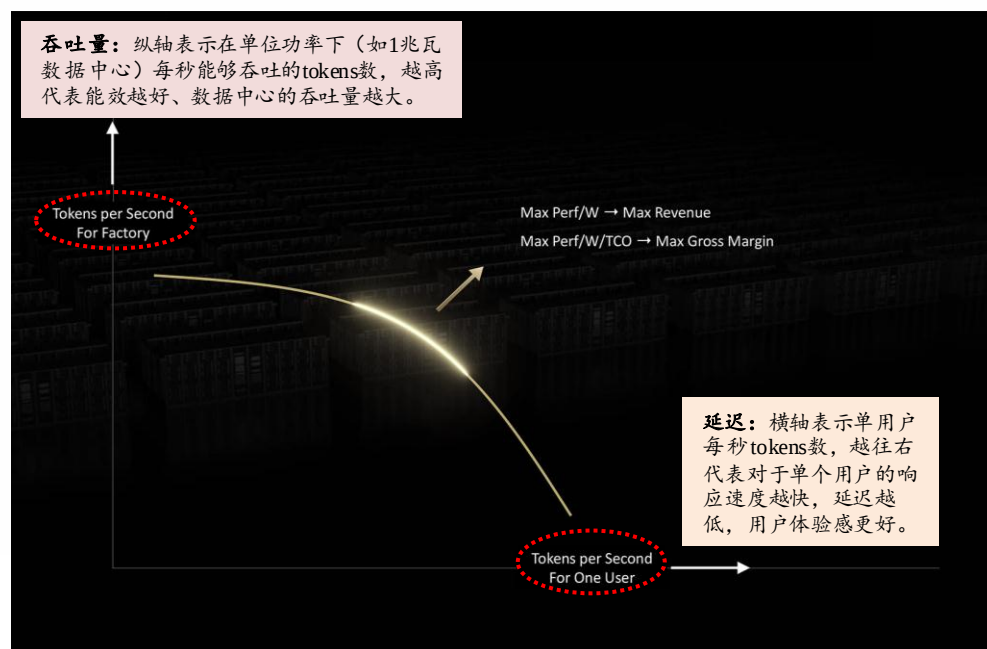
海外云厂商有关数据中心“电力瓶颈”和“产出效率”表述

云厂商	“数据中心电力供应瓶颈”相关表述	“数据中心产出效率”相关表述
微软	过去几年公司短缺的并不是GPU或CPU本身，而是 空间和电力 ，即数据中心的承载能力；公司过去花费大量时间解决基础设施问题，现在也通过租赁形式建设这些长期资产。	在基础设施方面，公司正在打造一个“行星级云”和“AI工厂”， 目标是在每美元、每瓦特下最大化token效率，同时满足各类客户需求 。25Q3将GPT-4.1和GPT-5（两个使用最广的模型）的GPU token吞吐量提升30%以上。
亚马逊	目前整个 行业的瓶颈可能在电力供应上 ；未来在某个时点，也可能转向芯片短缺。	公司正在努力 最大化现有数据中心的功率使用 ，既降低成本，也回收电力用于其他更新的工作负载。Trainium相比其他方案在性价比上高出30%~40%，预计Trainium3将比Trainium2再提升约40%的性能。
谷歌	/	优化自建数据中心方式， 提高能效与利用率 ，数据中心的系统针对高性能、高可靠、大规模的训练与推理进行优化；在运行大规模集群时， 实现2倍的能效提升（FLOPs per watt） ，在如今电力资源稀缺的环境下，这意味着更大的容量。
甲骨文	公司于FY26Q1（截至2025/8/31）业绩会上表示，算力供应跟不上需求，是因为 数据中心建设和计算机部署速度有限 ，而GPU的供应没有问题。	/
CoreWeave	不是电力不足的问题，目前电力供应充足，并且认为未来几年都不会成为瓶颈，真正的难题是在 “powered-shell” 层面。以当今市场的算力供给水平，根本无法满足需求，而这种缺口（电力、数据中心空间等）还在恶化。	/
Nebius	短期来看，主要挑战在于 电力获取和供应链环节 。	Greenfield的建造成本比build-to-suit低，且不受长期租约限制,通过控制建筑设计，从电力引入到机架和服务器的安装，可以实现更低的总拥有成本，约比市场平均水平低20%左右，可以根据实际需求灵活调整分阶段部署。

2.2 数据中心建设目标：电力容量成为核心瓶颈，追求最大化每瓦特产出效率

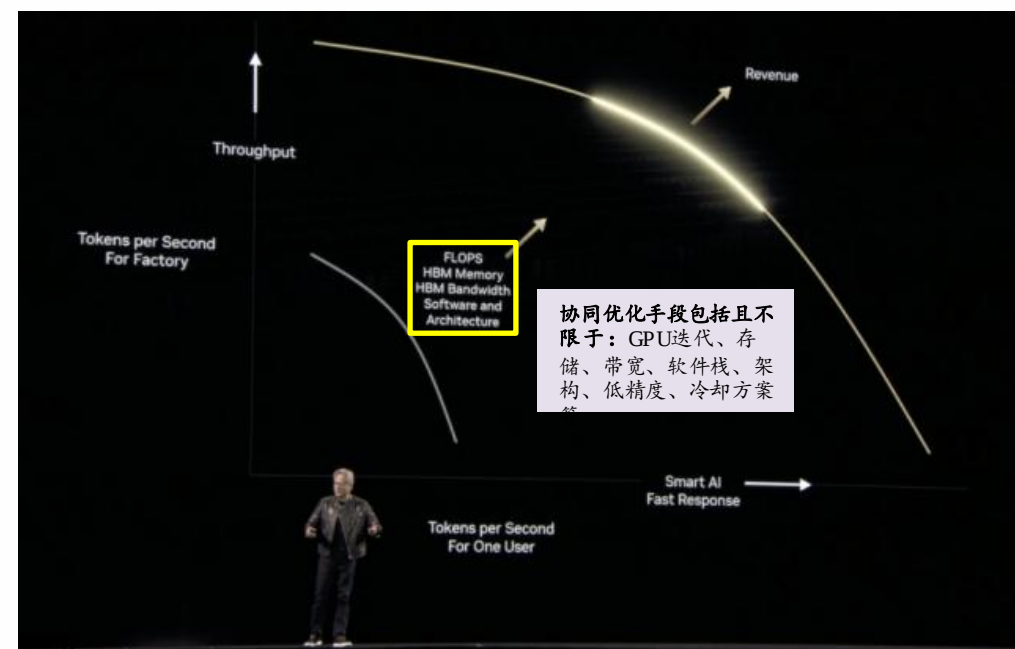
- ❑ **数据中心软硬件协同优化，帕累托前沿曲线加速上移。** 2025年，英伟达多次强调AI工厂的帕累托前沿曲线(Pareto Frontier)以及数据中心的极致协同设计(Extrem Co-Design)。AI工厂的帕累托前沿曲线揭示出数据中心在特定功率下的吞吐量与延迟之间的平衡，在该曲线上，改善一个目标意味着牺牲另一个目标，即“低延迟时吞吐量下降、高吞吐量时延迟上升”。数据中心的极致协同设计方面，包括芯片、计算系统、软件、模型架构、应用程序等各个环节，从而使得AI工厂在晶体管数量仅扩张2倍的情况下，仍然能够获得tokens产出效率10倍性能的提升。根据英伟达GTC大会，算力、存储、带宽、软件和架构等多维度的升级与优化，将推动帕累托前沿曲线向右上方推进，帮助AI工厂或数据中心实现收入的更大化。

AI工厂的帕累托前沿曲线：平衡吞吐量与延迟性能



资料来源：英伟达官网，西南证券整理

AI工厂的帕累托前沿曲线：通过协同优化推动曲线上移



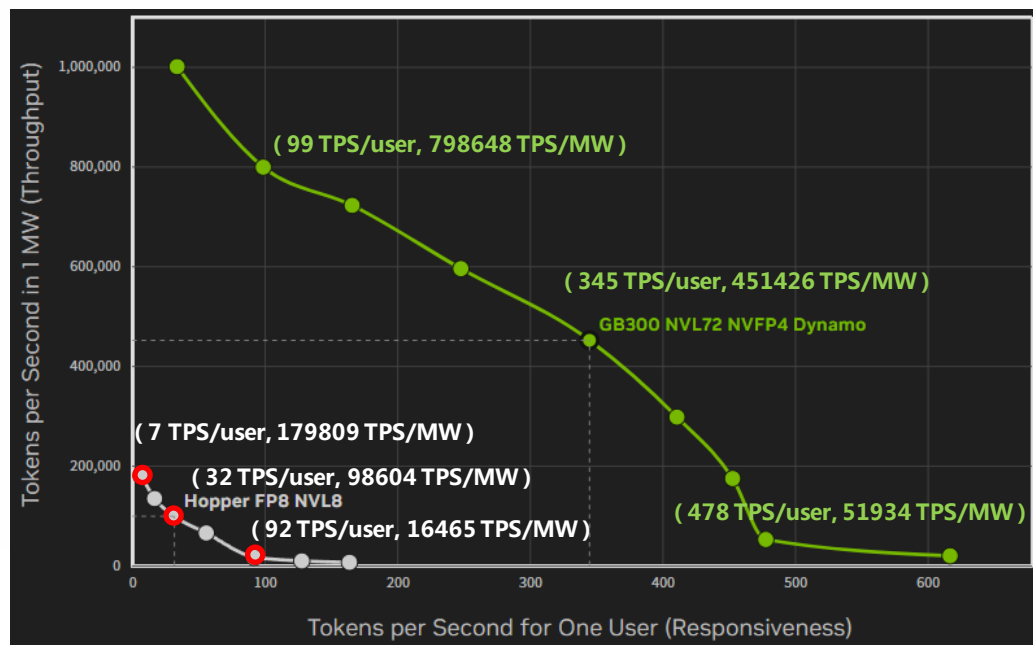
资料来源：英伟达GTC大会，西南证券整理

2.2 数据中心建设目标：电力容量成为核心瓶颈，追求最大化每瓦特产出效率

□ 算力芯片持续迭代，每瓦特产出效率大幅提升。

- **英伟达Hopper VS Blackwell**：根据英伟达官网信息，在1兆瓦功率限定下，若保持AI大模型为每位用户提供每秒90多个Tokens的生成速度，那么由Hopper系列芯片构建的数据中心每秒可以产生1.6万个Tokens，由GB300构建的数据中心每秒则能产出近80万个Tokens，实现约50倍增长，数据中心在芯片、存储、软件和架构等多维优化下，使得每瓦特下的Tokens产出效率实现大幅提升。
- **英伟达Blackwell VS Rubin**：根据2026年1月英伟达CES大会，Rubin系列产出效率有望在Blackwell基础上再提升10倍。

英伟达Hopper与Blackwell帕累托前沿曲线对比



资料来源：英伟达官网，西南证券整理

英伟达Blackwell与Rubin帕累托前沿曲线对比



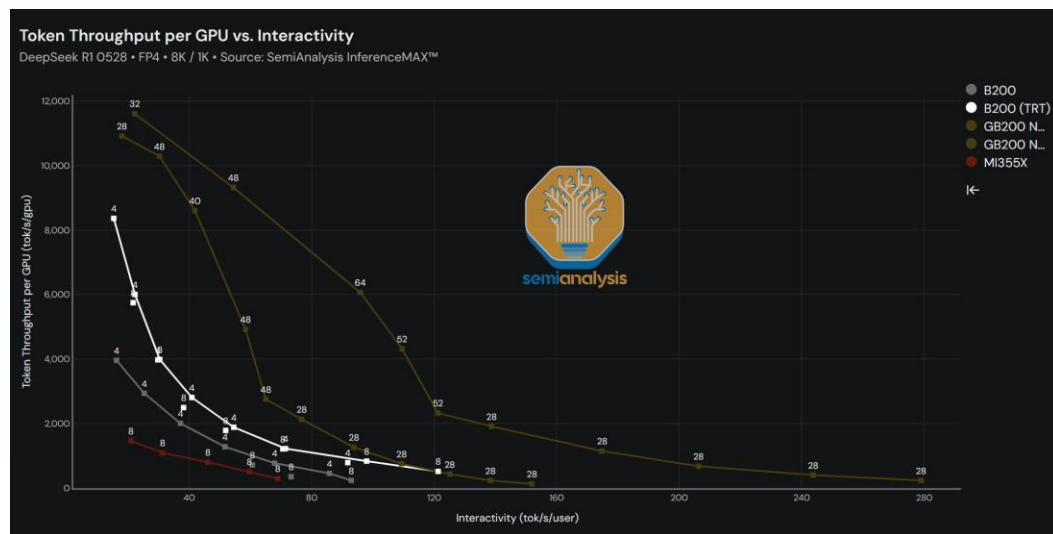
资料来源：英伟达CES大会，西南证券整理

2.2 数据中心建设目标：电力容量成为核心瓶颈，追求最大化每瓦特产出效率

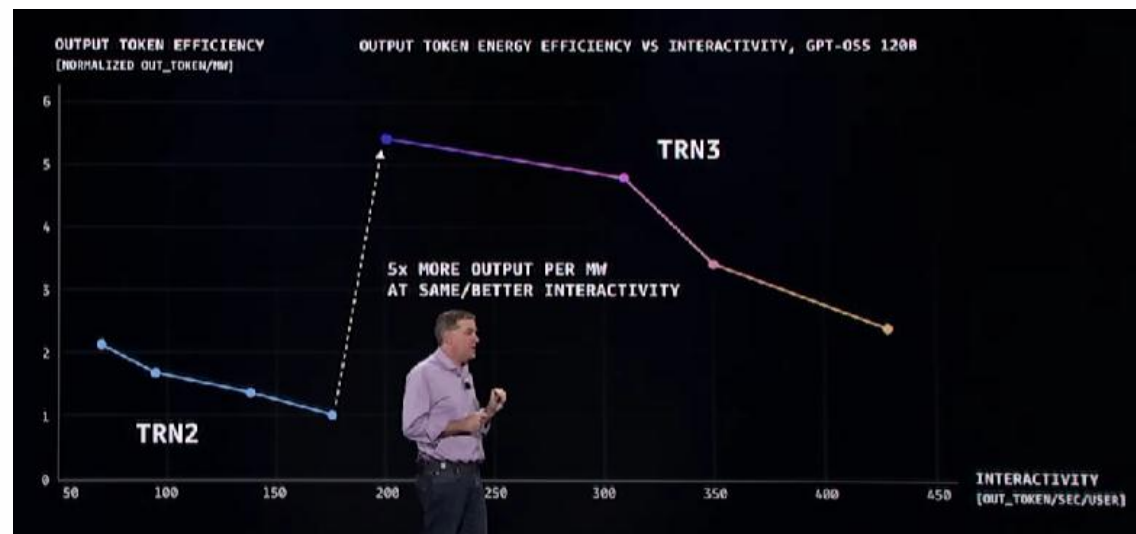
□ 算力芯片持续迭代，能效对比提升明显。

- **英伟达Blackwell VS AMD MI355**：根据SemiAnalysis基于其专有测试，得出Blackwell系列服务器在系统级创新和特定应用场景配置下，相较AMD的MI355系列芯片，能够带来较大数量级的性能提升。
- **亚马逊Trainium 2 VS Trainium 3**：2025年12月2日，亚马逊正式推出新一代AI训练芯片Trainium 3，并预告下一代Trainium 4的开发计划。Trainium 3是AWS首款采用3nm制程的AI芯片，专为下一代智能体、推理和视频生成应用App提供最佳Tokens经济效益而设计。根据AWS官方数据，搭载Trainium 3芯片的Trn3 UltraServer系统在性能上实现显著提升，Amazon EC2 Trn3 UltraSevers在能效比上每兆瓦所处理的token数量达前代产品的5倍。

