



通信行业研究

买入（维持评级）

行业周报
证券研究报告

通信组

分析师：张真桢（执业 S1130524060002）

zhangzhenzhen@gjzq.com.cn

韩国 800 万亿加码 AI 基建，国产链迎业绩兑现期

通信周观点：

1) 谷歌已对 Meta 使用其 Gemini 模型设置算力限制，Meta 转向自研 Muse Spark 模型以减少外部依赖。核心 AI 能力的外部供应风险正倒逼大厂加速自研。2) 韩国发布以半导体、物理 AI 与 AI 数据中心为三大支柱的产业计划，将在西南部投资 800 万亿韩元建设四座芯片工厂，目标五年内 DRAM 产能翻倍，规划至 2035 年 AI 数据中心总投入超 1000 万亿韩元。国家级战略加码，AI 基建长期确定性增强。3) Meta 拟向外部客户出售 AI 算力和模型访问权限，与云厂商直接竞争。我们认为算力外租本质是资产变现而非供给过剩，头部集中不改变 Capex 扩张逻辑。4) 美国商务部解除 Anthropic 旗下 Fable 5 和 Mythos 5 出口管制，模型合规化后，海外商用部署有望加速，推理算力需求向上弹性大。5) 6 月非超大规模云市场 GPU 租赁价格全面上涨，A100 环比+6.3%，H100 环比+3.7%，B200 环比+2.7%。算力供需持续偏紧，租赁价格有望维持高位。6) AXT 与 Coherent 签署三年期 InP 衬底供应协议，Coherent 预付约 2229 万美元。长协锁定上游核心材料，供应链可见度提升。7) 腾讯云宣布 DeepSeek-V4 正式版 7 月中旬上线，引入峰谷定价，高峰时段价格为平时 2 倍。国产大模型商业化加速，峰谷定价机制或成新趋势。8) 美团发布万亿参数大模型 LongCat-2.0，为业界首个在五万卡国产算力集群完成全流程训练与推理的模型。国产算力集群加速建设，产业链持续受益。9) 华为计划 Q4 向韩国推出昇腾 AI 芯片，售价仅英伟达四分之一，其中 950PR 推理性能达 H20 的 2.87 倍。国产芯片出海加速，性价比优势有望打开海外市场突破口。10) 星网锐捷预计上半年归母净利润为 3.1 亿元至 4.3 亿元，同比增长 46.27%至 102.90%，数据中心交换机业务显著放量。AI 网络设备需求高景气，交换机产业链充分受益。11) 杭电股份预计上半年归母净利润为 3.6 亿元至 4 亿元，较上年同期增长 852.03%至 957.82%，受益光纤光缆市场回暖量价齐升。业绩超预期，AI 数据中心拉动光纤需求增长。

细分赛道：

服务器：本周服务器指数+2.89%，本月以来，服务器指数-3.45%。韩国发布 800 万亿韩元半导体及 AI 基建投资计划，三星电子与 SK 海力士将新建四座晶圆厂。GPU 租赁价格延续全面涨势，A100、H100、B200 均价连续多月上行。算力需求持续旺盛，服务器产业链有望持续受益。

光模块：本周光模块（CPO）指数-6.10%，本月以来，光模块（CPO）指数-9.58%。Coherent 与 AXT 签署三年 6 英寸 InP 衬底供应协议，预付约 2,229 万美元，高速光芯片核心材料供应链价值凸显，利好光模块上游。

IDC：本周 IDC 指数-1.00%，本月以来，IDC 指数-5.33%。美团开源国产首个万卡级万亿参数大模型 LongCat-2.0，华为昇腾芯片计划四季度进军韩国市场。国产 AI 芯片出海进程加速，国产算力链及 IDC 板块增长可期。

光纤：本周光纤指数-1.00%，本月以来，光纤指数-5.33%。杭电股份上半年业绩超预期，受益光纤光缆市场回暖量价齐升。AI 数据中心集群规模扩张拉动光纤互联需求，光纤光缆产业链成长动能充足。

核心数据更新：

电信业务量收增速逐步提升。2026 年 1-5 月电信业务收入累计完成 7355 亿元，同比下降 1.9%。按照上年不变价计算的电信业务总量同比增长 7.9%。2026 年 5 月我国光模块出口数据同比+50.67%；1-5 月累计同比+18.6%。

投资建议与估值

建议关注国内 2950 亿美元 AI 基建计划带动的 IDC、服务器、国产 AI 芯片产业链，以及海外 AI Capex 持续超预期带动的光模块（特别是 NPO/1.6T）、服务器板块，同时关注 CPO 产业进度对光模块估值体系的重塑。

风险提示

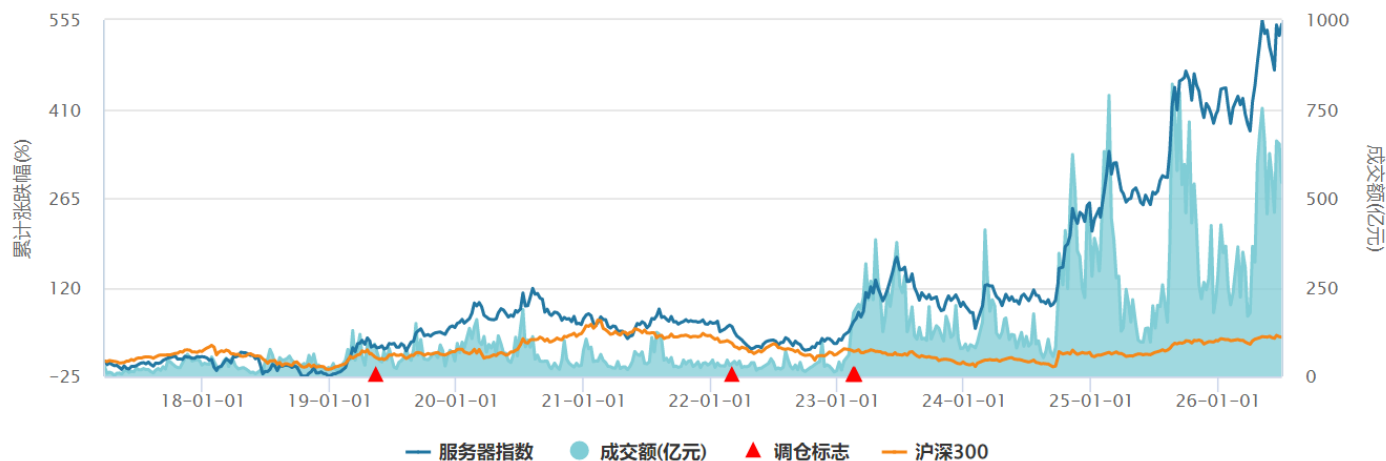
AI 建设不及预期、中美关税波动加剧、原材料供应不足。



一、细分行业观点

服务器：本周服务器指数+2.89%，本月以来，服务器指数-3.45%。韩国发布 800 万亿韩元半导体及 AI 基建投资计划，三星电子与 SK 海力士将新建四座晶圆厂。GPU 租赁价格延续全面涨势，A100、H100、B200 均价连续多月上行。算力需求持续旺盛，服务器产业链有望持续受益。

