AI 应用迎来多重催化，科技股受债市影响承压

2026 年 08 月 23 日

核心观点

- **双周主要指数整体回调，科技板块回调幅度强于大盘：**双周（2026.08.10-2026.08.21）主要指数中，S&P500 涨跌幅为-1.07%，纳指涨跌幅为-1.91%，道指涨跌幅为-1.00%；科技行业方面，以科技成长股为主的 Nasdaq-100 指数涨跌幅为-1.39%，费指涨跌幅为-4.99%，回调幅度强于大盘。年初至今涨跌幅方面，费指涨幅为 65.75%，仍明显好于大盘，且高于 2025 年同期（+12.52%）及 2024 年同期（+26.16%）水平，表明半导体板块中长期交易热度仍处于高位，双周回调主要系美债收益率上行，通胀担忧加剧引发降息预期降温，降低压制股市风险偏好。子行业方面，Nasdaq-100 中，双周涨幅靠前标的主要集中于存储、AI 应用等板块；双周跌幅靠前公司主要集中于半导体及硬件设备等板块。我们判断主要因素为，受部分 AI 应用公司业绩兑现影响，以及 OpenAI 增速放缓降低了先前 AI 取代 SaaS 行业的悲观情绪，AI 应用板块迎来多重催化；此外受债市影响，科技板块整体、尤其是半导体设备板块承压明显。

- **核心观点：**

  英伟达于 8 月 10 日联合 Apollo、贝莱德、黑石、Brookfield、高盛、KKR 六家机构签署谅解备忘录，拟搭建独立融资平台、长期撬动超 5000 亿美元第三方资本，将 GPU 纳入全球长期资本可投资的基础设施资产类别；8 月 17 日英伟达进一步锁定俄亥俄州 PORTS-Pike 科技园区的土地、电力与厂房资源，OpenAI 作为租户承租建设 AI 工厂（初始 4.25 IT-GW、可扩至 8.00 IT-GW），租期 20 年，英伟达同步向 SB Energy 投资 15 亿美元。**我们认为，供给侧层面，AI 工厂的竞争瓶颈已从先进制程外溢至土地、电力、厂房（LPS）等物理资源，LPS 提前锁定的重要性显著提升；英伟达以自身信用为数据中心资产融资增信，AI 基础设施的供给弹性将系统性抬升。产业链层面，算力资产化趋势将把订单与估值弹性从芯片设计进一步向 LPS 资源持有方、数据中心融资服务、能源配套及算力网络互连环节扩散，但同时现金流稳定程度可能在债务周期末端影响系统性风险。**

  OpenAI 于 8 月 21 日在官方开发者社区公告，未来三个月内将 GPT-5.6 Sol 模型的 API 及 Credit 价格下调超过 20%，其中输入价格降至每百万 token 4 美元、输出价格降至每百万 token 20 美元。**我们认为此次降价直接打通了 Agentic Workflow、长上下文理解、自主代码生成等高输出密度场景的单位经济模型（Unit Economics），将触发推理算力需求从训练集群向端侧与边缘侧扩散。对产业链而言，价值量占比将从训练芯片向推理芯片、推理优化框架、高速互连及 AI 应用层系统性迁移；同时，模型层利润率的压缩将倒逼上游算力侧进一步降低 TCO（总拥有成本），CPO、液冷等环节的战略价值持续抬升。**

- **投资建议：**建议关注相关龙头标的：光模块/光通信相关标的：Lumentum（LITE）,Coherent（COHR），Fabrinet（FN），Corning Inc.（GLW），Applied Optoelectronics（AAOI）等；设备制造相关标的：Ciena（CIEN），Cisco Systems（CSCO），Micron Technology（MU）等；运营商相关标的：AT&T（T），Verizon（VZ），T-Mobile US（TMUS）等；行业相关标的：Oracle（ORCL），Nvidia（NVDA）等。

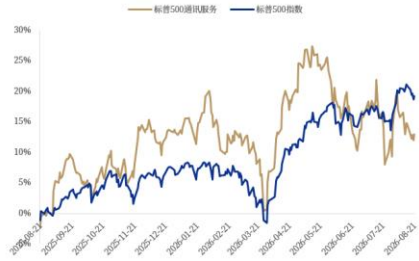
海外算力行业

推荐 维持评级

分析师

赵良毕
☎：010-8092-7619
✉：zhaoliangbi_yj@chinastock.com.cn
分析师登记编码：S0130522030003

相对标普 500 表现图 2026 年 08 月 21 日



资料来源：中国银河证券研究院

风险提示

1、技术落地与良率不及预期的风险；
2、CSP 厂商 Capex 不及预期的风险；
3、竞争加剧的风险；
4、政策摩擦的不确定性风险。

相关研究

1.【银河海外算力】行业深度_磷化铟：光互连“隐形基石”的产业链卡位与价值重估
2.【银河海外算力】行业深度_磷化铟供需缺口下的光器件产业变局和投资机遇
3.【银河海外算力】行业点评_住友、AXT 净利润激增，通美预备港股上市

仅供内部参考，请勿外传！

目录
Catalog

一、 双周行情回顾 3
（一） 海外主要指数回顾 3
（二） 各子行业表现情况 3
（三） 双周行业龙头标的表现 4
二、 双周主要事件梳理 5
（一） 设备制造 5
（二） 卫星通信 6
（三） 运营商 6
（四） 算力服务 6
（五） 液冷 7
（六） 行业应用 7
三、 行业数据跟踪 8
（一） 大模型 Tokens 调用量 8
（二） 热门 GPU 价格走势 9
（三） 子行业技术趋势 9
四、 行业大事前瞻 10
五、 投资建议 11
六、 风险提示 11

