下载日志已记录，仅供内部参考

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

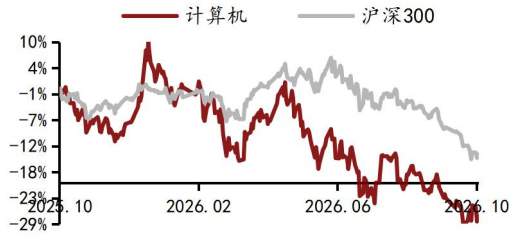
AI 芯片国产化率提升至 64%，Kimi 再次步入头部模型行列

——计算机行业月报

证券研究报告-行业月报 **强于大市(维持)**

发布日期：2026 年 10 月 29 日

计算机相对沪深 300 指数表现



资料来源：中原证券研究所，聚源

相关报告

《计算机行业月报：国产芯片发布计划大幅提前，算力投入呈现加速趋势》 2026-09-30

《计算机行业月报：国内资本开支加速明显，AI 硬件短缺趋势延续》 2026-08-31

《计算机行业月报：国产 AI 产业基本面持续向好》 2026-07-31

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 年前九个月份，集成电路进出口单价持续提升，同时出口单价连续 4 个月反超进口产品单价，国产芯片产品性能持续提升。2026 年上半年 AI 芯片国产化率进一步提升到了 64%，2026Q3 厂商业绩延续高增势头。DeepSeek 与昇腾的结合日趋紧密，正在通过开源方式，加快国产芯片替代进程。阿里新一代自研芯片 V900 将在 2027Q1 上市，较原先计划提前了 2 个季度，2027 年即将上市的昇腾 960 同样计划提前。由于逻辑折叠的加持，在相同性能下，麒麟 2026 在基准测试中，NPU 功耗下降 66%，GPU 下降 58%，性能核心 CPU 下降 41%。截至 2026 年 10 月 21 日，鸿蒙操作系统终端数已经突破 9500 万，华为预计 2026 年突破 1 亿用户规模。

- **AI：**月之暗面在 10 月发布了 Kimi K3.1 Pro，拥有 2.8 万亿参数，通过后训练加强和更多领域数据的强化学习，再次回归头部模型行列，其他国产模型主要位居第二梯队。以美国最强模型 Anthropic 的 Opus 5.5 的调用价格来看，其输入输出价格都约为同等级模型 Kimi-K3.1 pro 的 2 倍，在国产模型冲击和用户成本敏感度提升的背景下，近期美国闭源模型降价趋势明显。根据 OpenRouter 的模型调用数据，DeepSeek 份额 26%，连续 12 周位列第一大模型厂商。同时中国模型厂商份额达到 46%，暂领先较美国模型厂商（份额 44%）。10 月微软向约 60000 名员工（占比超 25%）开放自愿离职方案，以压缩成本、调配资源加码 AI，科技大厂裁员再呈加速迹象。9 月，美国企业模型采用率达到 55.9%，Anthropic 以 43.8%的份额再次超过 OpenAI（40.6%）。

- **给予行业强于大市的投资评级。**当前 AI 应用正在加速渗透至编程、影视游戏、数学研究、网络安全等领域，物理 AI 的突破进程也在加速。Agent 自主化的执行意味着模型被调用的频次也在大幅增长，让算力供应持续不足，算卡租赁价格继续呈现走高趋势。相对而言，受到缺芯因素的影响，国内算力紧张程度更高。随着 2027 年国产卡产能的释放和新卡的推出，国内算力需求有望得到更好的满足。我们继续对国内 AI 算力产业发展持乐观预期，但是考虑到外盘科技股整体位居高位，后续或受到 Anthropic 上市、2027 年资本开资担忧情绪等因素影响，A 股 AI 算力板块如要演绎独立行情或仍面临较大的信心挑战。**建议持续重点关注 AI 产业链上游及中游环节，包括半导体设备、芯片、存储、光通信、服务器、AIDC、云计算等环节方面，**推荐服务器厂商中科曙光（603019）、紫光股份（000938）、浪潮信息（000977），AIDC 厂商润泽科技（300442）、东阳光（600673）。同时建议积极关注正在积极开展 IPO 的超聚

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

下载日志已记录，仅供内部参考

变、平头哥、智谱、DeepSeek、月之暗面、长存控股。

风险提示：国内外科技领域巨头 IPO 带来的市场冲击；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。

下载日志已记录，仅供内部参考

内容目录

1. 行业数据 6
1.1. 行业数据：1-7 月软件行业收入稳定增长，利润增速回落明显 6
1.2. 2026 年高景气赛道：IC 设计、嵌入式软件、云计算+大数据服务、基础软件 7
1.3. 国产化：上半年 AI 芯片国产化率 64%，DeepSeek 开源加速国产适配 10
1.3.1. 国产芯片 10
1.3.2. 华为给出了芯片发展的新路径，逻辑折叠在性能、成本、功耗方面表现优异 20
1.3.3. 操作系统：鸿蒙 7 开启正式版升级，鸿蒙终端数量突破 9500 万 22
1.4. AI：头部模型继续呈现降价趋势 25
1.4.1. 中美大模型发展概况 25
1.4.2. 主流模型价格对比 26
1.4.3. 中美模型厂商收入 27
1.4.4. 厂商模型调用及国产模型替代情况 28
1.4.5. 模型与应用：大模型备案高度集中在发达区域 29
1.4.6. 科技企业裁员：2026 年前 9 个月美国因为 AI 取代占总裁员的比例达到 21% 30
1.4.7. 企业端 AI 应用情况 32
1.4.8. 国内 Token 使用情况 33
2. 河南计算机行业动态 34
2.1. 河南计算机行业数据跟踪 34
2.2. 河南上市公司行情回顾 35
3. 计算机行业行情表现 35
3.1. 行情回顾：10 月整体表现较弱 35
3.2. 估值：行业的估值低于历史均值水平 36
3.3. 计算机行业 ETF 行情回顾 36
3.4. 计算机行业可转债行情回顾 38
4. 行业观点与投资建议 38
5. 风险提示 40

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

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

下载日志已记录，仅供内部参考

图 12 ： 2026 年 1-7 月我国软件业务收入结构 9
图 13 ： 2026 年 1-6 月信息技术服务收入结构 9
图 14 ： 2026 年 1-6 月软件产品收入结构 9
图 15 ： 2022-2026 我国进出口集成电路单价（元/个） 10
图 16 ： 2022-2026 我国出口/进口集成电路比值 10
图 17 ： 2021-2026 年我国集成电路对进口依赖度和出口占比 11
图 18 ： 2023Q1-2025Q2 英伟达收入的区域分布情况（亿美元，按收入来源区域） 11
图 19 ： 2024Q3-2026Q2 英伟达收入的区域分布情况（亿美元，按企业总部所在地） 11
图 20 ： 2023-2026 我国 AI 芯片出货情况及国产化比率 13
图 21 ： 2026H1 我国加速芯片市场份额 13
图 22 ： 2017-2026 国内主要上市 AI 芯片企业收入及增速对比 14
图 23 ： DeepSeek V4 发布公告关于算力的表述 17
图 24 ： 麒麟 CPU 性能核心工作频率发展趋势 21
图 25 ： 麒麟芯片的晶体管密度和性能核心工作频率预测 21
图 26 ： 麒麟 2026 和麒麟 9030 Pro 在等效性能下的功率对比 21
图 27 ： 传统设计与韬定律的差异 22
图 28 ： 逻辑折叠示意图 22
图 29 ： 2022Q2-2026Q2 中国手机操作系统市场份额 24
图 30 ： 2022Q2-2026Q2 全球手机操作系统市场份额 24
图 31 ： 2022Q1-2026Q3 中国手机市场厂商份额 24
图 32 ： 2021Q1-2026Q2 中国平板电脑市场厂商份额 24
图 33 ： 不同地区的大模型智能水平对比（截至 2026.10.26，红色为中国模型） 26
图 34 ： 大模型智能水平及开源情况（截至 2026.10.27） 26
图 35 ： 海外模型及 AI 应用的年化收入（亿美元） 28
图 36 ： 国内模型及 AI 应用的年化收入（亿美元） 28
图 37 ： OpenRouter 各厂商模型调用量（百万 tokens） 29
图 38 ： OpenRouter 调用的模型区域份额 29
图 39 ： OpenRouter 各厂商视频模型调用量（百万 tokens） 29
图 40 ： OpenRouter 调用的视频模型区域份额 29
图 41 ： 全国大模型备案与登记数量 30
图 42 ： 各地大模型备案与登记数量 30
图 43 ： 大模型备案与登记数量的区域分布 30
图 44 ： 2025.7-2026.9 美国因 AI 被裁的人数及占裁员总数的比例（当月值，人） 32
图 45 ： 2023-2026 年美国因 AI 被裁的人数及占裁员总数的比例（累计值，人） 32
图 46 ： Ramp AI 指数（美国企业模型采用率） 32
图 47 ： 美国企业每个员工每月的 AI 支出（美元） 33
图 48 ： 美国各行业员工每月的 AI 支出中位数 33
图 49 ： 2026 年 9 月美国各行业在人工智能模型、订阅、代币和 API 上投入的企业占比 33
图 50 ： 2024-2026 年中国我国日均 Token 调用量 34
图 51 ： 2024-2030 年中国 MaaS 市场 Tokens 规模预测 34
图 52 ： 河南上市公司近期股价涨跌幅表现（%） 35
图 53 ： 2026 年 10 月中信一级子行业涨跌幅（截至 2026 年 10 月 27 日） 36
图 54 ： 近 10 年中信计算机行业估值水平（截至 2026.10.27） 36

表 1 ： 2025 年 AI 芯片相关动态 12
表 2 ： 阿里真武芯片性能及发布计划 15

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

下载日志已记录，仅供内部参考

表 3 ：华为昇腾系列芯片发布计划及芯片性能（2025.9）.....16
表 4 ：华为超节点上市计划及参数（2025.9）.....16
表 5 ：海光芯片研发进程.....18
表 6 ：通过安全可靠测评的芯片产品.....18
表 7 ：鸿蒙纯血系统发展动态.....23
表 8 ：模型调用价格（每百万 Token，Artificial Analysis2026 年 10 月 27 日智能指数）.....27
表 9 ：海外主要科技厂商近期裁员情况.....31
表 10 ：2026 年 9 月计算机行业 ETF 行情回顾（截至 2026.10.27）.....37
表 11 ：2026 年 10 月计算机行业可转债行情回顾（截至 2026.10.28）.....38
表 12 ：行业重点企业的 IPO 推进情况.....39

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

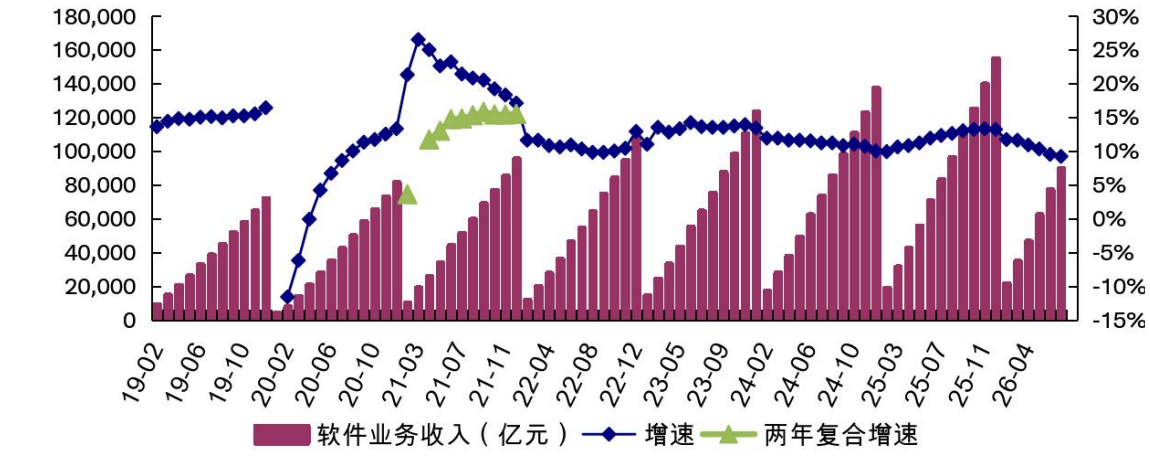
下载日志已记录，仅供内部参考

1. 行业数据

1.1. 行业数据：1-7 月软件行业收入稳定增长，利润增速回落明显

2026 年软件产业增速持续回落。根据工信部数据，2026 年 1-7 月软件业务收入 8.98 万亿元，同比增长 9.2%，较 2025 年回落了 4.0 PCT。

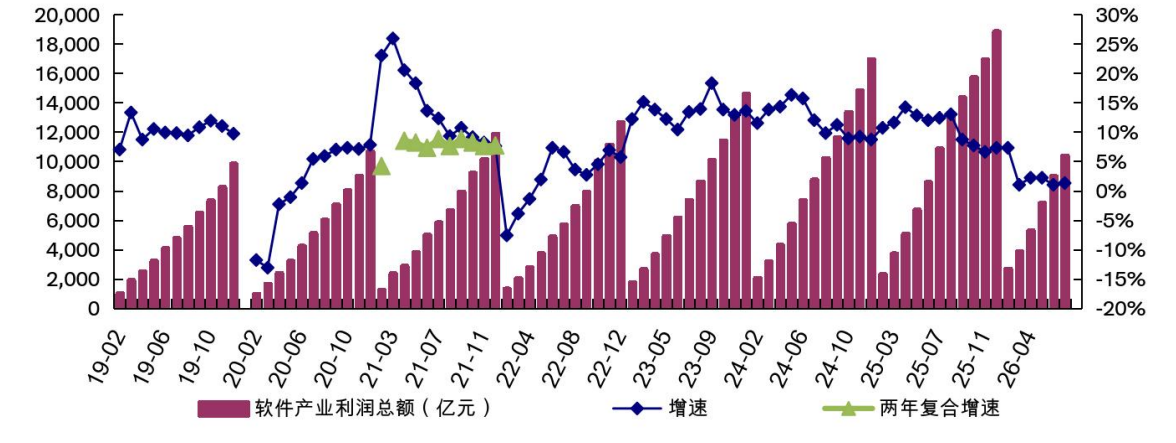
图 1：2019-2026 年我国软件业务收入及增速（月度累计值）



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下滑较为明显。2026 年 1-7 月软件业务利润总额 10396 亿元，同比增长 1.3%，较 2025 年回落 6.0 PCT，低于收入增速 7.9 PCT。考虑到 AI 应用的快速增长，我们认为利润下滑或源于企业在 AI 方面的投入和开支加大所致。

图 2：2019-2026 年我国软件业务利润总额及增速（月度累计值）

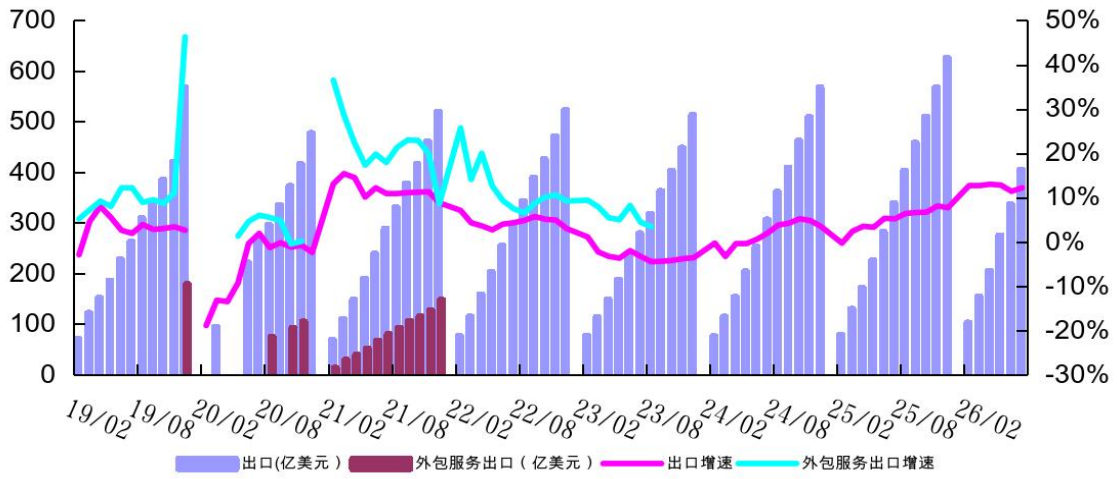


资料来源：工信部，中原证券研究所

2026 软件出口整体仍在加速趋势中。2026 年 1-7 月软件业务出口金额 406.1 亿美元，同比增长 12.2%，较 2025 年增长 4.5 PCT，软件行业出口金额约占行业收入的 3.3%。