英伟达Blackwell系列芯片性能对比



亚马逊Trainium系列芯片性能对比

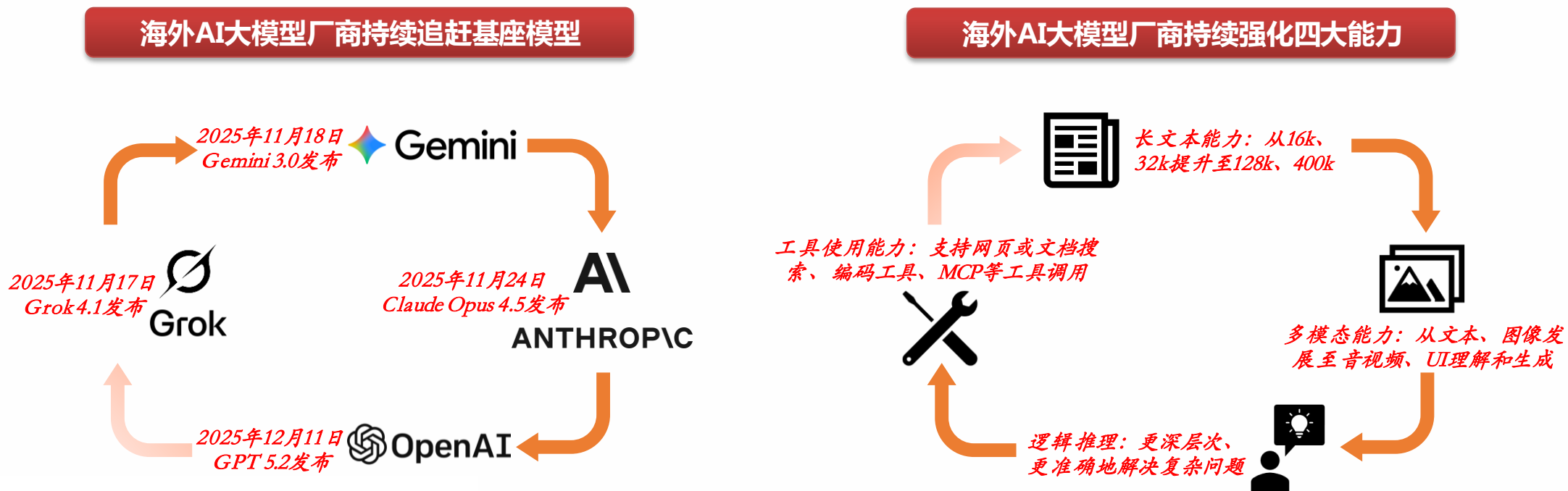


目录

- ◆ 一、资本开支节奏：资本开支预期上调，未来现金流压力加大
- ◆ 二、数据中心建设：电力容量面临限制，追求每瓦特产出效率
- ◆ 三、模型能力演进：研究端产品端并进，工程化能力继续增强
- ◆ 四、AI云业务增速：算力容量加速释放，AI云服务迈向扩张期

3.1 模型侧：AI模型算力需求持续，研究端与产品端并进

- **AI大模型竞争持续，演进方向聚焦四大能力。**当前，AI基础大模型竞争已进入高强度、快迭代阶段，2025年11月至12月，Grok 4.1、Gemini 3.0、Claude Opus 4.5、GPT-5.2相继发布，模型竞争形成追赶态势，反映头部大模型厂商在算力投入、模型训练与工程化能力上持续加码；未来，基座大模型与特定领域的持续优化仍是主要竞争方向。从模型能力上看，大模型迭代主要围绕**长文本、多模态、逻辑推理、工具使用**四大能力，追求更长的上下文以满足个性化记忆，多模态能力从文本和图像扩展至音视频理解生成能力，推理能力则向更深层次与更高准确率演进，同时，通过工具调用、结合MCP等机制强化未来AI大模型在具体任务上的执行能力。



3.1 模型侧：AI模型算力需求持续，研究端与产品端并进

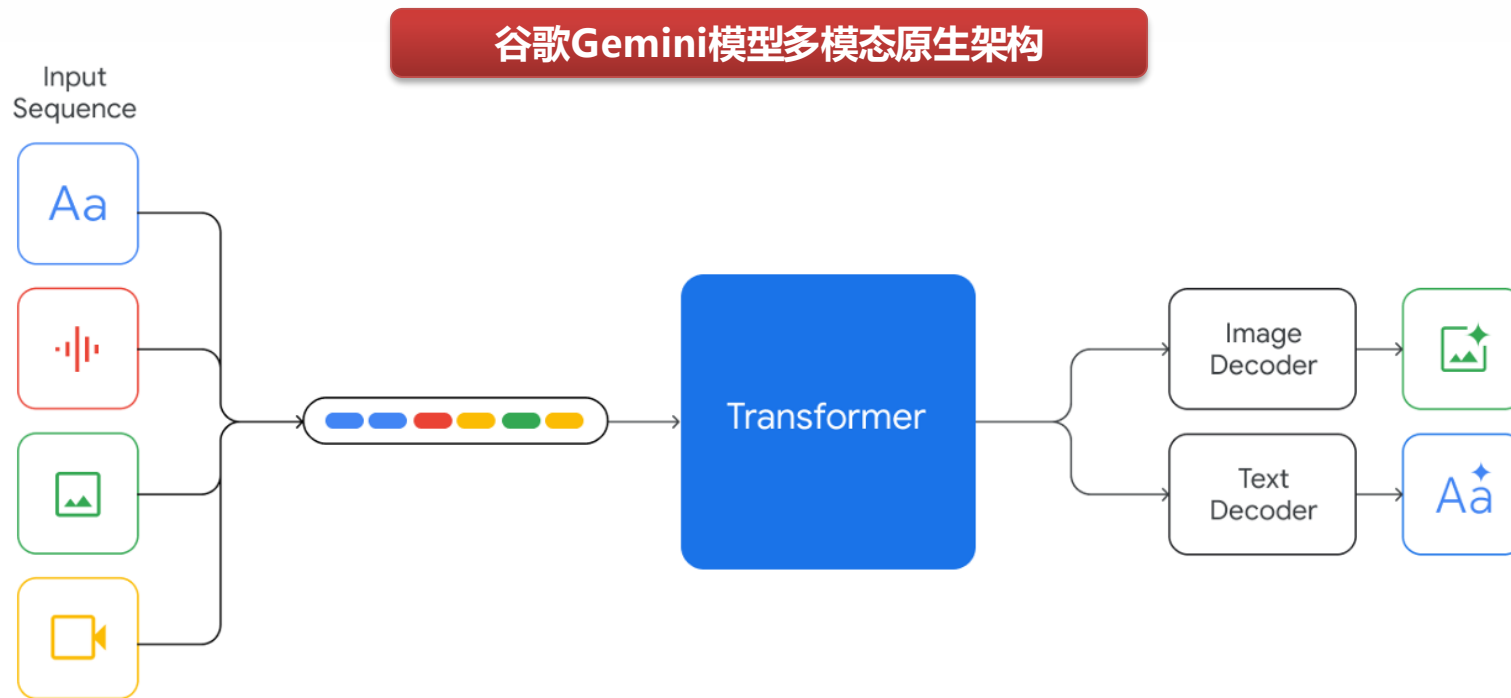
□ **训练算力：主流模型厂商继续进行预训练，同时关注持续学习和强化学习。**2024年，大模型厂商在预训练环节投入算力较高；2025年，大模型竞争格局收敛，主流厂商基于此前的基座模型进行持续打磨，算力消耗逐步向后训练的强化学习倾斜；2026年，训练算力需求有望持续增长，核心玩家将**继续预训练**，强化学习仍会持续，持续学习有望盛行。**1) 预训练：**随着竞争格局逐步收敛，留在牌桌上的模型厂商未来仍将继续预训练，探索更高效的模型架构和训练配方，预训练算力需求有望持续。**2) 中训练：**中期训练可用于延长模型知识的截止日期、提升特定领域的能力、或为高计算量的强化学习做准备。通常，在高度专业化的科学领域中，中训练能够使模型在后训练后达到更高的智能水平。**3) 后训练：**后训练阶段的典型技术为强化学习，主要针对多模态和长文本能力再训练、对智能体产品进行打磨与适配、与人类偏好对齐，未来AI大模型对中训练及后训练阶段的算力需求有望稳健增长。

AI大模型训练三阶段对比

维度	对比指标	预训练（Pre-training）	中训练（Mid-training）	后训练（Post-training）
简介	核心训练目标	通用能力：建立通用语言与世界知识	领域能力：增强专业能力、任务适应性	行为能力：对齐人类偏好与交互风格
	典型训练技术	Transformer、语言建模、并行训练技术	Full training、 Continual training 、PEFT、上下文扩展	SFT、RLHF、DPO
模型	单次训练时长	单次训练时间长（数周至数月）	单次训练时长居中（数天至数周）	单次训练时间短（数小时至数天）
	训练频率	频率低，通常间隔数月或每年1次集中分布式训练	频率高，针对某些领域进行持续迭代和优化	频率中等，目的在于与人类指令偏好对齐
数据集	规模/体量	大	中	小
	质量/复杂度	要求大而全、质量要求相对宽松	专业性强，结构化、干净、领域一致性高	精细标注、人类偏好数据质量要求严格
	公开性/可得性	公开语料，例如书籍/百科/网页等资料	专业领域的数据集	指令数据、人类偏好反馈数据（人工参与成本高）
算力	消耗特征	一次性高峰	高频、持续	中频、间歇
	对集群要求	大型GPU集群，大规模、分布式、高并行	中等到大规模	中小规模，但需要人类参与对其，时间成本较高
	未来需求展望	留在牌桌上的核心玩家将继续进行预训练	对智能体/上下文/专业领域的持续训练	算力需求稳定

3.1 模型侧：AI模型算力需求持续，研究端与产品端并进

- **模型架构：原生多模态对训练算力及模型结构要求更高，但也许更符合第一性原理。** 1) **多模态堆叠**：过去，大语言模型和扩散模型原本设计为单模态模型，而后经过修改或扩展，逐步具备处理其他模态的能力，多模态堆叠成为主流方法。2) **多模态原生**：2023年12月，谷歌发布Gemini大模型，支持原生多模态，从一开始就对多模态进行统一的预训练，再用额外的多模态数据进行微调和对齐，使Gemini系列模型可以互通和推理多种模态信息。其中，谷歌Gemini 3模型被明确描述为“原生多模态+ 稀疏MoE Transformer”架构，能够统一处理文本/图像/音频/视频和百万tokens长上下文，但在产品层面仍然通过工具/检索/IDE等模块处理任务，形成“原生多模态大脑+模块化外骨骼”的工程形态。此外，部分海外AI实验室仍在探索除Transformer之外的模型架构，但在更优美的模型架构出现之前，Transformer仍有较多工程化能力的提升空间。



3.1 模型侧：AI模型算力需求持续，研究端与产品端并进

❑ **数据使用：在有限数据下优化模型架构，发挥更高的单位算力和数据价值。**当前，在高质量互联网数据有限、暴力美学边际收益递减的情况下，大模型厂商需要在模型架构、数据结构、数据利用率上进行更聪明的设计，让每单位数据发挥更大的价值，在有限的数据下冲击更高的智能上限。此外，随着Scaling Law从预训练延伸至强化学习和持续学习，模型训练对数据集的要求也将发生转变，预训练阶段更强调“对的数据”——即数据的准确性；而在中训练和后训练阶段，将更加注重“好的数据”——即带有思考调整过程、任务推理经验的数据，带经验的好数据有望成为新需求。

OpenAI GPT-5.2支持文本和图像输入

5.2

GPT-5.2

Default

©

The best model for coding and agentic tasks across industries

Compare

Try in Playground

REASONING

Higher

SPEED

Fast

PRICE

\$1.75 · \$14

Input · Output

INPUT

Text, image

OUTPUT

Text

GPT-5.2 is our flagship model for coding and agentic tasks across industries. Learn more in our [latest model guide](#).

