下载日志已记录，仅供内部参考

美股短期调整持续，AI 应用发展加速

2026 年 10 月 25 日

核心观点

- **双周美股主要指数全数回调，费指跌幅扩大：**双周（2026.10.12-2026.10.23）主要指数中，S&P500 涨跌幅为-2.40%，纳指涨跌幅为-5.16%，道指涨跌幅为-2.43%；科技行业方面，以科技成长股为主的 Nasdaq-100 指数涨跌幅为-6.24%，费指涨跌幅为-14.26%，表现弱于大盘。年初至今涨跌幅方面，费指涨幅为+52.19%，好于大盘，且高于 2025 年同期（+37.50%）及 2024 年同期（+22.89%）水平，半导体板块年初至今的绝对涨幅仍居各主要指数前列。子行业方面，Nasdaq-100 中，双周涨幅靠前标的主要集中于企业软件与数据服务、消费必需品及医疗保健等板块；双周跌幅靠前公司主要集中于半导体设备、存储与 AI 算力服务等板块。我们判断主要因素为，一方面，随着超大规模厂商资本开支占经营现金流的比重上升，市场对算力投资的融资结构与回报周期重新定价，半导体设备、存储与光通信等前期强势环节的估值出现调整；另一方面，资金转向现金流相对稳定的下游应用及周期性板块，软件订阅、消费必需品与电信运营类标的。

- **核心观点：**台积电公布第三季度业绩，季度营收同比增长 50.9%、净利同比增长 68.7%，7 纳米及以下制程占晶圆收入 79%，并指引第四季度营收 523-543 亿美元，先进制程与存储的产能利用率仍处于高位，AI 加速器与高容量内存的订单预期尚未下修；SK 海力士在 OCP 峰会展示 256GB CMM-DDR5，支持 CXL 3.2、带宽 80GB/s，约束环节已由晶圆产能转向先进封装、HBM 堆叠与 CXL 内存的容量爬坡，其中 CXL 内存池化使单机架的内存扩展摆脱 DIMM 插槽限制，为长上下文推理提供了新的容量路径。

  SpaceX 披露第三季度营收 146 亿美元、同比增长 161%，星链用户达 1,360 万；AST SpaceMobile 计划年底前在五国提供初始间歇性天基服务，全球网关超过 60 个。我们认为，卫星直连由星座建设进入网络就绪阶段，网关密度与覆盖小区数量成为商用开通的前置条件，SpaceX 与 AST SpaceMobile 分别以自建运营商与分成合作两种路径推进，频谱资源与终端生态的兼容性决定长期份额。微软宣布自研 Maia 加速器自 2027 年起扩至多吉瓦并成为主要推理系统；Anthropic 推出 Claude Duet 研究预览版，Meta 发布开放权重的 Muse Spark Open。主动式智能体与本地化部署同时推进，Claude Duet 的持续监测模式将单用户调用频次由会话驱动转为轮询驱动，Muse Spark Open 则在桌面级集群上实现了 8,160 亿参数模型的分布式推理，推理负载的分布结构正由云端集中转向分层部署。

- **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：台积电（TSM），ASML（ASML），Lam Research（LRCX），Broadcom（AVGO），Micron Technology（MU）等；运营商相关标的：AT&T（T），Verizon（VZ），T-Mobile US（TMUS），Comcast（CMCSA）等；算力服务与行业应用相关标的：Microsoft（MSFT），Meta（META），Oracle（ORCL），CoreWeave（CRWV），Nebius（NBIS），Nvidia（NVDA），Vertiv（VRT），Trane Technologies（TT）等

海外算力行业

推荐 维持评级

分析师

赵良毕

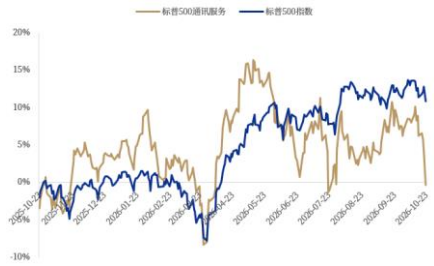
☎：010-8092-7619

✉：zhaoliangbi_yj@chinastock.com.cn

分析师登记编码：S0130522030003

相对标普 500 表现图

2026 年 10 月 23 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】首次覆盖报告_Lumentum（LITE.O）_全球全栈光电 IDM 龙头，弹性兑现在即

2.【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

3.【银河海外算力】行业双周报_美股新财季开启，供需缺口预计推高增速二阶导

下载日志已记录，仅供内部参考

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 7
（五） 液冷 8
（六） 行业应用 8
三、 行业数据跟踪 9
（一） 大模型 Tokens 调用量 9
（二） 热门 GPU 价格走势 10
（三） 子行业技术趋势 11
四、 行业大事前瞻 12
五、 投资建议 12
六、 风险提示 13