下载日志已记录，仅供内部参考

图 3：2019-2026 年我国软件业务出口数据



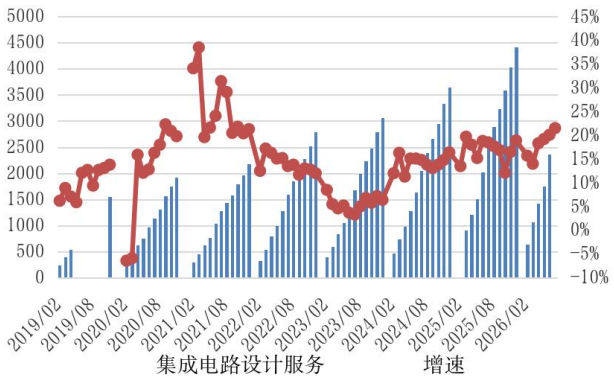
资料来源：工信部，中原证券研究所

1.2. 2026 年高景气赛道：IC 设计、嵌入式软件、云计算+大数据服务、基础软件

IC 设计：2026 年 1-7 月同比增长 21.5%，较 2025 年增长了 2.6 PCT，高于软件行业整体增速 12.3 PCT，维持了较高的景气度，2026 年有望继续受益于芯片产业的高景气和国产化趋势。

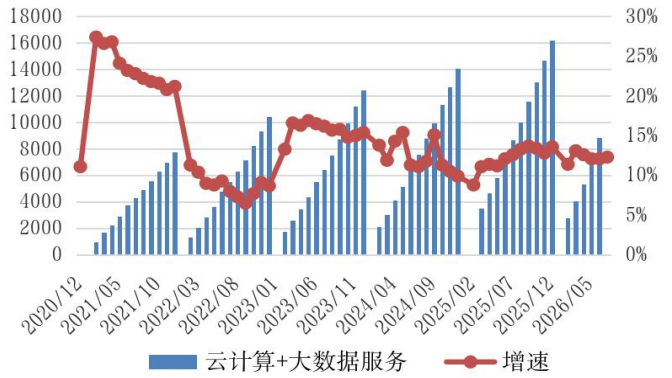
云计算+大数据服务：2026 年 1-7 月同比增长 12.3%，较 2025 年下降了 1.3 PCT，高于软件行业整体增速 3.1 PCT，2026 年预计仍将受益于 AI 应用快速发展趋势。

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

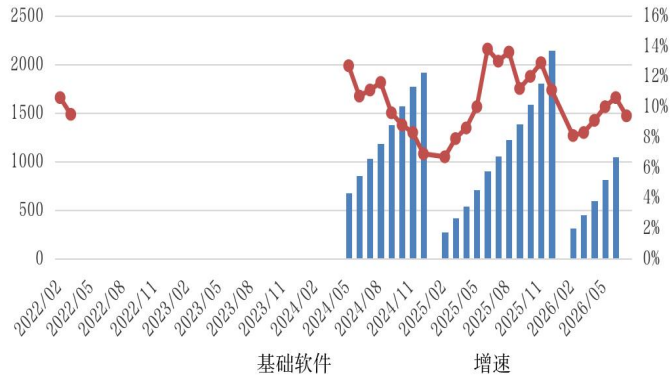
基础软件：2026 年 1-7 月同比增长 9.4%，较 2025 年下降了 1.7 PCT，高于软件行业整体增速 0.2 PCT。

工业软件产品：2026 年 1-7 月同比增长 8.4%，较 2025 年下降了 1.3 PCT，处于历史相对低位的水平。

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

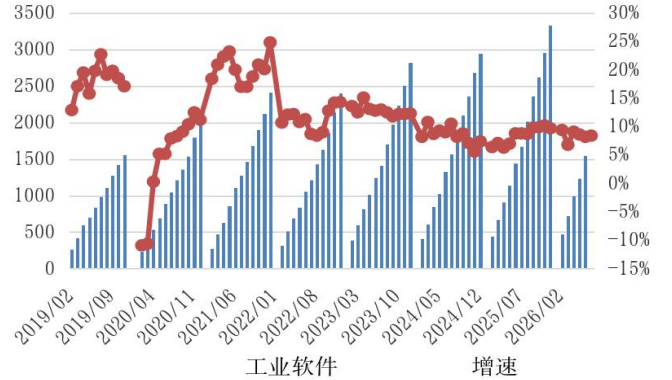
下载日志已记录，仅供内部参考

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

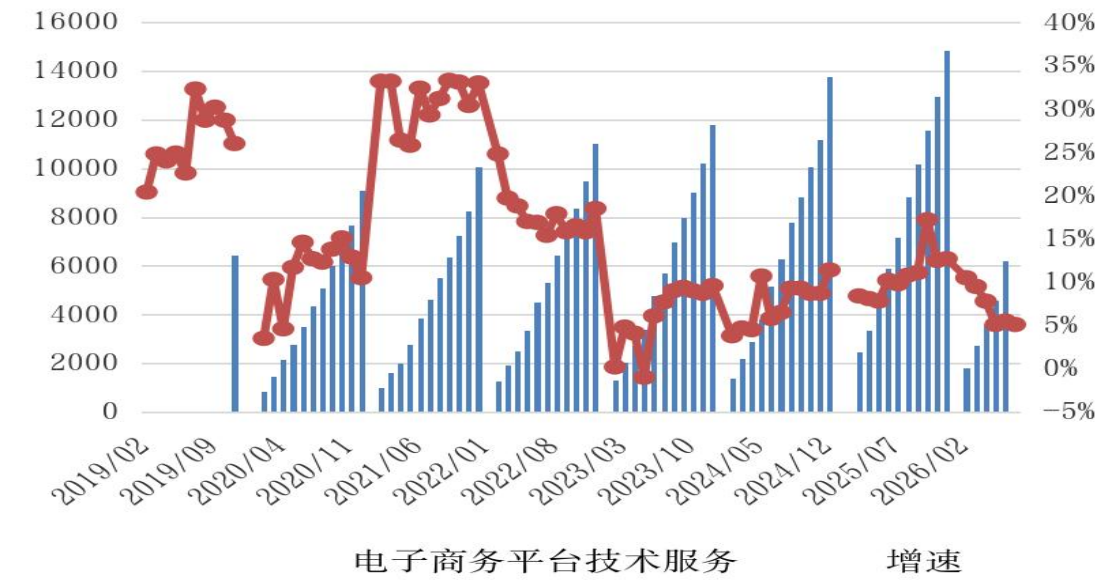
图 7：2019-2026 工业软件收入及增速（亿元）



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 1-7 月同比增长 5.1%，较 2025 年下降了 7.6 PCT，低于软件行业整体增速 4.1 PCT，受到消费端景气度的拖累。

图 8：2019-2026 电子商务平台技术服务收入及增速（亿元）



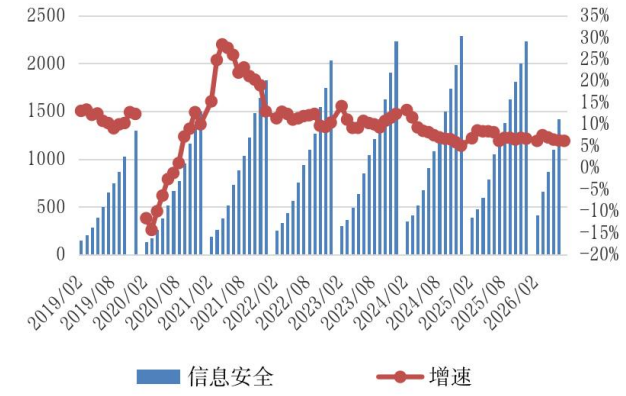
资料来源：工信部，中原证券研究所

信息安全：2026 年 1-7 月同比增长 6.2%，较 2025 年下降 0.5 PCT，低于行业 3.0 PCT，处于相对地位水平。

嵌入式系统软件：2026 年 1-7 月同比增长 10.5%，较 2025 年提升 1.2 PCT，高于行业平均增速 1.3PCT，或受到端侧 AI 应用的加持。

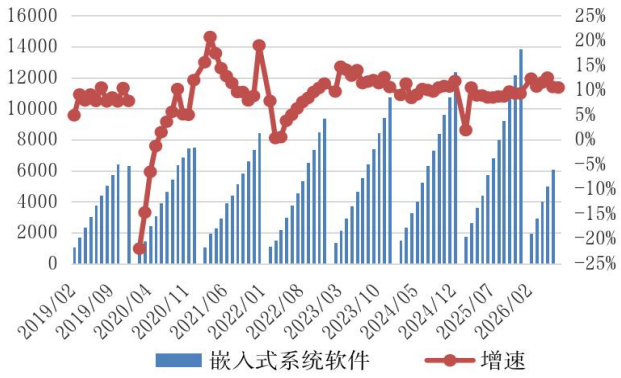
下载日志已记录，仅供内部参考

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

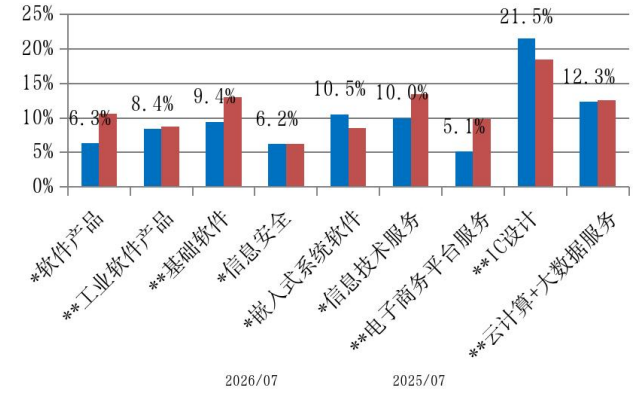
图 10：2019-2026 嵌入式系统软件收入及增速（亿元）



资料来源：工信部，中原证券研究所

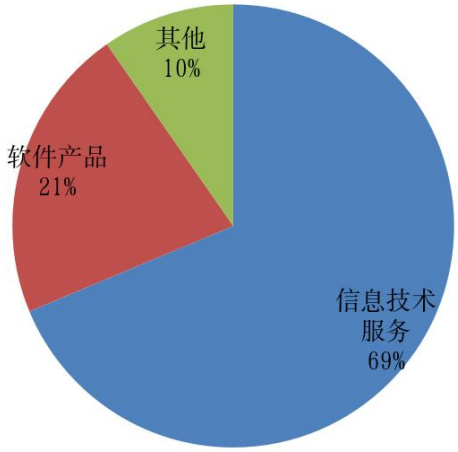
软件业务中，信息技术服务收入占比持续提升。2026 年 1-7 月信息技术服务收入增速 10.4%，高于软件业务整体增速 0.9 PCT，占比软件业务整体收入比重提升至 68.5%。

图 11：2026 与 2025 年 1-7 月软件子行业增速对比



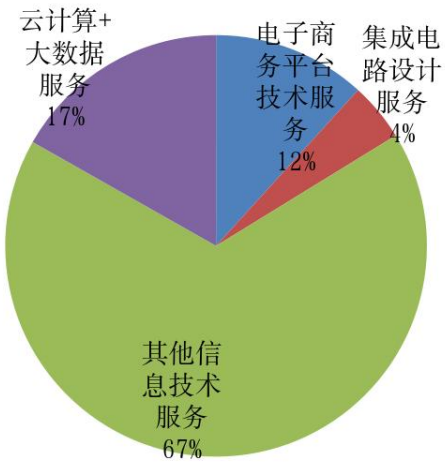
资料来源：工信部，中原证券研究所（*为一级子行业，**为二级子行业）

图 12：2026 年 1-7 月我国软件业务收入结构



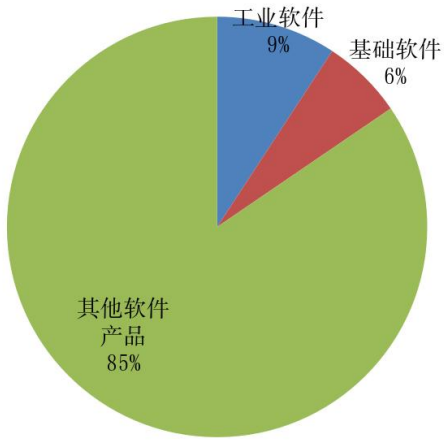
资料来源：工信部，中原证券研究所

图 13：2026 年 1-6 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 1-6 月软件产品收入结构



资料来源：工信部，中原证券研究所

下载日志已记录，仅供内部参考

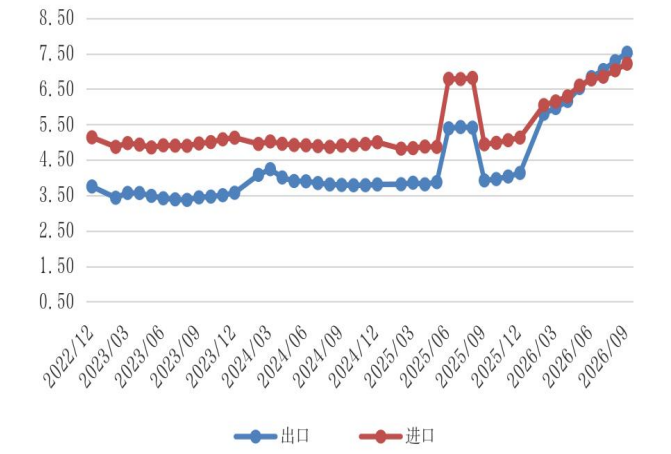
1.3. 国产化：上半年 AI 芯片国产化率 64%，DeepSeek 开源加速国产适配

1.3.1. 国产芯片

1.3.1.1. 国产芯片产品性能持续提升，上半年芯片自给率达到 24%

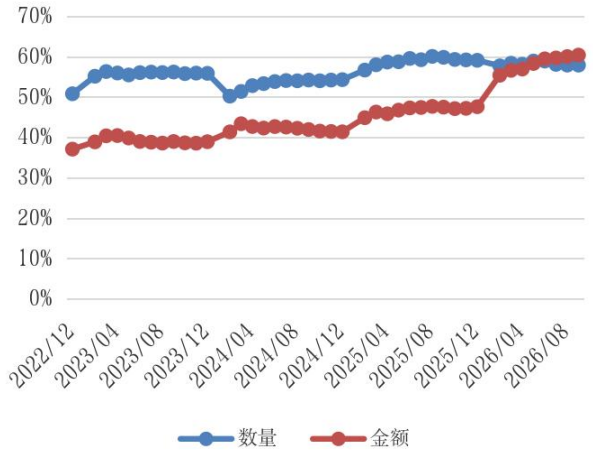
2026 年前九个月份，集成电路进出口单价持续提升，同时出口单价连续 4 个月反超进口产品单价。进口数量和金额分别同比增长 7%和 57%，出口数量和金额分别同比增长 3.8%和 99.1%。其中出口集成电路平均单价与进口电路平均单价价差从上一年同期的-1.02 元/个增长到 0.30 元/个，出口产品单价连续 4 个月实现反超。同时将我国出口和进口集成电路对比来看，出口集成电路数量占比进口数量上总体稳定，但是金额占比方面呈现持续提升趋势，已经由 2022 年的 37.1%提升到了 2026 年 1-9 月的 60.4%，**这表明我国的集成电路产品能力在持续提升。**

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



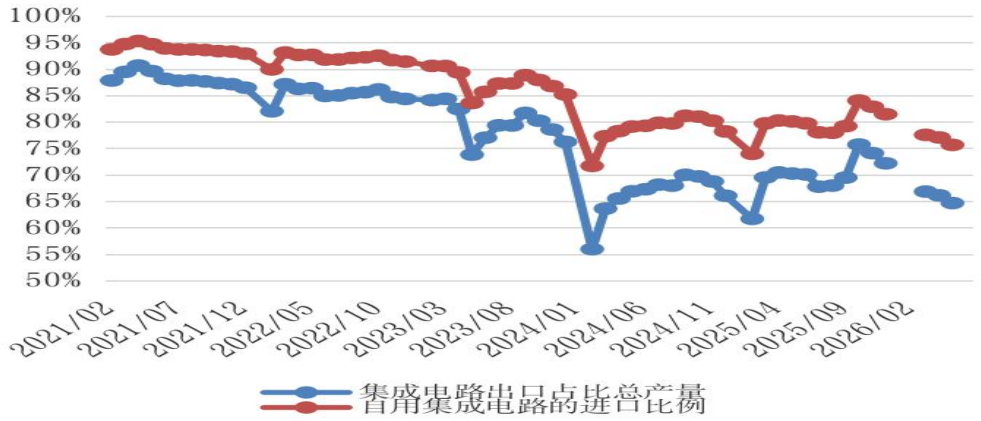
资料来源：海关总署，中原证券研究所

集成电路自给化比例升至 24%。2026 年 1-5 月，我国集成电路在进口依赖度比例和出口在总产量中占比分别为 76%（即国产化占比 24%）和 65%，分别较 2025 年下降了 7 PCT 和 5 PCT。考虑到当前产业链供应紧张，同时国内对于国产芯片的支持力度较高，后续国产替代的趋势将进一步凸显，两个指标 2026 年有望进一步降低。