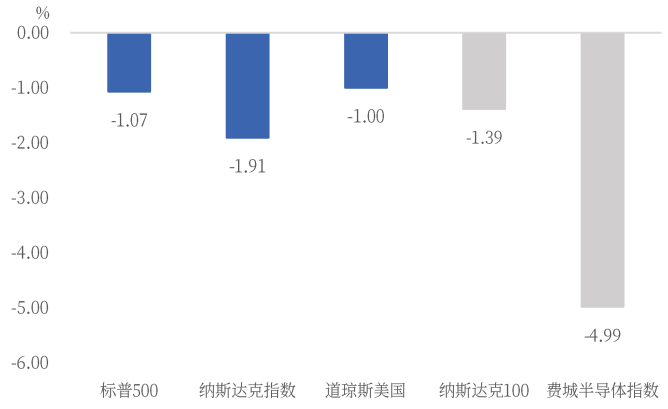
仅供内部参考，请勿外传！

一、双周行情回顾

（一）海外主要指数回顾

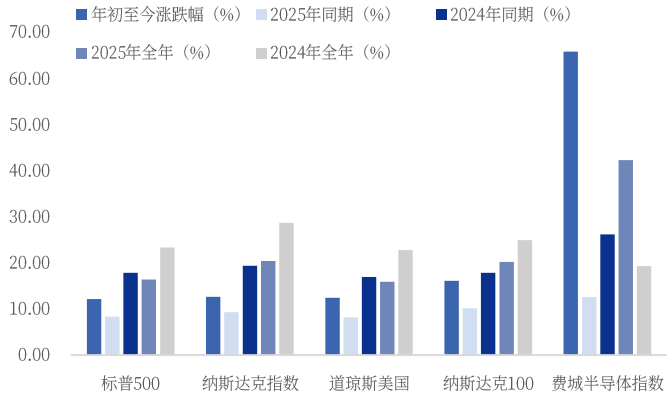
双周主要指数整体回调，科技板块回调幅度强于大盘：双周（2026.08.10-2026.08.21）主要指数中，S&P500 涨跌幅为-1.07%，纳指涨跌幅为-1.91%，道指涨跌幅为-1.00%；科技行业方面，以科技成长股为主的 Nasdaq-100 指数涨跌幅为-1.39%，费指涨跌幅为-4.99%，回调幅度强于大盘。年初至今涨跌幅方面，费指涨幅为 65.75%，仍明显好于大盘，且高于 2025 年同期（+12.52%）及 2024 年同期（+26.16%）水平，表明半导体板块中长期交易热度仍处于高位，双周回调主要系美债收益率上行，通胀担忧加剧引发降息预期降温，降低压制股市风险偏好。

图1：各主要指数双周涨跌幅对比



资料来源：Wind，中国银河证券研究院

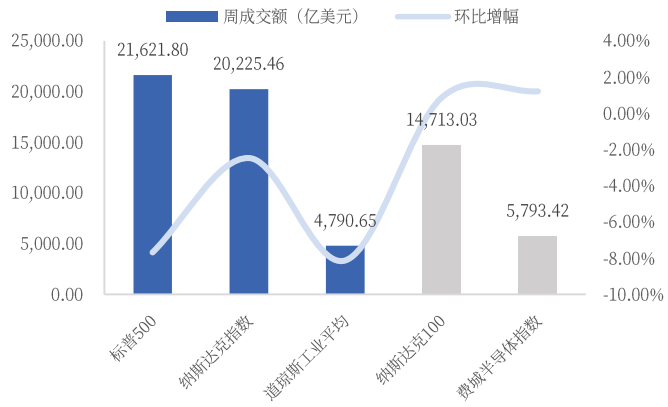
图2：各主要指数年初至今涨跌幅对比



资料来源：Wind，中国银河证券研究院

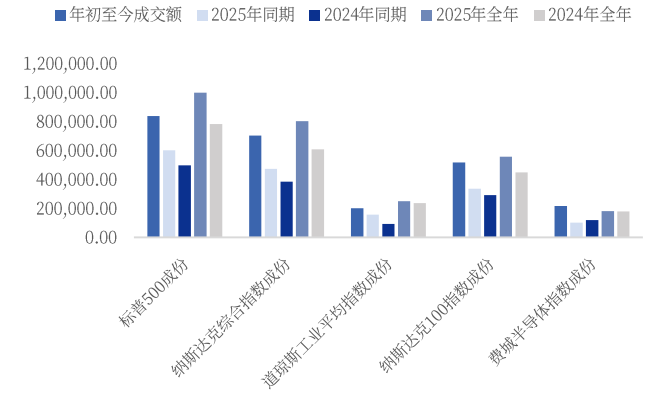
成交额方面，双周科技指数成交额绝对值不及大盘，标普 500、纳指及 Nasdaq-100 成交额分别为 21621.80 亿美元、20225.46 亿美元、14713.03 亿美元；环比来看，标普 500、纳指及道指成交额较前期均回落，Nasdaq-100 及费指小幅上升。相比前两年同期，当前年度成交额整体仍处于较高水平，市场交投活跃度好于 2025 及 2024 年同期水平。