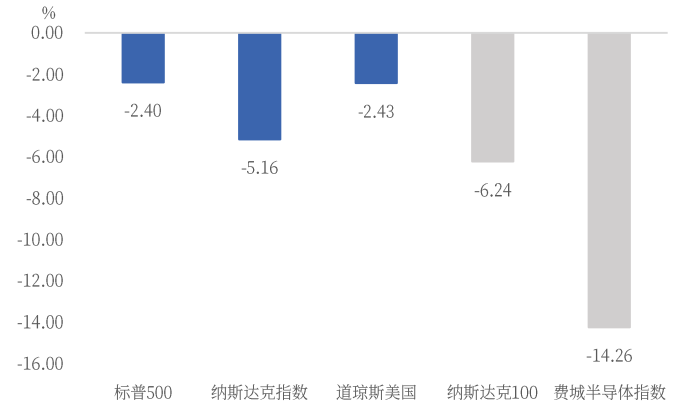
下载日志已记录，仅供内部参考

一、双周行情回顾

（一）海外主要指数回顾

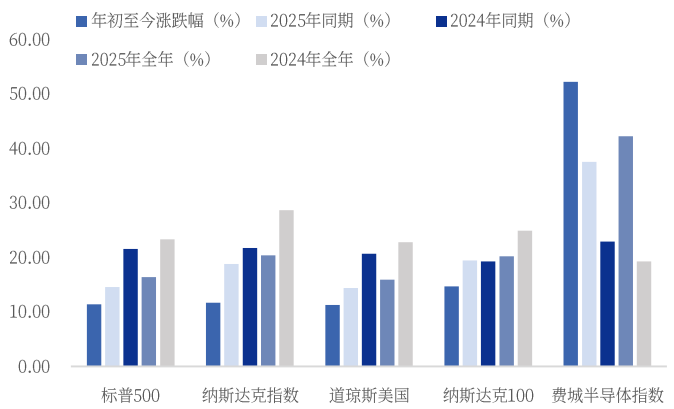
双周美股主要指数全数回调，费指跌幅扩大：双周（2026.10.12-2026.10.23）主要指数中，S&P500 涨跌幅为-2.40%，纳指涨跌幅为-5.16%，道指涨跌幅为-2.43%；科技行业方面，以科技成长股为主的 Nasdaq-100 指数涨跌幅为-6.24%，费指涨跌幅为-14.26%，表现弱于大盘。年初至今涨跌幅方面，费指涨幅为+52.19%，好于大盘，且高于 2025 年同期（+37.50%）及 2024 年同期（+22.89%）水平，半导体板块年初至今的绝对涨幅仍居各主要指数前列。

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

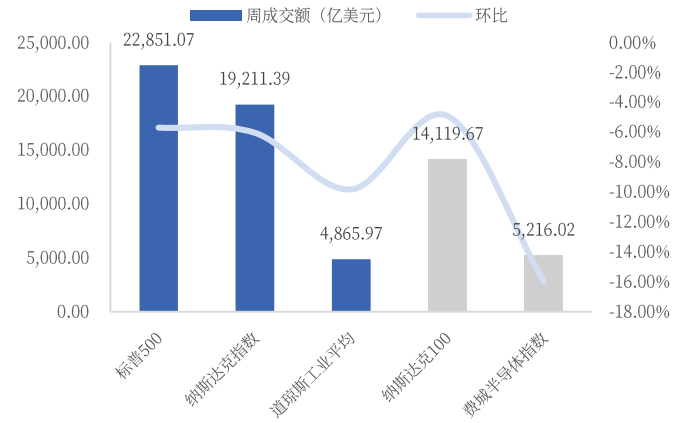
图2：各主要指数年初至今涨跌幅对比



资料来源：Wind，中国银河证券研究院

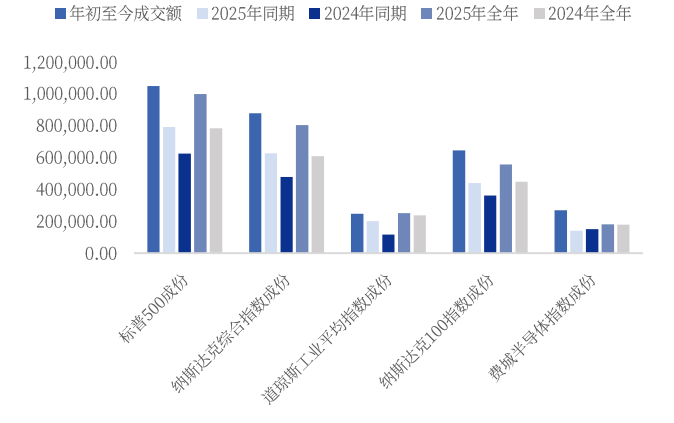
成交额方面，双周科技指数成交额绝对值不及大盘，最新一周标普 500、纳指及 Nasdaq-100 成交额分别为 22,851.07 亿美元、19,211.39 亿美元、14,119.67 亿美元；环比来看，上述指数成交额较前一周均有所回落，其中费指降幅最大，与半导体板块交投活跃度回落的方向一致。成交额与指数表现同向回落，半导体与算力板块的资金参与度较前期有所下降。

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

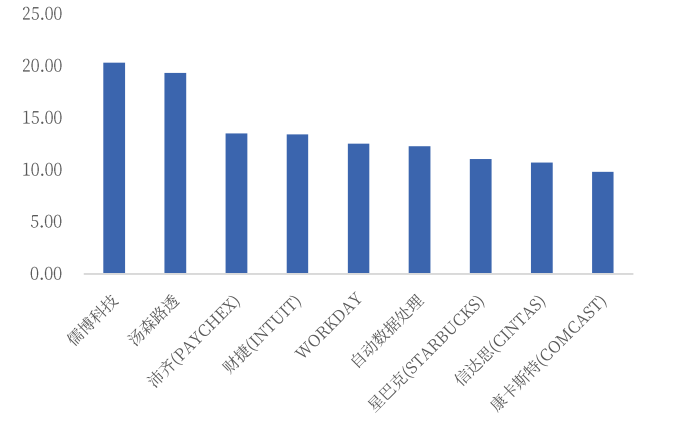
（二）各子行业表现情况

子行业方面，Nasdaq-100 中，双周涨幅靠前标的主要集中于企业软件与数据服务、消费必需

下载日志已记录，仅供内部参考

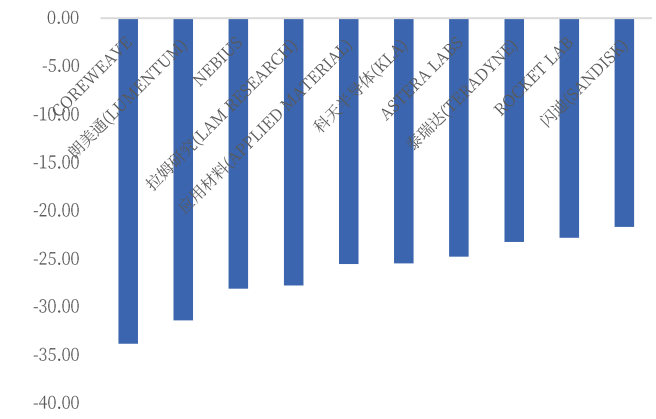
品及医疗保健等板块；双周跌幅靠前公司主要集中于半导体设备、存储与 AI 算力服务等板块。我们判断主要因素为，一方面，随着超大规模厂商资本开支占经营现金流的比重上升，市场对算力投资的融资结构与回报周期重新定价，半导体设备、存储与光通信等前期强势环节的估值出现调整；另一方面，资金转向现金流相对稳定的下游应用及周期性板块，软件订阅、消费必需品与电信运营类标的。

