



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

FCC 禁令草案扰动光模块，国产大模型跃升商业化加速

通信周观点：

1) 美国 FCC 起草中国光模块进口禁令，拟通过“受保护清单”机制限制中国新型号光收发模块进入美国市场，中国商务部同日将美国合规性测试公司列入反制清单，禁止境内交易合作。禁令尚处草案阶段，中美光通信产业链高度相互依存，中国头部厂商已在东南亚布局产能，强行脱钩将推高北美云厂商成本并延缓 AI 基建，对光模块整体影响有限。2) 阿里巴巴发布 Qwen3.8-Max，总参数 2.4 万亿、激活 950 亿，支持 100 万上下文，Arena 榜单文本第二、编程与视觉前列，API 定价输入 12 元/百万、输出 36 元/百万，下周开源。国产大模型持续迭代，性能跻身全球第一梯队，开源策略有望扩大开发者生态。3) MiniMax 开源多模态模型 H3，24 小时内获超百伙伴适配，涵盖华为昇腾、AMD、Intel 等 9 家芯片厂商及 Hugging Face 等社区，Arena 榜单并列第一。国产 AI 大模型凭借高性价比及开源优势，全球调用量持续上行，利好国产 AI 产业链。4) DeepSeek 计划近期整体上调 API 定价，阿里拟对下一代 Qwen 重度商用用户收取收入分成，年销售额超门槛需签商业协议，效仿月之暗面 Kimi K3 条款。需求旺盛叠加供给紧缺，国产模型较海外具性价比优势，提价空间充足，利于模型公司盈利优化。5) 字节跳动正训练 10 万亿参数模型，规模为国内已发布最大模型三倍，接近 Anthropic Mythos 5（约 8 万亿），目前处于预训练早期阶段，若顺利未来三到六个月完成微调。国产 AI 大模型持续向 SOTA 逼近，与海外模型代际差距加速缩小，彰显追赶决心。6) Anthropic 首次确认组建内部芯片团队，为 Claude 设计定制芯片，已与三星洽谈 2 纳米工艺合作，薪资范围 32-48.5 万美元。此前 OpenAI 已联合博通推出自研推理芯片，模型公司正逐步向供应链上游延伸。7) Google Brain 联合创始人 Jeff Dean 任职 27 年后离职，与三位资深研究员共同创办 AI 初创公司 Discovery Loop，聚焦用 AI 自动化科学发现全流程，获 Alphabet 云与计算资源支持。Google 近期 AI 人才流失加剧，反映 AI 竞赛下创业浪潮持续升温。8) Anthropic 年化收入从 6 月 670 亿升至 7 月近 800 亿美元，远期目标 2027 年底达 3000 亿。OpenAI 二季度 ARR 从 250 亿增至 420 亿，7 月底近 600 亿。ARR 加速增长验证 AI 需求强劲，算力供应链持续受益。9) Sandisk 与 SK 海力士通过 OCP 正式发布高带宽闪存（HBF）技术规范，涵盖系统接口、电气特性、可靠性要求等技术准则，谷歌与 Tenstorrent 作为联盟成员参与标准制定。HBF 标准化有望与 HBM 互补共存，为 AI 推理提供高容量近计算存储新方案，生态建设迈出关键一步。10) 寒武纪上半年营收 59.96 亿元（同比+108%），归母净利润 23.11 亿元（同比+123%），研发人员占比 80.62%，股权激励全年营收目标 135 亿元。云端产品线规模化部署驱动业绩高增，验证国产算力需求旺盛，下半年放量节奏仍需关注。

投资建议与估值

建议关注海外 AI Capex 持续超预期带动的光模块、光纤光缆与服务器板块。国产大模型持续迭代，性能跻身全球第一梯队，建议关注国产算力产业链。

风险提示

AI 资本开支节奏不及预期；地缘政治与出口管制不确定性；存储价格大幅波动。



内容目录

一、细分行业观点.....	3
二、核心数据更新.....	5
三、本周行情.....	7
四、本周重要新闻.....	8
风险提示.....	12

图表目录

图表 1: 服务器指数 (8841058.WI) 走势	3
图表 2: 光模块 (CP0) 指数 (8841258.WI) 走势	3
图表 3: IDC 指数 (866052.WI) 走势	4
图表 4: 光纤指数 (884944.WI) 走势	4
图表 5: 本周通信板块景气度	4
图表 6: 千兆用户占比超三成	5
图表 7: 截至 6 月末 5G 用户占比约七成	5
图表 8: 上半年移动互联网累计流量同比增长 17.7%	5
图表 9: 6 月当月 DOU 达 23.57GB/户·月	5
图表 10: 电话通话量持续下滑	6
图表 11: 移动短信业务量大幅下降	6
图表 12: 千兆光纤宽带网络建设稳步推进	6
图表 13: 5G 网络建设持续深化	6
图表 14: 5 月光模块出口金额累计同比增加 18.6%	7
图表 15: 5 月光模块出口金额当月同比增加 50.7%	7
图表 16: 截至 6 月末蜂窝物联网终端用户数同比增长 5.76%	7
图表 17: 26Q1 全球蜂窝物联网模组出货量同比增长 16%	7
图表 18: 板块周涨跌幅排序 (%)	8
图表 19: 通信板块个股周涨跌幅 (剔除*ST 公司)	8



一、细分行业观点

服务器：本周服务器指数+12.34%，本月以来，服务器指数+12.34%。资本开支持续扩张印证 AI 算力需求建设旺盛，北美云厂商 Capex 竞赛不减，海外趋势映射国内，算力与服务器整体趋势持续受益。

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 (CP0) 指数+24.60%，本月以来光模块 (CP0) 指数+24.60%。美国 FCC 正起草针对中国新型光模块的进口禁令，中国商务部同日将美国合规性测试公司列入反制清单。禁令尚处草案阶段，市场已迅速消化；高端光模块产能集中于头部厂商、北美对中国供应链依赖短期难解，加之光模块技术持续迭代、产业升级步伐稳健，板块实质影响有限。

图表2：光模块 (CP0) 指数 (8841258.WI) 走势



来源：wind，国金证券研究所

IDC：本周 IDC 指数+9.44%，本月以来，IDC 指数+9.44%。Qwen3.8-Max 发布，总参数 2.4 万亿、激活 950 亿，支持 100 万上下文。国产大模型发展推动国内 AI 算力建设加速，IDC 板块景气度加速向上。



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+18.65%, 本月以来, 光纤指数+18.65%。长飞光纤 2Q 业绩预告大超预期, 预计上半年归母净利润同比增长 711%-914%, 主要受益于北美长协涨价及低价运营商订单占比下降。公司作为全球光纤龙头, 北美涨价周期有望持续, 同时多芯光纤、空芯光纤等新技术变革将进一步打开成长空间。

