仅供内部参考，请勿外传！

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

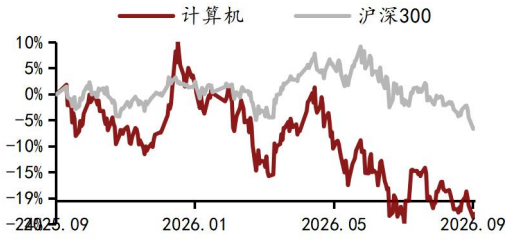
国产芯片发布计划大幅提前，算力投入呈现加速趋势

——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 09 月 30 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

《计算机行业半年度策略:模型、应用齐发力，国产算力迎价值重估》 2026-06-30

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：**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%。在 9 月第二批 AI 训练推理芯片的安全可靠评测结果发布后，华为、百度、阿里、壁仞、海光、天数、沐曦、摩尔线程 8 家芯片通过了测评，其中华为昇腾 950 成为首个 II 级安全可靠 AI 芯片。鸿蒙 7 在 9 月底开启正式版升级，鸿蒙终端数量突破 8500 万，2026Q2 在国内手机操作系统市场份额 24%，较 Q1 上升 5 PCT。

- **算力：**排名前 10 的智算企业及机构智算总规模达到 1266EFLOPS，占全国总规模的 57.9%，呈现高度集中化特征。根据 SemiAnalysis 数据，到 2026 年底，国内头部厂商的算力约为字节 5.3GW、阿里 3.9GW、腾讯 2.4GW、百度约 1.1GW、华为约 1.0GW。阿里宣布了 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。

- **AI：**由于当前 Anthropic 和 OpenAI 刚进行了一轮更新，国内模型由于发布节奏问题，与头部差距有短暂拉大。Anthropic 的 Opus 5.5、OpenAI 的 GPT-6 Astra、Meta 的 Muse Spark 1.3、小米的 MiMo-V2.6-Pro、阿里的 Qwen3.8-Max 都有了不错的表现。头部模型降价趋势明显，9 月发布的 Opus 5.5 又在 Opus 5 基础上再降价 20%，OpenAI 对 9 月新推出的 GPT-6 Sol 和 Luna 再次降价约 50%。最新的 OpenRouter 的模型调用周数据显示，中国模型厂商份额达到 45%，连续两周领先较美国模型厂商（份额 41%）。

- **给予行业强于大市的投资评级。**从当前国内 AI 产业发展来看，资本开支加大的趋势明显。国内虽然算力短缺，但是模型在综合能力和性价比的比拼当中成为了更多用户的选择，调用份额持续提升。以 WorkBuddy 为代表的 AI 办公产品，正在加速 AI 在 B 端的普及进程。因而在这一趋势下，包括字节、阿里在内的头部厂商的投资步伐加快明显，新一代国产芯片上市时间节点也大幅前置。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）、东阳光（600673）。同时建议积极关注正在积极开展 IPO 的超聚变、平头哥、智谱、**

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

仅供内部参考，请勿外传！

DeepSeek、Kimi、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。

仅供内部参考，请勿外传！

内容目录

1. 行业数据 7
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 7
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 8
1.3. 国产化：首个全国产十万卡 AI 超集群落地，昇腾 950DT 提前至 8 月上线 11
1.3.1. 国产芯片 11
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 21
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 8500 万 23
1.4. 算力：Q2 国内资本开支投入加速，全球硬件短缺趋势延续 25
1.4.1. 服务器：服务器企业中报情况普遍向好，原材料短缺持续影响供需 25
1.4.2. 数据中心：2026 年我国智算建设集中化趋势明显 29
1.4.3. 云计算：受到 AI 需求拉动，整体增速仍在加快 33
1.4.4. 资本开支：Q2 国内资本开资明显提速 35
1.4.5. 算力租赁价格 40
1.5. AI：头部模型继续呈现降价趋势 41
1.5.1. 中美大模型发展概况 41
1.5.2. 中美模型厂商收入 43
1.5.3. 厂商模型调用及国产模型替代情况 44
1.5.4. 模型与应用：大模型备案高度集中在发达区域 45
1.5.5. 科技企业裁员：2026 年前 8 个月美国因为 AI 取代占总裁员的比例达到 22% 46
1.5.6. 企业端 AI 应用情况 48
1.5.7. 国内 Token 使用情况 49
2. 河南计算机行业动态 50
2.1. 河南计算机行业数据跟踪 50
2.2. 河南上市公司行情回顾 50
3. 计算机行业行情表现 51
3.1. 行情回顾：9 月整体表现较弱 51
3.2. 估值：行业的估值低于历史均值水平 51
3.3. 计算机行业 ETF 行情回顾 52
3.4. 计算机行业可转债行情回顾 53
4. 行业观点与投资建议 54
5. 风险提示 55

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） 7
图 2 ：2019-2026 年我国软件业务利润总额及增速（月度累计值） 7
图 3 ：2019-2026 年我国软件业务出口 8
图 4 ：2019-2026 集成电路设计服务收入及增速（亿元） 8
图 5 ：2020-2026 云计算+大数据服务收入及增速（亿元） 8
图 6 ：2022-2026 基础软件收入及增速（亿元） 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

仅供内部参考，请勿外传！

图 7 ：2019-2026 工业软件收入及增速（亿元） 9
图 8 ：2019-2026 电子商务平台技术服务收入及增速（亿元） 9
图 9 ：2019-2026 信息安全收入及增速（亿元） 10
图 10 ：2019-2026 嵌入式系统软件收入及增速（亿元） 10
图 11 ：2026 与 2025 年 1-7 月软件子行业增速对比 10
图 12 ：2026 年 1-7 月我国软件业务收入结构 10
图 13 ：2026 年 1-6 月信息技术服务收入结构 10
图 14 ：2026 年 1-6 月软件产品收入结构 10
图 15 ：2022-2026 我国进出口集成电路单价（元/个） 11
图 16 ：2022-2026 我国出口/进口集成电路比值 11
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） 12
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） 12
图 20 ：2023-2025 我国 AI 芯片出货情况及国产化比率 14
图 21 ：2025 年我国加速芯片市场份额 14
图 22 ：2017-2026H1 国内主要上市 AI 芯片企业收入及增速对比 15
图 23 ：DeepSeek V4 发布公告关于算力的表述 18
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 22
图 26 ：麒麟 2026 和麒麟 9030 Pro 在等效性能下的功率对比 22
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 ：2022Q1-2026Q2 中国手机市场厂商份额 25
图 32 ：2021Q1-2026Q2 中国平板电脑市场厂商份额 25
图 33 ：2017-2025 全球服务器市场规模（亿美元） 25
图 34 ：2023Q2-2026Q2 全球服务器市场规模（亿美元） 25
图 35 ：2026Q2 全球不同类型服务器（亿美元） 26
图 36 ：2026Q2 全球不同 CPU 类型服务器（亿美元） 26
图 37 ：2026Q2 全球不同区域服务器收入增速 26
图 38 ：2026Q2 全球服务器区域分布 26
图 39 ：2026Q2 全球服务器厂商份额 27
图 40 ：2026Q2 全球服务器市场头部厂商收入（亿美元） 27
图 41 ：2018-2025 中国服务器市场规模（亿美元） 27
图 42 ：中国加速服务器增长预测（亿美元） 27
图 43 ：2025 中国服务器市场规模（亿美元） 28
图 44 ：2025 年中国 X86 服务器厂商市场份额 28
图 45 ：2021Q3-2026Q2 服务器厂商存货规模及增速（百万元） 29
图 46 ：2021Q3-2026Q2 服务器厂商预付账款规模及增速（百万元） 29
图 47 ：2026 年底全球分区域数据中心容量分布 29
图 48 ：2017-2026 我国智能算力规模（EFLOPs） 30
图 49 ：2026 年 6 月底我国智算区域 30
图 50 ：我国在用标准机架分布 30
图 51 ：我国在用智算规模（FP16）分布 31
图 52 ：各省上架率情况 31

仅供内部参考，请勿外传！

图 53 ：2019Q1-2026Q1 云巨头的相关业务收入及增速 33
图 54 ：甲骨文云基础设施收入及增速（亿美元） 34
图 55 ：甲骨文云基础设施业务年度收入预测（亿美元） 34
图 56 ：2019Q4-2026Q1 阿里云收入及其增速 34
图 57 ：2017Q1-2026Q2 金山云收入及其增速 34
图 58 ：2024-2025 中国 AI 云市场规模（亿元） 35
图 59 ：2025 中国 AI 云市场细分市场份额 35
图 60 ：2025H1 中国 AI 云市场格局 35
图 61 ：2025 中国 AI 云市场格局 35
图 62 ：4 大科技厂商资本投入（亿美元） 36
图 63 ：4 大科技厂商资本投入增速 36
图 64 ：2020-2026 年阿里单季度资本开支及增速 38
图 65 ：2020-2026 年腾讯单季度资本开支及增速（亿元） 38
图 66 ：中国超大规模数据中心容量（2017-2028） 39
图 67 ：中国超大规模数据中心容量（2017-2028） 40
图 68 ：2026 年中国超大规模数据中心容量 40
图 69 ：GPU 租赁价格趋势（1 年期合同） 41
图 70 ：不同地区的大模型智能水平对比（截至 2026.9.28，红色为中国模型） 42
图 71 ：大模型智能水平及开源情况（截至 2026.9.28） 43
图 72 ：海外模型及 AI 应用的年化收入（亿美元） 44
图 73 ：国内模型及 AI 应用的年化收入（亿美元） 44
图 74 ：OpenRouter 调用的模型份额分布（百万 tokens） 45
图 75 ：OpenRouter 调用的模型份额分布 45
图 76 ：全国大模型备案与登记数量 46
图 77 ：各地大模型备案与登记数量 46
图 78 ：大模型备案与登记数量的区域分布 46
图 79 ：2025.7-2026.8 美国因 AI 被裁的人数及占裁员总数的比例（当月值，人） 48
图 80 ：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） 48
图 81 ：Ramp AI 指数（美国企业模型采用率） 48
图 82 ：美国企业每个员工每月的 AI 支出（美元） 48
图 83 ：美国各行业员工每月的 AI 支出中位数 49
图 84 ：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 49
图 85 ：2024-2026 年中国我国日均 Token 调用量 49
图 86 ：2024-2030 年中国 MaaS 市场 Tokens 规模预测 50
图 87 ：河南上市公司近期股价涨跌幅表现（%） 51
图 88 ：2026 年 9 月中信一级子行业涨跌幅（截至 2026 年 9 月 28 日） 51
图 89 ：近 10 年中信计算机行业估值水平（截至 2026.9.29） 52

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 16
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） 17
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 19
表 7 ：鸿蒙纯血系统发展动态 23
表 8 ：2026H1 企业服务器相关业务对比（亿元） 28

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

仅供内部参考，请勿外传！

表 9 ：2026 年企业数据中心相关业务对比（亿元） 32
表 10 ：海外科技厂商算力相关投入 36
表 11 ：海外科技厂商算力相关投入 39
表 12 ：模型调用价格（每百万 Token，Artificial Analysis2026 年 9 月 24 日智能指数）4
表 13 ：海外企业切换到国产模型的案例 44
表 14 ：海外主要科技厂商近期裁员情况 47
表 15 ：2026 年 9 月计算机行业 ETF 行情回顾（截至 2026.9.28） 52
表 16 ：2026 年 9 月计算机行业可转债行情回顾（截至 2026.9.28） 54
表 17 ：行业重点企业的 IPO 推进情况 55

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

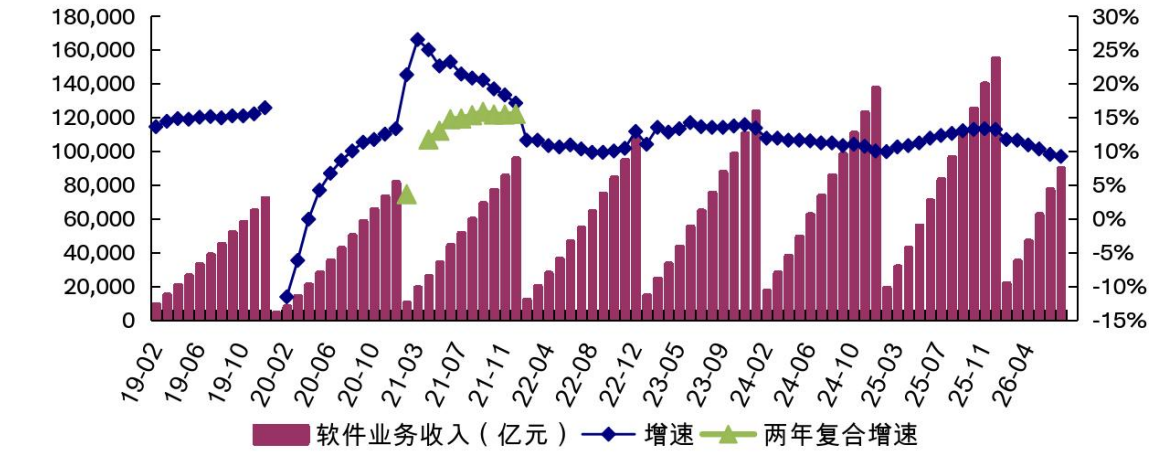
仅供内部参考，请勿外传！

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。

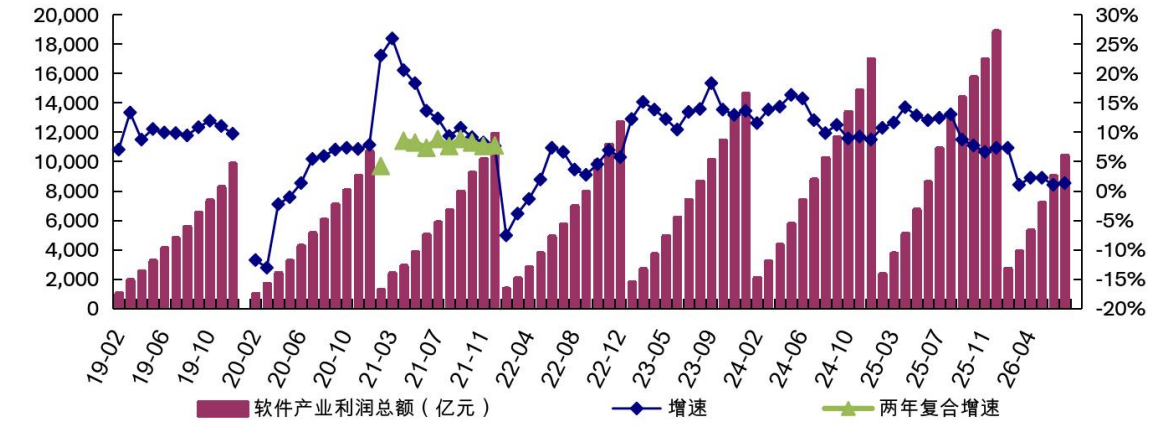
图 1：2019-2026 年我国软件业务收入及增速（月度累计值）



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）

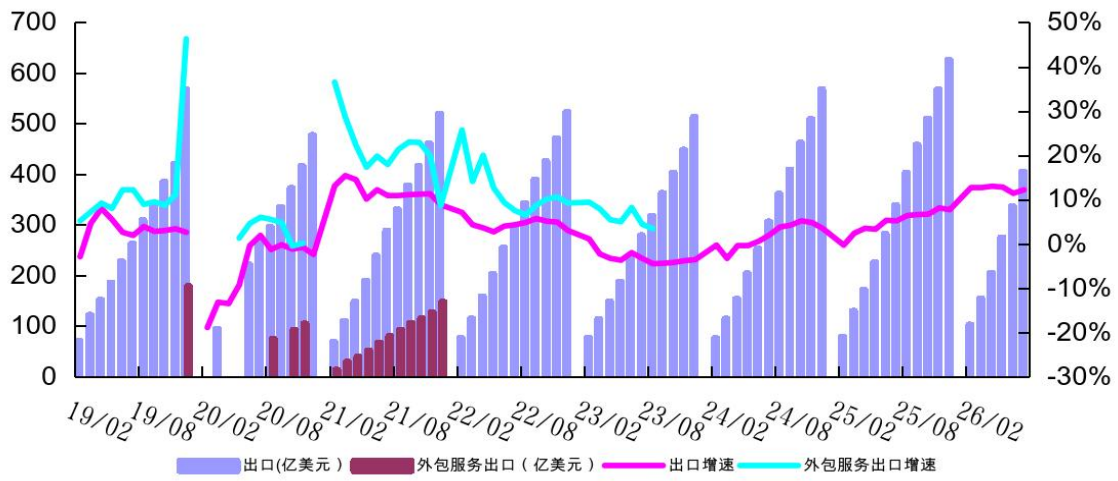


资料来源：工信部，中原证券研究所

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%。

仅供内部参考，请勿外传！

图 3：2019-2026 年我国软件业务出口数据



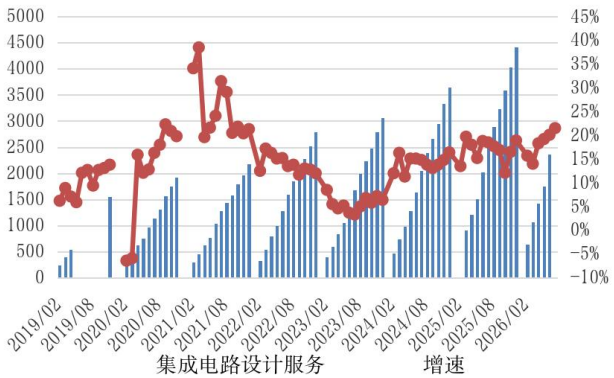
资料来源：工信部，中原证券研究所

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。

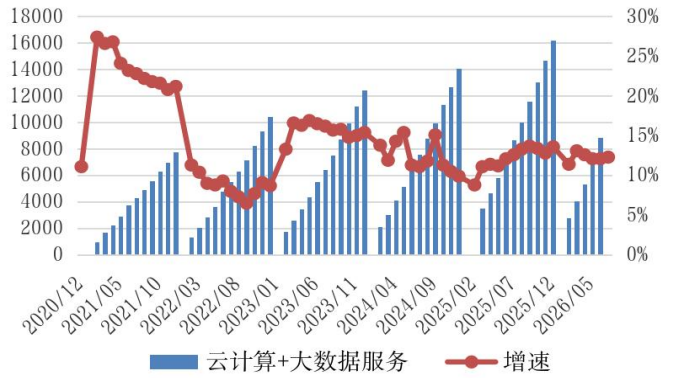
云计算+大数据服务：2026 年 1-7 月同比增长 12.3%，较 2025 年下降了 1.3 PCT，高于软件行业整体增速 3.1 PCT，2026 年预计仍将受益于 AI 应用快速发展趋势。

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

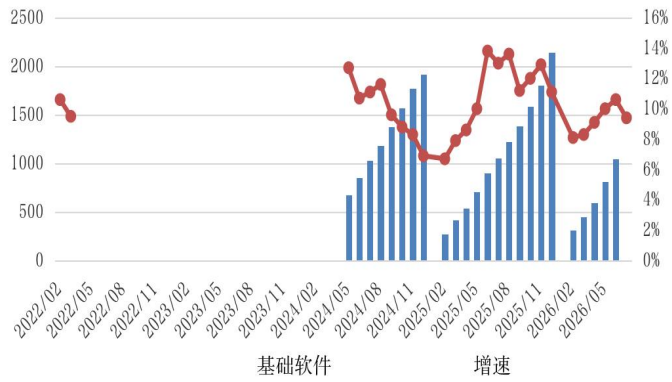
基础软件：2026 年 1-7 月同比增长 9.4%，较 2025 年下降了 1.7 PCT，高于软件行业整体增速 0.2 PCT。

