

分析师：唐月
登记编码：S0730512030001
tangyue@ccnew.com 021-50586737

AI 应用大发展令算力总体供不应求，国产 AI 芯片厂商迎来重要发展机遇

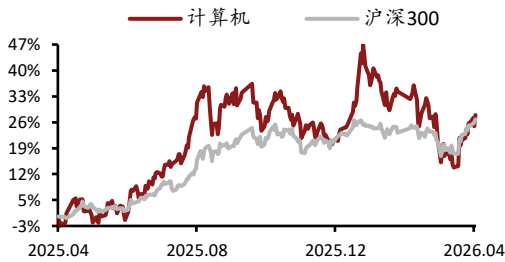
——计算机行业月报

证券研究报告-行业月报

强于大市(维持)

计算机相对沪深 300 指数表现

发布日期：2026 年 04 月 23 日



资料来源：中原证券研究所，聚源

相关报告

- 《计算机行业月报：中国 AI 超级周开启，算力呈现提价趋势》 2026-02-13
- 《计算机行业分析报告：DeepSeek 近期成果分析及 V4 影响力预测》 2026-01-29
- 《计算机行业月报：AI 应用全面加速，DeepSeek V4 有望深刻改变全球 AI 的竞争格局》 2026-01-22

联系人：李智

电话：0371-65585629

地址：郑州郑东新区商务外环路 10 号 18 楼

地址：上海浦东新区世纪大道 1788 号 T1 座 22 楼

投资要点：

- 国产化：**2026 年前三个月份，我国集成电路出口数量和金额分别同比增长 13%和 73%，出口集成电路金额占比进口金额占由 2022 年的 37.1%提升到了 2026 年前两个月份的 56.6%。2025 年下半年，我国 AI 芯片国产化比率从 2025 上半年的 35%提升到了 46%，提升趋势明显。国产 AI 芯片厂商 2025 年业务景气度持续提升，其中 2 家实现盈利。2026 年国产 AI 芯片在加速发展的过程中，有望开启向海外市场的拓展。
- 算力：**2025Q4 全球服务器市场中，美国同比增速 72.4%，而中国增速仅为 17.7%，较 2024Q4 93.3%的增速下滑明显，我们认为主要源于缺芯因素影响了服务器需求的释放。2026 年随着移动集采大单落地和郑州 6 万卡超算集群的建设，超节点正在成为发展趋势。2025 年算力建设受制于芯片供给，26 年在 AI 应用增长和国产芯片供给能力提升的双加持下，行业算力建设速度有明显加快迹象，第三方数据中心厂商开启积极建设步伐。
- AI：**目前国内模型在能力上主要集中在第二梯队，随着厂商新模型的陆续发布，整体水平上更加集中，且性价比优势更加凸显。Anthropic 年支出超 100 万美元的企业客户数从 2 月的超 500 家增长到 4 月超 1000 家，业务增长弹性在 2026 年充分显现。3 月豆包日均 token 使用量突破 120 万亿，季度环比增长 140%。AI 应用需求爆发式增长，拉动算力价格持续上涨，模型厂商、云厂商发布多轮涨价通知。
- 给予行业强于大市的投资评级。2026 年 AI 应用发展速度整体超预期，带来了整体算力产业供应链的紧张。2026 年来看，在海外芯片供给受限的大背景下，国内 AI 芯片也迎来了性能、产能的双方改善，有望持续受益于市场结构的变化，并最终改善国内算力的供给能力。建议关注有大规模智算中心交付计划的字节跳动数据中心核心供应商润泽科技（300442），有产业链一体化优势同时持股优质企业的服务器厂商中科曙光（603019）。同时建议积极关注正在积极开展 IPO 的长鑫科技、芯和半导体、合见工软、超聚变、平头哥、智谱。**

风险提示：国际局势的不确定性；海外 AI 产业竞争格局变化带来市场调整风险。

内容目录

| | |
|---|-----------|
| 一、 行业数据 | 5 |
| 1.1. 行业数据：2026 年 1-2 月软件行业收入增速回落，但是出口加速趋势明显 | 5 |
| 1.2. 2026 年 1-2 月高景气赛道：IC 设计、嵌入式软件 | 6 |
| 1.3. 国产化：2025 年下半年 AI 芯片国产化率进一步提升到了 46% | 9 |
| 1.3.1. 国产芯片 | 9 |
| 1.3.2. 纯血鸿蒙系统终端数量突破 5000 万 | 14 |
| 1.4. 算力：2025 年底我国智算规模较年中翻倍 | 16 |
| 1.4.1. 英伟达将在 2026 年下半年出货 Rubin，2027 年将采用的 4-Die 方案或遇阻 | 16 |
| 1.4.2. 自研芯片：正在模型厂商和云厂商成为算力供应的重要补充 | 17 |
| 1.4.3. 服务器：芯片供应影响了国内服务器增长的节奏 | 21 |
| 1.4.4. 数据中心：2025 年底我国智算规模较年中翻倍，2026 年有加速迹象 | 27 |
| 1.4.5. 云计算：受到 AI 需求拉动，整体增速仍在加快 | 28 |
| 1.4.6. 资本开支 | 31 |
| 1.5. AI：AI 应用需求爆发式增长，带来算力整体短缺 | 35 |
| 1.5.1. 头部模型概况 | 35 |
| 1.5.2. 模型调用及原生应用 | 37 |
| 1.5.3. 模型厂商业绩 | 40 |
| 1.5.4. 涨价情况：AI 应用需求爆发式增长，拉动算力价格持续上涨 | 43 |
| 1.5.5. 科技企业裁员：2026 年裁员规模明显提升 | 45 |
| 二、 河南计算机行业动态 | 45 |
| 1.6. 河南计算机行业数据跟踪 | 45 |
| 1.7. 河南上市公司行情回顾 | 46 |
| 2. 计算机行业行情表现 | 46 |
| 2.1. 行情回顾：国际局势影响下，行业剧烈调整后，4 月步入反弹走势 | 46 |
| 2.1. 估值：行业的估值超过历史均值水平 | 47 |
| 2.2. 计算机行业 ETF 行情回顾 | 47 |
| 2.3. 计算机行业可转债行情回顾 | 49 |
| 3. 行业观点与投资建议 | 50 |
| 4. 风险提示 | 51 |

图表目录

| | |
|-------------------------------------|---|
| 图 1：2019-2026 年我国软件业务收入及增速（月度累计值） | 5 |
| 图 2：2019-2026 年我国软件业务利润总额及增速（月度累计值） | 5 |
| 图 3：2019-2026 年我国软件业务出口数据 | 6 |
| 图 4：2019-2026 集成电路设计服务收入及增速（亿元） | 6 |
| 图 5：2020-2026 云计算+大数据服务收入及增速（亿元） | 6 |
| 图 6：2022-2026 基础软件收入及增速（亿元） | 7 |
| 图 7：2019-2026 工业软件收入及增速（亿元） | 7 |
| 图 8：2019-2026 电子商务平台技术服务收入及增速（亿元） | 7 |
| 图 9：2019-2026 信息安全收入及增速（亿元） | 8 |
| 图 10：2019-2026 嵌入式系统软件收入及增速（亿元） | 8 |
| 图 11：2026 与 2025 年 1-2 月软件子行业增速对比 | 8 |
| 图 12：2026 年 1-2 月我国软件业务收入结构 | 8 |

| | |
|---|----|
| 图 13: 2026 年 1-2 月信息技术服务收入结构 | 8 |
| 图 14: 2026 年 1-2 月软件产品收入结构 | 8 |
| 图 15: 2022-2026 我国进出口集成电路单价 (元/个) | 9 |
| 图 16: 2022-2026 我国出口/进口集成电路比值 | 9 |
| 图 17: 2024Q3-2025Q4 英伟达收入的区域分布情况 (亿美元) | 10 |
| 图 18: 2023-2025 我国 AI 芯片出货情况及国产化比率 | 11 |
| 图 19: 2025 年我国加速芯片市场份额 | 11 |
| 图 20: 2017-2026Q1 国内主要上市 AI 芯片企业收入及增速对比 | 12 |
| 图 21: Atlas 350 加速卡参数 | 13 |
| 图 22: 2022Q2-2025Q4 中国手机操作系统市场份额 | 15 |
| 图 23: 2022Q2-2025Q4 全球手机操作系统市场份额 | 15 |
| 图 24: 2022Q1-2025Q4 中国手机市场厂商份额 | 16 |
| 图 25: 2021Q1-2025Q4 中国平板电脑市场厂商份额 | 16 |
| 图 26: 英伟达 GPU 芯片核心指标变化 | 16 |
| 图 27: Vera Rubin NVL144 (2025 年旧计划) | 17 |
| 图 28: Vera Rubin NVL72 (2026 年实际发布产品) | 17 |
| 图 29: 包含 9,216 颗 Ironwood 的 TPU 集群 | 18 |
| 图 30: 2023Q4-2025Q4 博通 AI 业务收入 (亿美元) | 20 |
| 图 31: 23Q1-25Q4 Marvell 数据中心业务收入 (亿美元) | 20 |
| 图 32: 平头哥算卡与英伟达特供芯片的性能对比 | 21 |
| 图 33: 平头哥规模部署情况 | 21 |
| 图 34: 2017-2025 全球服务器市场规模 (亿美元) | 22 |
| 图 35: 2025Q4 全球不同类型服务器 (亿美元) | 22 |
| 图 36: 2025Q4 全球不同区域服务器收入增速 | 22 |
| 图 37: 2025Q4 全球服务器市场头部厂商收入 (亿美元) | 22 |
| 图 38: 2018-2025 中国服务器市场规模 (亿美元) | 23 |
| 图 39: 中国加速服务器增长预测 (亿美元) | 23 |
| 图 40: 2025 中国服务器市场规模 (亿美元) | 23 |
| 图 41: 2025 年中国 X86 服务器厂商市场份额 | 23 |
| 图 42: 2021Q2-2025Q4 紫光和浪潮存货增长情况 (亿元) | 24 |
| 图 43: scaleX640 超节点 | 25 |
| 图 44: 曙光数创浸没式液冷的技术进展 | 26 |
| 图 45: 不同冷却技术与 PUE 相关性 | 26 |
| 图 46: GPU 和 CPU 功耗提升趋势 | 27 |
| 图 47: 2017-2025 我国智能算力和在用数据中心机架数规模 | 27 |
| 图 48: 2019Q1-2025Q4 海外云巨头的相关业务收入及增速 | 29 |
| 图 49: 甲骨文云基础设施收入及增速 (亿美元) | 29 |
| 图 50: 甲骨文云基础设施业务年度收入预测 (亿美元) | 29 |
| 图 51: 2016-2025 年我国及全球云基础设施服务的支出规模及增速 (亿美元) | 30 |
| 图 52: 2025Q3 我国云基础设施服务市场结构 | 30 |
| 图 53: 2019Q4-2025Q4 阿里云收入及其增速 | 31 |
| 图 54: 2017Q1-2025Q4 金山云收入及其增速 | 31 |
| 图 55: 2025H1 中国 AI 云市场格局 | 31 |
| 图 56: 2024-2030 年 AI 云市场结构及规模 | 31 |
| 图 57: 4 大科技厂商资本投入 (亿美元) | 32 |
| 图 58: 4 大科技厂商资本投入增速 | 32 |

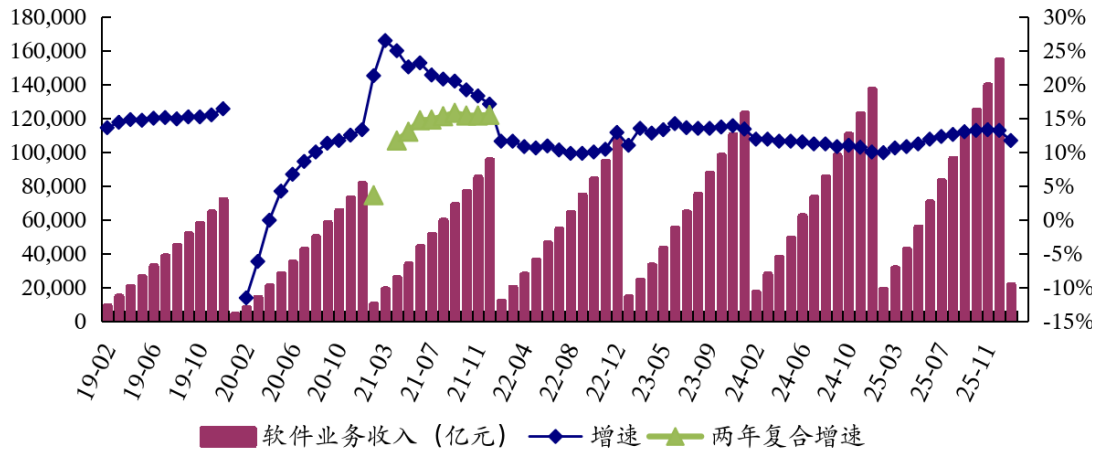
| | |
|---|----|
| 图 59: 2020-2025 年阿里单季度资本开支及增速 | 35 |
| 图 60: 2020-2025 年腾讯单季度资本开支及增速 (亿元) | 35 |
| 图 61: 大模型智能水平 (截至 2026.4.23) | 36 |
| 图 62: 大模型智能水平随时间增长的情况截至 (2026.4.23) | 36 |
| 图 63: 大模型价格对比 (2026.4.23) | 37 |
| 图 64: 大模型价格与智能的综合对比 (2026.4.23) | 37 |
| 图 65: OpenRouter 中模型的每周使用情况 (截至 4 月 23 日) | 38 |
| 图 66: 2025 年 12 月 AI 原生 APP 月活跃用户规模 (亿) | 39 |
| 图 67: 豆包大模型调用量增长情况 (万亿) | 39 |
| 图 68: OpenAI 的年化收入和算力增长情况 | 40 |
| 图 69: 2022-2025 智谱与 MiniMax 收入及增速 (亿元) | 42 |
| 图 70: 2022-2025 智谱与 MiniMax 研发费用 (亿元) | 42 |
| 图 71: 2025 年 MiniMax 分区域收入结构 | 42 |
| 图 72: 2025 年 MiniMax 分区域收入结构 | 42 |
| 图 73: 2025 智谱分产品毛利率 | 43 |
| 图 74: 2025 智谱分产品收入结构 | 43 |
| 图 75: H100 GPU 租赁价格趋势 (1 年期合同) | 44 |
| 图 76: 河南上市公司近期股价涨跌幅表现 (%) | 46 |
| 图 77: 2026 年 3 月中信一级子行业涨跌幅 | 47 |
| 图 78: 近 10 年中信计算机行业估值水平 (截至 2026.4.21) | 47 |
| | |
| 表 1: 2025 年 AI 芯片相关动态 | 10 |
| 表 2: 华为昇腾系列芯片发布计划及芯片性能 (2025.9) | 12 |
| 表 3: 华为超节点上市计划及参数 (2025.9) | 13 |
| 表 4: 鸿蒙纯血系统发展动态 | 14 |
| 表 5: 最新发布 AI 芯片及参数 | 18 |
| 表 6: 2025 年企业服务器相关业务对比 (亿元) | 24 |
| 表 7: 主要厂商的超节点产品上市情况 | 25 |
| 表 8: 2025 年企业数据中心相关业务对比 (亿元) | 28 |
| 表 9: 海外科技厂商算力相关投入 | 33 |
| 表 10: AI 应用用户数增长情况 | 38 |
| 表 11: 大模型厂商融资及估值情况 | 41 |
| 表 12: 大模型厂商融资及估值情况 | 43 |
| 表 13: 模型厂商和云计算涨价情况 | 44 |
| 表 14: 海外主要科技厂商近期裁员情况 | 45 |
| 表 15: 2026 年 3 月计算机行业 ETF 行情回顾 | 48 |
| 表 16: 2026 年 3 月计算机行业可转债行情回顾 | 49 |
| 表 17: 行业重点企业的 IPO 推进情况 | 50 |

一、行业数据

1.1. 行业数据：2026 年 1-2 月软件行业收入增速回落，但是出口加速趋势明显

2026 年软件产业增速略有回落。根据工信部数据，2026 年 1-2 月软件业务收入 2.15 万亿元，同比增长 11.7%，较 2025 年回落了 1.5 PCT。

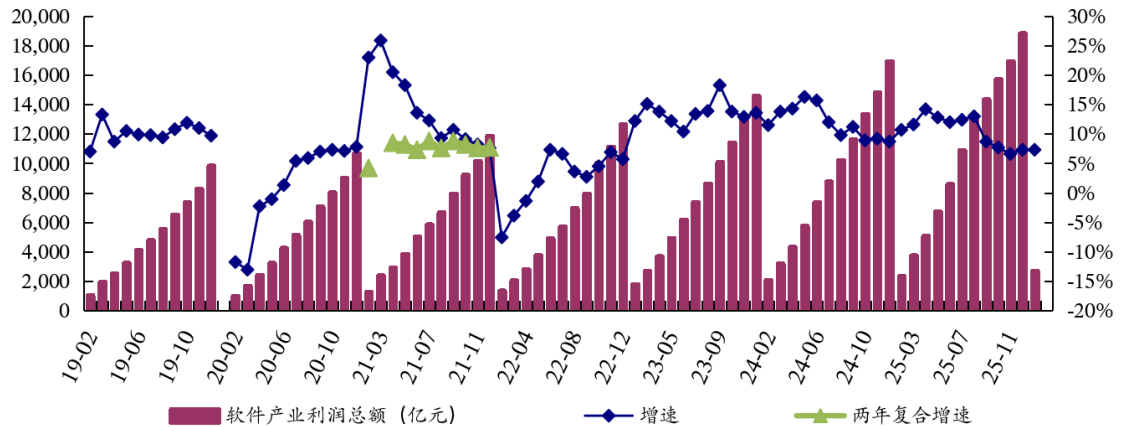
图 1：2019-2026 年我国软件业务收入及增速（月度累计值）



资料来源：工信部，中原证券研究所

利润方面，2026 年软件行业利润增速仍处于相对低位的水平。2026 年 1-2 月软件业务利润总额 2693 亿元，同比增长 7.3%，与 2025 年持平，低于收入增速 4.4 PCT。

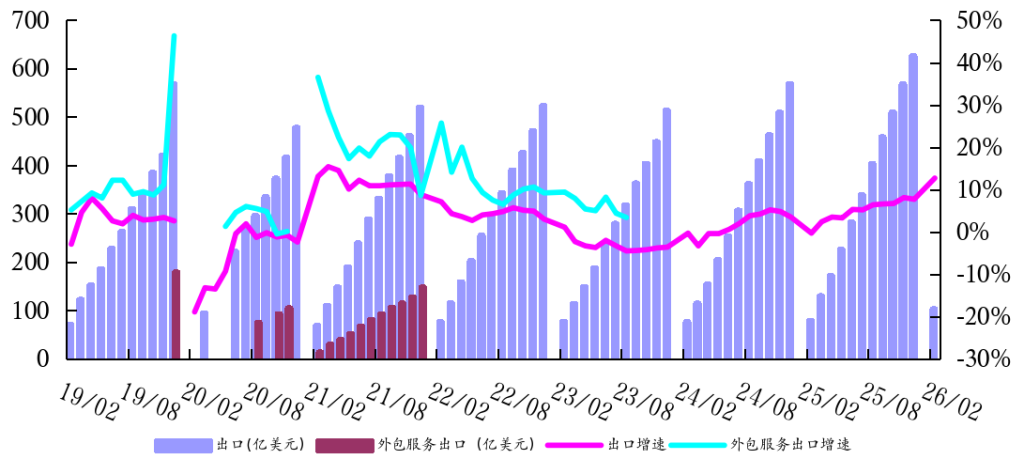
图 2：2019-2026 年我国软件业务利润总额及增速（月度累计值）



资料来源：工信部，中原证券研究所

2026 软件出口加速趋势明显。2026 年 1-2 月软件业务出口金额 103.8 亿美元，同比增长 12.7%，较 2025 年增长 5.0 PCT，软件行业出口金额约占行业收入的 3.5%。

图 3：2019-2026 年我国软件业务出口数据



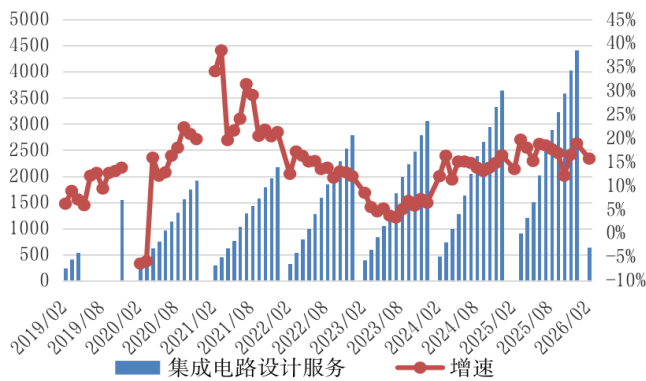
资料来源：工信部，中原证券研究所

1.2. 2026 年 1-2 月高景气赛道：IC 设计、嵌入式软件

IC 设计：2026 年 1-2 月同比增长 15.7%，较 2025 年下降了 3.2 PCT，高于软件行业整体增速 4.0 PCT，仍然保持了较高的景气度，2026 年有望继续受益于芯片产业的高景气度。

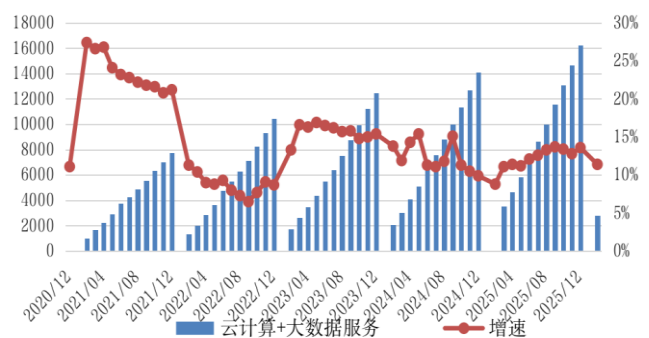
云计算+大数据服务：2026 年 1-2 月同比增长 11.4%，较 2025 年下降了 2.2 PCT，低于软件行业整体增速 03 PCT，2026 年预计仍将受益于 AI 应用快速发展趋势。

图 4：2019-2026 集成电路设计服务收入及增速（亿元）



资料来源：工信部，中原证券研究所

图 5：2020-2026 云计算+大数据服务收入及增速（亿元）

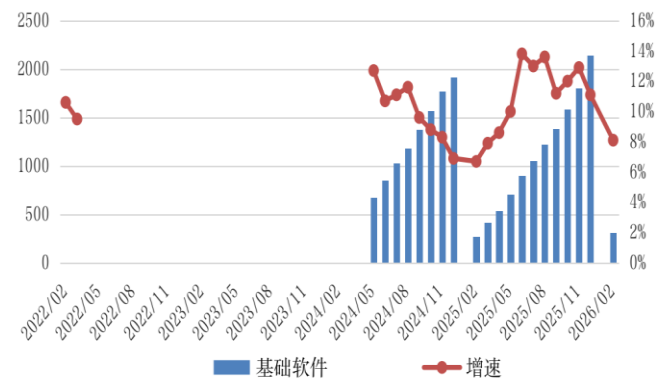


资料来源：工信部，中原证券研究所

基础软件：2026 年 1-2 月同比增长 8.1%，较 2025 年下降了 3.0 PCT，低于软件行业整体增速 3.6 PCT。

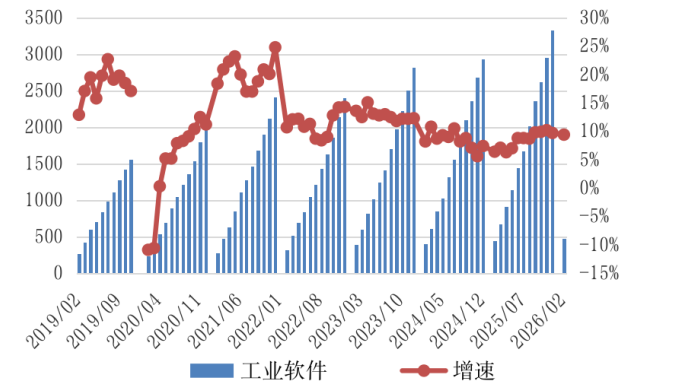
工业软件产品：2026 年 1-2 月同比增长 9.4%，较 2025 年下降了 0.3 PCT，保持相对稳定的增速水平。

