下载日志已记录，仅供内部参考

美联储加息抬升无风险利率，AI 步入理性验证

2026 年 09 月 20 日

核心观点

- **双周美股主要指数表现分化，科技板块强于大盘：**双周（2026.09.07-2026.09.18）主要指数中，S&P500 涨跌幅为-0.88%，纳指涨跌幅为+0.06%，道指涨跌幅为-1.02%；科技行业方面，以科技成长股为主的 Nasdaq-100 指数涨跌幅为+0.34%，费指涨跌幅为+1.59%，表现强于大盘。年初至今涨跌幅方面，费指涨幅为+68.31%，表现远超大盘，且远高于 2025 年同期（+26.07%）及 2024 年同期（+16.38%）水平，表明半导体板块中长期交易热度仍处于高位。双周市场走势主要受美联储货币政策转向影响：美联储于 9 月 16 日加息 25 个基点至 3.75%-4.00%，为 2023 年 7 月以来首次加息，点阵图显示政策利率预测中值上移、会后声明表述偏鹰，10 年期美债收益率上行至约 4.95%，压低市场风险偏好并压制利率敏感板块；同期 AI 算力资本开支密集验证，支撑半导体与 AI 基础设施相关标的走强。子行业方面，Nasdaq-100 中，双周涨幅靠前标的主要集中于半导体、硬件厂商等板块；双周跌幅靠前公司主要集中于电力公用事业、媒体及消费互联网等板块。我们判断主要因素为，AI 算力资本开支延续高强度验证，叠加超大规模算力集群对定制加速器与高速互联的持续投入，半导体等环节景气度居前。

- **核心观点：**

  美联储于 9 月 16 日加息 25 个基点至 3.75%-4.00%，为 2023 年 7 月以来首次加息。我们认为，分母端：无风险利率中枢系统性上移，且抬升的不只是政策利率。长端 10 年期美债 9 月中旬盘中触及 5.02%，创 2007 年以来新高，其驱动已不单是政策预期，而是油价高企、美债规模的供给压力、AI 基础设施企业大举发债与国债争夺资金所共同推高的期限溢价，意味着即便美联储停止加息，长端利率也难快速回落。对权益资产而言，贴现率抬升首先压缩的是市盈率，尤其是依赖远期现金流的高估值成长股。**分子端：**AI 叙事并未降温，产业景气仍高涨，只是定价逻辑从"远期愿景"切换到"当期兑现"。三季度以来海外利率预期反复与板块拥挤度上升，已推动全球主要科技指数完成一轮较为充分的估值消化，为修复留出空间。加息落地后首个交易日，存储、CPU、代工、光互联等算力产业链环节普遍占优：当期盈利强劲、投资回报周期短，使其对分母端利率上行具备明显钝感。**因此，分化并未终结，定价框架正在切换。有当期现金流与业绩验证的算力硬件（先进制程、存储、光互联、设备材料）受益于利空出尽；依赖远期叙事、缺乏盈利支撑的标的则被持续挤出，AI 行业步入理性验证阶段。**

- **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

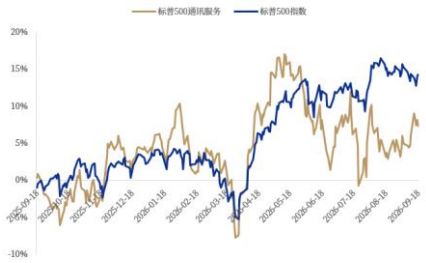
☎：010-8092-7619

✉：zhaoliangbi_yj@chinastock.com.cn

分析师登记编码：S0130522030003

相对标普 500 表现图

2026 年 09 月 18 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】行业双周报_OpenAI 发布 GPT-6 Astra，推理需求上修

3.【银河通信】行业点评_产业韧性凸显，美联储加息不改 AI 长期逻辑

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 4
（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 7
（四） 算力服务 7
（五） 液冷 9
（六） 行业应用 9
三、 行业数据跟踪 10
（一） 大模型 Tokens 调用量 10
（二） 热门 GPU 价格走势 11
（三） 子行业技术趋势 12
四、 行业大事前瞻 13
五、 投资建议 13
六、 风险提示 14

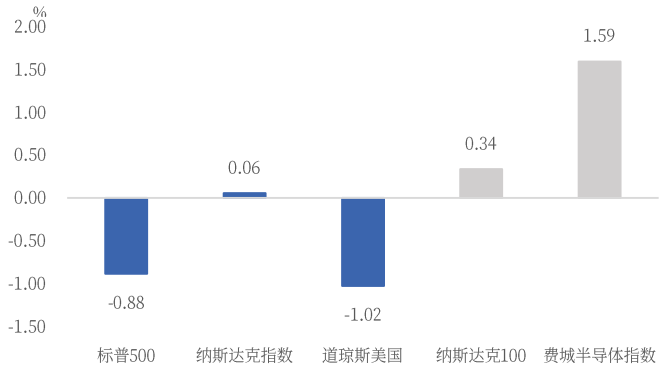
下载日志已记录，仅供内部参考

一、双周行情回顾

（一）海外主要指数回顾

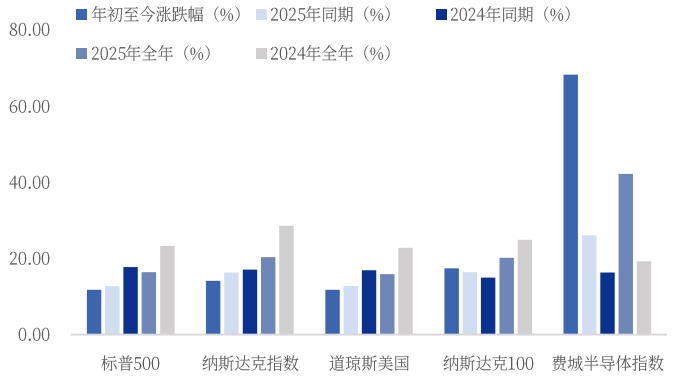
双周美股主要指数表现分化，科技板块强于大盘：双周（2026.09.07-2026.09.18）主要指数中，S&P500 涨跌幅为-0.88%，纳指涨跌幅为+0.06%，道指涨跌幅为-1.02%；科技行业方面，以科技成长股为主的 Nasdaq-100 指数涨跌幅为+0.34%，费指涨跌幅为+1.59%，表现强于大盘。年初至今涨跌幅方面，费指涨幅为+68.31%，表现远超大盘，且远高于 2025 年同期（+26.07%）及 2024 年同期（+16.38%）水平，表明半导体板块中长期交易热度仍处于高位。双周市场走势主要受美联储货币政策转向影响：美联储于 9 月 16 日加息 25 个基点至 3.75%-4.00%，为 2023 年 7 月以来首次加息，点阵图显示政策利率预测中值上移、会后声明表述偏鹰，10 年期美债收益率上行至约 4.95%，压低市场风险偏好并压制利率敏感板块；同期 AI 算力资本开支密集验证，支撑半导体与 AI 基础设施相关标的走强。