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

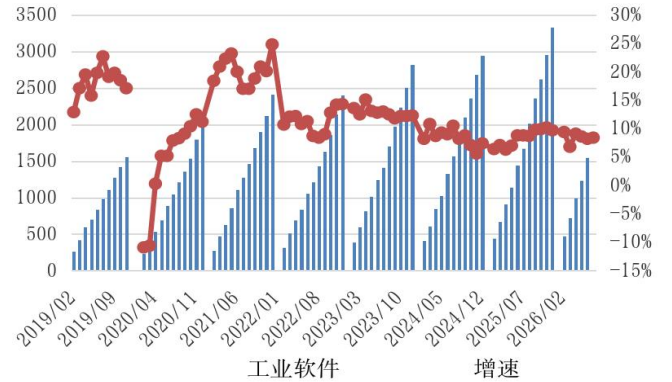
仅供内部参考，请勿外传！

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

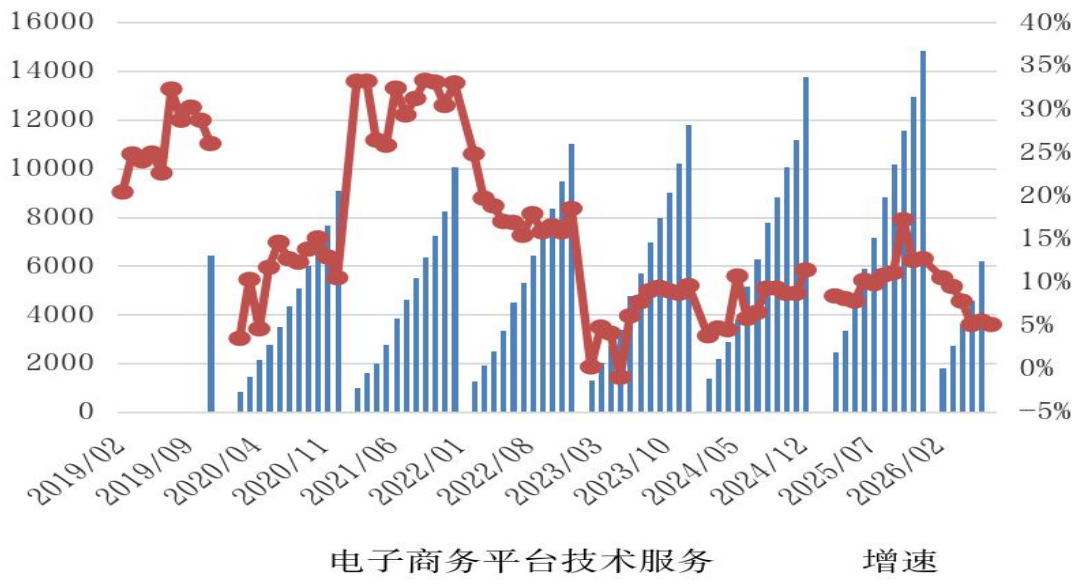
图 7：2019-2026 工业软件收入及增速（亿元）



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）



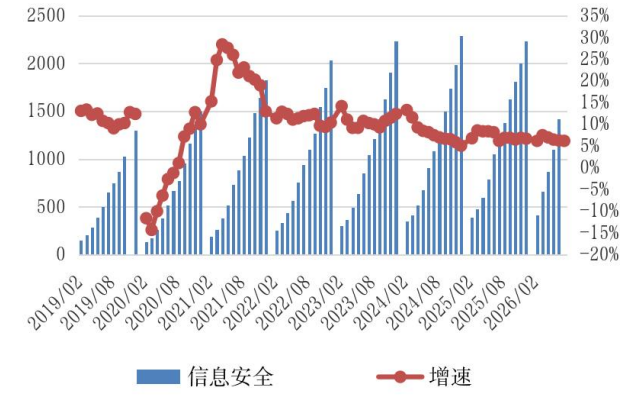
资料来源：工信部，中原证券研究所

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。

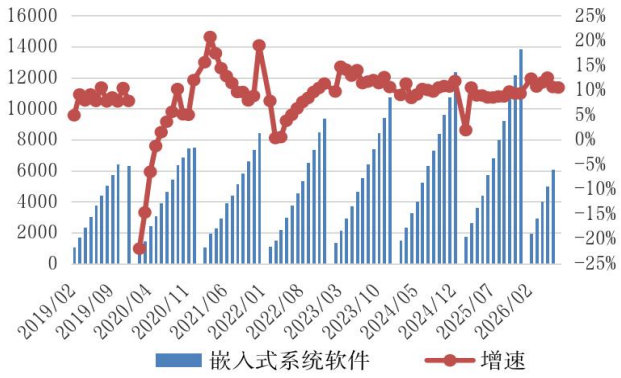
仅供内部参考，请勿外传！

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

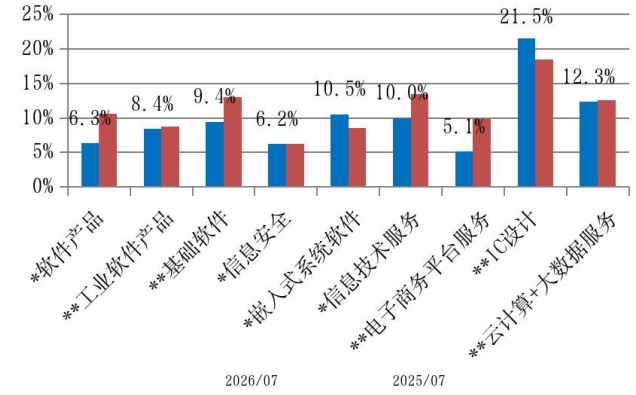
图 10：2019-2026 嵌入式系统软件收入及增速（亿元）



资料来源：工信部，中原证券研究所

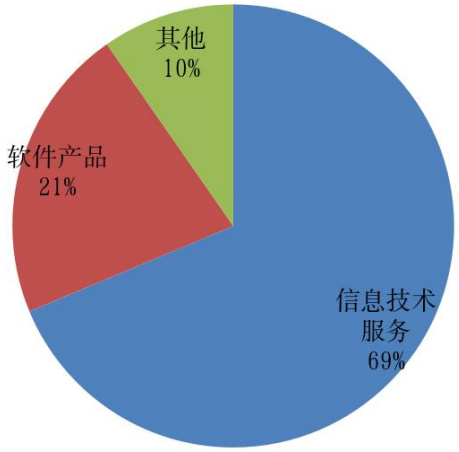
软件业务中，信息技术服务收入占比持续提升。2026 年 1-7 月信息技术服务收入增速 10.4%，高于软件业务整体增速 0.9 PCT，占比软件业务整体收入比重提升至 68.5%。

图 11：2026 与 2025 年 1-7 月软件子行业增速对比



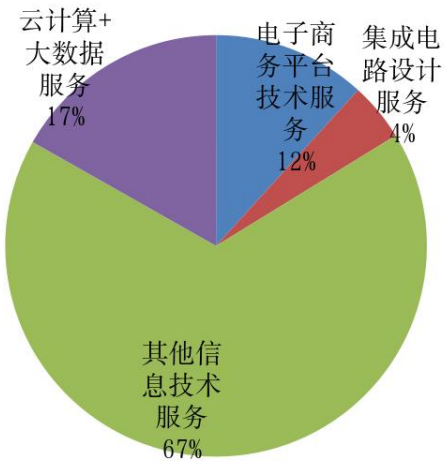
资料来源：工信部，中原证券研究所（*为一级子行业，**为二级子行业）

图 12：2026 年 1-7 月我国软件业务收入结构



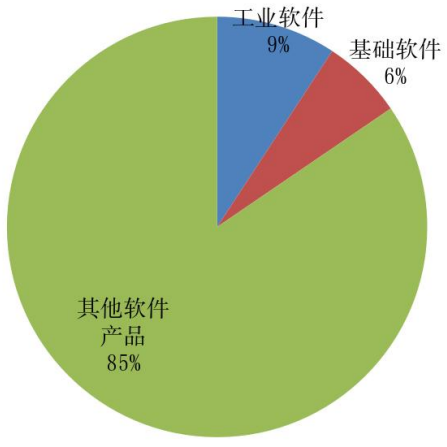
资料来源：工信部，中原证券研究所

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

仅供内部参考，请勿外传！

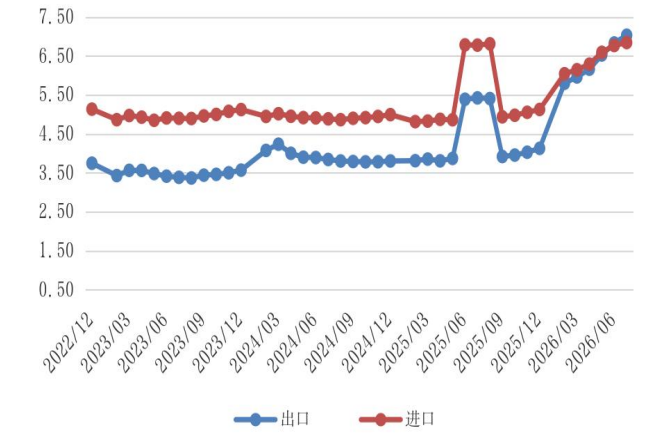
1.3. 国产化：首个全国产十万卡 AI 超集群落地，昇腾 950DT 提前至 8 月上线

1.3.1. 国产芯片

1.3.1.1. 国内芯片产品性能持续提升，上半年芯片自给率达到 24%

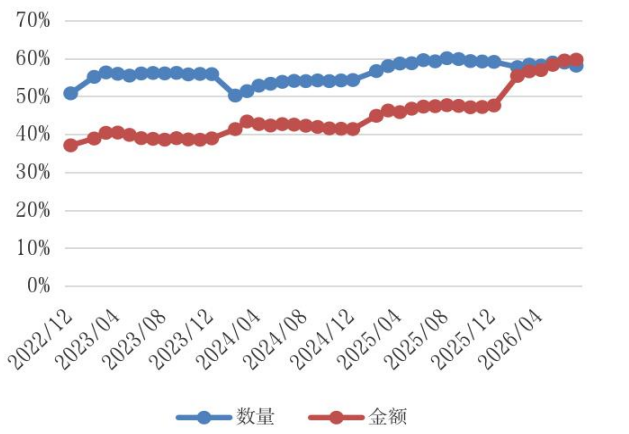
2026 年前七个月份，集成电路进出口单价持续提升，同时出口单价连续 2 个月反超进口产品单价。进口数量和金额分别同比增长 8%和 52%，出口数量和金额分别同比增长 6.2%和 91.6%。其中出口集成电路平均单价与进口电路平均单价价差从上一年同期的-1.40 元/个增长到 0.19 元/个，出口产品单价连续 2 个月实现反超。同时将我国出口和进口集成电路对比来看，出口集成电路数量占比进口数量上总体稳定，但是金额占比方面逐年提升，已经由 2022 年的 37.1%提升到了 2026 年 1-7 月的 59.6%，**这表明我国的集成电路产品能力在持续提升。**

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

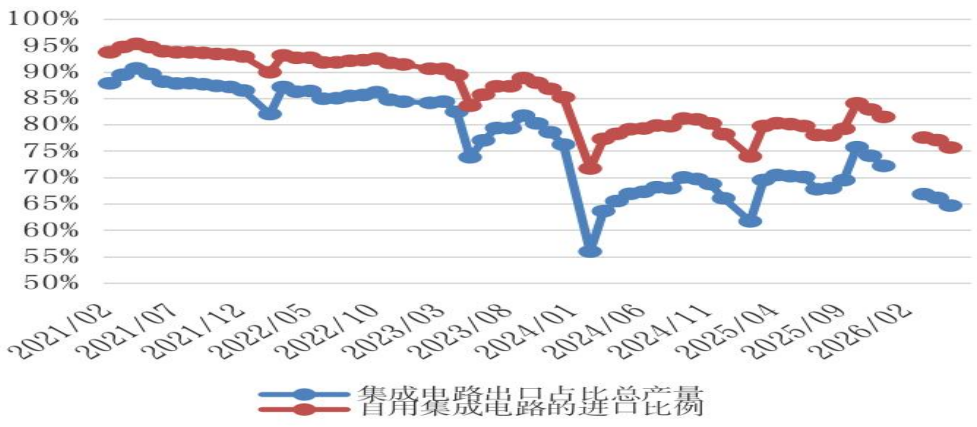
图 16：2022-2026 我国出口/进口集成电路比值



资料来源：海关总署，中原证券研究所

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。

图 17：2021-2026 年我国集成电路对进口依赖度和出口占比



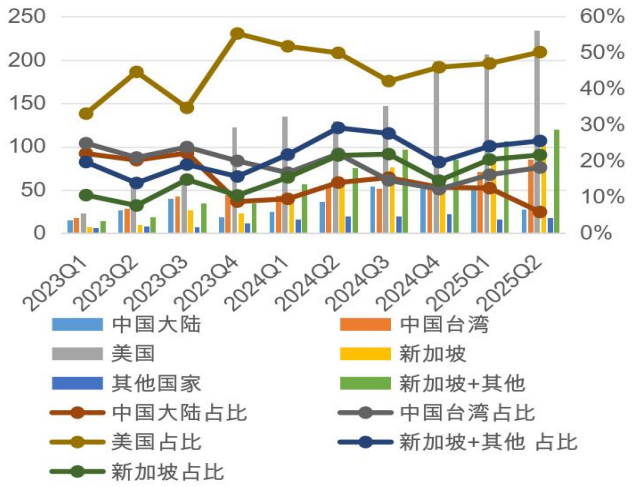
资料来源：工信部，海关总署，中原证券研究所

仅供内部参考，请勿外传！

1.3.1.2. 芯片国产化仍受益于政策发展趋势

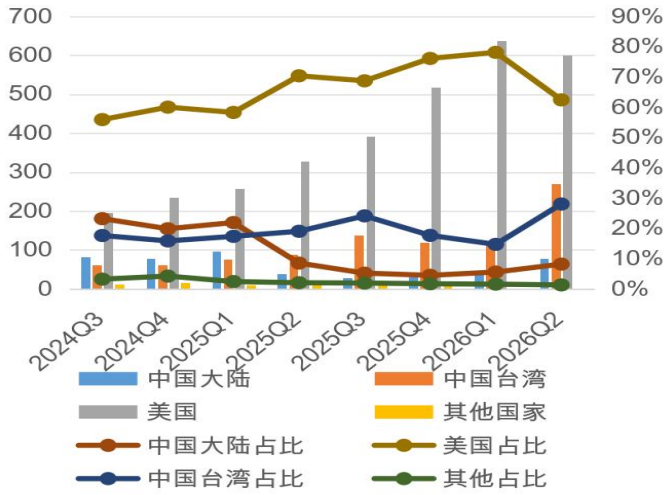
Q2 英伟达来自中国大陆企业的收入增长明显，但是仍然没有超过前期高位水平。2023 年 Q1 到 Q3，英伟达来自中国大陆收入占比分别为 22%、20%、22%，Q3 金额达到了 40.3 亿美元，Q4 在 A800 和 H800 禁令影响下该比例降低到了 9%。而后，中国大陆占比约为 10%-15%之间波动，在 H20 禁令的影响下，该比例 Q3 降低到了 6%，金额也从上个季度的 55.2 亿美元降低到了 27.7 亿美元。2025Q3 开始，英伟达对分区域的收入划分口径进行了修改，从客户账单所在地改为了总部所在地。根据新旧口径差额的估算，我们看到中国大陆企业不断加大在其他地区对于英伟达的采购，从 2024 年的 20.2 亿美元增长到 2024 年的 79.4 亿美元，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）

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。虽然近期有关于 H200 小批量进口的报道（根据 IDC 数据，2025 年我国 AI 芯片出货量超过 400 万片），但是我们认为对于总体国家支持国产芯片发展的方向影响有限。

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

仅供内部参考，请勿外传！

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。
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。

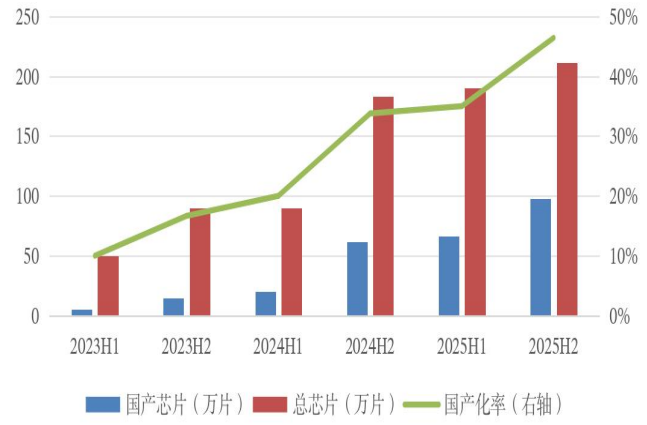
资料来源：快科技，和讯网，新京报网，观察者网，罗斯卫星通讯社，C114，闻华财讯，新浪财经，爱集微，云头条，中原证券研究所

仅供内部参考，请勿外传！

1.3.1.3. 2025 年下半年 AI 芯片国产化率进一步提升到了 46%，2026Q2 厂商延续高增势头

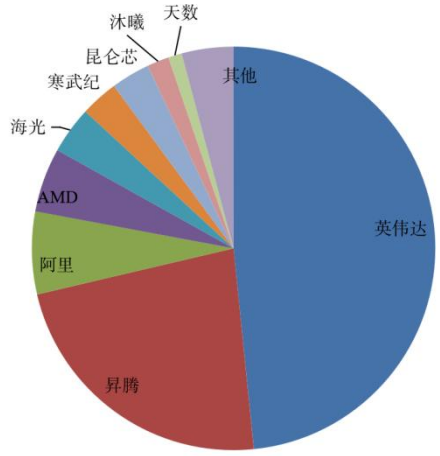
2026 年 AI 芯片国产替代趋势有望继续加快。根据 IDC 数据，2025 年下半年，我国 AI 芯片国产化比率从 2025 上半年的 35%提升到了 46%，提升趋势明显。考虑到英伟达芯片对国内实质性断供，同时国产芯片在性能和产能方面的提升，我们预计 2026 年国产 AI 芯片的占比将获得较大比例的提升。

图 20：2023-2025 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 21：2025 年我国加速芯片市场份额

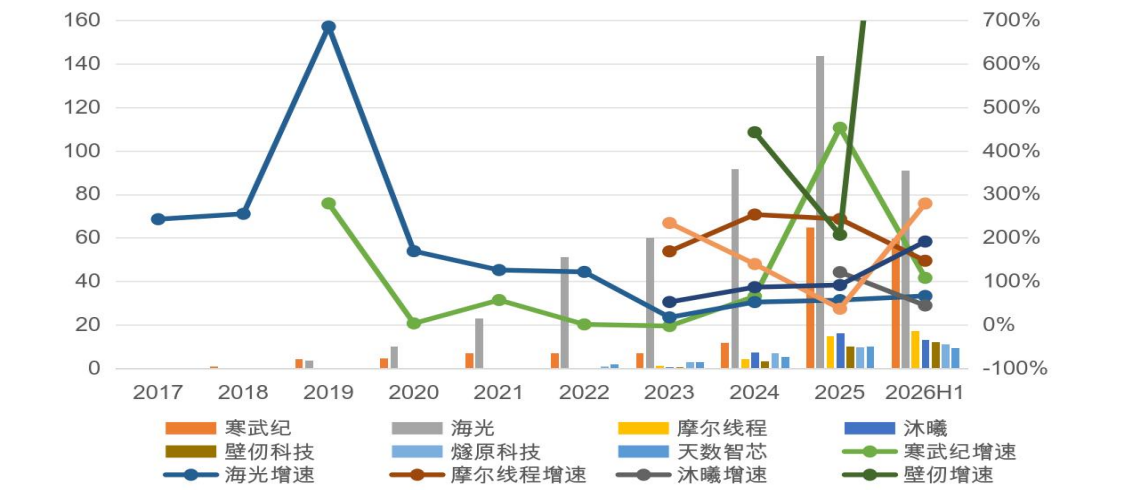


资料来源：IDC，中原证券研究所

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）。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。根据 2026H1 财报，海光信息收入增长 67%，寒武纪增速 108%，摩尔线程收入增长 147%，都在 Q2 取得了不错的业绩增速；壁仞科技增长 1998%，燧原科技增长 279%，壁仞科技同比增长 45%，都已经超过 2025 年全年收入水平；天数智芯同比增长 192%，也已经接近 2025 年全年收入水平。