图 6：2022-2026 基础软件收入及增速（亿元）



资料来源：工信部，中原证券研究所

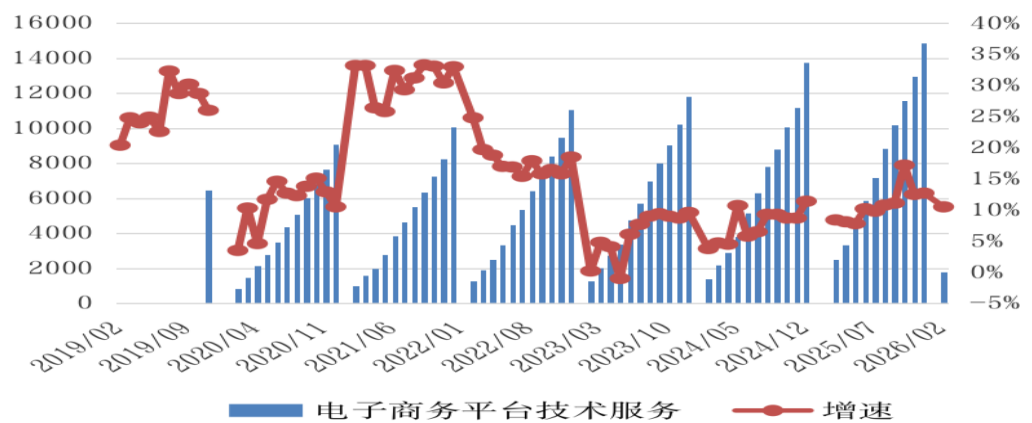
图 7：2019-2026 工业软件收入及增速（亿元）



资料来源：工信部，中原证券研究所

电子商务平台技术服务：2026 年 1-2 月同比增长 10.5%，较 2025 年下降了 2.2 PCT，低于软件行业整体增速 1.2 PCT。

图 8：2019-2026 电子商务平台技术服务收入及增速（亿元）

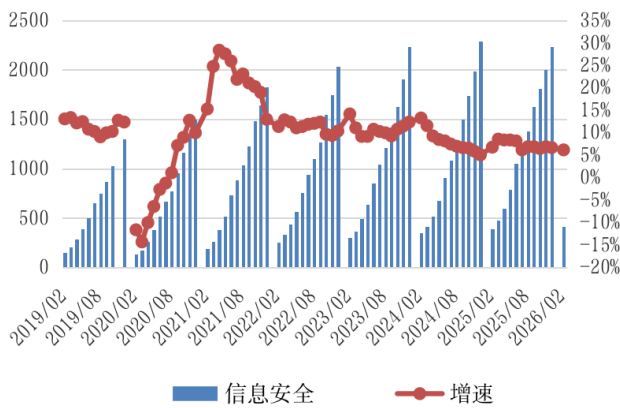


资料来源：工信部，中原证券研究所

信息安全：2026 年 1-2 月同比增长 6.2%，较 2025 年回落 0.5 PCT，仍然低于行业 5.5 PCT，处于相对地位水平。

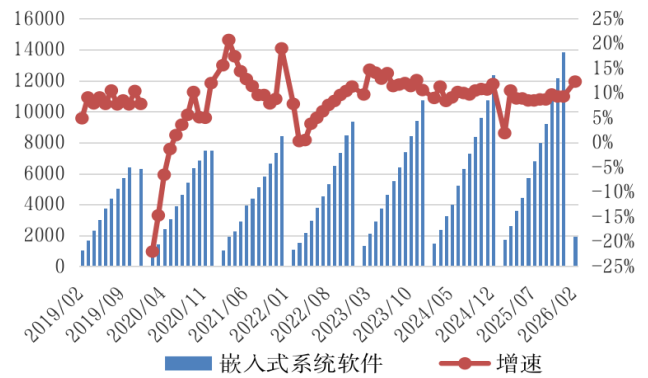
嵌入式系统软件：2026 年 1-2 月同比增长 12.3%，较 2025 年提升 3.0 PCT，处于相对高位水平，同时高于行业平均增速 0.6 PCT。

图 9：2019-2026 信息安全收入及增速（亿元）



资料来源：工信部，中原证券研究所

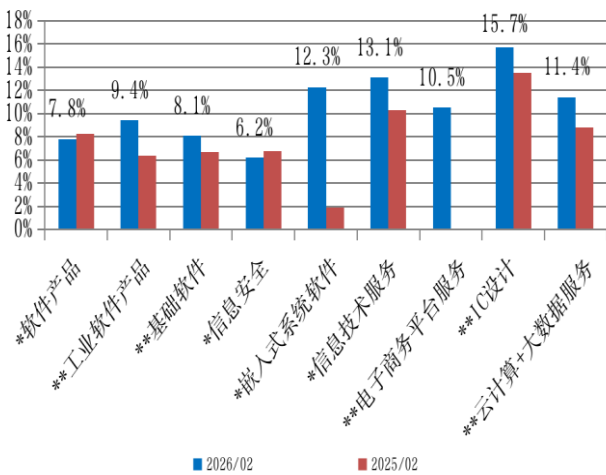
图 10：2019-2026 嵌入式系统软件收入及增速（亿元）



资料来源：工信部，中原证券研究所

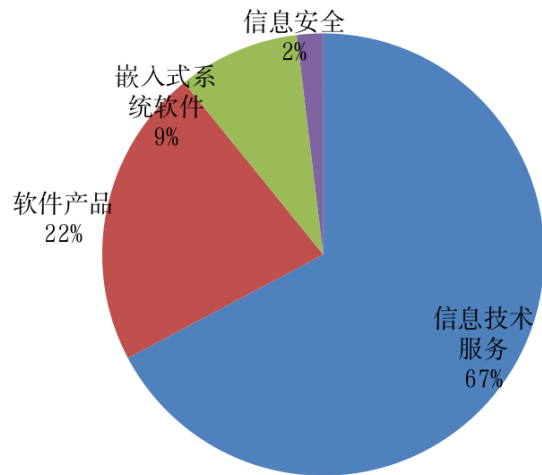
软件业务中，信息技术服务收入占比持续提升。2026 年 1-2 月信息技术服务收入增速 13.1%，高于软件业务整体增速 1.4 PCT，占比软件业务整体收入比重提升至 67.2%。

图 11：2026 与 2025 年 1-2 月软件子行业增速对比



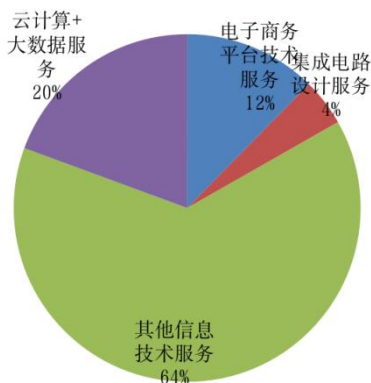
资料来源：工信部，中原证券研究所（*为一级子行业，**为二级子行业）

图 12：2026 年 1-2 月我国软件业务收入结构



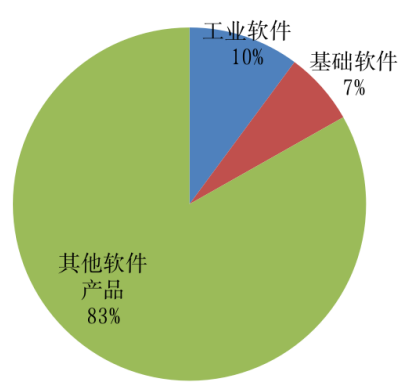
资料来源：工信部，中原证券研究所

图 13：2026 年 1-2 月信息技术服务收入结构



资料来源：工信部，中原证券研究所

图 14：2026 年 1-2 月软件产品收入结构



资料来源：工信部，中原证券研究所

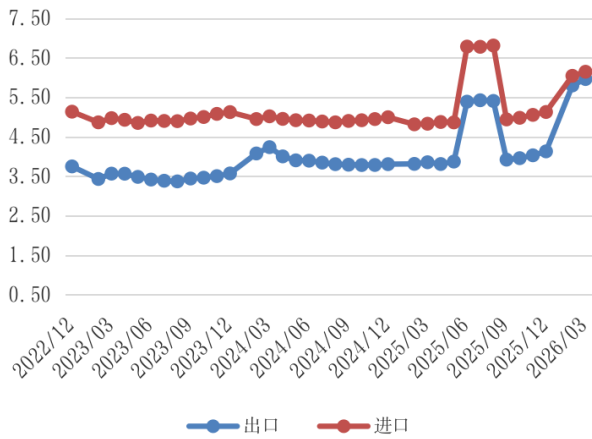
1.3. 国产化：2025 年下半年 AI 芯片国产化率进一步提升到了 46%

1.3.1. 国产芯片

1.3.1.1. 国内芯片产品性能持续提升，2026 年呈现量价齐升趋势

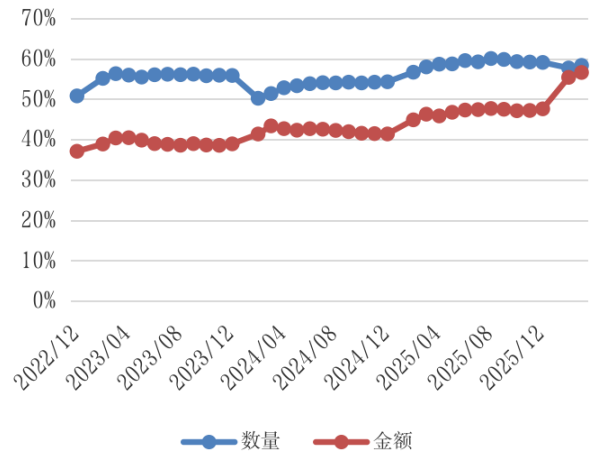
2026 年前三个月份，集成电路进出口整体呈现量价提升的局面。进口数量和金额分别同比增长 11%和 41%，出口数量和金额分别同比增长 13%和 73%。其中出口集成电路平均单价较进口电路平均单价价差从上一年 0.97 元/个缩小到 0.18 元/个。同时将我国出口和进口集成电路对比来看，出口集成电路数量占比进口数量上总体稳定，但是金额占比方面逐年提升，已经由 2022 年的 37.1%提升到了 2026 年前两个月份的 56.6%，这表明我国的集成电路产品能力在持续提升。

图 15：2022-2026 我国进出口集成电路单价（元/个）



资料来源：海关总署，中原证券研究所

图 16：2022-2026 我国出口/进口集成电路比值

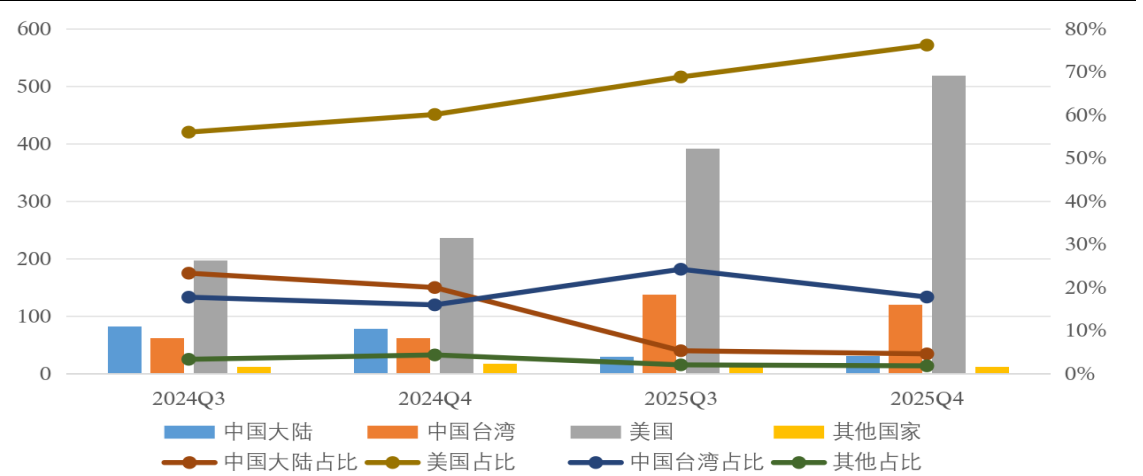


资料来源：海关总署，中原证券研究所

1.3.1.2. 2025 年 AI 芯片禁运，英伟达对中国客户的销售额下滑 60%

H20 禁运后的两个季度中，英伟达来自中国大陆公司收入下滑 60%以上。2025Q3 开始，英伟达对分区域的收入划分口径进行了修改，从客户账单所在地改为了总部所在地，在口径修改以后，2024Q3、Q4 英伟达来自中国大陆的收入分别增长了 27.26 亿美元和 22.84 亿美元。在 H20 禁令出台以后，其在 2025Q3 和 Q4 来自中国大陆的收入分别降低到了 29.73 亿美元和 30.60 亿美元，同比下滑 63%和 61%，即使全部当做中国大陆公司在海外下单贡献的收入，也仅有 9%和 34%的同比涨幅。

图 17: 2024Q3-2025Q4 英伟达收入的区域分布情况 (亿美元)



资料来源: 英伟达, 中原证券研究所 (英伟达财报时段较以上周期晚一个月)

从最新的中美政策来看, 英伟达 H200 目前仍受到较多的政策限制, 处于限制采购和强监管的状态, AMD 的 MI308 率先实现了少量的供货, 继续给国产厂商留下了较多的发展机会。

表 1: 2025 年 AI 芯片相关动态

| 时间 | 内容 |
|--------------|--|
| 2025. 3. 25 | 美国商务部将 54 家中国实体列入清单, 包括了浪潮集团旗下 6 家公司、宁畅信息、中科可控旗下的服务器品牌 Suma |
| 2025. 4. 16 | 英伟达表示美国政府已经于 4 月 14 日通知公司, H20 芯片未来在出口至中国时需要“无限期”申请许可证 |
| 2025. 6. 3 | 报道称英伟达将为中国研发 B30 芯片, 其采用 Blackwell 架构, GDDR7 显存, 而非 HBM, 也不会采用台积电的先进封装技术, 首度支持多 GPU 扩展, 允许用户通过连接多组芯片来打造更高性能的计算集群。采用 GB20X 芯片, 也就是 RTX 50 系列的芯片, 其售价预计在 6500 美元至 8000 美元之间, 远低于 H20 芯片的 1 万至 1.2 万美元。 |
| 2025. 7. 15 | 英伟达宣布美国已批准 H20 芯片销往中国 |
| 2025. 7. 31 | 国家网信办就 H20 算力芯片后门安全风险约谈英伟达公司。 |
| 2025. 8. 1 | 美国商务部 BIS 因人手短缺等因素陷入运作混乱, 几乎处于瘫痪状态, 包括英伟达拟对华出口的大批 H20 芯片在内的数千份许可申请被搁置。截至 7 月底尚无任何许可发放, 涉及订单金额达数十亿美元。 |
| 2025. 8. 6 | 英伟达声明其芯片中不存在后门、终止开关和监控软件。 |
| 2025. 8. 11 | 特朗普对外证实, 他已要求向 AI 芯片大厂英伟达、AMD 对其销往中国大陆的芯片征收营收的 15% 作为许可费。 |
| 2025. 8. 13 | 路透社报道, 美国秘密在人工智能相关芯片中设置跟踪器, 以发现“芯片被转运至中国的情况”。 |
| 2025. 8. 18 | 英伟达计划将特朗普额外收取的 15% 费用所带来的成本转嫁给中国客户。 |
| 2025. 8. 19 | 路透社报道, 英伟达正在为中国市场开发暂名 B30A 的 Blackwell 架构芯片, 其原始计算能力可能仅为 B300 一半, 预计将配备高带宽内存以及英伟达的 NVLink 技术, 以实现处理器之间的高速数据传输, 预计最快于 9 月向中国客户提供测试样品。 |
| 2025. 9. 15 | 英伟达违反反垄断法, 市场监管总局依法决定实施进一步调查。 |
| 2025. 11. 6 | The Information 报道称, 白宫已通知其他联邦机构, 不会允许英伟达向中国出售其最新的减配版人工智能 (AI) 芯片。 |
| 2025. 11. 21 | 路透社消息, 特朗普政府正在考虑批准英伟达 H200 人工智能 (AI) 芯片对中国的销售。 |
| 2025. 12. 8 | 美国总统特朗普示, 美国将允许英伟达向中国“经批准的客户”出售 H200 人工智能芯片。 |
| 2025. 12. 23 | 快科技报道, 阿里巴巴正考虑向 AMD 下单订购 4-5 万颗 MI308 AI 芯片。 |
| 2026. 1. 17 | 英伟达供应商因中方禁令暂停 H200 芯片关键组件的生产。 |
| 2026. 1. 28 | 路透社报道称, 字节跳动、阿里巴巴和腾讯三家企业合计获准采购逾 40 万颗 H200 芯片, 但批复附带一定条件, 相关要求仍在制定中。 |

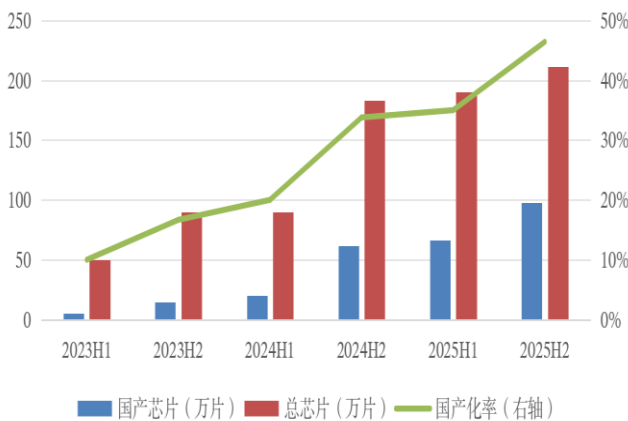
| | |
|-------------|---|
| 2026. 1. 29 | 英伟达黄仁勋表示，他过去几天在访问中会见了客户和官员，但没有收到任何 H200 芯片的新订单。 |
| 2026. 2. 3 | AMD 财报披露 4 季度 MI308 出货到中国的收入约为 3.9 亿美元，预计 2026Q1 约有 1 亿美元的 MI308 销往中国。 |
| 2026. 2. 3 | 美国政府对字节开出了 H200 一人一策的准入清单：若希望继续执行 3.2 万颗 H200 订单，须缴纳 30% 关税（约 2.8 亿美元）、接受美方技术审计并签署“非军用、非超算”最终用户承诺书。 |
| 2026. 3. 5 | 英伟达已正式停止生产专为中国市场定制的 H200 人工智能芯片。 |
| 2026. 3. 17 | 英伟达 CEO 黄仁勋在 GTC 2026 全球开发者大会上亲口确认，公司已收到来自中国客户的 H200 采购订单，正重启相关生产，供应链已启动运转。 |

资料来源：快科技，和讯网，新京报，观察者网，俄罗斯卫星通讯社，C114，中原证券研究所

1.3.1.3. 2025 年下半年 AI 芯片国产化率进一步提升到了 46%，寒武纪也实现了盈利

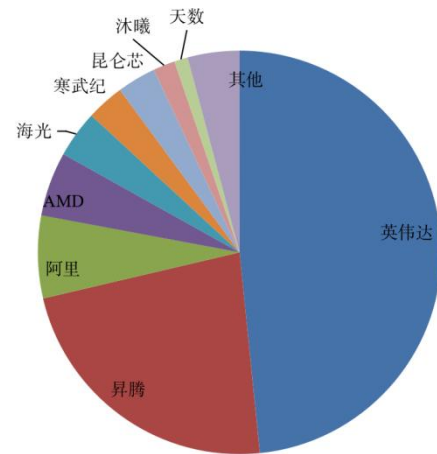
2026 年 AI 芯片国产替代趋势有望继续加快。根据 IDC 数据，2025 年下半年，我国 AI 芯片国产化比率从 2025 上半年的 35% 提升到了 46%，提升趋势明显。考虑到英伟达芯片对国内实质性断供，同时国产芯片在性能和产能方面的提升，我们预计 2026 年国产 AI 芯片的占比将获得较大比例的提升。

图 18：2023-2025 我国 AI 芯片出货情况及国产化比率



资料来源：IDC，中原证券研究所

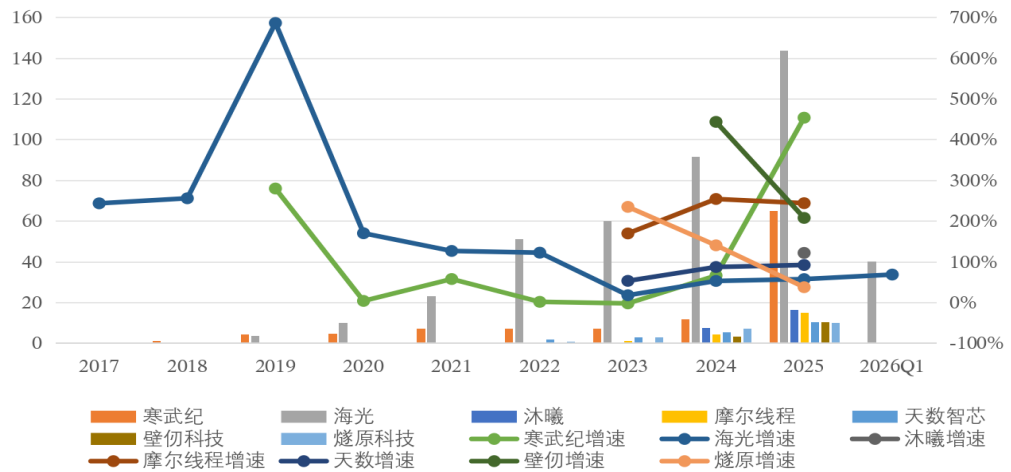
图 19：2025 年我国加速芯片市场份额



资料来源：IDC，中原证券研究所

从年报数据来看，继 2024 年的收入高增长以后，国产 AI 芯片厂商 2025 年业务景气度持续提升，其中 2 家实现盈利。目前 6 家 AI 芯片厂商都已经公布了年报或业绩快报，其中 2025 年业绩增速最快的为寒武纪，实现收入 64.97 亿元，同比增长 453%，同时寒武纪 2025 年首次实现盈利，归母净利润 20.59 亿元，毛利率和净利率分别为 55% 和 32%；海光信息的 AI 芯片深算三号投入市场，并且收入增速从 2025Q2 的 45% 逐季增长到了 2026Q1 的 68%，2025 年收入和净利润规模分别为 144 亿元和 25 亿元。其余有 4 家为规模都在 10-20 亿之间，增速分别为沐曦 121%、摩尔线程 243%、壁仞 207%、天数 92%。燧原科技规模 9.9 亿，增速 37%，增长动力相对较弱。

图 20: 2017-2026Q1 国内主要上市 AI 芯片企业收入及增速对比



资料来源: 上市公司财报, 中原证券研究所

1.3.1.4. 2026 年国产 AI 芯片在加速发展的过程中, 有望开启向海外市场的拓展

蚂蚁集团 3 月 11 日发表的论文显示, 其采用阿里巴巴和华为芯片在内的国产芯片与 MoE 技术, 成功将大模型训练成本降低 20%, 性能媲美英伟达 H800 方案。

阿里: 阿里巴巴在 3 月的财报电话会议上表示, 平头哥自研 AI 芯片已经实现规模化量产, 截至 2026 年 2 月, 已经累计规模化交付 47 万片。对比 IDC 阿里国内出货 26.5 万片数据来看, 还有 20 万片差额, 同时平头哥实现了年化过百亿的营收规模。这一交付规模意味着阿里正在成为国内对标谷歌的, 同时具有较强自研芯片能力、云服务能力、AI 模型能力的厂商。

华为: 3 月华为发布了搭载全新昇腾 950PR (Ascend 950PR) 处理器的 AI 训练推理加速卡 Atlas 350。与 910C 相比, 950PR 实现了在低精度数据格式、向量算力、互联带宽及自研 HBM 等方面实现大幅提升。

表 2: 华为昇腾系列芯片发布计划及芯片性能 (2025.9)

| | 昇腾910C | 昇腾950PR | 昇腾950DT | 昇腾960 | 昇腾970 |
|------|------------------------------|--|-----------------------------|---|---|
| 时间 | 2025Q1 | 2026Q1 | 2026Q4 | 2027Q4 | 2028Q4 |
| 微架构 | SIMD | SIMD/SIMT | | SIMD/SIMT | SIMD/SIMT |
| 数值类型 | FP32/HF32/FP16/BF16/ INT8 | FP32/HF32/FP16/BF16/ FP8/MXFP8/HiF8/MXFP4 | | FP32/HF32/FP16/BF16/FP8/ MXFP8/HiF8/MXFP4/HiF4 | FP32/HF32/FP16/BF16/FP8/ MXFP8/HiF8/MXFP4/HiF4 |
| 互联带宽 | 784GB/s | 2TB/s | | 2.2TB/s | 4TB/s |
| 算力 | 800TFLOPS FP16 | 1PFLOPS FP8, 2PFLOPS FP4 | | 2 PFLOPS FP8, 4 PFLOPS FP4 | 4 PFLOPS FP8, 8 PFLOPS FP4 |
| 内存 | 128GB, 3.2TB/s | 128GB, 1.6TB/s 自研 HiBL 1.0 | 144GB, 4TB/s 自研 HiZQ 2.0 | 288GB, 9.6TB/s | 288GB, 14.4TB/s |

资料来源: 华为, 中原证券研究所

根据 3 月公布的数据, 搭载了 950PR 的 Atlas 350 加速卡, 算力为 1.56P FP4 精度和 800T mxFP8 精度, 是 H20 的 2.87 倍, 是目前国内唯一支持 FP4 低精度的推理产品; 内存 112GB,

是 H20 的 1.12 倍，多模态生成速度可以提升 60%；带宽到了 1.4TB/s；功耗为 600W，是 H20 的 1.5 倍。同时在 950PR 发布的同时，昆仑、华鲲振宇、神州鲲泰、长江计算、宝德、软通华方、百信 7 家服务器厂商也在会上发布了该加速卡的服务器产品。

图 21: Atlas 350 加速卡参数



资料来源：华为，上海证券报，中原证券研究所

在出货方面，华为昇腾 384 超节点到 2025 年 9 月累计出货量超过 300 套，2026 年 1 月已经累计出货超 550 套，广泛应用于互联网大模型训练与推理、金融实时反欺诈、能源调度、港口智能排泊、智能制造质检等场景，在部分场景中实现模型训练 MFU（模型算力利用率）提升 30%~45%，或推理效率数量级增长。根据 2025 年 12 月中国移动招标信息来看，昇腾 384 超节点售价 1.35 亿元，这意味着超节点业务在 2025 年华为 8809 亿元的收入中已经占据了可观的比例。

根据 2026 年 3 月的报道，阿里巴巴、字节跳动、腾讯等科技巨头已提前下单华为新一代芯片，订单规模达数十万颗。

按照此前的计划，华为还将在 2026Q4 发布昇腾 950DT，在内存上实现较大的提升，在 2027Q4 发布昇腾 960，2028Q4 发布昇腾 970，将国产芯片性能给带来持续提升的空间。

2026 年 3 月，华为在西班牙世界移动通信大会上展示了 Atlas 950 SuperPoD 以及业界首款通算超节点 TaiShan 950 SuperPoD 超节点，Atlas 950 最大支持 8192 卡通过灵衢互联，具备超大带宽、超低时延和内存统一编址等关键特点。作为业界最大的超节点，Atlas 950 原计划在 2026 年下半年上市，有望加快华为超节点向海外市场的拓展。