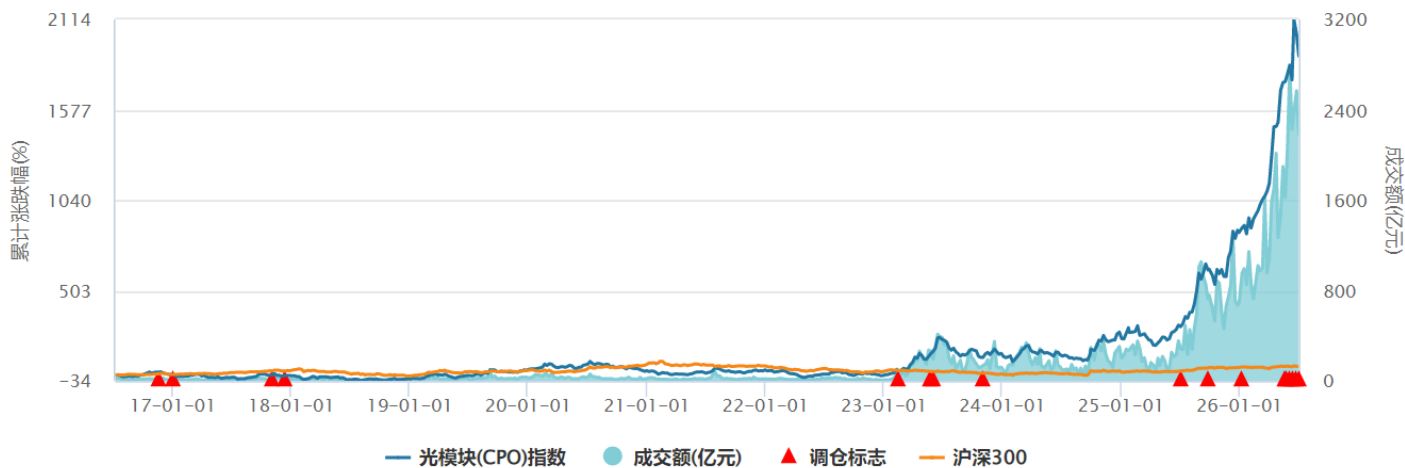
图表1：服务器指数 (8841058.WI) 走势



来源：wind，国金证券研究所

光模块：本周光模块（CPO）指数-6.10%，本月以来，光模块（CPO）指数-9.58%。Coherent 与 AXT 签署三年 6 英寸 InP 衬底供应协议，预付约 2,229 万美元，高速光芯片核心材料供应链价值凸显，利好光模块上游。

图表2：光模块(CPO)指数 (8841258.WI) 走势

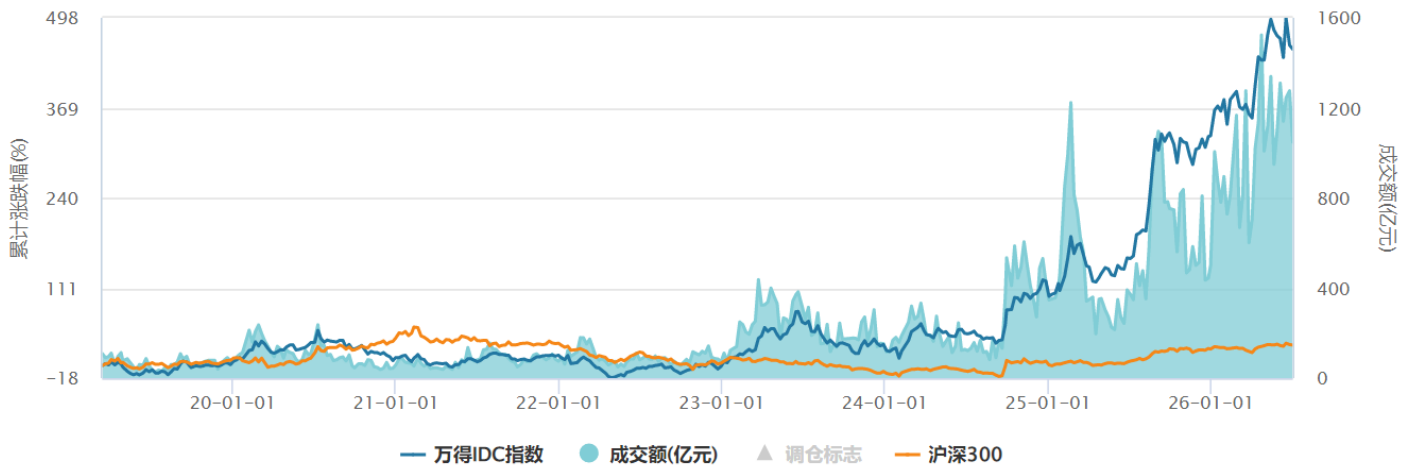


来源：wind，国金证券研究所

IDC：本周 IDC 指数-1.00%，本月以来，IDC 指数-5.33%。美国开源国产首个万卡级万亿参数大模型 LongCat-2.0，华为昇腾芯片计划四季度进军韩国市场。国产 AI 芯片出海进程加速，国产算力链及 IDC 板块增长可期。



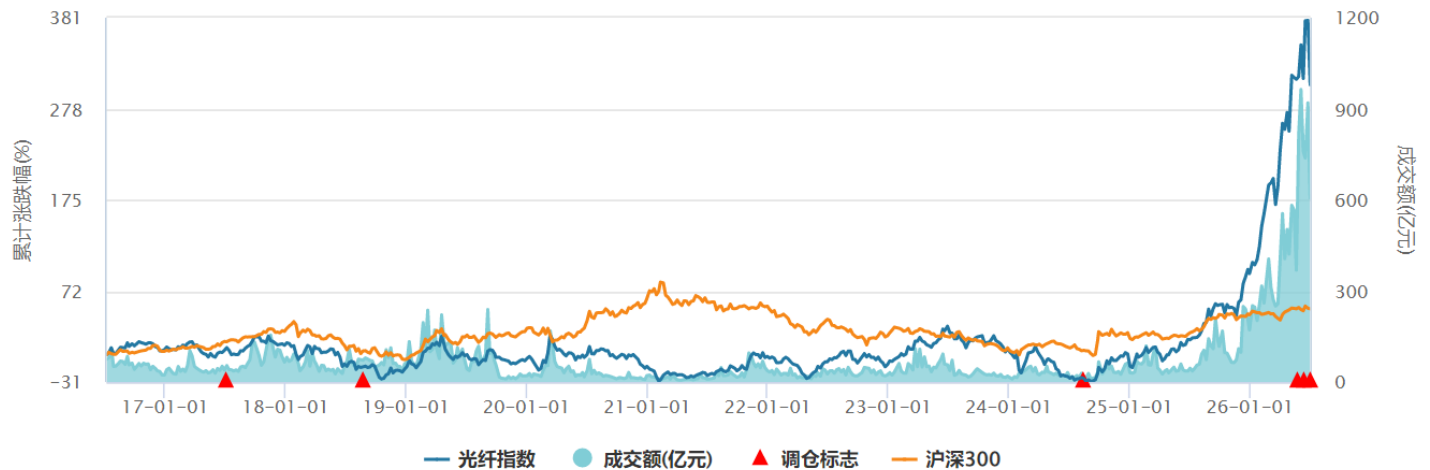
图表3: IDC指数 (866052.WI) 走势



来源: wind, 国金证券研究所

光纤: 本周光纤指数-1.00%, 本月以来, 光纤指数-5.33%。杭电股份上半年业绩超预期, 受益光纤光缆市场回暖量价齐升。AI 数据中心集群规模扩张拉动光纤互联需求, 光纤光缆产业链成长动能充足。

图表4: 光纤指数 (884944.WI) 走势



来源: wind, 国金证券研究所

图表5: 本周通信板块景气度

板块	景气度指标	本期景气度说明
运营商	稳健向上	5G 投资周期结束, 电信行业端承压, 但云与 IDC 业务放量接力成长, 整体景气度稳健向上。
光模块	稳健向上	AXT 与 Coherent 签署三年 InP 晶圆供应协议, 获 2229 万美元预付款支持, 光通信下游需求依然旺盛。
服务器	稳健向上	GPU 租赁价格延续全面涨势, A100、H100、B200 均价连续多月上行, 服务器板块稳健向上。
交换机	稳健向上	星网锐捷发布 2Q 业绩预告, 利润环比高增, 交换机板块稳健向上。
连接器	稳健向上	数据中心正驱动 G. 657. A1/A2 等高规格、高附加值光纤产品需求快速增长, 近期日本光纤龙头藤仓 (Fujikura) 谋划涨价。AI 驱动光纤、连接器等需求增长。
IDC	加速向上	豆包全系列产品当前日均 Token 消耗量突破 180 万亿, 较 26 年初 50 万亿日均规模环比增长 260%, 较 26 年 3 月末 120 万亿环增 50%。国内 Tokens 及推理需求爆发增长, IDC 建设需求依旧旺盛。
物联网	加速向上	苹果加码推进端侧 AI, 依托自研芯片在 iPhone、Mac 等终端设备本地运行 AI 模型, 同时正考虑收购端侧 AI 初创公司 Liquid AI。
液冷	高景气维持	谷歌正与中国英维克等企业洽谈采购数据中心液冷系统, 液冷板块再受催化。