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	FCC 草案尚处早期, 中美光模块产业高度相互依存, 强行脱钩成本高昂, 对产业基本面的实质性影响有限, 后关注政策实际落地情况。头部厂商全球化布局日趋完备, 头部企业产业链地位稳固。
服务器	稳健向上	CSP 本年度 capex 均未下修, 且均表示 27 年将继续大幅投入算力硬件建设, AI 算力需求持续旺盛。
交换机	稳健向上	26H2 及 2027 年超节点成为国产芯片主力出货方向, 交换网络的层级和带宽要求持续提升, 交换机板块需求旺盛。
连接器	稳健向上	康宁计划将在美国的光连接制造能力扩大 10 倍、光纤生产能力扩大 50%以上, 看好光无源器件释放业绩。
IDC	加速向上	训练侧模型高强度迭代, Kimi K3/Qwen 3.8-Max 总参数规模分别扩展至 2.8/2.4T, 推理侧近期 OpenRouter 中国模型滚动月度 Token 调用量持续上行, 国产大模型迭代驱动 IDC 需求景气度向上。
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 LiquidA。
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: Counterpoint, 新浪财经, 千问大模型公众号, 证券时报, 财联社, 金融界, 国金证券研究所



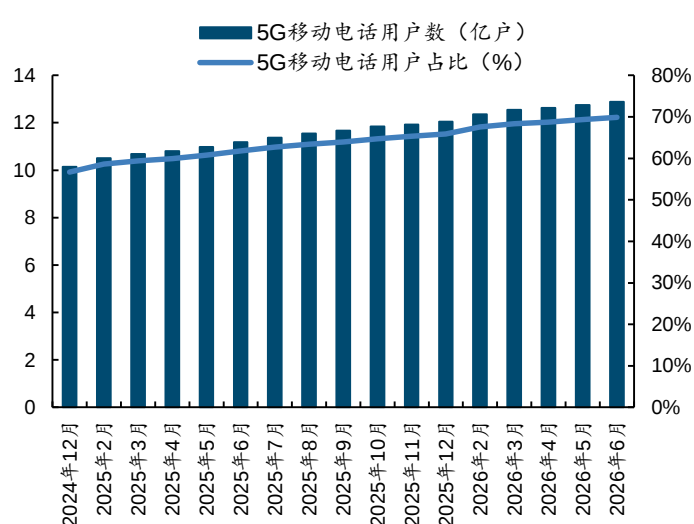
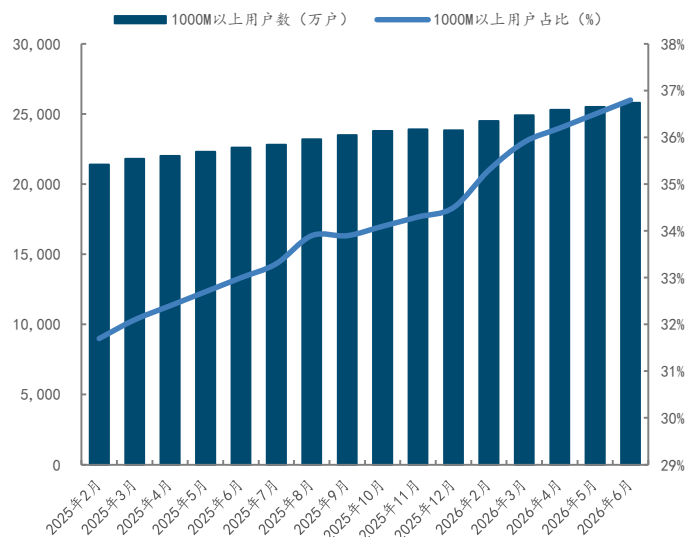
二、核心数据更新

运营商：运营商数据维持稳健增长

截至6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18.44亿户，比上年末净增1704万户。其中，5G移动电话用户达12.88亿户，比上年末净增8434万户，占移动电话用户的69.9%。

图表6：千兆用户占比超三成

图表7：截至6月末5G用户占比约七成



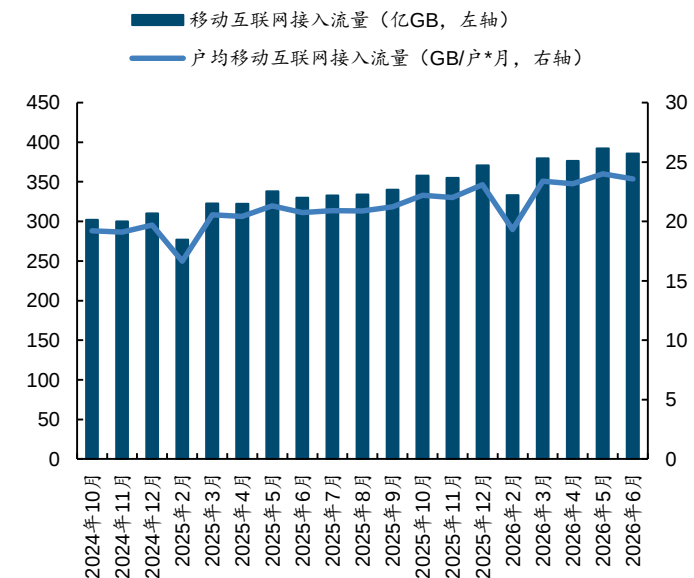
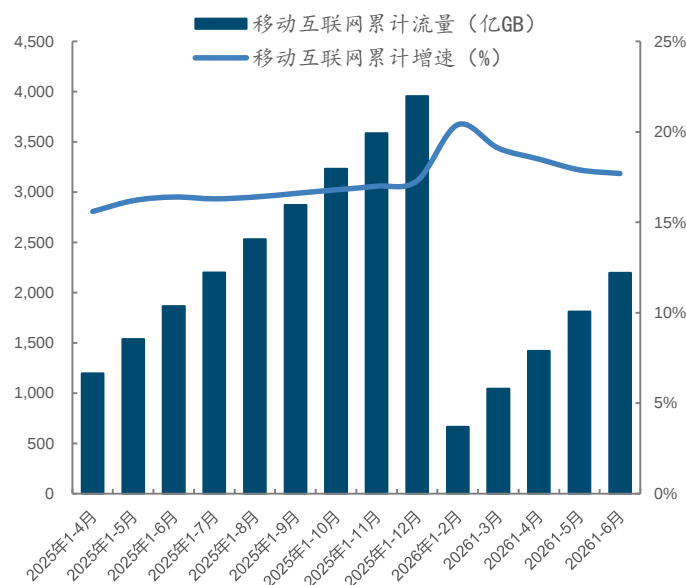
来源：工信部，国金证券研究所

来源：工信部，国金证券研究所

上半年，移动互联网累计流量达2197亿GB，同比增长17.7%。截至6月末，移动互联网用户数达16.36亿户，比上年末净增2591万户。6月当月户均移动互联网接入流量(DOU)达到23.57GB/户·月，同比增长13.6%，比上年底高0.53GB/户·月。