图3：各主要指数双周成交额对比



资料来源：Wind，中国银河证券研究院

图4：各主要指数年初至今成交额对比（亿美元）



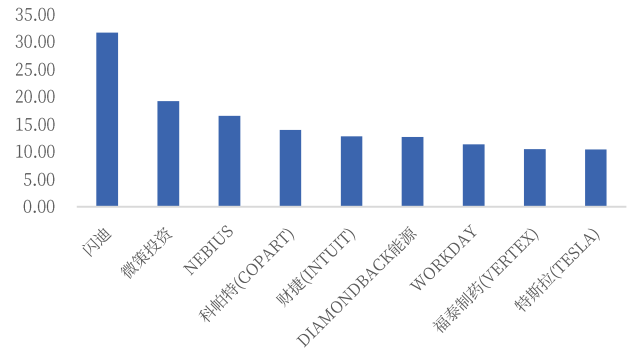
资料来源：Wind，中国银河证券研究院

（二）各子行业表现情况

仅供内部参考，请勿外传！

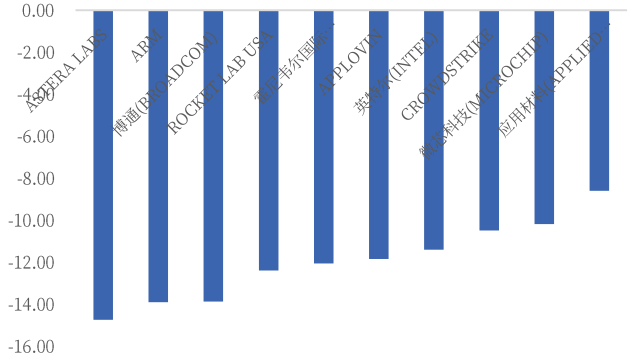
子行业方面，Nasdaq-100 中，双周涨幅靠前标的主要集中于存储、AI 应用等板块；双周跌幅靠前公司主要集中于半导体及硬件设备等板块。我们判断主要因素为，受部分 AI 应用公司业绩兑现影响，以及 OpenAI 增速放缓降低了先前 AI 取代 SaaS 行业的悲观情绪，AI 应用板块迎来多重催化；此外受债市影响，科技板块整体、尤其是半导体设备板块承压明显。

图5：Nasdaq-100 中双周涨幅前十名标的汇总（%）



资料来源：Wind，中国银河证券研究院

图6：Nasdaq-100 中双周跌幅前十名标的汇总（%）

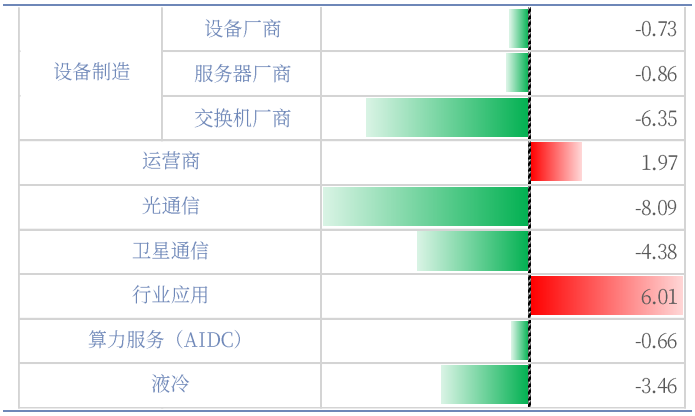


资料来源：Wind，中国银河证券研究院

（三）双周行业龙头标的表现

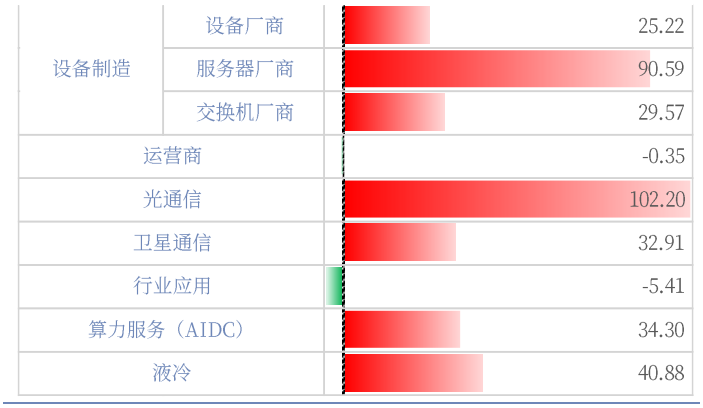
根据我们对于海外算力行业划分子板块，海外算力板块子行业包括设备制造（设备厂商、服务器厂商、交换机厂商）、运营商、光通信、卫星通信、行业应用、算力服务（AIDC）、液冷七大板块。双周内板块普遍回调，仅运营商与行业应用板块录得上涨：行业应用板块涨幅为 6.01%，运营商板块涨幅为 1.97%；回调板块方面，跌幅居前的板块分别为光通信、交换机厂商和卫星通信，跌幅分别为 8.09%、6.35%、4.38%。

图7：双周细分子板块涨跌幅情况（%）



资料来源：Wind，中国银河证券研究院

图8：年初至今子板块涨跌幅（%）



资料来源：Wind，中国银河证券研究院

上述行业分类包含的公司中，双周涨幅居前的行业龙头标的主要包括 Supermicro、Nebius、Figma、Hubspot、Workday 等，涨幅分别为 19.63%、16.58%、16.36%、14.05%、11.34%。回调标的按照跌幅降序分别为 Coherent、Fabrinet、TSS、AXT、OCC，跌幅分别为 23.64%、22.35%、22.29%、20.14%、14.77%，主要集中于光通信板块。

仅供内部参考，请勿外传！

二、双周主要事件梳理

（一）设备制造

■ NVIDIA（NVDA.O）：携手六大金融机构搭建 AI 算力融资平台，拟撬动超 5000 亿美元第三方资本

NVIDIA 于 8 月 10 日宣布，已与阿波罗全球管理（Apollo）、贝莱德（BlackRock）、黑石（Blackstone）、布鲁克菲尔德（Brookfield）、高盛（Goldman Sachs）及 KKR 六家全球顶级金融机构签署谅解备忘录，共同打造独立的 AI 算力融资平台，目标逐步调动超过 5000 亿美元的第三方资本，用于 AI 基础设施建设。平台将 NVIDIA 算力与全栈 AI 基础设施转化为全球资本可投资的资产类别，为前沿 AI 实验室、企业与 AI 云服务商提供规模化融资支持。

公司 CEO 黄仁勋表示，AI 时代"算力就是营收"，NVIDIA 算力凭借广泛采用、灵活复用及 CUDA 软件生态的持续升级，天然适合作为基础设施资产进行融资。此次合作将引入机构信贷、保险资金与私募资本，帮助客户在不完全依赖自身投入的前提下大规模获取稀缺算力资源，并减轻其资产负债表压力。贝莱德 CEO 拉里·芬克认为，数据中心融资才刚刚起步，有望成为金融工程的下一个前沿领域。

该平台的落地标志着 AI 基础设施融资模式的重要转变：NVIDIA 算力被纳入全球长期资本可投资的基础设施资产类别，面向 AI 数据中心建设的机构信贷、保险资金与私募资本有望持续流入，算力正从"芯片采购"向"基础设施资产"演进。