400,000 context window

128,000 max output tokens

Aug 31, 2025 knowledge cutoff

Reasoning token support

Pricing

Pricing is based on the number of tokens used, or other metrics based on the model type. For tool-specific models, like search and computer use, there's a fee per tool call. See details in the [pricing page](#).

Text tokens

Per 1M tokens · Batch API price

Input

\$1.75

Cached input

\$0.175

Output

\$14.00

Quick comparison

Input

Cached input

Output

GPT-5.2

\$1.75

GPT-5

\$1.25

GPT-5 mini

\$0.25

仅支持文本、图像模态

谷歌Gemini 3.0支持原生多模态输入

Model information

3 Pro

Status

Preview

Input

Output

Input tokens

1M

Output tokens

64k

Knowledge cutoff

January 2025

Tool use

Function calling

Structured output

Search as a tool

Code execution

Best for

Agentic

Advanced coding

Long context understanding

Multimodal understanding

Algorithmic development

Availability

Gemini App

Google Cloud / Vertex AI

Google AI Studio

Gemini API

Google AI Mode

Google Antigravity

Documentation

[View developer docs](#)

Model card

[View model card](#)

原生支持文本、图像、视频、音频、文档的多模态处理

3.1 模型侧：AI模型算力需求持续，研究端与产品端并进

❑ **产品特性：训练数据带来模型能力分化，推理阶段有望进一步放大差异化特征。**在AI基础设施追求通用性(fungible)的同时，AI大模型与应用产品正在走向分化，模型正在追求智能的非同质化(non-fungible)。在模型训练阶段，中训练和后训练使用的数据集将带来模型能力的差异，各模型将擅长不同的专业和领域，从而产生不同特质的模型。在此背景下，每个模型在推理阶段产生的token也会不同，从而带来不同的智能，这种差异化智能将在后续的反馈和循环中进一步放大，从而催生不同的AI用例。

AI大模型在训练和推理阶段数据使用类型对比

预训练：使用文本/代码/工具调用等数据，代表互联网海量原始数据

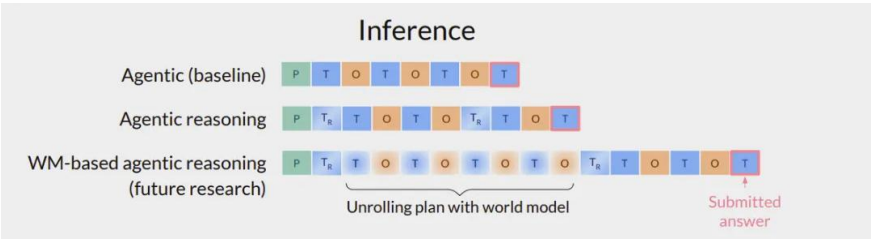
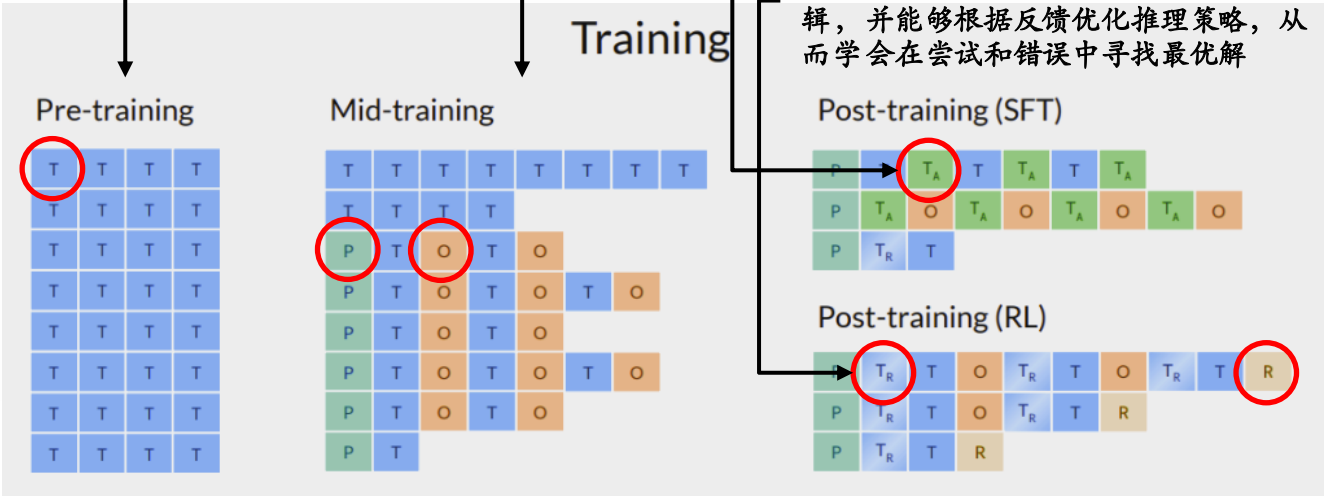
中训练：相较预训练阶段，中训练增加提示词/上下文、观察结果数据，模型能够获得外部环境的反馈

后训练—SFT（监督微调）：增加带有人工标注的文本/代码数据，使模型能够模仿优秀的正确答案

后训练—RL：增加带有推理过程的数据和奖励信号，使模型能够学习思考逻辑，并能够根据反馈优化推理策略，从而学会在尝试和错误中寻找最优解

推理阶段：

- 基础智能体：简单的“提示词-回答-观察-回答”循环；
- 智能体推理：在回答前引入思考；
- 基于世界模型的智能体推理：通过内置的世界模型模拟多种可能性，具备长程规划能力。

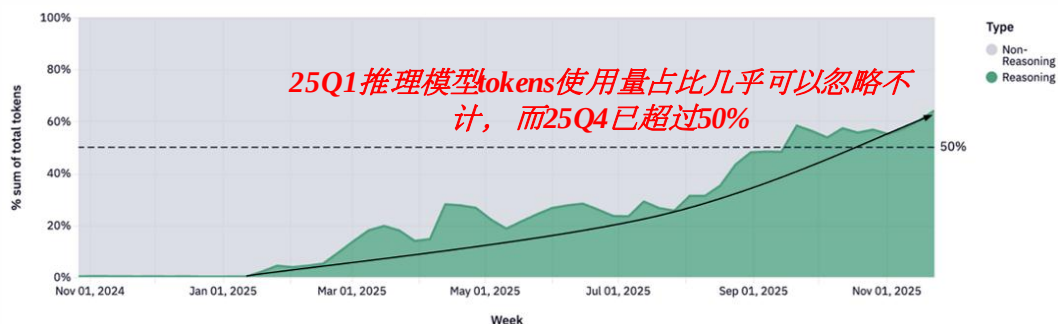


Types of data			
T	Text / code / tool call	P	Prompt (or context) text / code
T _R	Text / code with reasoning	T _A	Human annotated text / code
O	Observations	R	Rewards

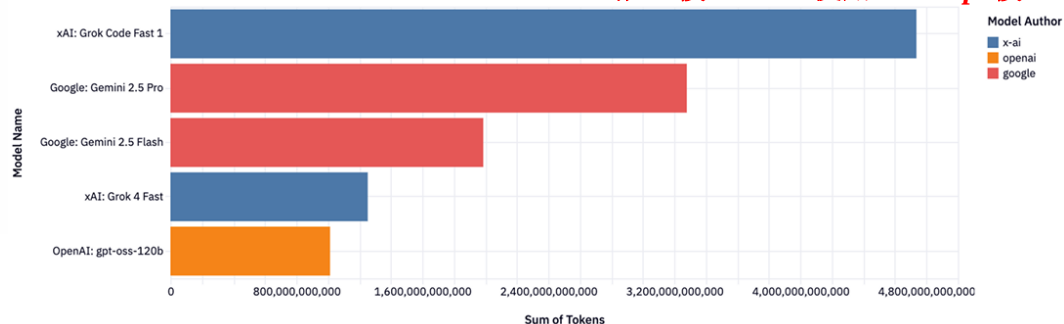
3.2 需求侧：当前推理模型成为主要驱动，未来工具调用需求可期

- ❑ **推理模型爆发**：AI大模型正从“单轮回答”走向“长思考、重推理”，推理模型参数更多、结构更复杂，训练阶段需要更多算力；同时，推理阶段因为思考得更久、生成的内容更长，因此tokens消耗也显著放大。根据OpenRouter平台数据，2025年初，推理模型tokens消耗量占比几乎为零，而25Q4占比已超过50%，推理模型成为2025年tokens消耗量显著增长的主要驱动因素。
- ❑ **工具调用渗透**：AI产品形态正从“聊天机器人”向“智能代理”转变，未来，随着大模型工具使用能力增强、调用频率提升，工具调用带来的tokens消耗量将持续增长。根据OpenRouter平台数据，25Q1工具调用tokens消耗量占比极低，而25Q4达15%。

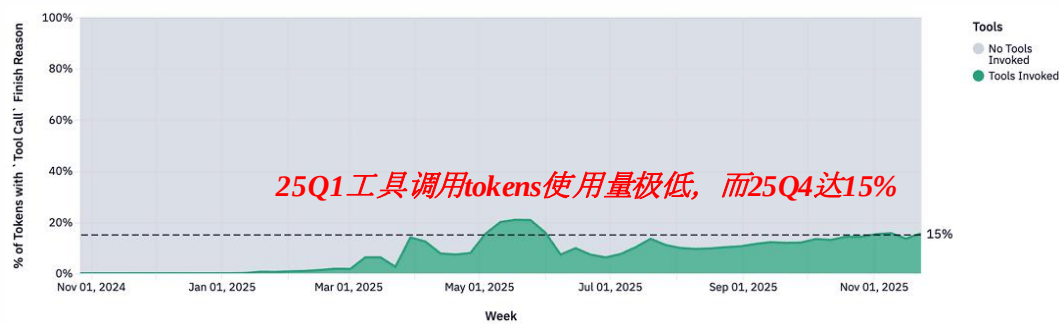
OpenRouter推理模型tokens消耗量占比提升



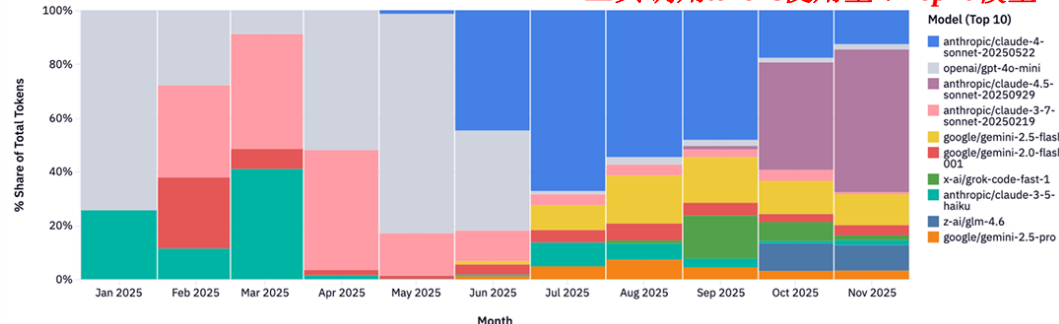
推理模型tokens使用量：Top5模型



OpenRouter工具调用tokens消耗量占比提升



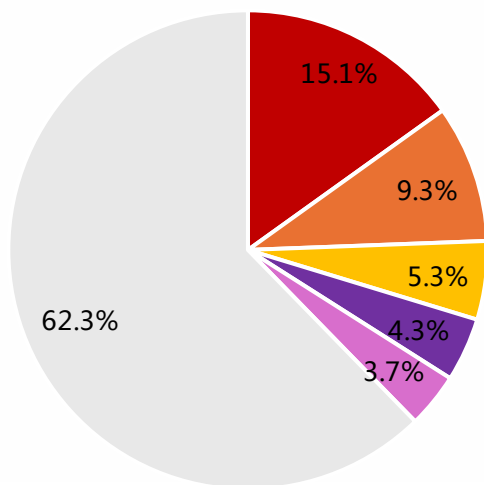
工具调用tokens使用量：Top10模型



3.2 需求侧：当前推理模型成为主要驱动，未来工具调用需求可期

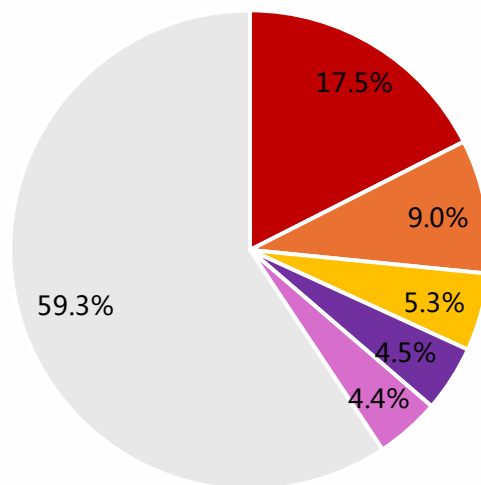
- **头部地区流量占比较高，同时长尾需求特征明显。**根据VisualCapitalist、Resourcera、Anthropic披露数据，2025年，ChatGPT、Gemini、Claude的用户或流量分布在地域结构上呈现出“核心国家/地区占主导、其他地区长尾需求明显”的特征。除美国、印度、巴西等核心地区，其他地区流量占比约60%，反映AI大模型市场长尾需求特征，可满足多语言、多场景、碎片化需求。
- **ChatGPT**：美国、印度、巴西、英国及印尼等市场具备更广泛的用户基础，其他地区占比62.3%。
- **Gemini**：美国与印度合计占比接近27%，反映其在英语市场及人口大国中的渠道优势。
- **Claude**：美国占比超过五分之一，同时日本、韩国等高质量发达市场占比进入前列，用户结构更偏向高付费能力与企业/开发者导向。