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（%）

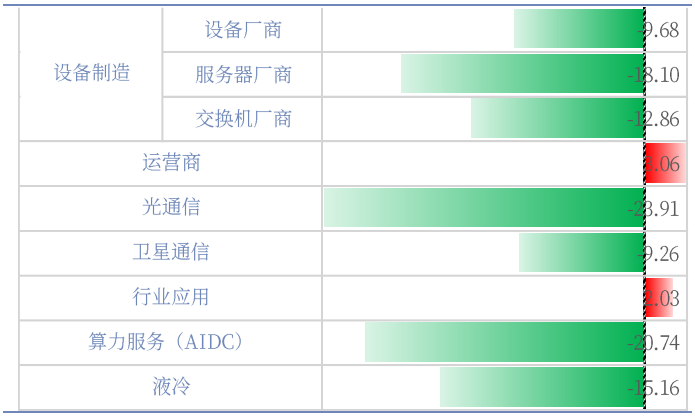


资料来源：Wind，中国银河证券研究院

（三）双周行业龙头标的表现

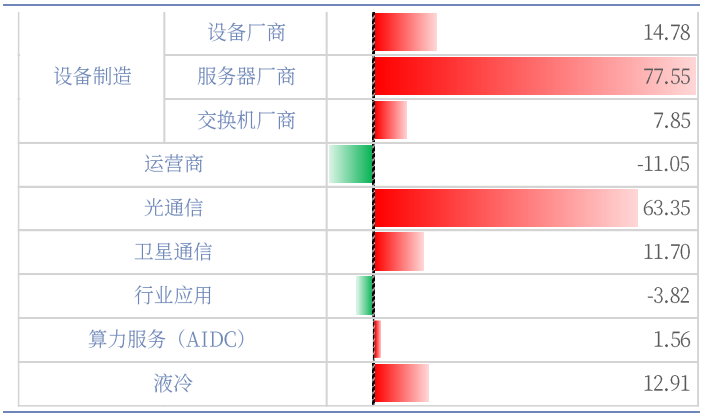
根据我们对于海外算力行业划分子板块，海外算力板块子行业包括设备制造（设备厂商、服务器厂商、交换机厂商）、运营商、光通信、卫星通信、行业应用、算力服务（AIDC）、液冷七大板块。双周内板块表现分化，实现上涨的板块为运营商与行业应用，涨幅分别为 3.06%、2.03%；回调板块方面，跌幅居前的板块分别为光通信、算力服务（AIDC）、服务器厂商，跌幅分别为 23.91%、20.74%、18.10%。

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 HubSpot、Northrop Grumman、Workday、IBM、Comcast 等，涨幅分别为 18.65%、14.71%、12.54%、11.64%、9.83%。回调标的按照跌幅降序分别为 AOI、Flex、Fabrinet、CoreWeave、Coherent，跌幅分别为 38.43%、36.12%、35.23%、33.92%、33.05%，主要集中于光通信、算力服务（AIDC）板块标的。

下载日志已记录，仅供内部参考

二、双周主要事件梳理

（一）设备制造

■ TSMC（TSM.N）：2026 年第三季度营收 NT$1.49 万亿、净利 NT$7,630.6 亿，第四季度指引 523-543 亿美元

TSMC 于 10 月 15 日公布 2026 年第三季度业绩。季度合并营收约 NT$1.4942 万亿，同比增长 50.9%；税后净利 NT$7,630.6 亿，同比增长 68.7%；毛利率 66.4%，营业利益率 58.7%。以美元计，季度营收 467.2 亿美元，同比增长 41.2%。制程结构方面，2 纳米制程占晶圆收入 8%、3 纳米占 33%、5 纳米占 28%，7 纳米及以下合计占 79%。公司指引第四季度营收 523-543 亿美元，毛利率区间 65%-67%，营业利益率区间 56.5%-58.5%，汇率假设为 1 美元兑 31.8 新台币。

7 纳米及以下制程占晶圆收入近八成，说明 AI 加速器与高速互连对先进产能的占用仍在延续，与上期月度营收中呈现的制程集中趋势一致。第四季度指引中值 533 亿美元较第三季度继续上行，对应环比约 14%的增长；毛利率区间 65%-67%的中值略低于本季的 66.4%，反映海外新厂投产初期的成本摊销与汇率波动的双重影响。需要关注的是，制程迁移至 2 纳米后单位面积成本上升，若客户在 2027 年进入拉货节奏切换期，收入确认的季节性波动可能放大。

■ ASML（ASML.O）：第三季度净销售额 117 亿欧元、净利 50 亿欧元

ASML 于 10 月 14 日披露 2026 年第三季度业绩。季度净销售额 117 亿欧元，净利润 50 亿欧元，对应净利率约 42.7%。EUV 光刻设备是 7 纳米以下先进逻辑与高端 DRAM、HBM 产能的前置设备，其交付节奏通常领先于晶圆厂资本开支六至十二个月，因此当季收入规模对下游 2027 年的扩产强度具有指向意义。

本季净销售额与净利润的规模，与台积电同季公布的 79%先进制程收入占比形成上下游印证，说明先进制程扩产仍处于设备交付的密集期，光刻环节尚未出现订单递延的迹象。需要关注的是，光刻设备的订单能见度取决于客户 2027 年的资本开支规划，而超大规模厂商的算力投资正逐步向自研芯片与推理侧倾斜，若先进逻辑的扩产节奏在 2027 年放缓，设备环节的收入确认节奏将滞后反映；此外，High-NA EUV 的导入成本与良率爬坡，仍是影响单台设备价值量与客户决策节奏的关键变量。

■ SK 海力士（000660.KS）：OCP 峰会展示 256GB CMM-DDR5，CXL 3.2 带宽 80GB/s

SK 海力士于 10 月 16 日披露，在 10 月 12 日至 15 日于圣何塞举行的 OCP Global Summit 2026 上展出第二代 CMM-DDR5（CXL 内存模组）。该模组单颗容量 256GB，支持 CXL 3.2 规范，带宽 80GB/s，能效较上一代改善约 30%。同场还展出了 HBM4、HBM4E 与 HBM3 的实物样品，以及 256GB 3DS RDIMM、192GB SOCAMM2 与 128GB DDR5 MRDIMM 等面向 AI 服务器的高容量内存产品。