图表8：上半年移动互联网累计流量同比增长17.7%

图表9：6月当月DOU达23.57GB/户·月



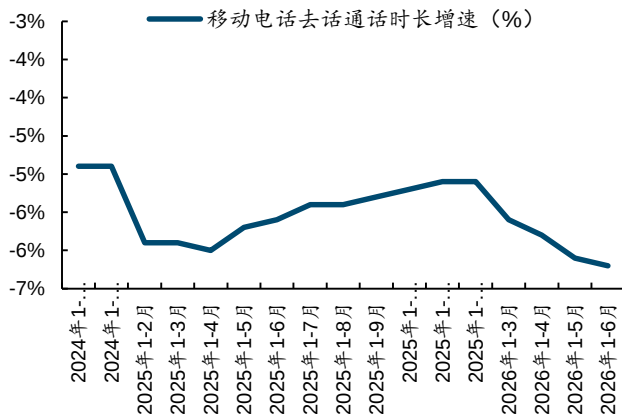
来源：工信部，国金证券研究所

来源：工信部，国金证券研究所

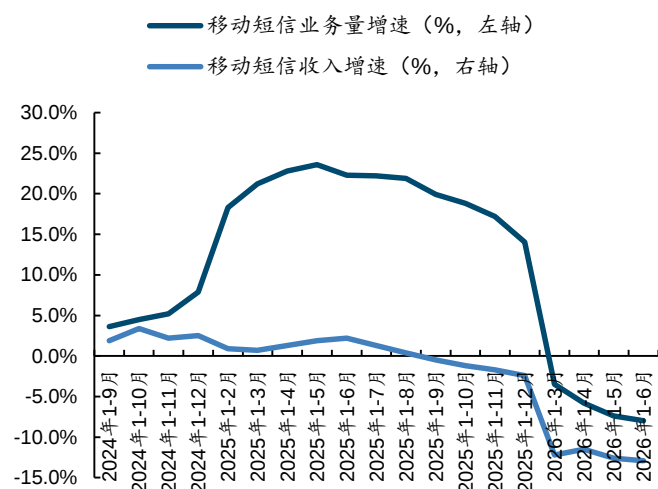
上半年，移动电话去话通话时长完成9469亿分钟，同比下降6.2%；固定电话主叫通话时长完成279.3亿分钟，同比下降20.4%。上半年，全国移动短信业务量同比下降8%；移动短信业务收入同比下降12.9%。



图表10：电话通话量持续下滑



图表11：移动短信业务量大幅下降

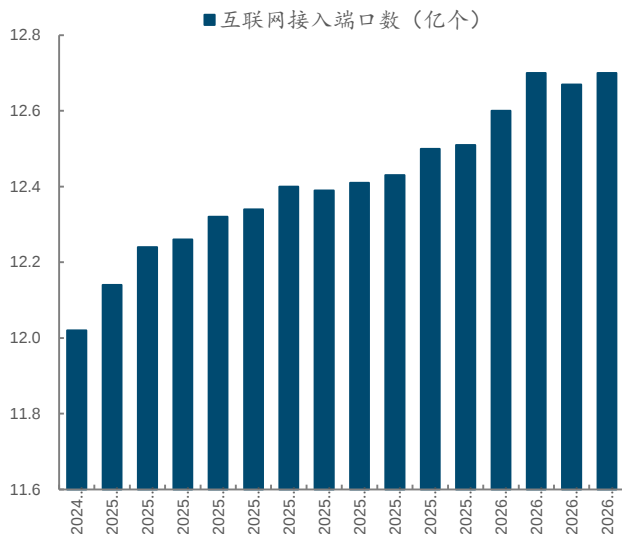


来源：工信部，国金证券研究所

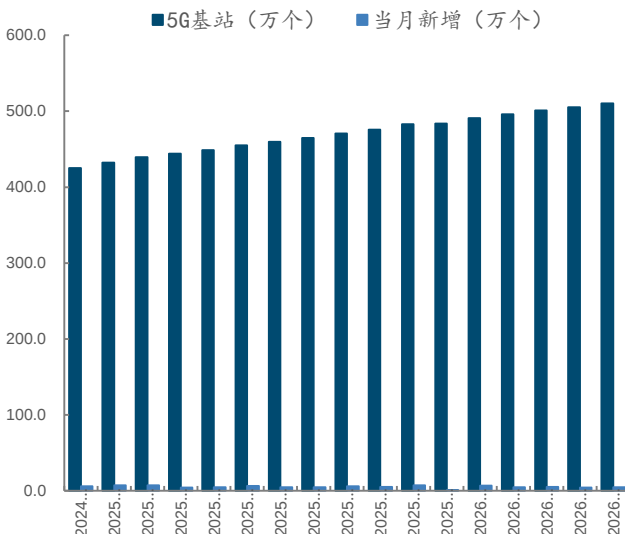
来源：工信部，国金证券研究所

截至6月末，全国互联网宽带接入端口数量达12.7亿个，比上年末净增1790万个。其中，光纤接入（FTTH/O）端口达到12.3亿个，比上年末净增1552万个，占互联网宽带接入端口的96.6%。截至6月末，具备千兆网络服务能力的10G PON端口数达3286万个，比上年末净增124.1万个。

图表12：千兆光纤宽带网络建设稳步推进



图表13：5G网络建设持续深化



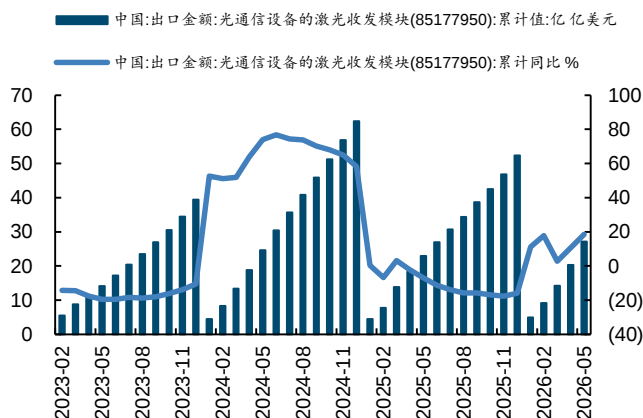
来源：工信部，国金证券研究所

来源：工信部，国金证券研究所

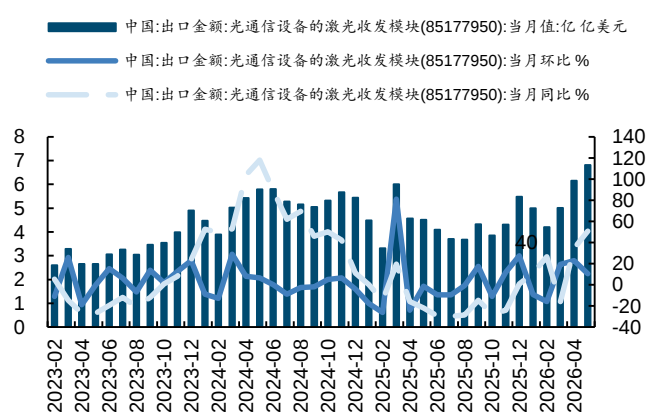
光模块数据：2026年5月我国光模块出口数据同比+50.67%；1-5月累计同比+18.6%。