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

下载日志已记录，仅供内部参考

图 17：2021-2026 年我国集成电路对进口依赖度和出口占比

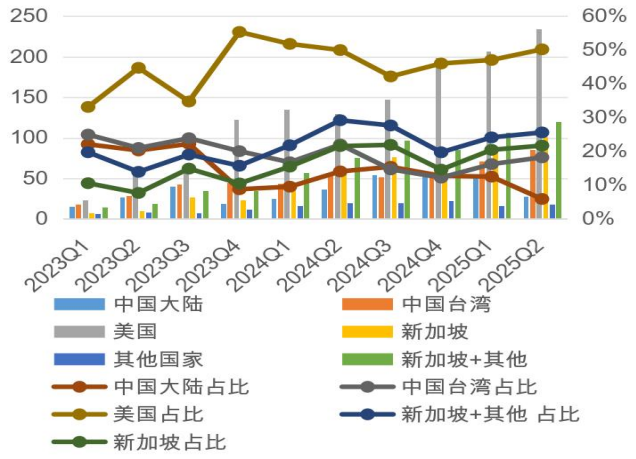


资料来源：工信部，海关总署，中原证券研究所

1.3.1.2. 芯片国产化仍受益于政策发展趋势

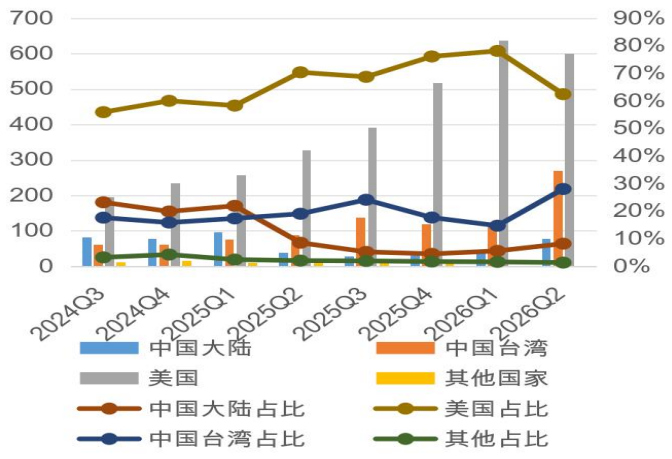
Q2 英伟达来自中国大陆企业的收入增长明显，但是仍然没有超过前期高位水平。2023 年 Q1 到 Q3，英伟达来自中国大陆收入占比分别为 22%、20%、22%，Q3 金额达到了 40.3 亿美元，Q4 在 A800 和 H800 禁令影响下该比例降低到了 9%。而后，中国大陆占比约为 10%-15%之间波动，在 H20 禁令的影响下，该比例 Q3 降低到了 6%，金额也从上个季度的 55.2 亿美元降低到了 27.7 亿美元。2025Q3 开始，英伟达对分区域的收入划分口径进行了修改，从客户账单所在地改为了总部所在地。根据新旧口径差额的估算，我们看到中国大陆企业不断加大在其他地区对于英伟达的采购，从 2025Q1 达到 41.3 亿美元，在监管政策变化后，2025Q2 降为 12.2 亿美元。2026Q2，在国内互联网厂商加大算力投入的大背景下，英伟达来自中国大陆企业的收入（包含其海外建设的算力）增加至 78.8 亿美元，同比大幅增长 98%，但是仍然没有超过前期高位的 96.59 亿美元，我们认为在国内算力需求爆发性增长的过程中，国产卡仍然在扮演越来越重要的地位。

图 18：2023Q1-2025Q2 英伟达收入的区域分布情况（亿美元，按收入来源区域）



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）

图 19：2024Q3-2026Q2 英伟达收入的区域分布情况（亿美元，按企业总部所在地）



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）

下载日志已记录，仅供内部参考

从最新的中美政策来看，5 月美国总统访华期间，美国批准向多加中国巨头出售 H200 芯片，但是随后特朗普在返程中证实中国拒绝采购 H200 芯片，目的是全力发展国产芯片。作为一个标志性事件，本次政策对国产芯片的支持态度进一步明确，也将给后续国产芯片的推广和应用进一步扫清障碍。

由于算力短缺问题较为严重，后续国内互联网厂商仍然在积极开展海外芯片采购。但是由于互联网厂商大规模采购的芯片总体性能偏低，采购量受限，后续随着 2027 年国产新一代芯片的提前上市，国产芯片性能的进一步提升，算力短缺现象有望得到较大的缓解，因而我们认为这些采购行为对于国家支持国产芯片发展的总体方向影响有限。

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，相关具体要求仍在制定中。
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。

下载日志已记录，仅供内部参考

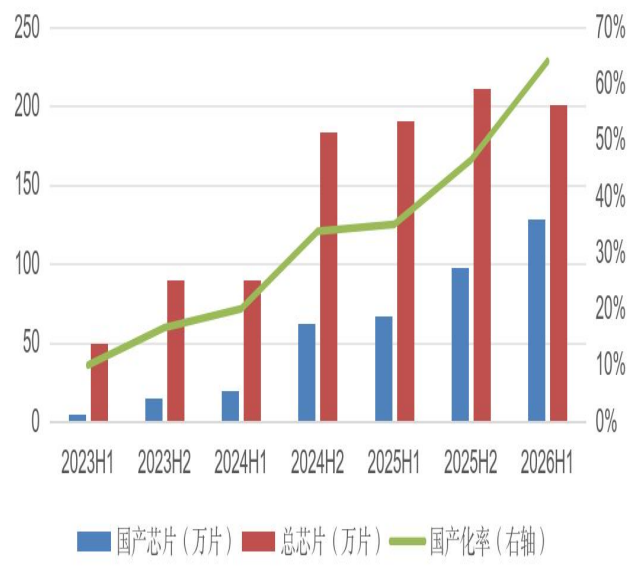
2026. 5. 15	美国方面已批准英伟达（Nvidia）向约 10 家中国公司出口其 H200 人工智能（AI）芯片，获准企业名单中包括阿里巴巴、腾讯及字节跳动等多家行业巨头。这是美方首次较大规模批准对华销售高端 AI 芯片。
2026. 5. 17	特朗普结束访华行程后在空军一号上向记者证实，中国拒绝采购英伟达 H200 芯片。他透露，中国之所以选择不批准相关采购，目的是全力发展自家芯片产业。
2026. 5. 22	国家发改委表示指导国产大模型加大力度适配国产算力芯片，在保持快速发展的同时，确保自主可控、向善发展、行稳致远。
2026. 5. 31	美国商务部工业和安全局（BIS）表示，即使实体位于中国境外，也将对总部设在中国的实体强制执行先进芯片的许可证要求。
2026. 7. 15	中兴通讯旗下全资子公司中兴康讯电信、专业服务器制造商 Maginfra，以及金山云子公司珠海横琴云祥智盛网络科技三家企业，近期正式拿到美国商务部发放的出口许可，被允许分别采购英伟达和 AMD 旗下的先进 AI 芯片产品。
2026. 7. 29	美国商务部工业与安全局一名发言人向《Axios》表示，该机构正在调查英伟达 Blackwell 芯片潜在的出口违规行为。
2026. 8. 19	英国金融时报报道，中国已允许英伟达 H200 芯片小批量进入内地，字节跳动和腾讯近几周分别收到约 1 万颗 H200 处理器，其他数家中国科技企业也有望在短期内获得类似供应。
2026. 10. 21	根据快科技报道，字节跳动、阿里巴巴、腾讯已订购超过 300 万张英伟达面向企业级与工作站市场的新款 RTX PRO 5500 GPU，总标价格超过 350 亿美元。由于中资科技企业需求远超原定产能，英伟达曾告知合作伙伴，可能只能供应客户所订购数量的 5%至 10%。

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，闻华财讯，新浪财经，爱集微，云头条，中原证券研究所

1.3.1.3. 2026 年上半年 AI 芯片国产化率进一步提升到了 64%，2026Q3 厂商延续高增势头

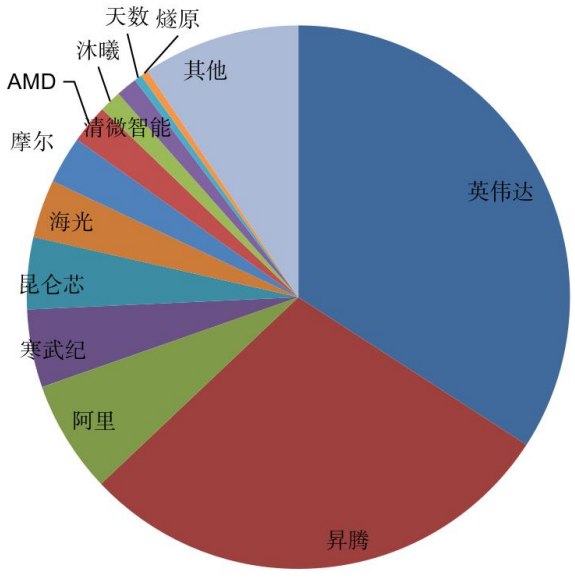
2026 年 AI 芯片国产替代趋势继续加快。根据 IDC 数据测算，2026 年上半年，我国 AI 芯片国产化比率从 2025 下半年的 46%提升到了 64%，提升趋势明显。考虑到英伟达芯片对国内供应受限，同时国产芯片在性能和产能方面的提升，我们预计 2026 年下半年到 2027 年国产 AI 芯片的占比还将获得较大比例的提升。

图 20：2023-2026 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 21：2026H1 我国加速芯片市场份额

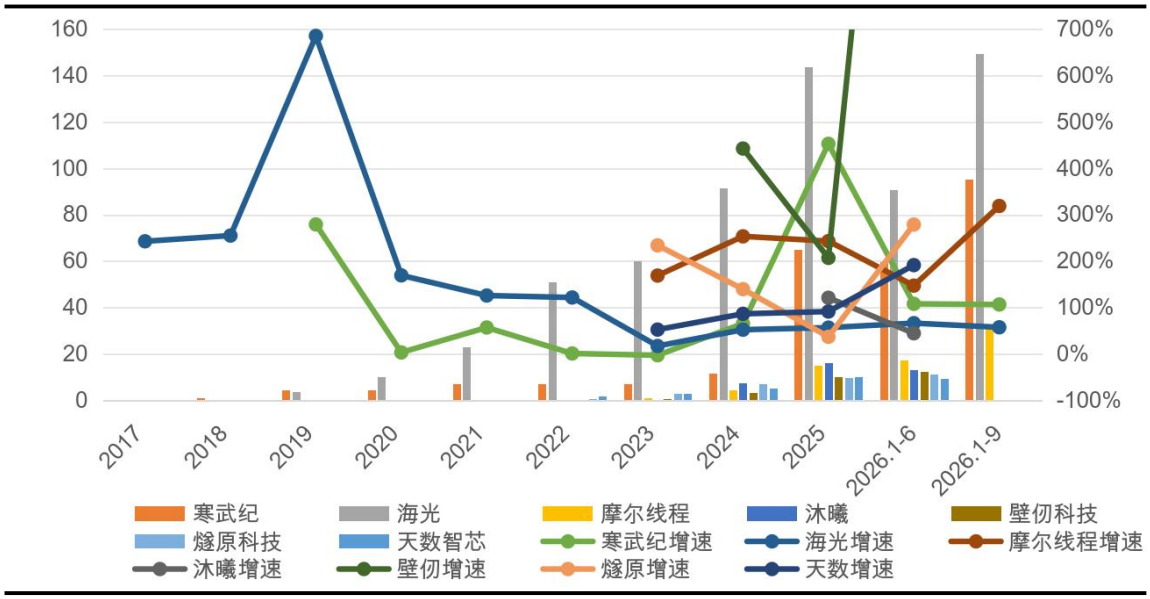


资料来源：IDC，中原证券研究所

下载日志已记录，仅供内部参考

从财报数据来看，继 2025 年的收入高增长以后，国产 AI 芯片厂商 2026 年业务景气度持续提升，其中 4 家实现盈利（2 家为金融资产公允价值变动因素）。从 3 季报来看，海光前 3 季度收入 149.55 亿元，同比增长 58%，第三季度收入 58.56 亿元，同比增长 45%，依然是收入最多的独立上市芯片厂商；寒武纪前 3 季度收入 95.26 亿元，同比增长 107%，Q3 收入 35.31 亿元，同比增长 105%，继续保持了较快的增速水平；摩尔线程前 3 季度上涨 319%至 32.89 亿元，同时在 2026 年实现扭亏为盈。

图 22：2017-2026 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（2026H1 收入按照业绩预告中位数）

6 月 30 日，美团发布 LongCat-2.0 大模型，并表示该模型已在五万卡国产算力集群上完成从零预训练到推理的全流程，将国产算力用于万亿参数模型训练。

8 月 26 日，智谱宣布 GLM-5.3-Flash 所承载的流量都由国产芯片承载，调用超过 10 万张国产芯片集群用于改模型推理服务，硬件效率及单 token 成本已经达到了与主流英伟达 GPU 相当的水平，这也意味着国产芯片的推理服务能力已经获得了核心客户的认可。

9 月 30 日，DeepSeek 开源了面向昇腾平台的 TileLang 编译工具、计算库和分布式通信库，完善了芯片的软件生态。10 月 20 日，DeepSeek 又开源了内部推理框架 DeepInfer，兼容英伟达与华为昇腾平台，可以实现模型的部署在两种硬件之间的灵活切换。同时，9 月 4 日，彭博社消息表示，DeepSeek 计划在内蒙古建设大型数据中心，部署至少 16 万颗华为昇腾 950DT 芯片，用于模型推理。**DeepSeek 与昇腾的结合日趋紧密，正在通过开源方式，进一步加快国产芯片的替代进程。**

1.3.1.4. 昇腾 960 和真武 V900 上市计划大幅提前

阿里：目前，平头哥已经在算力、网力、存力芯片上都实现了自研，包括真武系列 AI 芯片、倚天系列 CPU 芯片（ARM）、ICN Switch 互联芯片、磐脉系列智能网卡、镇岳系列存储主控

下载日志已记录，仅供内部参考

芯片，实现了 AI 集群的核心芯片全覆盖。**阿里成为国内少数可以对标谷歌的厂商，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。**

真武芯片上市进度提前了半年。继 1 月发布真武 810E、5 月发布 M890 以后，阿里在 9 月 22 的云栖大会上发布了新一代自研芯片 V900，算力提升至真武 M890E 的 3 倍，显存提升至 216GB，带宽达 1200GB，支持 FP8 与 FP4 精度，并将在 2027Q1 上市，较原先计划提前了 2 个季度，因而我们认为阿里原计划 2028 年第三季度发布的 J900 芯片也有望提前上市。新芯片构建的千卡超节点，内置真武 V900、ICN Switch、磐脉智能网卡及镇岳 SSD 主控芯片，基于它构建的单一集群可扩展至 50 万卡，可支撑前沿模型的训练和推理。阿里真武 M890 超节点已成功适配 Qwen3.8、Kimi K3 等 2 万亿参数大模型（国内首个），并在 Q3 规模化上线阿里云数据中心。

平头哥自研 AI 芯片真武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；截至 4 月，这一数值进一步增长到 56 万片，服务了 400 多家客户 20 个行业，其中在智能驾驶领域部署数量超过 13 万片，金融行业部署超过 10 万片。对比 IDC 阿里国内出货 26.5 万片数据来看，已经有了 29.5 万片差额，这意味着 2026 年前 4 个月阿里 AI 芯片或已经超过了 2025 年全年出货量水平，同时平头哥实现了年化过百亿的营收规模。8 月 M890 超节点上线阿里云，截至 9 月真武系列芯片已服务超 650 家企业客户，覆盖超 20 个行业。