仅供内部参考，请勿外传！

图 22：2017-2026H1 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）

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。

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。

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

仅供内部参考，请勿外传！

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

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

仅供内部参考，请勿外传！

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
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。

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

仅供内部参考，请勿外传！

图 23：DeepSeek V4 发布公告关于算力的表述



API 访问模型名	输入（缓存命中）	输入（缓存未命中）	输出	上下文长度
deepseek-v4-pro	1 元	12 元	24 元	1M
deepseek-v4-flash	0.2 元	1 元	2 元	

*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。

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			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

仅供内部参考，请勿外传！

						求
	深算四号	2023	2026			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
		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

仅供内部参考，请勿外传！

		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
			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

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

仅供内部参考，请勿外传！

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。

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

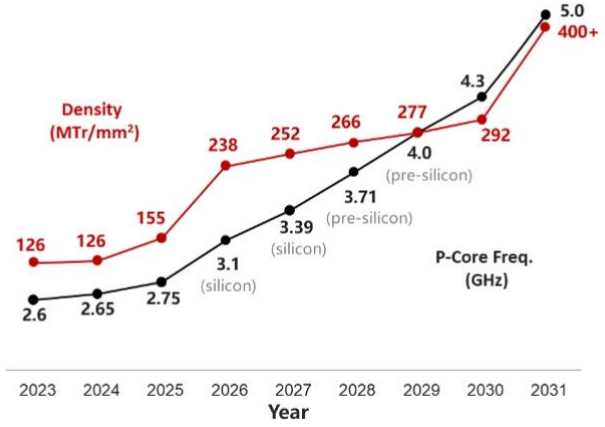
未来 10 年，逻辑折叠有望从局部关键路径扩展到全规模多层折叠，且在成本方面具有经济效益。预计到 2031 年，基于韬(τ)定律的高端芯片晶体管密度将达到 1.4 纳米制程的同等水平。值得关注的是，华为的 τ 缩放路线图还是在成本降低 30%的前提下实现的。

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%，

仅供内部参考，请勿外传！

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

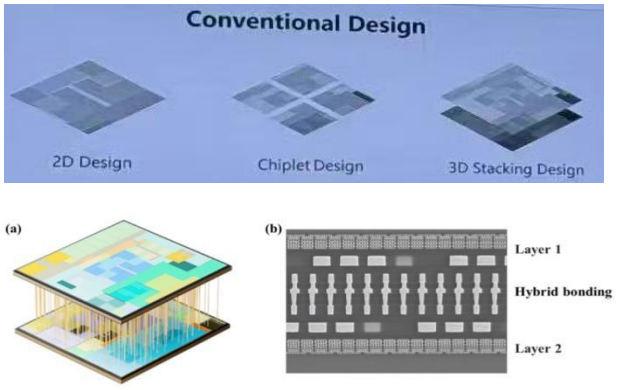
图 26：麒麟 2026 和麒麟 9030 Pro 在等效性能下的功率对比

	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

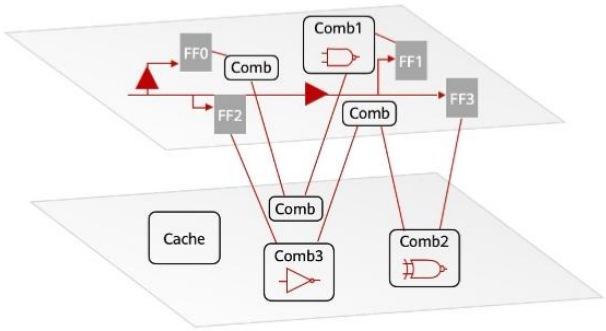
传统的 3D 堆叠路线是将独立的芯片进行物理堆叠；而在逻辑折叠架构中，上下有源层不再是独立的单芯片，而是不可分割的单个芯片的组成部分。上下逻辑层通过整体设计，利用极高密度的面对面混合键合技术，实现了接近芯片内部互联级别的垂直连接，满足了高效灵活的逻辑设计需求，进而大幅提升了有效晶体管密度。同时，华为的实践表明，相对于 2D 解决方案，逻辑折叠可以实现更好的成本效益。

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 堆叠为长期发展趋势，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

仅供内部参考，请勿外传！

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 8500 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，可以完成多步骤任务执行和 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。

截至 2026 年 9 月 7 日，鸿蒙操作系统终端数已经突破 8500 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

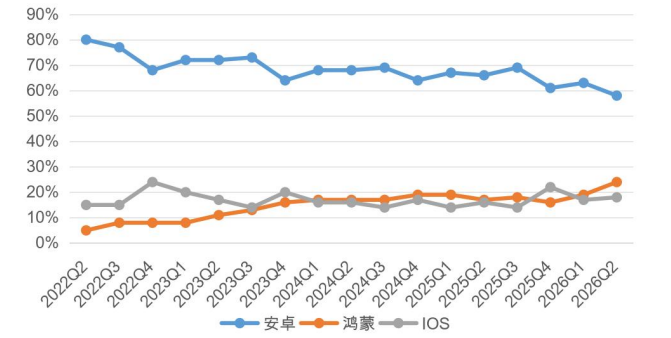
仅供内部参考，请勿外传！

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

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

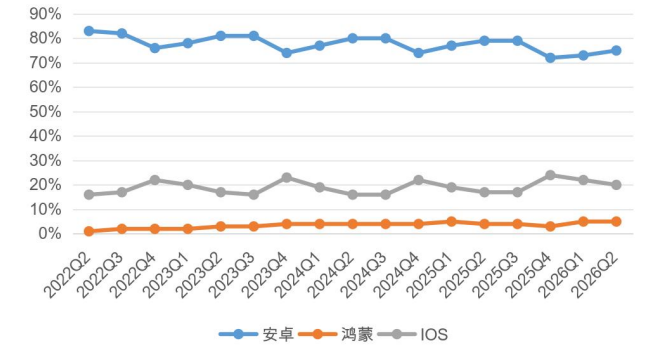
2026Q2 鸿蒙在手机操作系统市场份额回升明显，在中国市场重回第二。根据 Counterpoint 的数据，2026Q2 国内手机操作系统市场上，鸿蒙份额 24%，较上个季度上升 5 PCT，较 IOS 的优势进一步拉大，稳坐国内第二位置。全球市场来看，鸿蒙同样表现突出，份额继续保持 5%，继续位居全球第三，与 IOS 的差距缩小了 2 PCT。

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。

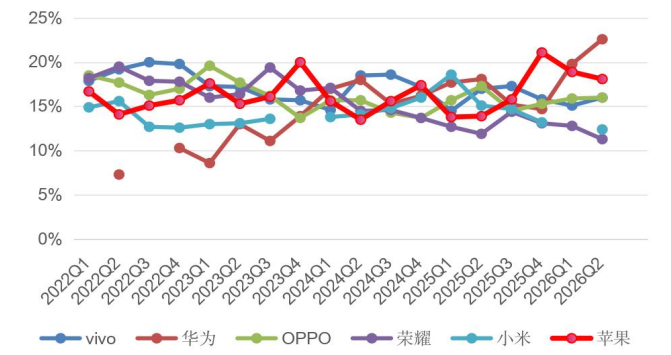
（1）手机：根据 IDC 数据，2026Q2 华为以 22.6%的出货量份额位居中国手机市场第 1，在行业下滑 4.3%的背景下，同比增长 19.4%，取得了优异的表现。

（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%，表现不及行业水平。

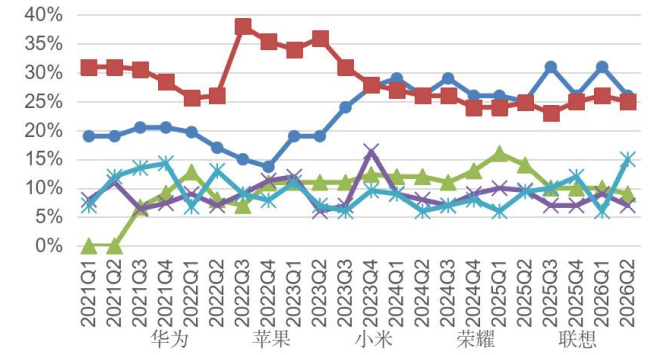
仅供内部参考，请勿外传！

图 31：2022Q1-2026Q2 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、鸿蒙智行和乾崑智驾、重构 AI 数据中心、液冷超快充。**

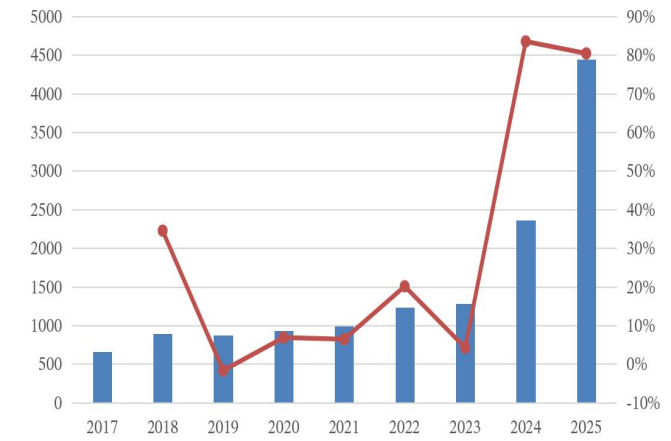
1.4. 算力：Q2 国内资本开支投入加速，全球硬件短缺趋势延续

1.4.1. 服务器：服务器企业中报情况普遍向好，原材料短缺持续影响供需

1.4.1.1. 全球市场：2026Q2 非 GPU 服务器和非 X86 服务器加速增长中

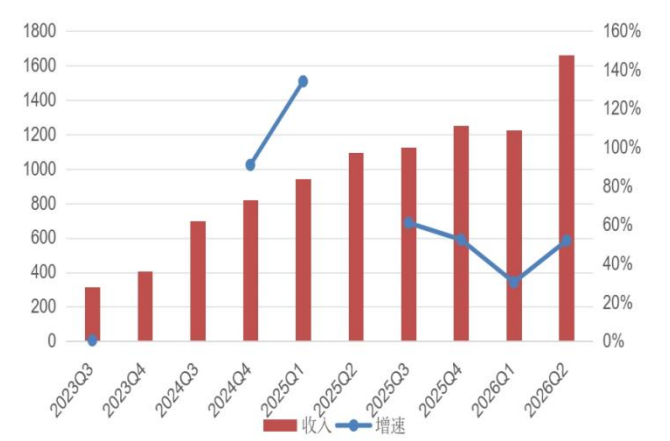
Q2 全球服务器市场重回快速增长通道。根据 IDC 数据，2025 年全球服务器市场规模达到 4441 亿美元，同比增长 80.4%。2026Q2 全球服务器市场规模达到 1663 亿美元，同比增长 52%，环比提升了 22 PCT。

图 33：2017-2025 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 34：2023Q2-2026Q2 全球服务器市场规模（亿美元）



原：IDC，中原证券研究所

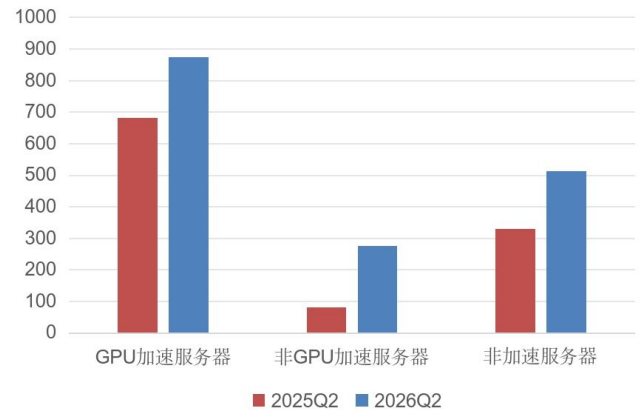
从服务器结构来看，增长动力主要来自非 GPU 加速服务器和非 X86 服务器。

仅供内部参考，请勿外传！

（1）非加速服务器和加速服务器分别增长 55.8%和 50.4%，基本相当。加速服务器中，非 GPU 加速服务器增速达到了 237.6%，呈现加速增长趋势；而 GPU 服务器增长 28.1%，占比 76%，仍然是加速服务器的主力品种。

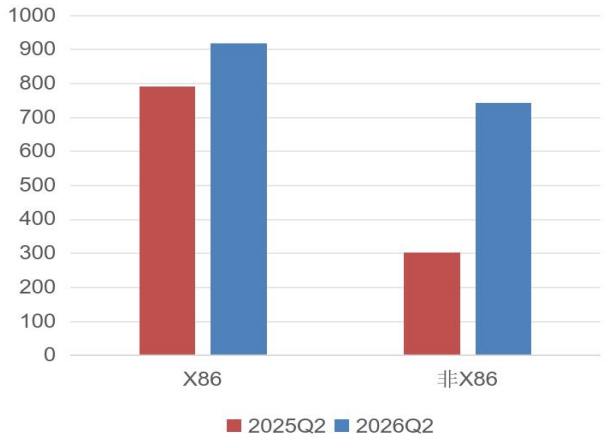
（2）Q2 X86 和非 X86 分别增长了 16.1%和 146%，非 X86 的占比已经达到了 48%，环比提升了 4 PCT。

图 35：2026Q2 全球不同类型服务器（亿美元）



资料来源：IDC，中原证券研究所

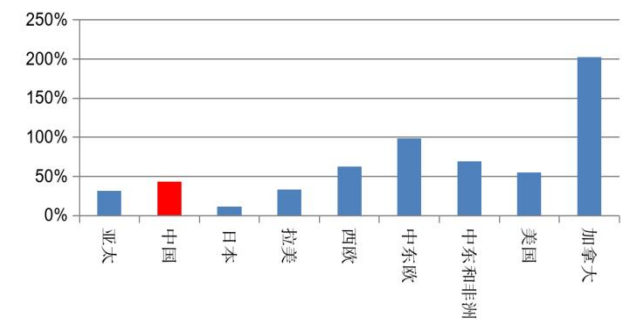
图 36：2026Q2 全球不同 CPU 类型服务器（亿美元）



资料来源：IDC，中原证券研究所

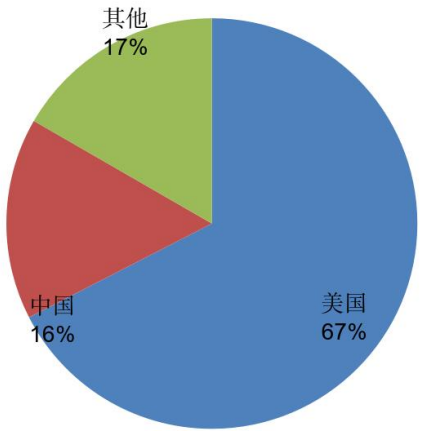
分区域来看，2026Q2 增长最快的地区为加拿大及中东欧，同比增速分别为 202.6%和 98.3%，而美国增速 55%领先行业整体增速，同期中国增速 43%，落后于全球整体水平。从规模上来看，Q2 美国占全球服务器销售额的 67%，中国占比 16%，全球算力的两极化趋势明显。

图 37：2026Q2 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 38：2026Q2 全球服务器区域分布

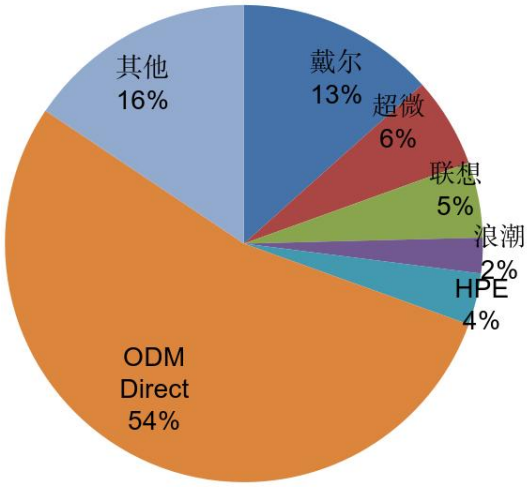


资料来源：IDC，中原证券研究所

分厂商来看，头部的戴尔、超微分别实现了 165%和 97%的增速，同样位于美国的 HPE 是增速 46%，联想增速 99.6%同样获得较快的增长，ODM 厂商仅增长了 35.2%，品牌厂商优势扩大更为明显，而国产厂商浪潮 8%，表现相对较弱。

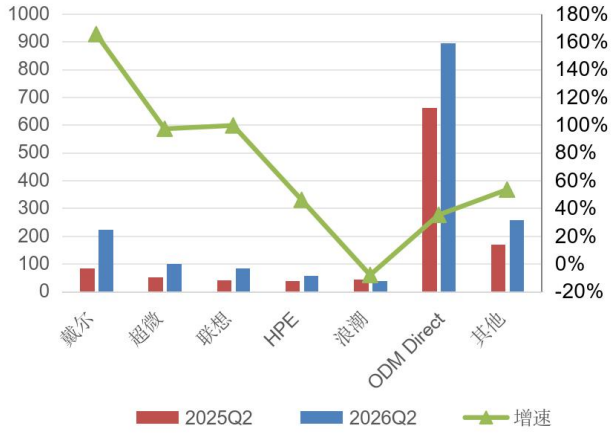
仅供内部参考，请勿外传！

图 39：2026Q2 全球服务器厂商份额



资料来源：IDC，中原证券研究所

图 40：2026Q2 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

1.4.1.2. 国内市场：服务器企业中报情况普遍向好，厂商加大存货和预付账款趋势明显

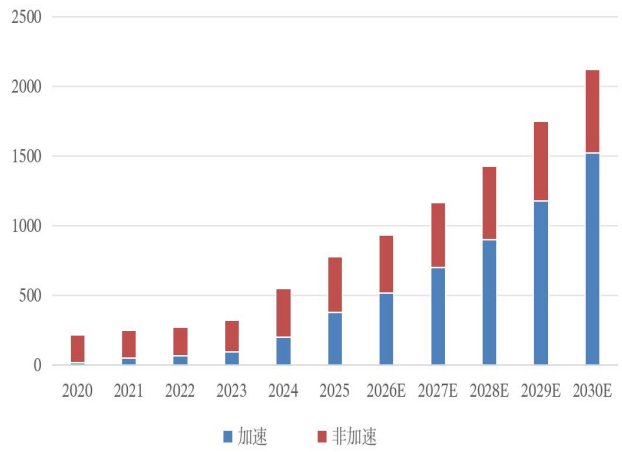
中国市场上，加速服务器也正在成为整体市场的增长引擎，IDC 预计到 2030 年其在整体服务器市场占比将达到 70%以上。