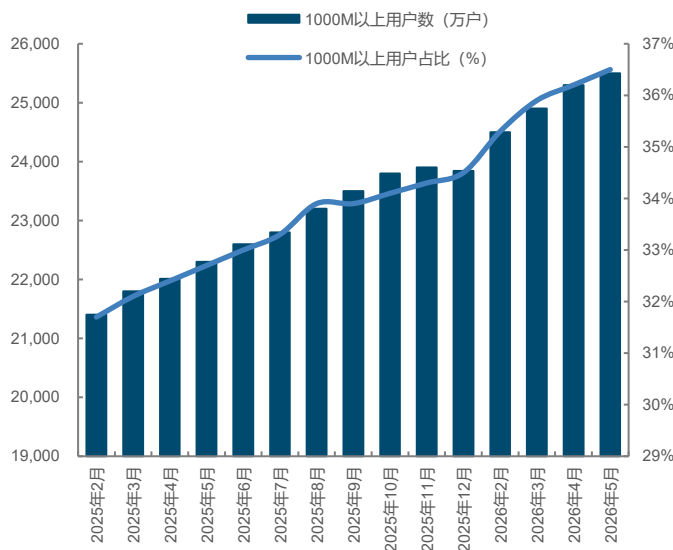
来源: TradingView, 财联社, 新浪财经, 国金证券研究所

二、核心数据更新

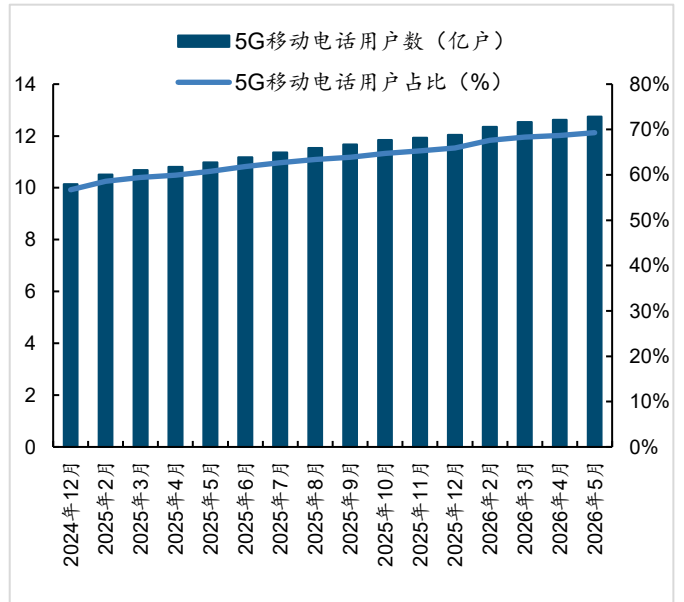
运营商: 运营商数据维持稳健增长

截至5月末, 三家基础电信企业及中国广电的移动电话用户总数达 18.41 亿户, 比上年末净增 1408 万户。其中, 5G 移动电话用户达 12.75 亿户, 比上年末净增 7077 万户, 占移动电话用户的 69.3%。

图表6: 千兆用户占比超三成



图表7: 截至5月末5G用户占比超六成

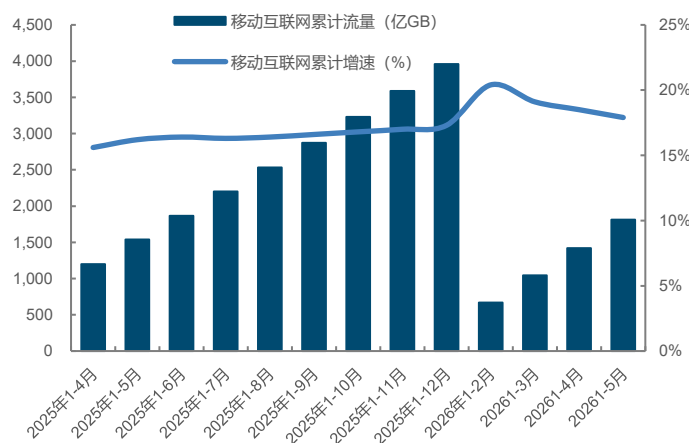


来源: 工信部, 国金证券研究所

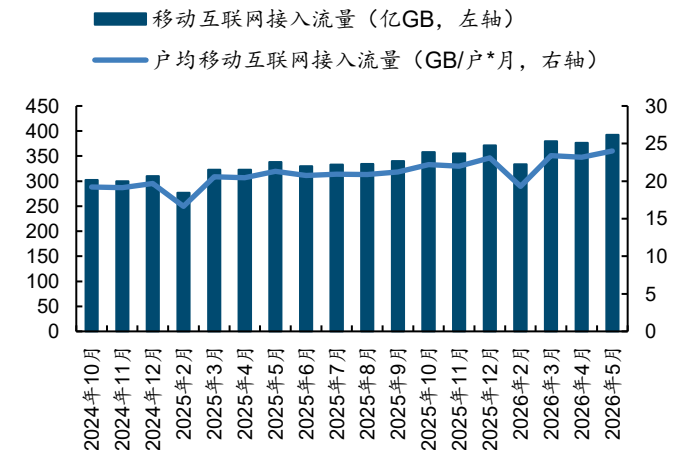
来源: 工信部, 国金证券研究所

前5个月, 移动互联网累计流量达 1812 亿 GB, 同比增长 17.9%。截至5月末, 移动互联网用户数达 16.34 亿户, 比上年末净增 2412 万户。5月当月户均移动互联网接入流量 (DOU) 达到 24.01GB/户·月, 同比增长 12.7%, 比上年底高 0.97GB/户·月。

图表8: 前5个月移动互联网累计流量同比增长 17.9%



图表9: 5月当月DOU达 24.01GB/户·月



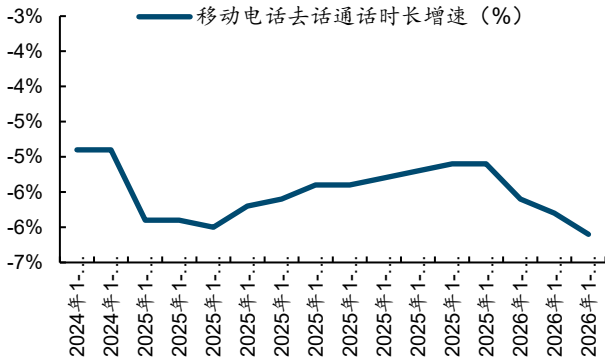
来源: 工信部, 国金证券研究所

来源: 工信部, 国金证券研究所

前5个月, 移动电话去话通话时长完成 7897 亿分钟, 同比下降 6.1%; 固定电话主叫通话时长完成 230.5 亿分钟, 同比下降 21.2%。前5个月, 全国移动短信业务量同比下降 7.4%; 移动短信业务收入同比下降 12.6%。