表 3: 华为超节点上市计划及参数 (2025.9)

| | Atlas 900 | Atlas 950 | Atlas 960 |
|------|-----------|-----------|-----------|
| 上市时间 | 2025.3 | 2026Q4 | 2027Q4 |

| 昇腾卡 | 昇腾 910C | 昇腾 950DT | 昇腾 960 |
|-------|-----------------|-------------------------------|--------------------------------|
| 昇腾卡数量 | 384 | 8192 | 15488 |
| 计算机柜 | 12 | 128 | 176 |
| 互联机柜 | 4 | 32 | 44 |
| 占地面积 | | 1000 平方米 | 2200 平方米 |
| 算力 | 300 PFLOPS BF16 | 8 EFLOPS FP8 16 EFLOPS FP4 | 30 EFLOPS FP8 60 EFLOPS FP4 |
| 互联带宽 | | 16.3PB/s | 34PB/s |
| 内存容量 | | 1152TB | 4460TB |
| 训练性能 | | 4.91M TPS | 15.9M TPS |
| 推理性能 | | 19.6M TPS | 80.5M TPS |
| 互联协议 | 灵衢 1.0 | 灵衢 2.0 | |

资料来源：华为，中原证券研究所

1.3.2. 纯血鸿蒙系统终端数量突破 5000 万

鸿蒙用户突破 5000 万，年内有望更新到 7.0 版本。2024 年 10 月鸿蒙 5.0 正式发布并在手机端应用，2025 年 10 月 22 日升级到鸿蒙 6.0，2026 年 3 月鸿蒙 5 以上操作系统终端数量突破 5000 万，同时 6 月有望升级到鸿蒙 7。

表 4：鸿蒙纯血系统发展动态

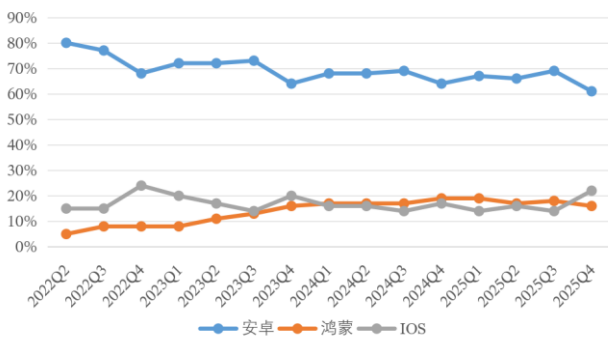
| 类型 | 时间 | 事件 |
|-------------|---|--|
| 系统发布 | 2024. 10. 22 | 纯血鸿蒙 5.0 (HarmonyOS NEXT) 正式发布，彻底摆脱安卓，并在手机端应用 |
| | 2025. 5. 19 | 首款鸿蒙电脑搭载鸿蒙 5.0 操作系统正式发布 |
| | 2025. 6. 20 | 正式发布 HarmonyOS 6，并同步启动开发者 Beta 测试。 |
| | 2025. 9. 27 | 鸿蒙 6.0 Release 版本正式发布 |
| | 2025. 10. 22 | HarmonyOS 6 正式发布 |
| | 2026. 6 | 鸿蒙 7 有望发布。 |
| 生态建设 | 2025. 5. 8 | 鸿蒙电脑有 300 多款应用完成适配 |
| | 2025. 5. 19 | 鸿蒙电脑有 1000 多款应用完成适配 |
| | 2025. 6. 6 | 鸿蒙电脑生态应用数量已经突破 2000 款 |
| | 2025. 6. 20 | 预计到 6 月底鸿蒙电脑生态应用突破 2500 款，超过 3 万个鸿蒙应用和元服务在全速开发和积极更新状态中 |
| 用户量 | 2025. 7. 30 | 鸿蒙 5 操作系统的终端数量突破了 1000 万 |
| | 2025. 8. 25 | 鸿蒙 5 操作系统终端数量突破 1200 万 |
| | 2025. 9. 20 | 鸿蒙 5 操作系统终端数量突破 1700 万 |
| | 2025. 9. 29 | 鸿蒙 5 操作系统终端数量突破 2000 万 |
| | 2025. 10. 17 | 鸿蒙版微信安装量已突破 2100 万 |
| | 2025. 10. 22 | 鸿蒙 5 操作系统终端数量突破 2300 万 |
| | 2025. 12. 22 | 鸿蒙 5 以上操作系统终端数量突破 3200 万 |
| | 2025. 12. 30 | 鸿蒙 5 以上操作系统终端数量突破 3600 万 |
| 2026. 1. 25 | 华为应用市场 App Gallery 显示，小艺输入法安装量正式突破 4000 万次，这意味着 HarmonyOS 5/6 终端设备数也已突破 4000 万。 | |

2026.3.23 鸿蒙5以上操作系统终端数量突破5000万。

资料来源：每日新闻，IT之家，每日经济新闻，华为终端微信号，快科技，中原证券研究所

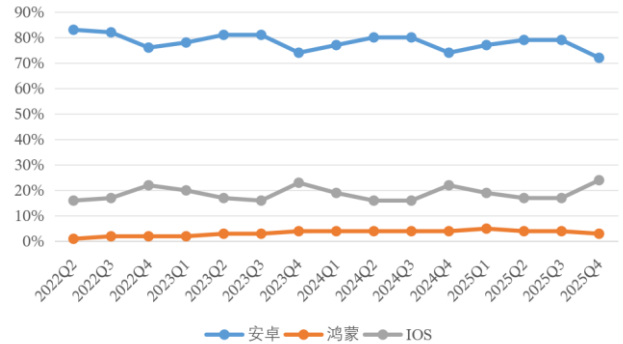
2025Q4 鸿蒙在手机操作系统市场份额下滑，在中国市场被 IOS 超越。根据 Counterpoint 的数据，2025Q4 国内手机操作系统市场上，鸿蒙份额 16%，较上个季度下降 2 PCT，同时因为苹果手机用户的大涨，安卓和鸿蒙系统份额都受到不同程度的影响，IOS 操作系统也在 Q4 份额大涨了 8 PCT，以 22%的份额超过了鸿蒙。全球市场来看，IOS 同样表现突出，份额上涨了 7 PCT，鸿蒙份额下滑 1 PCT 至 3%，继续位居全球第三。

图 22: 2022Q2-2025Q4 中国手机操作系统市场份额



资料来源：Counterpoint，中原证券研究所

图 23: 2022Q2-2025Q4 全球手机操作系统市场份额



资料来源：Counterpoint，中原证券研究所

与此前的操作系统不同，鸿蒙覆盖了桌面端、移动端、物联网终端，是一个面向万物智联时代所有终端的操作系统。鸿蒙可以实现一次开发、多端部署，可以有效降低 70%的终端适配成本。同时鸿蒙通过多端联动，实现应用的跨端流转和多端协同，将带来更好的用户使用体验。

纯血鸿蒙更加注重用户的隐私和信息安全，在安全访问机制方面与此前的操作系统有根本性区别。以安卓对比来看，此前应用借助用户访问权限的开放获取了更加广泛的相关信息，纯血鸿蒙直接取消了这其中 9 类不合理权限，并从管理权限变为了管理数据。

目前在鸿蒙 5 和鸿蒙 6 上，微信等已经适配的应用还在持续开发的过程中，部分功能缺失的问题影响了部分客户对于纯血鸿蒙系统的使用体验，进而影响了此前华为部分终端产品的销量。

(1) 手机：根据 IDC 数据，2025Q4 华为以 14.7%的出货量份额位居中国手机市场第 4，连续 3 个季度增速下滑，Q4 增速-10.3%低于行业-0.8%的增速水平。

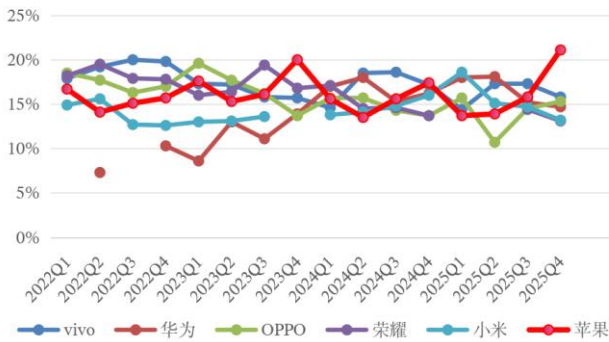
(2) 台式机和笔记本：2025Q4 华为位居国内台式机和笔记本市场第二，份额 11%，较上个季度增长 2PCT。

(3) 平板电脑：2025Q4 华为在国内平板电脑市场占据 26%的份额，位居第一，较 Q3 下降了 5PCT。

虽然鸿蒙 5 和鸿蒙 6 仍然处于发展阵痛期，但是随着用户量的增长、应用功能的完善、AI 能力赋能、芯片供应的改善和性能的提升，我们继续看好华为后续发展。从苹果系对客户的依附能力来看，随着纯血鸿蒙系统在易用性的持续突破，其独特的生态联动和系统联动能力也将

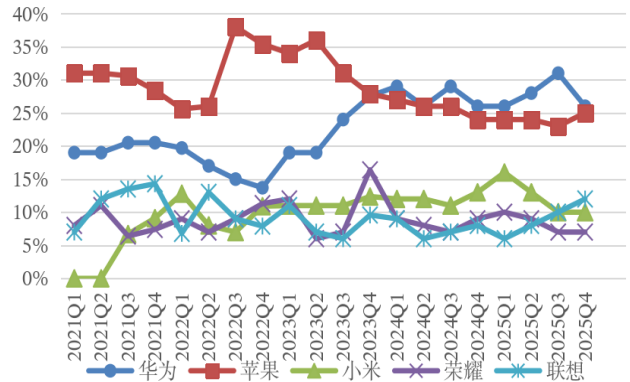
转化为强大的客户捆绑能力。

图 24：2022Q1-2025Q4 中国手机市场厂商份额



资料来源：IDC，中原证券研究所

图 25：2021Q1-2025Q4 中国平板电脑市场厂商份额



资料来源：Omdia，中原证券研究所

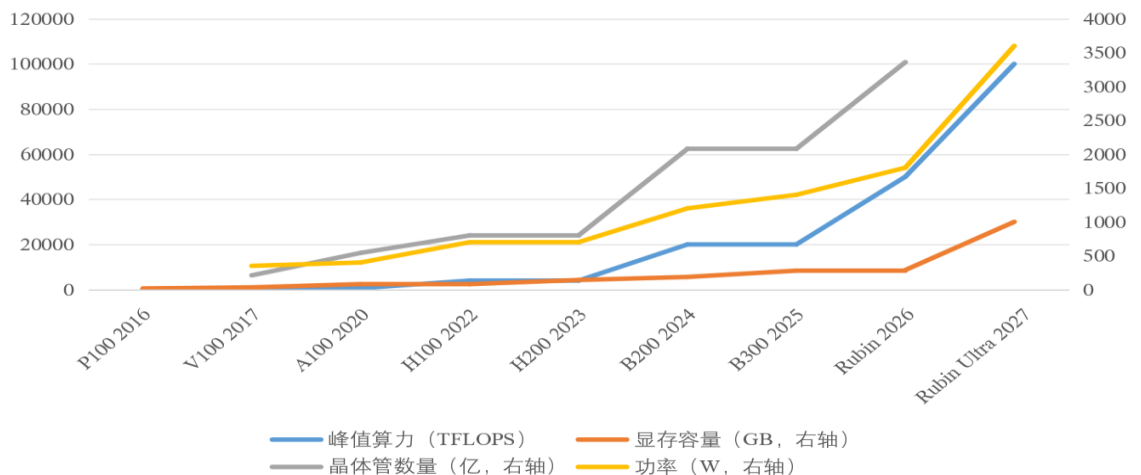
鸿蒙仍将是华为重点投入和发展的方向之一。2025 年初孟晚舟提出未来三年华为要与经济规律逆周期，加大战略纵深投入，错位发展，在根技术上压强式投入。她还表示鸿蒙生态处于量变到质变的关键历史节点，华为将围绕鸿蒙、鲲鹏、昇腾、云计算等业务，面向生态伙伴开发平台能力。在 2026 年新年致辞中，华为提到 2026 年主战场将包括人工智能行业应用、鲲鹏和昇腾、智联、鸿蒙、鸿蒙智行和乾崮智驾、重构 AI 数据中心、液冷超快充。

1.4. 算力：2025 年底我国智算规模较年中翻倍

1.4.1. 英伟达将在 2026 年下半年出货 Rubin，2027 年将采用的 4-Die 方案或遇阻

以 2025 年发布的 B300 芯片来看，英伟达当前量产的最新芯片峰值算力为 20 PFLOPS (FP4 Tensor Core, 稀疏)，功率 1400W，晶体管数量 2080 亿，显存为 HBM3e，容量 279GB，带宽 8TB/s。

图 26：英伟达 GPU 芯片核心指标变化



资料来源：英伟达，中原证券研究所 (Rubin Ultra 仍然按照 4die 进行封装数据估算)

2026 年下半年英伟达即将量产的 Rubin 架构芯片来看，峰值算力为推理 NVFP4 50 PFLOPS，训练 NVFP4 35PFLOPS，晶体管数量 3360 亿，显存为 HBM4，容量为 288GB，带宽 22TB/s 提升较为明显。

从英伟达的超节点来看，根据 2025 年发布的计划，2026 年英伟达将发布的是 144 个 GPU 的 Vera Rubin NVL144，2027 年将发布的是 576 个 GPU 的 Rubin Ultra NVL 576，但是 2026 年 1 月实际发布的是 72 个 GPU 的 Vera Rubin NVL72。虽然芯片数量缩减了一半，但是实际发布的超节点总体算力和 NVLink 带宽相当，HBM4 的显存容量缩减一半，显存带宽提升 37%。

后续来看，英伟达还将在 2027 年发布 Rubin Ultra 芯片，2028 年英伟达计划发布新架构 Feynman 的芯片。

图 27: Vera Rubin NVL144 (2025 年旧计划)



资料来源：英伟达，中原证券研究所

图 28: Vera Rubin NVL72 (2026 年实际发布产品)



资料来源：英伟达，中原证券研究所

随着降低精度 (FP4) 和工艺提升 (3nm Rubin, 1.6nm Feynman) 两条路径逐步遇到增长瓶颈，后续英伟达芯片性能增长速度有放缓的风险。为此我们看到 B200 开始，英伟达将 2 个 GPU 裸片 (die) 进行封装，而到了 Rubin Ultra 将 4 个 die 进行封装，但是根据 3 月 31 日 Trendforce 报道，Rubin Ultra 或将因为成本放弃 4-Die 方案，继续采用 2-Die 方案，或将影响 Rubin Ultra 芯片的各项性能指标的提升。同时，从 B200 开始，英伟达推出了由 2 个 GPU 和 1 个 CPU 共同组成的 GB200 芯片和超节点 GB200 NVL72，并开始更多展示在 Scale up 方向上的性能提升。

1.4.2. 自研芯片：正在模型厂商和云厂商成为算力供应的重要补充

为了降低在芯片端的支出和能耗需求，同时减少对英伟达的依赖，大型云厂商都在陆续开启芯片自研，并在 2025 年较好地实现对部分英伟达 GPU 需求的替代，2026 年行业格局有望进一步改变。

表 5：最新发布 AI 芯片及参数

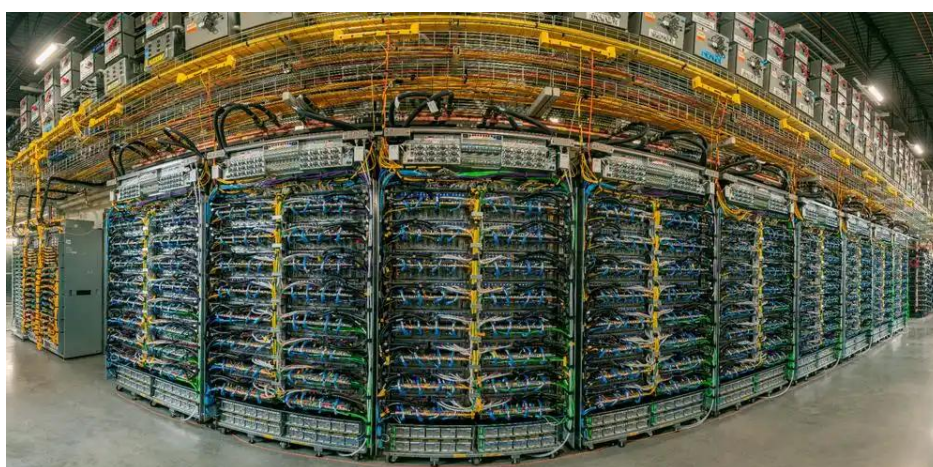
| | 英伟达 | 华为 | AMD | 谷歌 | 亚马逊 | 微软 | 阿里 |
|---------------|----------------------------------|-------------------|-----------------|-------------------|--------------------|------------------|---------|
| 芯片 | B300 | 昇腾 950PR | MI 355X | Ironwood | Trainium3 | Maia 200 | 真武 810E |
| 发布时间 | 2025. 3 | 2025. 9 | 2025. 6 | 2025. 4 | | | 2026. 1 |
| 出货时间 | 2025. 10 | 2026Q1 | 2025. 6 | 2025. 11 | 2025. 12 | 2026. 1 | |
| 微架构 | Blackwell Ultra | SIMD/SIMT | CDNA 4 | | | | |
| 制程 | 4nm | | 3nm | 3nm | 3nm | 3nm | |
| 最大功率 | 1400W | 600W | 1400W | 980W | | | 400W |
| 晶体管数量 | 2080 亿 | | 1850 亿 | | | 1400 亿 | |
| 互联带宽 | 1. 8TB/s | 1. 4TB/s | | 1. 2TB/s | 2. 2-2. 56TB/s | 2. 8TB/s | 700GB/s |
| 算力 | 15/20 PFLOPS* FP4 Tensor Core | 1.56PFLOPS FP4 | 20PFLOPS FP4 | 4614TFLOPs FP8 | 2517 TFLOPS FP4 | 10 PFLOPS FP4 | |
| 内存容量 | 279GB | 112GB | 288GB | 192GB | 144GB | 216GB | 96GB |
| 内存带宽 | 8TB/s | 1. 6TB/s | 8TB/s | 7. 4TB/s | 4. 9TB/s | 7TB/s | |
| 内存 | HBM3e | HiBL 1.0 | HBM3e | HBM3e | HBM3e | HBM3e | HBM2e |
| 超级算力集群 芯片数 | | | | 9216 | | | |

资料来源：英伟达，AMD，华为，谷歌，亚马逊，新浪财经，中原证券研究所（*稠密算力/稀疏算力）

1.4.2.1. 海外厂商：谷歌上调出货预期

(1) 谷歌：从 2013 年就开始自主研发云端 AI 加速芯片，并在 2015 年推出了自研的 TPU（张量处理器）。2025 年 11 月 6 日，谷歌正式发布了其第七代 TPU Ironwood，是谷歌首款专为推理设计的 TPU，其峰值算力 4614TFLOPs，内存 192GB，内存带宽 7.4TB/s，可扩展至 9216 片芯片集群，峰值算力可达 42.5EFLOPS，性能是 v6e（Trillium）的 4 倍以上，在特定应用中能提供比 GPU 高出 1.4 倍的美元性能。

图 29：包含 9,216 颗 Ironwood 的 TPU 集群



资料来源：谷歌，财联社，中原证券研究所

同时，作为云厂商中自有 AI 芯片占比最高的厂商，谷歌 2023 年自用的 TPU 芯片量已经突破了 200 万颗，已经成为了仅次于英伟达的全球第二大数据中心 AI 芯片厂商。2025 年谷歌

还宣布计划进入数据中心 CPU 市场，将实现微处理器和加速器芯片皆自研覆盖的发展格局。目前，谷歌的 Gemini、Veo、Imagen 以及 Anthropic 的 Claude 都在 TPU 上训练和服务。

2026 年 4 月谷歌宣布将 2026 年 TPU 芯片出货量目标大幅上调 50% 至 600 万颗。2025 年 12 月，Meta 与谷歌展开合作，将在 2026 年租用谷歌 TPU 进行 AI 训练，2027 年将直接采购谷歌的 TPU，Anthropic 2026 年将采购 1 吉瓦算力的谷歌 TPU，2027 年需求将增至 3 吉瓦。

(2) **亚马逊**:在 2013 年提出了开发定制硬件策略,2018 年推出了基于 ARM 架构的 CPU Graviton, 2019 年发布首代推理芯片 Inferentia, 2020 年发布了训练芯片 Trainium。2023 年末谷歌发布了 Trainium 2, 主要对标英伟达旗下 A100/H100 等旗舰 AI 训练芯片, 性价比比其他 GPU 供应商高出 30% 到 40%, 并已成为 Anthropic 等公司新一代模型训练的支柱。2025 年 12 月 4 日, 亚马逊发布自研芯片 **Trainium3** 和基于该芯片的 Trn3 UltraServers 服务器, 采用 3nm 工艺, 性能较 Trainium 2 提升 40%, 集成了 144GB 高带宽内存。集成 144 颗新一代芯片的 Trainium3 UltraServer 服务器, 计算能力是上一代产品的 4.4 倍, 内存带宽是上一代产品的 4 倍, 能够将能效提升 40%, 同时大幅提高了产品部署能力上限, 最多能部署多达百万颗 Trainium 3 芯片, 将此前的天花板提高了 10 倍。

同时, 亚马逊已经在全力研发 Trainium4, 并且已经进入深度设计阶段, 与 Trainium3 相比, Trainium4 将在所有维度实现巨大跃升: FP4 计算性能提升 6 倍、内存带宽提升 4 倍、高带宽内存容量提升 2 倍, 以支持全球最大规模模型的训练需求。

Trainium 的主要客户是 Anthropic、理光 (Ricoh) 等公司。

(3) **Meta**: 2026 年 3 月, 博通称下一代 MTIA 已开始出货, Meta 计划 2027 年及以后实现数吉瓦的定制加速器算力部署。

(4) **微软**: 2023 年 11 月, 微软推出了自研 AI 芯片 Maia 100 和 CPU 芯片 Cobalt 100, Maia 100 采用 5nm 工艺, 1050 亿颗晶体管。2026 年 1 月, 微软推出第二代人工智能芯片 Maia 200, 采用 3nm 工艺, 每颗芯片 1400 亿个晶体管, 是微软迄今为止部署的最高效推理系统, 其每美元性能比微软目前部署的新一代硬件提升了 30%, FP4 性能是第三代 Amazon Trainium 的 3 倍, FP8 性能则超过谷歌第七代 TPU。

微软目前在设计下一代芯片 Maia 300, 预计将于 2026 年下半年进入量产阶段。

(5) **OpenAI**: 2025 年 10 月, OpenAI 与博通合作生产其首款内部 AI 芯片, OpenAI 负责芯片设计, 博通从 2026 年下半年开始开发和部署芯片, 整体计划在 2029 年底完成价值 10GW 的定制芯片。2026 年 3 月博通表示 OpenAI 将在 2027 年部署超 1 吉瓦的首款自研定制 AI 处理器。

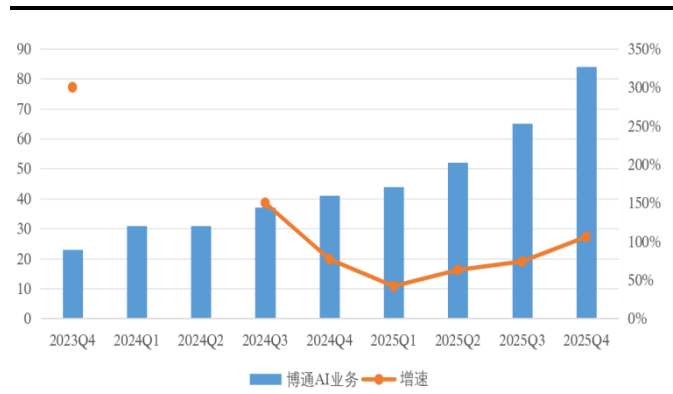
1.4.2.2. ASIC 厂商: 博通 2027 AI 芯片规模将超过 1000 亿美元

作为 AI 厂商和云厂商自研芯片的主要生产厂商，ASIC 厂商相关业务整体呈现出扩大趋势。

2025Q4 ASIC 厂商博通的 AI 业务实现收 84 亿美元，同比增长 106%，环比增长 29%，预计 2026Q1 AI 芯片营收将达到 107 亿美元，环比增长超 27%，同比增长 143%，预计到 2027 年其 AI 芯片收入将超过 1000 亿美元。博通量产客户包括了谷歌、Meta、字节，在 Q2 财报披露期间宣布拿下了第四个量产客户 100 亿美元订单，预计 2026 年开始交付。

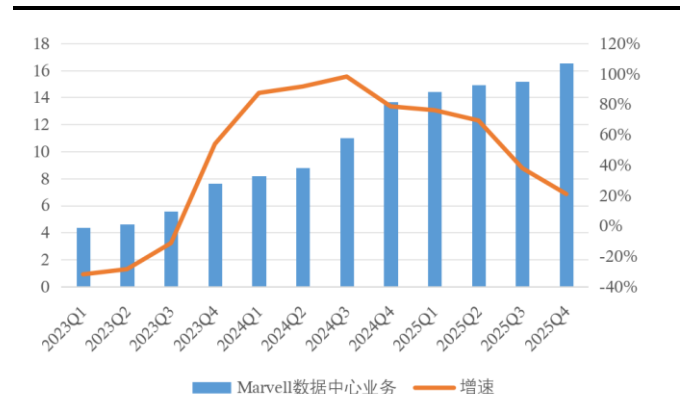
第二大 ASIC 厂商 Marvell 2025Q4 数据中心业务收入 16.5 亿美元，同比增长 20.86%，环比增长 9%，增长主要来自光互联产品带动，而定制 ASIC 业务出现环比下滑，收入主要来自 2024 年量产的亚马逊 Trainium 2，2026 年的定制 ASIC 业务增长将主要来自亚马逊的 Trainium3，同时在按计划推进 Maia 300 的研发，预计将在 2026 年下半年进入量产阶段。