在 V900 发布的同时，阿里还宣布了 5 万亿至 10 万亿参数级别的下一代模型计划，并计划将全球数据中心容量从 2022 年的 2GW 提升至 2032 年的 20GW，同时明确把“超级智能”作为探索方向。

2026 年 7 月 23 日，阿里平头哥宣布开源自研 AI 软件栈 T-Head SAIL，这是继 2025 年 8 月华为 CANN 全面开源以后，国内又一大国产 AI 芯片 AI 软件栈的开源行为，将有助于国产芯片的适配，对抗英伟达在 CUDA 上的生态壁垒。

表 2：阿里真武芯片性能及发布计划

	810E	M890	V900	J900
发布时间	2026.1	2026.5	2026.9	
上市时间		2026.8	2027Q1	2028Q3
最大功率	400W			
互联带宽	700GB/s	800GB/s	1200GB/s	
内存容量	96GB	144GB	216GB	
内存	HBM2e			
数值类型	BF16/INT8	FP32-FP4	FP8/FP4	
Scale-up 规模	16 卡	64 卡	千卡	
单集群上限	万卡	13 万卡	50 万卡	
较上代性能提升		3 倍	3 倍	

资料来源：阿里，中原证券研究所

下载日志已记录，仅供内部参考

华为：9 月 17 日，在昇腾 AI 产业峰会上，华为昇腾 950 超节点步入规模商用阶段，同时业界首个采用 NPO 光引擎的超节点 960 正式发布。同时华为表示昇腾 960 将提前上市，实现性能翻番，其中昇腾 960DT 将于 2027Q1 发布，昇腾 960PR 将于 2027Q3 发布，后续昇腾 970 和 980 将分别在 2028 年和 2029 年发布。

表 3：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	昇腾980
量产时间	2025Q1	2026.4	2026.9	2027Q1/2027Q3	2028Q4	2029
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

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，2026 年 7 月累计落地 750 多套，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。

2026 年 7 月，昇腾 950 超节点（Atlas 950 SuperPoD）首次真机亮相，作为业界最大规模超节点，昇腾 950 超节点真机 1024 卡集群部署，提供 1 EFLOPS FP8 、2 EFLOPS FP4 澎湃算力，拥有 256TB 全局统一内存编址空间，依托 TB 级 NPU 互联超大带宽与 3μs 超低 RTT 时延，突破集群内计算、存储等各类资源的通信瓶颈，具备超大带宽、超低时延、统一内存编址三大核心技术优势，专为万亿级大模型训练与高并发推理设计。

表 4：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	2026.9	2027Q3
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

下载日志已记录，仅供内部参考

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

按照目前更新的计划，华为原本计划在 2026Q4 量产昇腾 950DT，将提前至 8 月上线华为云，在内存容量和带宽上较 950PR 实现较大提升。华为还将在 2027Q4 量产昇腾 960，2028Q4 量产昇腾 970，将国产芯片性能给带来持续提升的空间。4 月 24 日，DeepSeek-V4 发布的同时，DeepSeek 表示受限于高端算力，目前 Pro 的服务吞吐十分有限，预计下半年昇腾 950 超节点批量上市后，Pro 的价格会大幅下调。

图 23：DeepSeek V4 发布公告关于算力的表述



资料来源：DeepSeek，中原证券研究所

海光：2025 年海光发布深算三号，具有 FP64、FP32、FP16、TF32、BF16 和 INT8 等多算力精度，受到了顾客的认可，2026 年还有望发布深算四号。

在海光芯片的基础上，7 月中国首个全国产十万卡 AI 超集群曙光 8000 正式落成。2024 年系统研制，2025 年完成工程建设，今年 2 月郑州 3 万卡上线试运行，4 月扩展到 6 万卡 AI4S 集群，7 月 10 万卡部署落地，具有芯片、计算、存储、网络、散热、管理、服务等全链路全国产自研能力。中科曙光已同步将曙光 8000 接入国家超算互联网，以算力服务的方式向科研机构和产业界开放，7 月已完成 300 余项应用适配，覆盖材料、电磁、量子、生物医药、天文气象等 20 多个领域，累计完成 70 余次万卡规模算力测试。中科曙光还与北京科学智能研究院达成战略合作，正式启动第二套全国产十万卡超智融合算力系统研制与建设，正在开启从示范性工程向规模化复制的进程。同时，中科曙光也致力于在十五五期间，建成以曙光 9000 为基础的全球最大的超级科学智能体。

9 月 28 日，海光信息公告中科曙光董事长兼总经理历军担任公司董事长，这意味着两家公司在后续协同将进一步紧密。特别是在超节点成为发展趋势的大背景下，算力先进性的比拼已经从单芯片性能提升到超节点其他部件的整体协作，海光与曙光的结合将提供给市场更优的产

下载日志已记录，仅供内部参考

品组合。

表 5：海光芯片研发进程

类型	产品	启动研发	发布	核心数	线程	备注
CPU	海光一号	2016.3	2018	32	64	消化吸收
	海光二号	2017.7	2019	32	64	开启自主设计，浮点性能翻倍
	海光三号	2018.2	2022.6	32	64	C86-3G，整体性能提升 30%
	海光四号	2019.7	2023.10	64	128	C86-4G，首次采用自研 CPU 微架构
	海光五号	2022	预计 2026	128	512	C86-4G 增强版
DCU	深算一号	2018.10	2021	4096		7nm
	深算二号	2020.1	2023.9			性能提升超过 100%，推出了 AI 专用版，增强对 FP16 和 INT8 等低精度计算格式的支持
	深算三号 BW1000	2023.9	2025			FP64、FP32、FP16、TF32、BF16 和 INT8 等多算力精度，显存带宽与容量完美匹配千亿参数大模型训练与推理需求
	深算四号	2023				聚焦 AI 算力
	深算五号					

资料来源：海光信息公告，中国政府采购新闻网，中科南京信息高铁研究院，中原证券研究所

1.3.1.5. CPU III 级和 AI 芯片 II 级安全可靠测评认证首次发放

9 月 20 日，中国信息安全测评中心与国家保密技术测评中心发布《安全可靠测评结果公告（2026 年第 3 号）》，公布了安全可靠测评体系建立以来，第二批 AI 训练推理芯片的评测结果。其中有华为、百度、阿里、壁仞、海光、天数、沐曦、摩尔线程 8 家芯片通过了安全可靠测评，其中华为昇腾 950 成为首个 II 级安全可靠 AI 芯片，有望在后续采购中获得更大的竞争优势。

同时本次公布的 CPU 测评结果，龙芯 2 款和华为 1 款芯片获得了 III 级安全可靠认证，这也是首次发放 III 级安全可靠认证。

表 6：通过安全可靠测评的芯片产品

日期	类型	厂商	产品	安全可靠等级
2026.9.20	CPU	龙芯	龙芯3A6000(C)	III级
			龙芯3C3000	III级
		华为海思	鲲鹏950	III级
			麒麟9001X	II级
		兆芯	兆芯处理器KH-50000	II级
		中电科申泰	威鑫H8100型微处理器	II级
		海光	海光C86-4G/H	II级
	AI训练推理芯片	华为海思	昇腾950	II级
		百度	昆仑芯M100	I级
2026.5.26	AI训练推理芯片	华为海思	昇腾310	I级
			昇腾910	I级

下载日志已记录，仅供内部参考

		阿里平头哥	真武M530	I级
			真武M890	I级
		壁仞科技	壁砺™ 166	I级
		海光信息	DCU-3G	I级
		天数智芯	KCC-V100X 芯片	I级
		沐曦股份	MXC600	I级
		摩尔线程	PH100	I级
2025.9.12	CPU	龙芯中科	龙芯3B6000M	II级
			龙芯2K1500	I级
		华为海思	海思麒麟9000X	II级
		飞腾	飞腾腾锐D3000M	I级
2025.3.14	CPU	飞腾	飞腾腾云S5000C-E	II级
		龙芯中科	龙芯3B6000	II级
			龙芯3C6000	II级
		中电科申泰	申威威鑫 H8000	II级
			申威WY831（GC版）	I级
		华为海思	麒麟X90	II级
		兆芯	兆芯处理器KX-U6980S	I级
			兆芯处理器KX-6940S	I级
2024.9.30	CPU	兆芯	兆芯处理器KX-6000G	I级
			兆芯处理器KX-7000	I级
2024.5.20	CPU	飞腾	飞腾腾云S5000C	II级
			飞腾腾珑E2000	II级
			飞腾腾锐D3000	II级
		龙芯中科	龙芯3A5000（DA版）	II级
			龙芯3A6000	II级
			龙芯3C5000	II级
			龙芯3D5000	II级
			龙芯2K2000	I 级
		海光信息	海光处理器C86-4G	II级
			海光处理器C86-4G-L	I 级
		华为海思	鲲鹏920 V200	II级
			麒麟9000C	II级
		无锡先进技术研究院	申威SW-WY831型微处理器	I 级
		兆芯	兆芯处理器KH-40000	I 级
2023.12.26	CPU	华为海思	鲲鹏920	I级
			盘古M900	I级
			麒麟9006C	I级
			麒麟990	I级
		龙芯中科	龙芯3C5000L	I级
			龙芯3A4000/3B4000	I级
			龙芯3A4000/3B4000	I级
			龙芯3A5000/3B5000	I级
		无锡先进技术研究院	申威1621	I级
			申威SW421	I级

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

下载日志已记录，仅供内部参考

			申威3231	I级
		飞腾	飞腾腾锐D2000	I级
			飞腾FT-2000	I级
			飞腾FT-2000+	I级
			飞腾腾云S2500	I级
		海光信息	海光C86-3G	I级
			海光2号C86 3230/3250/3280/5280/7250/7260/7280/7285	I级
		兆芯	兆芯ZX-E KX-U6780A/KH-37800D/KX-6640MA/KX-6640A	I级
			兆芯ZX-D KX-U5580	I级

资料来源：中国信息安全测评中心，中原证券研究所

1.3.2. 华为给出了芯片发展的新路径，逻辑折叠在性能、成本、功耗方面表现优异

2026 年 5 月 25 日，华为半导体业务部总裁何庭波在国际电路与系统研讨会（ISCAS 2026）上，正式发布“韬（τ）定律”，同时，华为还在当天发布了以何庭波为署名作者的论文《A Time Scaling Theory for Multi-Layer Electronic Systems》。7 月 3 日，该论文的更新版本发布，并公布了更多工程细节和芯片性能实测数据。9 月 4 日，何庭波又发布了《Huawei’s τ Chip Was Supposed to Melt?》，重点解释了 τ 芯片为何不因堆叠而过热的问题。

在受到美国禁令影响先进支撑受限的背景下，华为基于过去 6 年量产的 381 款芯片的实践提出了“韬（τ）定律”，其核心思想是用时间缩放替代“摩尔定律”的几何缩放，在当前摩尔定律逐渐失效、光刻工具接近物理极限的困境下，为半导体行业找到了新的发展路径。

而对于智算中心而言，其通过成百上千个芯片协同工作，且当前 AI 发展中，算力使用的瓶颈更多呈现在数据传输中。因而智算中心的 τ 缩放中，华为通过系统架构（统一总线）、近封装光学引擎（高密度光互联节点引擎 Hi-ONE）、封装本身的拓扑重组（3D 折叠）3 个层面来提升。其中 3D 折叠将改变在传统 2.5D 人工智能芯片中，逻辑芯片被 HBM、光 I/O、电源环绕逻辑芯片的布局，将迁移到垂直表面，形成垂直集成堆叠的结构，并实现同步拓展。约在 2030 年，昇腾 990 将把逻辑折叠引入到 AI 芯片中，之后 3D 折叠将成为 2035 年以前智算中心 τ 缩放的主要载体。2026-2035 年，预计在这个过程中硬件集成度的提升将超过 100 倍。

对于逻辑折叠而言，华为手机芯片的实践已经验证了该方向的发展潜力片。2026 年开始，华为借助逻辑折叠开启手机芯片的 τ 缩放，麒麟 2026 芯片（2026 年下半年伴随 Mata 90 机型发布）借助局部关键路径折叠，将可以把晶体管密度从 155 MTr/mm²提升至 238 MTr/mm²，这是过去需要 3 年几何级数扩展才能实现的提升幅度，芯片核心频率也从 2.75GHz 提升至 3.1GHz。

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

下载日志已记录，仅供内部参考

图 24：麒麟 CPU 性能核心工作频率发展趋势

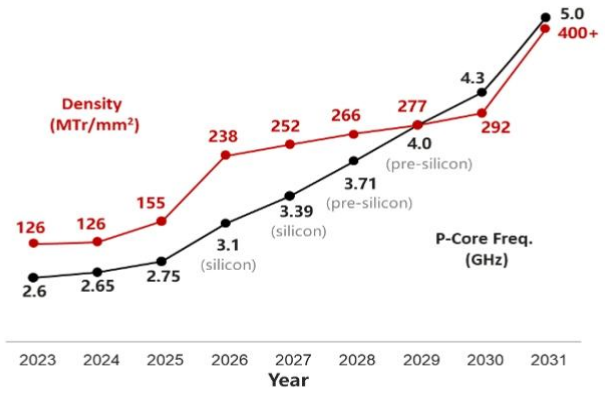
	SoC	Architecture	Frequency (GHz)	State
2023	Kirin9000s	Planar	2.6	Mass product
2024	Kirin9020	Planar	2.65	Mass product
2025	Kirin9030 pro	Planar	2.75	Mass product
2026	Kirin 2026	LogicFolding	3.1	Silicon
2027	Kirin 2027	LogicFolding	3.39	Silicon
2028	Kirin 2028	LogicFolding	3.71	Pre-silicon
2029	Kirin 2029	LogicFolding	4	Pre-silicon

资料来源：华为，中原证券研究所

未来 10 年，逻辑折叠有望从局部关键路径扩展到全规模多层折叠，且在成本方面具有经济效益。预计到 2031 年，基于韬(τ)定律的高端芯片晶体管密度将达到 1.4 纳米制程的同等水平。值得关注的是，华为的 τ 缩放路线图还是在成本降低 30%的前提下实现的。

相对于前代的麒麟 9030 Pro 而言，通过逻辑折叠带来的性能提升，如果麒麟 2026 将其电压降低来实现与麒麟 9030 Pro 相同的性能，那么在等性能下，麒麟 2026 的功耗降低 41%，同时功率密度下降 5.6%。

图 25：麒麟芯片的晶体管密度和性能核心工作频率预测



资料来源：华为，中原证券研究所

图 26：麒麟 2026 和麒麟 9030 Pro 在等效性能下的功率对比

	Kirin9030 Pro	Kirin2026
Temperature	25 ℃	25 ℃
Voltage	1.1 V	0.9 V
Frequency	2.75 GHz	2.5 GHz
Normalized power at iso-performance	1	0.59
Normalized area	1	0.625
Normalized power density	1	0.944

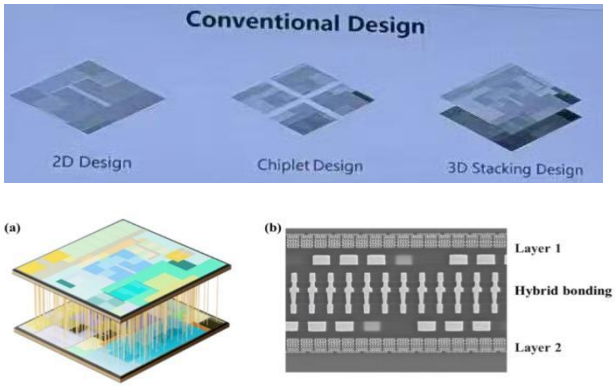
资料来源：华为，中原证券研究所

传统的 3D 堆叠路线是将独立的芯片进行物理堆叠；而在逻辑折叠架构中，上下有源层不再是独立的单芯片，而是不可分割的单个芯片的组成部分。上下逻辑层通过整体设计，利用极高密度的面对面混合键合技术，实现了接近芯片内部互联级别的垂直连接，满足了高效灵活的逻辑设计需求，进而大幅提升了有效晶体管密度。同时，华为的实践表明，相对于 2D 解决方案，逻辑折叠可以实现更好的成本效益。