图 41：2018-2025 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 42：中国加速服务器增长预测（亿美元）

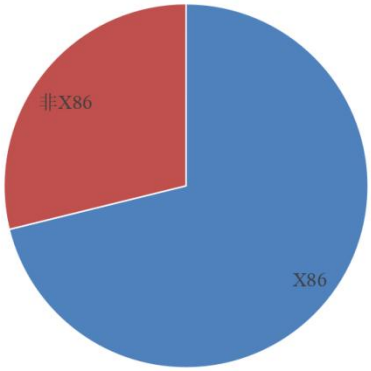


资料来源：IDC，中原证券研究所

2025 年我国服务器市场中，X86 占比超过七成，其中浪潮仍然牢居市场第一，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。

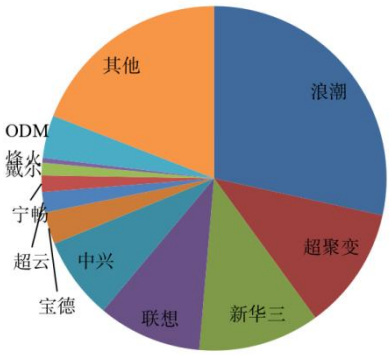
仅供内部参考，请勿外传！

图 43：2025 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 44：2025 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 2026 年半年报来看，企业业务普遍呈现出较好的增长势头。海外布局的工业富联和联想集团，上半年服务器相关业务同比分别增长了 76%和 98%。对标国内市场的企业中，中科曙光呈现了收入和净利润的同步上涨，增速分别为 28%和 27%；软通动力收入增长 40%；紫光股份旗下新华三收入增长 37%，利润增长 29%；浪潮信息收入增速为 5%，利润增长 270%；超聚变收入增速为 21%，毛利率提升较为明显；神州数码服务相关业务同比大幅增长 112%。

表 8：2026H1 企业服务器相关业务对比（亿元）

公司	相关业务	2026H1						2025		
		收入	增速	净利润	增速	毛利率	毛利额	收入	增速	人员增速
工业富联	云计算		76%					6027	89%	7%
	云服务商 AI 服务器		230%						300%+	
	GPU AI 机柜出货量		320%							
	ASIC AI 机柜出货量		300%							
	ASIC CPU 服务器出货量		250%							
浪潮信息	-	843	5%	30	270%	8.7%	74	1648	43%	-1%
联想集团	基础设施解决方案	579	98%	53*				1212	37%	2%
紫光股份	新华三	500	37%	24	29%			760	38%	-12%
华勤技术	计算及数据业务	433				6.9%	30	754	52%	20%
超聚变	-	373	21%	21		15%	58	582	32%	12%
中科曙光	-	75	28%	9.7	33%	27%	20	150	14%	
	IT 设备业务	67	28%			25%	16	125	7%	1%
宝德	（慧博云通拟收购）							109	11%	
软通动力	计算产品与智能电子	95	40%			7.7%	7.3	158	22%	
神州数码	自有品牌产品	66	112%			7.7%	5.0	74	62%	-6%
广电运通	广电五舟	7.4	-14%	0.10	4%			18	5%	3%
拓维信息	计算机、通信和其他电子设备制造业	1.9	-33%			34%	0.65	8.5	-59%	
	湘江鲲鹏	1.5	-34%	0.2	127%			6.3	-60%	

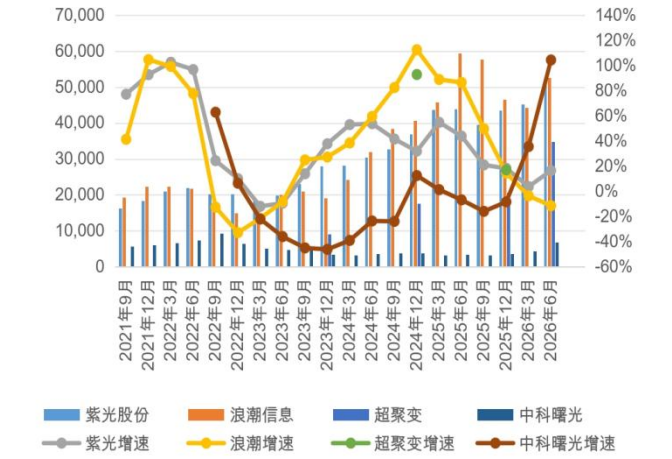
资料来源：上市公司公告，Wind，中原证券研究所（*运营利润）

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

仅供内部参考，请勿外传！

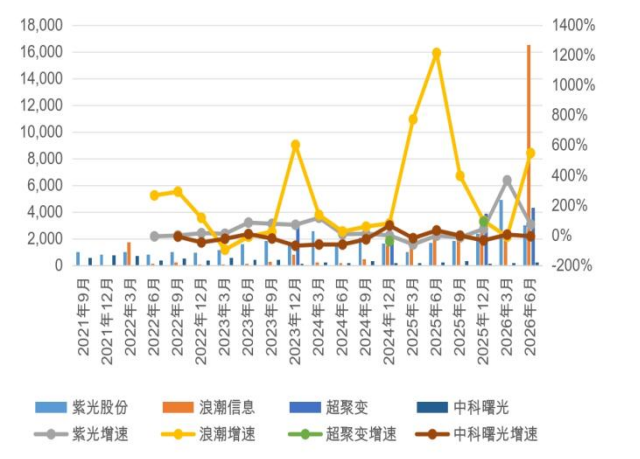
在持续凸显。从存货来看，紫光 2026 年前两个季度分别环比增长了 4%和 13%；浪潮 Q1 存货环比下滑了 5%，Q2 环比增加了 19%；中科曙光前两个季度环比分别增长了 23%和 57%，增长最为迅速；超聚变 6 月末存货较年初增长了 69%，仅次于中科曙光。从两家头部服务器厂商预付账款数据对比来看，紫光在 Q1 预付账款环比提升 105%以后，Q2 略有下滑 39%；而浪潮虽然 Q1 的预付账款环比下滑 37%，但是 Q2 同比增长 769%，呈现了数量级的增长，背后反映出当前服务器厂商对于零部件锁单的强烈需求。

图 45：2021Q3-2026Q2 服务器厂商存货规模及增速（百万元）



资料来源：上市公司公告，中原证券研究所

图 46：2021Q3-2026Q2 服务器厂商预付账款规模及增速（百万元）



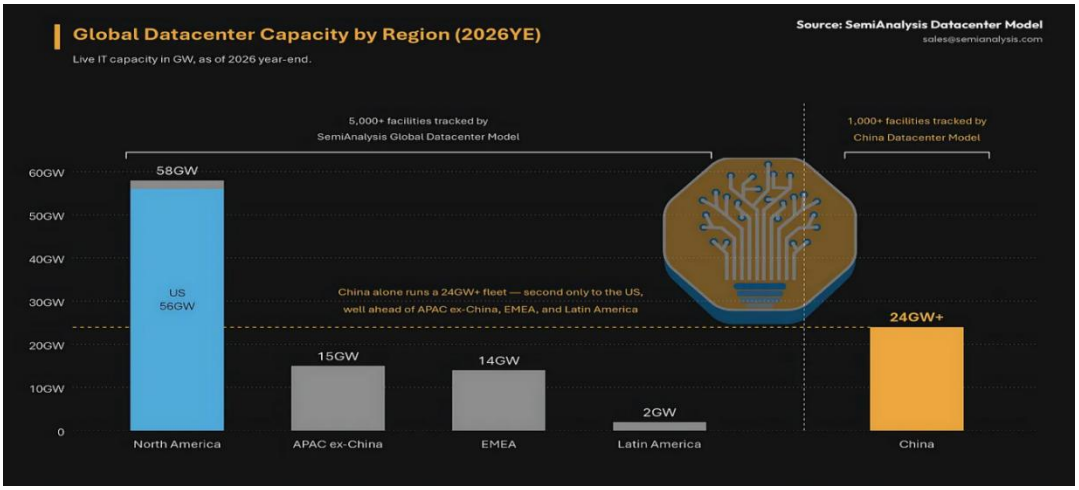
资料来源：上市公司公告，中原证券研究所

1.4.2. 数据中心：2026 年我国智算建设集中化趋势明显

1.4.2.1. 我国数据中心建设概况

全球算力建设上，呈现了中美两强格局。根据 SemiAnalysis 的预测，到 2026 年底全球算力分布上来看，美国以 56GW 的装机容量领先，中国将达到 24GW，亚洲除中国以外地区 15GW，欧洲、中东和非洲（EMEA）地区为 14GW。

图 47：2026 年底全球分区域数据中心容量分布

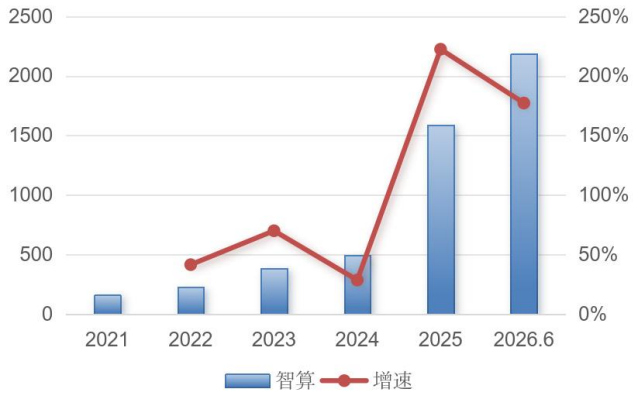


资料来源：SemiAnalysis，中原证券研究所

仅供内部参考，请勿外传！

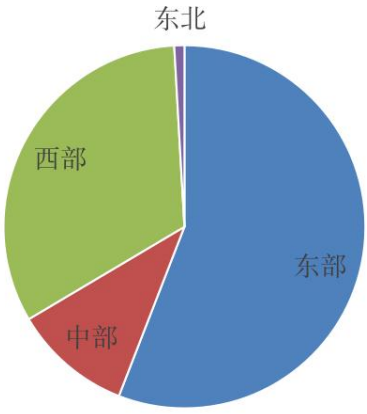
2026 年我国智能算力在快速增长过程中。根据工信部数据，2026 年上半年，我国智能算力规模达到 2185EFLOPS，同比增长 177%，环比增速为 16%。从区域分布上来看，智算高度集中在东部的经济发达区域和西部的能源福集区域，占比分别为 55.9%和 32.6%。同时，全国算力设施整体上架率达到 71.4%，这较 2024 年 7 月出台的《数据中心绿色低碳发展专项行动计划》提出的，到 2025 年底全国数据中心整体上架率不低于 60%的目标来看，上架率有了明显的提升。

图 48：2017-2026 我国智能算力规模（EFLOPs）



资料来源：工信部，原证券研究所

图 49：2026 年 6 月底我国智算区域分布



资料来源：工信部，中原证券研究所

在用标准机架方面，全国呈现“一超多强”的阶梯式分布，区域差距显著。截至 2026 年 6 月，河北在用标准机架达 250 万架，江苏、广东、贵州紧随其后，在用标准机架均超过 140 万架，是东部数字经济核心区与西部算力枢纽的典型代表。

图 50：我国在用标准机架分布

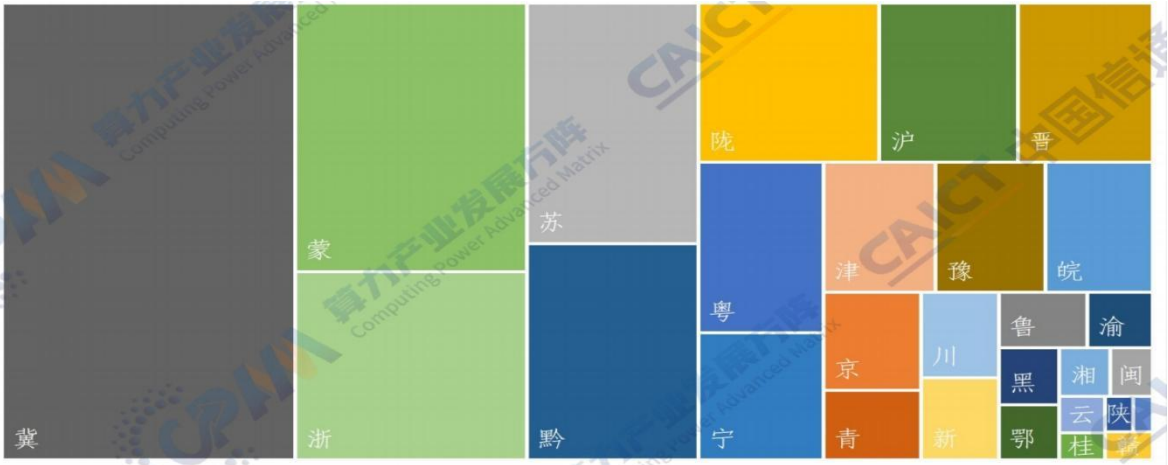


资料来源：信通院，中原证券研究所

在智算方面，资源集中性特征更为突出。截至 2026 年 6 月，河北省智算规模达 556EFLOPS，内蒙古、浙江、江苏、贵州均超过 150EFLOPS。北方能源型省份凭借成本成为核心承载区，东部的浙江、江苏凭借应用需求拉动快速增长。

仅供内部参考，请勿外传！

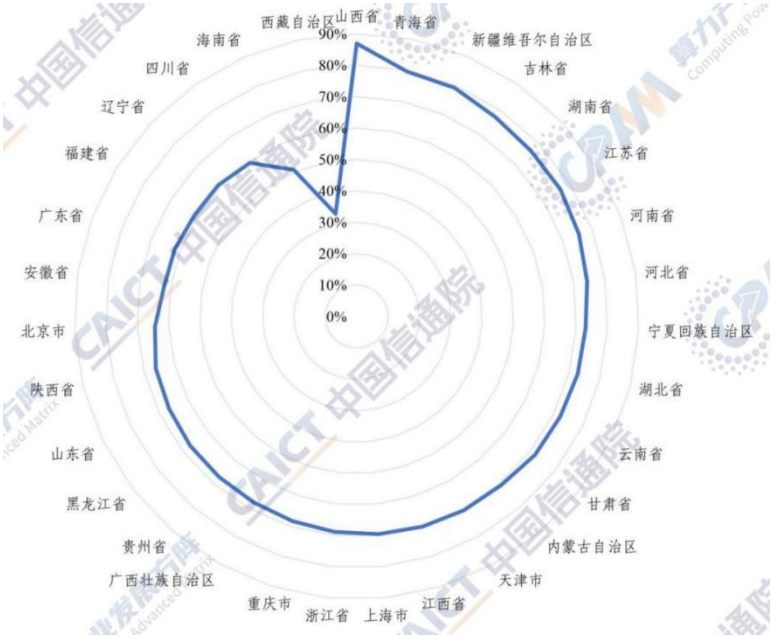
图 51：我国在用智算规模（FP16）分布



资料来源：信通院，中原证券研究所

从实际算力使用情况来看，山西、青海、新疆、吉林、湖南、江苏、河南上架率均超过 75%，其中江苏、山西在机架数和上架率上都位居前列，而河北、内蒙古、贵州大量新建的算力还处于爬坡期，机架上架率尚未充分释放。

图 52：各省上架率情况



资料来源：信通院，中原证券研究所

截至 2026 年 6 约，我国已建成千卡智算设施 4561 个，万卡智算设施 52 个，排名前 10 的智算企业及机构智算总规模达到 1266EFLOPS，占全国总规模的 57.9%，呈现高度集中化特征。

1.4.2.2. 数据中心厂商发展情况

2026 年厂商之间的差距逐步拉大，有多笔大单落地。世纪互联 2026 年上半年累计新签约

仅供内部参考，请勿外传！

订单总量达到 862MW，公司计划 2026 年交付 450MW，2027 年上半年交付 252MW，建设全面加快，同时其 2026 年上半年新增了约 1.8GW 的批发资源容量，其中包括了 500MW 的海外资源，总批发资源容量达到了 4GW。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，同时与客户签署了多笔 5 年期算力服务订单，订单总规模达到 390-460 亿元。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。

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 8 月设立，发行规模 17.955 亿元。尚杭科技拟登陆北交所。

表 9：2026 年企业数据中心相关业务对比（亿元）

公司	相关业务	2026H1					算力规模				2025	
		收入	增速	净利润	增速	毛利率	使用率	算力规模	算力增长	总体规划	收入	增速
中国移动	算力服务	529	14%									
	—数据中心	190	13%					3.75GW				9%
	——AIDC		486%					7GW	5GW			
	—云算服务		18%									
	——智算服务	53	130%									
	—云算应用		5.9%									
中国电信	智算		95%									
	AIDC		9.3%					3.2GW			345	
中国联通	数据中心						72%	2.75GW	250MW		281	9%
万国数据		65	15%	35	404%	28%	79.2%	68 万平米	1.7 万平米		114	11%
世纪互联		55	17%	-6.7	-168%	20%	74%-65%	1007MW+5 万架	118MW	4GW+5 万架	100	21%
秦淮数据	（东阳光）							799MW		4GW	64	6%
润泽科技		37	50%	12	36%	47%	90%	750MW	0MW	6GW	57	30%
宝信软件	服务外包	18	7%			43%					43	3%
城地香江		15	15%	-0.06	-117%	17%					39	138%
奥飞数据		15	28%	2	123%	34%		382M	40MW		25	16%
光环新网	IDC 及其增值服务	12	9%			23%	60%	361MW	114MW	1GW	22	7%
恒润股份	算力行业/润六尺	11	4%	0.88	290%						17	744%
郑州合盈	（豫能控股）						60%-79%	579MW			17*	
数据港		7.8	-4%	0.85	1%	29%		401MW	0MW		17	0%
科华数据	IDC 服务	6.9	13%			13%					13	4%
杭钢股份	数字产业经济	4.2				21%		69.08 机柜			8	82%
尚航科技		3.3	32%	0.17	-55%	17%		300MW			5	13%
大位科技		2.6	26%	-0.7	-202%	16%					4	8%

仅供内部参考，请勿外传！

美利云		1.7	-1%	0.3	65%	43%					3.4	-62%
	云业务	1.6	-1%			43%					3.2	4%
迈信林	提供服务	1.2	197%			24%					1.3	612%
群兴玩具	智算服务	0.3	70%			-11%					0.5	100%