图 30：2023Q4-2025Q4 博通 AI 业务收入（亿美元）



资料来源：博通财报，中原证券研究所（博通财报披露时间较以上周期晚一个月）

图 31：23Q1-25Q4 Marvell 数据中心业务收入（亿美元）



资料来源：Wind，中原证券研究所（Marvell 财报时段较以上周期晚一个月）

1.4.2.3. 国产互联网厂商：截至 2 月阿里自研芯片已经累计交付 47 万片

(1) 阿里：2026 年 1 月 29 日，阿里平头哥上线了自研芯片真武 810E 的产品信息，该 PPU 采用自研并行计算架构和片间互联技术，配合全栈自研软件栈，实现软硬件全自研。其内存为 96G HBM2e，片间互联带宽达到 700 GB/s，可应用于 AI 训练、AI 推理和自动驾驶。其关键参数超过 A800，与英伟达 H20 相当，其目前已经在阿里云实现了多个万卡集群部署，服务了 400 多家客户。

阿里 🐼 巴巴在 3 月的财报电话会议上表示，平头哥自研 AI 芯片已经实现规模化量产，截至 2026 年 2 月，已经累计规模化交付 47 万片。4 月阿里表示目前已有 30 多家车企和智驾方案提供商在阿里云上开展智驾研发，实际使用的平头哥自研“真武”PPU 已突破 10 万卡。4 月 3 日，由中国电信联合阿里云建设的粤港澳大湾区首个“真武”万卡智算集群在韶关数据中心正式上线，实现了从芯片、云平台到模型应用的全链路自主研发。

图 32：平头哥算卡与英伟达特供芯片的性能对比

| 厂商 | 型号 | 显存容量 | 显存类型 | 片间带宽 (GB/s) | PCIe | 功耗 (W) |
|-----|------|------|-------|-------------|--------|--------|
| 平头哥 | PPU | 96G | HBM2e | 700 | 5.0×16 | 400 |
| 英伟达 | A800 | 80G | H20 | 800 | 4.0×16 | 400 |
| 英伟达 | H100 | 80G | HBM3 | 900 | 5.0×16 | 550 |

李强在甘肃、青海调研

资料来源：新闻联播，中原证券研究所

图 33：平头哥规模部署情况

| 序号 | 项目名称 | 品牌 | 设备(台) | 算力卡(张) | 总算力(P) |
|-----|-------|-----|-------|--------|--------|
| 1 | 阿盟五万卡 | 平头哥 | 1024 | 16384 | 1945 |
| 2 | 中科院 | 沐曦 | 512 | 4096 | 984 |
| 3 | 北京航仪 | 壁仞 | 83 | 1328 | 450 |
| 4 | 中发芯策 | 影驰 | 121 | 1936 | 200 |
| 阿盟的 | | | 111 | 1776 | 2002 |
| 1 | / | 太初 | 800 | 12800 | 336 |
| 2 | / | 摩尔 | 120 | 1920 | 262 |
| 3 | / | 摩尔 | 100 | 1600 | 204 |

资料来源：新闻联播，中原证券研究所

(2) 百度：百度在 2011 年启动了 FPGA AI 加速器项目，2017 年部署过万片，2018 年从 FPGA 转向 AI 芯片研发，并在当年推出首款基于 XPU 架构的云端 AI 芯片“昆仑 1”，采用 14nm 工艺，专注 AI 推理，2020 年实现了大规模量产部署。2021 年 8 月，昆仑芯 2 代 AI 芯片 R200（高性能推理和训练加速卡）量产，采用 7nm 制程，128 TFLOPS FP16 和 256 TOPS INT8 算力，32GB 高速内存，512GB/s 内存带宽，同年实现大规模部署。2025 年 4 月，百度第三代昆仑芯 P800 速卡已成功点亮 3 万卡 AI 集群。在下一代芯片技术路线图方面，百度针对大规模推理场景优化设计的 M100 将于 2026 年上市，面向超大规模的多模态模型的训练和推理任务的 M300 预计 2027 年上市。2026 年 1 月 1 日，昆仑芯已经向港交所递交了主板上市申请。

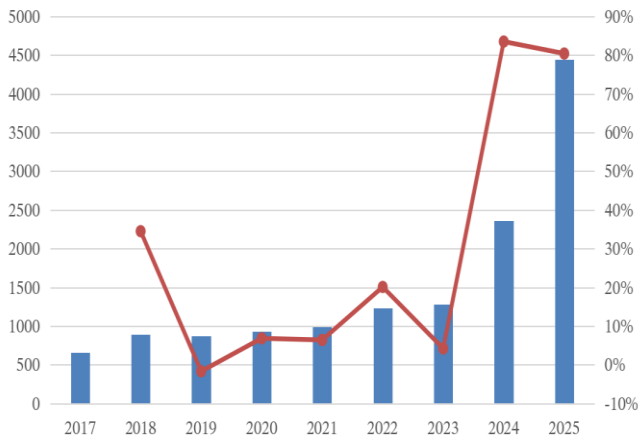
(3) 字节：2026 年 2 月报道称字节正在研发人工智能芯片，并与三星洽谈代工事宜。其计划在 3 月底前收到芯片样品，2026 年计划生产至少 10 万枚用于 AI 推理任务的芯片，并逐步扩产至 35 万枚。

1.4.3. 服务器：芯片供应影响了国内服务器增长的节奏

1.4.3.1. 全球市场：2025Q4 美国继续引领，国内缺芯影响下增速放缓

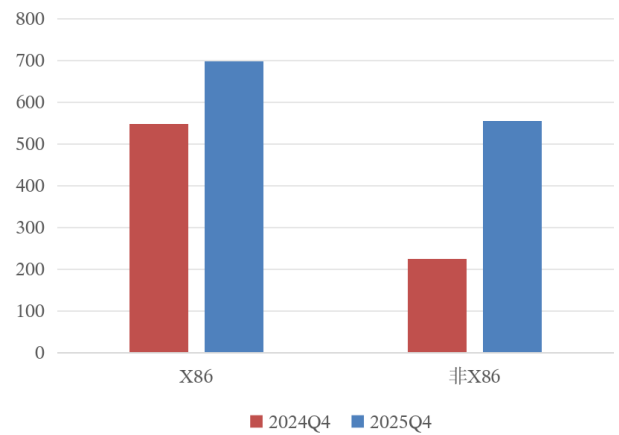
全球服务器市场呈现快速增长，其中非 X86 服务器已经成为重要的拉动因素。2025 年全球服务器市场规模达到 4441 亿美元，同比增长 80.4%。2025Q4 全球服务器市场规模达到 1253 亿美元，同比增长 52.4%，其中 X86 和非 X86 分别增长了 16.9% 和 146.7%，非 X86 的占比已经达到了 44%。

图 34: 2017-2025 全球服务器市场规模 (亿美元)



资料来源: IDC, 中原证券研究所

图 35: 2025Q4 全球不同类型服务器 (亿美元)

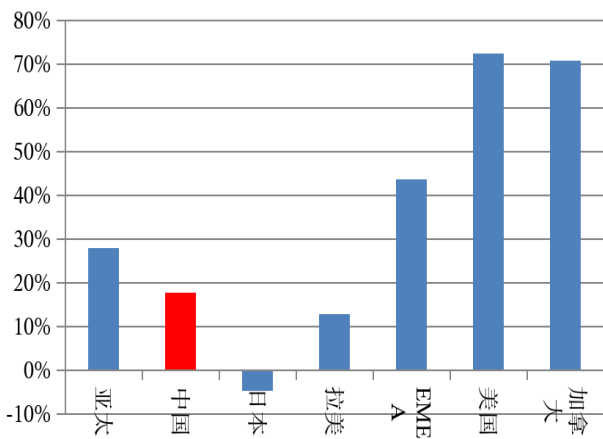


资料来源: IDC, 中原证券研究所

分区域来看, 2025Q4 增长最快的地区为美国和加拿大, 同比增速分别为 72.4% 和 70.7%, 而同期中国增速仅为 17.7%, 较 2024Q4 93.3% 的增速下滑明显, 我们认为主要源于缺芯因素影响了服务器需求的释放。

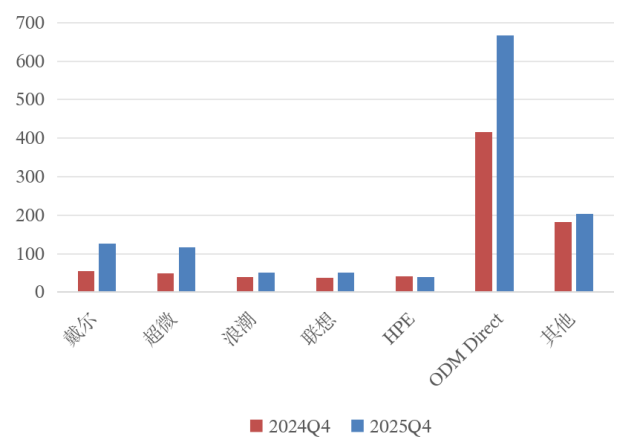
分厂商来看, 头部的戴尔、超微分别实现了 127% 和 134% 的增速, 同样位于美国的 HPE 是仅有的收入下滑的厂商, ODM 厂商同样实现了 61% 的增长, 深度受益于 AI 的发展, 而国内的两家厂商浪潮和联想都增长 34%, 受到了国内市场增速放缓的拖累。

图 36: 2025Q4 全球不同区域服务器收入增速



资料来源: IDC, 中原证券研究所

图 37: 2025Q4 全球服务器市场头部厂商收入 (亿美元)

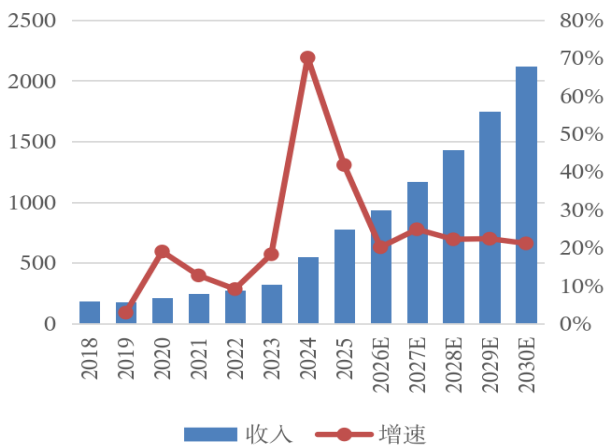


资料来源: IDC, 中原证券研究所

1.4.3.2. 国内市场: 2030 年加速服务器预计占比将达到 70%

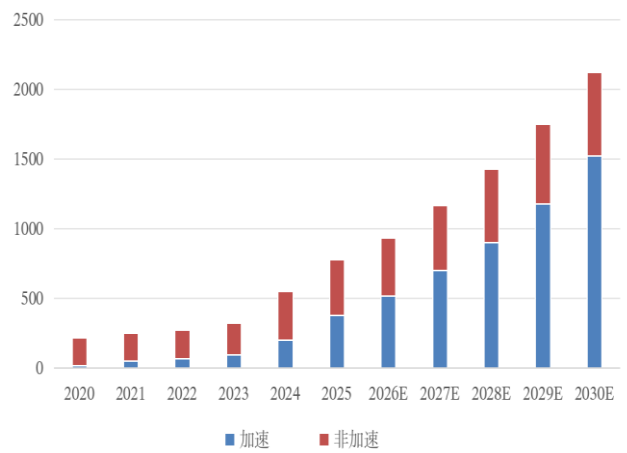
中国市场上, 加速服务器也正在成为整体市场的增长引擎, IDC 预计到 2030 年其在整体服务器市场占比将达到 70% 以上。

图 38: 2018-2025 中国服务器市场规模 (亿美元)



资料来源: IDC, 中原证券研究所

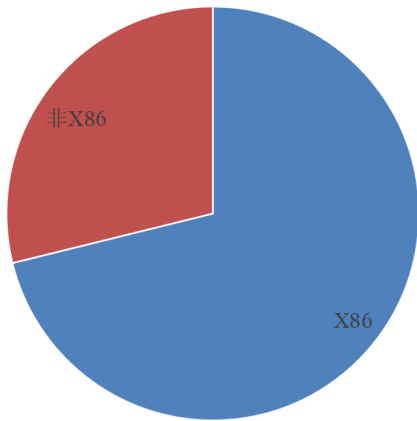
图 39: 中国加速服务器增长预测 (亿美元)



资料来源: IDC, 中原证券研究所

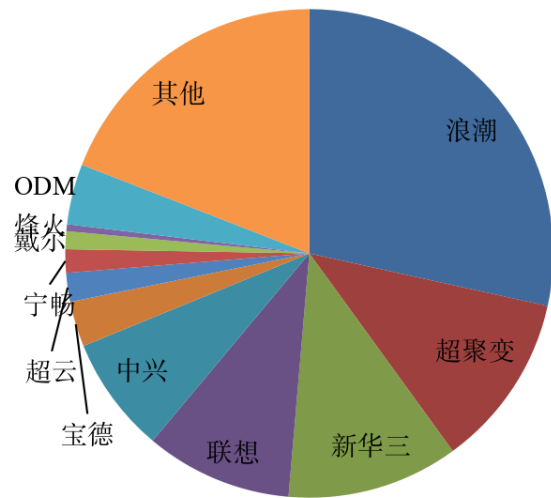
2025 年我国服务器市场中, X86 占比超过七成, 其中浪潮仍然稳居市场第一, 份额 31%; 第二梯队为超聚变、新华三、联想、中兴, 份额分别为 13%、13%、11%、9%; 第 6 到第 10 分别为宝德、超云、宁畅、戴尔、烽火; ODM 份额为 4%, 不同于海外 ODM 份额持续扩大, 国内服务器市场中, ODM 份额呈现缩小趋势。

图 40: 2025 中国服务器市场规模 (亿美元)



资料来源: IDC, 中原证券研究所

图 41: 2025 年中国 X86 服务器厂商市场份额



资料来源: IDC, 中原证券研究所

从 2025 年年报来看, 工业富联云计算业务增速从上年的 64% 增长到了 89%, 呈现进一步加快的趋势, 深度受益于北美云服务商 AI 业务的需求, 也明显跑赢其他国内厂商; 浪潮信息在禁令是影响下增速从上年的 74% 下滑到 43%; 联想的基础设施解决方案业务增速也从上年的 51% 略有下滑到了 43%; 新华三增速从上年 6% 提升到了 38%, 加速迹象明显; 中科曙光增速加快, 从上年的 -8% 增长到了 14%, 其收入规模虽然较小, 但是毛利率明显领先同行, 同时净利润水平与浪潮相当。

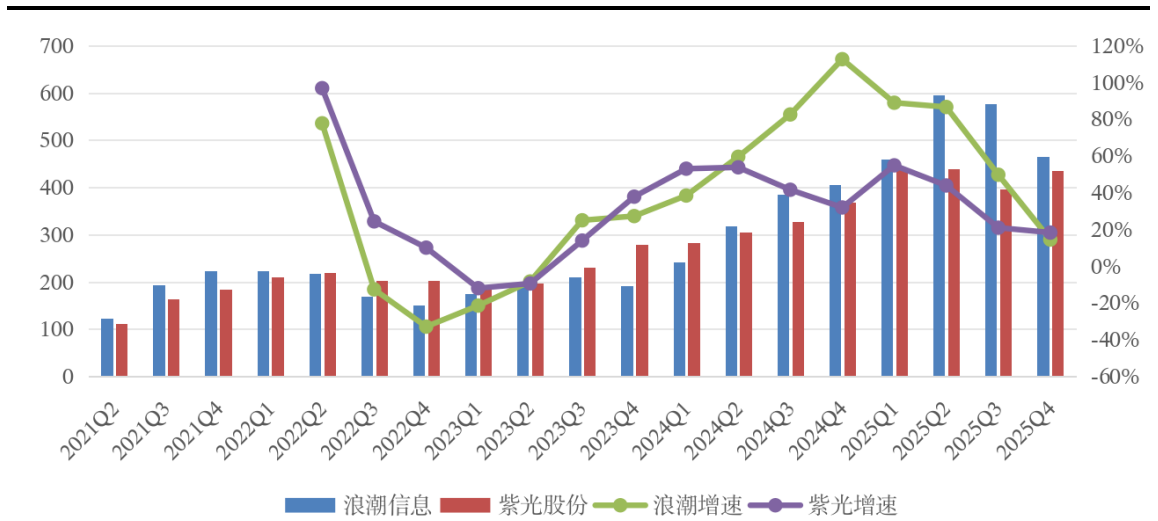
表 6：2025 年企业服务器相关业务对比（亿元）

| 公司 | 相关业务 | 2025 年 | | | | | | | 2024 年 | |
|------|------------------|--------|-------|------|-----|------|-----|------|--------|-----|
| | | 收入 | 增速 | 净利润 | 增速 | 毛利率 | 毛利额 | 人员增速 | 收入 | 增速 |
| 工业富联 | 云计算 | 6027 | 89% | | | 5.7% | 345 | 7% | 3194 | 64% |
| | 云服务商 AI 服务器 | | 300%+ | | | | | | | |
| 浪潮信息 | - | 1648 | 43% | 24 | 5% | 4.9% | 80 | -1% | 1148 | 74% |
| | 中国大陆收入 | | | | | | | | 807 | 43% |
| 联想集团 | 基础设施解决方案 | 1212 | 37% | | | | | 2% | 887 | 51% |
| 超聚变 | - | | | | | | | | 435 | 53% |
| 紫光股份 | 新华三 | 760 | 38% | 32 | 12% | | | -12% | 551 | 6% |
| 中科曙光 | - | 150 | 14% | 22 | 14% | 31% | 46 | | 131 | -8% |
| | IT 设备业务 | 125 | 7% | | | 27% | 34 | 1% | 117 | -8% |
| 宝德 | (慧博云通拟收购) | | | | | | | | 100 | 8% |
| 软通动力 | 计算产品与智能电子 | | | | | | | | 129.2 | |
| 神州数码 | 自有品牌产品 | 74 | 62% | 0.78 | 10% | | | -6% | 45.8 | 20% |
| 广电运通 | 广电五舟 | 18 | 5% | 0.28 | 73% | | | 3% | 17 | 31% |
| 拓维信息 | 计算机、通信和其他电子设备制造业 | | | | | | | | 20.7 | 17% |
| | 湘江鲲鹏 | | | | | | | | 15.8 | 2% |

资料来源：上市公司公告，Wind，大河网，中原证券研究所

由于 AI 应用带来的需求释放，目前上游产品供应紧张局势明显，供应端的不稳定性问题在 2026 年还将进一步凸显。从两家头部服务器厂商来看，浪潮存货在 2025Q2 见顶后，连续两个季度环比下滑，而紫光股份的存货在 Q3 环比下滑 10% 后，Q4 环比上升 10%，补货能力略好于浪潮。

图 42：2021Q2-2025Q4 紫光 and 浪潮存货增长情况（亿元）



资料来源：上市公司公告，中原证券研究所

1.4.3.3. 超节点：移动集采大单落地，郑州建 6 万卡超算集群

(1) 超节点落地超算

以 scaleX 640 为基础的国家超算互联网核心节点 AI 算力集群扩容至 6 万卡。

2 月 5 日，由中科曙光提供的 3 套 scaleX 万卡超集群系统在国家超算互联网郑州核心节点同步上线试运行。这是全国首个实现 3 万卡部署并实际投入运营的最大国产 AI 算力池，全面覆盖万亿参数模型训练、高通量推理、AI for Science 等大规模 AI 计算场景。scaleX 万卡超集群就是以 scaleX 640 超节点为基础建设的，scaleX 640 采用开放式架构，通过超高密度刀片设计配合浸没相变液冷，将单机柜算力密度提升到传统模式的约 20 倍，PUE 压低至接近 1.04 的高效水平。

4 月 14 日，该计算集群扩容至 6 万卡，可支持 8/16/32/64 位宽的全精度计算，打破了传统超算与智算的壁垒。随着这一集群的落地运行，国家超算互联网平台也构建起了链接超 300 万 CPU 和超 20 万 GPU 的 AI4S 计算基础设施，并接入了全国一体化算网调度体系。

图 43: scaleX640 超节点



资料来源：曙光数创公告，中原证券研究所

(2) 运营商集采超节点

4 月 12 日，中国移动 2026-2027 年人工智能超节点设备集中采购项目公布了招标结果。本次集采采购人工智能超节点，指定华为 CANN 生态，总计设计 AI 加速卡 6208 张，对应超节点 776 套，对应采购金额约为 20 亿元。本次为运营商集团层面的首次启动 AI 超节点集采，被视为运营商算力建设告别服务器简单堆叠的传统模式，转而步入超节点的规模化部署阶段，因而具有风向标意义。

本次中标的企业共有 5 家，其份额分别为：河南昆仑（超聚变）27.78%，长江计算（烽火）22.22%，华鲲振宇 19.44%，宝德 16.67%，华启智慧（大华股份）13.89%。

表 7: 主要厂商的超节点产品上市情况

| 上市时间 | 厂商 | 产品 | GPU | GPU 数量 | 机柜数量 | 备注 |
|----------|-----|-------------|-------|--------|------|-----------------------------------|
| 2024. 11 | 英伟达 | GB200 NVL72 | GB200 | 72 | 1 | 算力 180 EFLOPS FP16/360 PFLOPS FP8 |

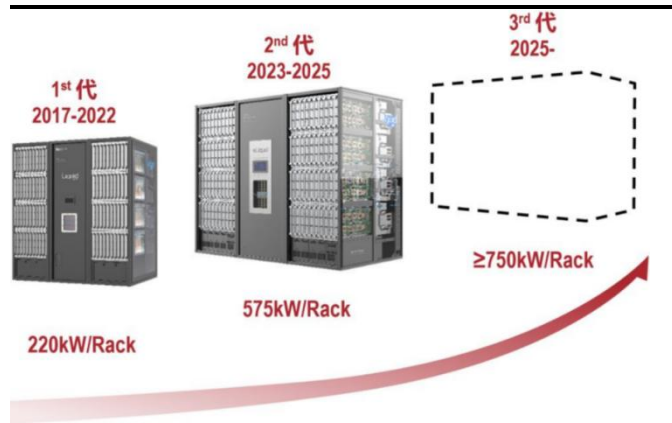
| | | | | | | |
|-----------|-----|--------------------|----------|-------|--------|-----------------------------------|
| 2025.3 | 华为 | Atlas 900 | 昇腾 910C | 384 | 12+4 | 算力 300 PFLOPS BP16, 业内最大超节点 |
| 2025.4 | 百度 | 天池 64 | | 64 | | 天池 32、天池 64 上市 |
| 2025.7 | 英伟达 | GB300 NVK72 | GB300 | 72 | 1 | 算力 360 PFLOPS FP8/1.08 EFLOPS FP4 |
| 2025.9 | 阿里 | 磐久 AI Infra2.0 | 多种 AI 芯片 | 128 | 1 | |
| 2025.11.6 | 曙光 | ScaleX640 | 多种 AI 芯片 | 640 | 1 | 浸没相变液冷, 算力密度业内新高 |
| 2026H1 | 百度 | 天池 256 | | | | |
| 2026H2 | 英伟达 | Vera Rubin NVL144 | Rubin | 144 | | 算力 1.2 EFLOPS FP8/3.6 EFLOPS FP4 |
| 2026H2 | 百度 | 天池 512 | | | | 能完成万亿参数模型训练 |
| 2026Q4 | 华为 | Atlas 950 | 昇腾 950DT | 8192 | 128+32 | 算力 8 EFLOPS FP8/16 EFLOPS FP4 |
| 2027H2 | 英伟达 | Rubin Ultra NVL576 | Rubin | 576 | | 算力 5 EFLOPS FP8/15 EFLOPS FP4 |
| 2027Q4 | 华为 | Atlas 960 | 昇腾 960 | 15488 | 176+44 | 算力 30 EFLOPS FP8/60 EFLOPS FP4 |
| 2028 | 百度 | 天池千卡级 | | 千卡 | | |

资料来源：英伟达，华为，阿里，曙光，百度，中原证券研究所

1.4.3.4. 液冷：浸没式液冷上限再获突破

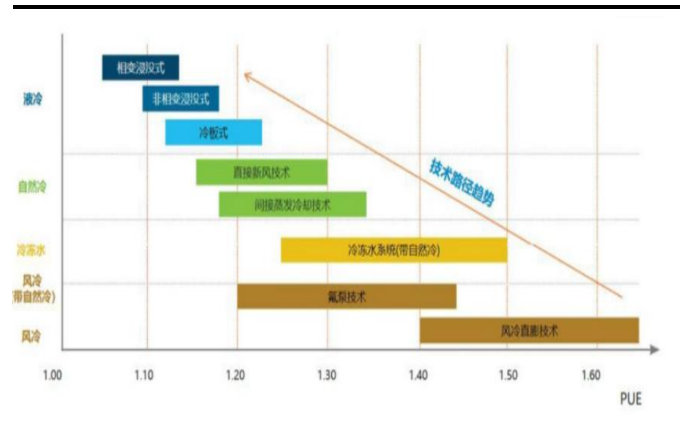
4月8日，曙光数创正式发布了全球首个 MW 级浸没式液冷整机柜及其基础设施解决方案——C8000 V3.0，最高支持的单机柜功率进一步提升到了 900kW。采用自研国产冷媒，可实现全年自然冷却。并首次规模化应用金刚石铜导热材料，导热率提升 80%，助力芯片性能提升 10%。机房面积节省超 85%，将数据中心 PUE 降至 1.04 以下。