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

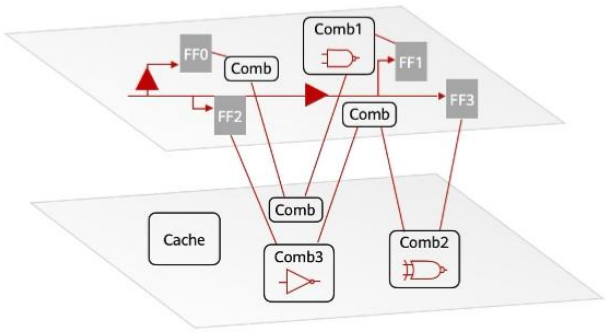
下载日志已记录，仅供内部参考

图 27：传统设计与韬定律的差异



资料来源：华为，中原证券研究所

图 28：逻辑折叠示意图



资料来源：华为，中原证券研究所

在散热方面，逻辑折叠同样展现出了优异的表现。由于功耗不光用于计算，在典型智能机负载下，九成用于数据动态传输的过程。因而逻辑折叠在重构电路空间关系的同时，也可以通过缩短连接和传输距离，达到降低功耗和控制散热的目的。在相同性能下，麒麟 2026 在基准测试中，NPU 功耗下降 66%，GPU 下降 58%，性能核心 CPU 下降 41%。

在逻辑折叠的基础上，3D 折叠也将成为长期发展趋势，带来逻辑与内存从解耦走向再融合。当前逻辑芯片通过 2.5D 扇出承载的内存、互联架构、电源，信号和电流都要通过芯片边缘到达计算资源内部，因而会带来计算能力（与面积 N^2正相关）与内存带宽、互联、供电（与周长 4N 正相关）发展的不匹配。最终芯片将走向 3D 折叠，将供电、存储、光学传输从外围迁移至芯片表面，通过从解耦到在融合的方式，实现资源根据需求实现可扩展性的释放。

1.3.3. 操作系统：鸿蒙 7 开启正式版升级，鸿蒙终端数量突破 9500 万

继 2024 年 10 月华为纯血鸿蒙推出以来，已经进行了 3 个大的版本更新，其中在 2026 年 6 月推出的鸿蒙 7，已经将盘古大模型深度嵌入到了系统内核，成为继进程管理、内存管理、文件系统后的第四个核心子系统，同时 9 月 28 日鸿蒙 7 系统开启正式版升级。

相对于鸿蒙 6 而言，鸿蒙 7 的 AI 功能的提升主要体现在 2 个方面：

AI 的实现方式：鸿蒙 6 中 AI 功能实现主要依靠云端，而鸿蒙 7 可以实现复杂任务的本地 NPU 推理，既减少了延迟，又减少了向云端传输可能带来的隐私风险。鸿蒙 7 中盘古端侧大模型实现了 30B 的参数落地，而此前端侧模型参数普遍在 1-3B 之间，给端侧 AI 能力带来了极大的提升。随着 2026 年下半年麒麟 2026 和 Mate 90 的发布，鸿蒙 7 优势将得到芯片和终端支撑，有望与同行产生代差。

主动性和多端联动：鸿蒙 6 中小艺主要通过呼唤方式调取，难以实现多应用的调取；而鸿蒙 7 主动研判用户需求，可以实现多任务执行和 APP 调用，还能实现多端联动，进一步发挥鸿蒙全场景的系统优势。

下载日志已记录，仅供内部参考

截至 2026 年 10 月 21 日，鸿蒙操作系统终端数已经突破 9500 万，华为预计 2026 年突破 1 亿用户规模。

表 7：鸿蒙纯血系统发展动态

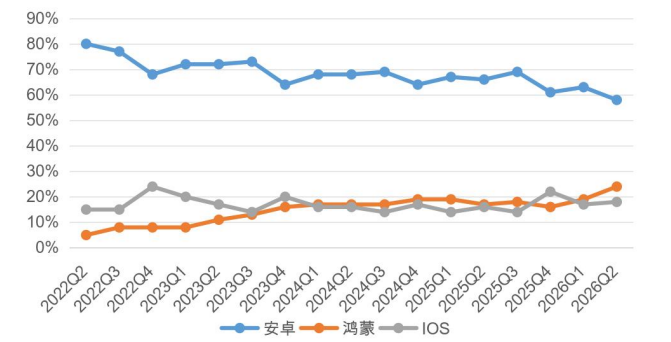
类型	时间	事件
系统发布	2024. 10. 22	纯血鸿蒙 5.0（HarmonyOS NEXT）正式发布，彻底摆脱安卓，并在手机端应用
	2025. 5. 19	首款鸿蒙电脑搭载鸿蒙 5.0 操作系统正式发布
	2025. 6. 20	正式发布 HarmonyOS 6，并同步启动开发者 Beta 测试。
	2025. 9. 27	鸿蒙 6.0 Release 版本正式发布
	2025. 10. 22	HarmonyOS 6 发布
	2026. 6. 12	HarmonyOS 7 发布。
生态建设	2025. 5. 8	鸿蒙电脑有 300 多款应用完成适配
	2025. 5. 19	鸿蒙电脑有 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
	2026. 6. 12	华为应用市场可获取应用和服务突破 40 万款，其中超过 2.3 万款应用已完成鸿蒙原生适配，平台注册开发者超过 1100 万
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。
	2026. 5. 14	支持鸿蒙 6 操作系统终端数量突破 6000 万
	2026. 6. 12	支持鸿蒙 6 操作系统终端数量突破 6600 万，生态设备总量超过 11.9 亿台
	2026. 7. 2	支持鸿蒙 6 操作系统终端数量突破 7000 万
	2026. 8. 20	支持鸿蒙 6 操作系统终端数量突破 8000 万
	2026. 9. 7	鸿蒙终端设备数量突破 8500 万
	2026. 10. 1	鸿蒙终端设备数量突破 9000 万
	2026. 10. 21	鸿蒙终端设备数量突破 9500 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

2026Q2 鸿蒙在手机操作系统市场份额回升明显，在中国市场重回第二。根据 Counterpoint 的数据，2026Q2 国内手机操作系统市场上，鸿蒙份额 24%，较上个季度上升 5 PCT，较 IOS 的优势进一步拉大，稳坐国内第二位置。全球市场来看，鸿蒙同样表现突出，份额继续保持 5%，继续位居全球第三，与 IOS 的差距缩小 2 PCT。

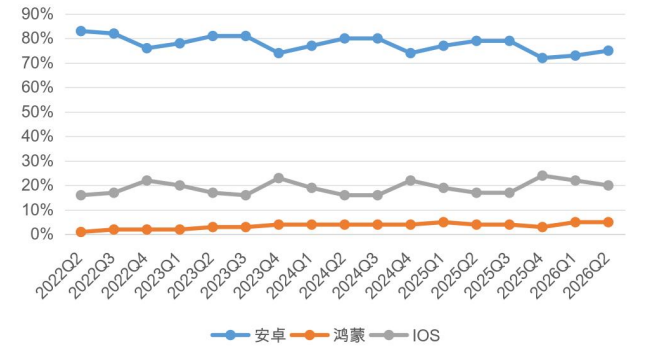
下载日志已记录，仅供内部参考

图 29：2022Q2-2026Q2 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 30：2022Q2-2026Q2 全球手机操作系统市场份额



资料来源：Counterpoint，中原证券研究所

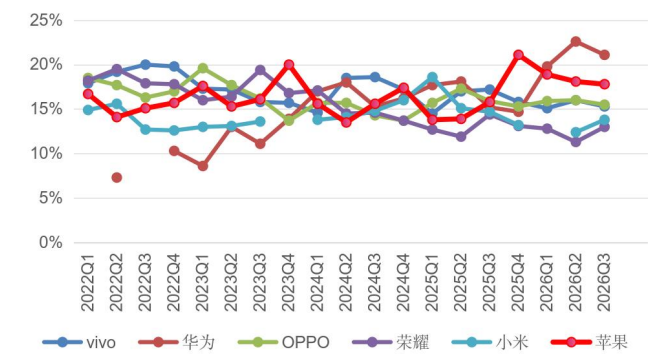
2026 年，华为的手机、PC 都实现了较高的增速水平，下半年有望在芯片、终端、操作系统、AI 模型的全面加持下再入佳境。

（1）手机：根据 IDC 数据，2026Q3 华为以 21.1%的出货量份额位居中国手机市场第 1，在行业下滑 15.4%的背景下，同比增长 17.3%，是唯一一个实现两位数增长的厂商。

（2）台式机和笔记本：2026Q2 华为位居国内台式机和笔记本市场第二，份额 16%，较上个季度增长与上季度持平，以 83%的增速明显领先其他头部厂商。

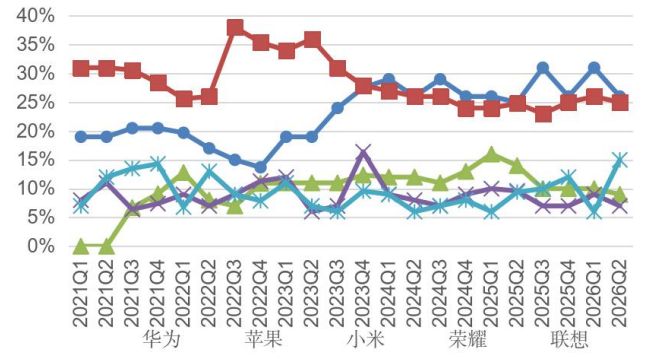
（3）平板电脑：2026Q2 华为在国内平板电脑市场占据 26%的份额，位居第一，较 Q1 下降了 5PCT，以-18%的增速落后于行业总体水平。2026Q2 华为在全球平板电脑市场份额 7.7%，同比下滑 16%，表现不及行业总体水平。

图 31：2022Q1-2026Q3 中国手机市场厂商份额



资料来源：IDC，中原证券研究所

图 32：2021Q1-2026Q2 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、鸿蒙智慧[illegible]、重构 AI 数据中心、液冷超快充。**

下载日志已记录，仅供内部参考

1.4. AI：头部模型继续呈现降价趋势

1.4.1. 中美大模型发展概况

第一梯队厂商来看：

（1）Anthropic：9 月新推出的 Opus 5.5、Sonnet 5.5 和 Fable 5.1 全球领先；

（2）Meta：在 10 月发布了 Muse Wonder Full，成为了全球排名第 3 的模型，成绩显著；

（3）月之暗面：继 7 月 Kimi K3 的优异表现以后，月之暗面在 10 月发布了 Kimi K3.1 Pro，拥有 2.8 万亿参数，通过后训练加强和更多领域数据的强化学习，再次回归头部模型行列；

（4）OpenAI：9 月发布的 GPT-6 Astra 位居其后；

（5）谷歌：在 10 月推出 Gemini 4 Argon，也借此重回大模型第一梯队行列。

第二梯队厂商中：

（1）xAI：在 9 月推出 Grok 4.7，在速度和价格方面较 Anthropic 和 OpenAI 模型具有优势；

（2）小米：在 9 月推出了 MiMo-V2.6-Pro，位居国产模型和开源模型第二；

（3）阿里：作为更新频率较快的厂商，阿里 9 月发布的 Qwen3.8-Max，紧随 MiMo-V2.6-Pro，位居国产第三；

（4）DeepSeek：在 9 月 10 日推出了 DeepSeek-V4.1 Falsh，在 10 月 17 日又进行了更新，代码和 Agent 能力明显提升，同时缩短了任务运行时间和输出长度；

（5）智谱：8 月推出的 GLM-5.3 与 Qwen3.8-Max 智能指数持平，同在 8 月推出的 GLM-5.3-Flash 拥有了性价比，同时具有多模态能力；

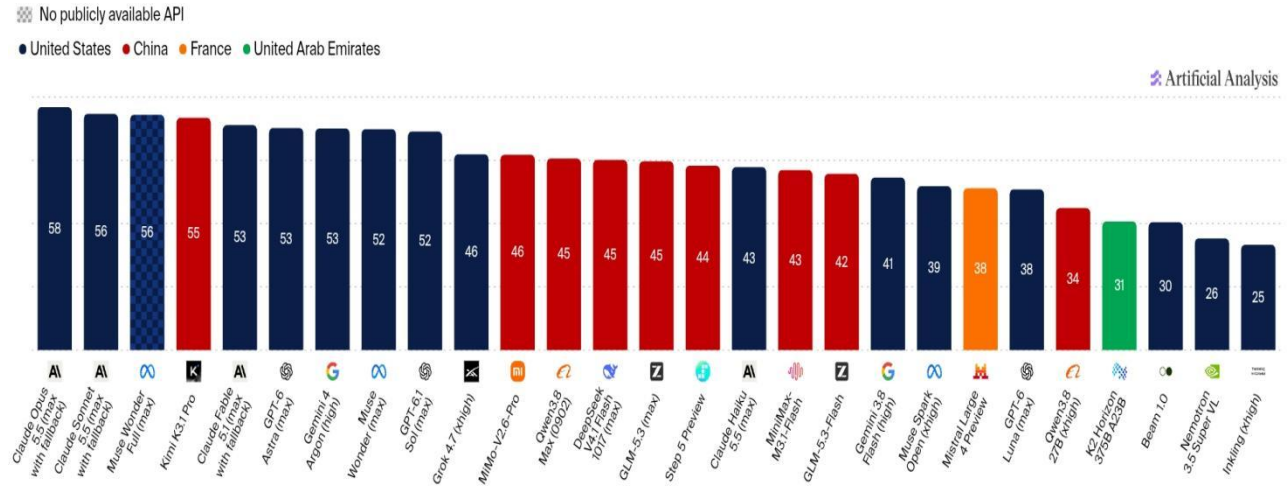
（6）MiniMax：10 月推出了 MiniMax-M3.1-Flash，定位为 Coding/Agent 轻量级模型。

总体来说国内模型能力在线，呈现多点开花的发展格局，对海外模型的可替代性较强。从模型能力对比来看，中国模型能力总体处于全球第二梯队，略落后于美国模型厂商。由于中美算力供给的差距，我们认为当前的竞争格局还将继续维持数月。



下载日志已记录，仅供内部参考

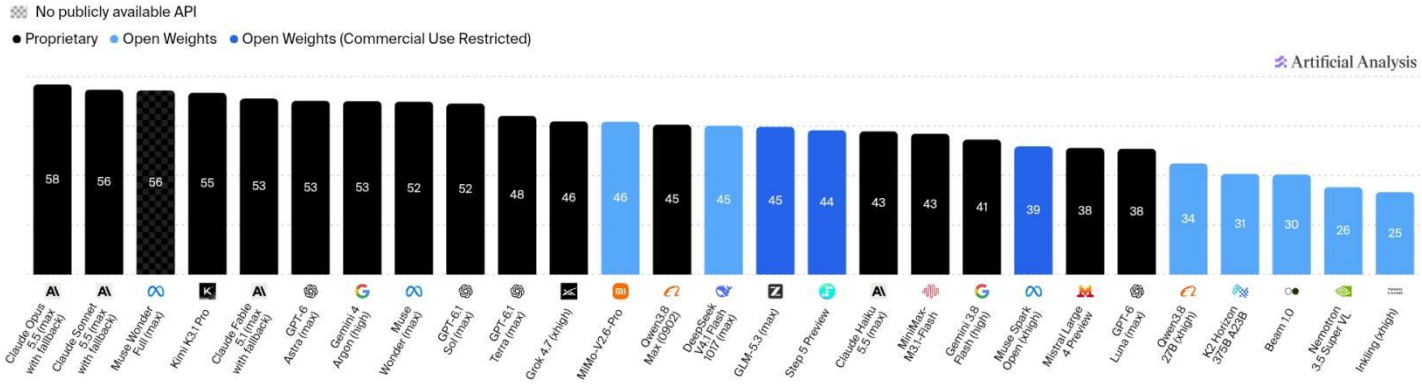
图 33：不同地区的大模型智能水平对比（截至 2026.10.26，红色为中国模型）



资料来源：Artificial Analysis，中原证券研究所

从发展路线来看，美国模型聚焦闭源路线，中国模型处于开源的第一梯队。OpenAI、xAI、Meta 在早期的主要模型上都进行了开源，而后转回了闭源路线。而国内模型厂商成为了开源路线的坚定执行者，阿里千问的开源模型在 Hugging Face 社区下载量第一；DeepSeek 坚持全栈开源，开启了开源模型极致降本增效发展路径，推动技术平权和传统 AI 巨头的技术封闭生态；Kimi-K3.1 Pro、MiMo-V2.6-pro、Qwen3.8 Max 能力逼近海外顶级闭源模型。**由于开源模型生态强大，加快应用落地，也有助于模型更加快速的发展，让国产开源模型与海外顶级模型的比拼中，占据了更大的主动权。**