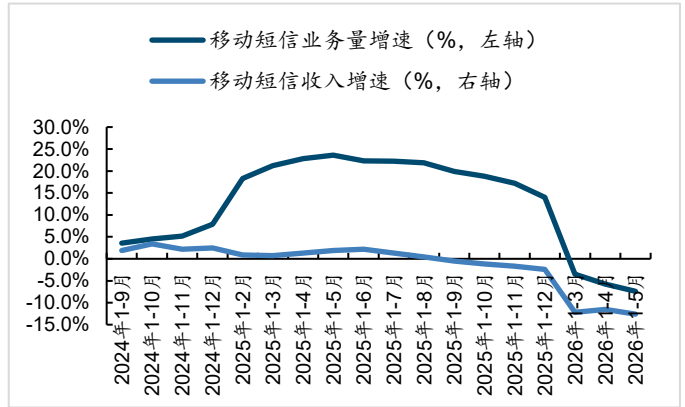


图表10: 电话通话量持续下滑



来源: 工信部, 国金证券研究所

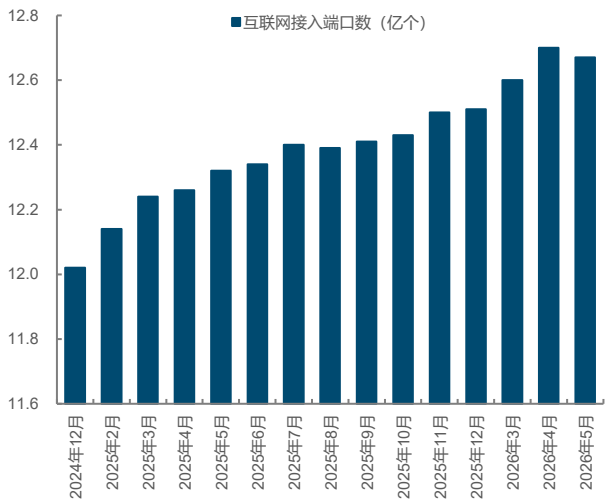
图表11: 移动短信业务量大幅下降



来源: 工信部, 国金证券研究所

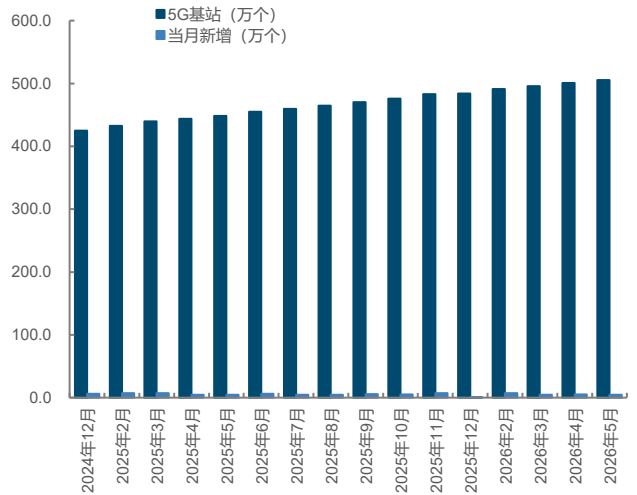
截至5月末, 全国互联网宽带接入端口数量达 12.67 亿个, 比上年末净增 1581 万个。其中, 光纤接入 (FTTH/O) 端口达到 12.24 亿个, 比上年末净增 1386 万个, 占互联网宽带接入端口的 96.7%。截至5月末, 具备千兆网络服务能力的 10G PON 端口数达 3257 万个, 比上年末净增 94.4 万个。

图表12: 千兆光纤宽带网络建设稳步推进



来源: 工信部, 国金证券研究所

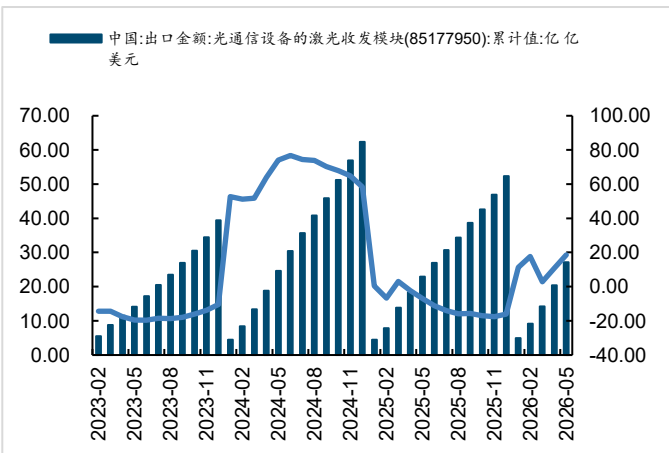
图表13: 5G 网络建设持续深化



来源: 工信部, 国金证券研究所

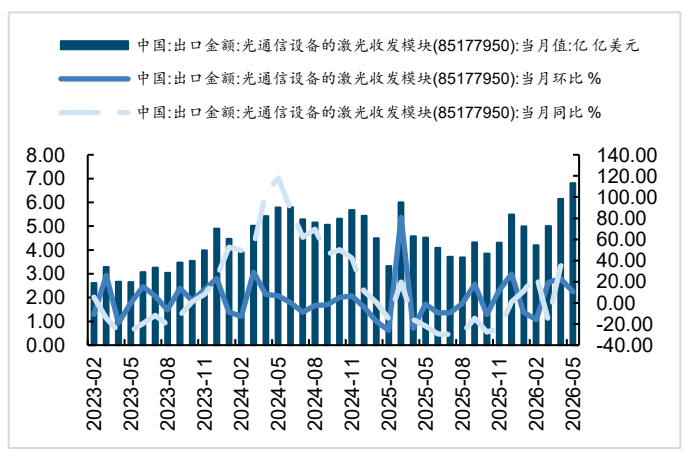
光模块数据: 2026 年 5 月我国光模块出口数据同比+50.67%; 1-5 月累计同比+18.6%。

图表14: 5月光模块出口金额累计同比增加 18.6%



来源: wind, 国金证券研究所

图表15: 5月光模块出口金额当月同比+50.67%

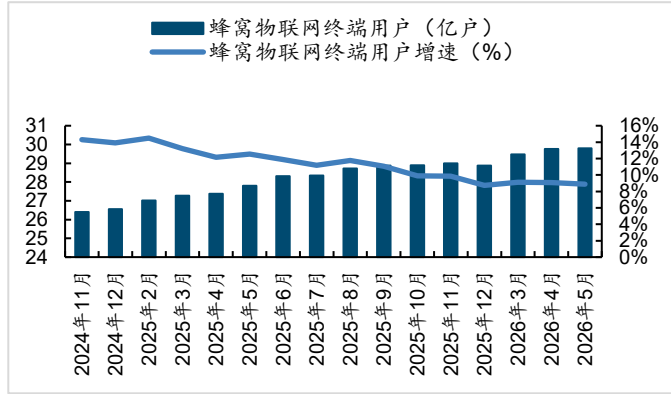


来源: wind, 国金证券研究所



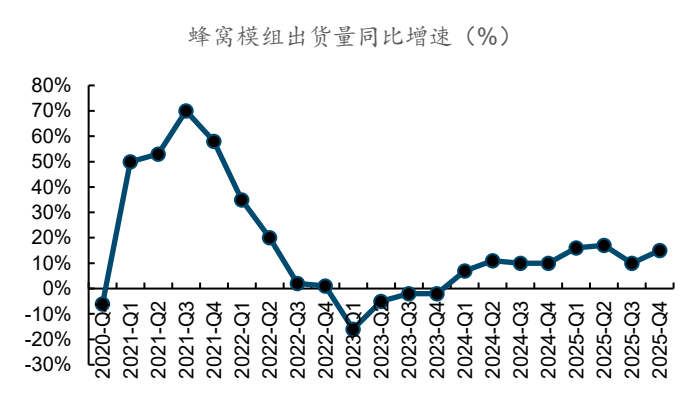
物联网数据：截至5月末，三家基础电信企业发展移动物联网终端用户 29.8 亿户，比上年末净增 9167 万户。互联网电视（IPTV、OTT）用户数达 4.09 亿户，比上年末净增 124.3 万户。

图表16：截至5月末蜂窝物联网终端用户数同比增长7.19%



来源：工信部，国金证券研究所

图表17：2025年全球蜂窝物联网模组出货量同比增长15%

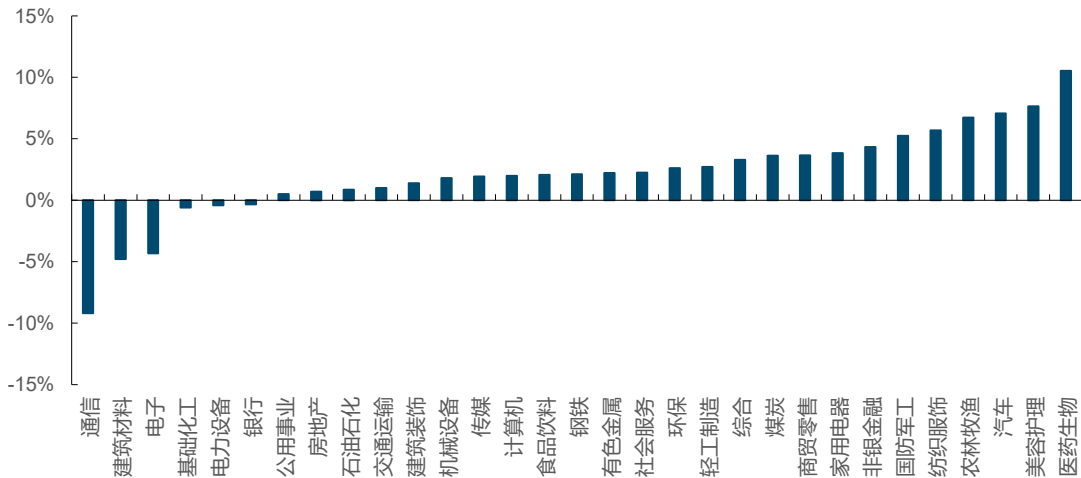


来源：国金证券数字未来实验室，RFID 世界网，C114 通信网，国金证券研究所

三、本周行情

回顾本周行情（6月29日-7月3日），参考申万一级行业划分，通信板块涨跌幅为-9.17%，排名全行业第31。

图表18：板块周涨跌幅排序 (%)