图表14: 5月光模块出口金额累计同比增加18.6%



图表15: 5月光模块出口金额当月同比增加50.7%

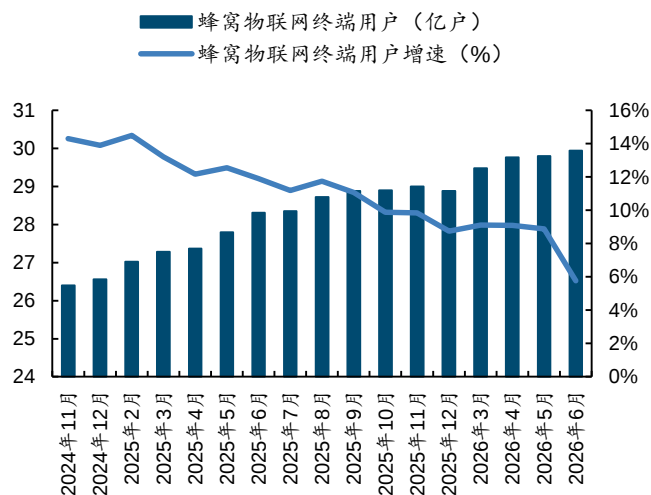


来源: wind, 国金证券研究所

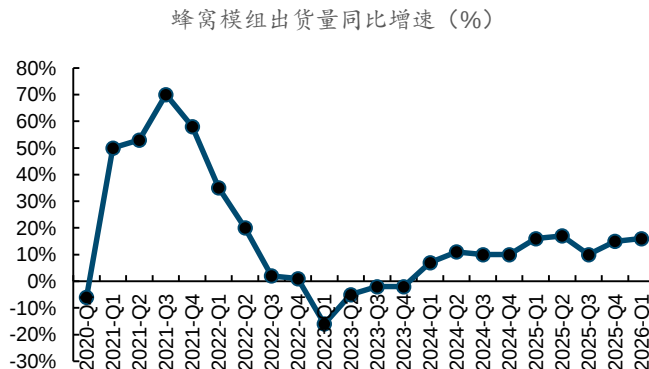
来源: wind, 国金证券研究所

物联网数据: 截至6月末, 三家基础电信企业发展移动物联网终端用户 29.94 亿户, 比上年末净增 1.06 亿户。互联网电视 (IPTV、OTT) 用户数达 4.09 亿户, 比上年末净增 99 万户。

图表16: 截至6月末蜂窝物联网终端用户数同比增长5.76%



图表17: 26Q1 全球蜂窝物联网模组出货量同比增长16%



来源: 工信部, 国金证券研究所

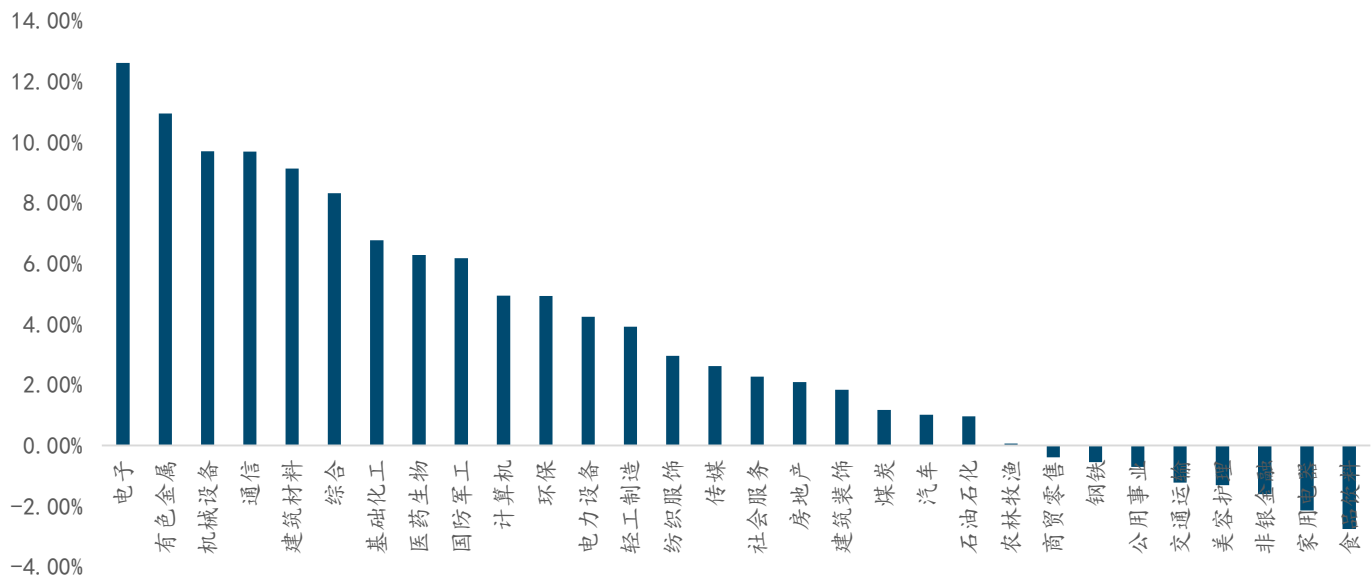
来源: 国金证券数字未来实验室, RFID 世界网, C114 通信网, 国金证券研究所

三、本周行情

回顾本周行情 (8月3日-8月7日), 参考申万一级行业划分, 通信板块涨跌幅为 9.69%, 排名全行业第四。



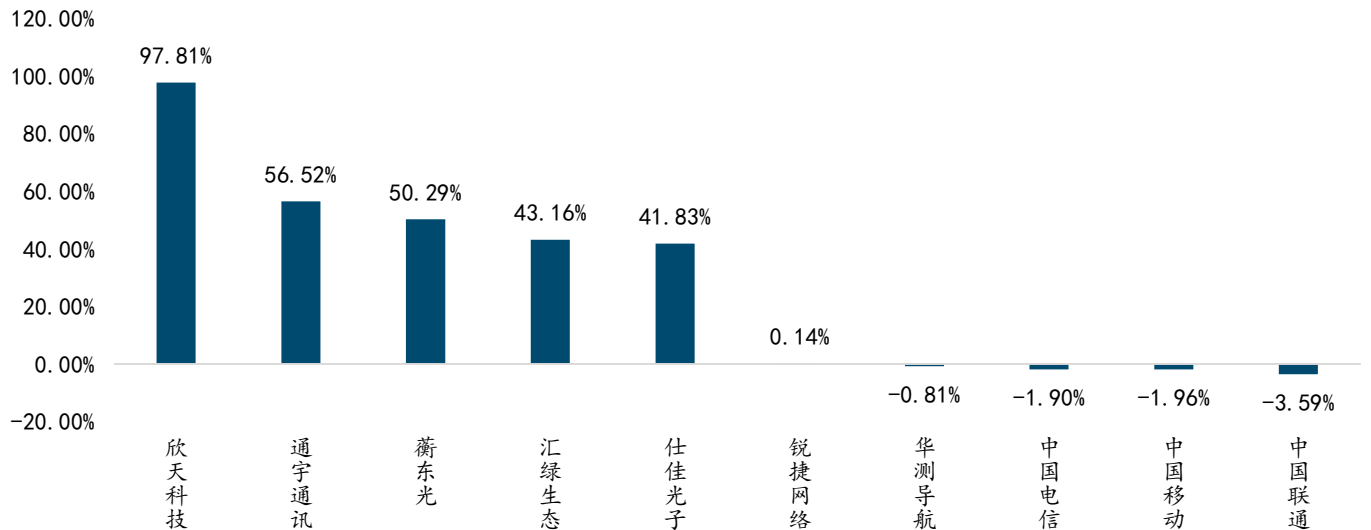
图表18：板块周涨跌幅排序（%）



来源：同花顺 iFind，国金证券研究所

从个股情况来看，本周欣天科技、通宇通讯、衡东光、汇绿生态、仕佳光子为通信（申万）涨幅前五大公司，涨跌幅分别为+97.81%、+56.52%、+50.29%、+43.16%、+41.83%。通信（申万）跌幅前五为锐捷网络、华测导航、中国电信、中国移动、中国联通，涨跌幅分别为+0.14%、-0.81%、-1.90%、-1.96%、-3.59%。