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

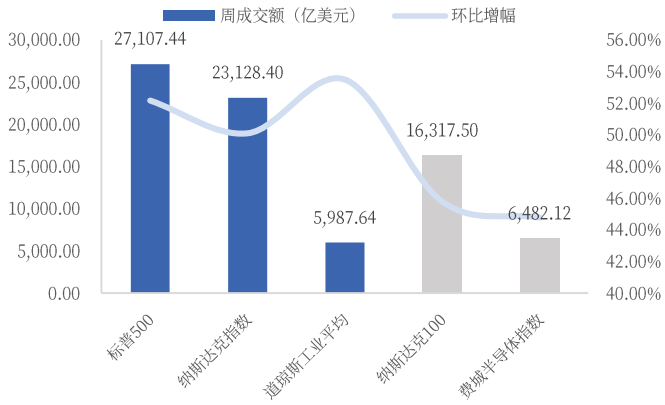
图2：各主要指数年初至今涨跌幅对比



资料来源：Wind，中国银河证券研究院

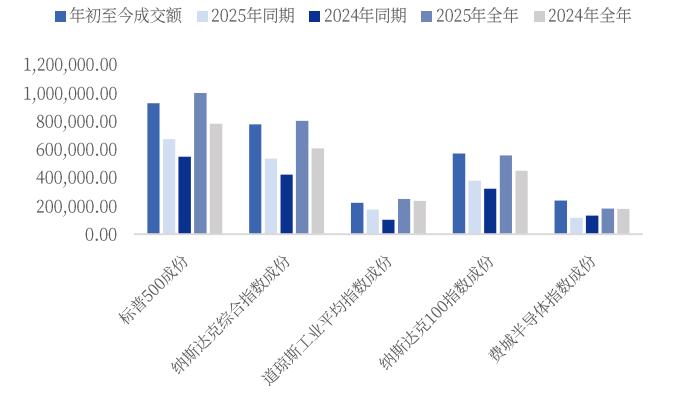
成交额方面，双周科技指数成交额绝对值不及大盘，环比均上升。相比前两年同期，当前年度成交额整体仍处于较高水平，市场交投活跃度好于 2025 及 2024 年同期水平。

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）



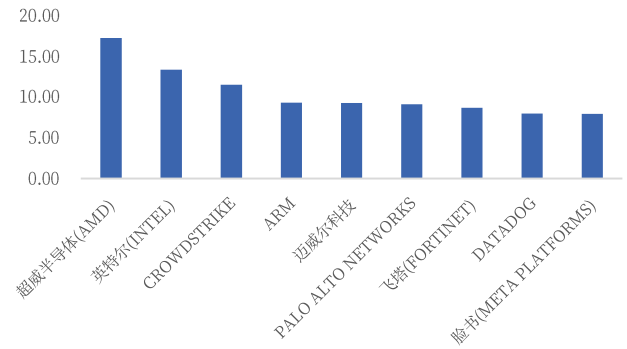
资料来源：Wind，中国银河证券研究院

下载日志已记录，仅供内部参考

（二）各子行业表现情况

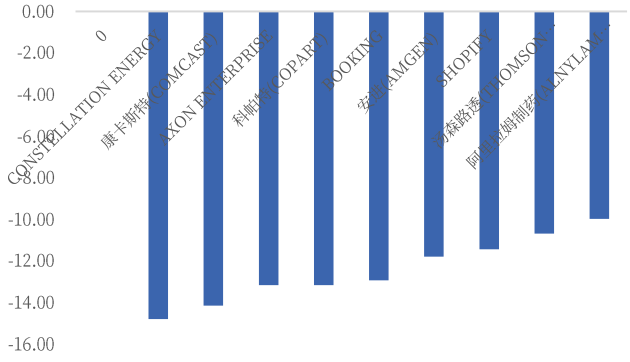
子行业方面，Nasdaq-100 中，双周涨幅靠前标的主要集中于半导体、硬件厂商等板块；双周跌幅靠前公司主要集中于电力公用事业、媒体及消费互联网等板块。我们判断主要因素为，AI 算力资本开支延续高强度验证，叠加超大规模算力集群对定制加速器与高速互联的持续投入，半导体等环节景气度居前。

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（%）

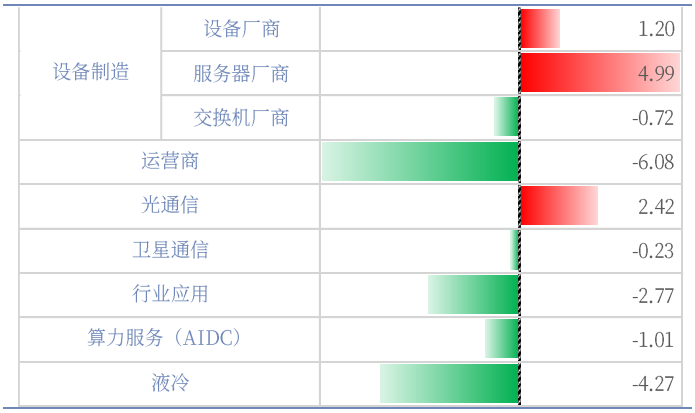


资料来源：Wind，中国银河证券研究院

（三）双周行业龙头标的表现

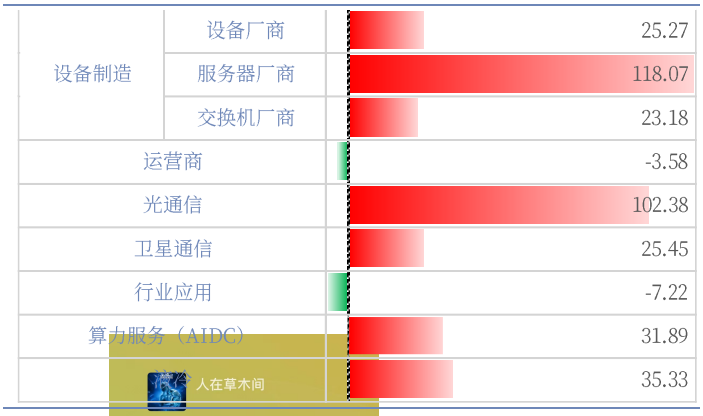
根据我们对于海外算力行业划分子板块，海外算力板块子行业包括设备制造（设备厂商、服务器厂商、交换机厂商）、运营商、光通信、卫星通信、行业应用、算力服务（AIDC）、液冷七大板块。双周内板块表现分化，仅设备厂商、服务器厂商和光通信板块录得上涨，涨幅分别为 1.20%、4.99%、2.42%；回调板块方面，跌幅居前的板块分别为运营商、液冷、行业应用，跌幅分别为 6.08%、4.27%、2.77%。