CXL 内存模组的意义在于把内存容量从每 CPU 插槽绑定的形态扩展为可池化的共享资源，256GB 单模组容量与 80GB/s 带宽使单机架的内存扩展能力不再受限于 DIMM 插槽数量，对长上下文推理与向量检索等内存密集型负载具有直接影响。结合本期微软自研推理芯片与 Meta 开放权重模型的进展，推理侧的容量需求正由 HBM 向 DDR5 与 CXL 分层扩散。需要关注的是，CXL 的商用节奏取决于服务器平台对 CXL 3.x 的支持进度，以及内存池化在操作系统与调度层的软件成熟度。

下载日志已记录，仅供内部参考

■ **Lam Research（LRCX.O）：季度营收 82.19 亿美元、环比增 22%，营业利益率 40.0%**

Lam Research 于 10 月 21 日公布截至 2026 年 9 月 27 日的季度业绩。季度营收 82.19 亿美元，环比增长 22%；GAAP 毛利率 52.3%，较上季度提升 60 个基点；营业利益率 40.0%，提升 260 个基点；稀释每股收益 2.19 美元，环比增长 21%。非 GAAP 口径下，毛利率 52.4%、营业利益率 40.2%、稀释每股收益 2.20 美元。

刻蚀与沉积设备是 3D NAND 层数堆叠与 DRAM 电容结构升级的核心工艺环节，本季营业利益率环比提升 260 个基点，说明产品结构正向高深宽比刻蚀与先进封装相关设备倾斜。需要关注的是，设备厂商的收入确认通常滞后于存储原厂的资本开支决策一到两个季度，本期营收的环比增长对应的是上半年存储合约价上行之后的扩产落地；若 2027 年存储厂商的资本开支增速回落，设备环节的订单增速可能先于收入见顶。

（二）卫星通信

■ **SpaceX：第三季度营收 146 亿美元，星链用户达 1,360 万**

SpaceX 于 10 月 20 日披露第三季度业绩（季度截至 9 月 30 日）。季度营收 146 亿美元，同比增长 161%；营业利润 28 亿美元、净利润 44 亿美元，上年同期为净亏损；调整后 EBITDA 为 88 亿美元。星链消费者用户达 1,360 万，季度净增 160 万，月度 ARPU 为 69 美元、环比提升 5%。分业务看，连接业务营收 52.48 亿美元、同比增长 69%，其中企业及政府客户 21.27 亿美元、增长 91%；超级智能与算力（SI）业务营收 76 亿美元、环比增长 195%。季度资本开支 195.93 亿美元，其中 SI 业务 164.44 亿美元；截至季末算力装机达 1.7GW，上年同期为 0.6GW。

单季资本开支 195.93 亿美元、其中算力业务 164.44 亿美元，意味着公司已进入星座扩容与算力基础设施并行投入的阶段，其融资能力与现金流结构对卫星互联网行业的资本形成具有示范效应。连接业务中企业及政府客户增长 91%，高于消费者业务，说明收入结构正由零售宽带向高单价、长约的政企市场倾斜；同期星链 ARPU 环比提升 5%至每月 69 美元，但较上年同期的 83 美元仍有回落。需要关注的是，用户净增对应的终端补贴与获客成本未在披露中拆分，若 ARPU 随新兴市场用户占比上升继续下行，用户增长对收入的边际贡献将递减。

■ **AST SpaceMobile（ASTS.O）：推进五国初始服务，全球网关超 60 个**

AST SpaceMobile 于 10 月 20 日披露服务准备进展。公司计划在 2026 年底前在美国、加拿大、日本、英国与沙特阿拉伯五国提供初始的间歇性天基蜂窝宽带服务；全球网关数量已超过 60 个，可支撑超过 60 家移动网络运营商、覆盖超过 30 亿用户。美国境内已启用超过 3,000 个覆盖小区，年底前计划激活约 5,600 个。频谱方面，公司已持有北美 45MHz 的 MSS 中频资源，并在各目标市场拥有 60MHz 的 S 频段优先权，目标在美国取得约 100MHz。制造端，BlueBird 卫星的生产已推进至第 50 颗。

网关与覆盖小区的数量是卫星直连从技术验证走向商用计费的关键前置条件，超过 60 个网关使公司具备了与多国运营商同步开通服务的基础设施条件。与 SpaceX 自建移动网络运营商的路径不同，AST 采取与地面运营商分成的模式，其商业化速度取决于合作运营商的终端适配与计费系统对接进度。需要关注的是，初始服务定位为“间歇性”，实际可用时长受卫星过顶窗口与在轨卫星数量约束，在 BlueBird 星座规模达到设计值之前，用户体验与 ARPU 的提升空间有限。

（三）运营商

下载日志已记录，仅供内部参考

■ **AT&T（T.N）：第三季度营收 323 亿美元，后付费电话净增 37.9 万**

AT&T 于 10 月 21 日公布 2026 年第三季度业绩。季度营收 323 亿美元，同比增长 5.3%；合并服务收入同比增长 3.8%，为四年多以来的最高水平；无线服务收入同比增长 5%，后付费电话 ARPU 同比提升 4%。用户方面，后付费电话净增 37.9 万，流失率 0.92%；高级家庭互联网净增 69.6 万，其中光纤 41.6 万、Internet Air 28.0 万，该业务收入 29.75 亿美元、同比增长 23.4%。季度资本开支 56 亿美元，自由现金流约 50 亿美元；公司将 2026 年调整后每股收益指引上调至约 2.40 美元，全年资本开支指引 230-240 亿美元。

在卫星直连与固定无线接入同时挤压传统宽带份额的背景下，AT&T 本季度的增长仍主要来自无线服务与光纤，说明地面网络在容量与时延上的优势短期内难以被替代。全年资本开支指引 230-240 亿美元与每股收益指引上调同时出现，反映公司在维持网络投入的同时已进入自由现金流释放阶段。需要关注的是，光纤与固定无线的用户增长伴随较高的获客补贴，若 ARPU 提升幅度不足以覆盖补贴成本，宽带业务的单位经济模型改善空间有限。