图 44：曙光数创浸没式液冷的技术进展



资料来源：曙光数创公告，中原证券研究所

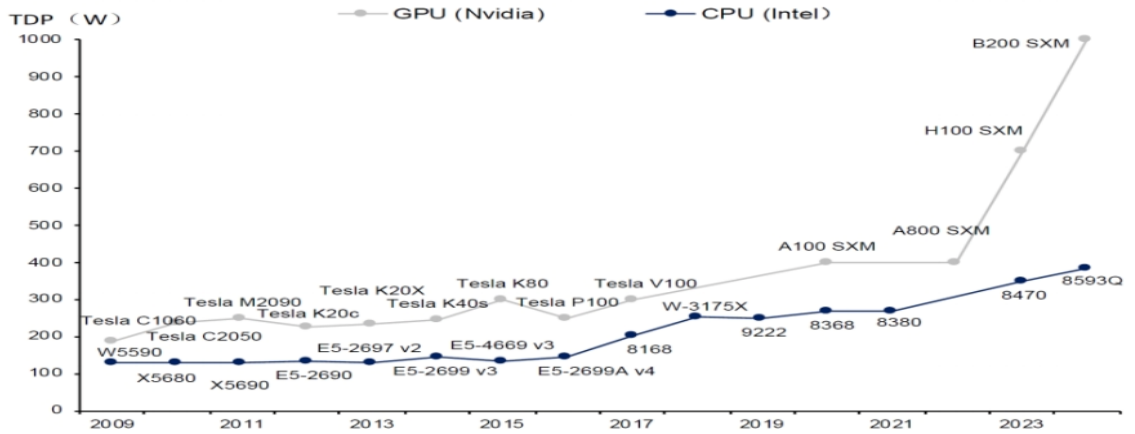
图 45：不同冷却技术与 PUE 相关性



资料来源：《中兴通讯液冷技术白皮书》，曙光数创公告，中原证券研究所

后摩尔定律时代下，芯片的制程技术发展趋缓，计算芯片多以提高核心数量等方式提高算力，导致计算芯片如 CPU 和 GPU 的功耗不断攀升，也倒逼了液冷技术的应用。

图 46: GPU 和 CPU 功耗提升趋势

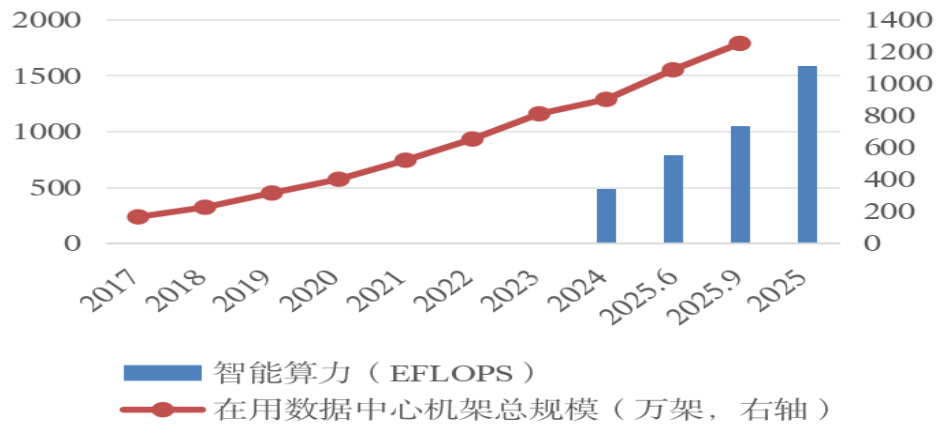


资料来源: 曙光数创公告, 中原证券研究所

1.4.4. 数据中心: 2025 年底我国智算规模较年中翻倍, 2026 年有加速迹象

2025 年我国总体算力加速增长过程中。根据工信部数据, 2025 年我国智能算力规模达到 1590EFLOPS, 较 2025 年中 788EFLOPS 的规模已经实现了翻倍。

图 47: 2017-2025 我国智能算力和在用数据中心机架数规模



资料来源: 工信部, 中原证券研究所

2026 年行业算力建设速度有明显加快迹象, 第三方数据中心厂商开启积极建设步伐。润泽 2026 年在香港、印尼分别获得了 240MW 和 360MW 的算力储备, 开启了境外布局的步伐。3 月媒体报道世纪互联获得字节 500MW 大单, 同时世纪互联计划 2026 年投入 100-120 亿元, 重点支撑 450MW-500MW 的年度交付计划。东阳光在 1 月完成对秦淮 30% 的股权收购以后, 又于 2 月发起收购秦淮剩余 70% 股权的事宜。

在融资方面, 考虑到数据中心是重资产业务, 需要较强的资金支持, 2026 年厂商加速融资迹象明显, REITs 作为新融资手段和资产变现的通道受到了更多的青睐。润泽科技 2 月公告发起新一轮两个数据中心的 REITs 融资, 同时计划发行可转债。世纪互联 1 月江苏及内蒙古两个

IDC 私募 REITs 获批，拟发行金额合计 107.59 亿元。奥飞数据在 1 月终止 2025 年发起的定增计划以后，其发起规模 21 亿元私募 REITs 在 4 月获受理。尚杭科技拟登陆北交所。

从财报相关数据来看，2025 年反映出的特点包括：

(1) 2025 年，世纪互联、润泽科技、光环新网的算力增速较快。

(2) 秦淮、润泽、宝信依然保持了行业最优秀的毛利率水平，进而保证了盈利能力，数据港 Q4 毛利率提升明显，而万国、世纪互联依然亏损。

(3) 较早的 AIDC 业务布局，让润泽 2025 年收入已经获得了较高的增速水平，2026Q1 有进一步提速迹象。

表 8：2025 年企业数据中心相关业务对比（亿元）

| 公司 | 相关业务 | 2025 年 | | | | | | | | | 2024 年 | |
|------|------------|--------|------|------|-------|------|---------|--------------|-------|------|--------|------|
| | | 收入 | 增速 | 净利润 | 增速 | 毛利率 | 使用率 | 算力规模 | 算力增长 | 总体算力 | 收入 | 增速 |
| 中国移动 | 数据中心 | | 9% | | | | | 3.75GW | | | | |
| 中国电信 | AIDC | 345 | | | | | | 3.2GW | | | | |
| 中国联通 | 数据中心 | 281 | 9% | | | | 72% | 2.75GW | 250MW | | 259 | 7% |
| 万国数据 | | 114 | 11% | -14 | -142% | 23% | 75.5% | 709 万平米 | 9 万平米 | | 103 | 4% |
| 世纪互联 | | 100 | 21% | -3 | -300% | 22% | 70%-64% | 889MW+5.2 万架 | 404MW | 10GW | 83 | 11% |
| 秦淮数据 | (东阳光) | 64 | 6% | 17 | 26% | 42%* | | 799MW | | 4GW | 61 | |
| 润泽科技 | | 57 | 30% | 50 | 182% | 46% | 90% | 750MW | 220MW | 6GW | 44 | 0.3% |
| 宝信软件 | 服务外包 | 43 | 3% | | | 43% | | | | | 37 | 5% |
| 奥飞数据 | | | | | | | | | | | 22 | 62% |
| 城地香江 | | | | | | | | | | | | |
| 光环新网 | IDC 及其增值服务 | | | | | | 60% | 361MW | 114MW | 1GW | 21 | -6% |
| 恒润股份 | 算力行业/润六尺 | 17 | 744% | 0.3 | | 5% | | | | | 2 | |
| 郑州合盈 | (豫能控股) | 17* | | | | | 60%-79% | 579MW | | | | |
| 数据港 | | 17 | 0% | 1.4 | 5% | 38% | | 401MW | 30MW | | 18 | 12% |
| 科华数据 | IDC 服务 | | | | | | | | | | 12 | -5% |
| 尚杭科技 | | 5 | 13% | 0.7 | 1% | 27% | | 300MW | | | 5 | -2% |
| 杭钢股份 | 算力子公司 | | | | | | | | | | 0 | |
| 大位科技 | | 4 | | -0.2 | | 17% | | | | | | |

资料来源：上市公司公告，Wind，中原证券研究所（*年化处理后数值）

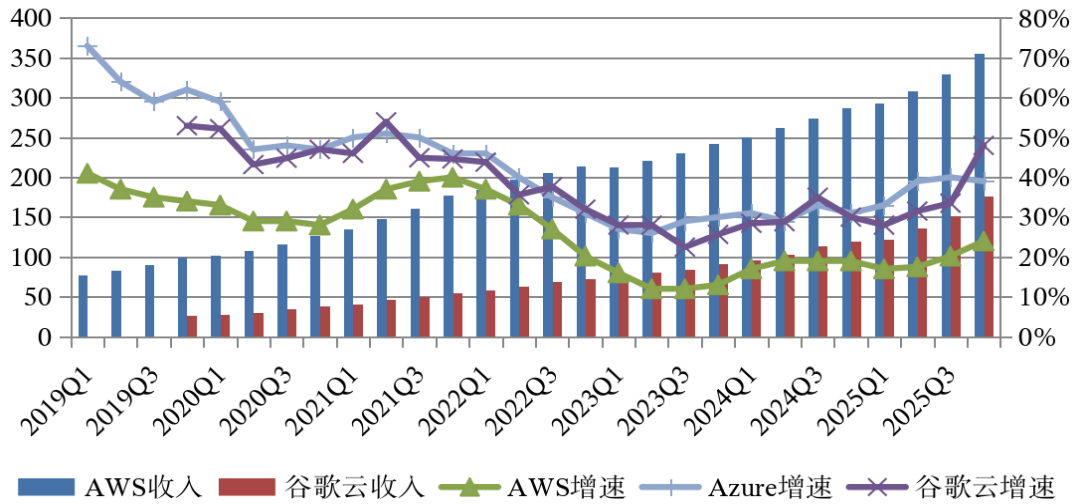
1.4.5. 云计算：受到 AI 需求拉动，整体增速仍在加快

1.4.5.1. 全球：云厂商云业务增速整体加快，但是厂商的分化更加明显

全球云巨头整体呈现了业务加速趋势，但是分化加剧。根据 2025Q4 相关公司财报，亚马

逊的 AWS 实现收入 356 亿美元，同比增长 24%，较上个季度提升了 3.8 PCT。微软 Azure 和谷歌云分别增长了 39%和 48%，增速明显快于 AWS，分别较上个季度增长了-1 和 14.5 PCT，分化加剧。

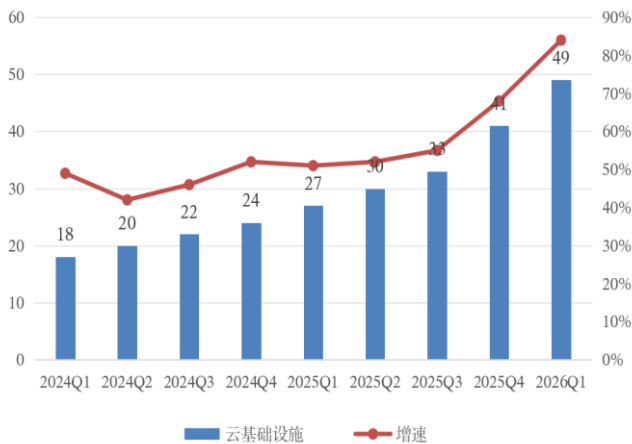
图 48：2019Q1-2025Q4 海外云巨头的相关业务收入及增速



资料来源：上市公司财报，中原证券研究所

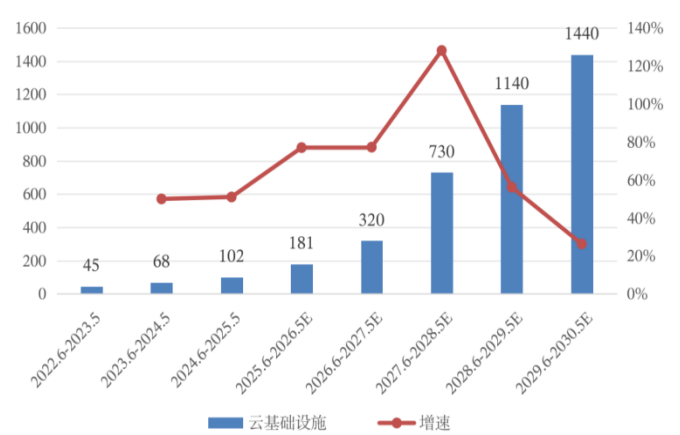
甲骨文云基础设施业务增速加快至 84%。2026 财年 3 季度，甲骨文云基础设施营收 49 亿美元，同比增长 84%，较上个季度上涨 16 PCT，该季度甲骨文向客户交付了超过 400MW 的产能。同时甲骨文剩余履约义务达到 5530 亿元，同比增长 325%，较上财季增长 290 亿美元。同时甲骨文表示未来三年已锁定超 10GW 的电力和数据中心容量，其中超过 90%完全由合作伙伴提供资金。受到巨额合同的需求支撑，2025 年 9 月甲骨文预计 2026 财年云基础设施营收增长超过 77%至 180 亿美元，2027-2030 财年将分别为 320 亿美元、730 亿美元、1140 亿美元、1440 亿美元。

图 49：甲骨文云基础设施收入及增速（亿美元）



资料来源：甲骨文财报，中原证券研究所（财报时点早一个月）

图 50：甲骨文云基础设施业务年度收入预测（亿美元）



资料来源：甲骨文财报，中原证券研究所

1.4.5.2. 国内：阿里云 AI 相关产品收入连续 10 个月三位数增长

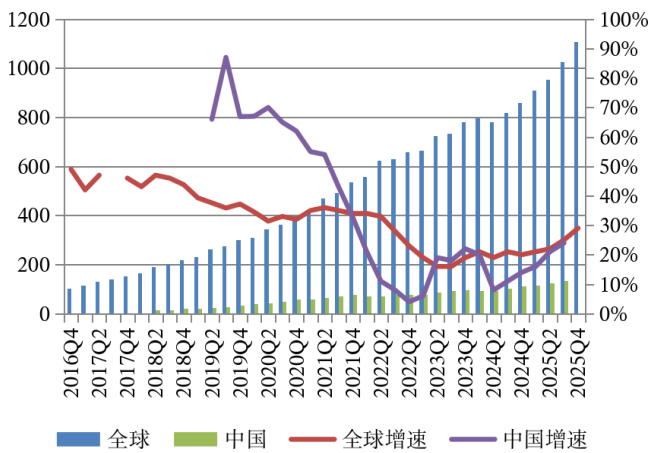
国内云业务连续 5 个季度提速。根据 Omdia 数据，2025Q3 我国云基础设施服务上的支出总计达到 134 亿美元，同比增长 24%，较上个季度增长 3PCT，连续 5 个季度增速回升，略低于全球 25% 的增速水平。我们预计随着国内大模型的快速发展和 AI 应用的持续落地，国内云业务的增速有望持续提升。

从云基础设施服务的厂商中，阿里、火山引擎呈现两强增长格局。

根据 Omdia 数据，2025Q3，阿里仍然牢居第一，占比 36% 较上季度增长 2PCT；华为云份额达到 16%，10 月华为云研发组织被切换到 ICT 下，后续或将更多让位于硬件端的发展；腾讯云以 9% 的份额位居第 3，由于优先对内供给，GPU 供给紧张，对市场份额产生负面影响，考虑到腾讯削减了全年资本开支目标，我们预计其云业务份额有进一步下滑趋势。

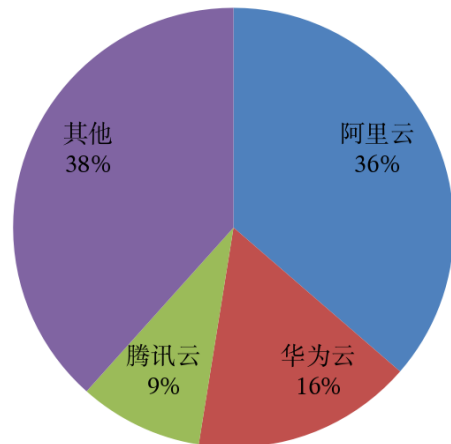
2021 年正式开启云服务的字节，2023 年下半年推出了火山方舟(MaaS 大模型服务平台)，2024 年中推出了豆包大模型。据《划重点》报道，火山引擎 2024 年营收超过 120 亿，增长超过 60%，2025 年营收目标超过 250 亿，目标增速超过 100%。

图 51：2016-2025 年我国及全球云基础设施服务的支出规模及增速（亿美元）



资料来源：Canalys，中原证券研究所

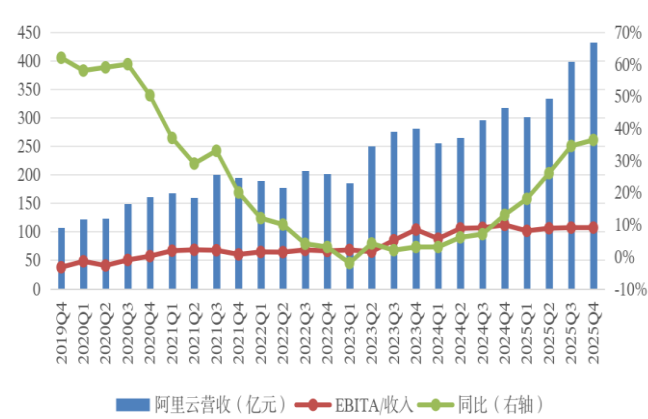
图 52：2025Q3 我国云基础设施服务市场结构



资料来源：Omdia，中原证券研究所

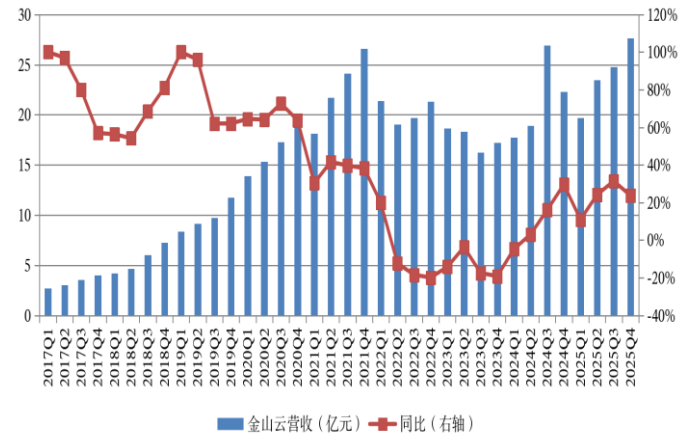
从厂商业绩来看，在 AI 推动下，阿里云增速持续提升，AI 相关收入连续 10 个月三位数增长。其中 2025Q4 阿里云收入 432.84 亿元，增速达到了 36%，较上个季度提升了 2PCT，其中 AI 相关收入连续 10 个季度实现同比三位数增长。2025Q4 金山云同比增长 23.71%，处于较高增速水平，其中智算云业务总收入同比增长 95%，至 9.3 亿元。

图 53：2019Q4-2025Q4 阿里云收入及其增速



资料来源：阿里财报，中原证券研究所

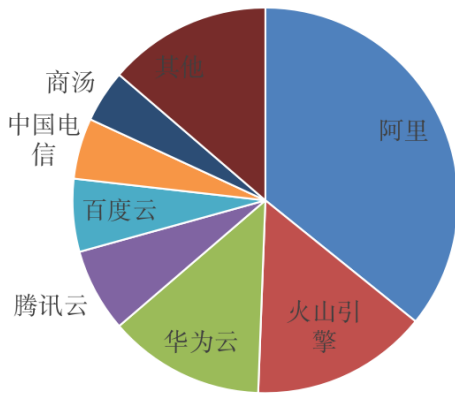
图 54：2017Q1-2025Q4 金山云收入及其增速



资料来源：金山云财报，中原证券研究所

在 AI 云加速增长趋势明显。根据 Omdia 数据，2024 年市场规模为 208.3 亿元，2025 年上半年达到了 223 亿元，超过了 2024 年全年。Omdia 预测，2025 年中国 AI 云市场规模达到 518 亿元，同比增长 148%，2030 年将达到 1930 亿元。2025 年上半年，AI 云市场前 7 大厂分别为阿里、字节、华为云、腾讯云、百度云、中国电信、商汤，其中阿里云份额为 35.8%。

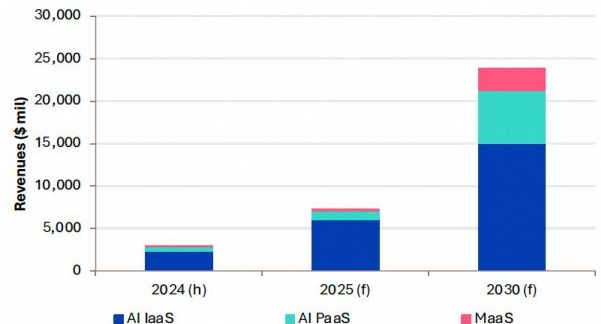
图 55：2025H1 中国 AI 云市场格局



资料来源：Omdia，中原证券研究所

图 56：2024-2030 年 AI 云市场结构及规模

AI cloud market, China, 2024-2030



Source: Omdia

© 2025 Omdia

资料来源：Omdia，中原证券研究所

阿里、字节目前都对 MaaS 业务给予了积极的预期。

阿里：2026 年 1 月 14 日，阿里云提出 2026 年发展目标是拿下全年中国 AI 云市场增量的 80%，同时阿里也判断 2026 年增量的 10% 都会大于 2025 年全年。3 月 19 日，阿里又宣布未来 5 年，阿里包含 MaaS 在内的云和 AI 商业化年化收入突破 1000 亿美元。

字节：此前火山引擎定下 2026 年 MaaS 业务收入超百亿元的目标，2025 年公司因为这部分业务增长迅猛，两度上调收入目标，按照目前的发展速度，公司千亿营收目标大概提前实现。

1.4.6. 资本开支

考虑到当前 AI 应用步入快速发展期，推动算力进入到了供不应求阶段。当前算力的两大需求主要来自美国和中国，两者受到各自政策和产业基础的影响，逐渐呈现出硬分叉状态：

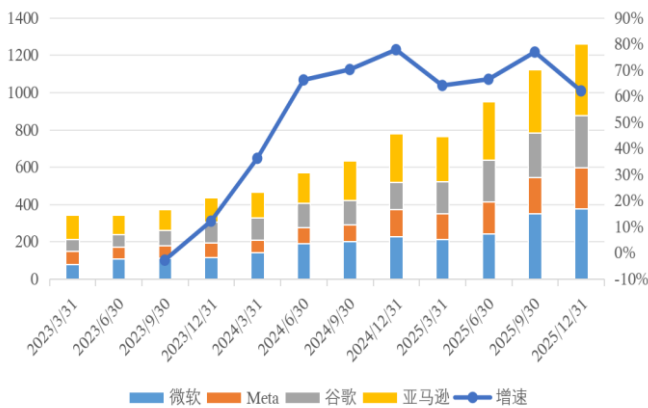
美国: Anthropic 引领的应用需求快速增长, OpenClaw 带动下 AI 应用呈现加速落地趋势。2024-2025 年美国科技厂商的资本开支持续快速增长, 但是在旺盛的需求背后, 算力基础设施建设进度受到电力网络的制约越来越明显, 同时以甲骨文为代表的厂商短期资本开支过大, 现金流短缺迹象明显, 引起了市场的担忧。

中国: 受到美国芯片出口管制政策的影响, 国内持续处于芯片缺芯状态, 抑制了部分采购需求。国产芯片从 2025 年底以来已经有较大的性能提升, 但是仍然受制于产能瓶颈, 或有望在 2026 年下半年获得较大的改善。同时, 2026 年以来, 应用端的需求释放明显, 带来算力总体供应不足, 涨价趋势明显。因而我们判断 2026 年, 中国算力开支有望获得较大增长, 并由此带来产业链上下游整体的投资机会。

1.4.6.1. 海外资本开支: 海外科技企业资本开支超预期, 引起市场担忧

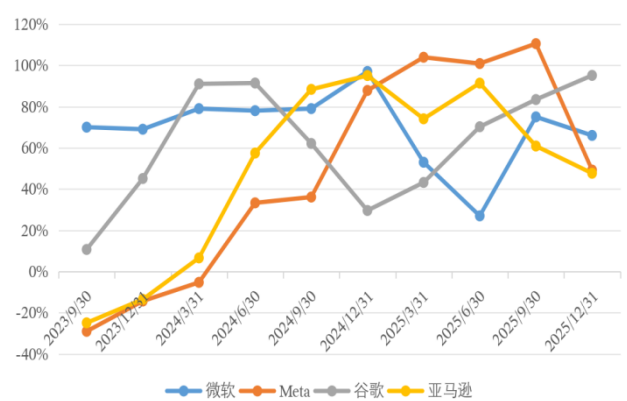
从海外市场来看, 2025 年美国 4 大科技厂商 (亚马逊、微软、谷歌、Meta) 的资本开支持续增长, 海外 AI 基础设施投入仍在加大。2025Q4 4 大厂商资本开支再创新高, 总计达到 1260 亿美元, 同比增长 62%, 仍然保持了较高增速水平。其中亚马逊资本投入 385 亿美元, 增长 48%, 仍然是投入最大的厂商; 微软资本开支 375 亿美元, 同比增长 66%; 谷歌资本开支 279 亿美元, 增长 95%, 在模型能力领先的同时也开始更加积极的 AI 投入; Meta 资本开支达到了 221 亿美元, 同比增速 49%, 较上个季度下滑明显。

图 57: 4 大科技厂商资本投入 (亿美元)



资料来源: 微软, Meta, 谷歌, 亚马逊, 中原证券研究所

图 58: 4 大科技厂商资本投入增速



资料来源: 微软, Meta, 谷歌, 亚马逊, 中原证券研究所

厂商 2026 年资本开支整体大幅超出了市场预期, 也引发了市场担忧。随着海外科技巨头接连公布了 2025Q4 业绩, 我们看到了资本开支都呈现出了上调趋势, 对 2026 年 AI 应用需求给予积极预期。其中亚马逊在 2025 年 1310 亿美元资本开支基础上, 2026 年计划资本开支 2000 亿美元; Meta 在 2025 年 722 美元资本开支基础上, 2026 年计划资本开支 1150-1350 亿美元; 谷歌在 2025 年 914 美元资本开支基础上, 2026 年计划资本开支 1750-1850 亿美元; 甲骨文 2025 年 12 月预计 2026 财年资本开支超过 500 亿美元, 较 6 月预期值翻倍。

资本开支持续增长的背后, 云厂商的积压订单持续增长, 显示需求在持续扩大。2025Q4, 甲骨文剩余履约价值 (RPO) 同比大增 433% 至 5233 亿美元; 亚马逊订单积压同比增长 40%