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 AMD、Hewlett Packard Enterprise、联想集团、AXT、Intel 等，涨幅分别为+17.22%、+17.14%、+14.28%、+13.61%、+13.36%。回调标的按照跌幅降序分别为 Charter Communications、Comcast、Hubspot、UiPath、Vertiv，跌幅分别为 15.67%、14.16%、12.71%、11.85%、11.08%，主要集中于运营商、行业应用及液冷板块标的。

人在草木间

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

■ Micron（MU.O）：展示全球首款 512GB DDR5 RDIMM，双路服务器内存容量突破 12TB

Micron 于 9 月 15 日宣布，已在多个服务器平台成功演示全球首款 512GB DDR5 寄存式双列直插内存模块（RDIMM），最高速率达 9,200 MT/s。该模块采用垂直互连封装技术，通过硅通孔（TSV）垂直堆叠 DRAM 裸片，在单台 24 插槽双路服务器中可实现最高 12TB 的 DDR5 内存容量；能效方面，相比采用 4 条 128GB RDIMM 的组合方案，运行功耗降低超过 60%；在 Spark 支持向量机（SVM）数据分析等内存受限场景下，性能较 256GB DDR5 配置最高提升 1.4 倍。AMD 与 Intel 均在验证该模块，公司预计于 2027 年下半年依据客户需求进入量产。

AI 推理、内存数据库与智能体工作负载的快速普及，持续推高服务器对大容量、高带宽内存的需求。512GB RDIMM 通过让更大规模数据集常驻主内存，减少了对低速存储层的依赖与数据搬运开销，是缓解 AI 服务器"内存墙"约束的新路径。存储原厂的竞争焦点正由单纯制程微缩转向封装密度与能效，后续重点跟踪高容量 DDR5 与 HBM 之间的产能分配及其价格走势。

■ ASML（ASML.O）：与台积电、三星、英特尔同步推进 High-NA EUV 落地，启动荷兰第二产业园区

ASML 于 9 月 8 日同日发布三份公告，披露与三家主要客户在 High-NA EUV 上的落地节奏：公司与台积电建立行业合作倡议，推动 EUV 光掩模由 6 英寸向 12 英寸升级，目标 2031 年建成 12 英寸掩模试产线、2033 年支持其进入先进节点生产，台积电计划自 2030 年起在先进制程量产中采用 High-NA 技术；三星计划于 2028 年前在业内首次将 High-NA EUV 用于 DRAM 大规模生产；英特尔披露 High-NA 已在量产线上累计处理超过 100 万片晶圆，覆盖设备认证、研发及酷睿 Ultra 系列 3 处理器部分层级的量产。此外，公司宣布在荷兰 Brainport 地区启动第二座产业园区建设，规划建筑面积约 35 万平方米、可容纳最多 2 万名员工。

High-NA EUV 的导入节奏直接决定 2 纳米以下制程的成本曲线，12 英寸掩模有望进一步提升光刻机产能并消除拼接限制。三家客户给出的时间表存在明显差异（存储 2028 年、逻辑 2030 年），既反映各家公司产品结构与制程节点的不同，也表明先进制程与先进存储的设备投入已进入可规划阶段。后续重点跟踪 High-NA 设备的确认收入节奏、掩模生态建设进展，以及光刻环节对上游零部件与材料的拉动。

■ Ciena（CIEN.N）：投资者论坛给出 2029 财年目标，营收复合增速约 30%

Ciena 于 9 月 16 日至 17 日在渥太华研发中心举办 2026 年投资者论坛，公布截至 2029 财年的三年财务目标：2026 至 2029 财年营收年复合增长率约 30%（对应 2029 财年约 140 亿美元）、调整后毛利率约 50%、调整后营业利润率 32%至 35%、自由现金流利润率约 20%；公司预计 2027 财年营收为 83 亿至 84 亿美元，同比增长至少 30%。公司同时宣布自 2027 财年第一季度起启用光系统、互连、全球服务、路由及其他四个新报告分部。据公司在论坛披露，其可服务市场（TAM）预计由 2026 年约 270 亿美元翻倍至 2029 年约 520 亿美元，订单（orders）连续两年翻倍、2026 年再增至少 50%（合计约为 2024 年的三倍），2026 财年末积压订单（backlog）约 100 亿美元，当前需求明显超过供给。

Ciena 的三年指引是光通信产业链对 AI 带宽需求最具约束力的量化承诺之一。公司援引的行业假设为：超大规模云厂商 2026 年资本开支约 1.2 万亿美元、2027 年增至至少 1.6 万亿美元，数

下载日志已记录，仅供内部参考

据中心带宽需求年复合增速约 36%，其中推理工作负载增速（约 36%）已超过训练（约 23%）。管理层将增速瓶颈归因于供给侧而非需求侧，并援引其经营现金流较 2024 年增长约 5 倍，以此支撑供应链扩张与订单交付。后续重点跟踪其订单转化节奏与扩产进度。

■ **Coherent（COHR.N）：携全栈光子方案亮相 ECOC 2026，聚焦 CPO 与 NPO 架构**

Coherent 于 9 月 8 日宣布，公司将参加 9 月 21 日至 23 日在西班牙马拉加举行的欧洲光通信大会（ECOC 2026），首席营销官 Sanjai Parthasarathi、首席技术官 Julie Sheridan Eng 等将参与多场主题论坛，议题覆盖下一代 AI 光互连、共封装光学（CPO）、纵向扩展与横向扩展网络架构，以及数据中心冷却与能效；公司将在展台展示面向 AI 数据中心与高速光通信的最新方案，包括覆盖光生成、光束整形、传输到探测完整信号链、支持 CPO 与 NPO 架构的 PhotonLink 平台。公司于 9 月 17 日另宣布扩展其可插拔光线路系统组合，新增全 C 波段放大能力。