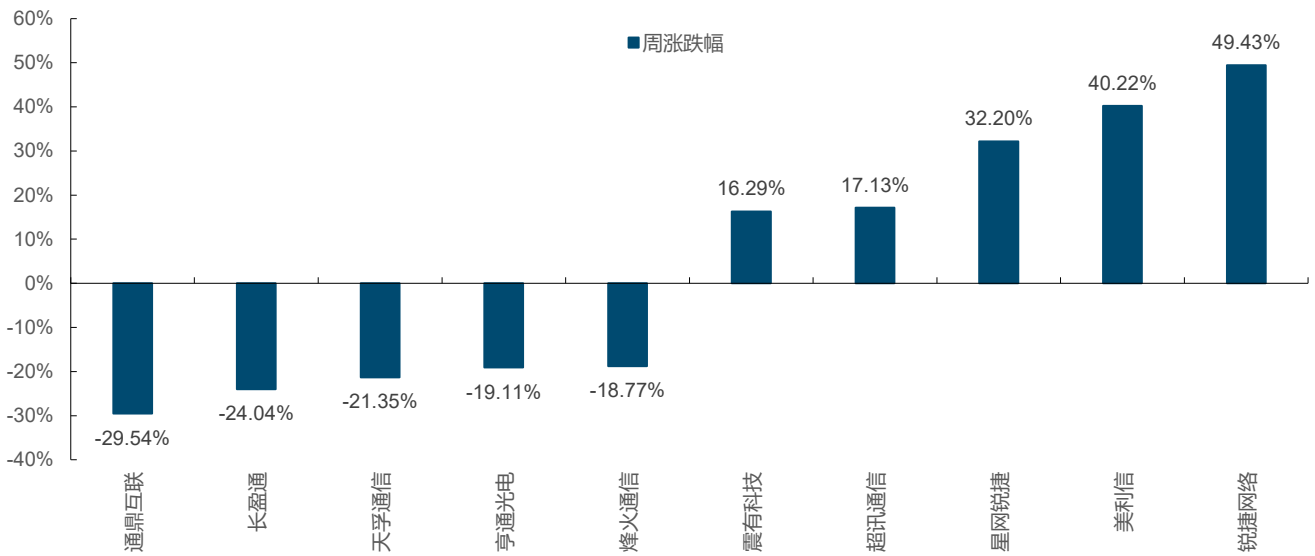


来源：wind，国金证券研究所

从个股情况来看，本周锐捷网络、美利信、星网锐捷、超讯通信、震有科技为通信（申万）涨幅前五大公司，涨跌幅分别为+49.43%、+40.22%、+32.20%、+17.13%、+16.29%。通信（申万）跌幅前五为通鼎互联、长盈通、天孚通信、亨通光电、烽火通信，涨跌幅分别为-29.54%、-24.04%、-21.35%、-19.11%、-18.77%。



图表19：通信板块个股周涨跌幅（剔除*ST公司）



来源：wind，国金证券研究所

四、本周重要新闻

4.1 行业新闻

韩国产业投资计划：

6月29日华尔街见闻讯，韩国政府于本周一宣布了迄今规模最大的半导体与人工智能产业投资计划，将半导体、物理AI和AI数据中心定位为产业升级的“三角支柱”，并明确提出要在五年内使DRAM产能实现翻番，从而抢占AI革命的主导地位。在具体部署上，西南部地区将承接约800万亿韩元的企业投资，用于兴建四座存储芯片晶圆厂，其中三星与SK海力士各负责两座；同时，面向AI数据中心的资金投入预计到2035年累计超过1000万亿韩元，忠清地区则单独获得81万亿韩元以构建先进封装产业集群。此外，政府规划在15年内投入30万亿韩元，覆盖从研发、设计到制造和验证的完整半导体产业链，并将东南及大庆地区定位为材料、零部件与设备供应枢纽。企业层面，SK海力士明确承诺出资400万亿韩元建设新芯片基地，三星方面也透露将在光州和忠清分别布局芯片生产与HBM专用工厂。市场对此反应积极，韩国综合股价指数早盘曾深跌逾3%，随后反转收涨约0.5%，SK海力士股价也从近6%的跌幅中回升，创业板指数更是一度冲高超过8%。李在明在发布会上表示，此次大规模投资并非行政强制的结果，而是政企双方利益高度一致的自发合作，未来将在青瓦台设立专门的项目主管，由其直接督导这三大超级项目的推进与落实。

GPU租赁价格：

7月3日财联社讯，摩根大通援引彭博数据指出，6月非超大规模云厂商市场的GPU租赁价格延续全面涨势。A100平均租赁价格环比上涨6.3%，达1.63美元/GPU小时，已连续第五个月上行；H100均价环比上涨3.7%，至2.72美元/GPU小时，连续第七个月攀升；B200均价环比上涨2.7%，升至5.33美元/GPU小时。三类GPU同步涨价，显示AI算力需求不仅集中于头部云厂商，初创及企业客户仍在激烈争夺资源。与此同时，高端GPU间的价格溢价持续收窄——H100对A100的价格比从4月的1.77倍降至6月的1.67倍，B200对H100的价格比从2.04倍降至1.96倍，较2025年9月指数推出时的2.58倍明显回落。这反映出市场正从“盲目追逐最新产品”转向更精细的性能与成本优化，但绝对价格仍处高位，算力供需紧张局面尚未缓解。

手机厂商出货目标：

6月30日新浪财经讯，据《日经亚洲》报道，由于零部件短缺与成本上涨，中国主流智能手机厂商再次下修2026年出货目标。小米下调约30%至9500万部，OPPO和vivo降至9000万部以下，荣耀表示难以维持2025年7100万部的水平。供应链人士称，多数厂商出货目标较年初预测下调15%-20%，部分达30%。核心原因是DRAM、PCB等关键零部件全面涨价短缺，有厂商坦言无法确保新品零部件供应。AI基础设施扩张正挤占手机产能，例如英伟达Vera Rubin NVL72单柜使用54TB LPDDR5X，相当于数千部旗舰机用量，降低手机优先级；联发科和高通也转向数据中心业务，联发科已因成本上调产品价格。Counterpoint于6月1日将2026年全球出货预测下调至同比下降13.9%，约10.8亿部。

晶圆代工：

据TrendForce集邦咨询6月30日调查，AI服务器、通用服务器及边缘AI需求升温，晶圆代工产能向AI倾斜。8英寸制程受益于AI Power订单及台积电、三星减产，产能利用率强势回升，2026年前十大业者8英寸平均利用率升至



88%，下半年达 90%，代工价格一季度至二季度全面涨 5%-15%，业者酝酿第三波涨价。12 英寸成熟制程方面，短期 AI Power 带动 55nm 以上需求，65/55nm 中介层、40/28nm FPGA 等需求增温；中长期台积电成熟制程减产、力积电转单效应，叠加硅桥/中介层、DTC、IPD、PIC、NAND CMOS 等新需求开案，产能能见度延伸至 2027 年。晶圆厂扩产但优先利润较佳产品，低毛利产线受压缩，原物料涨价加剧成本压力，部分制程已在二、三季度调涨 5%-10%，并计划 2027 年全面涨价。尽管消费电子客户协商暂缓，但原物料通膨、减产及 AI 排挤产能，2027 年涨价难以避免。

ECTC 2026 综述：

SemiAnalysis 于 7 月 3 日发布 ECTC 2026 (IEEE 电子元件与技术会议) 年度综述，聚焦先进封装技术如何突破 AI 加速器在尺寸、互连、散热和功耗方面的瓶颈。随着晶体管密度提升放缓，先进封装已成为主要扩展方向，但 AI 加速器规模持续扩大，封装本身正面临多重限制。

英特尔发布下一代 EMIB-T (带硅通孔)，凸点间距从 45 μm 缩小至 36/35 μm ，密度提升 65%，验证范围正扩展至 4.5 倍掩模板封装，预计 2026 年底完成认证，该技术预计用于谷歌 TPU v9。EMIB-T 通过 TSV 实现垂直供电，直流压降降低 68-80%；桥接内 MIM 电容密度达 500nF/mm²，PDN 交流阻抗降低 82% 以上。英特尔已完成 240mm \times 240mm 四分之一面板测试，但展台样品存在严重翘曲。此外，英特尔还展示了业界首款 510mm \times 515mm 24 层玻璃核心面板，内嵌 EMIB 桥及光波导，热冲击后无裂纹，但工艺尚不成熟。