（四）算力服务

■ **Microsoft（MSFT.O）：Azure Maia 扩至多吉瓦级，2027 年起成为主力推理系统**

Microsoft 于 10 月 12 日宣布，自研 AI 加速器 Maia 200 已投产并服务 Copilot，年内将扩展至 GitHub Copilot 与 Microsoft Foundry，目前已在北美、欧洲与澳大利亚部署，单位功耗的 Token 产出较主流 GPU 基础设施最高提升 1.7 倍。公司计划自 2027 年起将 Maia 系列扩展至多吉瓦（multi-gigawatt）规模，使其成为自有产品与服务的主要推理系统。下一代 Maia 300 每兆瓦 Token 产出再提升 40%，在 96 颗加速器的机架配置下，机架级吞吐最高为 Maia 200 的 7 倍；单颗 Maia 300 配备 8 个 HBM4 堆栈、带宽 18.0TB/s、热设计功耗 950 瓦，单机架 96 颗对应 167 千瓦、27.6TB HBM 与 1.7PB/s 带宽，计算芯粒采用台积电 N2P 工艺，计划 2027 年量产。

自研加速器由试点走向“主要推理系统”，意味着超大规模厂商的算力采购结构将从通用 GPU 的单极格局转向“通用 GPU 训练、自研 ASIC 推理”的分工，对 GPU 厂商的推理侧份额形成中期分流。单位功耗 Token 产出提升 1.7 倍与机架级吞吐提升 7 倍两项指标，指向的是推理成本而非峰值算力的竞争，这与本期模型侧单位价格下行的方向一致。需要关注的是，自研芯片的软件栈成熟度与开发者生态仍是制约因素，若主流 GPU 生态的迁移成本高于预期，自研 ASIC 的实际部署节奏可能慢于规划。

■ **Meta（META.O）：第三季度资本开支 349 亿美元，全年指引收窄至 1,350 亿美元起**

Meta 于 10 月 21 日公布 2026 年第三季度业绩。季度营收 646 亿美元、同比增长 26%，其中广告收入 630.6 亿美元；当季资本开支约 349 亿美元，自由现金流为-21 亿美元；经营利润约 130 亿美元、同比下降 37%，其中包含 99 亿美元的青少年在线安全诉讼和解费用。公司将 2026 年全年资本开支指引由 1,300-1,450 亿美元收窄至 1,350-1,450 亿美元，并指引第四季度营收 730-760 亿美元。

单季 349 亿美元的资本开支已超过当季自由现金流，自由现金流转负是本期 AI 资本开支融资担忧的直接来源之一，亦与半导体、光通信与算力服务板块的估值调整方向一致。资本开支指引下限由 1,300 亿美元上调至 1,350 亿美元，说明投入节奏并未因现金流压力而放缓，而是向区间上沿收敛。需要关注的是，经营利润同比下降 37%中虽有一次性诉讼费用的影响，但折旧摊销随算力资产规模上升而对利润表形成的持续压力将在 2027 年进一步显现，资本开支的回报周期正成为估值定价的核心变量。

下载日志已记录，仅供内部参考

■ **Firebird：获 4.5 万颗 GB GPU 出口许可，哈萨克斯坦二期规划约 500MW**

NVIDIA 于 10 月 22 日披露，Firebird 与哈萨克斯坦政府签署投资协议，取得 4.5 万颗 NVIDIA Grace Blackwell GPU 的对美出口许可。项目首期与 Kazakhtelecom 在 Data Center Valley 共建，目标于 2027 年第三季度投入商用；二期规划将使潜在 AI 基础设施容量达到约 500 兆瓦。Kazakhtelecom 已在阿斯塔纳与阿拉木图部署 400 颗 H200 与 480 颗 B300，并计划明年在 Data Center Valley 与 Ekibastuz 各部署超过 1,000 颗 GPU，后者采用 Vera Rubin 平台。

4.5 万颗 Grace Blackwell 的单笔出口许可是本期规模最大的算力基础设施承诺之一，其意义在于主权 AI 需求正从采购意向进入设备许可与园区建设阶段，为 GPU 厂商提供了北美超大规模厂商之外的第二增长曲线。二期约 500 兆瓦的容量规划对应的是电力与散热配套的同步建设，机架级液冷与高压直流供电成为必选项。需要关注的是，主权 AI 项目的落地节奏受出口管制政策、当地电网容量与融资结构三重约束，首期 2027 年第三季度的商用目标存在延后可能。

（五）液冷

■ **NVIDIA（NVDA.O）：OCP 发布机架级盲插液冷，单机架支持 250kW**

NVIDIA 于 10 月 12 日在 OCP Global Summit 2026 上宣布，通过 MGX 平台向 OCP 贡献 Vera Rubin 机架与计算托盘设计。MGX 1.2 规范新增机架级盲插液冷连接、液冷 5,000 安培母线以及无风扇液冷计算托盘，单机架支持最高 250 千瓦。同期，NVIDIA 与 AMD、Google、Meta 共同发布 Open Data Center Specification v0.7，并与 Google、Meta、Microsoft 发布 OCP Grid Resilience Specification，要求电网扰动后 2 秒内恢复接近满载运行。Meta 同场展示了 Spectrum-6 102.4Tb/s 交换机与 IcePack 100T 液冷机架，支持超过 10 万颗 GPU。

机架级盲插液冷把冷却接口从设备级搬到机架级，使数据中心可以按整机架的方式交付与维护，这是单机架功率向 250 千瓦演进的前提条件；5,000 安培液冷母线则解决了高功率密度下的供电路径问题。规范层面的联合发布说明液冷与供电的接口标准正由单一厂商主导转向多厂商共识，有利于降低整机柜方案的集成成本。需要关注的是，电网韧性规范提出的 2 秒恢复要求对储能与不间断电源配置提出更高要求，短期内将抬升单机架的配套投入。

■ **Trane Technologies（TT.N）：收购加拿大 Cancoil，扩充数据中心换热产能**

Trane Technologies 于 10 月 21 日签署协议，收购加拿大 Cancoil Thermal Corporation。Cancoil 成立于 1983 年，总部位于安大略省，拥有 450 余名员工，产品包括流体冷却器、风冷冷凝器与定制翅片管换热器，在安大略设有生产设施并在美国得克萨斯州 Jacksonville 具备产能，此前为 Trane 的供应商。交易预计在 2026 年底前完成交割，财务条款未予披露。