随着光互连由机架间（scale-out）向机架内（scale-up）延伸，光器件的可服务位置与单系统价值量同步抬升。市场对 CPO 是否将替代传统可插拔模块存在分歧，但从光信号链的角度看，激光器、光子组件、光引擎与封装环节的总体需求趋于增加，价值分配向掌握核心光子器件能力的厂商集中。后续重点跟踪 ECOC 期间各厂商公布的 CPO/NPO 量产时间表与客户验证进展。

（二）卫星通信

■ **Rocket Lab（RKLB.O）：完成第 96 次 Electron 发射，累计为 Synspective 部署 12 颗 StriX 卫星**

Rocket Lab 于 9 月 19 日成功执行第 96 次 Electron 火箭发射任务（代号"Owl By The Dozen"），从新西兰发射场 1 号复合体升空，将日本 Synspective 公司的第 12 颗 StriX 合成孔径雷达（SAR）卫星直接送入 572 公里近地轨道，任务全部成功。该任务系 2026 年以来 Electron 的第 17 次发射，Rocket Lab 在商业小型运载领域的发射频次与可靠性继续保持领先。

Synspective 正在建设由约 30 颗 StriX 卫星组成的 SAR 对地观测星座，目前已由 Rocket Lab 累计部署 12 颗，仍有 15 次专用发射在列，计划于 2030 年前完成星座组网。SAR 技术可在昼夜及云层条件下获取高分辨率地表图像，主要用于灾害响应、基础设施监测与资源管理。Rocket Lab 表示，稳定的发射批次有助于公司规划火箭量产与发射场排期。

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。但需关注的是，双周内公司股价累计上涨 0.48%，涨幅明显弱于同属算力硬件板块的标的，Neutron 中型火箭研发进展及后续发射合同执行情况后续值得关注。

■ **SpaceX：披露百亿美元 AI 算力托管订单，星舰将首次入轨并部署量产版 V3 星链**

SpaceX 首席财务官 Bret Johnsen 于 9 月 10 日出席 2026 年高盛 Communacopia+科技大会时披露，公司近期新签一笔大额 AI 算力托管订单，协议将于 12 月 1 日正式生效，可为公司带来每月约 11.1 亿美元的新增收入，对应年化增量约 133 亿美元，客户身份未予披露。叠加该笔订单，公司对年底达成 1,000 亿美元年度经常性收入（ARR）目标保持信心。公司同时披露，星舰第 14 次飞行任务将成为其首次产生收入的飞行，并首次将星舰送入地球轨道、部署量产版 V3 星链卫星。

以自有发射能力与低轨星座为基础，SpaceX 正由卫星互联网运营商向算力托管服务商延伸。百亿美元级别的签约规模进一步印证 AI 算力需求的广度已从传统云计算厂商扩散至卫星互联网等新兴运营商，也反映出在地面电力与并网受限背景下，具备能源与基础设施自建能力的主体在算力供给上的相对优势。后续重点跟踪该订单的客户结构与交付节点，以及星舰复用进度对单位入轨成本的实际影响。

下载日志已记录，仅供内部参考

（三）运营商

■ Verizon（VZ.N）：与 Corning 签署至 2032 年光纤供应协议，采购逾 8000 万英里高密度光纤

Verizon 于 9 月 8 日宣布与 Corning 达成数十亿美元多年期供应协议，2027 年至 2032 年间，Corning 将向 Verizon 交付超过 8000 万英里高密度光纤及连接解决方案，产品包括 Corning® Contour™ Flow Cable，可在既有管道资源内部署更多光纤。该批光纤将同时用于家庭与企业宽带接入，以及 AI 数据中心之间的长距骨干网建设，支撑 Verizon 4000 万至 5000 万宽带覆盖目标。

Verizon Business 首席执行官 Kyle Malady 表示，公司正在部署单一融合架构，将移动连接与宽带统一为一体化体验，同时构建 AI 超大规模客户所需的高容量、低时延骨干传输底座；Corning 光通信高级副总裁 Mike O'Day 表示，本次供货方案同时覆盖美国家庭高速光纤接入与 AI 数据中心互联需求，并将在美国本土扩大制造产能。

该协议将宽带接入与 AI 数据中心互联纳入同一张融合网络，显示运营商正把光纤资产从传统宽带基建重新定位为 AI 骨干网的核心物理底座。在 AI 工作负载对近零时延与大吞吐的刚性要求下，高密度光纤的长期供给锁定体现出产业链对需求弹性的判断；后续重点跟踪 Verizon AI Connect 战略下长距走廊的建设进度及超大规模客户签约情况。

■ Deutsche Telekom（DTEGY.OO）：Digital X 2026 提出五大 AI 优先级，拟竞标欧盟 AI 超级工厂

Deutsche Telekom 于 9 月 8 日至 10 日在科隆举办 Digital X 2026，CEO Tim Höttges 在会上宣布公司将筹备欧盟 AI Gigafactory（AI 超级工厂）计划竞标并提供相应基础设施建设能力，同时继续扩建位于慕尼黑的工业 AI 云（Industrial AI Cloud）。欧盟该计划拟建设最多七座 AI 超级工厂，公共资金规模最高 100 亿欧元、带动最少 200 亿欧元私人投资，申请截止日期为 11 月 12 日，预计 2027 年初作出决定。

Höttges 同时提出德国 AI 竞争力的五项重点，强调数字主权应体现为对数据、流程与技术的掌控能力，而非技术孤立；公司披露 AI 目前已承担其约 60%的客户问询处理。Digital X 2026 吸引超过 6000 名参与者，议题涵盖数字主权、网络安全、AI 智能体与云基础设施。

欧洲运营商正由传统连接服务向主权 AI 基础设施延伸，以数据中心、网络与能源资产切入算力供给。若欧盟 AI 超级工厂计划落地，欧洲区域的 GPU 算力与配套基础设施投资有望形成新的增量需求；后续重点跟踪竞标结果及工业 AI 云的建设节奏。