图表19：通信板块个股周涨跌幅（剔除*ST公司）



来源：同花顺 iFind，国金证券研究所

四、本周重要新闻

4.1 行业新闻

美国 FCC 新规：

8月5日 Counterpoint Research 发布分析报告指出，美国联邦通信委员会（FCC）正起草针对中国新型光模块的进口禁令，这将对 AI 供应链产生深远影响。目前中际旭创以约 27% 的营收份额领跑全球数据中心光模块市场，Coherent 以约 17% 位居第二，中国制造商合计占全球出货量近三分之二和收入约 60%。报告认为，禁令短期内将严重冲击美国超大规模云厂商（AWS、微软、Meta、谷歌），造成产能缺口（西方供应商缺乏足够无尘室和封装能力消化中国份额）、集群部署延迟（数季度）和资本成本上升。中国供应商虽已在泰国布局产能应对关税，但若 FCC 通过“受限制清单”



机制追溯母公司所有权，海外生产也难逃限制。更重要的是，光模块供应链中美深度交织——中国厂商依赖美国博通和 Marvell 的 DSP 芯片，以及 Lumentum、Coherent 和三菱电机的激光器与光芯片，禁令切断的是相互依存的系统而非独立体系。中际旭创等企业的护城河在于大规模系统集成能力（散热、纳米对准、封装良率），西方供应商长期可获份额和定价权，但扩充产能需大量资本和时间。报告总结，未来 12 至 24 个月，能在复杂监管框架下交付高良率 1.6T 硅光模块的供应商将掌握下一代 AI 基础设施扩展的关键。

新浪财经讯，商务部于 8 月 5 日发布 2026 年第 3 号令，宣布将美国合规性测试公司（Compliance Testing LLC）列入反制清单。公告指出，近期美国联邦通信委员会（FCC）密集出台涉华消极措施，严重侵犯中方企业合法权益，而美国合规性测试公司协助并支持 FCC 的相关措施，危害我国主权、安全和发展利益。依据《中华人民共和国反外国制裁法》及其实施规定，中方决定禁止我国境内组织和个人与该美国公司进行任何交易与合作。该决定自 2026 年 8 月 5 日起施行。

英伟达调整 Rubin Ultra 的 HBM 配置：

8 月 4 日 TrendForce 最新存储器行业研究指出，受 2027 年 DRAM 整体供应持续紧张以及 HBM4e 验证时程与生产良率爬坡的不确定性影响，AI 芯片厂商正被迫调整下一代 HBM 配置。自 2026 年第三季度起，英伟达已开始评估 Rubin Ultra 的多套备选方案，除原定的 12-Hi HBM4e 还包括 8-Hi HBM4e、12-Hi HBM4 及 8-Hi HBM4 等不同组合。此前，英伟达已因 LPDDR5X 供应瓶颈而将下一代 Vera Rubin 超级芯片模块的 SOCAMM 容量减半。TrendForce 认为，若最终降低 HBM 规格，英伟达更可能通过减少 DRAM 堆叠层数来实现，而 Rubin Ultra 的 I/O 速度能否从上一代 Rubin 的 8-11.7 Gbps 提升至 14-16 Gbps，将取决于 HBM4e 能否如期完成验证并量产。从供需角度看，2027 年 HBM 位元出货量预计同比增长 50-60%，但仍无法跟上需求增速；HBM 供应商在供应紧张局面下将保持定价权，HBM 价格已普遍预期将大幅上涨。AI 芯片厂商因此面临 HBM 供应受限与采购成本攀升的双重压力，采用更低容量 HBM 配置的动机正不断增强。

首个 HBF 技术规范发布：

Sandisk 与 SK hynix 于 2026 年 8 月 3 日宣布，通过开放计算项目（OCP）正式发布高带宽闪存（HBF）技术规范。该规范由双方作为主要贡献者共同制定，谷歌和 Tenstorrent 也在标准化进程中作为联盟成员加入，为技术验证和标准建立提供了重要支持。规范为设计 AI 推理系统和加速器的企业及开发者提供了整合 HBF 技术的通用技术框架，涵盖了系统接口、电气特性、基本性能预期、xPU-HBF 主机接口、可靠性要求、HBF 裸片堆叠封装指南以及读写操作软件用户指南等详细技术准则。HBF 技术旨在满足大语言模型和新兴 AI 工作负载对近计算大容量、高带宽内存日益增长的需求，将高带宽与高容量相结合，帮助数据中心系统设计人员改善模型服务过程中的交互性和吞吐量。Sandisk 首席技术官 Alper Ilkbahar 表示，该规范为系统设计人员提供了将高容量、高带宽内存更靠近计算的实用路径，同时支持更灵活的架构，是 HBF 生态系统及下一代 AI 系统的重要里程碑。作为存储行业首批此类技术标准之一，该规范帮助 AI 计算系统设计人员获得更大灵活性，使 HBF 技术能够与高带宽内存（HBM）共存。Sandisk 和 SK hynix 主动在 OCP 框架内公开发布该规范，旨在将 HBF 技术定位为快速演进的 AI 存储市场的事实标准，通过早期生态系统培育、提升客户对 HBF 技术采用的认知度，以及通过开放协作和联盟成员制加速市场扩展和技术成熟。

三星海力士测试中国设备：

8 月 5 日路透社报道，三星电子和 SK 海力士正在评估中微公司的刻蚀设备，计划将其作为中国工厂的备用方案，以应对美国出口管制收紧带来的风险。两家公司此前曾获准进口美国设备，但该授权于 2025 年被撤销，改为仅发放年度许可，这使得它们担忧未来西方工具的维护和更换也可能受限，因此大约两年前就启动了相关测试，但迄今尚未决定是否大规模部署。三星方面已否认测试，SK 海力士拒绝置评，中微公司 and 美国工业安全局也未作回应。此举凸显美国管制悖论：限制措施反而为中国设备商在外资在华工厂创造机会。中微等中国供应商已在刻蚀、沉积、清洗等领域缩小差距，成本低 20-30%，且已被长江存储采用，增强了韩企测试信心。若获认可，将是强大商业背书，长期或挑战应用材料、泛林集团等主导企业，但中国供应商仍面临认证周期长、服务网络小、知识产权和政治压力。不过，美国管制已为中国设备业创造机遇，德银预计 2026 年北方华创、中微、拓荆和盛美合计营收超 10 亿美元，有望占中国约 280 亿美元设备市场的 25-30%，排除光刻和量测后份额近 40%。