至 2440 亿美元；谷歌云业务订单积压 2400 亿美元，同比增长超过一倍；微软 CRPO 为 6250 亿美元，同比增长 110%。2026Q1，甲骨文 RPO 达到 5530 亿美元，同比增长 325%，较上季度增加 290 亿美元。

表 9：海外科技厂商算力相关投入

| 公司 | 时间 | 详情 |
|---------------|--------------|--|
| 微软 | 2024. 3. 31 | 与 OpenAI 正计划建设一个名为“星际之门”的特殊数据中心，为 OpenAI 提供强大的计算支持，计划最快在 2028 年启动该项目，并且 2030 年之前会进一步扩建，所需要的电力可能高达 5 千兆瓦。“星级之门”超算会配备百万专用的服务器芯片，项目成本预计高达 1150 亿美元。 |
| | 2025. 1. 4 | 将在本财年（6 月结束）投资 800 多亿美元建设人工智能负载数据中心，以训练人工智能模型，并部署人工智能和基于云的应用程序。 |
| | 2025. 5. 1 | Q2 的资本投入绝对额会比本季有所提升，2026 财年资本支出会继续增长，只是同比增速会放缓。 |
| | 2025. 7. 31 | 考虑到对云和人工智能产品的强劲需求信号以及大量合同积压，公司将继续在资本支出和运营费用方面进行投资，预计 2026 财年第一财季的资本支出将超过 300 亿美元。与 25 财年相比，2026 财年资本支出增长将放缓，短期资产组合更多。 |
| | 2025. 9. 4 | 每年投入 750-800 亿美元 |
| | 2025. 10. 30 | 2026 财年第一财季的资本支出 349 亿美元。随着需求加速增长，公司在 GPU 和 CPU 方面的支出也在增加。因此，总支出将环比增长，预计 2026 财年的增长率将高于 2025 财年。 |
| OpenAI 甲骨文 | 2025. 1. 21 | 美国总统特朗普在白宫宣布，由美国甲骨文公司、OpenAI 和软银共同出资投建，三家公司组成的联合企业将投资 5000 亿美元，在美建设名为“星际之门”的 AI 基础设施。该项目计划在美国建设数据中心，以支持人工智能技术发展。 |
| | 2025. 5. 27 | 甲骨文计划斥资 400 亿美元采购英伟达的 GB200 芯片，以支持其在德克萨斯州阿比林市建设的“星际之门”项目中首个启动的数据中心，该中心预计将在 2026 年年中全面投入运营。 |
| | 2025. 7. 22 | OpenAI 与甲骨文达成协议，将共同在美国开发 4.5GW 新增“星际之门”数据中心算力。据介绍，结合德州阿比林市在建的首个“星际之门”基地，此次与甲骨文的合作将使 OpenAI 开发中的 AI 数据中心总容量突破 5GW，可支持超 200 万枚芯片运行。 |
| 甲骨文 | 2025. 6. 12 | 预计 2026 财年资本开支超过 250 亿美元。 |
| | 2025. 9. 10 | 预计 2026 财年资本开支大约为 350 亿美元。 |
| | 2025. 12. 11 | 预计 2026 财年资本开支大约为 500 亿美元。 |
| OpenAI | 2025. 9. 5 | The Information 报道，OpenAI 大幅上调了到 2029 年的预期现金消耗至 1150 亿美元，较半年前的预期激增了 800 亿美元。 |
| | 2025. 9. 22 | 与 NVIDIA 达成 10GW 的协议，并获得后者高达 1,000 亿美元的投资承诺。 |
| | 2025. 10. 6 | 与 AMD 签署 6GW 协议，后者将向 OpenAI 提供高达 1.6 亿股普通股的认股权证。 |
| | 2025. 10. 13 | OpenAI 与博通合作开发 10GW 规模的定制 AI 芯片和网络系统机架。 |
| | 2026. 1. 15 | OpenAI 与 Cerebras 联合部署 750MW 晶圆级系统，交易价值超过 100 亿美元，于 2028 年之前完成，将成为全球规模最大的高速 AI 推理平台。 |
| 亚马逊 | 2024. 10. 29 | 亚马逊正在核能上押下重注，包括在美国三个州投资超过 520 亿美元，以帮助推动其庞大数据中心的扩张。 |
| | 2024. 12. 16 | AWS 计划追加投资约 100 亿美元，在俄亥俄州扩建数据中心基础设施。新数据中心将包含计算机服务器、数据存储驱动、网络设备和其他形式的技术基础设施，用于支持包括 AI 和机器学习在内的云计算。 |

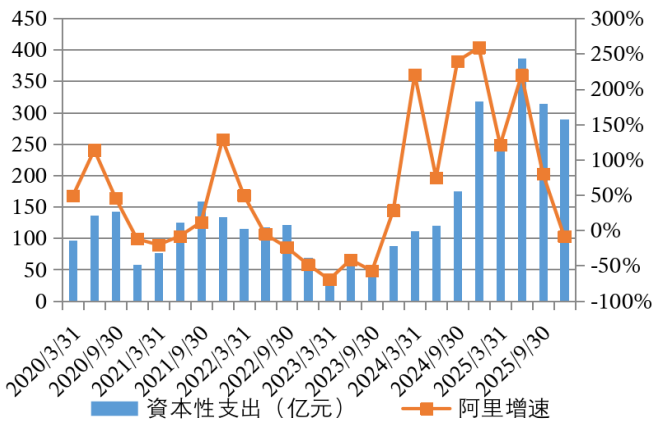
| | | |
|------|--------------|---|
| | 2025. 2. 6 | 预计将在 2025 年投入超过 1000 亿美元 ，高于 2024 年的 830 亿美元。 |
| | 2025. 6. 4 | 计划在美国北卡罗来纳州投资约 100 亿美元，扩展其位于当地的数据中心基础设施以支持 AI 和云计算技术。 |
| | 2025. 8. 1 | 2 季度资本支出高达 314 亿美元，其中大部分用于 AWS 的 AI 相关投资，这一投入水平将在今年剩余时间内维持相近水平 |
| | 2025. 10. 30 | 预计 2025 年全年资本支出将达到约 1250 亿美元 ，并将在 2026 年进一步增加 |
| | 2026. 2. 5 | 在现有业务需求强劲，以及 AI、芯片、机器人和近地轨道卫星等重大机会的推动下，预计 2026 年将投入约 2000 亿美元 的资本支出 |
| 谷歌 | 2024. 12. 11 | 谷歌与能源公司 Intersect Power 和投资公司 TPG Rise Climate 合作，计划在本十年内投资 200 亿美元 在美国建设多个“工业园区”，这些园区将配备可再生能源发电设施，以支持数据中心运营。第一个园区预计将在 2026 年部分投入运营，并于 2027 年全面完工。 |
| | 2025. 2. 4 | 计划在 2025 年投资约 750 亿美元 。 |
| | 2025. 7. 16 | 将在未来两年内投资 250 亿美元在美国最大的电力网区域建设数据中心，以支持其在人工智能领域的大规模布局。 |
| | 2025. 7. 23 | 将年度资本支出计划上调至约 850 亿美元 ，并预告明年将继续增加投入，公司表示这一决策源于其云计算服务的强劲需求。 |
| | 2025. 9. 4 | 两年内投入 2500 亿美元 |
| | 2025. 10. 30 | 预计 2025 年的资本支出将在 910 亿至 930 亿美元 之间。预计 2026 年资本开支将有显著增加 |
| | 2026. 2. 5 | 2026 年资本支出介于 1750 亿美元至 1850 亿美元 ，较 2025 年增长近一倍。资本支出用途，约 60% 用于服务器、其余 40% 用于数据中心和网络设备等长周期资产，比率与 2025 年相当。 |
| Meta | 2025. 1. 24 | 计划在 2025 年投入 600-650 亿美元 ，用于 AI 相关的资本支出，计划在 2025 年上线 1GW 的算力，预期到年底时能够拥有 130 万块 GPU。 |
| | 2025. 5. 1 | 下调了 2025 年全年总支出的预估范围，从 1140 亿美元至 1190 亿美元之间下调到 1130 亿美元至 1180 亿美元之间，其中用于生成式 AI 以及核心业务方面的投资范围从 600-650 亿美元之间上调到 640-720 亿美元 之间。 |
| | 2025. 7. 31 | 将大力推进人工智能基础设施建设，计划 2025 年相关资本支出（含融资租赁本金支付）达 660-720 亿美元 ，其中值较上年同比增长约 300 亿美元。这一积极的投资增长态势将延续至 2026 年，以持续提升在线产能，满足人工智能研发及业务运营的需求。 |
| | 2025. 8. 27 | 特朗普表示 Meta 计划在路易斯安娜州乡村兴建大型数据中心项目，总投资将达到 500 亿美元 。 |
| | 2025. 9. 4 | 到 2028 年将在美投资 6000 亿美元 。 |
| | 2025. 10. 30 | 将 2025 年资本支出指引上调至 700-720 亿美元 ，预计 2026 年资本支出的绝对额增长将显著高于 2025 年 |
| | 2026. 1. 29 | 2025 年资本支出为 722.2 亿美元 ，预计 2026 年将明显增加，预计将在 1150-1350 亿美元 之间，同比增长主要源于为支持 Meta 超级智能实验室的工作和核心业务而增加的投资。 |
| 苹果 | 2025. 9. 4 | 承诺投资 6000 亿美元 。 |
| xAI | 2026. 1. 9 | 计划斥资超 200 亿美元 ，在密西西比州绍斯黑文市兴建一座数据中心。 |

资料来源：财联社，科创板日报，福布斯中国，第一财经，IT 之家，新智元，金融界，环球网，The Information，同花顺，中原证券研究所

1.4.6.2. 国内资本开支：2025 年仍受制于芯片供给，整体需求有望更好地释放

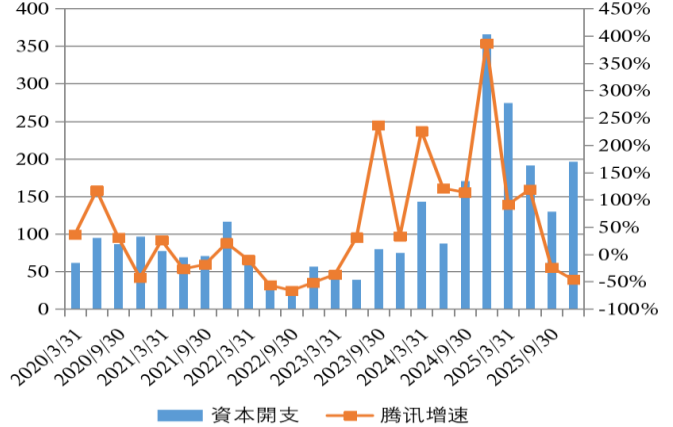
对比海外来看，国内科技厂商在 Q4 都出现了资本开支同比下滑，芯片供应短缺问题较为突出。腾讯 Q4 资本开支 196.32 亿元，同比下滑 46%，环比增长 51%，3 季度见底后回升。阿里 Q4 资本开支 290 亿元，同比下滑 8.7%，环比下滑 7.9%，受到芯片供应的影响，也出现了环比下滑。但是阿里表示考虑到推理需求爆发式增长，此前 3 年 3800 亿元的投资可能显得偏小，会在确保财务健康的前提下，动态调整资本开支节奏。

图 59：2020-2025 年阿里单季度资本开支及增速



资料来源：阿里财报，中原证券研究所

图 60：2020-2025 年腾讯单季度资本开支及增速 (亿元)



资料来源：腾讯财报，中原证券研究所

根据澎湃新闻，字节跳动在 2025 年下半年大幅上调了 AI 相关投入，覆盖高端 AI 芯片采购、底层模型研发等多个核心环节，巨额开支直接导致全年净利润同比下滑超 70%。2026 年技术资源投入会进一步扩大，从 2025 年的 1500 亿元增长到约 1600 亿元，其中 850 亿用于 AI 芯片采购，750 亿用于数据中心及配套设施建设，其中海外投入金额为 500 亿。

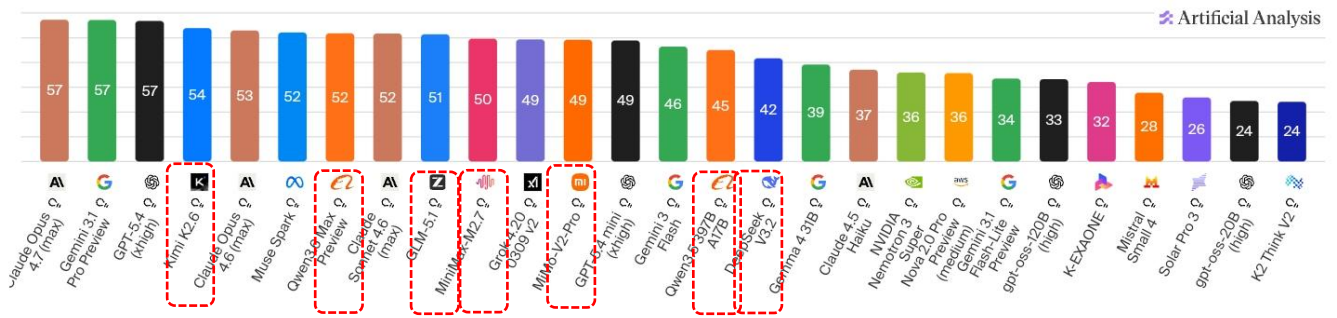
1.5. AI: AI 应用需求爆发式增长，带来算力整体短缺

1.5.1. 头部模型概况

1.5.1.1. 头部模型的跑分对比

根据 Artificial Analysis 的排名数据来看，随着 Meta 新模型的发布，目前国内模型中 Kimi 的 K2.6 暂居第一，阿里 Qwen3.6 Max Preview 紧随其后，智谱的 GLM-5.1、MiniMax 的 M2.7、小米的 MiMo-V2-Pro 位居国产模型第三到第五名，DeepSeek 的 V3.2 和其他模型的差距仍然较为明显。考虑到 DeepSeek V4 有 4 月下旬发布预期，DeepSeek 或能在模型能力上有较大提升。

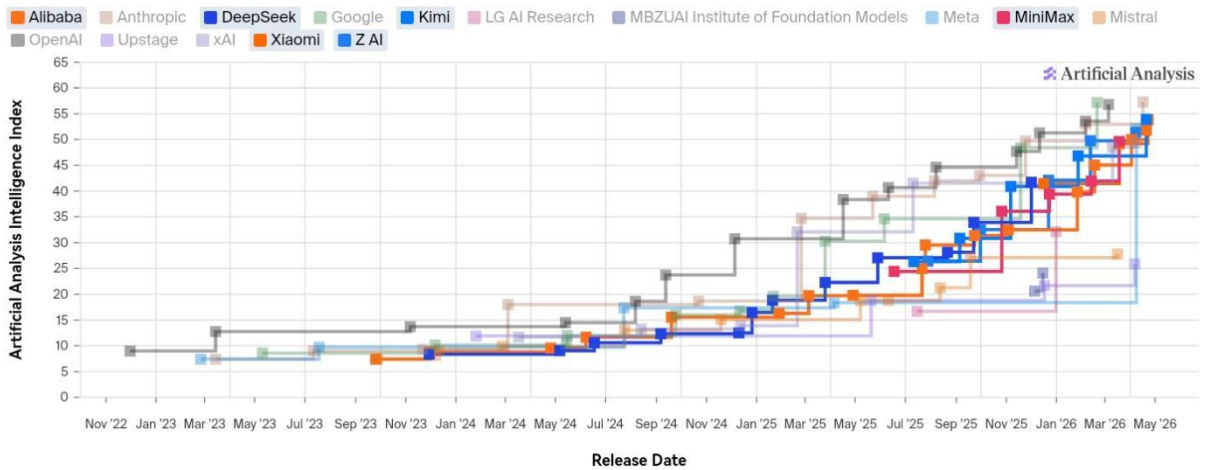
图 61：大模型智能水平（截至 2026. 4. 23）



资料来源：Artificial Analysis，中原证券研究所

目前国内模型在能力上主要集中在第二梯队，随着厂商新模型的陆续发布，整体水平上更加集中。

图 62：大模型智能水平随时间增长的情况截至（2026. 4. 23）

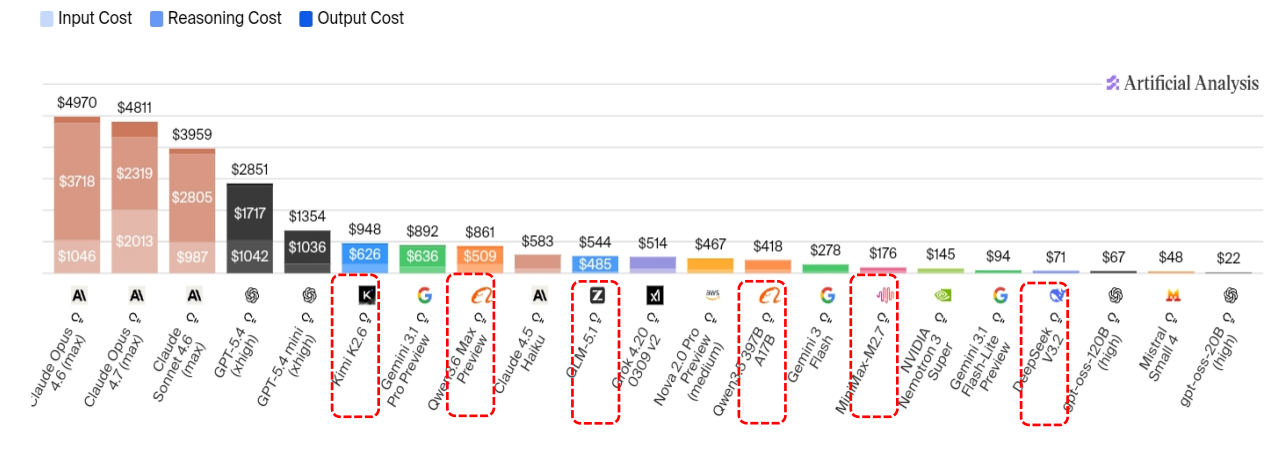


资料来源：Artificial Analysis，中原证券研究所

1.5.1.2. 模型价格：国产厂商的优势明显

从当前模型的成本情况来看，目前海外模型价格普遍较高，其中 Claude 的多个模型都明显高于其他竞争对手。由于在编程领域的不可或缺性，Claude 也借此成就了 2026 年营收的高增长，而国产头部模型的价格不足 Claude 的 1/5。

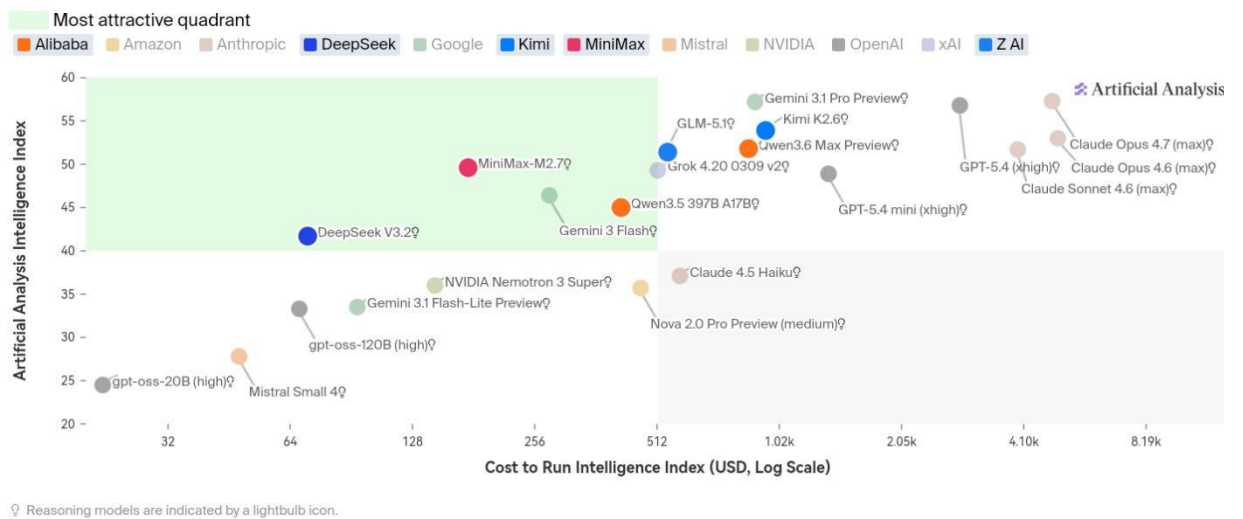
图 63：大模型价格对比（2026. 4. 23）



资料来源：Artificial Analysis，中原证券研究所

国产在性能持续提升的同时，性价比优势更加凸显。由于国内始终处于算力短缺状态，国产模型更加关注在软件调优层面的模型调优，催生出以 DeepSeek 为代表的国产突破，也成就了国产模型高性价比的特征。从 Artificial Analysis 的对比数据来看，国产模型都位于模型性能与价格对比图的对角线上方，高性价比优势明显。

图 64：大模型价格与智能的综合对比（2026. 4. 23）

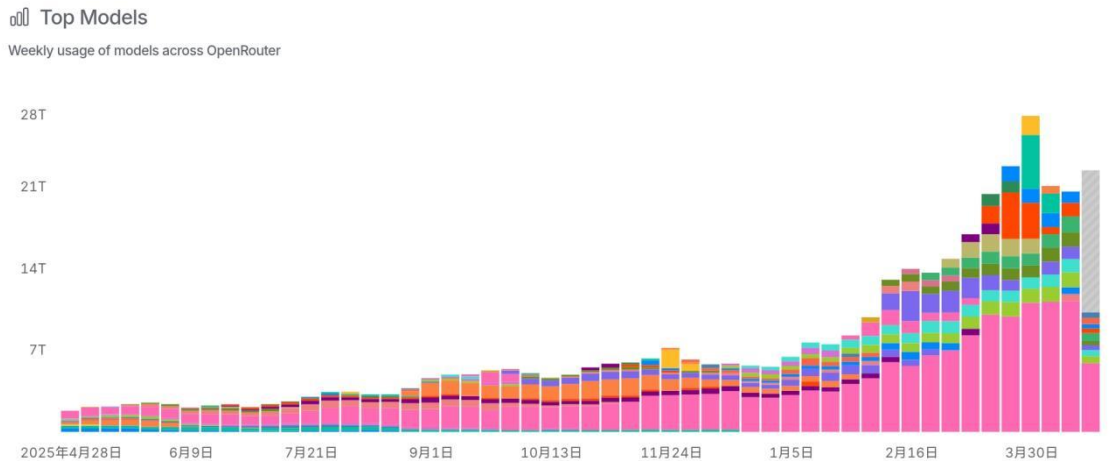


资料来源：Artificial Analysis，中原证券研究所

1.5.2. 模型调用及原生应用

OpenRoute 中模型的每周使用数据来看，截至 4 月 23 日，模型整体调用峰值出现在 4 月第一周，而而后的 2 周中略有回落，最近一周有望回升。

图 65: OpenRouter 中模型的每周使用情况 (截至 4 月 23 日)



资料来源: OpenRouter, 中原证券研究所

1.5.2.1. 海外原生应用: Anthropic 高净值用户数量 2 个月翻倍

在用户数据方面, OpenAI 依然拥有最多活跃用户, 但是 2026 年 Anthropic 的增长弹性更高。根据 2026 年 4 月 OpenAI 披露的数据, ChatGPT 周活跃用户数超过 9 亿, 其中付费订阅用户 5000 万, 900 万个付费企业。2 月 Gemini 应用程序月活跃用户突破 7.5 亿, 较 OpenAI 的差距进一步缩小。Anthropic 年支出超 100 万美元的企业客户数从 2 月的超 500 家增长到 4 月超 1000 家, 业务增长弹性在 2026 年充分显现出来。

表 10: AI 应用用户数增长情况

| 公司 | 时间 | 详情 |
|-------------|--|--|
| 微软 | 2023. 3 | Bing 日活跃用户达到了 1 亿 |
| | 2025. 10. 30 | 旗下所有产品中, 人工智能功能的月活跃用户已达 9 亿 |
| | 2026. 1. 28 | 面向消费者的 Copilot 产品日活同比增长接近 3 倍 |
| OpenAI | 2023. 1 | ChatGPT 月活用户突破 1 亿 |
| | 2023. 11 | ChatGPT 周活跃用户 1 亿 |
| | 2024. 8 | ChatGPT 周活跃用户 2 亿 |
| | 2024. 12 | ChatGPT 周活跃用户 3 亿 |
| | 2025. 2 | ChatGPT 周活跃用户 4 亿 |
| | 2025. 4 | ChatGPT 周活跃用户突破 5 亿 |
| | 2025. 7 | ChatGPT 周活跃用户数突破 7 亿, 用户每周发送的消息总量高达 180 亿条 |
| | 2025. 10 | ChatGPT 周活跃用户数突破 8 亿 |
| | 2025. 11 | OpenAI 达到 8 亿周活跃用户 (WAU), 约 1.9 亿日活跃用户 (DAU), 3500 万付费订阅者, 100 万企业客户 |
| | 2025. 12 | ChatGPT 周活跃用户数接近 9 亿 |
| 2026. 2. 27 | ChatGPT 周活跃用户已突破 9 亿, 个人订阅用户超过 5000 万, 已拥有 900 万个企业付费席位。Codex 周活跃用户达到 160 万, 2026 年初以来增至 3 倍多。 | |
| 谷歌 | 2024. 10 | Gemini 用户月活用户 9000 万 |
| | 2025. 3 | Gemini 用户月活用户已达 3.5 亿 |

| | | |
|-----------|---------|---|
| | 2025.7 | Gemini 应用的月活跃用户已突破 4.5 亿大关 |
| | 2025.10 | 每月处理的 token 量从 7 月的 980 万亿激增至超过 1300 万亿, 一年内增长超过 20 倍。 Gemini 应用程序月活跃用户突破 6.5 亿, 查询量比第二季度激增 3 倍。 |
| | 2026.2 | Gemini 月活跃用户突破 7.5 亿 |
| Anthropic | 2026.2 | 年支出超 100 万美元的企业客户数已突破 500 家 |
| | 2026.4 | 年支出超 100 万美元的企业客户数已突破 1000 家 |