（四）算力服务

■ Oracle（ORCL.N）：2027 财年第一季度业绩超预期，剩余履约义务升至 6640 亿美元

Oracle 于 9 月 10 日发布 2027 财年第一季度业绩（截至 2026 年 8 月 31 日），季度总营收 193.45 亿美元，同比增长 30%；其中云收入 116.07 亿美元，同比增长 62%，云基础设施（IaaS）收入 74 亿美元，同比增长 121%；GAAP 摊薄每股收益 1.56 美元，同比增长 55%，非 GAAP 摊薄每股收益 1.92 美元，同比增长 30%；GAAP 经营利润 67.28 亿美元，同比增长 57%；经营现金流 230 亿美元，同比增长 184%，自由现金流为-50 亿美元。

订单与产能方面，公司剩余履约义务（RPO）升至 6640 亿美元，同比增加 2090 亿美元、环比增加 260 亿美元，季度内新增 AI 云合同超过 300 亿美元，且大部分采用预付款或客户自带硬件等方式；公司当季向 AI 云客户交付超过 30 万张 GPU，接近上一财季的三倍，并新增 850 兆瓦数据

下载日志已记录，仅供内部参考

中心容量。公司同步将 2027 财年营收指引上调至不低于 900 亿美元、非 GAAP 每股收益指引 8.10 美元，资本开支维持 900 亿至 950 亿美元。

作为 AI 算力产业链资本开支与订单的先行指标，Oracle 的 RPO 与 IaaS 增速进一步确认 AI 基础设施需求仍快于供给。公司当季完成 200 亿美元 ATM 股权融资，自由现金流转负，反映大规模算力扩张对资金占用与融资结构的压力；后续重点跟踪订单向收入转化的节奏、资本开支强度以及毛利率随基础设施上量的变化。

■ CoreWeave（CRWV.O）：三季度以更高价格签约算力容量，同步启动 37 亿美元可转债发行

CoreWeave 于 9 月 17 日披露，自 2026 年 6 月 30 日以来，公司持续以更高价格签订新的算力容量合同；2026 年第三季度签署的短期限客户合约期限约为 3 至 6 个月，定价约为 4000 万美元/兆瓦，该价格以年化收入除以服务相关集群所需电力计算。公司重申第三季度初新增超过 250 亿美元客户净承诺（未计入截至 6 月 30 日的收入积压订单），签约电力总容量截至 8 月 11 日增至约 4.2 吉瓦，6 月 30 日约为 3.7 吉瓦。

9 月 17 日，公司宣布拟非公开发行 30 亿美元 2033 年到期可转换优先票据，后因需求上调至 37 亿美元，票面利率 2.875%，初始转股价约 97.85 美元/股，较 9 月 17 日收盘价 79.88 美元溢价约 22.50%，预计 9 月 22 日完成交割；公司另设立最多 3500 万股 A 类普通股的 ATM 增发计划，用于增强融资灵活性并推动信用状况向投资级迁移。

短期限、高单价的签约结构表明算力买方对即时交付的支付意愿显著抬升，供需缺口下算力服务商的定价权进一步增强；但短约也意味着收入能见度偏低、需按季重新定价，若需求转弱则面临向下重定价风险。叠加股权与可转债融资安排，算力服务商的融资能力与资本结构正成为下一阶段分化的关键变量。

■ Nebius（NBIS.O）：三个月内第二次提价，Nvidia GPU 按需价格上调 17%-21%

Nebius 于 9 月 17 日向客户发布调价通知，自 10 月 1 日起上调多项按需计算资源价格。其中 Nvidia GPU 涨幅为 17%至 21%：H100 由 3.85 美元/GPU 小时升至 4.50 美元，H200 由 4.50 美元升至 5.40 美元，B200 由 7.15 美元升至 8.50 美元，B300 由 7.85 美元升至 9.50 美元；纯 CPU 实例价格上调 25%，内存产品价格上调约 41%。预订多个月大规模集群的客户仍可获得承诺折扣。

这是 Nebius 三个月内的第二次提价，今年 5 月公司已对按需容量平均提价 29%、对可抢占容量提价 51%；以 B300 为例，其价格自 5 月前的约 6.10 美元/小时累计涨幅已达约 56%。管理层此前表示，需求之强劲使价格调整仍不足以平衡供需，公司曾对稀缺的 Blackwell 算力进行拍卖测试，客户实际支付价格较此前最高定价高出 15%至 20%。

本轮提价同时覆盖 Hopper 与 Blackwell 两代架构，表明算力紧缺并非局限于最新一代芯片，而是广泛分布于各类 AI 算力资源；结合 CoreWeave 同期以更高价格签约，AI 算力供不应求的行业格局得到交叉验证。后续重点跟踪提价对中小 AI 客户需求弹性的影响，以及算力服务商利润率与利用率的边际变化。

■ Google（GOOGL.O）：两年内向芬兰投资至少 130 亿欧元，配套 22 年期核电购电协议

Google 于 9 月 9 日宣布，计划在 2027 年至 2028 年向芬兰投入至少 130 亿欧元用于数据中心及配套数字基础设施，具体包括扩建位于东南部城市哈米纳的现有数据中心，并在卡亚尼、穆霍斯与瓦拉新建三座设施，为 Gemini、搜索、地图、YouTube 及云服务提供算力支撑；该项目为公司迄今在欧洲的单笔最大投资。能源配套方面，公司与芬兰能源公司 Fortum 签署 22 年期购电协议，为洛维萨核电站延寿提供长期收入保障；在卡亚尼配套建设 94 兆瓦电池储能系统以增强电网调峰

下载日志已记录，仅供内部参考

能力；并设立社区基金支持当地生物多样性、教育与劳动力发展。

据公司测算，项目建设阶段（2027—2028 年）将年均贡献芬兰 GDP 约 36 亿欧元、支撑全国超过 3.7 万个就业岗位；设施投运后每年可支撑约 7000 个岗位，相关岗位平均薪资高于芬兰工资中位数 24%。选址逻辑上，Google 将新建数据中心布局于芬兰北部而非南部，以利用既有电网基础设施与无碳能源资源，配合输电运营商 Fingrid 的并网规划降低系统性成本。

超大规模云厂商的算力布局正从电力充裕的既有枢纽向低碳电力富集区域迁移。Google 首次以"购电协议+储能+社区基金"的组合方式锁定长期能源供给，显示电力获取能力已成为数据中心选址的首要约束，也为北欧、加拿大等低碳电力富集地区的数据中心投资提供了可复制路径。后续重点跟踪其他超大规模厂商在欧洲的能源配套安排与并网节奏。