ChatGPT流量占比（按国家/地区）



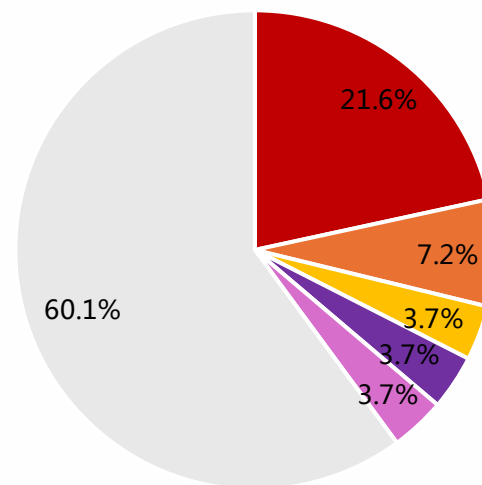
■ 美国 ■ 印度 ■ 巴西 ■ 英国 ■ 印度尼西亚 ■ 其他

Gemini用户占比（按国家/地区）



■ 美国 ■ 印度 ■ 印度尼西亚 ■ 巴西 ■ 越南 ■ 其他

Claude使用占比（按国家/地区）



■ 美国 ■ 印度 ■ 巴西 ■ 日本 ■ 韩国 ■ 其他

资料来源：visualcapitalist，西南证券整理

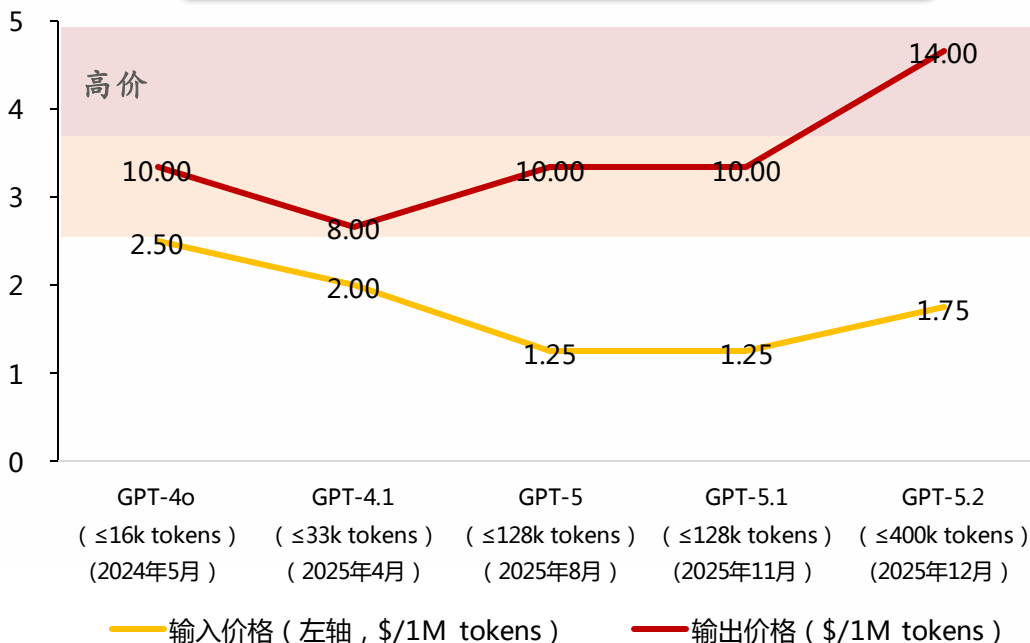
资料来源：resourcera，西南证券整理

资料来源：Anthropic，西南证券整理

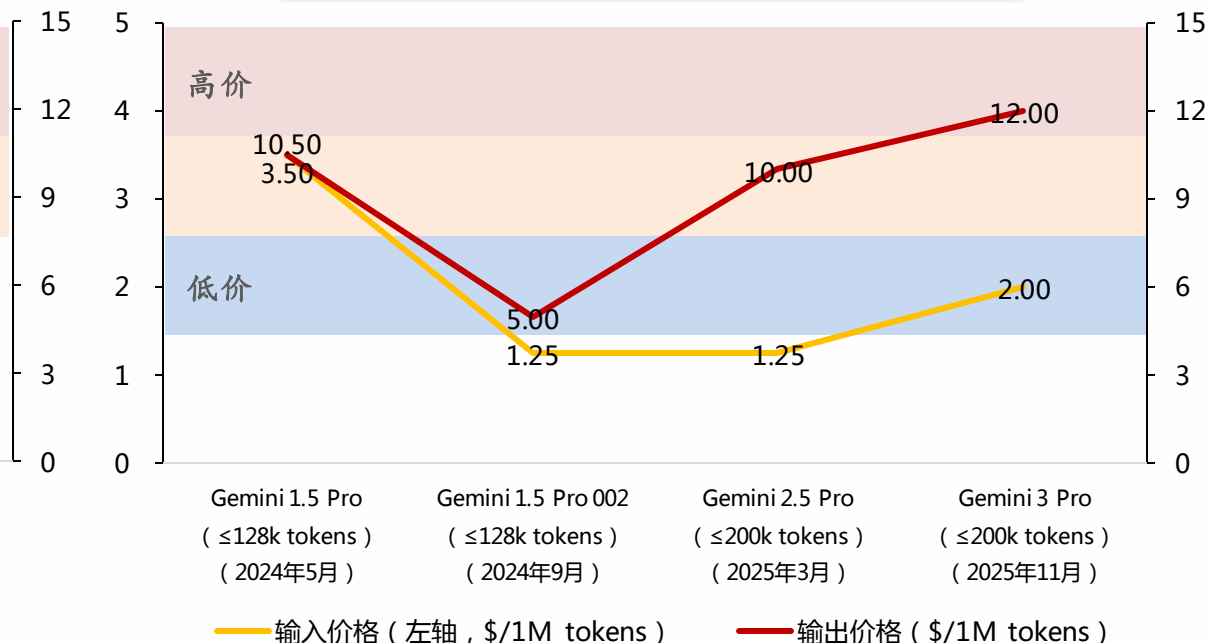
3.3 变现侧：大模型商业化持续推进，用户愿意为更强性能付费

- ❑ **OpenAI GPT系列模型**：从GPT-4o到GPT-5.2模型定价水平来看，2024年至今，OpenAI GPT系列模型输入/输出价格保持相对优势，主要得益于其模型性能相对领先，能够在较高定价水平下保持较强用户粘性。
- ❑ **Google Gemini Pro系列模型**：**1) 2024年上半年**：谷歌Gemini 1.5 Pro模型推出之初，其输入与输出价格均略高于GPT-4o模型；**2) 2024年下半年**：谷歌推出Gemini 1.5 Pro 002版本，同时大幅降价，意在通过低价渗透用户生态、保住市场份额。**3) 2025年**：谷歌凭借Gemini 2.5 Pro和3.0 Pro模型在长上下文、多模态、逻辑推理等方面的性能提升，开始重新调整策略，Gemini 2.5 Pro模型价格与GPT模型对齐，3.0 Pro模型实现反超，定价战略呈现“低价竞争→能力上位→提价反超”的变化节奏。

OpenAI GPT模型标准任务API价格



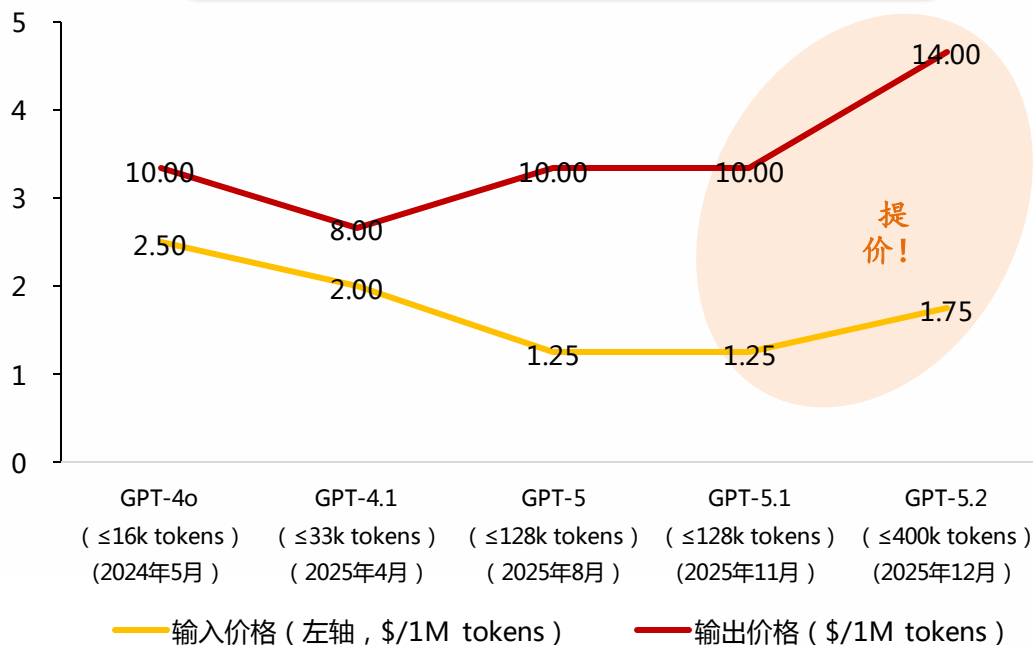
谷歌Gemini Pro模型标准任务API价格



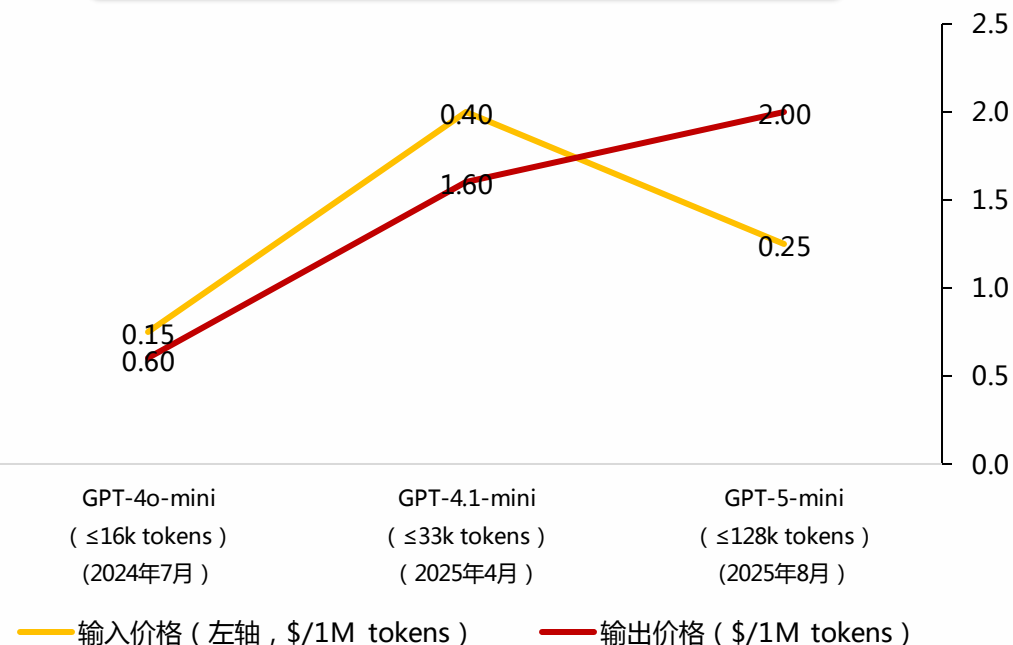
3.3 变现侧：大模型商业化持续推进，用户愿意为更强性能付费

- ❑ **OpenAI定价策略一：追求每token更低成本下的更高智能，以能力驱动高阶定价。**2025年12月，OpenAI推出GPT-5.2模型，API价格较GPT-5.1模型提升约40%，OpenAI认为GPT-5.2模型能够以较低的每token成本实现更高智能，从而在模型能力提升背景下采取涨价手段。
- ❑ **OpenAI定价策略二：打造差异化价格带模型，Mini系列主打性价比。**OpenAI采取跨模型系列的不同定价，在mini模型方面将定价中枢下移，形成差异化价格带，打造模型性价比，以扩大开发者与企业的可用面与调用规模。