图 34：大模型智能水平及开源情况（截至 2026.10.27）



资料来源：Artificial Analysis，中原证券研究所

1.4.2. 主流模型价格对比

在模型成本方面，国产模型价格优势非常明显。从 2023 年至今，国内模型始终就面临缺芯难题，更加关注软硬件协同发展，并带来了极致的性价比。以美国最强模型 Anthropic 的 Opus 5.5 的调用价格来看，其输入输出价格都约为同等级模型 Kimi-K3.1 pro 的 2 倍。但是根据 The Information 7 月的报道，在低廉的价格策略之下，DeepSeek-V4 的 API 业务仍然保持了超过 50%的毛利率水平。

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

下载日志已记录，仅供内部参考

在国产模型冲击和用户成本敏感度提升的背景下，近期美国闭源模型降价趋势明显，Anthropic 7 月发布的 Opus 5 仅为 Fable 5 的一半价格，9 月 Opus 5.5 又在此基础上再降低了 20%；OpenAI 也在 7 月 31 日对 GPT-5.6 Terra 和 Luna 进行了降价，降幅为 20%和 80%，8 月 21 日又对 GPT-5.6 Sol 下调 20%，9 月 23 日新推出的 GPT-6 Sol 和 Luna 再次降价约 50%，10 月 26 日推出的 GPT-6.1 Terra 也较上一代降价 50%；谷歌 8 月推出的 Gemini 3.7 flash 也为原 3.6 Flash 的一半；10 月 24 日，月之暗面的 Kimi-K3.1 pro 实现了加量降价，在能力提升的同时，价格降低了 1/3。

表 8：模型调用价格（每百万 Token，Artificial Analysis2026 年 10 月 27 日智能指数）

厂商	模型	发布时间	智能指数	输入	缓存输入	输出	缓存输出
OpenAI	GPT-6 Astra	2026.9	53	$10	$1	$50	
	GPT-6.1 Sol	2026.9	52	$2	$0.1	$10	
	GPT-6.1 Terra	2026.10	48	$1	$0.05	$5	
	GPT-6 Luna	2026.9	38	$0.1	$0.01	$0.5	
Anthropic	Opus 5.5	2026.9	58	$4	$0.20	$20	$5
	Sonnet 5.5	2026.10	56	$2	$0.1	$10	$2.5
	Fable 5.1	2026.9	53	$10	$0.25	$50	$12.5
	Haiku 5.5	2026.10	43	$0.1（0-100K） $0.5（100K 以上）	$0.01 $0.05	$0.5 $2.5	$0.125 $0.625
Meta	Muse Wonder	2026.10	52	$4		$20	
	Muse Spark 1.3	2026.9	48	$1.25	$0.15	$4.25	
谷歌	Gemini 4 Argon	2026.10	53	$2		$10	
	Gemini 3.8 flash	2026.9	41	$0.75		$3.75	
xAI	Grok4.7	2026.10	46	$2		$6	
Kimi	Kimi-K3.1 pro	2026.10	55	￥13	￥1.3	￥68	
阿里	Qwen3.8-max 0902	2026.9	45	￥12		￥36	
	Qwen3.8-27b	2026.8	34	￥3		￥12	
智谱	GLM-5.3	2026.8	45	￥8	￥2	￥28	
	GLM-5.3 Flash	2026.8	42	￥0.8	￥0.23	￥2.8	
DeepSeek	DeepSeek V4.1 Flash 1017*	2026.10	45	￥2	￥0.04	￥8	
小米	MiMo-V2.6-pro	2026.9	46	￥3	￥0.025	￥6	
	MiMo-V2.6 flash	2026.9	38	￥1	￥0.02	￥2	
MiniMax	MiniMax-M3.1 Flash	2026.10	45	￥1	￥0.1	￥4	
	MiniMax-M3	2026.6	30	￥4.2（0-512K） ￥2.1（512K-1M）	￥0.84 ￥0.42	￥16.8 ￥8.4	

资料来源：Anthropic，OpenAI，谷歌，xAI，阿里，智谱，Kimi，MiniMax，DeepSeek，小米，Artificial Analysis，中原证券研究所（*空闲时段执行半价）

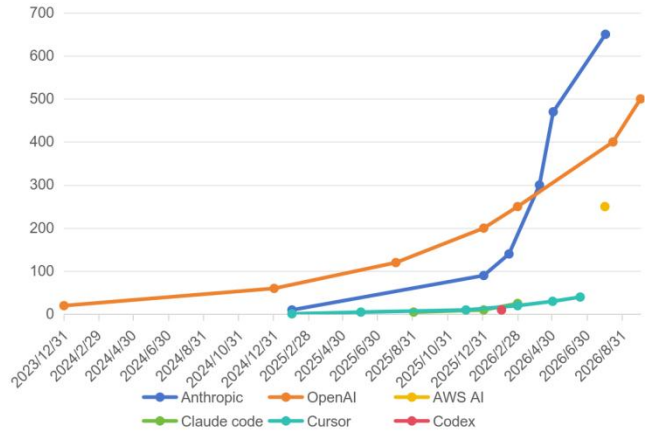
1.4.3. 中美模型厂商收入

从模型收入情况来看，美国模型因为价格高昂，能力最强，Agent 能力突出，实现了较大

下载日志已记录，仅供内部参考

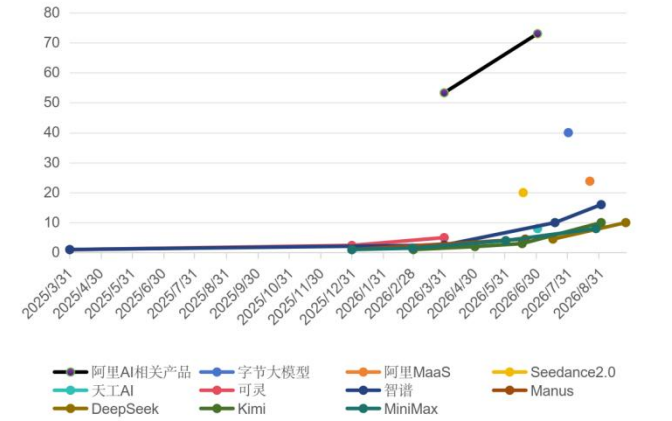
的收入规模。依靠 AI 编程领域的主导地位，龙头 Anthropic 2026 年收入快速增长，年化收入从年初的 10 亿美元，7 月底达到了 650 亿美元，在 Q2 调整后营业利润为正。OpenAI 年化收入在 2026 年 9 月达到 500 亿美元，但 Q2 亏损进一步加大至 123 亿美元。中国模型厂商中，阿里巴巴 AI 相关产品 6 月年化收入约 73 亿美元，字节跳动 7 月 30 日披露大模型年化收入 40 亿美元，根据 the Information 9 月 24 日报道，DeepSeek 年化收入达到 10 亿美元，智谱 8 月年化收入达到 16 亿美元，Kimi 8 月年化收入突破 10 亿美元，MiniMax 8 月年化收入达到 8 美元，体量上和美国模型厂商差距明显。

图 35：海外模型及 AI 应用的年化收入（亿美元）



资料来源：Anthropic，OpenAI，Cursor，中原证券研究所

图 36：国内模型及 AI 应用的年化收入（亿美元）



资料来源：字节，智谱，Manus，DeepSeek，Kimi，可灵，昆仑万维，中原证券研究所

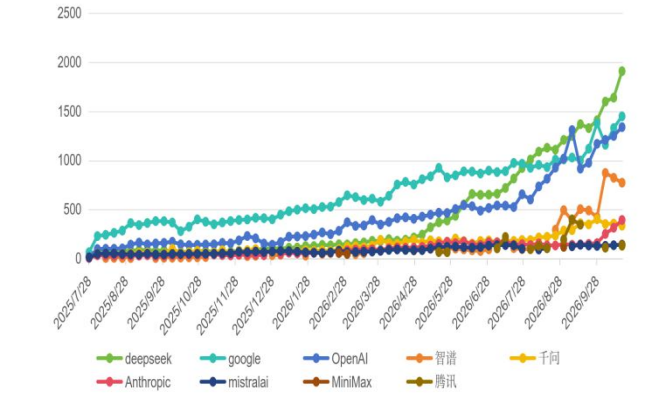
1.4.4. 厂商模型调用及国产模型替代情况

出于成本压力，海外用户对国产模型的需求提升。6 月 16 日消息，微软在成本压力下，先是将 Copilot Cowork 从无限制使用变为了 Token 计费，又在考虑在即将推出的 Copilot Cowork 低成本版本中用 DeepSeek V4 或其他开源模型代替 Ahthropic 和 OpenAI 的模型。7 月 3 日，微软旗下编程助手 GitHub Copilot 正式接入了 Kimi K2.7 Code，这是其首次接入开源模型。Kimi K2.7 Code 将逐步向 Copilot Pro、Pro+以及 Max 订阅方案开放，并将把支持范围扩大至 Copilot Business、Enterprise 版本及其他应用平台。8 月，路透社报道月之暗面正在与微软、亚马逊谷歌展开 Kimi K3 模型收入分层的谈判。

切换国产模型正在成为模型调用发展的趋势。根据 OpenRouter 的模型调用数据（2026 年 10 月 19 日-25 日），DeepSeek 份额 26%，连续 12 周位列第一大模型厂商。同时中国模型厂商份额达到 46%，暂领先较美国模型厂商（份额 44%）。

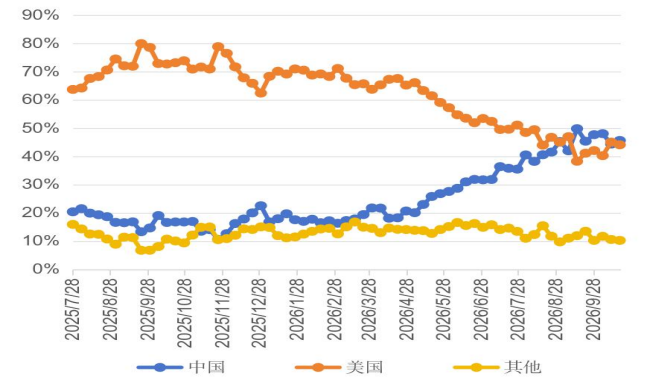
下载日志已记录，仅供内部参考

图 37：OpenRouter 各厂商模型调用量（百万 tokens）



资料来源：OpenRouter，中原证券研究所

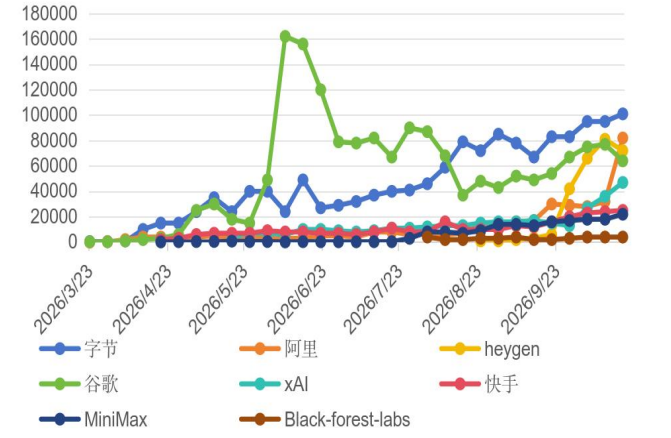
图 38：OpenRouter 调用的模型区域份额



资料来源：OpenRouter，中原证券研究所

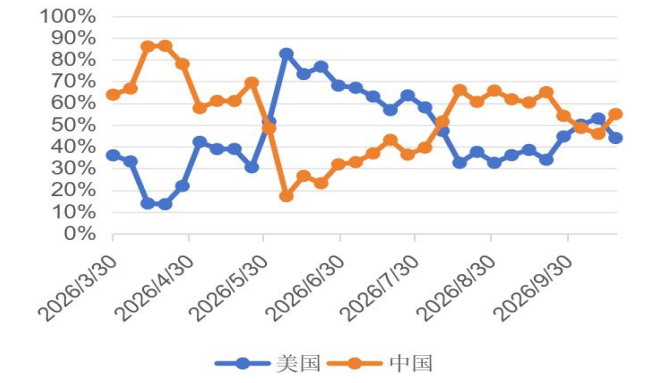
我国在视频模型调用量同样领先。根据 OpenRouter 的模型调用数据（2026 年 10 月 19 日-25 日），字节份额 24%，连续 10 周位列第一大模型厂商。同时中国模型厂商份额达到 55%，暂领先较美国模型厂商（份额 44%）。

图 39:OpenRouter 各厂商视频模型调用量(百万 tokens)



资料来源：OpenRouter，中原证券研究所

图 40：OpenRouter 调用的视频模型区域份额



资料来源：OpenRouter，中原证券研究所

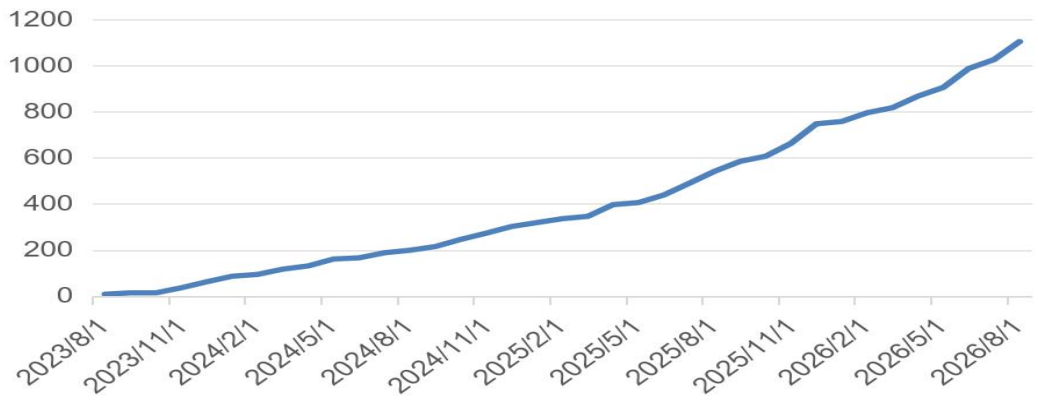
2026 年国产芯片性能提升，超节点优势持续释放，国产模型加速向国产芯片进行推理端和训练端算力的转移，模型端与国产芯片的协同效应更加紧密，后续产业发展的稳定性将持续提升。考虑到 2026 下半年昇腾 950 超节点批量上市后，DeepSeek 的价格有望下调，国产模型的性价比将进一步凸显，或将加速海外闭源模型厂商客户流失。

1.4.5. 模型与应用：大模型备案高度集中在发达区域

根据网信办数据，截至 2026 年 8 月，全国大模型备案与登记工作推进取得显著进展，目前已经有 30 个省市及国资委共计 1105 款大模型完成备案，另有 731 款大模型产品在属地完成登记。

下载日志已记录，仅供内部参考

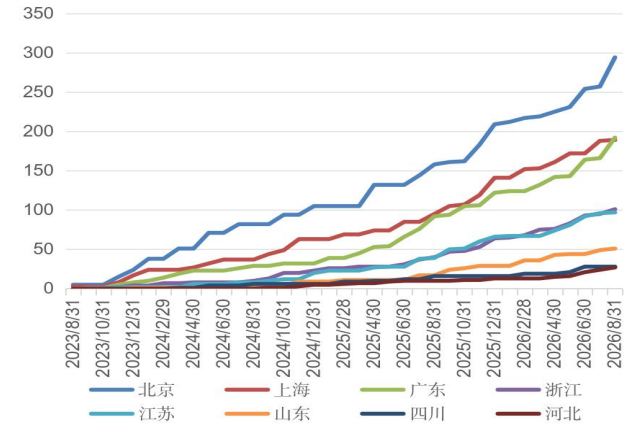
图 41：全国大模型备案与登记数量



资料来源：国家网信办，中原证券研究所

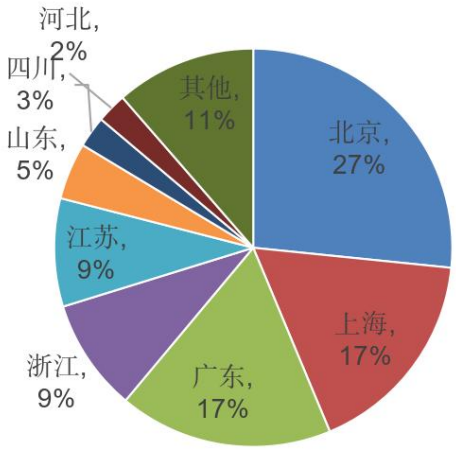
从地域分布看，北京市完成备案大模型数量遥遥领先，总计 294 款，占比 27%；上海市和广东省紧随其后，分别备案 189 款和 192 款，占比均为 17%，三地合计占比约为六成，成为我国大模型发展的核心区域。浙江省和江苏省分别备案 101 款和 97 款，均占比 9%，构成第二梯队。京津冀、长三角、珠三角等经济发达地区在政策支持、技术资源和产业基础上具备明显优势，成为大模型创新和应用的高地。