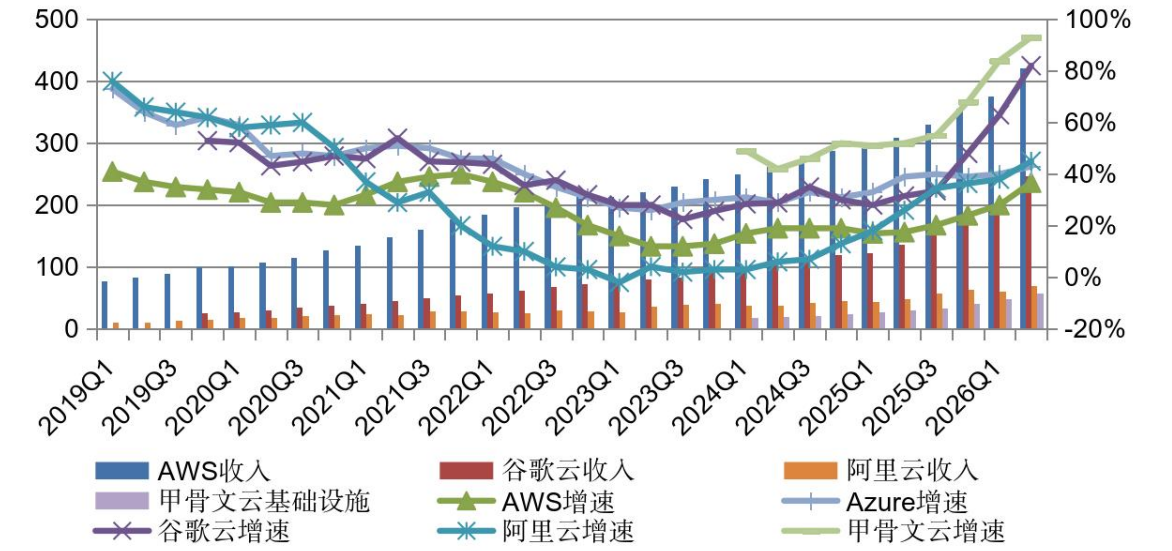
资料来源：上市公司公告，Wind，中原证券研究所（*年化处理后数值）

1.4.3. 云计算：受到 AI 需求拉动，整体增速仍在加快

1.4.3.1. 全球：云厂商云业务增速继续加快

全球云巨头继续呈现了加速趋势。根据 2026Q2 相关公司财报，亚马逊的 AWS 实现收入 422 亿美元，同比增长 36.7%，较上个季度提升了 8.7 PCT。微软 Azure 和谷歌云分别增长了 43%和 82%，增速明显快于 AWS，分别较上个季度增长了 3 和 19 PCT。

图 53：2019Q1-2026Q1 云巨头的相关业务收入及增速

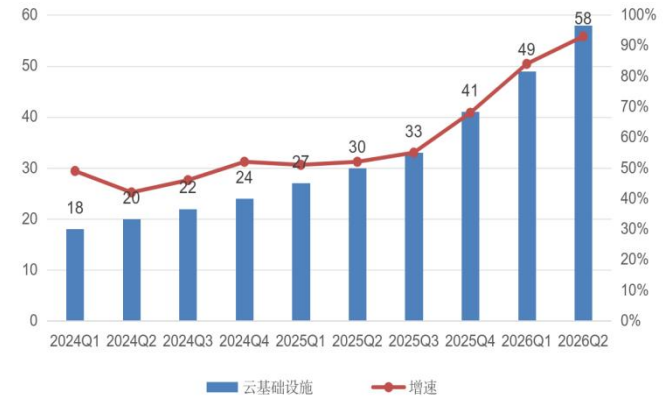


资料来源：上市公司财报，中原证券研究所

甲骨文云基础设施业务增速加快至 93%。2026 财年 3 季度，甲骨文云基础设施营收 58 亿美元，同比增长 93%，较上个季度上涨 9 PCT，仍然是云业务增速最快的厂商。同时甲骨文剩余履约义务达到 6380 亿元，同比增长 363%，较上财季增长 850 亿美元。受到巨额合同的需求支撑，2025 年 9 月甲骨文预计 2027-2030 财年基础设施营收将分别为 320 亿美元、730 亿美元、1140 亿美元、1440 亿美元。

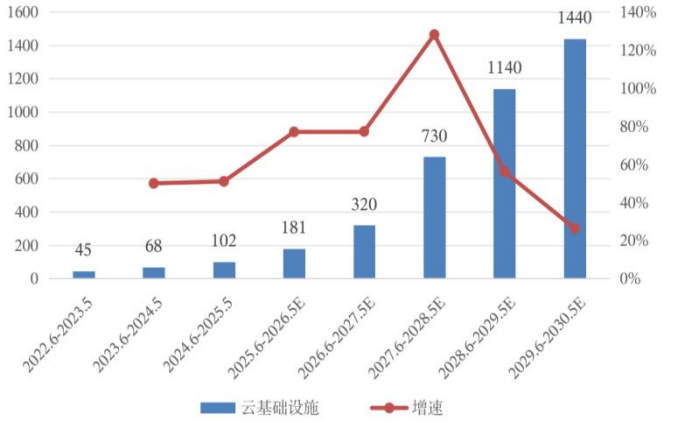
仅供内部参考，请勿外传！

图 54：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 55：甲骨文云基础设施业务年度收入预测（亿美元）

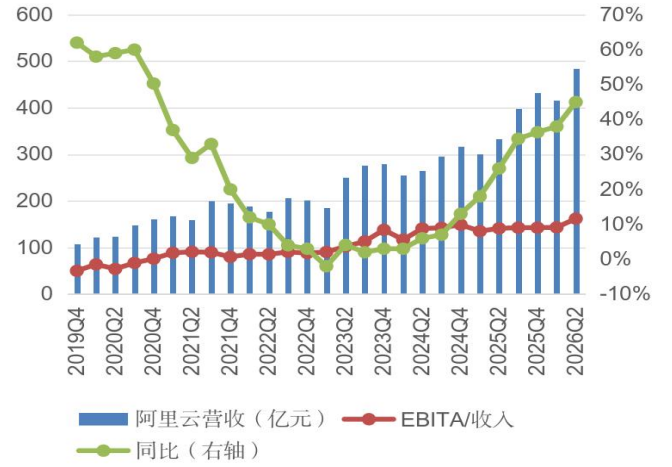


资料来源：甲骨文财报，中原证券研究所

1.4.3.2. 国内：阿里云 AI 相关产品收入连续 12 个月三位数增长

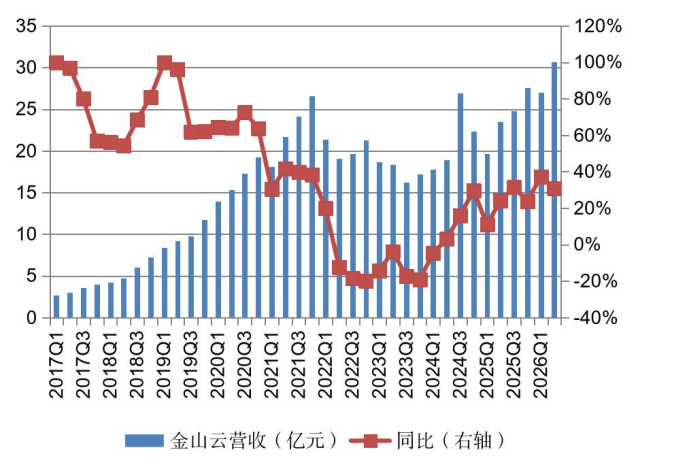
从厂商业绩来看，在 AI 推动下，阿里云增速持续提升，AI 相关收入连续 12 个月三位数增长。其中 2026Q2 阿里 AI 云与算力服务收入 484.37 亿元，增速达到了 45%，较上个季度提升了 7PCT，其中 AI 相关收入 123.76 亿元，连续 12 个季度实现同比三位数增长，截至 8 月 MaaS 业务的年化收入已经超过 160 亿元。2026Q2 金山云同比增长 31%，处于较高增速水平，其中智算云账单收入同比增长 82%，至 13.27 亿元。

图 56：2019Q4-2026Q1 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 57：2017Q1-2026Q2 金山云收入及其增速



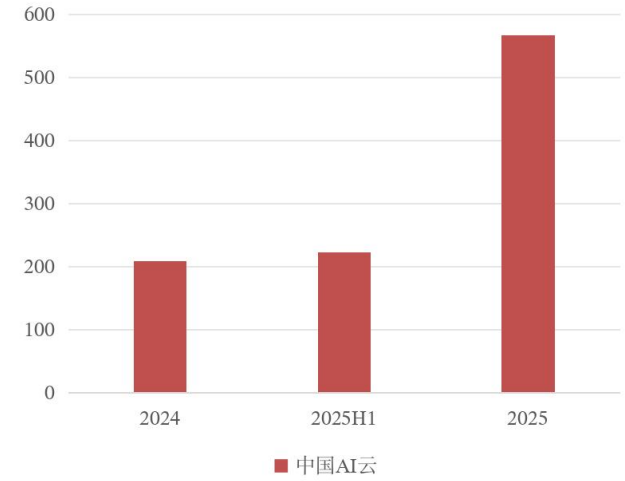
资料来源：金山云财报，中原证券研究所

在 AI 云加速增长趋势明显。根据 Omdia 数据，2025 年市场规模为 567 亿元，超出了此前 IDC 预期（518 亿），同比增长 172%，其中 2025 年下半年达到了 344 亿元，环比增长 54%。2025 年全年来看，云 AI 市场中有 69%为 IaaS，23%为面向模型的生产服务 MPS，8%为即开即用的模型 API 服务 MCS。

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

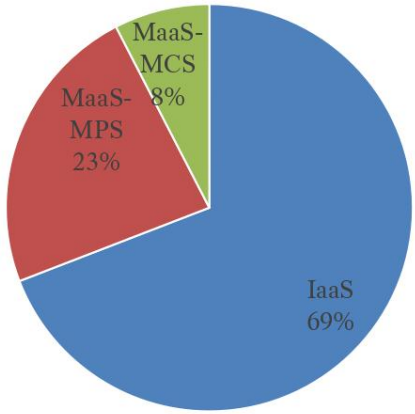
仅供内部参考，请勿外传！

图 58：2024-2025 中国 AI 云市场规模（亿元）



资料来源：Omdia，中原证券研究所

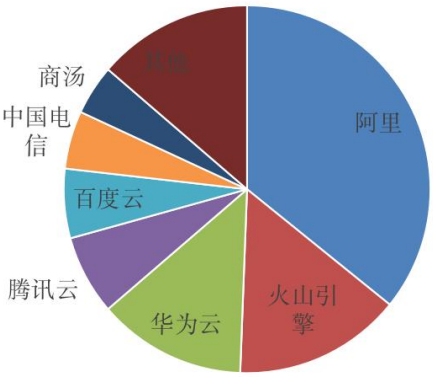
图 59：2025 中国 AI 云市场细分市场份额



资料来源：Omdia，中原证券研究所

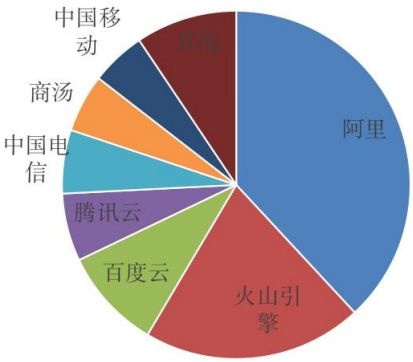
AI 云厂商中，阿里、字节表现最为优异。从 Omdia 给出的厂商份额数据来看，阿里以 38.1%的份额位列第一，在 IaaS 和 MaaS 中的 MPS 方向的份额都位于市场第一；火山引擎（字节）在 2025 下半年份额增长明显，份额从上半年的 14.8%，增长到了全年的 20.4%；上半年位于市场第三的华为，已经没有出现在全年头部厂商中；百度云全年超过了腾讯云，位列行业第三；腾讯 2025H2 份额仅 5.7%，不及上半年 7.0%的水平，在缺芯的情况下，其采取了优先保证内部需求的策略；后续厂商中，中国电信、商汤、中国移动全年都有不同程度的份额提升。

图 60：2025H1 中国 AI 云市场格局



资料来源：Omdia，中原证券研究所

图 61：2025 中国 AI 云市场格局



资料来源：Omdia，中原证券研究所

1.4.4. 资本开支：Q2 国内资本开资明显提速

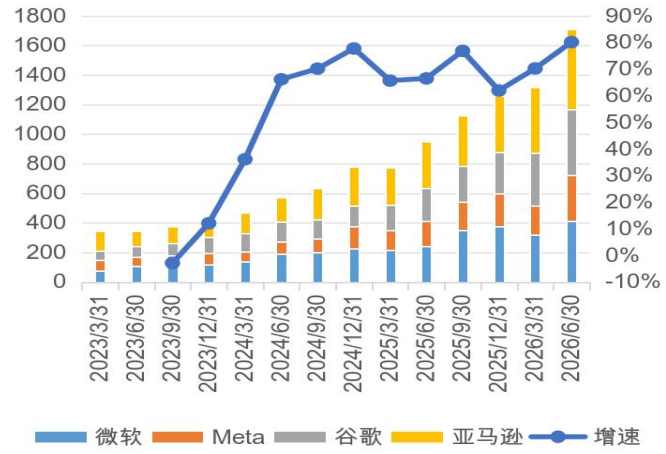
1.4.4.1. 海外资本开支：Q2 海外科技企业资本开支增速创新高

从海外市场来看，2026Q2 美国 大科技厂商（亚马逊、微软、谷歌、Meta）的资本开支增速创新高。2026Q2 4 大厂商资本开支再创新高，总计达到 1712 亿美元，同比增长 80%，增速创新高。其中亚马逊资本投入 542 亿美元，增长 69%，仍然是投入最大的厂商；谷歌资本

仅供内部参考，请勿外传！

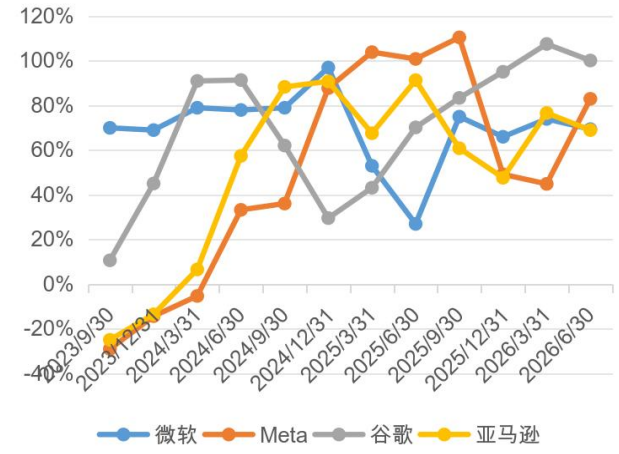
开支 449 亿美元，增长 100%，在云业务加快增长的同时资本开支增速持续领先；微软资本开支 410 亿美元，同比增长 69%；Meta 资本开支达到了 311 亿美元，同比增速 83%，Q2 加快投入步伐。

图 62：4 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 63：4 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

随着海外科技巨头资本开支整体继续呈现上调趋势，对 2027 年给予积极预期。其中亚马逊在 2025 年 1310 亿美元资本开支基础上，2026 年计划资本开支上调到 2200 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 1300-1450 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；微软资本开支约达 1900 亿美元，但是受到会计核算规则变更的影响，将降低到 1750 亿美元；甲骨文 2026 财年本开支 557 亿美元，计划在 2027 财年投入约 700 亿美元。**除甲骨文以外的 4 大科技厂商 2026 年资本开支约 7125 亿美元，甲骨文、谷歌、亚马逊都给出了对 2027 年积极的投入预期。**

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。2026Q2，微软 CRPO 为 6780 亿美元，同比增长 84%；甲骨文剩余履约价值（RPO）同比大增 362%至 6380 亿美元；谷歌云业务订单积压 5140 亿美元，同比增长 385%；亚马逊订单积压同比增长 154%至 4960 美元。

表 10：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。

仅供内部参考，请勿外传！

	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。
	2026.4.30	2026 日历年全年，微软预计资本开支约达 1900 亿美元，其中约 250 亿美元来自更高组件定价的影响。
	2026.7.30	受数据中心使用寿命会计政策变更（从 15 年延长至 25 年）影响，部分租赁从融资租赁转为经营租赁，2026 自然年资本支出预计将在 1750 亿美元左右。
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。
	2026.7.4	2026 财年资本支出达 557 亿美元，计划在 2027 财年投入约 700 亿美元
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
	2026.7.31	2026 年资本预期增长到 2200 亿美元，主要原因是内存价格上涨以及持续对 AI 和数据中心基础设施的投入。即便完成 2200 亿资本投入，年内算力产能仍无法完全满足客户订单，2027 年算力缺口将延续。
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。
	2026.4.30	上调 2026 年全年资本支出指导范围至 1800 亿美元至 1900 亿美元，并预计 2027 年的资本支出将比 2026 年显著增加。
	2026.7.22	全年资本支出预期上调至 1950 亿美元至 2050 亿美元区间，预计 2027 年继续显著增加
Meta	2025.1.24	计划在 2025 年投50 600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

仅供内部参考，请勿外传！

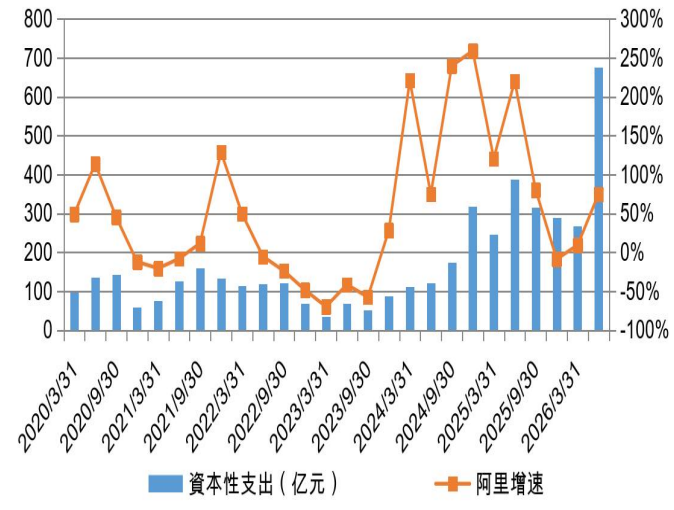
	至 1180 亿美元之间，其中用于生成式 AI 以及核心业务方面的投资范围从 600-650 亿美元之间上调到 640-720 亿美元之间。
2025.7.31	将大力推进人工智能基础设施建设，计划 2025 年相关资本支出（含融资租赁本金支付）达 660-720 亿美元，其中值较上年同比增长约 300 亿美元。这一积极的投资增长态势将延续至 2026 年，以持续提升在线产能，满足人工智能研发及业务运营的需求。
2025.8.27	特朗普表示 Meta 计划在路易斯安娜州乡村兴建大型数据中心项目，总投资将达到 500 亿美元。
2025.9.4	到 2028 年将在美投资 6000 亿美元。
2025.10.30	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。
2026.4.30	2026 年全年资本支出上调至 1250 亿美元至 1450 亿美元
2026.7.30	预计 2026 年资本支出（含融资租赁本金付款）将在 1300 亿美元至 1450 亿美元区间。2026 年 7 月新签数据中心租约约 680 亿，2027-2028 起租，租期 18-20 年。

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

1.4.4.2. 国内资本开支：2026Q2 国内资本开资呈现增长拐点

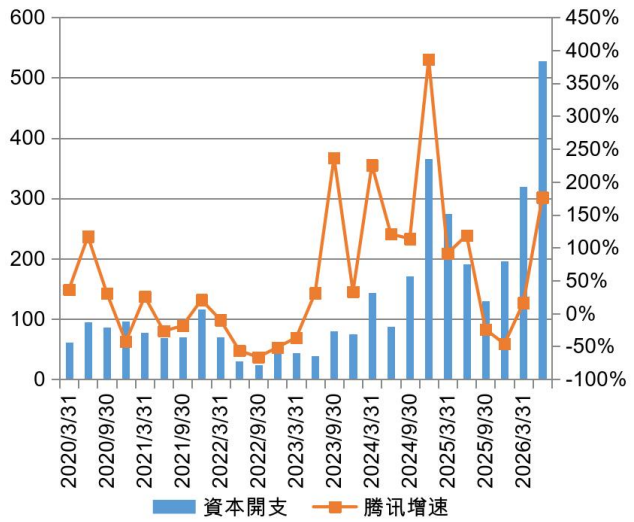
腾讯和阿里在 Q2 双双实现资本开支的快速增长。腾讯 Q2 资本开支 527.84 亿元，同比增长 176%，环比增长 65%，2025Q3 见底后连续 3 个季度回升，明显超出市场预期。阿里 Q2 资本开支 676.78 亿元，同比增长 75%，环比增长 151.7%，达到历史新高，开支规模超过腾讯。