（五）液冷

■ Colovore：里诺首个液冷 AI 数据中心投运，零水冷却支撑 5-600kW 机柜

Colovore 于 9 月 9 日宣布，其位于内华达州斯帕克斯市 RNO01 园区的首座液冷 AI 数据中心正式投运，该设施提供 6 兆瓦液冷基础设施，且已被客户全部租用。设施采用零水冷却设计，不使用任何市政供水，水使用效率（WUE）为 0，目标电能使用效率（PUE）为 1.3；基于 direct-to-chip 液冷架构，可支持单机柜 5 至 600 千瓦以上的部署密度，面向最新一代 GPU 与加速器平台。园区第二座数据中心已在建设中，预计 2027 年 2 月完工。

单机柜功率密度向数百千瓦乃至兆瓦级跃迁，使液冷从可选配置转为高密度部署的基础设计架构；"零水冷却"路径则在缓解用水约束的同时，降低了项目在缺水地区的选址门槛。该项目投运即满租的状态，反映液冷 AIDC 在电力与水资源的双重限制下仍具备较强的需求承接能力。后续重点跟踪其第二座数据中心的建设进度，以及此类专用液冷园区在硅谷以外区域的复制速度。

■ Schneider Electric：推出最高 2.5MW 预制电力模块，集成 NVIDIA GB300 液冷控制

Schneider Electric 于 9 月 16 日发布新一代标准化预制电力模块（Prefabricated Power Modules）与 Power Skid，单模块容量最高 2.5 兆瓦，为公司迄今容量最大的单箱式预制电力方案；产品提供 2 兆瓦、2.25 兆瓦与 2.5 兆瓦三档配置，基于公司与 NVIDIA 联合开发、面向 GB300 NVL72 平台的最新参考设计，集成电力管理与液冷控制。预制模块在工厂内整合 Galaxy VXL UPS、开关设备、冷却系统、锂离子电池、母线槽与配电系统并完成测试；公司表示，标准化方案可将预制基础设施的制造周期由传统的 12 至 14 个月压缩至最短 6 个月。产品即日起在欧洲上市，北美将于 2026 年晚些时候推出，随后扩展至亚太等市场，制造与测试在西班牙巴塞罗那 Sant Boi de Llobregat 工厂完成。

AI 数据中心的建设瓶颈已由芯片供给转向"供电—散热"一体化交付效率，预制化、模块化方案正成为压缩 time-to-power 的关键路径。厂商将电力管理与液冷控制纳入同一参考设计，意味着机电与热管理由分立采购转向集成交付，供应链价值分配向能够提供整体解决方案的头部集成商集中。后续重点跟踪预制模块在超大规模与新型云服务商（neocloud）客户中的渗透率，以及该类集成方案对液冷控制环节的价值拉动。

（六）行业应用

■ OpenAI：ChatGPT Work 上线 Data agent，企业数据分析与智能体工作流进一步打通

OpenAI 于 9 月 10 日在 ChatGPT Work 中推出 Data agent，可直接连接企业已授权的数据源，包括 Amazon Redshift、Databricks、Snowflake、MongoDB、ClickHouse、Google BigQuery、

下载日志已记录，仅供内部参考

Datadog 等，并可引入 Google Drive 与 SharePoint 中的文件；该功能沿用企业既有权限体系（含表、行、列级限制），由管理员决定可用的数据连接与可使用角色，无需编写 SQL 或学习新的分析工具即可完成数据探查并生成可编辑、共享与刷新的交互式仪表盘。

该功能可与企业既有 BI 工具协同，包括 Tableau、Power BI、Sigma、Oracle BI、ThoughtSpot 与 Omni；语义层可由 Databricks Genie Ontology、dbt、GitHub、Snowflake Horizon 及既有 BI 仪表盘提供，使分析沿用组织自身的指标定义与数据关系。据公司披露，OpenAI 内部几乎全部产品团队及超过三分之二的市场团队已在使用 Data agent。

Data agent 将 AI 能力从内容生成与问答延伸至企业数据接入、分析与决策执行的完整工作流，使大模型在传统 BI 与数据分析场景中承担更多职能。随着企业数据与智能体的结合程度提升，AI 商业化正由单纯的模型调用向"数据接入—分析—决策支持"的闭环演进，有望进一步提升单位企业客户的推理算力消耗。

■ **DeepSeek：发布 V4.1 Flash，552B MoE 非对称架构显著压缩推理成本**

DeepSeek 于 9 月 10 日正式发布 DeepSeek V4.1 Flash，该模型为 552B 参数的 MoE 模型，采用全新的 Causal Encoder–Decoder 非对称架构，输入激活参数仅 8B、输出激活参数 16B，并具备原生多模态视觉理解能力；与上一代相比，其 KV Cache 对 HBM 的需求降至 1/4、对 SSD 的需求降至 1/8，显著降低智能体类任务的缓存成本。V4.1 Flash 已同步上线 DeepSeek API，模型名称为 deepseek-flash。

定价与路由方面，新 API 价格于 9 月 10 日 12:00 起生效并延续峰谷定价机制，闲时价格为高峰时段价格的一半；自 9 月 14 日 12:00 起至未来 V4.1 Pro 上线之前，用户访问 deepseek-v4-pro 的请求将全部路由至 V4.1 Flash，并按 V4.1 Flash 单价计费。据公司披露，经多方测试，V4.1 Flash 在性能、费用、速度与总用时等指标上已全面超越 DeepSeek V4 Pro。

KV Cache 与激活参数的压缩直接降低了长上下文与智能体类任务的使用成本，契合 Agent 工作负载对推理成本高度敏感的特征。在 Anthropic、OpenAI 等厂商同期密集迭代旗舰模型的背景下，模型层的效率竞争与价格竞争持续压缩推理成本，有助于 AI 应用侧规模化放量，并推动推理算力需求从训练集群向端侧与推理优化环节扩散。