图 42：各地大模型备案与登记数量



资料来源：国家网信办，中原证券研究所

图 43：大模型备案与登记数量的区域分布



资料来源：国家网信办，中原证券研究所

1.4.6. 科技企业裁员：2026 年前 9 个月美国因为 AI 取代占总裁员的比例达到 21%

2026 年海外主要科技厂商都出现了较多的裁员，根据不完全统计，其中亚马逊约 3 万人，Meta 约 9500 人，微软超过 7600 人（10 月还有 6 万人收到自愿离职方案未计入），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。**

下载日志已记录，仅供内部参考

表 9：海外主要科技厂商近期裁员情况

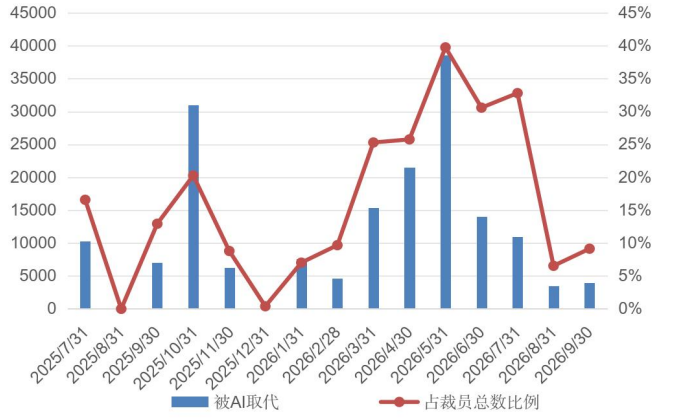
厂商	时间	人数	备注
亚马逊	2025.1	约 200	在北美商店部门裁员约 200 人，涵盖包括自有品牌、Prime 会员计划和消费品业务部门
	2025.10.28	约 14000	将裁减约 1.4 万个企业岗位，占其约 35 万名企业职员的 4%，此次裁员旨在提高决策效率，并将资源转向核心业务
	2026.1.28	约 16000	宣布调整组织架构，确认将裁减约 1.6 万个工作岗位。
	2026.4.7	约 14000	计划在 2026 年 5 月启动新一轮全球裁员，预计约 1.4 万名员工将受到影响。此次裁员覆盖 AWS 云服务、零售、人力资源等多个核心业务板块，中国区部分团队甚至可能被整体裁撤。
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，本次裁员波及多个团队，其中包括 GTIG 威胁情报小组、
	2026.6		谷歌已在过去两周内，对其云服务部门进行新一轮裁员，包括 GTIG 威胁情报小组、Mandiant 等，目的是向 AI 等增长领域重新投入资源。
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。
	2026.4	超 2625	微软推出公司历史上首个自愿退休计划，面向约 8750 名符合条件的美国员工，最终超过 30%选择接受。
	2026.6	200	微软中国启动新一轮裁员这波调整主要集中在 Azure 云业务研发团队，超 200 名员工受到影响。
	2026.7	4800	微软宣布裁员约 4800 人，约占员工总数的 2.1%。
	2026.10		向约 60000 名员工（超 25%）开放自愿离职方案，以压缩成本、调配资源加码 AI
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。
	2026.9.23		一轮裁员正式开始。一份内部文件显示，一些团队的裁员计划幅了两位数百分比。

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

下载日志已记录，仅供内部参考

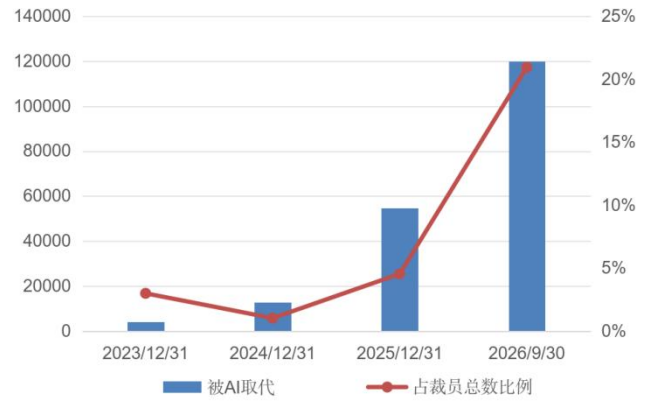
2026 年前 9 个月美国因为被 AI 取代占总裁员比例的 21%，8、9 月 AI 替代导致的裁员有所缓解。Challenger, Gray & Christmas 的数据显示，2026 年前 9 个月，美国因 AI 被裁员的人数总计达到 12.0 万人，占总计裁员比例的 21%。其中 9 月因 AI 被裁员人数 3961 人，占总裁员人数的 9%，处于近期低位水平。

图 44：2025.7-2026.9 美国因 AI 被裁的人数及占裁员总数的比例（当月值，人）



资料来源：Challenger, Gray & Christmas，中原证券研究所

图 45：2023-2026 年美国因 AI 被裁的人数及占裁员总数的比例（累计值，人）

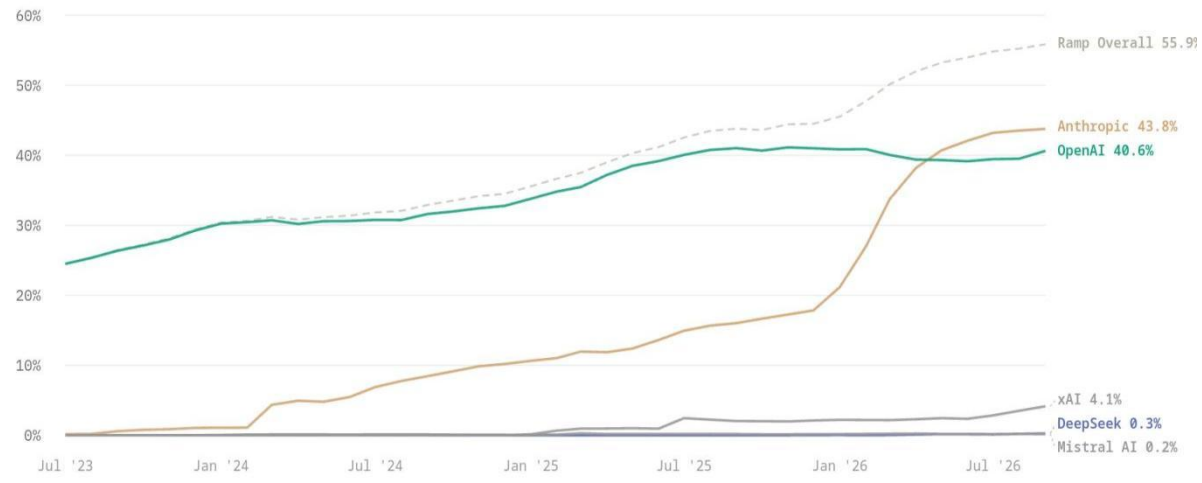


资料来源：Challenger, Gray & Christmas，中原证券研究所

1.4.7. 企业端 AI 应用情况

Anthropic 在企业业务领域成为美国第一大模型供应商。根据企业费用管理平台 Ramp 2026 年 9 月的数据，Anthropic 以 43.8%的份额再次超过 OpenAI（40.6%），成为了企业市场最大的模型供应商。

图 46：Ramp AI 指数（美国企业模型采用率）



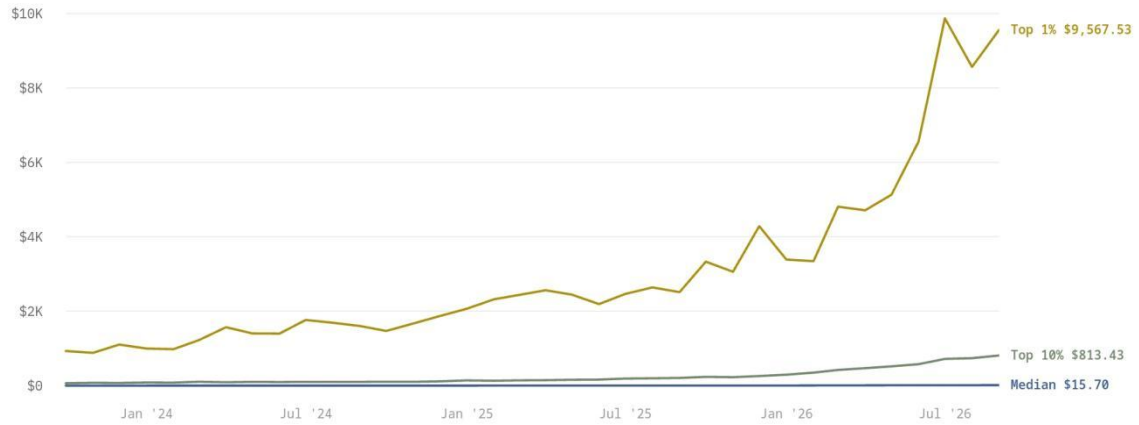
资料来源：Ramp，中原证券研究所

在 AI 用户的支出方面，top 1%的公司每个员工每月支出在 9 月略有回升至 9567 美元，top 10%的公司每个员工每月支出金额为 813.43 美元，环比提升了 4.8%；AI 中位公司每个员工每月支出为 15.7 美元每人每月，环比增长 5.2%。

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

下载日志已记录，仅供内部参考

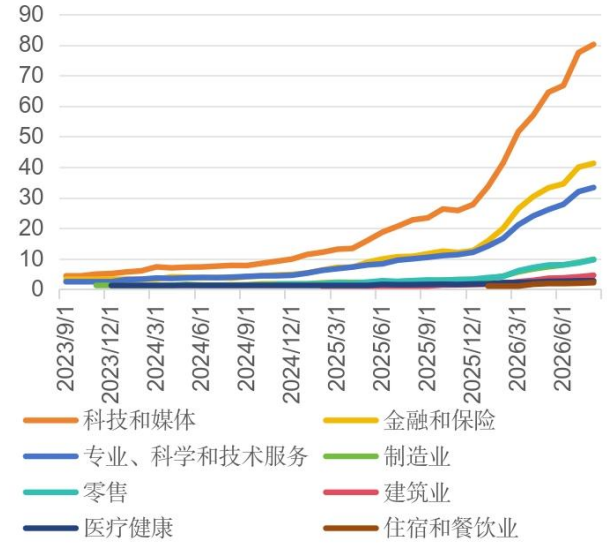
图 47：美国企业每个员工每月的 AI 支出（美元）



资料来源：Ramp，中原证券研究所

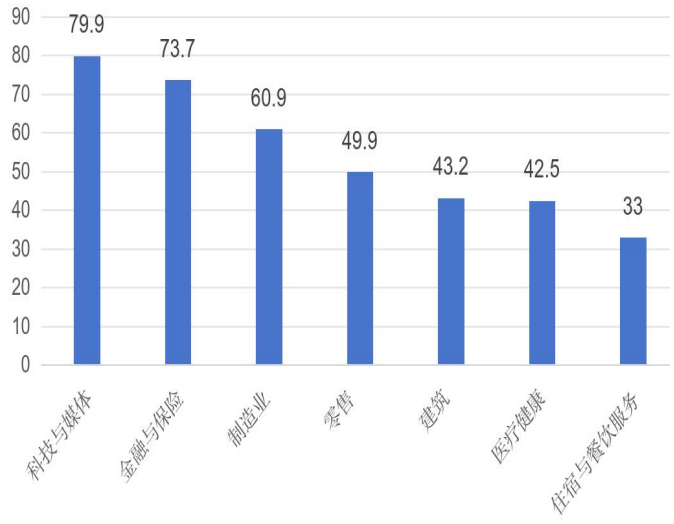
分行业来看，科技和媒体、金融和保险、制造业都已经有 50%以上的企业在进行 AI 相关投入。但是从金额上来看，科技和媒体、金融各行业员工的 AI 支出中位数明显领先其他行业，这与当前编程、视频领域 AI 的深度应用紧密相关；金融和保险行业无论还是普及度和员工 AI 开支都位居行业第二；专业、科学和技术服务在 AI 员工开支方面位居第三；制造业虽然 AI 普及率较高，但是人均贡献金额处于较低水平。

图 48：美国各行业员工每月的 AI 支出中位数



资料来源：Ramp，中原证券研究所

图 49：2026 年 9 月美国各行业在人工智能模型、订阅、代币和 API 上投入的企业占比



资料来源：Ramp，中原证券研究所

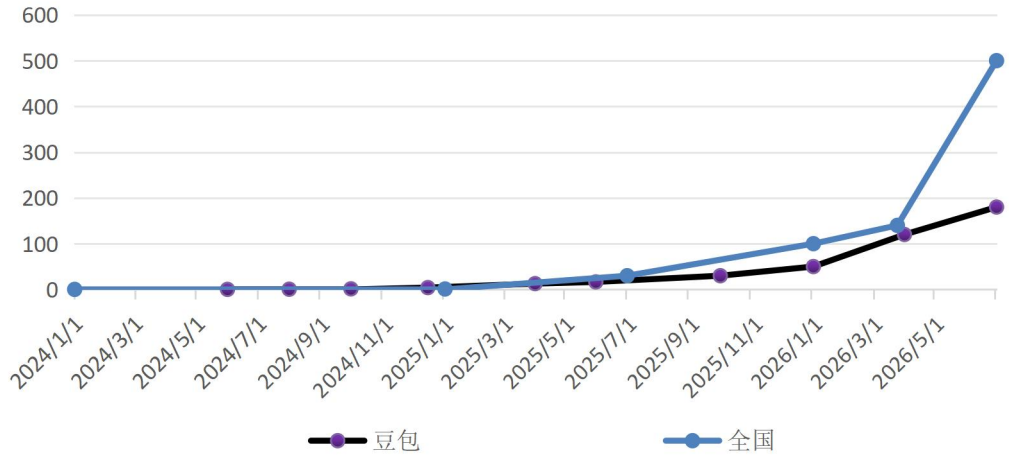
1.4.8. 国内 Token 使用情况

根据工信部和国家数据局的数据，截至 2026 年 6 月，我国日均词元调用量已突破 500 万亿，较 3 月的 140 万亿有数倍增长，较 2024 年初的 1000 亿和 2025 年底的 100 万亿都有明显的提升。

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

下载日志已记录，仅供内部参考

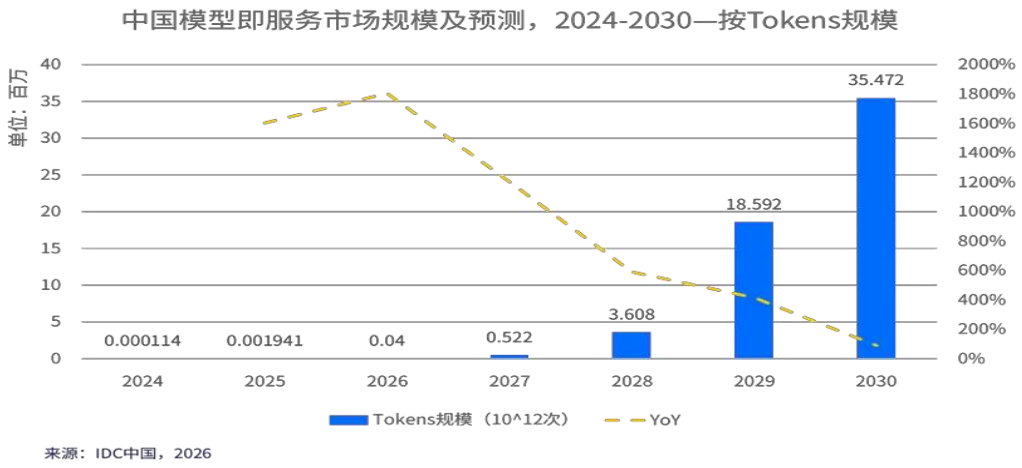
图 50：2024-2026 年中国我国日均 Token 调用量



资料来源：国家数据局，工信部，字节跳动，中原证券研究所

当前的 Token 使用仍处于初级阶段。根据 IDC 的数据，2025 年我国 MaaS 市场的 Token 消耗量为 1941 万亿次，约有 16 倍增长，2026 年有望继续增长到 40,000 万亿次，有望继续获得 20 倍的增长，但是 IDC 预测 2026-2030 年这一数值还将有近千倍的增幅。