图 64：2020-2026 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 65：2020-2026 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领域，开支直接导致全年净利润同比下滑超 70%。2026 年资本开支投入进一步扩大，原计划从 2025 年的 1500 亿元增长到约 1600 亿元，5 月又进一步提升至 2000 亿元。

仅供内部参考，请勿外传！

表 11：海外科技厂商算力相关投入

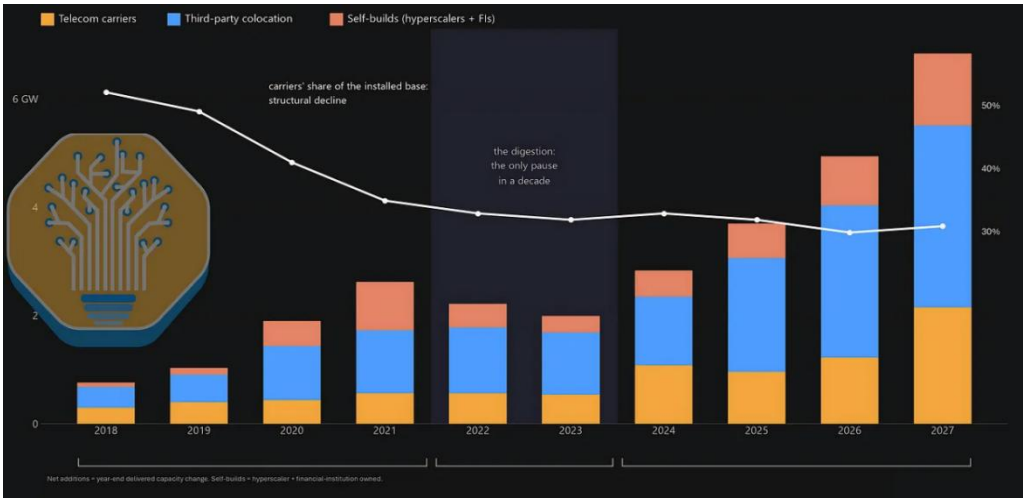
公司	时间	详情
字节跳动	2024.12.29	在 AI 上投入巨大，2024 年资本开支达到 800 亿元，2025 年资本开支有望达到 1600 亿元，旨在打造自主可控的大规模数据中心集群，其中约 900 亿元将用于 AI 算力的采购，700 亿元用于 IDC 基建以及网络设备。
	2025.1.24	2025 年将拨出超过 1500 亿元（206.4 亿美元）的资本支出，其中大部分将集中在人工智能（AI）上。
		2025 年三、四季度，字节大幅增加了 AI 算力采购、模型研发和数据中心基础设施的资本支出，全年资本开支超过 1500 亿元，其中约 900 亿砸向 AI 算力采购
	2025.12.23	初步计划 2026 年在 AI 领域投入 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。
	2026.5.9	2026 年资本开支上调至 2000 亿元，字节跳动 AI 基础设施的投入，包括自研大禹 AI 液冷整机柜、万卡级训练集群等。
	2026.5.28	将 2026 年 AI 相关资本开支内部评估上限抬升至 700 亿美元(约合人民币 4776.8 亿元)
阿里	2025.2.24	未来三年，阿里将投入超过 3800 亿元，用于建设云和 AI 硬件基础设施，总额超过过去十年总和。
	2026.5.13	未来五年 AI 基建投入远超 3800 亿，将采用自建+租赁+芯片销售多模式布局。
腾讯	2026.3.18	2025 年资本开支规模低于预期的原因在于先进 GPU 在中国供应受阻，最后通过租用弥补了一些算力需求，同时减少对外租售来保证自身的算力。2026 年预计还会增加资本开支，获取算力支撑模型训练和推理，同时有额外的算力对外租售。

资料来源：金融时报，The Information，爱集微，钛媒体，中原证券研究所

1.4.4.3. 我国算力发展趋势：互联网厂商主导，电价敏感度更高，字节算力规模最大

算力市场结构也在发生根本性变化，新增需求主要来自批发型数据中心，零售型数据中心利用率下滑明显，廉价的电力的吸引和更快传输网络的支持下，以内蒙古为代表的西部地区算力建设需求快速增长。从承运人上来看，运营商占比缩小，而自建数据中心和第三方数据中心占比持续提升。

图 66：中国超大规模数据中心容量（2017-2028）

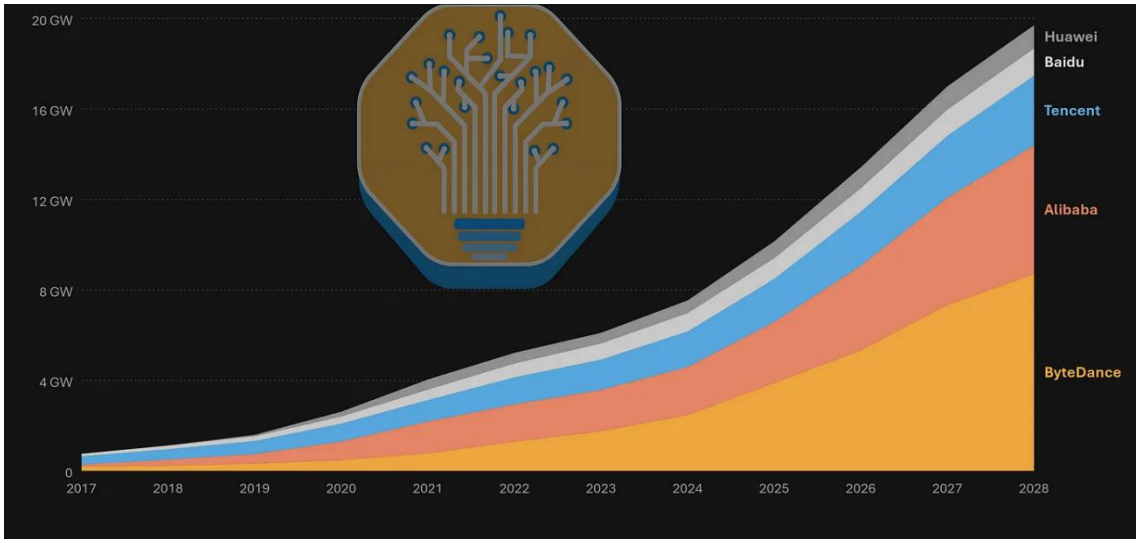


资料来源：SemiAnalysis，中原证券研究所

仅供内部参考，请勿外传！

根据 SemiAnalysis 数据，BATB（字节、阿里、腾讯、百度）资本开支从 2024 年的 350 亿美元，增长到 2025 年超过 500 亿美元，2026 年将达到 1000 亿美元。

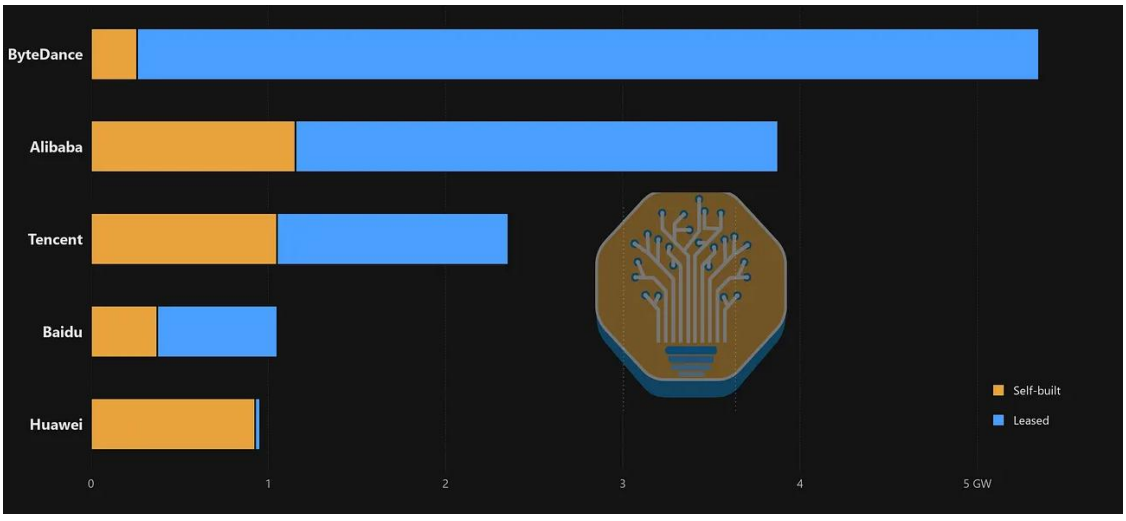
图 67：中国超大规模数据中心容量（2017-2028）



资料来源：SemiAnalysis，中原证券研究所

根据 SemiAnalysis 数据，到 2026 年底，字节将拥有约 5.3GW 数据中心，其中少部分为自建，大部分来自租赁算力；阿里巴巴作为国内最大的云厂商，数据中心规模仅次于字节，约为 3.9GW，自建数据中心占比也更高；腾讯数据中心约 2.4GW，自建比例更高；百度数据中心容量约 1.1GW，华为约 1.0GW，总体规模相当，但是华为基本都来自自建数据中心。

图 68：2026 年中国超大规模数据中心容量



资料来源：SemiAnalysis，中原证券研究所

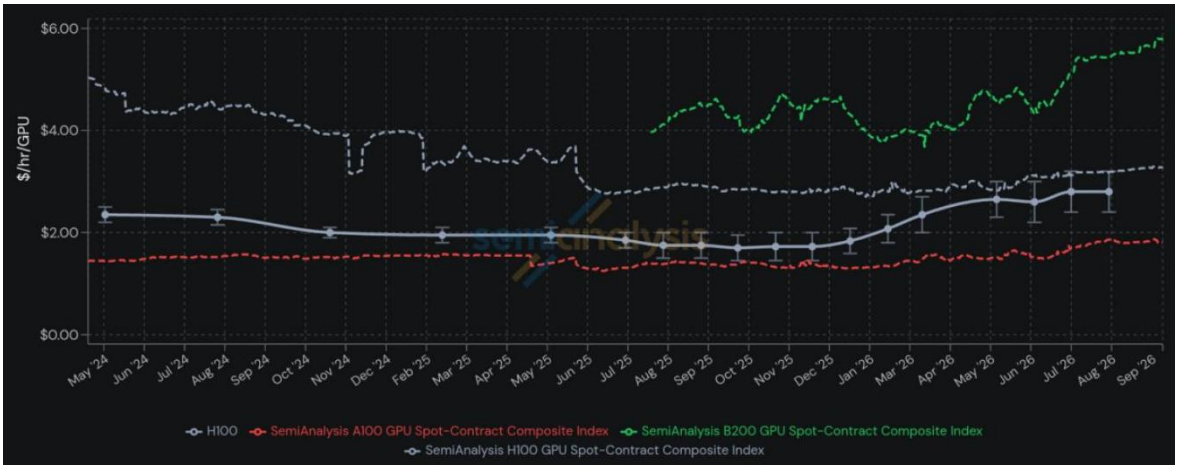
1.4.5. 算力租赁价格

根据 SemiAnalysis 发布的算力租赁价格趋势数据，A100 仍然处于近 2 年以来的高位，H100

仅供内部参考，请勿外传！

也处于 2025 年 6 月以来的高位水平，B200 的租赁价格还在持续提升的过程中。

图 69：GPU 租赁价格趋势（1 年期合同）



资料来源：SemiAnalysis，中原证券研究所

根据阿里巴巴 8 月业绩报告会，市场预测到 2030 年以前算力紧缺还将持续，阿里测算其资本开支 3 年回本，其 2018 年购买的 V100 和 2020 年购买的 A100 都还在满负荷运行。

1.5. AI：头部模型继续呈现降价趋势

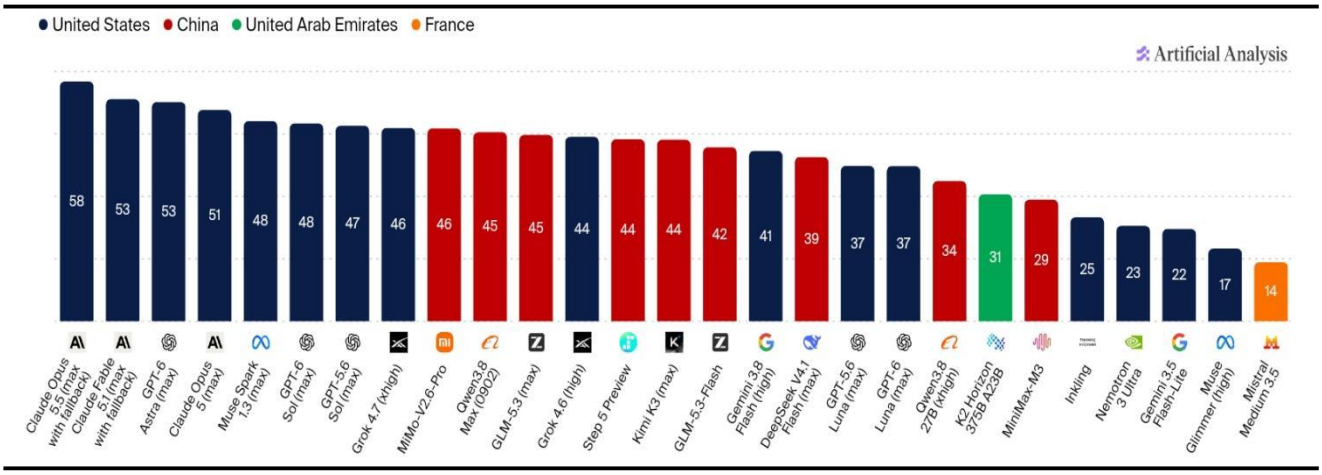
1.5.1. 中美大模型发展概况

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。Anthropic 9 月新推出的 Opus 5.5 和 Fable 5.1 全球领先，OpenAI GPT-6 Astra 紧随其后；Meta 的 Muse Spark 1.3，模型能力重回头部模型梯队，同等任务实际 Token 消耗明显降低（25%）；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；小米在 9 月推出了 MiMo-V2.6-Pro，位居国产模型和开源模型第一；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，位居国产第二；智谱 8 月推出的 GLM-5.3 与 Qwen3.8-Max 智能指数持平，GLM-5.3-Flash 形成了极致性价比同时具有多模态能力；Kimi 在 7 月推出的 Kimi K3 超过了 Opus 4.8，形成了对头部模型的市场威胁，当时位居开源模型第一位，但是由于更新节奏问题暂时有一定落后；谷歌原计划推出 Gemini 3.5 pro 持续跳票，或已搁置，转而更新 Flash 产品，后续或将通过 Gemini 4 来填补空挡；DeepSeek 在 9 月推出了 DeepSeek-V4.1 Falsh，仍具有性价比优势；MiniMax 6 月推出了 MiniMax-M3，但是模型能力与头部差距较为明显。**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。由于当前 Anthropic 和 OpenAI 刚进行了一轮更新，国内模型由于发布节奏问题，与头部差距有短暂拉大。**

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

仅供内部参考，请勿外传！

图 70：不同地区的大模型智能水平对比（截至 2026.9.28，红色为中国模型）



资料来源：Artificial Analysis，中原证券研究所

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。以美国最贵的模型 Anthropic 的 Opus 5.5 的调用价格来看，其输出价格约为 MiMo-V2.6-pro 的 22 倍，输入价格约为其 9 倍。但是在低廉的价格策略之下，根据 The Information 的报道，DeepSeek-V4 的 API 业务仍然保持了超过 50%的毛利率。**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%；谷歌 8 月推出的 Gemini 3.7 flash 也为原 3.6 Flash 的一半。**

表 12：模型调用价格（每百万 Token，Artificial Analysis2026 年 9 月 24 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 Sol	2026.9	48	$2	$0.2	$10	
	GPT-6 Luna	2026.9	37	$0.1	$0.01	$0.5	
Anthropic	Opus 5.5	2026.9	58	$4	$0.20	$20	$5
	Fable 5.1	2026.9	53	$10	$0.25	$50	$12.5
	Sonnet 5	2026.6	38	$2	$0.2	$10	$2.5
Kimi	Kimi-k3	2026.7	44	￥20	￥2	￥100	
xAI	Grok4.7	2026.9	46	$2	$0.5	$6	
阿里	Qwen3.8-max 0902	2026.9	45	￥12		￥36	
	Qwen3.8-27b	2026.8	34	￥3		￥12	
Meta	Muse Spark 1.3	2026.9	48	$1.25	$0.15	$4.25	
智谱	GLM-5.3	2026.8	45	￥8	￥2	￥28	
	GLM-5.3 Flash	2026.8	42	￥0.8	￥0.23	￥2.8	
谷歌	Gemini 3.8 flash	2026.9	41	$0.75		$3.75	
MiniMax	MiniMax-M3	2026.6	30	(0-512K) ￥2.1（512K-1M）	￥0.84 ￥0.42	￥16.8 ￥8.4	
DeepSeek	DeepSeek V4.1 Flash*	2026.9	40	￥2	￥0.04	￥8	
	DeepSeek V4-pro 0813*	2026.8	36	￥9	￥0.03	￥27	

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

仅供内部参考，请勿外传！

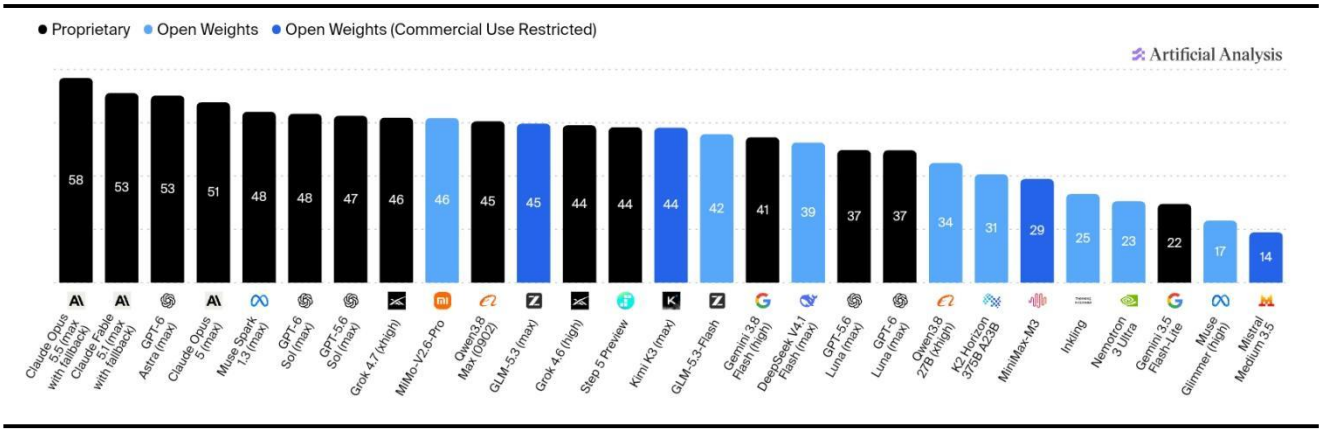
小米	MiMo-V2.6-pro	2026.9	46	¥3	¥0.025	¥6	
	MiMo-V2.6 flash	2026.4	22	¥1	¥0.02	¥2	