■ Super Micro（SMCI.O）：发布 2026 财年第四季度财报，新财年营收指引达 650 亿至 720 亿美元

Supermicro 于 8 月 11 日发布 2026 财年第四季度财报，季度净销售额 111.20 亿美元（公司披露$11,119.8 百万），同比增长 93.15%（去年同期$5,756.9 百万），环比增长 9.02%；非 GAAP 每股收益 1.70 美元，非 GAAP 毛利率 17.60%，较上季度（10.10%）大幅修复。公司 2026 财年全年净销售额 390.63 亿美元，同比增长 78.00%。公司受益于 AI 训练与推理算力基础设施的爆发式需求，服务器订单持续放量。

公司给出的 2027 财年第一季度营收指引为 145 亿至 155 亿美元，非 GAAP 摊薄每股收益指引为 1.01 至 1.10 美元（GAAP 口径为 0.89 至 0.98 美元）；公司进一步预计，2027 财年全年净销售额将达 650 亿至 720 亿美元。公司 CEO Charles Liang 表示，公司过去一年新增数百家企业及其他客户，创造逾 600.00 亿美元新订单，并以创纪录的积压订单进入 2027 财年；技术领先、制造规模与全球合规方面的持续投入正帮助客户更快速、更高效地部署 AI 基础设施。

公司官方同时披露，董事会正就与出口管制相关的特定交易开展独立审查，审查结果可能影响相关财务预测及历史期间业绩。公司上一财季经营现金流由负转正，达 7.47 亿美元，前期库存与产能投入正逐步回笼变现。

■ NVIDIA（NVDA.O）：锁定俄亥俄州 PORTS-Pike 园区 LPS 资源，OpenAI 将承租建设 AI 工厂

NVIDIA 于 8 月 17 日宣布，与能源开发商 SB Energy 达成合作，锁定位于美国俄亥俄州派克县 PORTS-Pike 科技园区的土地、电力与厂房（LPS）资源，该园区将专门用于承载 NVIDIA 算力系统，OpenAI 将作为租户在该场地建造并运营 AI 工厂，租期 20 年。公司 CEO 黄仁勋同日发表署名文章《保障智能基础设施安全》，指出 AI 基础设施竞争焦点已从先进芯片延伸至土地、电力与厂房。

仅供内部参考，请勿外传！

该园区初始部署规模为 4.25 IT-GW 的 AI 工厂产能，NVIDIA 保留将容量扩至 8 IT-GW 的选择权。NVIDIA 将向 SB Energy 投资 15 亿美元以支持其业务扩张，并为其约 4 吉瓦 LPS 基础设施提供长达 20 年的期权支持（限定于租赁与电力支付的特定部分）。据公司披露，PORTS-Pike 部署的每一代 AI 工厂系统约相当于 150 万张 NVIDIA GPU，对应约 1500 亿至 2000 亿美元收入；到 2030 年，OpenAI 已承诺部署的 NVIDIA AI 基础设施总量约 12 吉瓦（若扩展 PORTS-Pike 项目可增至 16 吉瓦），对应总计约 6000 亿美元的算力商业机会。

该合作由软银集团旗下 SB Energy 主导开发，规划新建至少 10 吉瓦发电能力并投资至少 42 亿美元于区域电网基础设施，产能预计 2028 年起分阶段上线。黄仁勋明确表示该安排并非"循环融资"，OpenAI 将自行支付租金，NVIDIA 的担保风险敞口将随租约履行逐年递减。这一模式标志着算力巨头正深度介入 AI 基础设施的"开发-建设-融资"环节。

（二）卫星通信

■ **Rocket Lab（RKLB.O）：完成第 93 次 Electron 发射，为 iQPS 部署第 9 颗 SAR 卫星**

Rocket Lab 于 8 月 20 日成功执行第 93 次 Electron 火箭发射任务（代号"The Lightning God Defends"），从新西兰发射场 1 号复合体升空，将日本 iQPS 公司的 QPS-SAR 卫星（SUSANOO-II）部署至约 575 公里近地轨道。该任务系 2026 年以来 Electron 的第 14 次发射，也是 Rocket Lab 为 iQPS 执行的第九次 QPS-SAR 卫星部署，全部实现 100%任务成功。

iQPS 正在建设由 36 颗合成孔径雷达（SAR）卫星组成的对地观测星座，SAR 技术可在昼夜及云层条件下获取高分辨率地表图像。本次发射后，iQPS 仍有 9 次专用发射预订于 Electron，计划于 2030 年前完成星座组网，Electron 持续担任其主力发射工具。此前 8 月 6 日，Rocket Lab 已完成年内第 13 次发射并为 iQPS 部署第 8 颗 SAR 卫星。

持续高频的发射节奏验证了商业小卫星发射需求的景气度，Rocket Lab 在 8 月单月完成两次任务，Neutron 中型火箭的研发进展及后续发射合同执行情况将于日后更新。

（三）运营商

■ **AT&T（T.N）：押注开源 AI 模型降本增效，智能路由管理每日 450.00 亿 Token 推理调用**

AT&T 首席数据与 AI 官 Andy Markus 于 8 月 12 日接受《华尔街日报》采访时表示，开源模型目前约占公司整体 AI 使用量的 25%，公司目标将这一比例提升至 70%-80%。AT&T 平均每日消耗约 450 亿个 AI tokens，已自建"智能路由器"系统，自动为具体任务选择最具成本效益的模型；从闭源专有模型切换至开源模型，已在部分应用场景中实现 80%-90%的成本节约。

公司内部拥有超过 1000 个 AI 应用场景，涵盖法务、财务等后台职能，现场技术人员支持及核心网络运营；客服通话记录的摘要与分析已完全由开源模型支撑。公司还基于电信行业数据定制开源模型 OTel，支撑 AI 智能体对网络故障的根因检测。Markus 表示，面对业界对"token 末日"的担忧，公司并不畏惧 token 成本增长，"这是我们觉得可以管理的事情"；开源模型可部署于公司自有数据中心而非租用云资源，进一步压缩 AI 推理成本。鉴于数据保护与知识产权考量，公司认为使用开源模型"不可协商"。