三星提出 8 层硅中介层方案用于 HBM4E，较预估需求减少 20% 层数，采用双信号/一地交错布线屏蔽高速信号。在热管理方面，三星比较了 TCB 与混合铜键合 (HCB)，HCB 使内部热阻降低约 12%，4 倍焊盘密度下堆叠级热阻降低 29%。随着 HBM 向 20 层、24 层演进及定制 HBM 将更多逻辑移入基底芯片，散热瓶颈正从 GPU-HBM 串扰转向堆叠自身热阻。三星还发布了垂直铜柱堆叠 DRAM 技术，完全避开 TSV，功耗降低 41%，速率从 8.6Gb/s 提升至 11.8Gb/s。

Marvell 展示定制 HBM4E 方案，通过将接口逻辑从加速器移至 HBM 基底芯片，缩短中介层通道长度 (从 6.5mm 降至 1.5mm)，主机 ASIC 面积减少约 60%，带宽达 4.1TB/s，英伟达 Feynman 和 AMD MI450/MI500 预计采用类似方案。在光学互连方面，Marvell 发布 OMIB (光学多芯片互连桥)，将 PIC 嵌入有机 RDL 中介层，支持 1.8Tbps/mm² 带宽密度，采用电吸收调制器 (EAM) 替代微环调制器，宣称更优热稳定性，但规模化制造仍是挑战。热特性对比显示，基板方案 PIC 温升不到 5 $^{\circ}\text{C}$ ，而硅中介层方案高达 25 $^{\circ}\text{C}$ 。

台积电展示在 CoWoS-R 平台上集成微柱直触硅液体冷却方案，在 8LPM 流速下可散热 5.3kW，较传统方案大幅提升。与 GUC 合作展示基于 8 层 CoWoS-R RDL 的 UCle-A x64 芯粒集成，在 32GT/s 下实现 0.77UI 眼宽，证明有机中介层可满足异构集成信号完整性要求。

微软在 GH200 GPU 上验证微通道直触硅冷却，报告 GPU 结到入口热阻降低 51-60%，封装总热阻降低 50%。6 个月内在约 4370 次观测中仅记录 9 次潜在堵塞事件，堵塞率随时间下降，无硅侵蚀现象；节点层面完成 3 周重复基准测试及 1 周连续稳定运行，集群级 MTBF 和可用性仍在测试中。

4.2 公司新闻

Anthropic：

36 氪 7 月 1 日讯，Anthropic 正式发布 Claude Sonnet 5，定位为“最具 Agent 属性的 Sonnet 模型”。该模型在推理、编程、工具使用及知识工作等方面较前代 Sonnet 4.6 均有提升，性能已接近 Opus 4.8，但成本更低。安全表现亦有改善，幻觉率与谄媚行为率均低于前代。定价方面，即日起至 8 月 31 日推广期内为输入 \$2/百万 token、输出 \$10/百万 token，之后恢复至输入 \$3、输出 \$15 的标准定价。不过，Sonnet 5 采用了全新 tokenizer，相同内容映射的 token 数增加约 1.0-1.35 倍。即日起已面向所有渠道正式可用。

7 月 1 日华尔街见闻讯，Anthropic 的企业协作工具 Claude Tag 在接入 Slack 后进一步集成至 Microsoft Teams，完成企业协作双平台布局。此举背后是 Anthropic 商业化急剧加速的防御性策略——微软和 Salesforce 若拒绝接入，企业客户可能为使用 Claude Tag 而迁移至其他平台，但这也意味着 Claude Tag 将在 Teams 中与微软 Copilot 正面竞争，可能分流后者订阅收入。分析认为，传统企业软件巨头在 AI 时代的护城河正被削弱，平台“管道化”风险上升，AI 独角兽在应用层的变现潜力需重新评估。

7 月 1 日新浪财经讯，美国商务部于 6 月 30 日正式解除对 Anthropic 旗下 Fable 5 和 Mythos 5 两款模型的出口管制。此前，Fable 5 因美国政府担忧其可能被“越狱”用于网络攻击，发布仅三天即遭全量停服。此次解禁的关键在于 Anthropic 调整策略，承诺构建更强的安全防护措施。然而，此举引发 OpenAI 等竞争对手呼吁建立规范化流程，以遏制政府在 AI 监管上临时性、不可预期的做法。

7 月 2 日 The Information 讯，据三位知情人士透露，Anthropic 已正式启动自研 AI 芯片项目，目前处于早期规划阶段，尚未进入详细设计及生产环节。公司正与三星洽谈采用其 2 纳米工艺及先进封装技术的潜在制造合作，并已聘请 OpenAI 自研芯片团队早期成员 Clive Chan 加入。Anthropic 方面表示，AWS Trainium、Google TPU 及英伟达 GPU 仍将是其未来算力扩张的核心依赖。此次合作若达成，将助力三星代工业务在先进 AI 芯片领域争取更多订单，缩小与台积电的差距。值得一提的是，三星及 SK 海力士、美光科技已于今年 5 月共同参与了 Anthropic 650 亿美元融资，反映出 Anthropic 与关键芯片供应商之间的深度绑定关系。

美团：

6 月 30 日，美团正式发布并开源新一代万亿参数大模型 LongCat-2.0。该模型总参数达 1.6T (平均激活约 48B，动态



范围 33B 至 56B)，是业界首个在五万卡国产算力集群上完成全流程训练与推理的万亿参数模型。模型从零开始预训练，预训练数据规模超过 30 万亿 tokens，覆盖中英等多语言及代码数据；原生支持 100 万 token 超长上下文，采用 LSA 稀疏注意力机制实现线性级计算复杂度。面对万卡级训练中的硬件故障、通信异常等难题，团队通过自研确定性算子、Bitwise 一致性验证及自动故障恢复等手段，将月均日故障率降低 70% 以上，训练 MFU 提升 1.5 倍，稳态日吞吐超过 1 万亿 tokens。在推理阶段，通过大规模专家并行聚合与零计算专家机制，实现低延迟解码。目前 LongCat-2.0 预览版已通过 OpenRouter 平台和 longcat.ai 向全球开发者开放，月调用量跻身 OpenRouter 全球前三。

昆仑芯：

6 月 28 日 The Information 报道，百度旗下 AI 芯片公司昆仑芯计划赴港 IPO，目标估值 500 亿美元，并要求基石投资者按认购金额 3 至 7 倍购买其芯片，这种捆绑模式在中国科技企业上市中罕见，凸显国内 AI 芯片竞争激烈。昆仑芯拥有百度这一内置客户，已用于训练文心大模型，产品主打推理芯片并兼容英伟达 Cuda 系统，便于开发者迁移，目标客户包括央企、大型互联网公司及 AI 模型开发者，腾讯已成重要外部客户。然而，公司未进入中国首批 AI 芯片安全可靠认证目录，不利争取国资买家，且过去依赖三星代工，正与中芯国际洽谈产能转移。若成功上市，昆仑芯将成为 2025 年 12 月以来第六家上市的国产 AI 芯片设计公司，此前曾寻求募资 20 亿美元。

DeepSeek：

7 月 3 日 IT 之家讯，腾讯云发布公告称，DeepSeek-V4 正式版“原厂直供”模型计划于 2026 年 7 月中旬在腾讯云大模型服务平台 TokenHub 和智能体开发平台正式上线。该模型由 DeepSeek 原厂直接提供，本次版本更新将带来多项功能优化与性能提升。伴随正式版发布，腾讯云将同步引入峰谷定价机制。高峰时段定义为每日 9:00 至 12:00 和 14:00 至 18:00 (北京时间)。以 DeepSeek-V4-Pro 为例，推理输出百万 tokens 平时价格为 6 元，高峰时段为 12 元；DeepSeek-V4-Flash 推理输出百万 tokens 平时价格为 2 元，高峰时段为 4 元。腾讯云表示，将在定价生效调整前 24 小时通过邮件、站内信、短信等方式向用户提前告知。