收购上游换热器供应商的核心目的是缩短关键部件的交付周期，在冷水机组与冷却分配单元需求集中释放的阶段，翅片管换热器的产能往往成为整机交付的瓶颈。以垂直整合方式锁定部件产能，反映液冷与热管理环节的竞争已从整机方案延伸至零部件供应链。需要关注的是，换热器属于定制化程度较高的部件，产能爬坡与工艺磨合需要时间，短期内对交付能力的改善幅度可能有限。

（六）行业应用

■ **Anthropic：推出 Claude Duet，主动式智能体进入办公场景**

Anthropic 于 10 月 21 日推出 Claude Duet 研究预览版，面向美国地区的 Pro 与 Max 订阅用户开放候补名单，提供 iOS 与 Android 独立应用及 claude.ai 入口，与原账户共用用量额度，底层模型为 Claude Opus 5.5。产品在每个早晨生成要点简报，持续监测已连接的应用并主动提醒；任

下载日志已记录，仅供内部参考

务运行于自带浏览器的虚拟机中，设备离线后可继续执行；可调用 Claude Marketplace 的 6,000 余个连接器与插件。在发送消息、付款、签署文件与删除操作前，系统会逐项请求确认，证券交易与账户间转账被列为不可授权的操作。

由被动响应转向主动监测并执行，意味着智能体的调用频次从用户发起转为持续轮询，单用户的 Token 消耗强度将高于对话式产品，对推理算力形成结构性增量需求。6,000 余个连接器与逐项确认机制，说明产品设计在扩展操作权限的同时以人工确认作为风险边界。需要关注的是，主动式智能体的实际采用率取决于误操作成本与用户信任的建立速度，研究预览阶段的额度共用安排也限制了对真实用量强度的观测。

■ **Meta：发布 Muse Spark Open，8,160 亿参数可在双机运行**

Meta 于 10 月 15 日发布开放权重的 Muse Spark Open 模型。模型总参数 8,160 亿，推理时激活约 420 亿参数，上下文窗口 100 万 Token，输入支持文本、图像、视频与音频。常规精度下模型约需 1.6TB 内存，Meta 与 Apple 合作通过 MLX 量化将体积压缩至约 493GB，借助 Thunderbolt 5 上的 RDMA 在两至四台 M5 Ultra Mac Studio 之间分布式推理，四机配置下的生成速度约为每秒 36.7 个 Token。模型已在 Hugging Face 开放下载，采用 Meta 自有许可。

在两至四台桌面级工作站上运行 8,160 亿参数模型，把大模型的部署门槛从数据中心级拉低到工作室级，其产业含义在于推理负载的分布结构可能从云端集中向“云端训练、边缘推理”分层。100 万 Token 上下文配合本地部署，对隐私敏感与低时延场景具有直接吸引力。需要关注的是，493GB 的量化体积意味着精度损失与显存带宽约束，每秒 36.7 个 Token 的生成速度仅适用于单用户或小并发场景，对云端推理需求的替代作用短期内有限。

三、行业数据跟踪

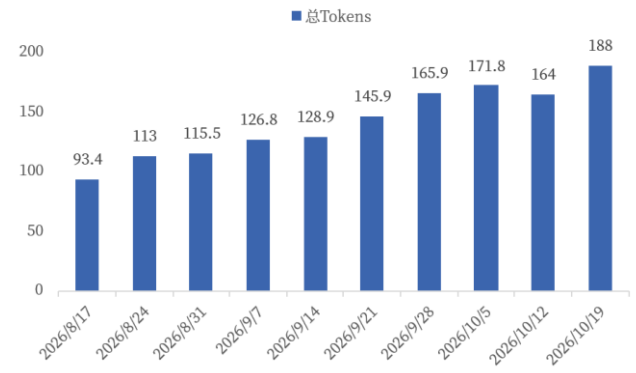
（一）大模型 Tokens 调用量

本期（统计口径为 2026.10.12-2026.10.25）模型总调用量合计约为 352.1 万亿 tokens，较上一双周（2026.09.28-2026.10.11）的 337.7 万亿 tokens 环比增长约 4.26%，在上一双周小幅回落后重回升势。本期两周分别录得 164.0T 与 188.0T，其中 10 月 19 日当周为近十周最高水平，较前一周环比增长约 14.63%；上一双周两周分别为 165.9T 与 171.8T。从结构上看，本周增量主要来自 DeepSeek V4.1 Flash 新版与 MiniMax M3.1 Flash（free）等新进入榜单的模型，周度调用量自 9 月下旬以来持续处于高位。

分模型厂商来看，OpenRouter 统计的最新一周（2026 年 10 月 19 日当周）各厂商文本请求量份额中，DeepSeek 占比最高，为 26.5%，位居首位；Google 位列第二，占比 20.1%；OpenAI 位列第三，占比 18.5%；Z.ai 位列第四，占比 10.7%；Anthropic 位列第五，占比 5.5%；Qwen、Mistral AI、MiniMax、Tencent 分别占比 4.7%、2.0%、2.0%、1.8%，其他厂商合计占比 8.3%。前三家合计占比 65.1%，前五家合计占比 81.3%。从请求量变化看，DeepSeek、Google、OpenAI 与 Anthropic 分别增长 16%、9%、7%、25%，Z.ai 与 Qwen 分别下降 6%，其他厂商合计增长 4%。