三、行业数据跟踪

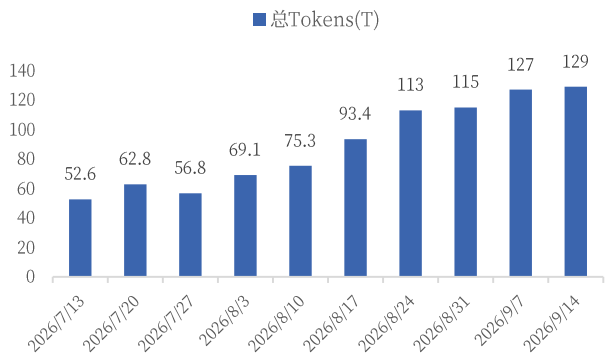
（一）大模型 Tokens 调用量

当期（统计口径为 2026.09.07-2026.09.20）模型总调用量合计约为 259.0 万亿 tokens，较上期双周（2026.08.24-2026.09.06）的 228.0 万亿 tokens 环比上升约 13.60%。近十周总调用量维持单边上行，本期两周分别录得 127T 与 132T，再度刷新周度峰值，但环比增速较上期明显收窄，AI 模型调用需求仍处扩张通道。

分模型厂商来看，最新一期 Token 市场份额中，DeepSeek 占比最高，为 25.4%，连续第三期位居首位；Google 位居第二，占比 16.3%；OpenAI 位居第三，占比 15.4%；Z-AI 位居第四，占比 12.1%；Tencent、Qwen、Mistral AI 分别占比 8.2%、6.6%、3.2%，Anthropic、NVIDIA 分别占比 2.1%、1.7%，其他模型合计占比 9.0%。DeepSeek、Google、OpenAI 合计占比为 57.1%，前五大模型厂商合计占比 77.4%，头部集中度较上期有所回落；结构上，DeepSeek、Z-AI、Tencent 份额提升，OpenAI、Google 份额下降，Mistral AI 与 NVIDIA 新进入榜，上期的 minimax 与 Meta-LLaMA 退出。

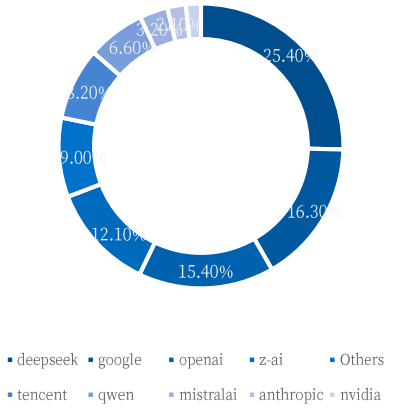
下载日志已记录，仅供内部参考

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 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 V4.1 Flash、GLM 5.3 Flash、Hy4 preview、GPT-5.6 Luna、DeepSeek V4 Flash 0731，调用量分别为 15.8T、14.1T、12.5T、9.72T、9.44T。其中，DeepSeek V4.1 Flash 以 15.8T 登顶（环比+219%），为本期最大增量来源，本期系其发布后首个完整统计周期；GLM 5.3 Flash 以 14.1T 升至第二位（环比+18%），规模延续扩张；Hy4 preview 以 12.5T 由前期首位回落至第三位（环比-26%）；GPT-5.6 Luna 以 9.72T 降至第四位（环比-47%），为本期前十中降幅最大的模型；DeepSeek V4 Flash 0731 以 9.44T 位居第五（环比-18%）。此外，Hy3、Nemotron 3 Ultra (free) 环比均+26%，GLM 5.3 环比+19%，增速靠前；MiMo-V2.5、DeepSeek V4 Flash 0423 环比分别-9%、-13%。厂商层面，DeepSeek 三款模型合计 29.01T，约占前十总调用量的 34.3%，为本期上榜数量最多的厂商，Tencent 与 Z-AI 各有两款入围。

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	deepseek	15.8T	219%
GLM 5.3 Flash	z-ai	14.1T	18%
Hy4 preview	tencent	12.5T	-26%
GPT-5.6 Luna	openai	9.72T	-47%
DeepSeek V4 Flash 0731	deepseek	9.44T	-18%
MiMo-V2.5	xiaomi	7.07T	-9%
Hy3	tencent	4.78T	26%
Nemotron 3 Ultra (free)	nvidia	4.49T	26%
DeepSeek V4 Flash 0423	deepseek	3.77T	-13%
GLM 5.3	z-ai	3T	19%

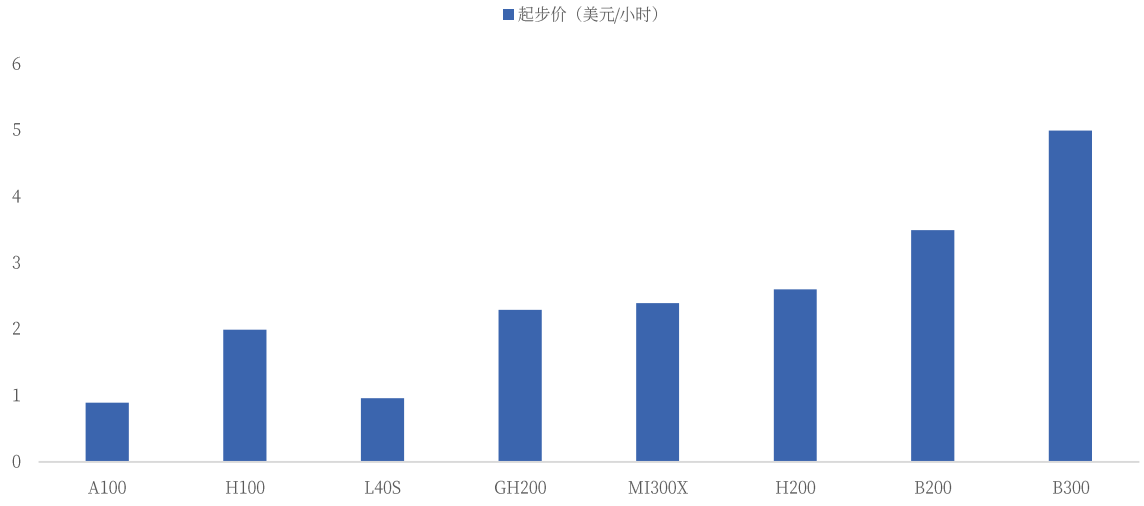
资料来源：OpenRouter，中国银河研究院

（二）热门 GPU 价格走势

下载日志已记录，仅供内部参考

GPU Rental Prices 数据显示（截止 2026.09.20），最新热门 GPU 起步价（美元/小时）情况如下图所示。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$0.89、$0.96；H100、GH200、MI300X、H200 居中，最低报价分别为每小时$1.99、$2.29、$2.39、$2.60；Blackwell 架构的 B200、B300 居于高位，最低报价分别为每小时$3.49、$4.99。整体来看，起步价随架构代际呈台阶式抬升，Blackwell 型号显著高于 A100、L40S 等成熟型号，同代内 B300 报价亦高于 B200，代际定价梯度清晰；各型号最低报价与上期持平、未见回落，显示高性能 AI 算力资源的供给稀缺性仍在延续。