图 51：2024-2030 年中国 MaaS 市场 Tokens 规模预测



资料来源：IDC，中原证券研究所

2. 河南计算机行业动态

2.1. 河南计算机行业数据跟踪

1-9 月，智能手机产品产量增长 13.6%，信息传输业投资增长 19.2%，计算机、通信和其他电子设备制造投资下滑 3.4%。

9 月，计算机、通信和其他电子设备制造业增长 20.8%。电子计算机整机 40.20 万台，同比增长 22.7%；智能手机 930.60 万台，同比下降 12.9%。

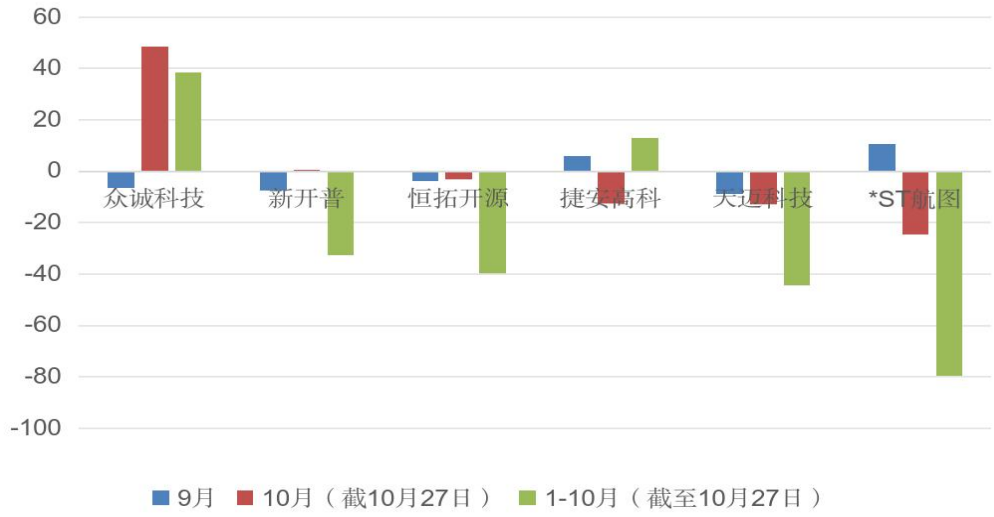
下载日志已记录，仅供内部参考

（数据来源：河南省统计局）

2.2. 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下跌。2026 年以来（截至 10 月 27 日）河南 6 家计算机公司中 2 家上涨、4 家下跌。众诚科技以 38.40%的涨幅领涨，其中 10 月上涨 48.58%；捷安高科 13.08%，10 月下跌了 12.64%。领跌个股为*ST 航图，下跌 79.64%，10 月下跌 24.66%;天迈科技下跌 44.23%，恒拓开源下跌 39.79%，新开普下跌 32.70%，全年表现较弱。

图 52：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

3. 计算机行业行情表现

3.1. 行情回顾：10 月整体表现较弱

10 月，计算机行业整体表现较弱。10 月（截至 27 日）中信计算机指数下跌了 6.43%，跑输上证指数 1.38 PCT，跑赢沪深 300 指数 0.64 PCT，跑赢创业板指数 12.30 PCT，在 30 个中信一级行业中排名第 17。

下载日志已记录，仅供内部参考

图 53：2026 年 10 月中信一级子行业涨跌幅（截至 2026 年 10 月 27 日）

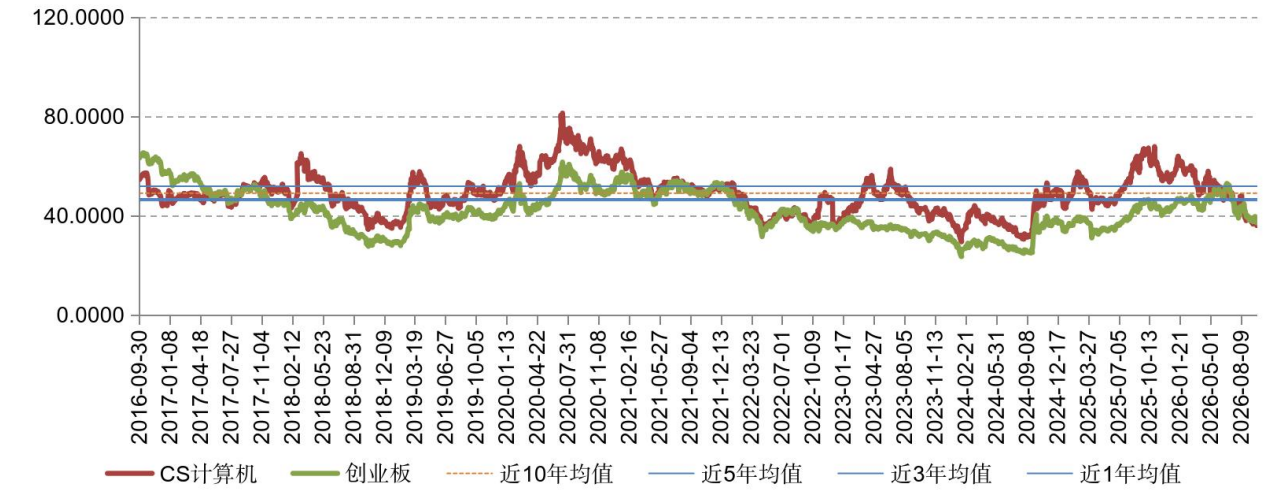


资料来源：Wind，中原证券研究所

3.2. 估值：行业的估值低于历史均值水平

行业的估值低于历史均值水平。根据 Wind 数据，2026 年 10 月 27 日中信计算机行业 TTM 整体法（剔除负值）估值为 32.72 倍，创业板估值 30.90 倍。行业近 1 年、3 年、5 年、10 年的平均估值分别为 50.07 倍、46.72 倍、46.00 倍、48.98 倍，当前估值低于历史均值水平，近 3 年最高点 67.87 倍和最低点 29.69 倍。

图 54：近 10 年中信计算机行业估值水平（截至 2026.10.27）



资料来源：Wind，中原证券研究所

3.3. 计算机行业 ETF 行情回顾

10 月行业 ETF 下跌为主，其中软件类 ETF 部分上涨，人工智能类 ETF 跌幅领先。

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

下载日志已记录，仅供内部参考

表 10：2026 年 9 月计算机行业 ETF 行情回顾（截至 2026.10.27）

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）
159899.SZ	招商中证全指软件 ETF	中证全指软件指数收益率	0.67	0.01	303
515230.SH	国泰中证全指软件 ETF	中证全指软件指数收益率	0.63	-0.03	2,336
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	0.18	-0.05	168
159852.SZ	嘉实中证软件服务 ETF	中证软件服务指数收益率	-0.11	-0.04	5,126
159586.SZ	南方中证全指计算机 ETF	中证全指计算机指数收益率	-1.13	0.01	127
159108.OF	博时工业软件 ETF	国证工业软件主题指数收益率	-1.28	0.05	98
159107.SZ	富国创业板软件 ETF	创业板软件指数收益率	-2.33	0.05	702
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	-5.03	0.00	64
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	-5.05	-0.02	36
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	-5.60	-0.02	97
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	-5.60	-0.02	119
516000.SH	华夏中证大数据产业 ETF	中证大数据产业指数收益率	-6.41	0.10	367
516700.SH	华宝中证大数据产业 ETF	中证大数据产业指数收益率	-6.43	0.07	110
515400.SH	富国中证大数据产业 ETF	中证大数据产业指数收益率	-6.44	0.06	1,012
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	-7.63	0.19	243
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	-7.64	0.19	77
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	-7.67	0.16	112
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	-7.67	0.16	438
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	-7.74	0.09	1,286
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	-8.75	0.07	1,015
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	-8.75	0.07	96
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	-8.75	0.07	853
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	-8.79	0.02	127
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	-8.80	0.02	143
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	-8.80	0.02	665
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	-8.80	0.02	2,222
589560.SH	汇添富上证科创板人工智能 ETF	上证科创板人工智能指数收益率	-8.80	0.01	136
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	-8.81	0.01	3,163
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	-8.82	-0.01	421
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	-10.75	0.35	272
560660.SH	新华中证云计算 50ETF	中证云计算 50 指数收益率	-10.83	0.18	42
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	-10.86	0.23	78
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	-10.96	0.13	195
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	-11.27	0.11	31
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	-13.00	0.43	48
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	-13.65	0.40	93
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	-13.77	0.26	220
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	-13.96	0.09	18

本报告版权属于中原证券股份有限公司 www.ccnew.com
请阅读最后一页各项声明

下载日志已记录，仅供内部参考

515070.SH	华夏中证人工智能 ETF	中证人工智能主题指数收益率	-13.98	0.05	1,809
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	-13.99	0.04	2,276
588780.SH	国联安科创芯片设计 ETF	上证科创板芯片设计主题指数	-17.02	0.14	2,344
159939.SZ	广发中证全指信息技术 ETF	中证全指信息技术指数	-17.51	0.33	386
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	-17.71	0.13	20
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	-18.90	0.15	18,239
159243.OF	招商创业板人工智能 ETF	创业板人工智能指数收益率	-19.25	0.32	453
159242.SZ	大成创业板人工智能 ETF	创业板人工智能指数收益率	-19.37	0.20	1,158
159279.SZ	华安创业板人工智能 ETF	创业板人工智能指数收益率	-19.45	0.12	646
159388.SZ	国泰创业板人工智能 ETF	创业板人工智能指数收益率	-19.47	0.10	1,547
159381.SZ	华夏创业板人工智能 ETF	创业板人工智能指数收益率	-19.49	0.09	4,265
159363.SZ	华宝创业板人工智能 ETF	创业板人工智能指数收益率	-19.51	0.07	13,393
159246.SZ	富国创业板人工智能 ETF	创业板人工智能指数收益率	-19.51	0.06	1,319
512330.SH	南方中证 500 信息技术 ETF	中证 500 信息技术指数	-19.68	0.32	98
159382.SZ	南方创业板人工智能 ETF	创业板人工智能指数收益率	-19.95	-0.37	2,289
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	-20.15	0.22	44

资料来源：Wind，中原证券研究所

3.4. 计算机行业可转债行情回顾

10 月计算机行业可转债以下跌为主，其中姚记转债上涨 3.45%，实现领涨，宏图转债下跌 25.30%，跌幅大幅领先。

表 11：2026 年 10 月计算机行业可转债行情回顾（截至 2026.10.28）

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
127104.SZ	姚记转债	3.4474	002605.SZ	姚记科技	53.2975	2030/01/25	A+
118052.SH	浩瀚转债	0.9095	688292.SH	浩瀚深度	74.1950	2031/03/13	A+
123104.SZ	卫宁转债	0.4074	300253.SZ	卫宁健康	22.0897	2027/03/16	AA
123261.SZ	普联转债	0.1283	300996.SZ	普联软件	67.9820	2031/12/05	A+
123138.SZ	丝路转债	-0.4181	300556.SZ	丝路视觉	81.5818	2028/03/02	A+
123196.SZ	正元转 02	-0.6283	300645.SZ	正元智慧	116.8031	2029/04/18	A+
123157.SZ	科蓝转债	-0.7177	300663.SZ	科蓝软件	61.1744	2028/08/30	AA-
123173.SZ	恒锋转债	-1.1332	300605.SZ	恒锋信息	39.8318	2028/12/30	A
118007.SH	山石转债	-2.0221	688030.SH	山石网科	36.3843	2028/03/22	A+
123131.SZ	奥飞转债	-6.1250	300738.SZ	奥飞数据	31.1857	2027/12/03	A+
118027.SH	宏图转债	-25.2957	688066.SH	*ST 航图	21.2642	2028/11/28	A

资料来源：Wind，中原证券研究所

4. 行业观点与投资建议

2026 年软件产业增速持续回落，行业利润增速下滑较为明显。2026 年 1-7 月软件业务收

下载日志已记录，仅供内部参考

入 8.98 万亿元，同比增长 9.2%，较 2025 年回落了 4.0 PCT。2026 年 1-7 月软件业务利润总额 10396 亿元，同比增长 1.3%，较 2025 年回落 6.0 PCT，低于收入增速 7.9 PCT。考虑到 AI 应用的快速增长，我们认为利润下滑或源于企业在 AI 方面的投入和开支加大所致。

重点关注子行业的主要数据和动态包括：

（1）国产化：2026 年前九个月份，集成电路进出口单价持续提升，同时出口单价连续 4 个月反超进口产品单价，国产芯片产品性能持续提升。2026 年上半年 AI 芯片国产化率进一步提升到了 64%，2026Q3 厂商业绩延续高增势头。DeepSeek 与昇腾的结合日趋紧密，正在通过开源方式，加快国产芯片替代进程。阿里新一代自研芯片 V900 将在 2027Q1 上市，较原先计划提前了 2 个季度，2027 年即将上市的昇腾 960 同样计划提前。由于逻辑折叠的加持，在相同性能下，麒麟 2026 在基准测试中，NPU 功耗下降 66%，GPU 下降 58%，性能核心 CPU 下降 41%。截至 2026 年 10 月 21 日，鸿蒙操作系统终端数已经突破 9500 万，华为预计 2026 年突破 1 亿用户规模。

（2）AI：月之暗面在 10 月发布了 Kimi K3.1 Pro，拥有 2.8 万亿参数，通过后训练加强和更多领域数据的强化学习，再次回归头部模型行列，其他国产模型主要位居第二梯队。以美国最强模型 Anthropic 的 Opus 5.5 的调用价格来看，其输入输出价格都约为同等级模型 Kimi-K3.1 pro 的 2 倍，在国产模型冲击和用户成本敏感度提升的背景下，近期美国闭源模型降价趋势明显。根据 OpenRouter 的模型调用数据，DeepSeek 份额 26%，连续 12 周位列第一大模型厂商。同时中国模型厂商份额达到 46%，暂领先较美国模型厂商（份额 44%）。10 月微软向约 60000 名员工（占比超 25%）开放自愿离职方案，以压缩成本、调配资源加码 AI，科技大厂裁员再呈加速迹象。9 月，美国企业模型采用率达到 55.9%，Anthropic 以 43.8%的份额再次超过 OpenAI（40.6%）。

给予行业强于大市的投资评级。当前 AI 应用正在加速渗透至编程、影视游戏、数学研究、网络安全等领域，物理 AI 的突破进程也在加速。Agent 自主化的执行意味着模型被调用的频次也在大幅增长，让算力供应持续不足，算卡租赁价格继续呈现走高趋势。相对而言，受到缺芯因素的影响，国内算力紧张程度更高。随着 2027 年国产卡产能的释放和新卡的推出，国内算力需求有望得到更好的满足。我们继续对国内 AI 算力产业发展持乐观预期，但是考虑到外盘科技股整体位居高位，后续或受到 Anthropic 上市、2027 年资本开资担忧情绪等因素影响，A 股 AI 算力板块如要演绎独立行情或仍面临较大的信心挑战。**建议持续重点关注 AI 产业链上游及中游环节，包括半导体设备、芯片、存储、光通信、服务器、AIDC、云计算等环节。个股方面，推荐服务器厂商中科曙光（603019）、紫光股份（000938）、浪潮信息（000977），AIDC 厂商润泽科技（300442）。同时建议积极关注正在积极开展 IPO 的超聚变、平头哥、智谱、DeepSeek、月之暗面、长存控股。**

表 12：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

下载日志已记录，仅供内部参考

	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	2026.2.11	港交所上市
	燧原科技	2026.9.11	科创板上市
	昆仑芯	2026.1.1	港交所递交上市申请
		2026.5.7	科创板辅导备案
	平头哥		计划重组，未来独立上市
内存	长鑫科技	2026.7.27	上市
	长存控股	2026.8.21	提交上市申报稿
服务器	超聚变	2026.9.21	第三次审核问询回复
大模型	智谱	2026.1.8	港交所上市
		2026.6.17	科创板 IPO 辅导备案验收
	MINIMAX	2026.1.9	港交所上市
		2026.5.29	签署 A 股辅导备案协议
	DeepSeek	2026.8.24	已启动 IPO 筹备，计划年内递交科创板上市申请，预计 2027 年正式挂牌。
	月之暗面	2026.9.3	已经保密形式向港交所递交 IPO 文件

资料来源：澎湃，Wind，中原证券研究所

5. 风险提示

国内外科技领域巨头 IPO 带来的市场冲击；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。

下载日志已记录，仅供内部参考

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。