（四）算力服务

仅供内部参考，请勿外传！

■ **CoreWeave（CRWV.O）：第二季度营收同比增长 112%，积压订单达 1040 亿美元**

CoreWeave 于 8 月 11 日公布 2026 年第二季度业绩，季度营收 25.75 亿美元（公司披露$2,575 百万），同比增长 112.46%（去年同期$1,212 百万）；调整后 EBITDA 15.10 亿美元，调整后 EBITDA 利润率 59.00%；GAAP 口径净亏损 6.26 亿美元，每股亏损 1.14 美元。

订单与容量扩张方面，截至季度末公司积压订单（revenue backlog）约 1040 亿美元，第三季度初又新获超过 250 亿美元的客户承诺（尚未计入上述订单）；公司活跃电力容量单季增加近 500 兆瓦至 1.5 吉瓦，总签约电力容量约 3.7 吉瓦，AI 算力交付与签约需求持续旺盛。

作为美股"新云"（neocloud）代表性标的，CoreWeave 的订单与容量数据与 Nebius、SMCI 等形成交叉验证，共同指向 AI 算力供不应求的行业格局。

■ **Nebius（NBIS.O）：二季度 AI 云收入暴增 514%，算力定价能力显著提升**

Nebius 于 8 月 12 日公布 2026 年第二季度财报，季度总营收 5.82 亿美元，同比增长 454.04%；其中 AI 云业务销售额 5.75 亿美元，同比增长 514%，占总营收约 98%。盈利能力大幅改善：调整后 EBITDA 为 2.36 亿美元（上年同期亏损 2100 万美元），调整后 EBITDA 利润率 41%，其中 AI 云业务调整后 EBITDA 利润率 50%；GAAP 口径持续经营净亏损 1.90 亿美元，每股亏损 0.68 美元。

公司本季度经营现金流 22.46 亿美元，并持续加大基础设施投入；公司重申 2026 年全年营收指引为 30 亿至 34 亿美元。据公司管理层披露，公司正以"有合约支撑、可盈利"的方式将强劲需求转化为增长，AI 算力供不应求格局下的定价能力与利润率弹性正在显现。

作为 AI 算力租赁（neocloud）代表性标的，Nebius 的业绩验证了 AI 算力供需偏紧下的定价能力与利润率弹性。

（五）液冷

■ **Vertiv（VRT.N）：马来西亚新厂首批液冷设备出货，海外产能持续扩张**

Vertiv 于 8 月 17 日宣布在意大利帕多瓦 Tognana 园区投资扩产，用于扩大数据中心冷却系统的制造与测试能力，预计 2026 年底将该园区冷水机组产能翻倍，新的大型测试实验室将于 2027 年初投入运营。

扩张背景是 AI 高密度机柜对液冷与电力管理需求的快速增长：单机柜功率持续攀升，direct-to-chip、CDU 与整机柜液冷成为新建 AI 数据中心的标准配置。公司 2026 年第二季度净销售额 32.70 亿美元，同比增长 24.00%；公司维持 2026 年全年净销售额指引 138 亿至 142 亿美元、调整后每股收益指引 6.65 至 6.75 美元，并继续扩充产能以应对 AI 与通用计算部署带来的电力与热管理需求。

液冷供应链正在从北美向亚太等区域延伸，本地化产能有助于缩短交付周期、缓解"需求强劲而交付受限"的行业矛盾。区域数据中心投资有望为液冷与电力设备厂商提供增量需求。

（六）行业应用

■ **OpenAI：下调 GPT-5.6 Sol API 价格逾 20%，大模型价格竞争延续**

OpenAI 于 8 月 21 日在官方开发者社区公告，未来三个月内将 GPT-5.6 Sol 模型的 API 及 Credit 价格下调超过 20.00%，其中输入价格降至每百万 token 4 美元、输出价格降至每百万 token 20 美元；API 现已开放使用，并逐步开放给符合条件的 ChatGPT Work 与 Codex 积分套餐，Pro、Plus 及 Business 订阅价格维持不变。

仅供内部参考，请勿外传！

此次降价是 OpenAI 首次下调其旗舰级（frontier）模型价格，延续了公司近期降低推理成本、助推智能体（Agent）工作负载商业化的定价策略。同期，OpenAI 还宣布因前沿模型能力触及关键安全阈值，暂停部分 Astra 模型的强化学习训练，并开放 Codex Harness 底层能力，推动 AI 竞争从模型能力向任务稳定执行的 Agent 平台方向延伸。

头部大模型厂商的定价竞争持续压缩推理成本，利好 AI 应用侧放量与商业化落地；同时，模型安全边界与 Agent 执行可靠性正成为新的竞争维度，企业级 AI 应用的关注重点正从模型能力转向数据上下文、任务执行与生产环境可管理性。

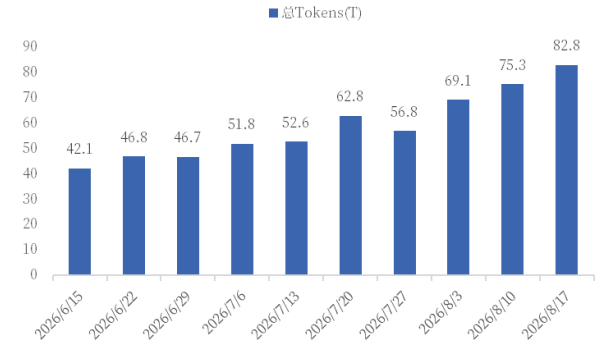
三、行业数据跟踪

（一）大模型 Tokens 调用量

当期(统计口径为 2026.08.10-2026.08.23)模型总调用量为 158.1 万亿 tokens，较上期的 125.9 万亿 tokens 环比上升约 25.58%。近十周总调用量整体呈稳步爬升态势，2026.08.10-2026.08.23 连续两周创新高（分别达 75.30T、82.80T），AI 模型调用需求加速释放。