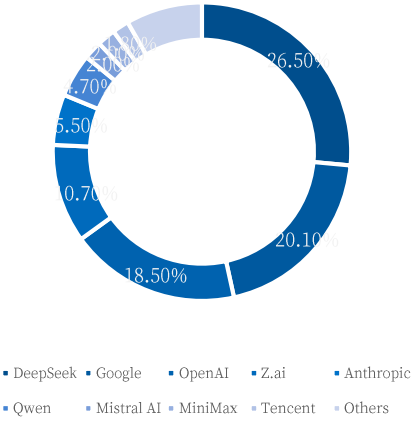
下载日志已记录，仅供内部参考

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：本周各模型市场份额（%，按请求数计）



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分别为 DeepSeek V4.1 Flash 的两个版本、MiniMax M3.1 Flash（free）、GLM 5.3 Flash 与 Hy4 preview，调用量分别为 33.56T、18.00T、13.68T、13.24T、8.97T。其中，DeepSeek V4.1 Flash 的 10 月 17 日版本以 33.56T 位居首位，为本期新进入榜单的版本；同系列 9 月 10 日版本以 18.00T 位居第二（环比-45.2%），反映 DeepSeek 内部用量向新版本迁移；MiniMax M3.1 Flash（free）以 13.68T 新进入榜、位列第三；GLM 5.3 Flash 以 13.24T 位列第四（环比+8.2%）；Hy4 preview 以 8.97T 位列第五（环比+18.7%）。前十其余模型依次为 GPT-6 Luna(8.35T, 环比+34.6%, 为本期前十中增速最高)、Claude Opus 5.5(8.27T, 环比+33.7%)、MiMo-V2.6-Flash（6.07T，环比-38.9%）与 Nemotron 3 Ultra（free）（5.85T，环比+17.8%）。厂商层面，DeepSeek 两款模型合计 51.56T，约占前十合计 115.99T 的 44.5%，为本期上榜规模居首的厂商；前十合计约 115.99T，其余模型合计约 72.05T。

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
DeepSeek V4.1 Flash 1017	deepseek	33.56T	new
DeepSeek V4.1 Flash 0910	deepseek	18.00T	-45.2%
MiniMax M3.1 Flash (free)	minimax	13.68T	new
GLM 5.3 Flash	z-ai	13.24T	8.2%
Hy4 preview	tencent	8.97T	18.7%
GPT-6 Luna	openai	8.35T	34.6%
Claude Opus 5.5	anthropic	8.27T	33.7%
MiMo-V2.6-Flash	xiaomi	6.07T	-38.9%
Nemotron 3 Ultra (free)	nvidia	5.85T	17.8%

资料来源：OpenRouter，中国银河研究院

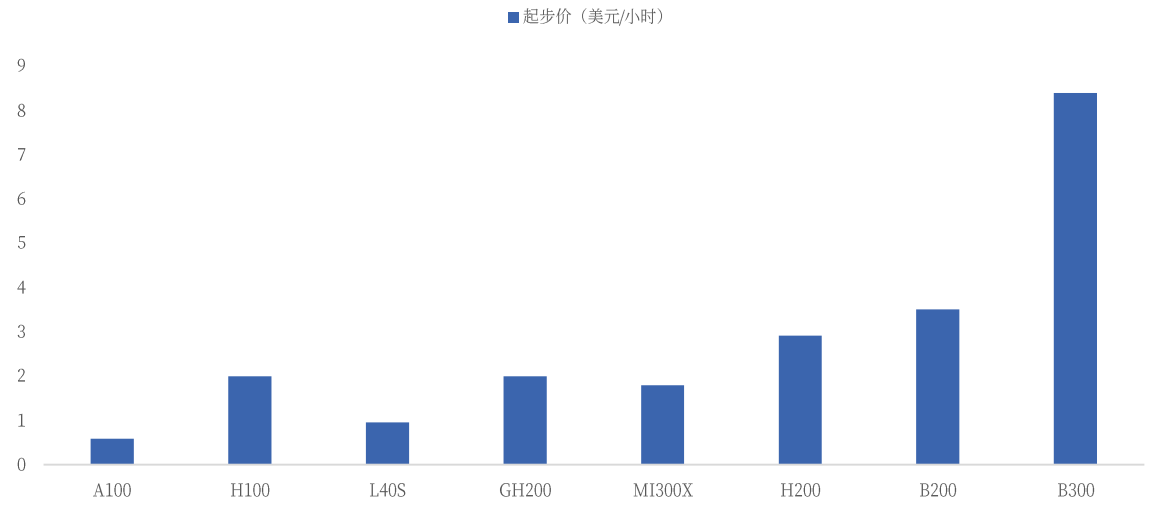
（二）热门 GPU 价格走势

Cloud-gpus.com 数据显示（截止 2026.10.25），最新热门 GPU 起步价（美元/小时）情况如

下载日志已记录，仅供内部参考

下图所示。按报价水平可分三档：A100、L40S 等成熟型号处于低位，最低报价分别为每小时$0.58、$0.95；MI300X、H100、GH200 居中，最低报价分别为每小时$1.79、$1.99、$1.99；H200 及 Blackwell 架构的 B200、B300 居于高位，最低报价分别为每小时$2.91、$3.50、$8.39。整体来看，起步价仍随架构代际呈台阶式抬升，同代内 B300 报价约为 B200 的 2.4 倍。与上期相比，八个型号中有七个的最低报价保持不变，仅 B300 由每小时$8.40 微降至$8.39，各代际报价中枢整体稳定，未出现随二级市场估值调整而松动的迹象。

图11：最新热门 GPU 起步价情况



资料来源：Cloud-gpus.com，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 算力的供给验证由需求端转向产能与制程端：台积电第三季度营收同比增长 50.9%、7 纳米及以下制程占晶圆收入 79%，第四季度营收指引 523-543 亿美元；ASML 第三季度净销售额 117 亿欧元；Lam Research 季度营收 82.19 亿美元、营业利益率 40.0%；SK 海力士在 OCP 展出 256GB CMM-DDR5，支持 CXL 3.2、带宽 80GB/s、能效改善约 30%。先进制程、光刻与高容量内存三条供给链的扩产仍在延续，约束由晶圆产能向先进封装与存储容量转移。
运营商	运营商的增长来源由用户扩张转向存量 ARPU 与光纤、固定无线：AT&T 第三季度营收 323 亿美元、同比增长 5.3%，后付费电话净增 37.9 万、流失率 0.92%，高级家庭互联网净增 69.6 万，2026 年调整后每股收益指引上调至约 2.40 美元，全年资本开支指引 230-240 亿美元。在卫星直连进入商用准备阶段的背景下，地面网络以容量与时延优势维持核心地位。
光通信	板块估值随资本开支融资担忧快速调整：双周内光通信板块下跌 24.24%，为七大板块中跌幅最大，Ciena、Lumentum、Coherent、Fabrinet 分别下跌 28.79%、31.47%、33.05%、35.23%，费指同期下跌 14.26%。产业侧，1.6T 相干与高速光模块的升级周期与器件产能排期未出现方向性变化，估值调整更多反映贴现率与资本开支预期的重定价；四季度业绩季的订单与毛利率指引是验证两者能否收敛的关键。
卫星通信	卫星直连的商业化由星座建设转向网络就绪与用户计费：SpaceX 第三季度营收 146 亿美元、同比增长 161%，星链消费者用户达 1,360 万、季度净增 160 万，季度资本开支 195.93 亿美元、其中算力业务 164.44 亿美元；AST SpaceMobile