长鑫存储：

6 月 29 日据路透社报道，三位知情人士透露，长鑫存储已与腾讯签署了一份价值超 200 亿元人民币（约合 29.4 亿美元）的长期供货协议，涵盖服务器用 DRAM 芯片数年的供应，协议期限为三至五年。在全球 DRAM 供应偏紧、价格飙升的背景下，企业将长期协议视为优先事项。该交易更多细节尚不明确，包括是否涉及高带宽内存 (HBM) 产品。交易正值长鑫存储科创板 IPO 前夕（已获批、拟募资 295 亿元），来自腾讯的订单是对其重大认可；另据消息人士称，长鑫存储正与其他大型互联网公司洽谈类似合作。行业层面，瑞银数据显示 Q1 DRAM 合约价环比飙升约 95%，上行周期预计持续至 2027 年末。长鑫存储作为全球第四大 DRAM 厂商（Q1 营收 508 亿元、净利润 250 亿元），目前在合肥及北京拥有三座 12 英寸晶圆厂、月产能约 30 万片，正扩建上海新厂及 HBM 封装厂，目标月产能翻倍至 60 万片。不过消息人士透露，其一季度 DDR5 良率较低，技术差距依然存在。

字节跳动：

6 月 29 日澎湃新闻讯，针对近期“豆包大模型正在内测社交功能”的传闻，豆包相关负责人于 6 月 29 日回应称，豆包在企业办公场景与飞书存在协同尝试，未来合作会更紧密，但没有传闻所说的社交功能计划。

星网锐捷：

7 月 3 日新浪财经讯，星网锐捷 (SZ002396) 于 7 月 2 日晚间披露 2026 年上半年业绩预告。公司预计上半年归母净利润为 3.1 亿元至 4.3 亿元，同比增长 46.27% 至 102.90%；扣非净利润为 2.1 亿元至 3.1 亿元，同比增长 11.76% 至 64.98%。结合一季度数据推算，公司二季度归母净利润环比增长 633% 至 955%，扣非净利润环比增幅更是高达 2122% 至 3228%，强势扭转了此前连续三个季度的业绩下滑态势。此次业绩大幅增长的核心驱动力，是子公司面向互联网客户的数据中心交换机业务实现显著放量；与此同时，产业基金公允价值变动及股权转让带来的投资收益也贡献了 1 亿元至 1.2 亿元的非经常性损益。值得关注的是，在业绩预告正式发布前，公司股价已提前连续两个交易日涨停。

杭电股份：

7 月 4 日证券时报讯，杭电股份 (603618) 7 月 3 日股价收报 49 元/股，当日上涨 2.96%，今年以来最大涨幅曾超过 600%；同日，公司披露业绩预告，预计 2026 年上半年实现归母净利润为 3.6 亿元至 4 亿元，较上年同期增长 852.03% 至 957.82%；实现扣非净利润为 3.55 亿元至 3.95 亿元，同比增长 1087.67% 至 1221.49%。公司以电线电缆为基石（是国家电网及南方电网主要供应商），同时将光通信与铜箔打造为战略增长极。其中，光通信板块由子公司富春江光电及其下属企业运营，已形成“光纤预制棒—光纤—光缆”的一体化产业链；报告期内，另一家二级全资子公司杭州永特信息技术有限公司的光纤销量随行业复苏同步增长，亦为公司营业利润做出了积极贡献。此次业绩大幅预增，主要得益于光纤光缆市场回暖带来的量价齐升。

SoftBank：

7 月 2 日软银集团与软银公司联合发布声明，双方将在美国特拉华州成立一家名为“SB Neo”的合资公司，于 2026 年 7 月设立。SB Neo 由软银公司持股 51%、软银集团持股 49%，为软银公司的合并子公司。该公司将依托软银集团正在推进的 10 吉瓦级能源和 AI 基础设施项目，于 2027 财年（截至 2028 年 3 月）起为美国超大规模云服务商等大型企业提供 AI 云服务，满足大规模 AI 模型训练和推理所需的算力资源。为应对美国 AI 算力的强劲需求，SB Neo 将分阶段扩大服务能力，并计划最终部署 10 吉瓦规模的 AI 基础设施。软银公司此前已于 2026 年 5 月起在日本提供基于“Infraia



AI Cloud OS"的 GPU 云服务测试版，将把相关经验应用于美国业务拓展。

AXT:

7月3日 TradingView 报道，AXT 与 Coherent 于 2026 年 6 月 25 日签署了一份为期三年的主开发与供应协议，涉及 6 英寸磷化铟 (InP) 晶圆衬底的大规模开发与供应。根据协议，Coherent 向 AXT 支付 22,288,500 美元预付款，与 AXT 在 2026 年至 2028 年间的产能承诺及北京工厂的扩产计划直接挂钩。若 Coherent 未达到最低采购量，则剩余预付款不予退还，且 AXT 有权终止协议；若 AXT 连续六个月以上未能履行产能承诺，Coherent 可终止协议并要求退还剩余预付款。

4.3 海内外大厂重点跟踪

华为:

据《联合早报》7月3日报道，华为计划今年第四季度向韩国市场推出昇腾系列 AI 芯片，售价仅为英伟达芯片的四分之一。主力产品包括已量产、面向推理场景的昇腾 950PR，以及计划同步面市、专为训练设计的昇腾 950DT。华为还将提供基于灵衢 2.0 互联技术的 Atlas 950 SuperPod 一体化方案，最大可扩展至 8192 颗芯片全互联。华为称 950PR 推理性能可达英伟达 H20 的 2.87 倍。目前华为已选定 SK Shieldus 等两家韩国分销商，但业内预计韩国市场对中国核心技术的敏感情绪仍是主要挑战。

阿里巴巴:

华尔街见闻 7 月 3 日讯，阿里巴巴内部正全面收紧对 Anthropic 的使用边界。由于近期 Claude Code 被曝存在严重的“植入后门”安全风险，阿里经综合评估后已将其列入高风险软件名单。自 7 月 10 日起，阿里将全面禁止内部员工在办公环境下使用 Claude Code，全员须卸载 Anthropic 相关产品，范围涵盖 Sonnet、Opus、Fable 等多个系列模型，并推荐使用内部工具“Qoder”作为替代方案。双方矛盾早有苗头——6 月 24 日 Anthropic 指控相关厂商利用约 2.5 万虚假账号进行“工业级模型蒸馏攻击”，随即收紧风控并对中国用户大规模封号；阿里亦已就美国国防部将其列入“中国军事企业名单”提起诉讼。Anthropic 指控引发业界争议，批评者认为模型蒸馏属合法技术，且 Anthropic 自身训练数据同样源于互联网爬取。

Meta:

6 月 28 日金融时报讯，由于 AI 需求持续挤压算力容量，谷歌已对 Meta 使用其 Gemini AI 模型设置了限制。今年 3 月左右，谷歌告知 Meta 无法提供后者希望购买的全部 Gemini 容量，此举已干扰并延迟了 Meta 部分内部 AI 项目。受此限制及精简 AI 成本推动，Meta 已鼓励员工更高效地使用 AI token。此次限制影响不仅限于 Meta，但 Meta 因其对谷歌模型的极高需求量而受到特别严重冲击。谷歌为应对需求压力，本月初与 SpaceX 签署了每月 9.2 亿美元的算力租赁协议。谷歌 CEO 皮查伊在 4 月财报会上承认公司短期内受到算力约束，称若能满足需求，云收入本可更高。Meta 选择 Gemini 的原因是其表现优于 Meta 自有的 Llama 开源模型。但近期 Meta 已开始优先采用自研的 Muse Spark 模型，该模型被认为在竞争力上与 Gemini 相当，有助于减少部分应用对外部模型的依赖。

7 月 1 日 Bloomberg 讯，Meta 正在制定一项云基础设施业务计划，拟向外部客户出售 AI 算力和模型访问权限，由此与亚马逊 AWS、微软 Azure 和谷歌云等头部云厂商展开直接竞争。据知情人士透露，Meta 考虑两种模式：其一，出售托管在 Meta 基础设施上各类 AI 模型（包括自研 Muse Spark 模型）的访问权限，类似 AWS 的 Bedrock 服务；其二，出售“原始”算力，类似于 CoreWeave 等 Neocloud 厂商的模式。该业务由 Meta 内部项目“Meta Compute”主导，由基础设施主管 Santosh Janardhan、超级智能实验室 AI 部门负责人 Daniel Gross 及 Meta 总裁 Dina Powell McCormick 共同领导。此前，Meta 已承诺投入数千亿美元用于数据中心和 AI 芯片等基础设施，包括与 CoreWeave、谷歌和甲骨文等公司的大型计算交易。Meta CEO 扎克伯格在 5 月股东电话会上表示，出售过剩算力“绝对在考虑之中”，但尚未实施的原因是公司目前自身仍有算力需求，“如果我们意识到建设过度了，这就是一个备选项”。