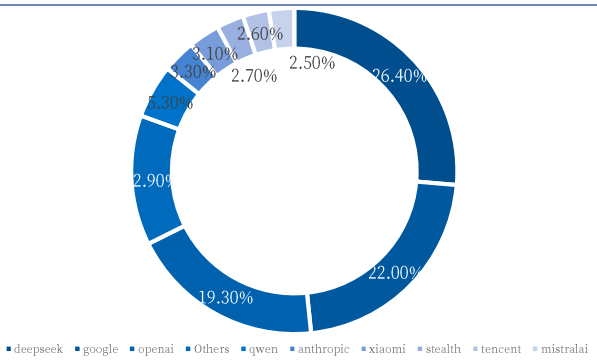
分模型厂商来看，最新一期 Token 市场份额中，DeepSeek 占比最高，为 26.40%；Google 位居第二，占比 22.00%；OpenAI 位居第三，占比 19.30%；Qwen、Anthropic 分别占比 5.30%、3.30%。此外，Xiaomi、Stealth、Tencent、Mistral AI 分别占比 3.10%、2.70%、2.60%、2.50%，其他模型合计占比 12.90%。整体来看，头部模型厂商市场份额仍较为集中，Google、DeepSeek 与 OpenAI 合计占比达到 67.70%，前五大模型厂商合计占比达到 76.30%。

图9：近十周每周总 Token 调用量变化（T）



资料来源：OpenRouter，中国银河证券研究院

图10：本周各模型 Token 市场份额（%）



资料来源：OpenRouter，中国银河证券研究院

分模型来看，最新一期调用量排名前五的模型分别为 DeepSeek V4 Flash 0731、MiMo - V2.5、Hy3、Ox Alpha、DeepSeek V4 Flash 0423，调用量分别为 11.6T，9.14T，8.56T，6.54T，5.3T。其中，DeepSeek V4 Flash 0731 以 11.6T 继续位居榜首，环比+5%，增速趋稳；MiMo-V2.5 从 3.79T 上涨至 9.14T（环比+138%），排名由第七位跃升至第二位，为本周最大增量来源；Ox Alpha 新进入榜单（6.54T），为 stealth 厂商模型，首次进入前十即位列第四；Hy3 从 9.95T 回落至 8.56T（环比-18%），排名由第二降至第三；GPT-5.6 Luna 基本持平（5.17T，-1%），排名稳定。

表1：本周热门模型 Token 调用量排名

模型	公司	当周 tokens 调用量	环比

仅供内部参考，请勿外传！

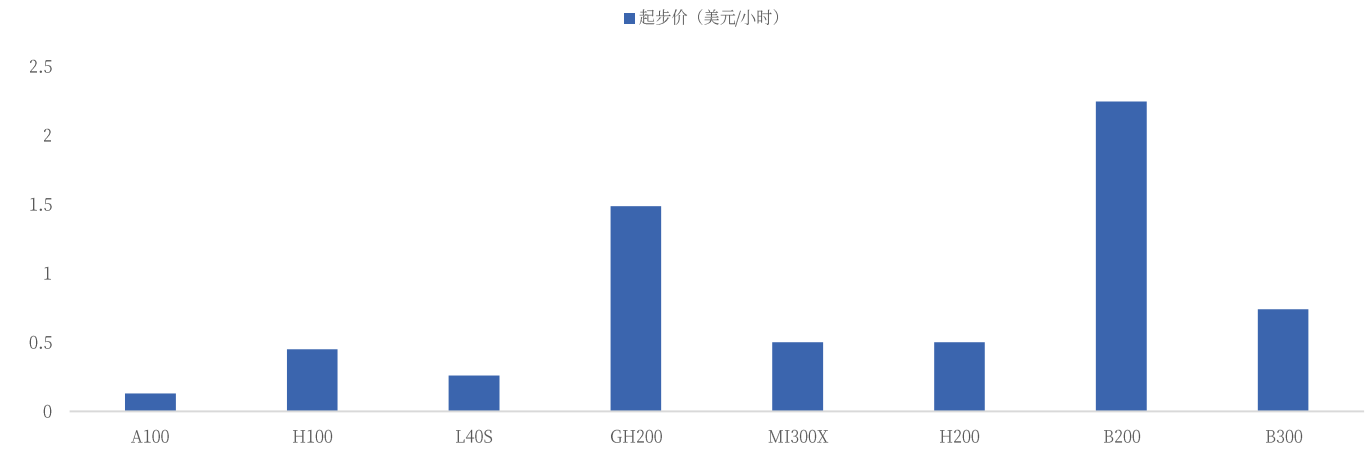
DeepSeek V4 Flash 0731	deepseek	11.6T	5%
MiMo - V2.5	xiaomi	9.14T	138%
Hy3	tencent	8.56T	-18%
Ox Alpha	stealth	6.54T	new
DeepSeek V4 Flash 0423	deepseek	5.3T	11%
GPT-5.6 Luna	openai	5.17T	-1%
Nemotron 3 Ultra (free)	nvidia	4.42T	122%
GLM5.2	z-ai	3.59T	-15%
Claude Opus 5	anthropic	2.09T	1%
DeepSeek V4 Pro 0423	deepseek	1.88T	-32%

资料来源：OpenRouter，中国银河研究院

（二）热门 GPU 价格走势

根据 Cloud GPU Tracker 数据，最新热门 GPU 起步价情况如下图所示。以 Nvidia A100 为例，最低报价为每小时 $0.13；H100 最低报价为每小时 $0.45，H200 最低报价为每小时 $0.50，B200 和 B300 最低报价分别为每小时 $2.25、$0.74。整体来看，新一代高端 GPU 价格仍明显高于 A100 等成熟型号，反映高性能 AI 算力资源供给仍具稀缺性，各型号起步价与上期基本持平。