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）

6 月马斯克曾经判断中国模型需要到 2027 年 1 月才能达到 Fable 5 级别。随着 7 月 15 日 Kimi K3 的推出，我们看到在 Frontend Code Arena 榜单中，Kimi K3 的前端开发能力已经超过了 Fable 5 和 GPT-5.6 Sol 位居榜首，同时在 Artificial Analysis 的智能水平排名中，紧随 Fable 5 和 Sol。**这意味着当前中国模型已经几乎与头部模型看齐，也意味着我们的赶超速度远快于市场的预期。在一轮海外模型更新后，我们也期待国产模型再次发布高性能的对标产品。**

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；MiMo-V2.6-pro、Qwen3.8 Max、GLM-5.3、Kimi-K3 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 71：大模型智能水平及开源情况（截至 2026.9.28）



资料来源：Artificial Analysis，中原证券研究所

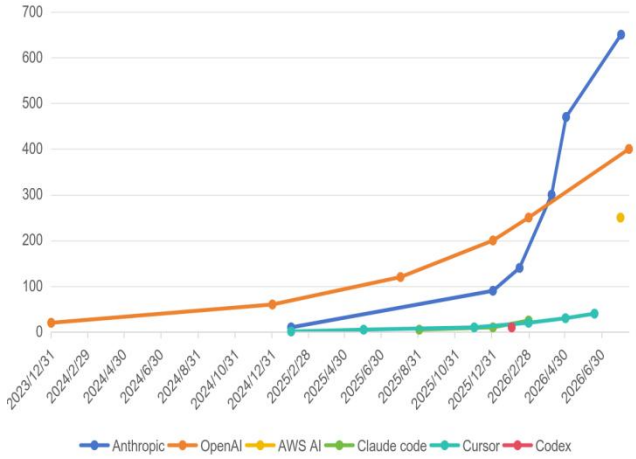
1.5.2. 中美模型厂商收入

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 8 月达到 400 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

仅供内部参考，请勿外传！

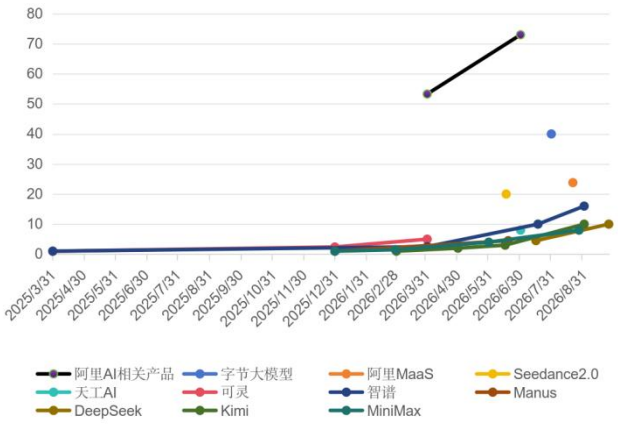
元，体量上和美国模型厂商差距明显。

图 72：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 73：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

1.5.3. 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。

表 13：海外企业切换到国产模型的案例

时间	厂商	行业	选择模型	内容
6.4	Lindy	AI 助手初创企业	DeepSeek V4	两个月前开始测试 DeepSeek 的 V4 模型，其创始人 Flo Crivello 称，DeepSeek 在处理收件箱、日历管理、邮件草拟和会议转录等任务上，表现与 Anthropic 的 Sonnet 相当，但成本仅约为后者的十分之一，仅此一项就为公司节省了数百万美元。
6.12	Detail	美国 AI 识别 Bug 企业	智谱 GLM	约 90%的工作负载从 Anthropic 的 Claude 和谷歌的 Gemini 转移到了自家定制模型以及中国的 GLM 模型上。
6.17	微软		DeepSeek V4 等	考虑在即将推出的 Copilot Cowork 低成本版本中用 DeepSeek 的模型取代 Ahthropic 和 OpenAI 的模型。
6.29	Coinbase	美国加密货币交易平台	GLM 5.2 和 Kimi K2.7	已通过内部 LLM 网关将代码审查、文档总结等常规任务的默认模型从 Anthropic 和 OpenAI 的前沿模型换成智谱的 GLM 5.2 和月之暗mi K2.7。通过更换默认模型、智能路由和强化缓存等措施，Coinbase 的 AI 支出已被压缩近一半。
7.3	微软		Kimi K2.7 Code	旗下编程助手 GitHub Copilot 正式接入了 Kimi K2.7 Code，这是其

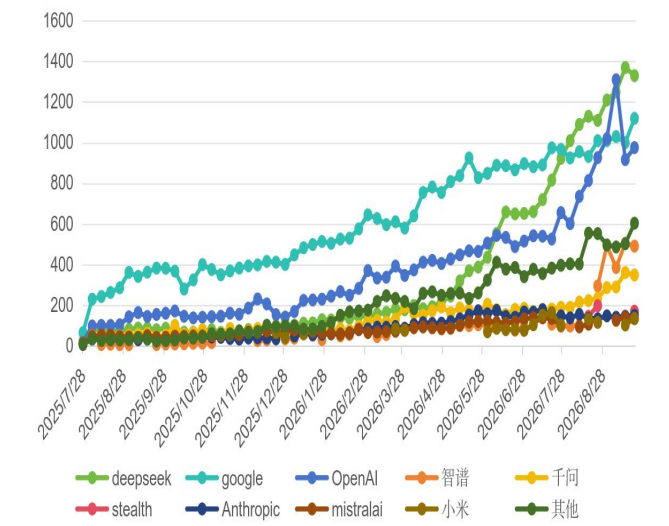
仅供内部参考，请勿外传！

				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。
7.13	DoorDash	美国最大外卖公司	Kimi K2.6	将“较低级别的工作”交给月之暗面的 Kimi K2.6 模型处理，而将 Anthropic 的 Fable 仅保留用于“最困难的工作”。新的组合比之前仅使用 Anthropic 的配置“成本更低，效果却大幅提升”。
7.13	西门子	德国科技巨头	DeepSeek 和智谱	使用的模型范围广泛，包括中国的 DeepSeek 和智谱以及美国前沿实验室、英伟达和法国 AI 集团 Mistral 的模型。
7.13	Airbnb	全球旅行房屋租赁平台		使用了“有限数量的中国来源模型”，并且通过“仅通过经批准的美国服务提供商”运行这些模型，得以保护其数据和运营。
8.26	微软、亚马逊、谷歌	云厂商	Kimi K3	月之暗面正与微软、亚马逊以及 Alphabet 旗下的谷歌洽谈营收分成协议。该协议将允许这几家美国云计算巨头托管其旗舰模型 Kimi K3。

资料来源：澎湃，《金融时报》，智东西，观察者网，金融界，路透社，中原证券研究所

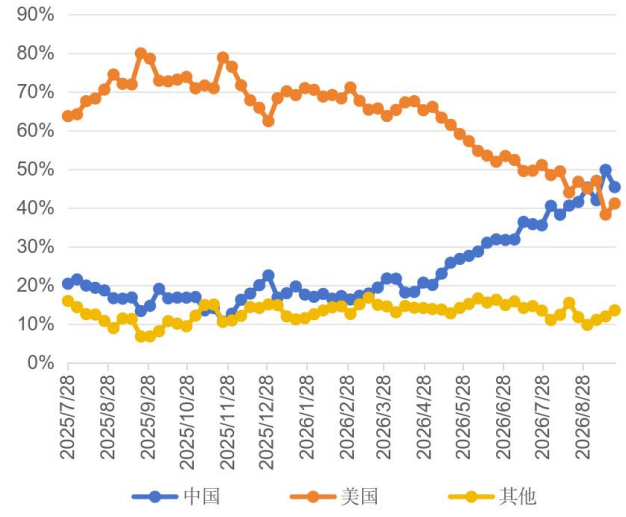
切换国产模型正在成为模型调用发展的趋势。根据 OpenRouter 的模型调用数据（2026 年 9 月 21 日-27 日），DeepSeek 份额 24%，连续 8 周位列第一大模型厂商。同时中国模型厂商份额达到 45%，连续两周领先较美国模型厂商（份额 41%）。

图 74：OpenRouter 调用的模型份额分布（百万 tokens）



资料来源：OpenRouter，中原证券研究所

图 75：OpenRouter 调用的模型份额分布



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 V4 Pro 的价格有望大幅下调，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。

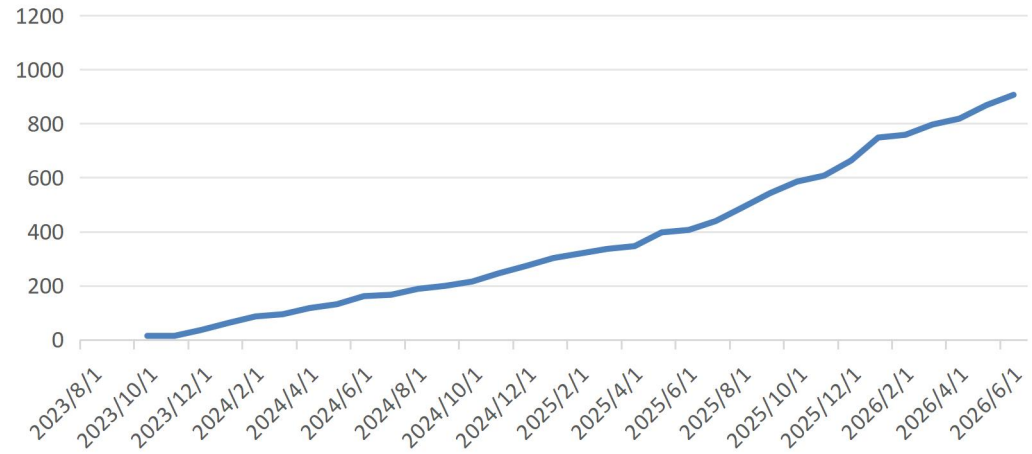
1.5.4. 模型与应用：大模型备案高度集中在发达区域

根据网信办数据，截至 2026 年 6 月，全国大模型备案与登记工作推进取得显著进展，目

仅供内部参考，请勿外传！

前已经有 30 个省市及国资委共计 988 款大模型完成备案，另有 598 款大模型产品在属地完成登记。

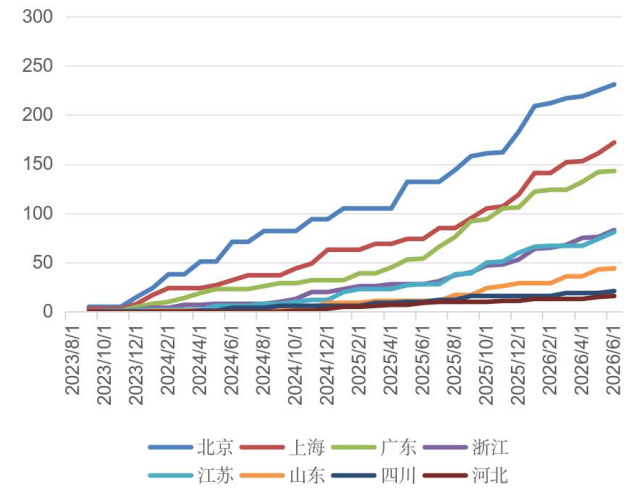
图 76：全国大模型备案与登记数量



资料来源：国家网信办，中原证券研究所

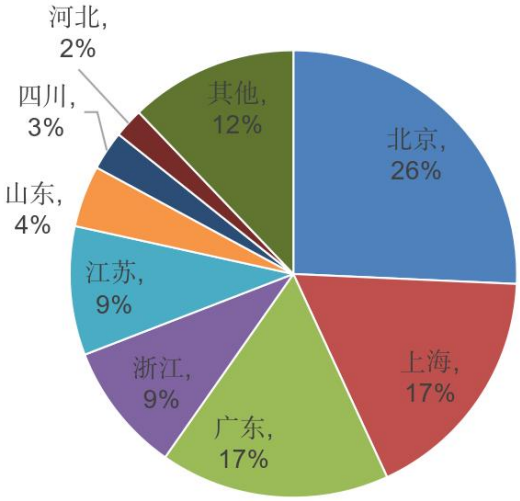
从地域分布看，北京市完成备案大模型数量遥遥领先，总计 254 款，占比 26%；上海市和广东省紧随其后，分别备案 172 款和 164 款，占比均为 17%，三地合计占比约为六成，成为我国大模型发展的核心区域。浙江省和江苏省分别备案 93 款和 92 款，均占比 9%，构成第二梯队。京津冀、长三角、珠三角等经济发达地区在政策支持、技术资源和产业基础上具备明显优势，成为大模型创新和应用的高地。

图 77：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 78：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

1.5.5. 科技企业裁员：2026 年前 8 个月美国因为 AI 取代占总裁员的比例达到 22%

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**

仅供内部参考，请勿外传！

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

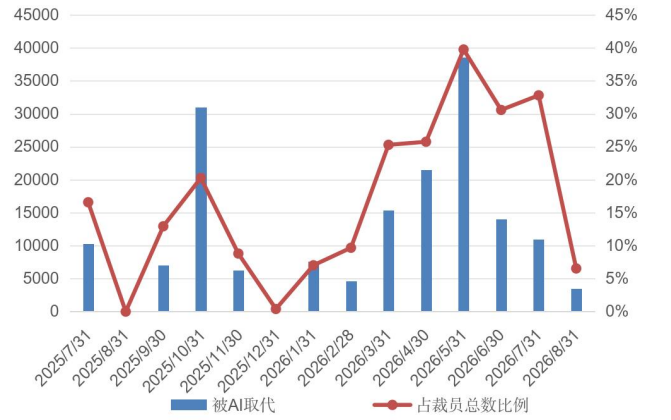
厂商	时间	人数	备注
亚马逊	2025.1	约 200	在北美商店部门裁员约 200 人，涵盖包括自有品牌、Prime 会员计划和消费品业务部门
	2025.10.28	约 14000	将裁减约 1.4 万个企业岗位，占其约 35 万名企业职员的 4%，此次裁员旨在提高决策效率，并将资源转向核心业务
	2026.1.28	约 16000	宣布调整组织架构，确认将裁减约 1.6 万个工作岗位。
	2026.4.7	约 14000	计划在 2026 年 5 月启动新一轮全球裁员，预计约 1.4 万名员工将受到影响。此次裁员覆盖 AWS 云服务、零售、人力资源等多个核心业务板块，中国区部分团队甚至可能被整体裁撤。
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。
	2026.1	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%。
	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。
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

2026 年前 8 个月美国因为被 AI 取代占总裁员比例的 22%，8 月 AI 替代导致的裁员有所缓解。Challenger, Gray & Christmas 的数据显示，2026 年前 8 个月，美国因 AI 被裁员的人数总计达到 11.6 万人，占总计裁员的 22%。其中 8 月因 AI 被裁员人数 3462 人，占总裁员人数的 7%，降幅较为明显。

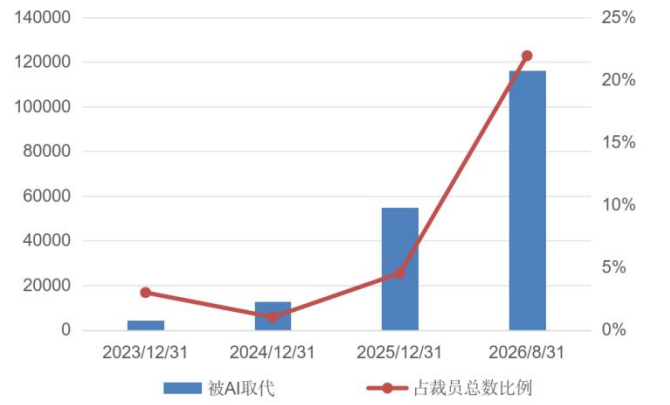
仅供内部参考，请勿外传！

图 79：2025.7-2026.8 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 80：2023-2026 年美国因 AI 被裁的人数及占裁员总数的比例（累计值，人）



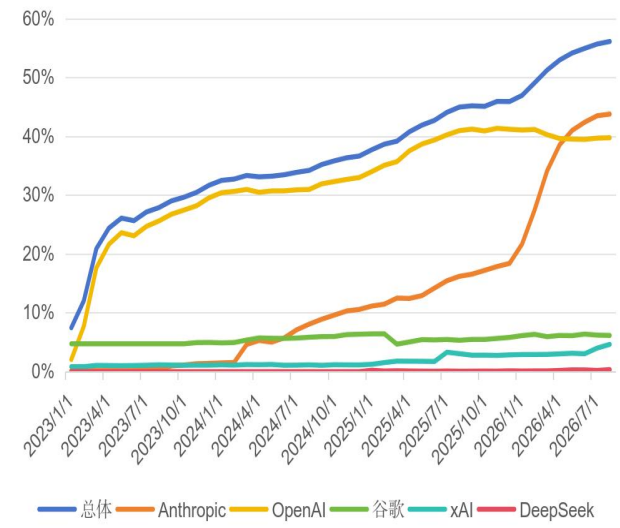
资料来源：Challenger, Gray & Christmas，中原证券研究所

1.5.6. 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 8 月的数据，Anthropic 以 43.8%的份额再次超过 OpenAI（39.8%），成为了企业市场最大的模型供应商。Anthropic 在金融、科技和专业服务等高技术行业中占据领先优势，而 OpenAI 仍在更广泛的企业群体中保持领先。

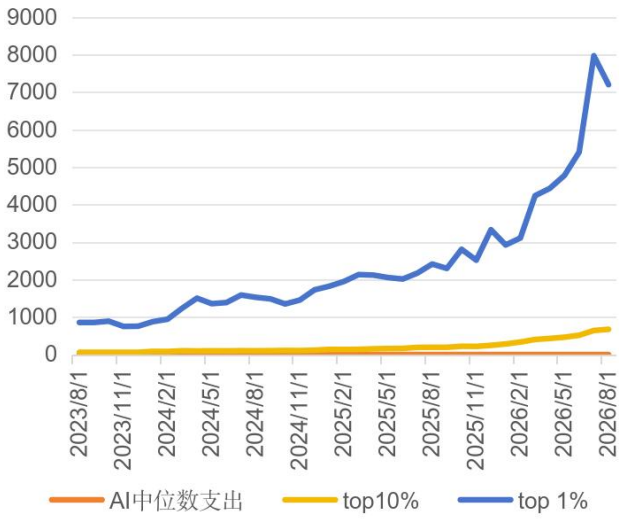
在 AI 用户的支出方面，top 1%的公司每个员工每月支出在 7 月快速上升 47.6%以后，8 月回落 9.7%至 7205 美元，我们认为可能的影响因素包括了头部模型的降价等，对整体行业收入增长带来了较大的负面影响；top 10%的公司每个员工每月支出金额为 675.6 美元，环比提升了 4.8%；AI 中位公司每个员工每月支出为 12.5 美元每人每月，环比增长 5.2%。