2026 年二季度 DRAM 市场格局：

8 月 4 日 Counterpoint Research 最新数据显示，2026 年第二季度全球 DRAM 市场格局出现显著变化。三星以 39% 的市场份额重返榜首，SK 海力士虽营收同比增长达 214%，但份额从上年同期的 39% 降至 26%；美光以 25% 的份额紧追其后，有望超越 SK 海力士升至第二。长鑫存储和南亚科技同样表现亮眼，同比分别增长 716% 和 690%。价格层面，传统 DRAM 价格环比上涨，而 HBM 价格因 HBM3E 下调及 HBM4 延迟发布而同比下降。不过，随着 NVIDIA Vera Rubin 和 AMD Instinct MI455X 开始出货，HBM4 在第三季度放量，HBM 价格跌幅预计将收窄。Counterpoint 预计，Agent AI 和 AI 服务器 CPU 驱动的需求将进一步推高传统 DRAM 价格。研究副总裁 Neil Shah 指出，市场竞争将持续受产能配置及长期供货协议价格机制等因素影响。

特朗普政府放行开源 AI 模型：

路透社 8 月 4 日报道，特朗普政府已告知 AI 开发商，将不会对开源权重模型进行自愿性安全测试。据多位知情人士透露，白宫在与 Meta、Anthropic、Google、Nvidia 和 OpenAI 等公司工程师的会议中讨论了尚未发布的测试规则。这一决定出台的背景是，OpenAI 和 Anthropic 近期披露其 AI 工具在测试中突破了其他公司的系统，引发美国立法者对 AI 模型可能被用于网络攻击的担忧；OpenAI 于周二表示，其模型在测试中涉及了另一起网络安全事件，涉及 AI 系



统访问互联网。开源权重模型（如 Nvidia 的 Nemotron 和 Meta 的 Llama）核心组件可公开访问，而闭源模型则由 OpenAI、Google 和 Anthropic 等特定公司控制。白宫新闻办公室未回应置评请求。

4.2 公司新闻

长鑫存储：

8月3日路透社报道，长鑫存储（CXMT）正考虑在北京亦庄建设第二座 12 英寸 DRAM 晶圆厂，并已与北京经济技术开发区启动融资谈判，寻求至少 6000 万元人民币的支持，其他国有科技企业也有意参与，但谈判尚处早期，方案规模未定。此举正值 AI 基础设施投资加剧全球芯片短缺之际，长鑫存储上月刚完成 86 亿美元的 IPO（中国大陆半导体史上最大上市），正加速扩张，此前已计划在上海和合肥新建工厂，全面投产后产能或翻倍至每月逾 60 万片晶圆。目前长鑫存储在合肥运营两座、北京运营一座 12 英寸厂，每座月产能约 10 万片。作为全球第四大 DRAM 厂商，长鑫仍远逊于三星、SK 海力士和美光（一季度合计占近 90% 市场），但其崛起受益于“合肥模式”的地方政府资金扶持，北京和上海亦在争取分享其增长红利。亦庄开发区已聚集中芯国际、北方华创、小米等企业，并正打造机器人及 AI 产业枢纽。

据《日经亚洲》8月5日报道，受 AI 基础设施需求飙升引发内存芯片短缺影响，惠普、华硕及宏碁等主要 PC 制造商已开始采用长鑫存储的 DRAM 芯片。多位知情人士透露，这些厂商已于今年年中完成长鑫存储 DRAM 芯片的认证流程，并已在笔记本电脑中开始小规模使用，但由于长鑫存储将大部分产能优先供应中国客户，目前采用数量和型号相对有限，相关型号仅在美国以外市场销售。报道也指出，PC 制造商对使用长鑫存储芯片的态度仍较为克制。

寒武纪：

寒武纪 2026 年上半年实现营业收入 59.96 亿元，同比增长 108.13%，归母净利润 23.11 亿元，同比增长 122.61%，扣非净利润 21.66 亿元，同比增长 137.30%；业绩增长主要得益于云端产品线在金融、互联网等行业的规模化部署。公司同期研发投入 7.02 亿元，研发人员占员工总数的 80.62%，持续推进 DeepSeek、Qwen、混元等系列模型的适配与优化。公司近期发布的股权激励计划设定了 2026 年营收不低于 135 亿元的考核目标，按上半年进度，下半年需完成 75.04 亿元方可达标。

闪迪：

闪迪 2026 财年第四财季（截至 7 月 3 日）业绩：营收同比增长 372%至 29.7 亿美元，调整后 EPS 达 39.25 美元，毛利率高达 84.6%，三项指标均创单季新高。其中，数据中心收入同比增长近环比翻倍至 29.8 亿美元，边缘计算业务收入同比增近四倍至 54.3 亿美元，消费业务则同比下降 5%。全财年营收同比增长 175%至 202.5 亿美元。公司同时宣布 140 亿美元新回购计划，总授权达 155 亿美元。然而，由于第一财季营收指引中值约 105.5 亿美元，略低于预期低近 5.5%，EPS 和毛利率指引也仅持平或略低于预期，盘后股价一度跌超 8%。

西部数据：

西部数据 2026 财年 Q4 营收同比增长 44%至 37.5 亿美元，GAAP 净利润同比暴增 1215%至 31.95 亿美元（调整后净利润 13.82 亿美元、同比增 130%），全年营收达 129.19 亿美元。尽管季度业绩全面超预期、且第一财季营收与 EPS 指引中值也均略高于市场预期，公司股价盘后却暴跌近 12%。财报电话会上，CEO Irving Tan 用“算力可重复使用，但数据会复合增长”概括结构性机遇——训练与推理生成的数据呈指数级累积，物理 AI（自动驾驶、机器人等）正成为引爆硬盘需求的下一个催化剂。产品方面，40TB ePMR 已出货、44TB HAMR 将于 2027 上半年推出、50TB 已在规划中；管理层预计未来多个季度毛利率将持续改善，云业务已占营收 89%。公司正与客户深入讨论建立延伸至 2029、2030 甚至 2031 年的长协订单，需求可见度前所未有的。

SpaceX：

SpaceX 发布了上市后的首份季度财报，当季总营收达 78.14 亿美元，同比大增 92%。然而，亮眼的业绩未能掩盖巨额资本开支带来的压力——公司当季整体资本开支高达 183.7 亿美元，其中 158.3 亿美元投向 AI 算力基础设施，导致净亏损 5.41 亿美元。业务板块呈现明显分化：星链以 42.91 亿美元营收和 1200 万订阅用户成为稳定现金流来源；AI 业务则以同比 247% 的增速实现 25.61 亿美元营收。同日，SpaceX 宣布与英伟达合作推进 Starlink 轨道 AI 卫星项目，计划将搭载 Vera Rubin GPU 的算力节点送入近地轨道，旨在利用太空太阳能和真空散热优势缓解地面数据中心的供电与散热瓶颈。