图11：最新热门 GPU 起步价情况



资料来源：CloudGPUTracker，中国银河证券研究院

（三）子行业技术趋势

表2：子行业技术趋势

行业	技术看点
设备制造	AI 基础设施竞争从单一芯片向"土地-电力-厂房（LPS）"全要素延伸。NVIDIA 通过 5000 亿美元融资平台与 PORTS-Pike 园区布局，将算力转化为可融资的基础设施资产；DSX AI 工厂架构成为新一代 AI 云建设蓝本，SB Energy 等运营商与能源商

仅供内部参考，请勿外传！

	加速入局。Super Micro 新财年指引 650-720 亿美元，机柜级、液冷化整机方案持续放量，核心竞争从单颗 GPU 性能扩展至计算、互联、供电、冷却与交付能力协同。
运营商	运营商技术重点由传统连接服务向 AI 云与 AI 基础设施服务延伸。SK Telecom 采用 NVIDIA DSX 建设吉瓦级 AI 云，首个 AI 工厂 2027 年上线；"电信网络成为国家 AI 基础设施"的叙事强化，运营商凭借网络、数据中心与能源资产切入 AI 算力供给。后续重点跟踪运营商 AI 云落地节奏、私有 5G 与边缘推理等企业 AI 场景的商业化进展。
光通信	AI 数据中心高速互联需求持续推动 800G 向 1.6T/3.2T 升级，CPO、硅光、EML、VCSEL 等技术路线仍是产业关注重点。当前光通信供应链核心矛盾逐渐从需求不足转向产能和关键材料保障，InP 晶圆、激光器等环节成为影响高速光模块扩产的重要因素。Lumentum、Coherent 等厂商持续受益于 AI 数据中心光互联需求增长，后续重点跟踪光模块升级节奏、CPO 商业化进展及上游材料供给情况。
卫星通信	卫星通信技术主线继续围绕 SAR 对地观测星座、D2D 直连与低轨通信展开。Rocket Lab 单月两次为 iQPS 部署 SAR 卫星，Electron 全年发射已达 14 次；SAR 星座具备全天候观测能力，商业遥感与国防需求共振。后续关注低轨通信星座部署节奏、D2D 商业化进展及发射供应链交付能力。
行业应用	企业 AI 应用持续从 Copilot 辅助工具向 Agentic AI 与自动化工作流演进，大模型定价竞争加剧（OpenAI 下调 GPT-5.6 Sol 价格逾 20.00%），推理成本下降利好应用放量。Codex Harness 等开放 Agent 平台的出现，标志 AI 竞争从模型聪明度转向任务稳定执行；企业侧关注重点从模型能力转向数据上下文、任务执行、流程编排与生产环境可管理性。
算力服务	AIDC 板块核心变量仍集中于 GPU 可获得性、数据中心容量、电力资源和交付速度，但本期新增"定价权"维度：Nebius 首次容量拍卖溢价 15%-20%、短期合同升至每兆瓦每年 4000 万-5000 万美元，显示供不应求格局下算力定价能力显著提升。CoreWeave 积压订单 1040 亿美元、Nebius 签约容量目标上调至 5GW，AI 云厂商从"容量扩张"转向"定价与利润质量"竞争。后续重点跟踪 CoreWeave、Nebius、Equinix 等厂商的容量扩张、客户签约与定价变化。
液冷	高功耗 GPU 与高密度 AI 机柜推动液冷加速渗透，CDU、direct-to-chip 与整机柜液冷成为新建数据中心标配。Vertiv 马来西亚新厂首批 CDU 出货、意大利 Tognana 园区扩产，显示液冷供应链正加快区域化布局以缩短交付周期；800VDC 高压直流供电与"电链+冷链"一体化方案成为 AI 数据中心设计的新方向。

资料来源：中国银河研究院整理

四、行业大事前瞻

表3：行业大事前瞻

日期	类型	事件/标的
8 月 26 日（周三）	财报	NVIDIA 财报
8 月 27 日（周四）	财报	Dell Technologies 财报
9 月 2 日（周三）	财报	Broadcom 财报
时间待定	上市进展	OpenAI IPO：可能从原先 2026 年 Q4 推迟至 2027 年，暂无官方时间表
预计 2026 年 Q4	上市进展	Anthropic IPO
2027 年初预测	产业事件	Dan Ives 预测 Tesla-SpaceX 2027 年初合并
持续跟踪	宏观	FOMC 利率路径：鹰派点阵图 vs 实际通胀走势，市场与美联储对赌
11 月 3 日	政治事件	美国中期选举

资料来源：中国银河研究院整理

仅供内部参考，请勿外传！

五、投资建议

英伟达于 8 月 10 日联合 Apollo、贝莱德、黑石、Brookfield、高盛、KKR 六家机构签署谅解备忘录，拟搭建独立融资平台、长期撬动超 5000 亿美元第三方资本，将 GPU 纳入全球长期资本可投资的基础设施资产类别；8 月 17 日英伟达进一步锁定俄亥俄州 PORTS-Pike 科技园区的土地、电力与厂房资源，OpenAI 作为租户承租建设 AI 工厂（初始 4.25 IT-GW、可扩至 8.00 IT-GW），租期 20 年，英伟达同步向 SB Energy 投资 15 亿美元。**我们认为，供给侧层面，AI 工厂的竞争瓶颈已从先进制程外溢至土地、电力、厂房（LPS）等物理资源， LPS 提前锁定的重要性显著提升；英伟达以自身信用为数据中心资产融资增信，AI 基础设施的供给弹性将系统性抬升。产业链层面，算力资产化趋势将把订单与估值弹性从芯片设计进一步向 LPS 资源持有方、数据中心融资服务、能源配套及算力网络互连环节扩散，但同时现金流稳定程度可能在债务周期末端影响系统性风险。**