SemiAnalysis 于 7 月 3 日发布报告，针对彭博社关于“Meta 可能转型为 Neocloud”的报道引发的市场波动进行深度分析。报告认为，市场对 Meta 出售算力的担忧被夸大——Meta 的数据中心与算力采购只会加速而非放缓，2027 年资本支出将高得惊人。仅 2026 年上半年，Meta 已在云服务和托管领域签约超 5GW 容量，且未计入其自建项目。报告指出，Meta 的算力有四大高价值用途：其一，前沿 AI 模型训练 (MSL 并未放弃)；其二，推荐系统扩展 (可扩大 10 倍以上，驱动广告收入加速增长)；其三，销售第三方模型服务 (类似 AWS Bedrock)，报告称 Meta 正与 Anthropic 进行最终谈判，以获取 Claude 私有实例访问权，并计划推出“Token 即服务”端点；其四，达成“SpaceX 式”高溢价算力租赁交易——SpaceX 开创的此类交易每兆瓦营收达同行 3-4 倍，仅 200MW 即可创造年收入 100 亿美元。报告认为，这种高可选性使 Meta 能够持续激进采购算力，即使 MSL 未能成功，仍有多种高利润变现渠道。此外，SemiAnalysis 预计 Meta 很快将宣布一笔百亿美元级别的 Anthropic 交易，以启动增长飞轮。

7 月 3 日路透社讯，Meta CEO 扎克伯格在内部全体会议上承认，AI 智能体技术进展慢于预期，公司对新架构的押注“尚未产生成果”，高管们对变革时机的判断存在失误。为筹集 AI 基础设施投入，Meta 于 5 月裁员约 10% 并将 7000 人调至 AI 团队，引发员工反弹。扎克伯格预计未来三到六个月内 AI 投资将开始产生更显著回报，今年 Meta AI 基础设施支出预计达 1450 亿美元。此外，Meta CTO 表示争议性鼠标追踪软件数据安全事件审查显示无员工数据被用于 AI 训练，该软件若重启将改为“自愿参与”方式。

OpenAI:



7月2日财联社讯，OpenAI 正就一项向美国政府提供 5% 股份的提议展开谈判，旨在获取资金支持并化解日益严峻的政治阻力。OpenAI CEO 山姆·奥尔特曼主张，让公众分享人工智能发展所带来的经济收益，是应对 AI 社会影响的最佳方式。该方案的核心是，美国主要 AI 开发商（包括 Anthropic、谷歌、Meta 等）应将 5% 的股权分配给类似阿拉斯加永久基金的公共投资载体，由政府持有并惠及公众。当前，AI 企业正面临来自公众与政界对数据中心建设、就业及网络安全影响的日益担忧。近期 OpenAI 与 Anthropic 均因美国审查而推迟了尖端模型发布。奥尔特曼已就此与特朗普总统、商务部长卢特尼克、财政部长贝森特及民主党参议员桑德斯等展开积极沟通。不过，相关谈判尚处“概念性”早期阶段，任何协议均需国会批准方可实施。

6月30日 The Information 讯，OpenAI 工程师本月初在一次内部沟通中透露已实现一项推理成本优化技术突破，可将现有模型运行成本压缩逾 50%。具体技术方案尚未公开，可能涵盖量化压缩、批处理调度及模型路由等方向。该技术已在小规模场景落地，用于支持 ChatGPT 免登录访客用户，推理集群一度仅需数百块英伟达 GPU 即可承载，不过 OpenAI 对该类用户设有限额，实际节省总量仍属可控。该优化效果或随年底更大参数模型上线而有所摊薄。OpenAI 同步推进与博通的定制推理芯片研发，以从硬件层面降低对英伟达的依赖。财务方面，OpenAI 一季度末毛利率为 39%，距年底 52% 的目标仍有距离，后三个季度须达 56% 以上。Anthropic 上半年销售增长已证明需求旺盛阶段利润率改善具备快速弹性，但 OpenAI 定价能力尚不如 Anthropic，其边际改善主要来自推理成本的直接压降，此次优化为其毛利率目标提供了更可落地的支撑。

英伟达：

英伟达于当地时间 7 月 1 日宣布推出一项全新的 AI 基础设施合作模式，通过与 AI 云服务商建立收入分成与信用支持机制，共同部署大规模、多租户 AI 工厂。随着 AI 从模型开发走向生产级推理，算力需求正加速转向需要持续运行、大规模生成 Token 的 AI 工厂。新模式下，AI 云服务商可采购英伟达基础设施并向客户销售 GPU 云服务，英伟达则同时获得标准的硬件销售收入和与算力使用量挂钩的持续性云收入分成。该计划已进入落地阶段，首批合作伙伴包括 Sharon AI（将部署最多 4 万块英伟达 Grace Blackwell GB300 GPU）和 Firmus Technologies（将在印度尼西亚巴淡岛建设 DSX AI 工厂园区，规划扩展至 360 兆瓦电力规模，并部署最多 17 万块英伟达 GPU）。英伟达表示，此举旨在帮助初创企业、模型开发者、企业及科研机构更快获得 AI 算力，无需经历数据中心选址、供电建设及硬件部署等漫长过程。Baseten、Fireworks AI 和 Together AI 等 AI 原生公司已展现出对这类规模化 AI 云容量的迫切需求。

风险提示

- 1、AI 商业价值不及预期的风险：目前 AI 市场应用仍处于初级阶段，盈利模式仍需探索，市场尚未成熟。若商业模式无法持续发展新客户，需求大幅减弱，或市场接受度偏低，可能对营业收入造成较大负面影响，损害相关公司的盈利能力及产品或服务的商业价值。
- 2、技术发展速度不及预期的风险：目前 AI 模型的使用仍受限于诸多因素，在特定领域无法达到预期的提高生产力效果。该领域目前仍面临较大的技术挑战，包括模型训练效果不稳定、算法不成熟等问题。若技术落地不及预期，可能影响 AI 的应用领域和运行效率，造成较大的投资损失。
- 3、供应链集中度过高的风险：AI 行业基础设施建设目前高度依赖某几家核心供应商，极易受到相关供应商供应短缺的影响。此外在训练方面，AI 技术依赖于大量优质数据的输入。不可靠、低质量的数据来源会一定程度上影响 AI 模型训练的性能，同时提高训练过程中的不可控成本。
- 4、行业监管加剧的风险：目前生成式 AI 工具仍存在法律、伦理、安全风险。AI 生成内容的产权问题仍存在较大争议。各国可能针对 AI 的使用及 AI 生成内容进行更严格的监管及抵制，影响投资预期，并阻碍 AI 技术在产业上进一步落地。公司面临法律诉讼和声誉受损等负面影响风险。
- 5、市场竞争加剧的风险：在如今巨头科技公司加大 AI 投入，大量创业公司涌入竞争的大环境下，技术的迅速迭代及新算法的涌现可能使得公司技术迅速落后竞争对手，影响相关公司的市场份额和投资回报的稳定性。



行业投资评级的说明：

买入：预期未来 3—6 个月内该行业上涨幅度超过大盘在 15%以上；

增持：预期未来 3—6 个月内该行业上涨幅度超过大盘在 5%—15%；

中性：预期未来 3—6 个月内该行业变动幅度相对大盘在 -5%—5%；

减持：预期未来 3—6 个月内该行业下跌幅度超过大盘在 5%以上。



特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于C3级(含C3级)的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海	北京	深圳
电话：021-80234211	电话：010-85950438	电话：0755-86695353
邮箱：researchsh@gjzq.com.cn	邮箱：researchbj@gjzq.com.cn	邮箱：researchsz@gjzq.com.cn
邮编：201204	邮编：100005	邮编：518000
地址：上海浦东新区芳甸路1088号 紫竹国际大厦5楼	地址：北京市东城区建国内大街26号 新闻大厦8层南侧	地址：深圳市福田区金田路2028号皇岗商务中心 18楼1806



**【小程序】
国金证券研究服务**



**【公众号】
国金证券研究**