图 81：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

图 82：美国企业每个员工每月的 AI 支出（美元）

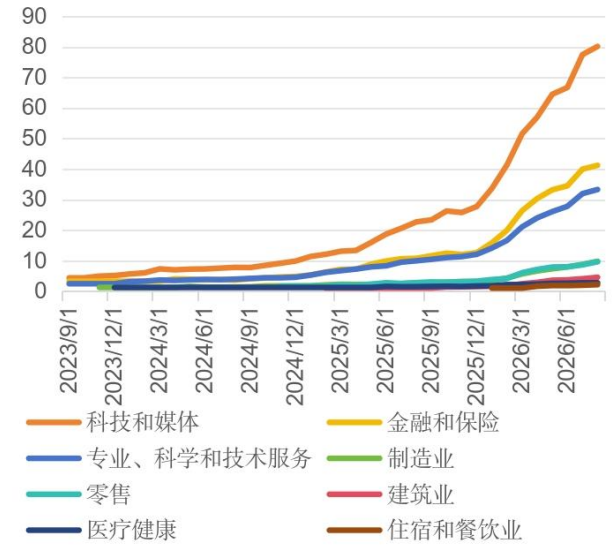


资料来源：Ramp，中原证券研究所

仅供内部参考，请勿外传！

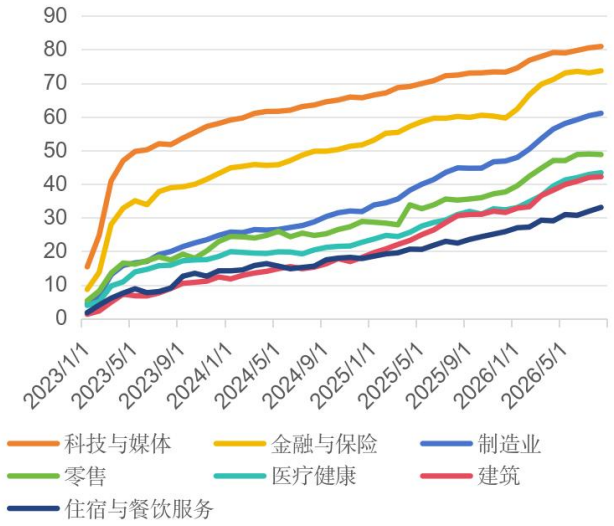
分行业来看，科技和媒体、金融和保险、制造业都已经有 50%以上的企业在进行 AI 相关投入。但是从金额上来看，科技和媒体、金融各行业员工的 AI 支出中位数明显领先其他行业，这与当前编程、视频领域 AI 的深度应用紧密相关；金融和保险行业无论还是普及度和员工 AI 开支都位居行业第二；专业、科学和技术服务在 AI 员工开支方面位居第三；制造业虽然 AI 普及率较高，但是人均贡献金额处于较低水平。

图 83：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 84：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

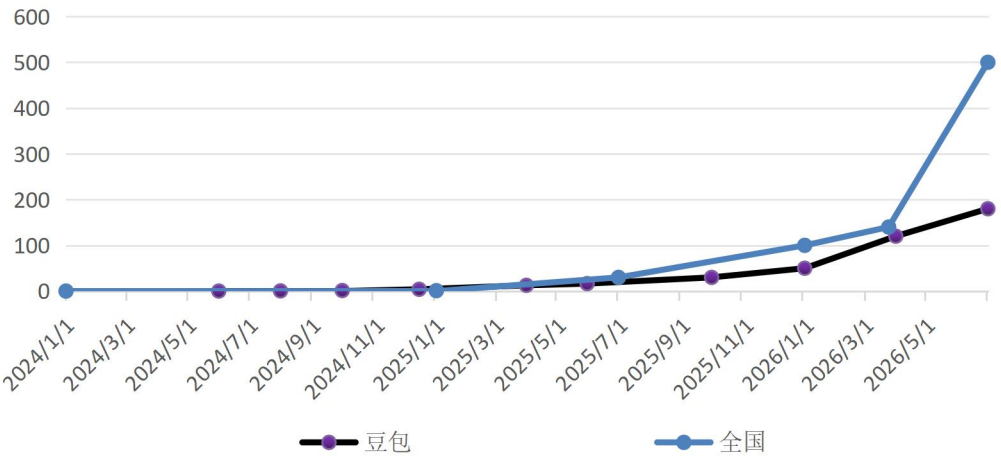


资料来源：Ramp，中原证券研究所

1.5.7. 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。

图 85：2024-2026 年中国我国日均 Token 调用量

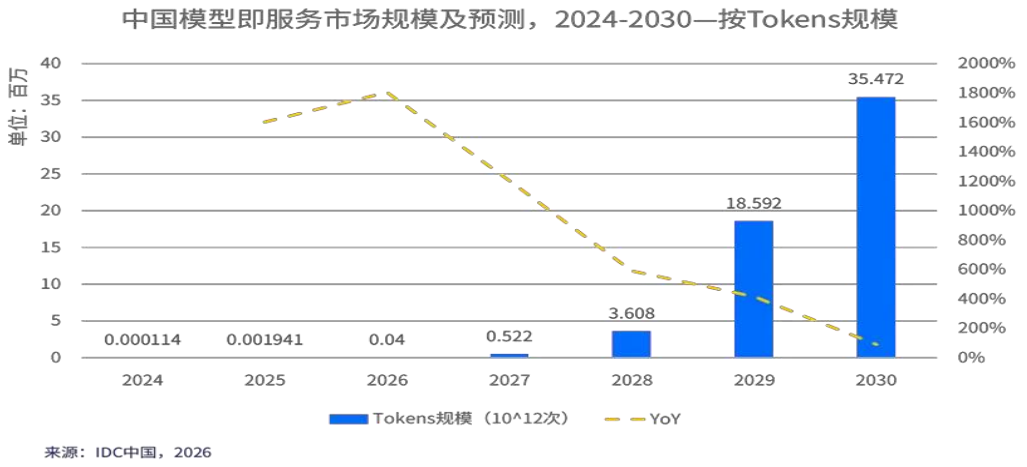


资料来源：国家数据局，工信部，字节跳动，中原证券研究所

仅供内部参考，请勿外传！

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 86：2024-2030 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

1-8 月，信息传输业投资增长 27.9%，计算机、通信和其他电子设备制造业投资下降 4.6%。

8 月，分产品看，智能手机产量增长 6.1%，计算机通信和其他电子设备制造业增加值增长 24.2%。在消费品以旧换新地方自主品类补贴活动有序实施带动下，限额以上单位计算机及其配套产品、可穿戴智能设备、智能手机类商品零售额分别增长 162.8%、70.5%、46.5%，分别比上月加快 63.8 个、46.1 个、6.1 个百分点。电子计算机整机 22.61 万台，同比下降 0.8%；智能手机 644.69 万台，同比增长 6.1%。

（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 9 月 28 日）河南 6 家计算机公司中 1 家上涨、5 家下跌。上涨公司为捷安高科 29.89%，继 8 约上涨 39.43%以后，其 9 月也获得了 6.41%的涨幅。其中跌幅增高的为*ST 宏图-74.54%，众诚科技-6.58%相对抗跌，恒拓开源-38.80%、新开普-34.32%、迈科技-36.75%跌幅都相对较高。

仅供内部参考，请勿外传！

图 87：河南上市公司近期股价涨跌幅表现（%）



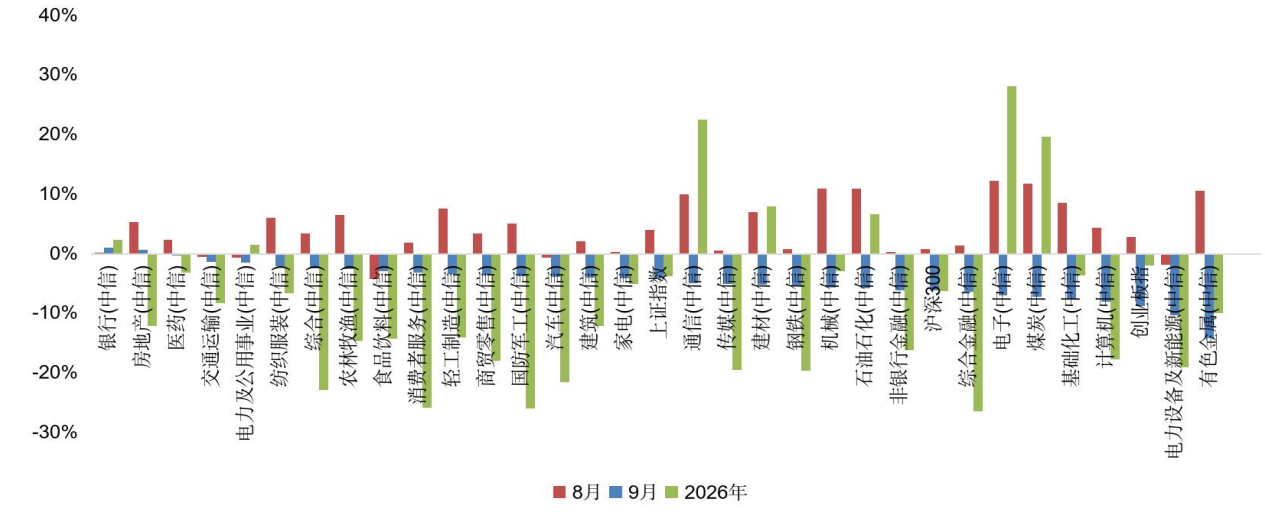
资料来源：Wind，中原证券研究所

3. 计算机行业行情表现

3.1. 行情回顾：9 月整体表现较弱

9 月，计算机行业整体表现较弱。9 月（截至 28 日）中信计算机指数下跌了 8.69%，跑输上证指数 3.98 PCT，跑输沪深 300 指数 1.91 PCT，跑赢创业板指数 0.63 PCT，在 30 个中信一级行业中排名第 28。

图 88：2026 年 9 月中信一级子行业涨跌幅（截至 2026 年 9 月 28 日）



资料来源：Wind，中原证券研究所

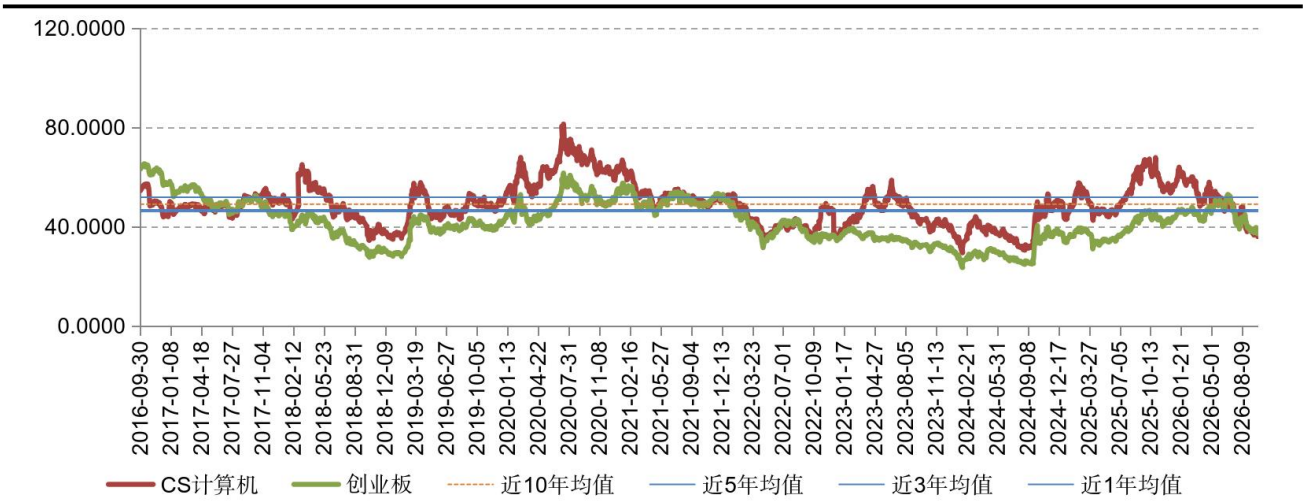
3.2. 估值：行业的估值低于历史均值水平

行业的估值低于历史均值水平。根据 Wind 数据，2026 年 9 月 29 日中信计算机行业 TTM 整体法（剔除负值）估值为 36.20 倍，创业板估值 37.39 倍。行业近 1 年、3 年、5 年、10 年

仅供内部参考，请勿外传！

的平均估值分别为 51.93 倍、46.84 倍、46.19 倍、49.12 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 89：近 10 年中信计算机行业估值水平（截至 2026.9.29）



资料来源：Wind，中原证券研究所

3.3. 计算机行业 ETF 行情回顾

9 月行业 ETF 系数下跌，其中大数据类 ETF 领跌，人工智能类 ETF 相对抗跌。

表 15：2026 年 9 月计算机行业 ETF 行情回顾（截至 2026.9.28）

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）
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	-5.39	0.05	62
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	-5.58	0.21	246
515070.SH	华夏中证人工智能 ETF	中证人工智能主题指数收益率	-5.76	0.03	2,120
159243.OF	招商创业板人工智能 ETF	创业板人工智能指数收益率	-5.95	0.30	817
512330.SH	南方中证 500 信息技术 ETF	中证 500 信息技术指数	-6.01	0.05	111
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	-6.07	0.07	20,929
159586.SZ	南方中证全指计算机 ETF	中证全指计算机指数收益率	-6.08	0.18	209
159382.SZ	南方创业板人工智能 ETF	创业板人工智能指数收益率	-6.11	0.14	3,687
159242.SZ	大成创业板人工智能 ETF	创业板人工智能指数收益率	-6.13	0.12	2,357
159279.SZ	华安创业板人工智能 ETF	创业板人工智能指数收益率	-6.15	0.09	990
159381.SZ	华夏创业板人工智能 ETF	创业板人工智能指数收益率	-6.17	0.07	5,775
159388.SZ	国泰创业板人工智能 ETF	创业板人工智能指数收益率	-6.18	0.07	2,459
159246.SZ	富国创业板人工智能 ETF	创业板人工智能指数收益率	-6.19	0.05	2,179
159363.SZ	华宝创业板人工智能 ETF	创业板人工智能指数收益率	-6.23	0.02	16,292
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	-6.36	0.06	537
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	-6.38	0.04	113
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	-6.39	0.04	3,721

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	-6.39	0.04	159
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	-6.39	0.03	178
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	-6.40	0.03	1,079
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	-6.40	0.02	2,376
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	-6.41	0.02	658
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	-6.45	-0.02	153
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	-6.51	-0.08	970
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	-6.58	0.21	140
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	-6.61	0.06	3,140
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	-6.63	0.16	185
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	-6.64	0.08	191
159107.SZ	富国创业板软件 ETF	创业板软件指数收益率	-7.18	0.04	827
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	-7.63	0.00	207
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	-7.64	0.09	117
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	-7.67	-0.03	62
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	-7.68	0.05	396
159899.SZ	招商中证全指软件 ETF	中证全指软件指数收益率	-7.72	0.17	483
560360.SH	万家中证软件服务 ETF	中证软件服务指数收益率	-7.75	0.27	89
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	-7.80	-0.07	274
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	-7.81	-0.15	3,683
515230.SH	国泰中证全指软件 ETF	中证全指软件指数收益率	-7.85	0.05	3,077
159939.SZ	广发中证全指信息技术 ETF	中证全指信息技术指数	-7.85	0.20	557
159852.SZ	嘉实中证软件服务 ETF	中证软件服务指数收益率	-8.00	0.01	6,317
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	-8.05	0.00	19
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	-8.26	0.25	110
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	-8.29	0.05	61
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	-8.30	0.04	69
560660.SH	新华中证云计算 50ETF	中证云计算 50 指数收益率	-8.34	0.17	83
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	-8.44	0.07	17
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	-8.61	0.30	326
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	-8.65	0.26	179
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	-8.74	0.16	122
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	-8.76	0.15	647
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	-8.77	0.13	2,285
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	-8.78	0.12	27
516000.SH	华夏中证大数据产业 ETF	中证大数据产业指数收益率	-8.79	0.22	542
515400.SH	富国中证大数据产业 ETF	中证大数据产业指数收益率	-8.91	0.11	1,384
516700.SH	华宝中证大数据产业 ETF	中证大数据产业指数收益率	-8.96	0.06	167

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

9 月计算机行业可转债以下跌为主，其中仅有宏图转债上涨（22.07%），跌幅较大的包括

仅供内部参考，请勿外传！

普联转债-6.97%、奥飞转债-9.59%。

表 16：2026 年 9 月计算机行业可转债行情回顾（截至 2026.9.28）

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
118027.SH	宏图转债	22.0741	688066.SH	*ST 航图	2.8659	2028/11/28	A
118007.SH	山石转债	-1.0633	688030.SH	山石网科	23.0328	2028/03/22	A+
118052.SH	浩瀚转债	-1.4901	688292.SH	浩瀚深度	52.1605	2031/03/13	A+
123173.SZ	恒锋转债	-1.8342	300605.SZ	恒锋信息	30.6142	2028/12/30	A
123138.SZ	丝路转债	-2.8438	300556.SZ	丝路视觉	108.4189	2028/03/02	A+
123104.SZ	卫宁转债	-3.4475	300253.SZ	卫宁健康	26.2552	2027/03/16	AA
123196.SZ	正元转 02	-4.2734	300645.SZ	正元智慧	95.5432	2029/04/18	A+
123157.SZ	科蓝转债	-6.0515	300663.SZ	科蓝软件	54.8160	2028/08/30	AA-
127104.SZ	姚记转债	-6.4248	002605.SZ	姚记科技	60.3883	2030/01/25	A+
123261.SZ	普联转债	-6.9719	300996.SZ	普联软件	64.4176	2031/12/05	A+
123131.SZ	奥飞转债	-9.5908	300738.SZ	奥飞数据	37.4330	2027/12/03	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：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%。在 9 月第二批 AI 训练推理芯片的安全可靠评测结果发布后，华为、百度、阿里、壁仞、海光、天数、沐曦、摩尔线程 8 家芯片通过了测评，其中华为昇腾 950 成为首个 II 级安全可靠 AI 芯片。鸿蒙 7 在 9 月底开启正式版升级，鸿蒙终端数量突破 8500 万，2026Q2 在国内手机操作系统市场份额 24%，较 Q1 上升 5 PCT。

（2）算力：排名前 10 的智算企业及机构智算总规模达到 1266EFLOPS，占全国总规模的 57.9%，呈现高度集中化特征。根据 SemiAnalysis 数据，到 2026 年底，国内头部厂商的算力约为字节 5.3GW、阿里 3.9GW、腾讯 2.4GW、百度约 1.1GW、华为约 1.0GW。阿里宣布了 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。

（3）AI：由于当前 Anthropic 和 OpenAI 刚进行了一轮更新，国内模型由于发布节奏问题，与头部差距有短暂拉大。Anthropic 的 Opus 5.5、OpenAI 的 GPT-6 Astra、Meta 的 Muse Spark

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

仅供内部参考，请勿外传！

1.3、小米的 MiMo-V2.6-Pro、阿里的 Qwen3.8-Max 都有了不错的表现。头部模型降价趋势明显，9 月发布的 Opus 5.5 又在 Opus 5 基础上再降价 20%，OpenAI 对 9 月新推出的 GPT-6 Sol 和 Luna 再次降价约 50%。最新的 OpenRouter 的模型调用周数据显示，中国模型厂商份额达到 45%，连续两周领先较美国模型厂商（份额 41%）。

给予行业强于大市的投资评级。从当前国内 AI 产业发展来看，资本开支加大的趋势明显。国内虽然算力短缺，但是模型在综合能力和性价比的比拼当中成为了更多用户的选择，调用份额持续提升。以 WorkBuddy 为代表的 AI 办公产品，正在加速 AI 在 B 端的普及进程。因而在这一趋势下，包括字节、阿里在内的头部厂商的投资步伐加快明显，新一代国产芯片上市时间节点也大幅前置。**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）、东阳光（600673）。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、Kimi、长存控股。**

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	2026.2.11	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。
	Kimi	2026.9.3	已经保密形式向港交所递交 IPO 文件
EDA	芯和半导体	2025.12.24	IPO 的辅导工作完成
	全芯制造	20256.4.10	IPO 的辅导工作完成
	合见工软	2026.7.1	IPO 辅导工作完成

资料来源：澎湃，Wind，中原证券研究所

5. 风险提示

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

仅供内部参考，请勿外传！

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。

仅供内部参考，请勿外传！

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。