OpenAI GPT系列标准版模型API价格



OpenAI GPT系列mini版模型API价格



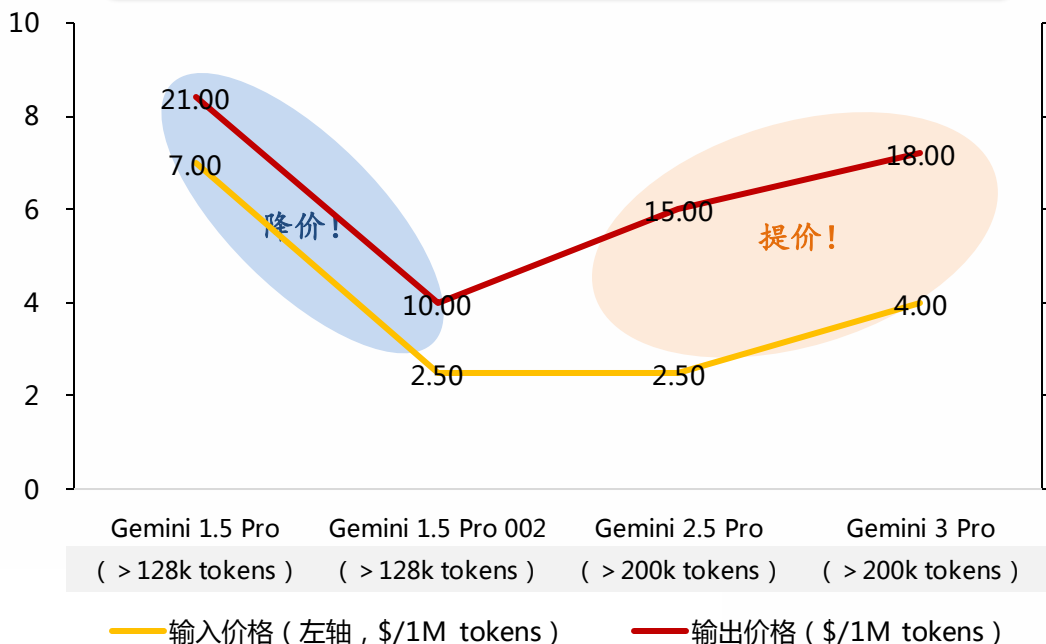
资料来源：OpenAI官网，西南证券整理

资料来源：OpenAI官网，西南证券整理

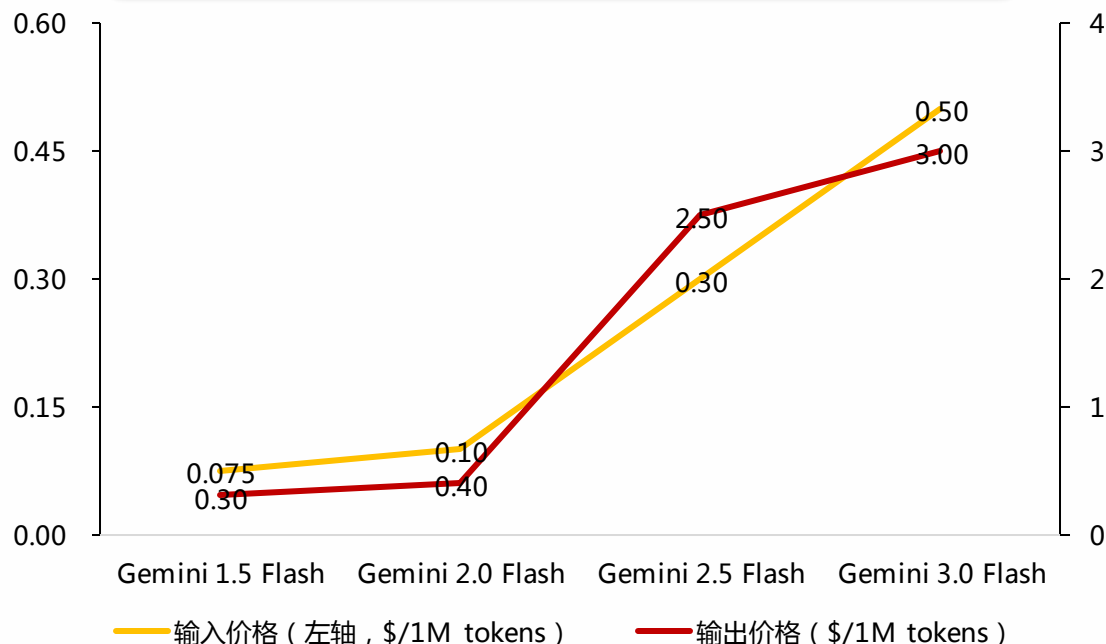
3.3 变现侧：大模型商业化持续推进，用户愿意为更强性能付费

- ❑ **谷歌2024 “防御型跟随年”**：Gemini 1.5 Pro初版定价较高，在竞争压力下，Gemini 1.5 Pro 002版本于2024年9月显著降价，输入/输出价格分别降低64%/52%，调整价格自2024年10月1日生效，输入与输出价格均降至历史最低区间。
- ❑ **谷歌2025 “生态级竞争年”**：进入2025年，谷歌Gemini 2.5 Pro、Gemini 3 Pro的推出标志着谷歌开始进入“性能溢价期”，两代模型在上下文长度、推理深度、多模态能力、专业任务表现上均出现跃升，使其具备重新定价能力，“性能换溢价”的逻辑意味着谷歌不再依赖价格补贴抢占市场份额，而能够凭借其领先技术和性能优势享有溢价空间，推动谷歌Gemini Pro系列模型定价实现扭转。

谷歌Gemini Pro模型长上下文版本API价格



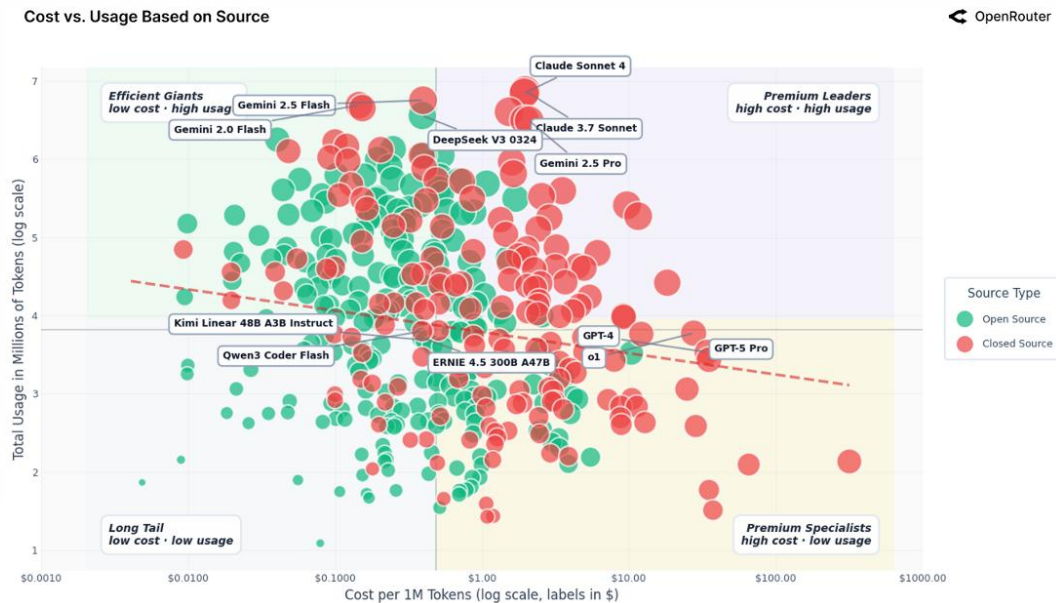
谷歌Gemini Flash模型标准任务API价格



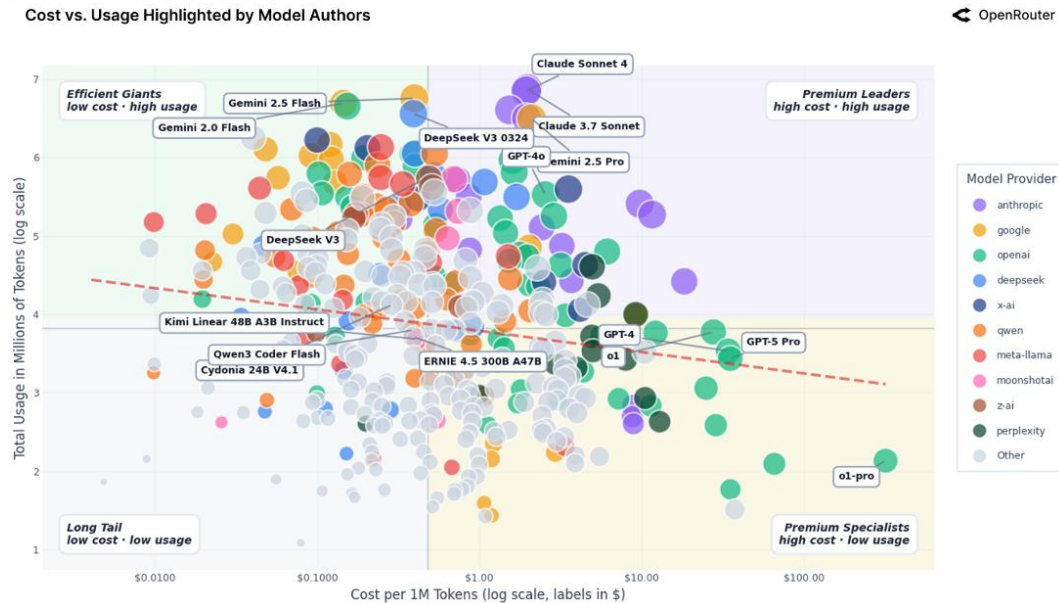
3.3 变现侧：大模型商业化持续推进，用户愿意为更强性能付费

- ❑ **用户侧：模型性能优先级高于定价，用户愿意为更强能力付费。**随着模型能力提升、Agent化程度提高，终端/企业用户越来越接受为AI模型的高性能支付溢价。当前，模型质量与能力的优先级往往优先于成本，便宜本身并不足以赢得使用。如果某个模型在能力上显著领先，或在可靠性与信任度上具有优势，用户愿意承担更高的成本。通常在许多关键 workflows 中，模型带来的价值远远大于API调用的费用。尽管不少开源模型的价格几乎为零，但由于不够可靠，最终仍然无法与具体工作负载形成良好匹配。
- ❑ **产业侧：杰文斯悖论持续有效，降价或为长期趋势。**当模型变得便宜且速度更快以后，用户开始在更多任务中更频繁地调用模型，最终消耗更多总tokens。随着单位token成本下降，模型被集成到各种系统中，总消耗量将继续提升。

AI模型成本与使用量情况（按开源/闭源划分）



AI模型成本与使用量情况（按模型提供商划分）



资料来源：OpenRouter，西南证券整理

资料来源：OpenRouter，西南证券整理

目录

- ◆ 一、资本开支节奏：资本开支预期上调，未来现金流压力加大
- ◆ 二、数据中心建设：电力容量面临限制，追求每瓦特产出效率
- ◆ 三、模型能力演进：研究端产品端并进，工程化能力继续增强
- ◆ 四、AI云业务增速：算力容量加速释放，AI云服务迈向扩张期

4.1 海外AI云服务画像：云服务差异化显现，目标客户群分化

□ **海外云厂AI服务差异化显现，提供多样AI模型及配套工具。**在AI时代下，云厂商对于云业务的定位，是成为汇集众多AI模型、AI智能体和配套工具的平台，因此，无论AI模型或者AI智能体的市场竞争格局如何，云厂商都能从中获益，而AI模型成为AI时代云厂商吸引或创造AI收入的核心驱动力；相关开发工具、数据库、存储等配套增值服务，将成为提升云业务利润率的重要手段。**1）微软Azure**：享有OpenAI API独家IP权利，配套工具完善；**2）亚马逊AWS**：长期与Anthropic Claude模型合作，定制类工具较多；**3）谷歌云GCP**：自研Gemini模型实力强劲，从文本到多模态布局全面；**4）甲骨文OCI**：打造AI模型超市，数据库附加价值大；**5）CoreWeave**：主要提供开源模型，存储服务持续完善，通过收购手段完善软件栈布局；**6）Nebius**：布局开源模型。

海外云厂商提供的AI云服务对比

AI云服务	提供的AI模型	提供的AI配套服务
微软	根据25Q3业绩会，微软向开发者和企业提供超过11,000个模型的访问，比任何厂商都多，包括OpenAI、Meta、Mistral AI、Cohere、DeepSeek、Stability AI旗下第三方模型，MAI/Phi系列自研模型模型，以及25Q3新引入的OpenAI GPT-5和xAI的Grok 4。	数据库与数据存储服务Azure Cosmos DB、Microsoft Fabric、Azure PostgreSQL、Azure SQL Database Hyperscale；应用开发与部署平台Azure App Services、AI应用开发平台Azure AI Foundry等。
亚马逊	通过Bedrock平台提供第三方模型Anthropic Claude、Meta Llama、DeepSeek、Mistral AI Pixtral Large等、自研模型Amazon Nova系列。	Amazon SageMaker支持模型定制；AgentCore、Strands等工具。
谷歌	根据25Q2业绩会，Vertex AI平台集成超过200个基础模型；自研模型包括Gemini大语言模型系列、Imagen、Nano Banana图像生成模型、Veo系列视频生成模型、Genie 3世界模型等；第三方模型Meta Llama、Anthropic Claude等模型。	数据云BigQuery，开发工具和平台如Agentspace，智能体开发工具Agent Development Kit（ADK），JAX/XLA/Pathways等编译器。
甲骨文	云基础设施能够访问最新最强大的AI模型，包括OpenAI GPT-5、xAI Grok、Google Gemini、Anthropic Claude、Meta Llama、Cohere系列模型，通过生成式AI服务提供，并能在应用中充分利用这些模型。	数据库服务、Gen AI Service、AI代理平台等。
CoreWeave	提供OpenAI的开源模型、Meta Llama、DeepSeek、Kimi K2、Qwen3等。	在不断完善AI工作负载配套的托管存储服务，例如CoreWeave AI对象存储、Kubernetes服务等；同时通过收购手段完善自身软件栈能力，例如观测平台Weights & Biases，无服务器强化学习工具OpenPipe等。
Nebius	Meta Llama、Qwen、DeepSeek、Mistral、Nematron、OpenAI的开源模型等。	迭代企业级平台Aether至3.0版本；与DDN、VAST等合作存储解决方案；开发工具与集成AI Studio、JupyterLab、MLflow、Slurm等。

4.1 海外AI云服务画像：云服务差异化显现，目标客户群分化

□ **头部客户订单额高，长尾客户需求广泛。**从海外云厂商客户画像来看，可大致分为“大型客户+AI原生创企+企业客户”三大客户类型。根据海外云厂商业绩会表述，目前头部客户订单合同金额较高，但客户类型逐步向多元化方向发展。其中，**微软Azure和亚马逊AWS**与大客户合作程度较深，前者受益于OpenAI，后者受益于Anthropic，两者企业客户基础扎实，企业长尾客户需求可观；**谷歌云**则依托前沿AI实验室和AI科研型客户，AI云服务需求旺盛；**Oracle OCI**通过数据库和基础设施双重驱动，围绕数据库等配套产品的企业客户需求具备一定粘性，2025年大客户增量较多；**CoreWeave**主要以大型客户订单为主，同时依托收购措施增强与企业级客户的关系；**Nebius**则以AI原生创企为核心，25H2通过大额订单验证基础设施能力，并开始拓展企业客户。