下载日志已记录，仅供内部参考

	全球网关超过 60 个、美国已启用超 3,000 个覆盖小区，计划年底前在五国提供初始间歇性服务。竞争维度由发射成本延伸至网关密度、频谱资源与终端生态兼容性。
行业应用	AI 应用的重心由模型能力转向主动式智能体，端侧本地推理成为并行的落地路径：Anthropic 推出 Claude Duet 研究预览版，以每日简报与持续监测实现主动执行，可调用 6,000 余个连接器；Meta 发布开放权重的 Muse Spark Open，总参数 8,160 亿、激活约 420 亿、上下文 100 万 token，经量化压缩后约 493GB，可在 2 至 4 台 M5 Ultra Mac Studio 上分布式推理，生成速度约 36.7 tokens/秒。应用侧的算力需求因此由单一云端推理分化为云端长上下文与端侧本地集群两类，端侧对内存容量与互联带宽的要求同步上升。
算力服务	AIDC 的核心变量由容量扩张转向资本开支的融资结构与回报周期：Meta 第三季度资本开支 349 亿美元、自由现金流为-21 亿美元，2026 年指引收窄至 1,350-1,450 亿美元；微软宣布自研 Maia 加速器自 2027 年起扩至多吉瓦并成为主要推理系统；Firebird 取得 4.5 万颗 Grace Blackwell 出口许可、哈萨克斯坦二期规划约 500MW。算力供给的增量正由北美超大规模厂商向主权 AI 项目扩散。
液冷	液冷由部件选型转为主机架级接口标准：NVIDIA 在 OCP 发布 MGX 1.2 规范，新增机架级盲插液冷连接、5,000 安培液冷母线与无风扇液冷计算托盘，单机架支持最高 250 千瓦；Trane Technologies 收购加拿大换热器厂商 Cancoil，以垂直整合锁定关键部件产能。功率密度提升使冷却接口的标准化与部件供应链的产能保障同时成为交付瓶颈。

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
10 月 26 日（周一）	财报	Celestica（CLS）2026 年第三季度业绩，关注 AI 网络与服务器业务的订单与毛利率
10 月 28 日（周三）	财报	Vertiv（VRT）、Amphenol（APH）与 T-Mobile US（TMUS）2026 年第三季度业绩
10 月 28 日（周三）	宏观	美联储 FOMC 利率决议，市场关注在关键经济数据发布节奏受扰下的政策路径
10 月 29 日（周四）	财报	Intel（INTC）与 KLA（KLAC）2026 年第三季度业绩，关注晶圆代工与设备订单指引
10 月下旬-11 月上旬	财报	海外算力产业链三季度财报季：微软、谷歌与亚马逊的资本开支指引，以及 Ciena、Coherent、Lumentum、Fabrinet、Arista Networks、AMD 的业绩与订单展望；应用与端侧环节关注 Apple、Qualcomm 的季度指引，以及管理层对端侧 AI 装机与推理负载的表述。
11 月 3 日（周二）	政治事件	美国中期选举
持续跟踪	产业事件	AST SpaceMobile 五国初始服务的商用开通进度；Firebird 哈萨克斯坦园区的设备到货与并网；微软 Maia 与 Meta 自研加速器的量产节奏；存储与光器件的合约价与产能排期更新；端侧落地节奏，包括 Meta 与运营商合作的 AI 眼镜上市推进、Windows 本地推理能力的推送范围，以及 Claude Duet 候补名单的开放进度。

资料来源：中国银河研究院整理

五、投资建议

台积电公布第三季度业绩，季度营收同比增长 50.9%、净利同比增长 68.7%，7 纳米及以下制程占晶圆收入 79%，并指引第四季度营收 523-543 亿美元，先进制程与存储的产能利用率仍处于高

下载日志已记录，仅供内部参考

位，AI 加速器与高容量内存的订单预期尚未下修；SK 海力士在 OCP 峰会展示 256GB CMM-DDR5，支持 CXL 3.2、带宽 80GB/s，约束环节已由晶圆产能转向先进封装、HBM 堆叠与 CXL 内存的容量爬坡，其中 CXL 内存池化使单机架的内存扩展摆脱 DIMM 插槽限制，为长上下文推理提供了新的容量路径。

SpaceX 披露第三季度营收 146 亿美元、同比增长 161%，星链用户达 1,360 万；AST SpaceMobile 计划年底前在五国提供初始间歇性天基服务，全球网关超过 60 个。我们认为，卫星直连由星座建设进入网络就绪阶段，网关密度与覆盖小区数量成为商用开通的前置条件，SpaceX 与 AST SpaceMobile 分别以自建运营商与分成合作两种路径推进，频谱资源与终端生态的兼容性决定长期份额。微软宣布自研 Maia 加速器自 2027 年起扩至多吉瓦并成为主要推理系统；Anthropic 推出 Claude Duet 研究预览版，Meta 发布开放权重的 Muse Spark Open。主动式智能体与本地化部署同时推进，Claude Duet 的持续监测模式将单用户调用频次由会话驱动转为轮询驱动，Muse Spark Open 则在桌面级集群上实现了 8,160 亿参数模型的分布式推理，推理负载的分布结构正由云端集中转向分层部署。

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：台积电（TSM），ASML（ASML），Lam Research（LRCX），Broadcom（AVGO），Micron Technology（MU）等；运营商相关标的：AT&T（T），Verizon（VZ），T-Mobile US（TMUS），Comcast（CMCSA）等；算力服务与行业应用相关标的：Microsoft（MSFT），Meta（META），Oracle（ORCL），CoreWeave（CRWV），Nebius（NBIS），Nvidia（NVDA），Vertiv（VRT），Trane Technologies（TT）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。

下载日志已记录，仅供内部参考

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） 10
图 10： 本周各模型市场份额（%，按请求数计） 10
图 11： 最新热门 GPU 起步价情况 11

表 1： 本周热门模型 Token 调用量排名 10
表 2： 子行业技术趋势 11
表 3： 行业大事前瞻 12

下载日志已记录，仅供内部参考

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