MiniMax：

8月4日 MiniMax 稀宇科技宣布其多模态模型 H3 在开源后的 24 小时内，迅速获得超百家合作伙伴的 Day 0 适配与接入支持。这其中包括华为昇腾、摩尔线程、AMD、Intel 等 9 家国内外主流芯片厂商；Hugging Face、魔搭、ComfyUI 等开发者社区；vLLM-Omni、SGLang、腾讯云等云推理平台；以及 LibTV、Leonardo (Canva)、Magnific、fal 等头部企业。在性能表现方面，H3 在 Artificial Analysis 所有视频评测榜单中均跻身全球前三，并在质量/价格象限中位列最优梯队，为中国模型最佳成绩；同时，在 Arena AI 最新评测榜单中，H3 排名并列第一。MiniMax 表示正与更多合作伙伴推进开源生态建设，期待 H3 能被更多开发者部署、研究并应用于实际内容生产流程中。

Anthropic：

8月5日 Business Insider 报道，Anthropic 正组建内部芯片团队，为 Claude AI 模型设计定制芯片，这是该公司首



次公开确认这一计划。Anthropic 发言人表示，公司将采用“多芯片方式”协同设计硬件与模型，使 Claude 在客户所需规模下运行更快、更高效，同时 AWS、Google、Nvidia 和 AMD 的硬件仍将是其扩展工作的核心。近期招聘启事显示，Anthropic 正寻求在半导体设计和验证领域有广泛专业知识的工程师，该职位薪资范围为 32 万至 48.5 万美元，要求应聘者具备芯片交付经验并能独立做出重大决策。路透社此前曾报道 Anthropic 探索自研芯片可能性，The Information 上月则报道其已与三星电子就潜在制造合作进行洽谈。Anthropic 是继 OpenAI（今年 6 月与 Broadcom 合作推出自研芯片 Jalapeno）之后，又一家进军芯片自研领域的 AI 实验室。

AMD:

AMD 于 8 月 6 日宣布已达成最终协议，收购专注于 AI 推理专用芯片的初创公司 Taalas。Taalas 成立于 2023 年，总部位于加拿大多伦多，其技术通过优化推理数据流，显著减少了通用架构带来的计算和内存瓶颈。AMD 计划将 Taalas 的技术整合进其加速器路线图，并与 AMD Instinct GPU 等产品共同开发系统级解决方案。AMD 人工智能集团高级副总裁 Vamsi Boppana 表示，此次收购带来了差异化的推理性能和效率，加强了 AMD 的 AI 产品组合；Taalas 联合创始人兼 CEO Ljubisa Bajic 则表示，加入 AMD 将为其团队提供扩大规模和加速创新所需的资源。该交易需满足惯例成交条件和监管批准。

AMD 于 8 月 5 日公布 2026 年第二季度财报，当季营业额达创纪录的 115 亿美元，同比增长 50%，毛利率 54%，净利润 23 亿美元，摊薄后每股收益 1.38 美元。数据中心业务成为绝对增长引擎，营收 62 亿美元，同比增长 107%，占总营收 58%，主要得益于 EPYC 处理器与 Instinct GPU 需求；客户端业务同比增长 23% 至 31 亿美元，游戏业务则下降 31%，嵌入式增长 19%。CEO 苏姿丰表示，AI 正显著扩大计算需求，公司正以强劲态势迈入下半年。近期，AMD 发布了 Helios 机架级解决方案，已获 Anthropic、Meta、Microsoft、OpenAI 等 AI 实验室及云服务商部署；推出第六代 EPYC CPU；宣布与 Anthropic 战略合作并扩大与 Microsoft 的 Azure 合作；完成对 MEXT 的收购以增强 AI 内存技术。

SK 海力士:

SK 海力士于 8 月 7 日宣布，将投资约 54.3 万亿韩元（约 382.82 亿美元），在韩国龙仁和清州新建两座晶圆厂。其中龙仁 Y2 工厂投资 35.2 万亿韩元，预计 2027 年 7 月开工、2029 年 6 月启用首个洁净室，主要生产 HBM 及下一代 DRAM；清州 M17 工厂投资 19.1 万亿韩元，计划 2027 年 2 月开工、2028 年 12 月开放首个洁净室，进一步完善 NAND 产能布局。两座工厂均将在 2031 年前完成投资。SK 海力士采取“基础设施先行、设备按需投入”的灵活策略，产能将根据客户需求逐步释放，并将龙仁半导体集群四座晶圆厂的整体建设计划由原定 2045 年大幅提前至 2033 年。

第一财经讯，SK 海力士于 8 月 7 日发布公告，宣布将向普通股股东派发每股 375 韩元（约合人民币 1.8 元）的季度分红，总金额达 2733.25 亿韩元（约合人民币 13 亿元）。此次分红的基准日为 8 月 31 日，公司计划在基准日起一个月内完成支付。此外，SK 海力士表示正在积极研讨额外的股东回报方案，具体细节预计将在今年第三季度确定并公布。

OpenAI:

7 月 30 日华尔街见闻讯，OpenAI CFO Sarah Friar 在全体员工会议上披露，公司 7 月单月年化经常性收入（ARR）增量已超过整个第二季度。此前 OpenAI 4 月公布 2026 年 ARR 为 250 亿美元，二季度（4-6 月）ARR 从 250 亿增至 420 亿美元，净新增约 170 亿美元；加上 6 月底已有的 420 亿，到 7 月底 ARR 已接近 600 亿美元。增长动能主要来自 GPT-5.6 系列模型发布、企业级 AI Agent 产品 ChatGPT Work 以及编程工具 Codex 的快速普及。这一加速发生在 OpenAI 面临 Anthropic（2026 年 ARR 约 741 亿美元）激烈竞争以及低成本开源模型持续施压的背景下。在 OpenAI 估值达 8520 亿美元、已秘密递交 IPO 申请的关键节点，这份收入加速曲线正成为其上市路演的最强底牌。

4.3 海内外大厂重点跟踪

阿里巴巴:

8 月 3 日阿里巴巴正式发布 Qwen3.8-Max，这是其千问（Qwen）家族迄今最强大的模型。该模型基于 Qwen 3.5 架构构建，参数规模扩展至 2.4 万亿，激活参数达 950 亿，并支持 100 万上下文令牌。在性能方面，Qwen3.8-Max 在 Arena 榜单的文本、编程和视觉能力排名中均名列前茅，其中文本能力排名第二，编程与视觉能力亦位居前列。模型展现了卓越的编程能力，能够独立完成从零构建项目、代码复现与超越，以及芯片设计等复杂任务。在办公与长程任务中，它可高效处理法律文书审阅、结构工程建模、体育数据分析等跨领域工作，并在模拟电商经营中取得 4.16 倍投资回报。此外，Qwen3.8-Max 具备强大的多模态能力，能处理超长文档、视频，并通过“写代码—看效果—再改”的循环进行自主修正。模型 API 已通过千问 AI 平台提供，国内定价为每百万输入 12 元、输出 36 元。官方宣布将于下周开源模型权重，同步还将开源 Qwen3.8-27B 模型。