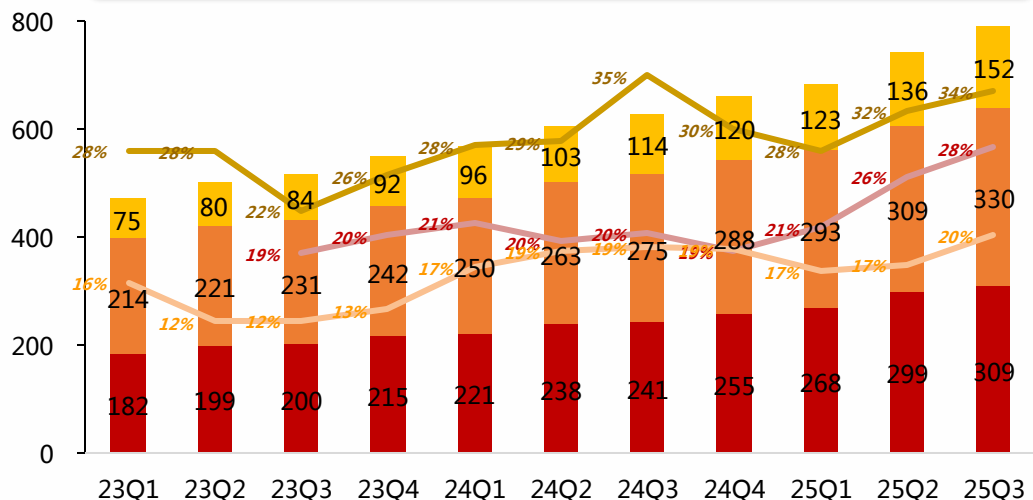
截至25Q3海外云厂商最新客户画像

客户类别	微软	亚马逊	谷歌	甲骨文	CoreWeave	Nebius
大型客户	OpenAI、Meta等	Anthropic等	全球前10大AI Lab中有9家选择Google Cloud；Anthropic也开始用TPU	OpenAI、Meta等	Meta、OpenAI等	Meta、微软
AI原生创企	Cohere、Mistral AI、Stability AI、xAI等	Elastic、12 Labs、Cohere Health、Vercel、Cursor等	AnyScale、Contextual AI、Safe Superintelligence等	Modal Labs等700+ 客户在使用GPU等AI加速器	Mistral、Poolside、Periodic Labs、Jasper、Cohere、Moonvalley、Hippocratic AI等	Lightning AI、DeepSeek、NexaHealth、Cursor AI、Black Forest Labs、World Labs等
企业级客户	Azure AI Foundry 25Q3已有8万家客户使用，包括财富500强中的80%	传统企业上云需求旺盛，同时目前正在构建智能体agents的公司数量非常可观	传统企业正在部署AI模型，同时当前超过70%的GCP现有客户均在使用公司AI产品	传统企业客户数据库等配套产品的需求、传统企业上云需求，特定行业企业专有云需求等	收购Weights & Biases对企业客户群的业务关系起到非常关键的作用	25Q2开始有实质性进展，代表企业包括Shopify、Cloudflare、Prosus、Cisco

4.2 海外AI云业务阶段：从验证期迈向扩张期，未来增长有望加速

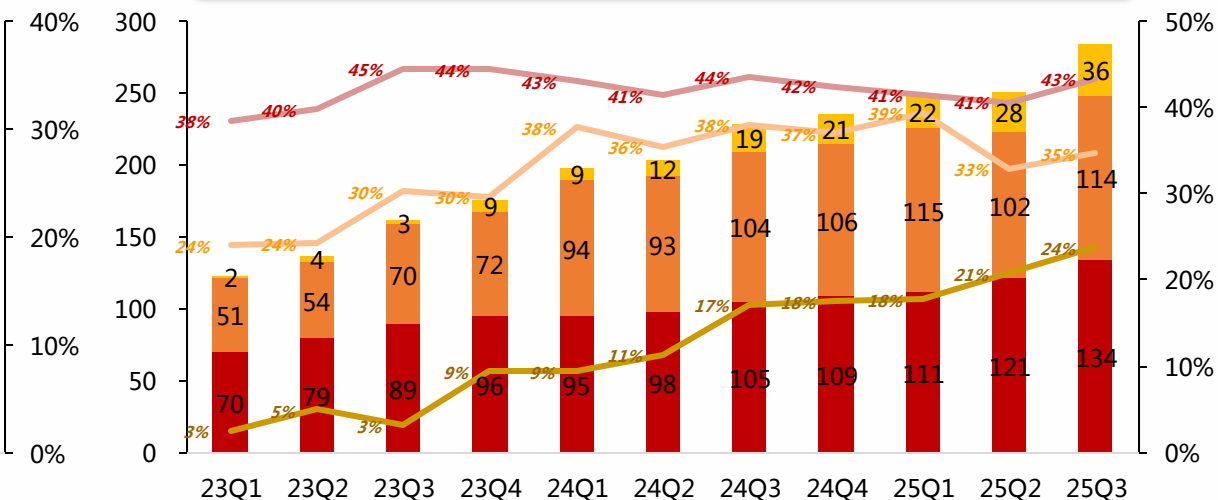
□ **云收入增长提速，经营利润率波动。** 23Q1至25Q3，海外三大云厂商云业务合计收入从471亿美元增长至791亿美元；25Q3合计收入同比增速提升至26%，自23Q3以来创历史新高，云业务增长整体呈现加速态势，其中，谷歌云增速较高、微软智能云收入增速次之、亚马逊AWS在高基数下增速略低。经营利润方面，23Q1至25Q3，海外三大云厂商云业务合计经营利润从123亿美元增长至284亿美元，整体经营利润率从26%提升至36%，其中，谷歌云在规模化扩张和AI业务持续推进下逐步改善，微软智能云和亚马逊AWS经营利润率近年来略有波动，主要系AI基础设施扩展带来折旧摊销成本增加所致。

23Q1-25Q3海外三大云厂商云业务收入及增速



■ 微软智能云收入（左轴，亿美元） ■ 亚马逊AWS收入（左轴，亿美元）
■ 谷歌云收入（左轴，亿美元） ■ 微软智能云收入yoy（右轴）
■ 亚马逊AWS收入yoy（右轴） ■ 谷歌GCP收入yoy（右轴）

23Q1-25Q3海外三大云厂商云业务经营利润及利润率

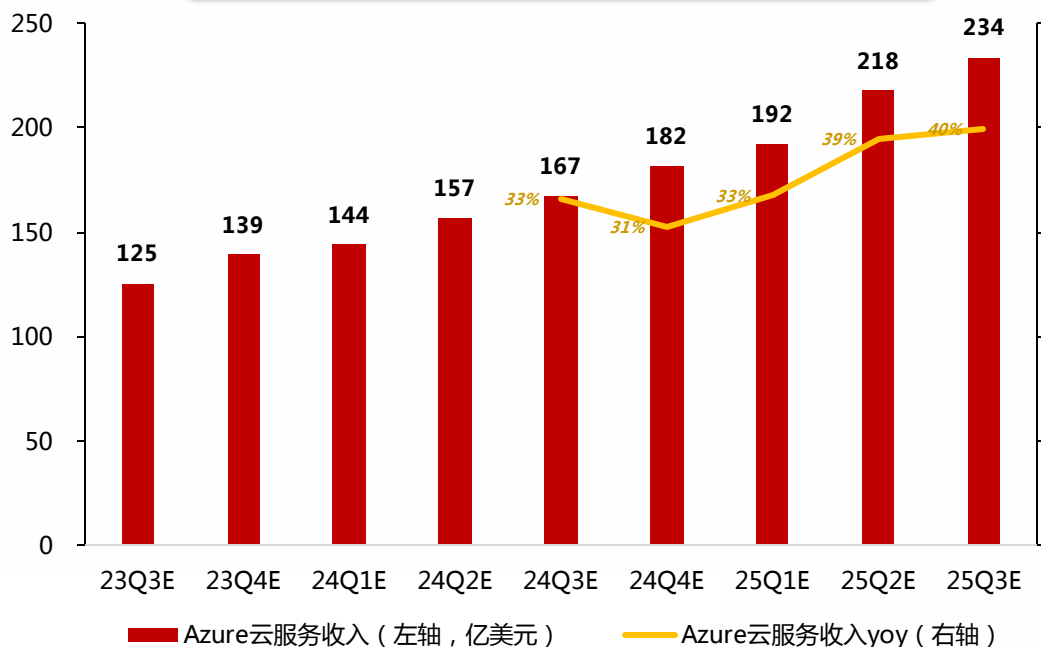


■ 微软智能云经营利润（左轴，亿美元） ■ 亚马逊AWS经营利润（左轴，亿美元）
■ 谷歌云经营利润（左轴，亿美元） ■ 微软智能云经营利润率（右轴）
■ 亚马逊AWS经营利润率（右轴） ■ 谷歌GCP经营利润率（右轴）

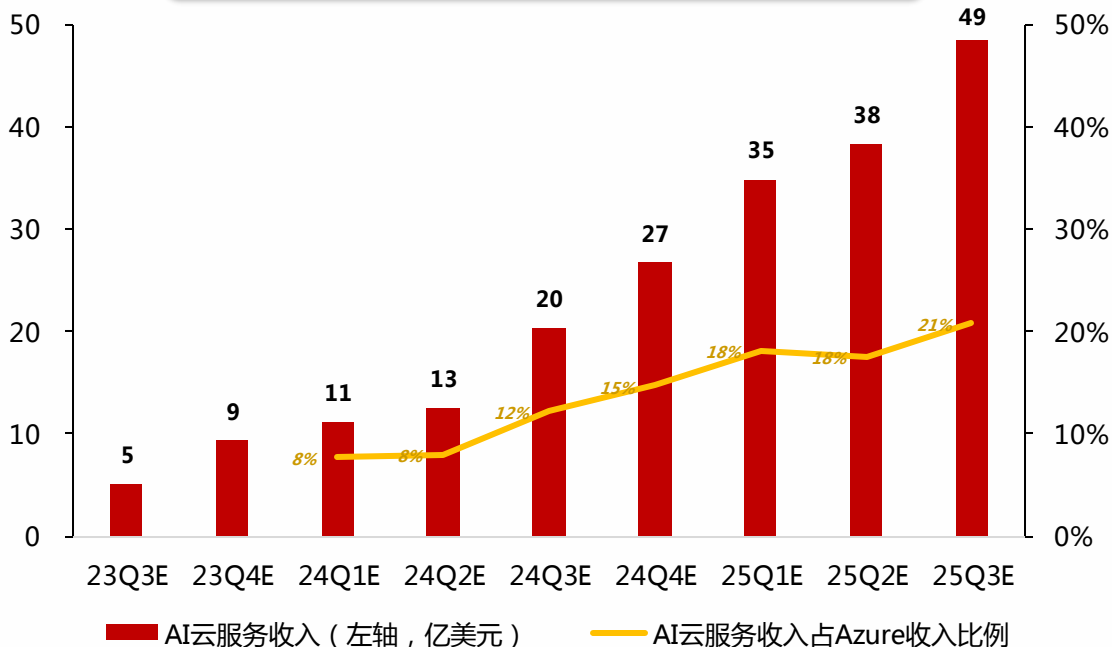
4.2 海外AI云业务阶段：从验证期迈向扩张期，未来增长有望加速

- **AI云服务或已贡献约20%云收入，逐步迈向规模化扩张阶段。**通常来说，在众多行业中，20%的渗透率意味着业务逐步跨过“早期采用者”时期、进入“规模化扩张”阶段，有望形成“需求加速、生态完善、成本下降”的正向循环，推动渗透率向40%~50%迈进。从海外云厂商AI贡献占比来看，截至目前，AI云服务可能已贡献约20%的云收入。
- **微软Azure：**23Q3至25Q1微软Azure云服务收入同比增速中AI云服务贡献的同比增速分别约为5%/9%/10%/11%/12%/13%/16%，若假设25Q2和25Q3微软AI云服务贡献的收入同比增速环比仅增加0.5pp，即约16.5%和17%，则25Q2和25Q3的AI云服务收入体量分别约为38和49亿美元，分别占Azure云业务收入的18%和21%，已达20%的渗透率拐点。