图11：最新热门 GPU 起步价情况



资料来源：GPU Rental Prices，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 算力供给瓶颈进一步向物理资源延伸：NVIDIA 于 9 月 9 日宣布与澳洲数据中心生态合作，规划到 2027 年最高实现 2GW 的 DSX AI 工厂容量，其中 Sharon AI 将部署最多 68,000 张 NVIDIA GPU。本期 Oracle 单季交付超过 30 万张 GPU、新增 850 兆瓦数据中心容量，剩余履约义务升至 6640 亿美元，算力需求可见性由单季延伸至两年以上维度。供给端竞争维度同步外溢，除 GPU 排产外，土地、电力与厂房（LPS）等物理资源及并网节奏正成为新的约束，算力供给的资产化与融资属性抬升。
运营商	运营商 AI 战略由“降本”与“平台化”进一步走向“资产化”：本期 Verizon 与 Corning 签署至 2032 年、超过 8000 万英里高密度光纤供应协议，将家庭宽带与 AI 数据中心长距骨干纳入同一融合架构，并直接对接 AWS 等 AI 超大规模客户；Deutsche Telekom 在 Digital X 2026 宣布拟竞标欧盟 AI 超级工厂计划并扩建慕尼黑工业 AI 云。运营商正凭借光纤、数据中心与能源资产切入 AI 基础设施供给。后续重点跟踪长距光纤走廊建设、超大规模客户签约及欧洲主权 AI 算力的落地节奏。
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、InP 晶圆等仍是产业关注重点。本期 Verizon 与 Corning 的 8000 万英里高密度光纤长单（含 Contour Flow Cable）显示干线光缆供需格局持续趋紧；双周内光通信板块（+2.68%）为七大板块中涨幅居前的硬件板块，AXT（+13.61%）、Intel（+13.36%）、Coherent（+12.59%）、Lumentum（+5.63%）领涨。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。

下载日志已记录，仅供内部参考

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 于 9 月 19 日完成 2026 年第 17 次 Electron 发射，为 Synspective 部署第 12 颗 StriX 卫星，Synspective 计划于 2030 年前完成约 30 颗卫星的星座组网，目前仍有 15 次专用发射在列；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	AI 应用由“模型能力竞争”转向“任务执行与数据闭环”：OpenAI 在 ChatGPT Work 上线 Data agent，打通企业数据源与 BI 工具并生成交互式仪表盘；DeepSeek 发布 V4.1 Flash（552B MoE，KV Cache 对 HBM 需求降至上一代的 1/4），模型层降本与效率竞争持续加速。本期双周内行业应用板块（-2.79%）随高估值软件整体回调，HubSpot（-12.71%）、UiPath（-11.85%）、Salesforce（-8.22%）回撤居前，与基本面兑现节奏形成背离。企业侧关注重点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。
算力服务	AIDC 板块核心变量由“容量扩张”转向“定价与资本结构”：本期 CoreWeave 三季度以约 4000 万美元/兆瓦（年化）签约 3 至 6 个月短期限合约，签约电力升至约 4.2 吉瓦，并同步启动 37 亿美元可转债发行与最多 3500 万股 ATM 增发计划；Nebius 三个月内第二次提价，Nvidia GPU 按需价格上调 17%-21%。Oracle 剩余履约义务升至 6640 亿美元、单季新增 AI 云合同超 300 亿美元。后续重点跟踪算力定价的可持续性、短约占比对收入能见度的影响，以及电力并网等资源约束的缓解节奏。
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 AI 数据中心标配；800VDC 高压直流供电与“电链+冷链”一体化方案仍是设计方向。本期液冷板块（-4.27%）随高估值硬件与利率敏感资产回调，Vertiv（-11.10%）跌幅居前。后续跟踪液冷渗透率提升、产能区域化布局及电力侧并购整合进展。

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
9 月 30 日（周三）	财报	Micron 公布 2026 财年第四季度财报
10 月 1 日	产业事件	Nebius 新一轮按需定价生效（Nvidia GPU 上调 17%-21%）
10 月 21 日（周三）	产业事件	NVIDIA GTC Berlin 2026，黄仁勋主题演讲
10 月下旬（预计）	宏观	美联储 FOMC 利率决议
持续跟踪	产业事件	GPT-6 Astra 向订阅用户全面开放节奏及推理需求放量；NVIDIA 5000 亿美元 AI 算力融资平台落地进展
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

五、投资建议

美联储于 9 月 16 日加息 25 个基点至 3.75%-4.00%，为 2023 年 7 月以来首次加息。我们认为，分母端：无风险利率中枢系统性上移，且抬升的不只是政策利率。长端 10 年期美债 9 月中旬盘中触及 5.02%，创 2007 年以来新高，其驱动已不单是政策预期，而是油价高企、美债规模的供给压力、AI 基础设施企业大举发债与国债争夺资金所共同推高的期限溢价，意味着即便美联储停止加

下载日志已记录，仅供内部参考

息，长端利率也难快速回落。对权益资产而言，贴现率抬升首先压缩的是市盈率，尤其是依赖远期现金流的高估值成长股。**分子端：**AI 叙事并未降温，产业景气仍高涨，只是定价逻辑从"远期愿景"切换到"当期兑现"。三季度以来海外利率预期反复与板块拥挤度上升，已推动全球主要科技指数完成一轮较为充分的估值消化，为修复留出空间。加息落地后首个交易日，存储、CPU、代工、光互联等算力产业链环节普遍占优：当期盈利强劲、投资回报周期短，使其对分母端利率上行具备明显钝感。**因此，分化并未终结，定价框架正在切换。有当期现金流与业绩验证的算力硬件（先进制程、存储、光互联、设备材料）受益于利空出尽；依赖远期叙事、缺乏盈利支撑的标的则被持续挤出，AI 行业步入理性验证阶段。**

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。

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） 11
图 10： 本周各模型 Token 市场份额（%） 11
图 11： 最新热门 GPU 起步价情况 12

表 1： 本周热门模型 Token 调用量排名 11
表 2： 子行业技术趋势 12
表 3： 行业大事前瞻 13

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