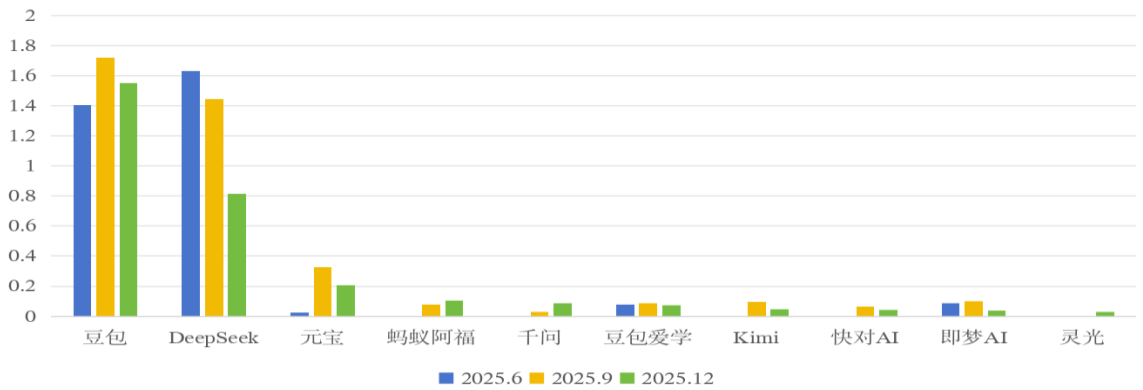
资料来源: 微软, OpenAI, 谷歌, 中原证券研究所

1.5.2.2. 国内原生应用: 豆包日均 token 使用量突破 120 万亿

国家数据局的数据, 中国日均词元 (Token) 调用量 3 月已突破 140 万亿, 这一数字较 2024 年初的 1000 亿和 2025 年底的 100 万亿都有明显的提升。

根据 QuestMobile 的数据, 12 月国内月活跃用户最高的 AI 原生 APP 分别为豆包 1.55 亿、DeepSeek 0.82 亿、元宝 0.21 亿、蚂蚁阿福 0.10 亿、千问 0.09。由于阿里在 Q4 加大了新产上线和推广的力度, 蚂蚁阿福和千问都步入了前 5 大 AI 原生 App。

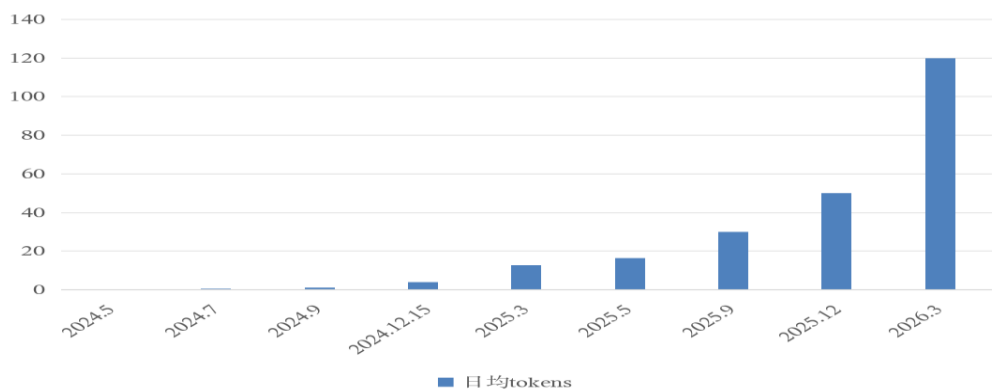
图 66: 2025 年 12 月 AI 原生 APP 月活跃用户规模 (亿)



资料来源: QuestMobile, 中原证券研究所

截至 2026 年 3 月, 豆包大模型的日均 token 使用量突破 120 万亿, 较 2025 年 12 月增长了 140%。

图 67: 豆包大模型调用量增长情况 (万亿)



资料来源: 字节, 中原证券研究所

截至 2 月底，近 1.4 亿用户通过千问 App 的智能体功能首次体验 AI 购物；2 月千问 C 端应用月活跃用户数超过 3 亿。

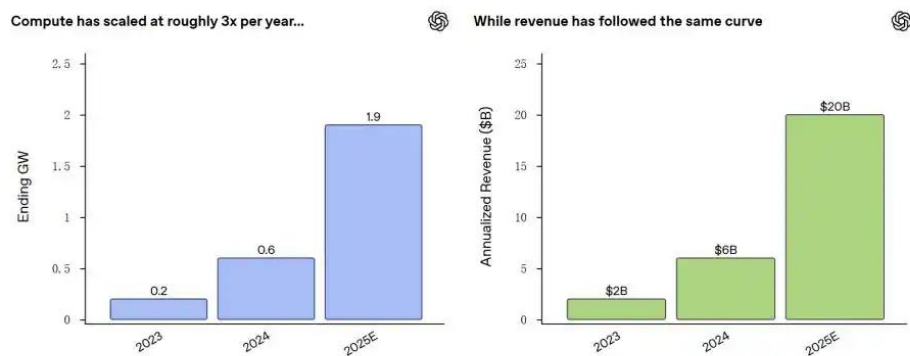
1.5.3. 模型厂商业绩

1.5.3.1. 美国模型厂商：3 家厂商积极寻求 2026 年上市

OpenAI 实现了 250 亿美元年化收入，算力约 1.9GW。OpenAI 年化收入从 2023 年的 20 亿美元增长到 2024 年的 60 亿美元，2025 年预计增长到 200 亿美元，三年增长了 10 倍，2026 年 4 月进一步增长到了 250 亿美元，单月最高收入突破 20 亿美元。OpenAI 算力从 2023 年的 0.2GW 增长到 2024 年的 0.6GW，2025 年预计 1.9GW 增长约 9.5 倍。

OpenAI 计划上市，目前仍然有较大的盈利压力。2025 年，OpenAI 全年收入 131 亿美元，亏损 80 亿美元，2026 年预计将亏损 250 亿美元，计划在 2026Q4 上市。2026 年 1 月 16 日 OpenAI 宣布在 ChatGPT 免费版和 Go 版本中投放广告。据 The Information 报道，OpenAI 当前 9 亿的用户中付费率仅 5%，近 90% 的用户位于美国、加拿大之外的区域，其平均收入贡献远低于北美，广告价值较低，给 OpenAI 以广告为核心的收入模式带来压力。在运营成本压力下，2 月 OpenAI 与五角大楼达成合作，3 月 OpenAI 宣布将于 4 月 26 日关停 Sora。

图 68：OpenAI 的年化收入和算力增长情况



资料来源：OpenAI，新智元，中原证券研究所

Anthropic 年化收入 300 亿美元，2026 年收入增速明显加快。Anthropic 2025 年年化收入从年初 10 亿美元跃升至年底逾 90 亿美元，2 月 13 日达到 140 亿美元，4 月 7 日达到 300 亿美元，已经超过 OpenAI。在营收方面，2024 年为 3.81 亿美元，2025 年营收达 45 亿美元，略低于此前近 47 亿美元的乐观预期。同时，Anthropic 最快有望在 2026 年开启 IPO 事宜。2 月 13 日，Anthropic 完成最新 300 亿美元的 G 轮融资，投后估值达到 3800 亿美元。

2026 年 2 月 2 日，马斯克旗下 SpaceX 宣布完成对 xAI 的收购，此次收购核心目标是打通“数据-AI 训练-太空算力-太空通信”全闭环，合并后的公司已经在 3 月递交了上市申请，或进

行 400-800 亿美元融资，估值有望达到 1.75 万亿美元。同时，4 月 21 日，SpaceX 表示 Cursor 已授予其 2026 年早些时候以 600 亿美元直接收购该公司的权利。

表 11：大模型厂商融资及估值情况

| 厂商 | 时间 | 融资金额 | 估值 | 其他 |
|-----------|-------------|----------|----------|---|
| Anthropic | 2026. 2. 13 | 300 亿美元 | 3800 亿美元 | G 轮融资，新加坡主权基金 GIC 和对冲基金 Coatue 领投，微软、英伟达、亚马逊和谷歌等科技巨头跟投 |
| OpenAI | 2015 | 10 亿美元 | | 成立 |
| | 2019. 7 | 10 亿美元 | | 微软投资，双方建立独家云计算合作关系，微软成为 OpenAI 的独家许可方，OpenAI 使用微软 Azure 云计算平台来开发和托管其产品，双方还同意共同构建新的 Azure AI 超级计算技术。 |
| | 2021. 1 | 20 亿美元 | | 微软再次投资，签署长期合作协议，随后微软推出 AI 编程助手 GitHub Copilot 和 Azure OpenAI 服务 |
| | 2022. 11 | 130 亿美元 | | 微软投资 |
| | 2024. 10 | 66 亿美元 | | |
| | 2025. 3 | 400 亿美元 | 3000 亿美元 | 从软银等机构融资，如果 OpenAI 在 2025 年内没有从一家营利性公司转型为一家公益公司，软银可以将其总投资从 400 亿美元减 50% 至 200 亿美元。 |
| | 2026. 2. 28 | 1100 亿美元 | 8400 亿美元 | 亚马逊投入 500 亿美元，英伟达和软银各投 300 亿美元。 |
| | 2026. 4. 1 | 1220 亿美元 | 8520 亿美元 | 由亚马逊、英伟达与软银领投，软银与风投机构 a16z 担任本轮融资的联合牵头人。其中个人投资者筹集超过 30 亿美元。 |

资料来源：微软，OpenAI，谷歌，中原证券研究所

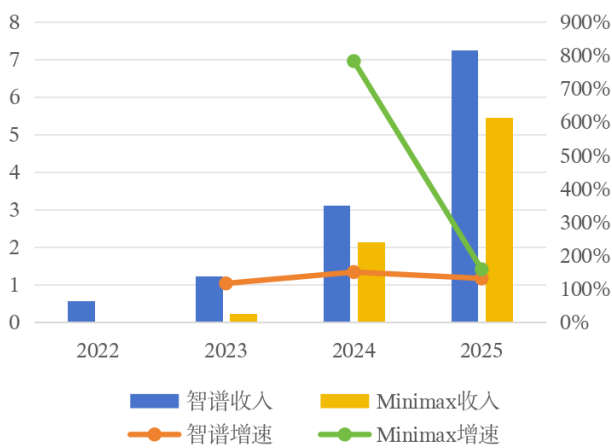
1.5.3.2. 中国模型厂商：2025 年智谱、Minimax 收入获得翻倍增长

智谱和 MiniMax 1 月在港股上市后，首次公布了年报。

营收：截至 2025 年，智谱和 MiniMax 收入分别为 7.24 亿元和 5.46 亿元，同比分别增长了 132% 和 159%，MiniMax 增速更快。

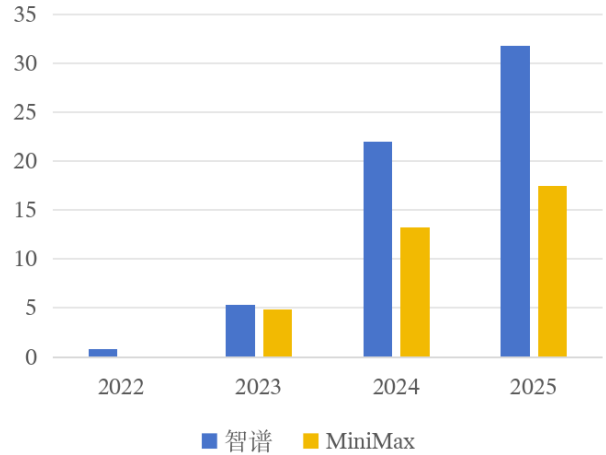
研发费用：由于两家公司还在业务投入期，在研发方面还有较大的支出需求。2025 年智谱和 MiniMax 的研发费用分别为 31.80 亿元和 17.46 亿元，分别同比增长了 45% 和 32%，MiniMax 的研发费用控制相对较好。

图 69：2022-2025 智谱与 MiniMax 收入及增速（亿元）



资料来源：上市公司公告，中原证券研究所

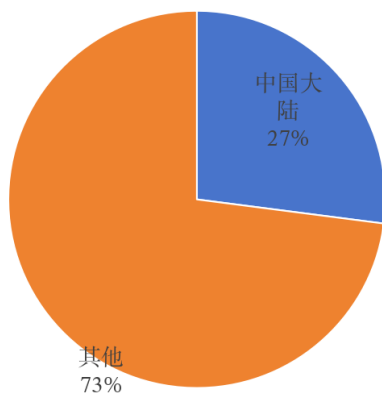
图 70：2022-2025 智谱与 MiniMax 研发费用（亿元）



资料来源：上市公司公告，中原证券研究所

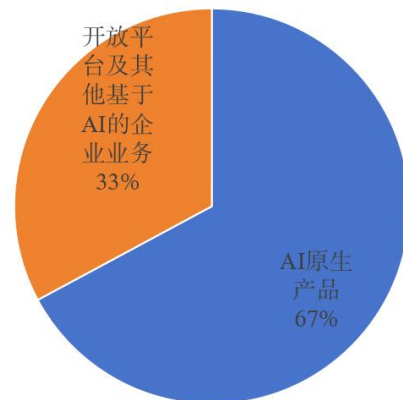
从 MiniMax 2025 年收入来看，其收入中仅有 27% 来自中国大陆，另有新加坡、美国等作为主要区域的海外收入占比达到了 73%，海外收入增速达到 167%，给公司带来了强劲增长动力。其产品收入结构来看，67% 来自 AI 原生产品，33% 来自开放平台及其他基于 AI 的企业业务。后者 1-9 月毛利率高达 69%，而前者在 2025 年毛利才转正，1-9 月毛利率达到 5%。

图 71：2025 年 MiniMax 分区域收入结构



资料来源：MiniMax，中原证券研究所

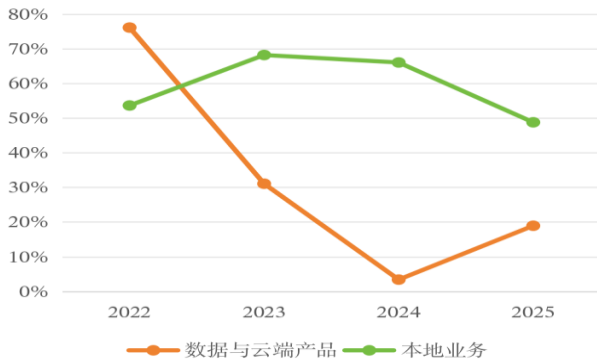
图 72：2025 年 MiniMax 分产品收入结构



资料来源：MiniMax，中原证券研究所

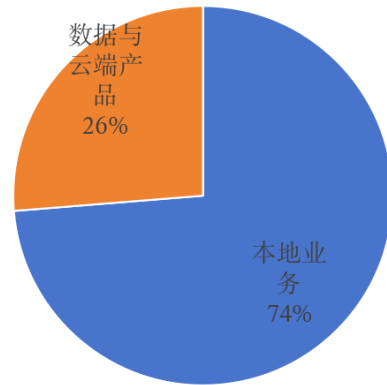
2025 年，智谱收入中 74% 来自本地业务，26% 来自数据与云端产品，毛利率分别为 49%、19%，而前者毛利率仍然呈现下滑趋势。

图 73：2025 智谱分产品毛利率



资料来源：智谱，中原证券研究所

图 74：2025 智谱分产品收入结构



资料来源：智谱，中原证券研究所

2026 年 1 月底以后，受 Kimi K2.5 模型及 Kimi Claw 爆火的双重影响，Kimi 20 天累计收入已超 2025 年全年总收入。增长主要受全球付费用户及 API 调用量大涨共同驱动，特别是海外付费用户保持高速增长，Kimi 海外收入已超国内。据界面新闻 3 月 30 日报道，K2.5 发布仅一个月，月之暗面 ARR（年度经常性收入）即突破 1 亿美元。

近期关于 DeepSeek 融资的新闻也获得了市场的关注。根据智东西报道，对于 DeepSeek 估值快速地从 100 亿美元提升到了 200 亿美元，同时阿里、腾讯也洽谈投资过程中。

根据澎湃新闻，字节跳动因为 AI 业务投入，2025 年净利润同比下滑超过 70%。

表 12：大模型厂商融资及估值情况

| 厂商 | 时间 | 融资金额 | 估值 | 其他 |
|----------|-------------|----------|-----------|--|
| DeepSeek | 2026. 4. 17 | 至少 3 亿美元 | 超 100 亿美元 | |
| | 2026. 4. 22 | | 超 200 亿美元 | 阿里、腾讯洽谈投资 |
| Kimi | 2023. 6 | | 3 亿美元 | 天使轮 |
| | 2026. 12 | 5 亿美元 | | C 轮 |
| | 2026. 2. 17 | 7 亿美元 | 超百亿美元 | 由阿里、腾讯、五源、九安等联合领投，创下国内企业从成立到估值超 100 亿美元的最快晋级速度 |
| | 2026. 3 | 10 亿美元 | 180 亿美元 | |

资料来源：Kimi，The Information，智东西，中原证券研究所

1.5.4. 涨价情况：AI 应用需求爆发式增长，拉动算力价格持续上涨

随着 AI 应用需求的持续增长，以智谱为代表的模型厂商进行了多轮提价，价格体系逐步看齐海外头部厂商，同时以阿里云、腾讯云、百度智能云为代表的云计算厂商也接连多轮涨价，来平衡硬件端原材料价格上涨带来的开支压力。

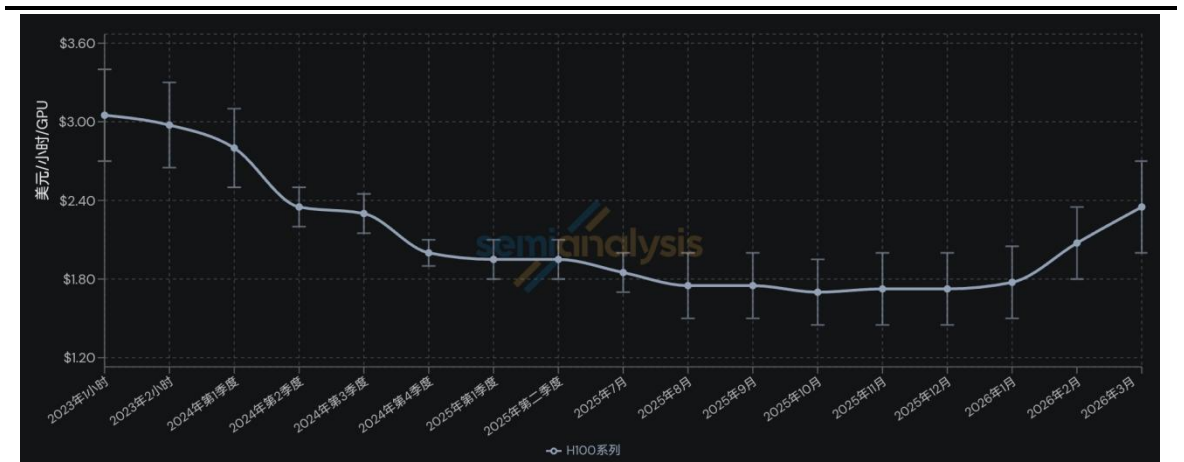
表 13: 模型厂商和云计算涨价情况

| 公司 | 时间 | 产品 | 涨价幅度 | 生效日期 |
|---------------------------|-----------------|-------------------------------|---------------------|-------|
| 阿里云 | 2026. 3. 18 | 平头哥真武 810E 等算力卡相关服务 | 5%-34% | 4月18日 |
| | | 文件存储产品 CPFS (智算版) | 30% | |
| | 2026. 4. 11 | Coding Plan Lite 基础套餐 | 停止续费和升级 | 4月13日 |
| | 2026. 4. 13 | DataWorks 标准版 | 调用 API 免费额度调整 | 4月23日 |
| | | DataWorks 专业版 | 调用 API 免费额度调整 | |
| | 2026. 4. 15 | 部分 MU (Model Unit) 模型单元的服务 | 2%-5% | 5月15日 |
| | | DDoS 原生防护 2.0 (包年包月) | 月 95:20%, 日 95:-50% | 7月15日 |
| DDoS 高防 (中国内地) | | 月 95:50%, 日 95:33% | | |
| | DDoS 高防 (非中国内地) | 25%-50% | | |
| 腾讯云 | 2026. 3. 11 | 第三方模型 GLM5、MiniMax2.5、Kimi2.5 | 结束限时免费公测 | 3月13日 |
| | | 混元系列模型 Tencen HY2.0 Instruct | 输入 463%, 输出 456% | |
| | | 混元系列模型 Tencen HY2.0 Think | 输入 430%, 输出 5200% | |
| | 2026. 4. 9 | AI 算力相关产品服务 | 5% | 5月9日 |
| | | 容器服务 TKE-原生节点相关产品服务 | 5% | |
| 弹性 MapReduce (EMR) 相关产品服务 | | 5% | | |
| 百度智能云 | 2026. 3. 18 | AI 算力相关产品服务 | 5%-30% | 4月18日 |
| | | 并行文件存储等 | 30% | |
| 智谱 | 2026. 2. 12 | GLM Coding Plan | 30%起 | |
| | 2026. 3. 16 | GLM-5-Turbo API | 20% | |
| | 2026. 4. 8 | OpenRouter GLM 系列 API | 10% | |
| | 2026. 4. 11 | 智谱 Coding Plan (海外版) | 80%-150% | |

资料来源: 智谱, 阿里云, 百度云, 腾讯云, 中原证券研究所

根据 SemiAnalysis 发布的数据, H100 一年期租赁合同价格从 2025 年 10 月的低点 1.70 美元/小时/GPU 飙升至 2026 年 3 月的 2.35 美元/小时/GPU, 涨幅达到了 38%。

图 75: H100 GPU 租赁价格趋势 (1 年期合同)



资料来源: SemiAnalysis, 中原证券研究所

1.5.5. 科技企业裁员：2026 年裁员规模明显提升

2026 年海外主要科技厂商都出现了较多的裁员，其中亚马逊约 3 万人，Meta 约 9500 人，微软约 1.1-2.2 万人，甲骨文约 3 万人。由于算力采购资本开支持续加重，AI 带来的编程效率的持续提升，科技厂商还有持续削减人员开支的趋势。

表 14：海外主要科技厂商近期裁员情况

| 厂商 | 时间 | 人数 | 备注 |
|------|------------|-----------|---|
| 亚马逊 | 2025.1 | 约 200 | 在北美商店部门裁员约 200 人，涵盖包括自有品牌、Prime 会员计划和消费品业务部门 |
| | 2025.10.28 | 约 14000 | 将裁减约 1.4 万个企业岗位，占其约 35 万名企业员工的 4%，此次裁员旨在提高决策效率，并将资源转向核心业务 |
| | 2026.1.28 | 约 16000 | 宣布调整组织架构，确认将裁减约 1.6 万个工作岗位。 |
| | 2026.4.7 | 约 14000 | 计划在 2026 年 5 月启动新一轮全球裁员，预计约 1.4 万名员工将受到影响。此次裁员覆盖 AWS 云服务、零售、人力资源等多个核心业务板块，中国区部分团队甚至可能被整体裁撤。 |
| Meta | 2025.1 | 3600 | 计划裁员约 5%，对象是“表现不佳的员工” |
| | 2025.10.22 | 600 | 将在其人工智能部门裁员约 600 人，希望通过此举减少层级，提高运营灵活性。 |
| | 2026.1 | 1500 | Meta 计划对元宇宙业务所在的 Reality Labs 部门裁员 10%，涉及大约 1500 人。 |
| | 2026.4.18 | 8000 | 计划于今年 5 月 20 日启动一轮影响全球 10% 员工的裁员，相当于近 8000 人 |
| 谷歌 | 2023.3 | 12000 | 谷歌母公司 Alphabet 全球裁员 6% |
| | 2025.2 | | 对人力运营与云计算部门进行了裁员，并向美国员工提供自愿离职计划 |
| | 2025.3 | | 对云计算团队进行了优化。裁减了运营支持岗位，将部分职位转移至印度、墨西哥城等低成本地区。 |
| | 2025.4 | 数百 | 在安卓系统与 Pixel 手机业务线裁员数百人，涉及 Chrome 开发、硬件工程及产品管理等核心岗位 |
| 微软 | 2023 | 超 10000 | |
| | 2025.1 | | 裁减安全、体验和设备、销售和游戏等部门的员工 |
| | 2025.5 | 约 6000 | 在全球范围内裁减不到 3% 的员工 |
| | 2025.7 | 约 9000 | 为了简化业务流程，减少管理层级。裁员人数不到公司全球总员工数的 4%，涉及不同部门、地区以及各个经验层级的员工。 |
| | 2026.1 | 1.1-2.2 万 | 计划于 1 月第三周启动新一轮全球裁员，规模预计在 1.1 万至 2.2 万人之间，占其全球 22 万名员工总数的 5% 至 10%。 |
| 甲骨文 | 2026.3.31 | 约 3 万 | 甲骨文在全球范围内启动突袭式大规模裁员，总计约 3 万名员工受波及。 |

资料来源：光明网，彭博社，极目新闻，新浪财经，中原证券研究所

二、河南计算机行业动态

1.6. 河南计算机行业数据跟踪

一季度，全省规模以上高技术制造业、工业战略性新兴产业增加值分别同比增长 23.4%、14.3%，比上年同期分别加快 9.3 个、3.7 个百分点。高技术制造业投资增长 4.5%。限额以上单位可穿戴智能设备、计算机及其配套产品零售额分别增长 67.4%、33.8%。