23Q3-25Q3微软Azure云服务收入测算



23Q3-25Q3微软Azure云服务AI贡献测算

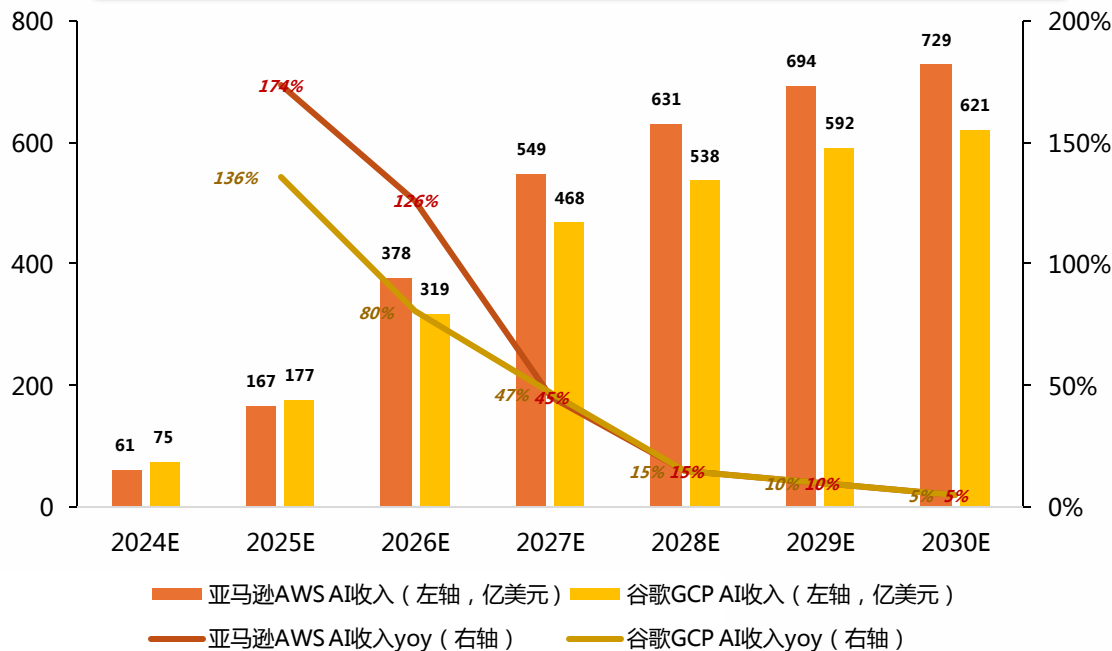


4.2 海外AI云业务阶段：从验证期迈向扩张期，未来增长有望加速

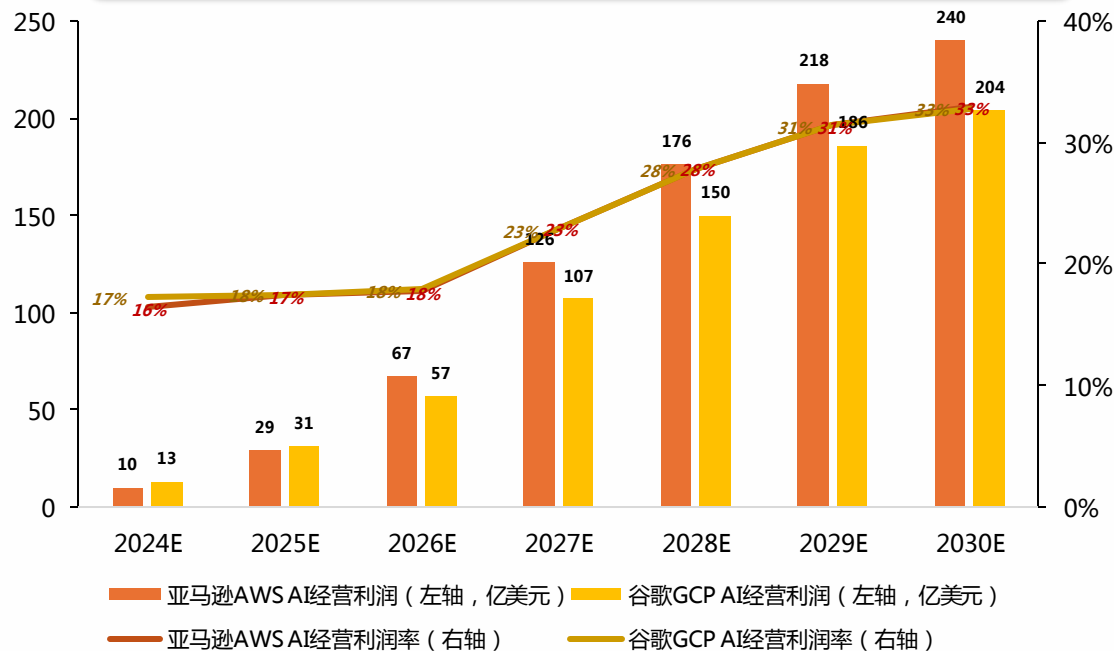
❑ AI云服务或已贡献约20%云收入，逐步迈向规模化扩张阶段。

- **亚马逊AWS**：根据NewStreet Research测算，预计2024年亚马逊AWS中AI实现收入61亿美元，约占2024年AWS全年收入的6%，预计AI云服务经营利润率约16.4%；2025年AI收入有望增长至167亿美元，经营利润率提升至17.4%。
- **谷歌GCP**：根据NewStreet Research测算，预计2024年谷歌云中AI实现收入75亿美元，约占2024年谷歌云全年收入的17%，AI云的经营利润率预计为17.3%；2025年AI云收入有望增至177亿美元，经营利润率预计为17.5%。

2024-2030年亚马逊AWS和谷歌GCP的AI收入预测



2024-2030年亚马逊AWS和谷歌GCP的AI经营利润预测



资料来源：NewStreet Research，西南证券整理

资料来源：NewStreet Research，西南证券整理

4.2 海外AI云业务阶段：从验证期迈向扩张期，未来增长有望加速

❑ **云订单频繁签订，合同额度高企。**2025年，AI云服务逐步进入“超大订单+长期基建”模式，合同金额从十亿级到千亿级不等，合同年限通常设为5至6年，合作范畴涵盖具体的芯片类型、AI模型厂商、电力规模等多种维度，大型云订单也逐步成为海外云厂商剩余履约合同金额（RPO）大增的主要驱动因素，而RPO作为未来云收入的领先指标，其率先实现高增表明未来云厂商的AI云服务可能将迈向加速扩张阶段。此外，根据SemiAnalysis数据，不同云订单对应的息税前利润率各不相同，其中，GCP-Anthropic和OCI-OpenAI两大合作项目的息税前利润率高于其他AI云订单，主要得益于大型集群带来的规模化效应、算力设施建设经验成熟；而CoreWeave/Nebius/Nscale/IREN等新云厂商或第三方数据中心提供方受制于租赁业务模式、定价权相对较弱等原因，利润率较低。

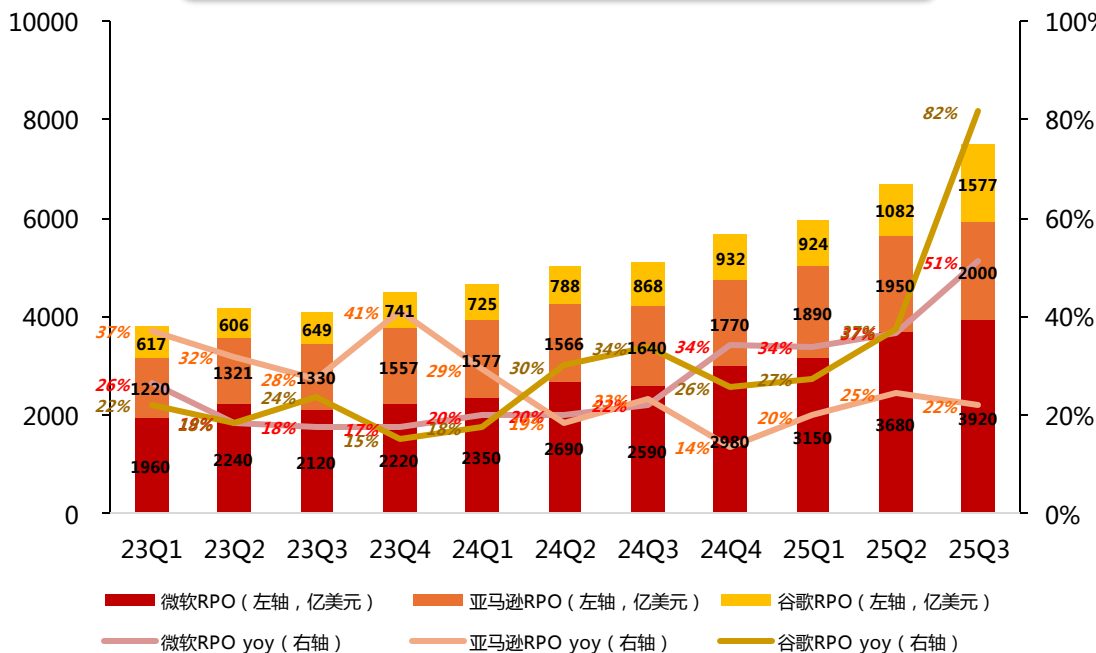
2025年海外主要AI算力大型订单

大型云订单	GCP- Anthropic	IREN- Microsoft	CoreWeave- Meta	OCI- OpenAI	Nebius- Microsoft	CoreWeave- OpenAI ①	CoreWeave- OpenAI ②	CoreWeave- OpenAI ③	Nscale- Microsoft ①	Nscale- Microsoft ②
公告时间	2025/11/9	2025/11/3	2025/9/30	2025/9/10	2025/9/8	2025/3/10	2025/5/15	2025/9/25	2025/9/16	2025/10/15
芯片类型	TPU v7	GB300	GB300*	GB300*	GB300	GB200*	GB200*	GB300*	GB300	GB300
合同总额（亿美元）	420	97	142	3000	174	119	40	65	62	140
合同年限（年）	5	5	6	5	5	5	5	5	5	5
年化收入（亿美元）	84	19	24	600	35	24	8	13	12	28
关键IT电力（MW）	788	170	230	4500	300	195	65	105	116	260
按照合同年限折旧下的 EBIT利润率	37.5%	20.0%	17.6%	37.1%	17.1%	24.5%	25.0%	24.6%	16.7%	17.3%

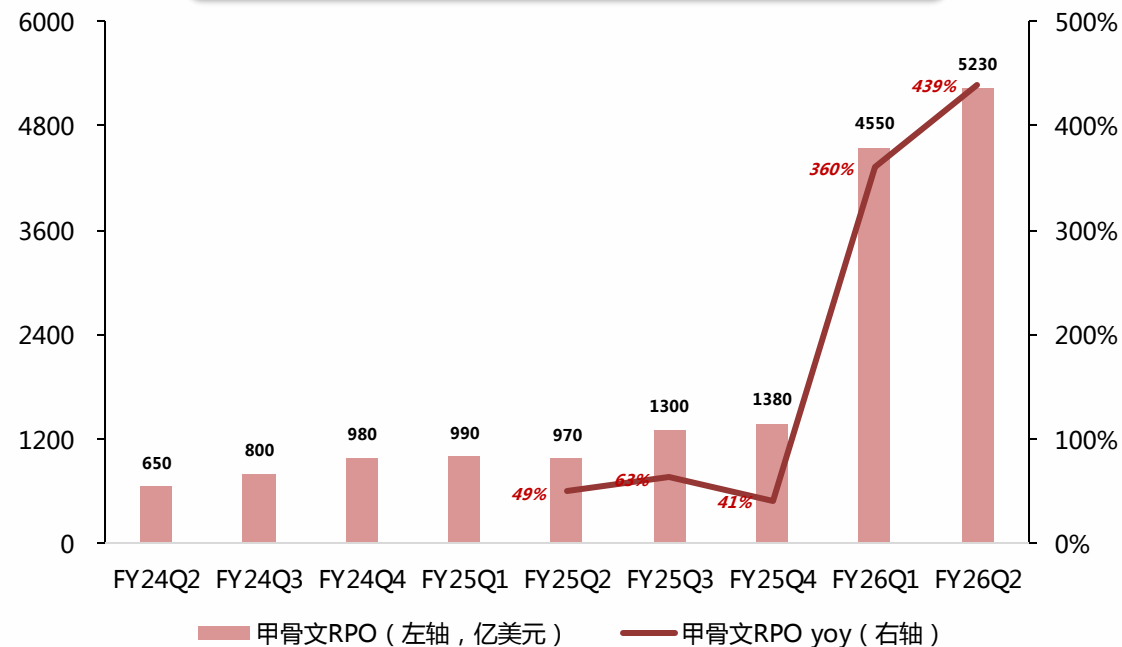
4.2 海外AI云业务阶段：从验证期迈向扩张期，未来增长有望加速

- **云积压订单高增，云需求信号强劲。** 23Q1至25Q3，海外云厂商剩余履约合同金额持续增长，其中，多数海外云厂商25Q3积压订单环比大增、并表示于25Q4签署大型订单。**1) 微软：**RPO增长来自众多客户，OpenAI只是其中一部分，截至25Q3的积压订单不包含10月宣布的OpenAI额外2500亿美元Azure承诺。**2) 亚马逊：**截至25Q3的积压订单不包括10月签署的几笔未披露新交易，这些交易规模合计已超过整个25Q3的成交量。**3) 谷歌：**积压订单的爆炸式增长主要由企业级AI的强劲需求所驱动；2025年前九个月，谷歌云签署的价值超过10亿美元的大额合同数量，已超过前两年总和，战略性大单是推高积压订单金额的直接原因。**4) 甲骨文：**RPO出现爆炸性增长主要得益于签下OpenAI、xAI、Meta等顶级客户及许多其他公司的云订单。**5) Coreweave：**25Q3新增积压订单超过250亿美元，主要来自与Meta等大客户的新合同。**6) Nebius：**2025年9月和11月分别与微软和Meta签署大单。

23Q1-25Q3海外三大云厂商剩余履约合同



FY24Q2-FY26Q2甲骨文剩余履约合同



4.3 海外云厂商算力规划：加速锁定电力容量，业务增速依赖产能上线节奏

❑ **算力产能供不应求，加速锁定电力容量。**从海外云厂商表态来看，各公司均表示当前算力产能供不应求，正加速锁定未来电力容量，且基本已锁定数吉瓦级别。**1) 微软**：2025年微软AI算力总量将增加80%以上，并计划在未来两年将数据中心总规模翻倍；**2) 亚马逊**：根据25Q3业绩会，亚马逊在过去一年里新增3.8GW产能，并预计到2027年底整体产能翻倍。**3) 谷歌**：暂无披露。**4) 甲骨文**：FY26Q1向客户交付400MW数据中心产能。**5) CoreWeave**：截至25Q3，公司签约电力总容量增加至2.9GW。**6) Nebius**：预计2026年签约电力总容量将增长至2.5GW。