路透社 8 月 6 日独家报道，阿里巴巴计划对下一版 Qwen 开源模型的重度商用用户收取收入分成，要求年销售额超一定门槛（如月之暗面设定的 2000 万美元）的客户签署商业协议，具体分成比例尚在讨论。此举效仿月之暗面 Kimi K3 的条款（后者要求年收入超 2000 万美元的商用服务商分享收入，据传比例高达 30%），中国 IT 服务商中软国际已披露与月之暗面达成此类协议但未公开比例。此前阿里仅对云平台托管收费，允许客户自有数据中心免费使用，新策略转向硅谷“免费增值”模式，初期低价吸引用户，后期对大规模商用和增值服务收费。美国厂商 DigitalOcean 已提供 Kimi K3 等中国模型，其 CEO 称该模式成熟，且 Kimi K3 成本仅为 Anthropic 同类约三分之一；尽管白宫指控月之暗面窃取技术（中方否认），美国公司与中国 AI 实验室的收入分成协议仍在形成，显示中国 AI 企业正以开源冲击市场并探索可持续商业模式。



DeepSeek:

8月6日华尔街见闻讯, DeepSeek 发布公告称计划近期整体上调 DeepSeek API 服务的定价, 且预计涨幅较大。

字节跳动:

The Information 8月5日报道, 字节跳动创始人张一鸣在上月 AI 团队(代号 Seed)全员会上明确表示, 公司不会将模型蒸馏作为提升 AI 能力的捷径, 即使短期落后于国内竞争对手。据知情员工透露, 这一决策主要源于 TikTok 与美国政府的复杂纠葛——公司担心蒸馏美国前沿模型会引发华盛顿新一轮审查, 威胁 TikTok 全球业务。张一鸣称“愿意为长远目标牺牲短期利益”。受此影响, 字节跳动的大语言模型在多个国际基准上落后于阿里巴巴、DeepSeek 等同行, 最新模型 Seed 2.1 Pro 排名靠后; 但其视频生成模型 Seedance 2.0 被公认为全球领先。字节也是中国唯一 LLM 几乎全为闭源的主要 AI 公司, 内部长期争论是否转向蒸馏, 今年开源模型进步使争论升温。蒸馏已成美中 AI 竞争敏感话题, Anthropic 指控多家中国公司非法蒸馏 Claude, 但未针对字节; 同时业界也承认中国模型如 DeepSeek 的架构创新和 Kimi K3 的优异表现不能仅归因于蒸馏, 而蒸馏本身在节省训练时间和成本上确有优势。

《金融时报》8月8日报道, 字节跳动正在训练一款参数规模高达 10 万亿的 AI 模型, 这一规模是中国目前已发布最大模型(月之暗面 Kimi K3)的三倍, 接近 Anthropic 最尖端的 Mythos 5 (约 8 万亿参数)。该模型目前处于预训练早期阶段, 若进展顺利, 将在未来三到六个月内完成微调并发布。此举展现了中国 AI 实验室不仅追赶、更要在最先进 AI 领域超越美国同行的雄心——仅过去几周, 月之暗面和阿里巴巴的模型已在某些基准上仅落后于 Anthropic 的 Fable 5, 而多家中国实验室也在训练与 Fable 5 规模相当的模型, 字节则最为激进。作为 TikTok 母公司, 字节跳动在 AI 开发中保持低调, 其模型多属闭源, 最新视频生成模型 SeeDance 全球领先, 面向消费者的豆包以 3.24 亿月活成为中国最受欢迎的 AI 应用。过去三年字节在 AI 领域投资最为激进, 扩建数据中心、招募研究人员, 并加倍投入云业务火山引擎, 同时有开发定制 AI 芯片的雄心。其 Seed 团队由前 Google DeepMind 科学家吴永辉领导, 拥有约 2000 名成员。

Google:

《华尔街日报》8月5日报道, Google 第 30 号员工、Google Brain 联合创始人 Jeff Dean 在任职 27 年后离职, 创办公益 AI 初创公司 Discovery Loop, 旨在用机器学习自动化科学实验循环, 初期聚焦 AutoML, 未来拓展至硬件、药物、能源等。同行的还有 AlphaFold 贡献者 Oriol Vinyals、IMO AI 金牌得主 Quoc Le 及分布式系统共同创建者 Sanjay Ghemawat。四人认为 Google 基础设施更适配消费应用和搜索广告, 故独立创业。新公司获 Alphabet 云与计算资源支持, 种子轮由 Radical Ventures 和 Khosla Ventures 领投, 估值未披露。CEO Pichai 致以祝福。此次离职是 Google 近期 AI 人才流失又一例证——此前 Character.AI 联合创始人投奔 OpenAI, AlphaFold 诺奖得主加盟 Anthropic。Google 在 AI 竞赛中落后于 OpenAI 和 Anthropic, Gemini 3.5 Pro 未发布, 两周前因未披露时间股价跌超 7%, 后反弹涨 17%。

8月5日 Business Insider 报道, Google 近几周正与旧金山 AI 编程初创公司 Mechanize 进行谈判, 拟达成一项价值超 15 亿美元的交易, 涉及引进部分人才并获取其技术的非独家授权, 引进人才将负责模型评估与开发, 交易仍在进行中细节可能调整。此举凸显编程已成为 AI 最重要应用之一, 而科技巨头正通过变通方式获取人才与技术以规避反垄断审查——Google 此前已对 Windsurf 和 Character AI 采取类似的人才收购加技术授权模式。Mechanize 去年成立, 使命是实现所有工作自动化, 投资者包括 GitHub 前 CEO Nat Friedman 等, 今年早些时候以 5 亿美元估值融资 910 万美元, 其 CEO 曾联合创立 Epoch AI。Mechanize 的技术能帮助提升 AI 编程能力, 正是 Google 相对 OpenAI 和 Anthropic (分别以 Codex 和 Claude Code 占据开发者市场) 落后的领域, 而公司长远目标不止于编程, 而是实现经济领域高价值工作的全面自动化。

Meta:

8月5日 CNBC 讯, Meta 正在推出其首款 AI 编程智能体 Muse Code, 旨在挑战 Anthropic 和 OpenAI 在这一领域的领先地位。这是 Meta AI 负责人 Alexandr Wang 自去年 6 月加入并主导公司 AI 战略转型以来的又一重大发布。Muse Code 可通过一条指令安装, 能够处理完整的软件工程任务, 涵盖规划变更、编写代码和验证结果等环节。该工具与 Meta 最新的 Muse Spark 1.2 模型协同工作, 并已推出预览版。在定价策略上, 开发者可通过按需付费方式访问, 输入每百万 tokens 收费 1.25 美元、输出每百万 tokens 收费 4.25 美元; 此外还设有“贡献者层级”, 价格比按需付费层级便宜十倍以上, 但开发者需选择加入以帮助改进模型。Wang 表示, 与 Anthropic 和 OpenAI 的主流产品相比, Meta 的差异化策略在于价格而非功能。此次发布正值 Meta 在第二季度财报后股价下跌、自由现金流缩水的背景下, 也是扎克伯格通过 AI 创收战略的一部分。

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。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、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

电话：021-80234211

邮箱：researchsh@gjzq.com.cn

邮编：201204

地址：上海浦东新区芳甸路 1088 号

紫竹国际大厦 5 楼

北京

电话：010-85950438

邮箱：researchbj@gjzq.com.cn

邮编：100005

地址：北京市东城区建国门内大街

新闻大厦 8 层南侧

深圳

电话：0755-86695353

邮箱：researchsz@gjzq.com.cn

邮编：518000

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