一季度，信息传输、软件和信息技术服务业 575.78 亿元，同比增长 4.0%。

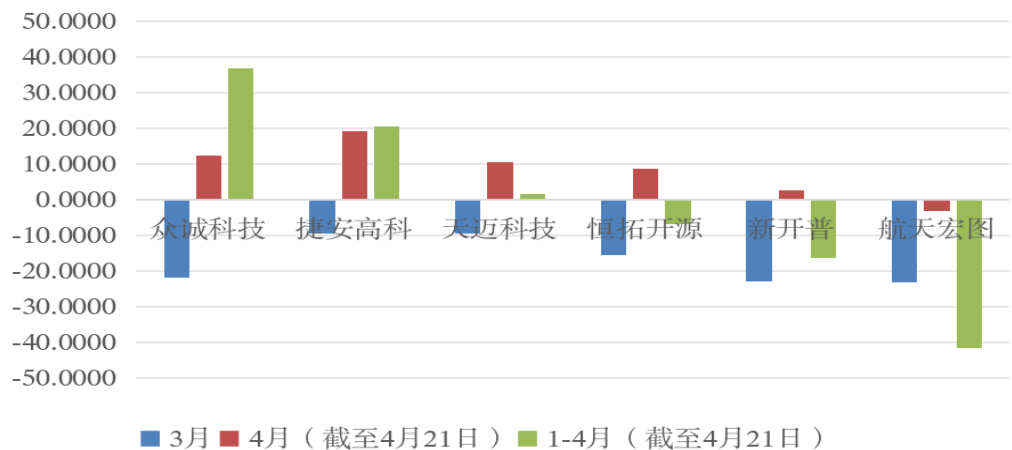
计算机、通信和其他电子设备制造业增长 13.6%，电子计算机整机 23.53 万台，同比增长 0.8%；智能手机 928.96 万台，同比增长 9.4%。

(数据来源：河南省统计局)

1.7. 河南上市公司行情回顾

2026 年以来河南计算机行业上市公司多数上涨。2026 年以来（截至 4 月 21 日）河南 6 家计算机公司中 3 家上涨、3 家下跌，其中众诚科技、捷安高科、天迈科技分别上涨 36.75%、20.41%、1.50%，航天宏图、新开普、恒拓开源分别下跌 41.58%、16.28%、6.88%。

图 76：河南上市公司近期股价涨跌幅表现 (%)



资料来源：Wind，中原证券研究所

2. 计算机行业行情表现

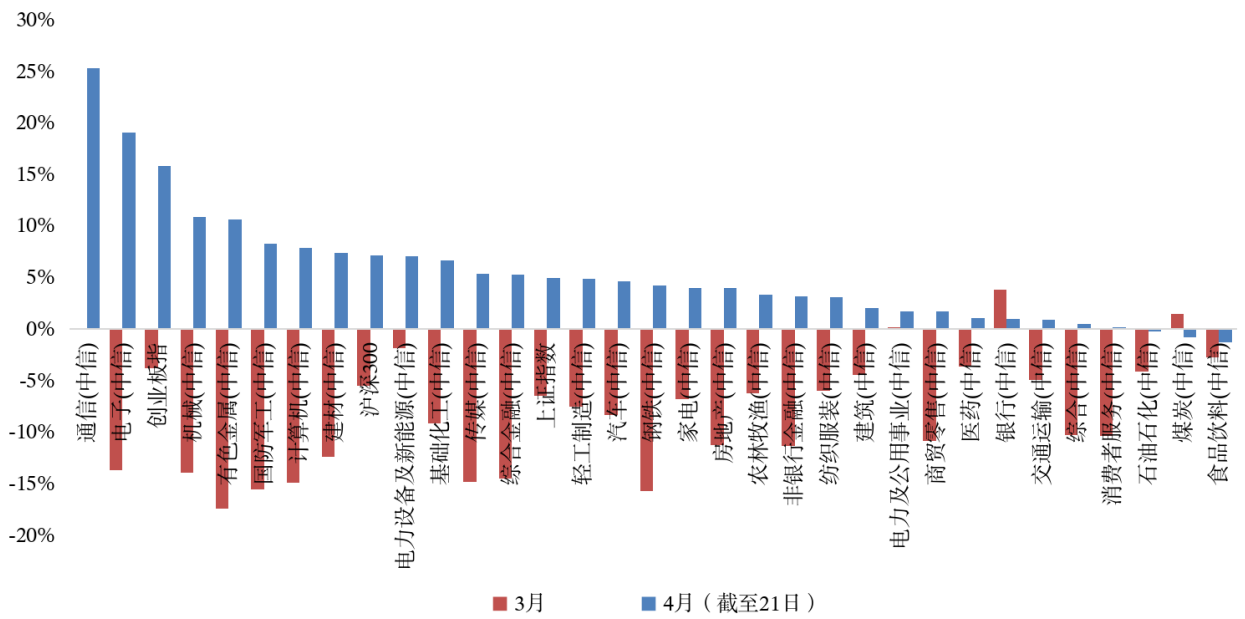
2.1. 行情回顾：国际局势影响下，行业剧烈调整后，4月步入反弹走势

在国际局势的影响下，3月行业调整幅度居前，4月整体呈现出反弹走势。

3月中信计算机指数下跌了 14.91%，跑输上证指数 8.40 PCT，跑输创业板指数 11.12 PCT，跑输沪深 300 指数 9.38 PCT，在 30 个中信一级行业中排名第 27。

4月（截至 21 日）中信计算机指数上涨了 7.81%，跑赢上证指数 2.85 PCT，跑输创业板指数 8.40 PCT，跑赢沪深 300 指数 0.67 PCT，在 30 个中信一级行业中排名第 6。

图 77：2026 年 3 月中信一级子行业涨跌幅

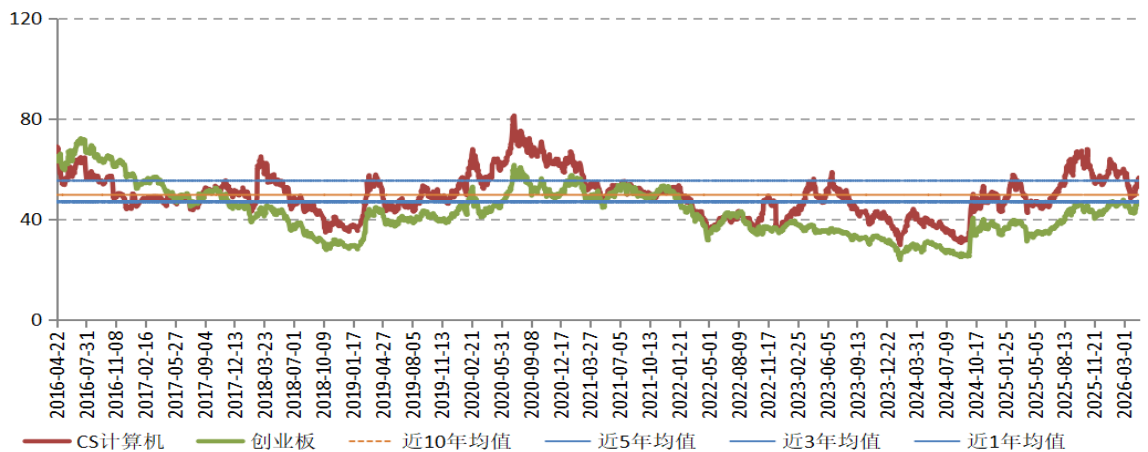


资料来源：Wind，中原证券研究所

2.1. 估值：行业的估值超过历史均值水平

行业的估值超过历史均值。根据 Wind 数据，2026 年 4 月 21 日中信计算机行业 TTM 整体法（剔除负值）估值为 55.77 倍，创业板估值 46.83 倍。行业近 1 年、3 年、5 年、10 年的平均估值分别为 55.36 倍、47.24 倍、46.71 倍、49.72 倍，当前估值已经超过历史均值水平，近 3 年最高点 67.87 倍和最低点 29.69 倍。

图 78：近 10 年中信计算机行业估值水平（截至 2026.4.21）



资料来源：Wind，中原证券研究所

2.2. 计算机行业 ETF 行情回顾

3月行业ETF皆呈现下跌行情，其中9只人工智能ETF领跌，跌幅在20%以上。

表 15：2026 年 3 月计算机行业 ETF 行情回顾

| 证券代码 | 证券简称 | 业绩比较基准 | 区间涨跌幅 (%) | 区间净值超越基准收益率 (%) | 区间日均成交额 (百万元) |
|-----------|--------------------|-------------------|-----------|-----------------|---------------|
| 159778.SZ | 鹏华中证工业互联网主题ETF | 中证工业互联网主题指数收益率 | -8.68 | 0.11 | 25 |
| 159738.SZ | 华泰柏瑞中证沪港深云计算产业ETF | 中证沪港深云计算产业指数收益率 | -8.87 | 0.08 | 554 |
| 517390.SH | 天弘中证沪港深云计算产业ETF | 中证沪港深云计算产业指数收益率 | -8.94 | 0.02 | 258 |
| 159273.SZ | 汇添富中证沪港深云计算产业ETF | 中证沪港深云计算产业指数收益率 | -9.10 | -0.15 | 692 |
| 159243.OF | 招商创业板人工智能ETF | 创业板人工智能指数收益率 | -9.74 | 0.19 | 661 |
| 560660.SH | 新华中证云计算50ETF | 中证云计算50指数收益率 | -9.81 | 0.18 | 350 |
| 159242.SZ | 大成创业板人工智能ETF | 创业板人工智能指数收益率 | -9.84 | 0.08 | 1,299 |
| 159279.SZ | 华安创业板人工智能ETF | 创业板人工智能指数收益率 | -9.86 | 0.07 | 968 |
| 159381.SZ | 华夏创业板人工智能ETF | 创业板人工智能指数收益率 | -9.90 | 0.03 | 7,834 |
| 159363.SZ | 华宝创业板人工智能ETF | 创业板人工智能指数收益率 | -9.90 | 0.02 | 14,384 |
| 159388.SZ | 国泰创业板人工智能ETF | 创业板人工智能指数收益率 | -9.92 | 0.01 | 2,431 |
| 159246.SZ | 富国创业板人工智能ETF | 创业板人工智能指数收益率 | -9.92 | 0.00 | 2,359 |
| 159382.SZ | 南方创业板人工智能ETF | 创业板人工智能指数收益率 | -9.97 | -0.04 | 2,408 |
| 159248.SZ | 万家中证人工智能主题ETF | 中证人工智能主题指数收益率 | -11.02 | 0.15 | 350 |
| 515070.SH | 华夏中证人工智能ETF | 中证人工智能主题指数收益率 | -11.15 | 0.02 | 4,229 |
| 515980.SH | 华富中证人工智能产业ETF | 中证人工智能产业指数收益率 | -12.72 | -0.05 | 8,239 |
| 159890.SZ | 招商中证云计算与大数据ETF | 中证云计算与大数据主题指数收益率 | -13.38 | 0.11 | 1,132 |
| 159739.SZ | 鹏华中证云计算与大数据主题ETF | 中证云计算与大数据主题指数收益率 | -13.38 | 0.10 | 457 |
| 516630.SH | 华夏中证云计算与大数据主题ETF | 中证云计算与大数据主题指数收益率 | -13.39 | 0.09 | 913 |
| 516510.SH | 易方达中证云计算与大数据ETF | 中证云计算与大数据主题指数收益率 | -13.40 | 0.08 | 3,489 |
| 159527.SZ | 广发中证云计算与大数据主题ETF | 中证云计算与大数据主题指数收益率 | -13.41 | 0.07 | 222 |
| 517800.SH | 方正富邦中证沪港深人工智能50ETF | 中证沪港深人工智能50指数收益率 | -13.61 | 0.00 | 292 |
| 560850.SH | 汇添富中证信息技术应用创新产业ETF | 中证信息技术应用创新产业指数收益率 | -14.41 | 0.09 | 149 |
| 562030.SH | 华宝中证信息技术应用创新产业ETF | 中证信息技术应用创新产业指数收益率 | -14.42 | 0.08 | 107 |
| 159939.SZ | 广发中证全指信息技术ETF | 中证全指信息技术指数 | -14.46 | 0.10 | 486 |
| 562560.SH | 华夏中证全指信息技术ETF | 中证全指信息技术指数收益率 | -14.55 | 0.01 | 20 |
| 516700.SH | 华宝中证大数据产业ETF | 中证大数据产业指数收益率 | -14.63 | 0.13 | 299 |
| 562920.SH | 易方达中证信息安全主题ETF | 中证信息安全主题指数收益率 | -14.63 | 0.09 | 180 |
| 159613.SZ | 嘉实中证信息安全主题ETF | 中证信息安全主题指数收益率 | -14.69 | 0.04 | 97 |
| 515400.SH | 富国中证大数据产业ETF | 中证大数据产业指数收益率 | -14.69 | 0.07 | 2,907 |
| 516000.SH | 华夏中证大数据产业ETF | 中证大数据产业指数收益率 | -14.70 | 0.06 | 550 |
| 560800.SH | 鹏扬中证数字经济主题ETF | 中证数字经济主题指数收益率 | -15.69 | 0.25 | 181 |
| 159311.SZ | 易方达中证数字经济主题ETF | 中证数字经济主题指数收益率 | -15.87 | 0.06 | 16 |
| 588260.SH | 华安上证科创板新一代信息技术ETF | 上证科创板新一代信息技术指数收益率 | -16.04 | 0.16 | 71 |
| 159899.SZ | 招商中证全指软件ETF | 中证全指软件指数收益率 | -16.18 | 0.10 | 544 |
| 515230.SH | 国泰中证全指软件ETF | 中证全指软件指数收益率 | -16.22 | 0.06 | 4,039 |
| 588770.SH | 摩根上证科创板新一代信息技术ETF | 上证科创板新一代信息技术指数收益率 | -16.28 | -0.08 | 317 |
| 512330.SH | 南方中证500信息技术ETF | 中证500信息技术指数 | -16.33 | -0.03 | 94 |

| | | | | | |
|-----------|------------------|---------------------------|--------|-------|-------|
| 159586.SZ | 南方中证全指计算机ETF | 中证全指计算机指数收益率 | -16.51 | 0.05 | 393 |
| 561010.SH | 华安中证全指软件开发ETF | 中证全指软件开发指数收益率 | -16.64 | 0.18 | 327 |
| 560360.SH | 万家中证软件服务ETF | 中证软件服务指数收益率 | -16.67 | 0.24 | 151 |
| 159852.SZ | 嘉实中证软件服务ETF | 中证软件服务指数收益率 | -16.90 | 0.00 | 9,538 |
| 159131.OF | 华宝中证港股通信息技术综合ETF | 经人民币汇率调整的中证港股通信息技术综合指数收益率 | -16.90 | 0.06 | 3,907 |
| 588780.SH | 国联安科创板芯片设计ETF | 上证科创板芯片设计主题指数 | -16.96 | 0.01 | 2,755 |
| 159108.OF | 博时工业软件ETF | 国证工业软件主题指数收益率 | -17.05 | 0.09 | 134 |
| 159107.SZ | 富国创业板软件ETF | 创业板软件指数收益率 | -17.37 | 0.03 | 1,289 |
| 588930.SH | 银华上证科创板人工智能ETF | 上证科创板人工智能指数收益率 | -20.99 | 0.28 | 1,441 |
| 589090.SH | 鹏华上证科创板人工智能ETF | 上证科创板人工智能指数收益率 | -21.04 | 0.23 | 193 |
| 589520.SH | 华宝上证科创板人工智能ETF | 上证科创板人工智能指数收益率 | -21.13 | 0.13 | 712 |
| 589380.SH | 富国上证科创板人工智能ETF | 上证科创板人工智能指数收益率 | -21.14 | 0.13 | 146 |
| 588730.SH | 易方达上证科创板人工智能ETF | 上证科创板人工智能指数收益率 | -21.19 | 0.07 | 2,771 |
| 588790.SH | 博时科创板人工智能ETF | 上证科创板人工智能指数收益率 | -21.25 | 0.01 | 5,637 |
| 589110.SH | 国泰上证科创板人工智能ETF | 上证科创板人工智能指数收益率 | -21.26 | 0.00 | 204 |
| 589010.SH | 华夏上证科创板人工智能ETF | 上证科创板人工智能指数收益率 | -21.28 | -0.02 | 1,706 |
| 588760.SH | 广发上证科创板人工智能ETF | 上证科创板人工智能指数收益率 | -21.29 | -0.03 | 4,379 |

资料来源: Wind, 中原证券研究所

2.3. 计算机行业可转债行情回顾

3月计算机行业可转债皆下跌,其中跌幅较大的包括奥飞转债下跌19.74%,新致转债下跌18.13%,姚记转债下跌15.61%。

表 16: 2026 年 3 月计算机行业可转债行情回顾

| 证券代码 | 证券简称 | 月涨跌幅 (%) | 正股代码 | 正股简称 | 转股溢价率 (%) * | 到期日期 | 发行时债项等级 |
|-----------|----------|----------|-----------|------|----------------|------------|---------|
| 123104.SZ | 卫宁转债 | -2.7779 | 300253.SZ | 卫宁健康 | 48.6154 | 2027-03-16 | AA |
| 123138.SZ | 丝路转债 | -5.7063 | 300556.SZ | 丝路视觉 | 95.6636 | 2028-03-02 | A+ |
| 123173.SZ | 恒锋转债 | -6.7241 | 300605.SZ | 恒锋信息 | 38.8985 | 2028-12-30 | A |
| 123196.SZ | 正元转02 | -7.6268 | 300645.SZ | 正元智慧 | 75.4076 | 2029-04-18 | A+ |
| 118007.SH | 山石转债 | -7.9644 | 688030.SH | 山石网科 | 49.5928 | 2028-03-22 | A+ |
| 123157.SZ | 科蓝转债 | -9.7006 | 300663.SZ | 科蓝软件 | 71.1456 | 2028-08-30 | AA- |
| 123059.SZ | 银信转债 | -10.2441 | 300231.SZ | 银信科技 | 11.9173 | 2026-07-15 | AA- |
| 123054.SZ | 思特转债 | -12.4651 | 300608.SZ | 思特奇 | 5.7964 | 2026-06-10 | AA- |
| 118027.SH | 宏图转债 | -12.9624 | 688066.SH | 航天宏图 | 19.5503 | 2028-11-28 | A |
| 118052.SH | 浩瀚转债 | -13.3601 | 688292.SH | 浩瀚深度 | 58.1212 | 2031-03-13 | A+ |
| 123261.SZ | 普联转债 | -14.4594 | 300996.SZ | 普联软件 | 63.7759 | 2031-12-05 | A+ |
| 127104.SZ | 姚记转债 | -15.6138 | 002605.SZ | 姚记科技 | 46.8764 | 2030-01-25 | A+ |
| 118021.SH | 新致转债(退市) | -18.1324 | 688590.SH | 新致软件 | 4.9160 | 2028-09-27 | A |
| 123131.SZ | 奥飞转债 | -19.7372 | 300738.SZ | 奥飞数据 | 23.2939 | 2027-12-03 | A+ |

资料来源: Wind, 中原证券研究所 (*选取 2026 年 4 月 21 日数据)

3. 行业观点与投资建议

2026 年软件产业增速略有回落。2026 年 1-2 月软件业务收入 2.15 万亿元,同比增长 11.7%,较 2025 年回落了 1.5 PCT。2026 年 1-2 月软件业务利润总额 2693 亿元,同比增长 7.3%,与 2025 年持平,低于收入增速 4.4 PCT

重点关注子行业的主要数据和动态包括:

(1) 国产化: 2026 年前三个月份,我国集成电路出口数量和金额分别同比增长 13%和 73%,出口集成电路金额占比进口金额占由 2022 年的 37.1%提升到了 2026 年前两个月份的 56.6%。2025 年下半年,我国 AI 芯片国产化比率从 2025 上半年的 35%提升到了 46%,提升趋势明显。国产 AI 芯片厂商 2025 年业务景气度持续提升,其中 2 家实现盈利。2026 年国产 AI 芯片在加速发展的过程中,有望开启向海外市场的拓展。

(2) 算力: 2025Q4 全球服务器市场中,美国同比增速 72.4%,而中国增速仅为 17.7%,较 2024Q4 93.3%的增速下滑明显,我们认为主要源于缺芯因素影响了服务器需求的释放。2026 年随着移动集采大单落地和郑州 6 万卡超算集群的建设,超节点正在成为发展趋势。2025 年算力建设受制于芯片供给,26 年在 AI 应用增长和国产芯片供给能力提升的双加持下,行业算力建设速度有明显加快迹象,第三方数据中心厂商开启积极建设步伐。

(3) AI: 目前国内模型在能力上主要集中在第二梯队,随着厂商新模型的陆续发布,整体水平上更加集中,且性价比优势更加凸显。Anthropic 年支出超 100 万美元的企业客户数从 2 月的超 500 家增长到 4 月超 1000 家,业务增长弹性在 2026 年充分显现。3 月豆包日均 token 使用量突破 120 万亿,季度环比增长 140%。AI 应用需求爆发式增长,拉动算力价格持续上涨,模型厂商、云厂商发布多轮涨价通知。

给予行业强于大市的投资评级。**2026 年 AI 应用发展速度整体超预期,带来了整体算力产业供应链的紧张。2026 年来看,在海外芯片供给受限的大背景下,国内 AI 芯片也迎来了性能、产能的双方面改善,有望持续受益于市场结构的变化,并最终改善国内算力的供给能力。建议关注有大规模智算中心交付计划的字节跳动数据中心核心供应商润泽科技(300442),有产业链一体化优势同时持股优质企业的服务器厂商中科曙光(603019)。同时建议积极关注正在积极开展 IPO 的长鑫科技、芯和半导体、合见工软、超聚变、平头哥、智谱。**

表 17: 行业重点企业的 IPO 推进情况

| 行业 | 公司 | 日期 | 进程 |
|-------|------|--------------|-----------|
| AI 芯片 | 摩尔线程 | 2025. 12. 5 | 上市 |
| | 沐曦股份 | 2025. 12. 17 | 上市 |
| | 壁仞科技 | 2026. 1. 2 | 港交所上市 |
| | 天数智芯 | 2026. 1. 8 | 港交所上市 |
| | 爱芯元智 | 2026. 1. 25 | 港交所披露上市材料 |

| | | | |
|-----|---------|--------------|------------------|
| | 燧原科技 | 2026. 4. 16 | 科创板首轮审核问询 |
| | 昆仑芯 | 2026. 1. 1 | 港交所递交上市申请 |
| | 平头哥 | | 计划重组，未来独立上市 |
| HBM | 长鑫科技 | 2025. 12. 30 | 获得科创板 IPO 的交易所受理 |
| 服务器 | 超聚变 | 2026. 1. 6 | 提交上市辅导备案 |
| 大模型 | 智谱 | 2026. 1. 8 | 港交所上市 |
| | | 2026. 2. 13 | 科创板辅导备案 |
| | MINIMAX | 2026. 1. 9 | 港交所上市 |
| EDA | 芯合半导体 | 2025. 12. 24 | IPO 的辅导工作完成 |
| | 全芯制造 | 2025. 12. 11 | 辅导备案登记 |
| | 合见工软 | 2025. 12 | 办理了上市辅导备案登记 |

资料来源：澎湃，Wind，中原证券研究所

4. 风险提示

国际局势的不确定性；海外 AI 产业竞争格局变化带来市场调整风险。

行业投资评级

强于大市：未来 6 个月内行业指数相对沪深 300 涨幅 10% 以上；

同步大市：未来 6 个月内行业指数相对沪深 300 涨幅 -10% 至 10% 之间；

弱于大市：未来 6 个月内行业指数相对沪深 300 跌幅 10% 以上。

公司投资评级

买入：未来 6 个月内公司相对沪深 300 涨幅 15% 以上；

增持：未来 6 个月内公司相对沪深 300 涨幅 5% 至 15%；

谨慎增持：未来 6 个月内公司相对沪深 300 涨幅 -10% 至 5%；

减持：未来 6 个月内公司相对沪深 300 涨幅 -15% 至 -10%；

卖出：未来 6 个月内公司相对沪深 300 跌幅 15% 以上。

证券分析师承诺

本报告署名分析师具有中国证券业协会授予的证券分析师执业资格，本人任职符合监管机构相关合规要求。本人基于认真审慎的职业态度、专业严谨的研究方法与分析逻辑，独立、客观的制作本报告。本报告准确的反映了本人的研究观点，本人对报告内容和观点负责，保证报告信息来源合法合规。

重要声明

中原证券股份有限公司具备证券投资咨询业务资格。本报告由中原证券股份有限公司（以下简称“本公司”）制作并仅向本公司客户发布，本公司不会因任何机构或个人接收到本报告而视其为本公司的当然客户。

本报告中的信息均来源于已公开的资料，本公司对这些信息的准确性及完整性不作任何保证，也不保证所含的信息不会发生任何变更。本报告中的推测、预测、评估、建议均为报告发布日的判断，本报告中的证券或投资标的价格、价值及投资带来的收益可能会波动，过往的业绩表现也不应当作为未来证券或投资标的表现的依据和担保。报告中的信息或所表达的意见并不构成所述证券买卖的出价或征价。本报告所含观点和建议并未考虑投资者的具体投资目标、财务状况以及特殊需求，任何时候不应视为对特定投资者关于特定证券或投资标的的推荐。

本报告具有专业性，仅供专业投资者和合格投资者参考。根据《证券期货投资者适当性管理办法》相关规定，本报告作为资讯类服务属于低风险（R1）等级，普通投资者应在投资顾问指导下谨慎使用。

本报告版权归本公司所有，未经本公司书面授权，任何机构、个人不得刊载、转发本报告或本报告任何部分，不得以任何侵犯本公司版权的其他方式使用。未经授权的刊载、转发，本公司不承担任何刊载、转发责任。获得本公司书面授权的刊载、转发、引用，须在本公司允许的范围内使用，并注明报告出处、发布人、发布日期，提示使用本报告的风险。

若本公司客户（以下简称“该客户”）向第三方发送本报告，则由该客户独自为其发送行为负责，提醒通过该种途径获得本报告的投资者注意，本公司不对通过该种途径获得本报告所引起的任何损失承担任何责任。

特别声明

在合法合规的前提下，本公司及其所属关联机构可能会持有报告中提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问等各种服务。本公司资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告意见或者建议不一致的投资决策。投资者应当考虑到潜在的利益冲突，勿将本报告作为投资或者其他决定的唯一信赖依据。