海外云厂商有关数据中心“供不应求”和“电力容量”表述

AI云服务	“云业务供不应求”相关表态	“数据中心电力容量”相关表述
微软	根据25Q3业绩会，公司预计 至少到26H1末仍将面临容量限制 ；而此前25Q2业绩会上表示，预计到25H2末持续受限，算力容量限制时间预期延长。	根据25Q2业绩会， 过去的12个月中，新增的数据中心产能就超过2GW 。根据25Q3业绩会， 2025年微软AI算力总量将增加80%以上 ，并计划在未来两年将数据中心总规模翻倍；且25Q4宣布将在威斯康星州建设全球最强大的AI数据中心Fairwater， 预计2026年上线，单体规模可达2GW 。
亚马逊	根据25Q3业绩会， 公司将继续积极投资扩充产能，因为看到强劲需求 。根据25Q2业绩会，目前 云服务的需求超过现有产能 ，公司预计短期内（未来几个季度）难以完全解决产能不足，但预计未来将逐季改善、对改善趋势保持乐观态度。	根据25Q3业绩会， 在过去一年里新增3.8GW产能，25Q4还会再增加超过1GW 。预计到 2027年底，整体产能将翻倍 。
谷歌	公司将努力增加产能并提高服务器部署速度，但预计 进入2026年时供需紧张仍将持续 。	/
甲骨文	根据FY26Q2业绩会， 公司持续面临显著且前所未有的云服务需求 。	在FY26Q1，公司向客户交付接近400MW的数据中心产能。根据FY25Q3业绩会，公司已签约的电力容量增长速度甚至快于数据中心数量的增长， 预计在2025年自然年内可用电力容量将翻倍，并在FY2026财年结束前实现三倍增长 。
CoreWeave	根据25Q3业绩会，公司表示自身依然处于 供应极度紧张 的市场环境中，对CoreWeave的AI云平台的需求远远超过现有产能。	根据25Q3业绩会，公司 有效供电规模 环比增加120MW， 达到约590MW ，同时 签约的电力总容量增加至2.9GW ，其中超过1GW的合同容量将在未来12至24个月陆续上线。预计 2025年年底有效供电规模将超过850MW 。
Nebius	根据25Q3业绩会，公司表示 需求非常强劲，所有可用算力容量已经售罄 （GPU产能已全部售罄，25Q4产能也接近售罄）；公司持续看到一个非常一致的趋势：每当公司上线新的算力产能，就会被迅速卖光。	根据25Q3业绩会，算力产能如今已成为营收增长的主要瓶颈，预计 到2025年年底，公司拥有220MW的已接入电力；预计到2026年底，有约800MW至1GW的接入电力；预计2026年签约电力总量将增长至2.5GW ，较25Q2提到的1GW有显著提升。

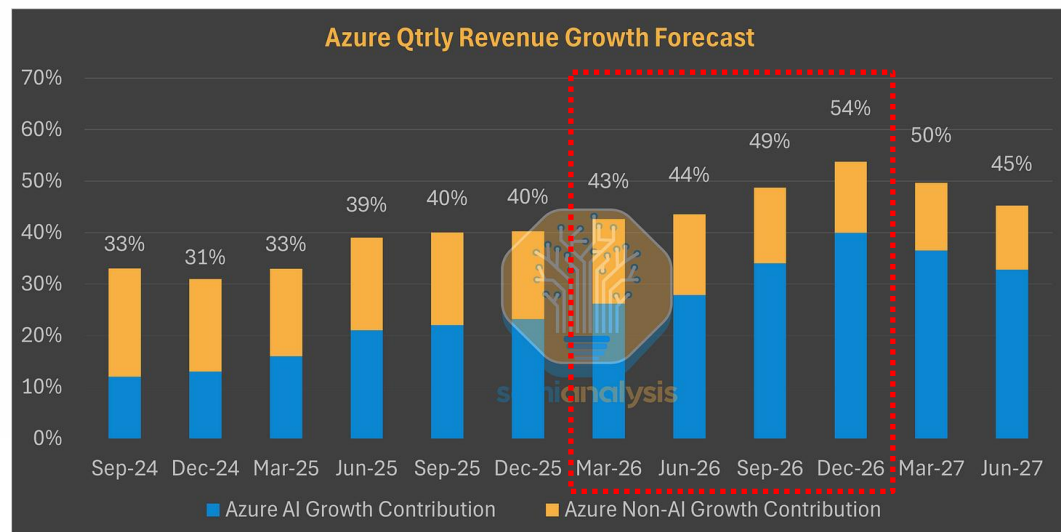
资料来源：各公司公告，西南证券整理

4.3 海外云厂商算力规划：加速锁定电力容量，业务增速依赖产能上线节奏

□ 未来云业务收入增速高度依赖产能上线节奏，2026年云业务伴随电力容量释放有望迎来增长加速。

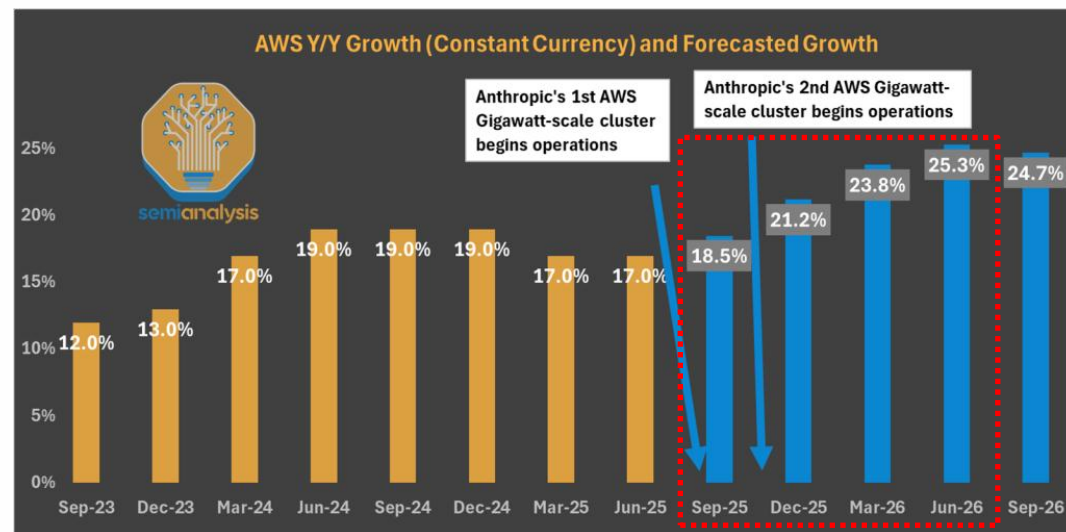
- **微软**：公司于25Q4宣布将在威斯康星州建设全AI数据中心Fairwater，预计2026年上线，单体规模可达2GW。根据SemiAnalysis预测，微软Azure云将于26Q1至26Q4迎来收入增长加速期，同比增速预计分别提升至43%/44%/49%/54%，Azure云服务有望连续四个季度实现加速增长。
- **亚马逊**：公司预计25Q4将增加超过1GW数据中心产能，Anthropic已在使用Project Rainier集群来训练和部署其领先的AI模型，预计到2026年底将运行超过100万枚Trainium2芯片，Trainium2已成为一个数十亿美元级业务。根据SemiAnalysis预测，随着亚马逊算力容量持续释放，AWS有望在25Q3至26Q2迎来加速增长，同比增速预计分别为18.5%/21.2%/23.8%/25.3%，其中，根据25Q3业绩公告，亚马逊AWS的25Q3实际收入同比增速（yoy+20%）已超出SemiAnalysis预计增速（yoy+18.5%）。

微软Azure云业务收入增速预测



资料来源：SemiAnalysis，西南证券整理

亚马逊AWS云业务收入增速预测



资料来源：SemiAnalysis，西南证券整理

相关标的

海外科技大厂及AI独角兽AI投入预期高企，未来现金流压力加大

AI投资从FOMO CapEx转向ROI CapEx

数据中心层面：

在面临电力容量限制下，数据中心建设更加强调co-design和fungible，以提升每瓦特tokens产出效率

芯片
英伟达
博通

存储
美光
闪迪

光通信
LITE
CRDO

AI模型与应用层面：

工程化能力持续提升，长文本和智能体成为重要迭代方向，AI商业化变现诉求增强

互联网大厂
(AI+广告) 谷歌
Meta

①Data infra
②高ROI场景(编程/客服/多模态)

关注AI独角兽IPO进展
OpenAI
Anthropic

AI云服务层面：

海外云厂商加速锁定电力容量，云业务增速高度依赖算力产能上线节奏，AI云服务未来增长有望加速

Tier1 云厂商
商：
亚马逊
谷歌
微软

Tier2 云厂商
商：
甲骨文
CoreWeave
Nebius

风险提示

- AI投资带来现金流压力风险；
- AI云业务增长不及预期风险；
- AI应用商业化变现不及预期风险等。



西南证券
SOUTHWEST SECURITIES

分析师：王湘杰
执业证号：S1250521120002
电话：0755-26671517
邮箱：wxj@swsc.com.cn

分析师：尤品柯
执业证号：S1250525050001
邮箱：ypk@swsc.com.cn

西南证券投资评级说明

报告中投资建议所涉及的评级分为公司评级和行业评级（另有说明的除外）。评级标准为报告发布日后6个月内的相对市场表现，即：以报告发布日后6个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。其中：A股市场以沪深300指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以恒生指数为基准；美国市场以纳斯达克综合指数或标普500指数为基准。

公司评级

买入：未来6个月内，个股相对同期相关证券市场代表性指数涨幅在20%以上
持有：未来6个月内，个股相对同期相关证券市场代表性指数涨幅介于10%与20%之间
中性：未来6个月内，个股相对同期相关证券市场代表性指数涨幅介于-10%与10%之间
回避：未来6个月内，个股相对同期相关证券市场代表性指数涨幅介于-20%与-10%之间
卖出：未来6个月内，个股相对同期相关证券市场代表性指数涨幅在-20%以下

行业评级

强于大市：未来6个月内，行业整体回报高于同期相关证券市场代表性指数5%以上
跟随大市：未来6个月内，行业整体回报介于同期相关证券市场代表性指数-5%与5%之间
弱于大市：未来6个月内，行业整体回报低于同期相关证券市场代表性指数-5%以下

分析师承诺

报告署名分析师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，报告所采用的数据均来自合法合规渠道，分析逻辑基于分析师的职业理解，通过合理判断得出结论，独立、客观地出具本报告。分析师承诺不曾因，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获取任何形式的补偿。

重要声明

西南证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会核准的证券投资咨询业务资格。

本公司与作者在自身所知情范围内，与本报告中所评价或推荐的证券不存在法律法规要求披露或采取限制、静默措施的利益冲突。

《证券期货投资者适当性管理办法》于2017年7月1日起正式实施，本公司签约客户使用，若您并非本公司签约客户，为控制投资风险，请取消接收、订阅或使用本报告中的任何信息。本公司也不会因接收人收到、阅读或关注自媒体推送本报告中的内容而视其为客户。本公司或关联机构可能会持有报告中提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行或财务顾问服务。

本报告中的信息均来源于公开资料，本公司对这些信息的准确性、完整性或可靠性不作任何保证。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可升可跌，过往表现不应作为日后的表现依据。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保持在最新状态。同时，本公司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。

参考之用，不构成出售或购买证券或其他投资标的的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，本公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

本报告及附录版权为西南证券所有，未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布。如引用须注明出处为“西南证券”，且不得对本报告及附录进行有悖原意的引用、删节和修改。未经授权刊载或者转发本报告及附

西南证券研究院

上海

地址：上海市浦东新区陆家嘴21世纪大厦10楼

邮编：200120

北京

地址：北京市西城区金融大街35号国际企业大厦A座8楼

邮编：100033

深圳

地址：深圳市福田区益田路6001号太平金融大厦22楼

邮编：518038

重庆

地址：重庆市江北区金沙门路32号西南证券总部大楼21楼

邮编：400025

西南证券机构销售团队

区域	姓名	职务	手机	邮箱	姓名	职务	手机	邮箱
上海	蒋诗烽	院长助理、研究销售部经理、 上海销售主管	18621310081	jsf@swsc.com.cn	欧若诗	销售岗	18223769969	ors@swsc.com.cn
	崔露文	销售岗	15642960315	clw@swsc.com.cn	蒋宇洁	销售岗	15905851569	jyj@swsc.com.cn
	李煜	销售岗	18801732511	yfliyu@swsc.com.cn	贾文婷	销售岗	13621609568	jiawent@swsc.com.cn
	汪艺	销售岗	13127920536	wyyf@swsc.com.cn	张嘉诚	销售岗	18656199319	zhangjc@swsc.com.cn
	戴剑箫	销售岗	13524484975	daijx@swsc.com.cn	毛玮琳	销售岗	18721786793	mwl@swsc.com.cn
	李嘉隆	销售岗	15800507223	ljlong@swsc.com.cn				
北京	李杨	北京销售主管	18601139362	yfly@swsc.com.cn	王宇飞	销售岗	18500981866	wangyuf@swsc.com.cn
	张岚	销售岗	18601241803	zhanglan@swsc.com.cn	王一菲	销售岗	18040060359	wyf@swsc.com.cn
	姚航	销售岗	15652026677	yhang@swsc.com.cn	张鑫	销售岗	15981953220	zhxin@swsc.com.cn
	杨薇	销售岗	15652285702	yangwei@swsc.com.cn	马冰竹	销售岗	13126590325	mbz@swsc.com.cn
广深	龚之涵	销售岗	15808001926	gongzh@swsc.com.cn	文柳茜	销售岗	13750028702	wlq@swsc.com.cn
	唐茜露	销售岗	18680348593	txl@swsc.com.cn	林哲睿	销售岗	15602268757	lzh@swsc.com.cn