OpenAI 于 8 月 21 日在官方开发者社区公告，未来三个月内将 GPT-5.6 Sol 模型的 API 及 Credit 价格下调超过 20%，其中输入价格降至每百万 token 4 美元、输出价格降至每百万 token 20 美元。**我们认为此次降价直接打通了 Agentic Workflow、长上下文理解、自主代码生成等高输出密度场景的单位经济模型（Unit Economics），将触发推理算力需求从训练集群向端侧与边缘侧扩散。对产业链而言，价值量占比将从训练芯片向推理芯片、推理优化框架、高速互连及 AI 应用层系统性迁移；同时，模型层利润率的压缩将倒逼上游算力侧进一步降低 TCO（总拥有成本），CPO、液冷等环节的战略价值持续抬升。建议关注相关龙头标的：**

光模块/光通信相关标的：Lumentum（LITE），Coherent（COHR），Fabrinet（FN），Corning Inc.（GLW），Applied Optoelectronics（AAOI）等；设备制造相关标的：Ciena（CIEN），Cisco Systems（CSCO），Micron Technology（MU）等；运营商相关标的：AT&T（T），Verizon（VZ），T-Mobile US（TMUS）等；行业相关标的：Oracle（ORCL），Nvidia（NVDA）等。

六、风险提示

1、 技术落地与良率不及预期的风险；

2、 CSP 厂商 Capex 不及预期的风险；

3、 竞争加剧的风险；

4、 政策摩擦的不确定性风险。

仅供内部参考，请勿外传！

图表目录

图 1： 各主要指数双周涨跌幅对比 3
图 2： 各主要指数年初至今涨跌幅对比 3
图 3： 各主要指数双周成交额对比 3
图 4： 各主要指数年初至今成交额对比（亿美元） 3
图 5： Nasdaq-100 中双周涨幅前十名标的汇总（%） 4
图 6： Nasdaq-100 中双周跌幅前十名标的汇总（%） 4
图 7： 双周细分子板块涨跌幅情况（%） 4
图 8： 年初至今子板块涨跌幅（%） 4
图 9： 近十周每周总 Token 调用量变化（T） 8
图 10： 本周各模型 Token 市场份额（%） 8
图 11： 最新热门 GPU 起步价情况 9

表 1： 本周热门模型 Token 调用量排名 8
表 2： 子行业技术趋势 9
表 3： 行业大事前瞻 10

分析师承诺及简介

本人承诺以勤勉的执业态度，独立、客观地出具本报告，本报告清晰准确地反映本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与、未来也将不会与本报告的具体推荐或观点直接或间接相关。

赵良毕，通信&中小盘首席分析师，科技组组长。北京邮电大学通信硕士，复合学科背景，2022 年加入中国银河证券。8 年中国移动通信产业研究经验，7 年证券从业经验。曾获得 2018/2019 年（机构投资者 II-财新）通信行业最佳分析师前三名，2020 年获得 Wind（万得）金牌通信分析师前五名，获得 2022 年 Choice（东方财富网）通信行业最佳分析师前三名。

免责声明

本报告由中国银河证券股份有限公司（以下简称银河证券）向其客户提供。银河证券无需因接收人收到本报告而视其为客户。若您并非银河证券客户中的专业投资者，为保证服务质量、控制投资风险、应首先联系银河证券机构销售部门或客户经理，完成投资者适当性匹配，并充分了解该项服务的性质、特点、使用的注意事项以及若不当使用可能带来的风险或损失。

本报告所载的全部内容只提供给客户做参考之用，并不构成对客户的投资咨询建议，并非作为买卖、认购证券或其它金融工具的邀请或保证。客户不应单纯依靠本报告而取代自我独立判断。投资者应自主作出投资决策并自行承担投资风险，任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。银河证券认为本报告资料来源是可靠的，所载内容及观点客观公正，但不担保其准确性或完整性。本报告所载内容反映的是银河证券在最初发表本报告日期当日的判断，银河证券可发出其它与本报告所载内容不一致或有不同结论的报告，但银河证券没有义务和责任去及时更新本报告涉及的内容并通知客户。银河证券不对因客户使用本报告而导致的损失负任何责任。

本报告可能附带其它网站的地址或超级链接，对于可能涉及的银河证券网站以外的地址或超级链接，银河证券不对其内容负责。链接网站的内容不构成本报告的任何部分，客户需自行承担浏览这些网站的费用或风险。

银河证券在法律允许的情况下可参与、投资或持有本报告涉及的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。银河证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

银河证券已具备中国证监会批复的证券投资咨询业务资格。除非另有说明，所有本报告的版权属于银河证券。未经银河证券书面授权许可，任何机构或个人不得以任何形式转发、转载、翻版或传播本报告。特提醒客户及公众投资者慎重使用公众媒体刊载的证券研究报告。

本报告版权归银河证券所有并保留最终解释权。

评级标准

评级标准		评级	说明
评级标准为报告发布日后的 6 到 12 个月行业指数（或公司股价）相对市场表现，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，北交所市场以北证 50 指数为基准，香港市场以恒生指数为基准。	行业评级	推荐：	相对基准指数涨幅 10%以上
		中性：	相对基准指数涨幅在-5%～10%之间
		回避：	相对基准指数跌幅 5%以上
	公司评级	推荐：	相对基准指数涨幅 20%以上
		谨慎推荐：	相对基准指数涨幅在 5%～20%之间
		中性：	相对基准指数涨幅在-5%～5%之间
		回避：	相对基准指数跌幅 5%以上

联系

中国银河证券股份有限公司 研究院

深圳市福田区金田路 3088 号中洲大厦 20 层

上海浦东新区富城路 99 号震旦大厦 31 层

北京市丰台区西营街 8 号院 1 号楼青海金融大厦

公司网址：www.chinastock.com.cn

机构请致电：

深广地区：苏一耘 0755-83479312 suyiyun_yj@chinastock.com.cn
程 曦 0755-83471683 chengxi_yj@chinastock.com.cn

上海地区：林 程 021-60387901 lincheng_yj@chinastock.com.cn
李洋洋 021-20252671 liyangyang_yj@chinastock.com.cn

北京地区：田 薇 010-80927721 tianwei@chinastock.com.cn
褚 颖 010-80927755 chuying_yj@chinastock.com.cn