

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

海外 AI 投入加大，国产基础软硬件加速发展

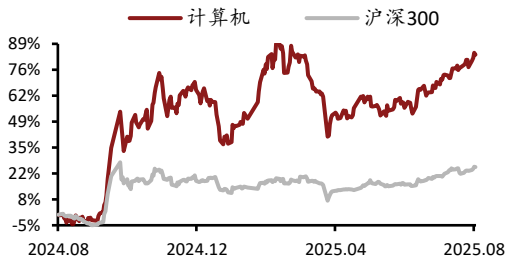
——计算机行业月报

证券研究报告-行业月报

强于大市(维持)

计算机相对沪深 300 指数表现

发布日期：2025 年 08 月 15 日



资料来源：中原证券研究所，聚源

相关报告

《计算机行业月报：EDA、H2O 禁令接连解除，鸿蒙电脑销售良好》 2025-07-18

《计算机行业半年度策略：持续掘金国产化、AI、智算三大赛道》 2025-06-20

《计算机行业月报：国产基础软硬件持续突破，算力互联互通将实现算力网络升级》 2025-06-17

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：

- **2025 年 1-6 月软件产业增速持续上升。**2025 年 1-6 月我国软件业务收入 7.06 万亿元，增长 11.9%，较 1-5 月回升了 0.7 PCT。行业利润总额 8581 亿元，增长 12.0%，较 1-5 月下降了 0.8 PCT。

- **重点关注子行业的主要数据和动态包括：**

(1) **AI：**8 月发布的 GPT-5 性能虽然有较大提升，价格也下降，但是并不是期待中有“代差”的产品，总体领先竞争对手，但是评分差距有限。7 月阿里开源了升级版模型 Qwen3-235B-A22B-Thinking-2507，成为了全球最强大的开源模型，整体实力居全球大模型第一梯队。2025 年 7 月大模型相关中标项目数量为 574 个，中标项目披露的金额约为 13.35 亿元，同比分别增长 422%和 540%。Q2 财报来看，AI 在广告、云业务等方向上给科技大商带来了较大的经济效益。随着模型能力提升面临发展瓶颈，我们预计后续大厂更多注意力聚焦在 AI 应用落地和算力成本控制方面。

(2) **国产化：**2025 年 1-6 月基础软件增长 13.8%，增速连续 4 个月呈现回升趋势，高于软件行业整体增速 1.9 PCT，下半年或将继续受益于“十四五”收官之年的需求释放，增速有进一步加快的可能性。H2O 解禁后，仍面临了较多政策不确定性，于此同时，国产 AI 芯片厂商 2025 上半年整体呈现出加速发展趋势，CANN 的开源有助于加快国产 AI 芯片的生态建设。

(3) **算力：**2025Q2，美国 6 大科技厂商（亚马逊、微软、谷歌、脸书、甲骨文、苹果）的资本开支再创新高，总计达到 999.73 亿美元，同比增长 77%，增速再创新高，同时海外大型科技厂商的资本开支计划呈现上调趋势。在 AI 业务的带动下，Q2 全球云巨头都呈现了业务加速趋势，甲骨文云基础设施业务预计 2026 财年营收增长超过 70%，即将借力 AI 成为全球第四大云基础设施服务厂商。

- **给予行业强于大市的投资评级。**在 AI 助力广告、云等业务发展的同时，海外 AI 投入加大趋势明显。随着头部大模型性能提升面临瓶颈，后续大厂有望投入跟多关注度在 AI 应用和算力成本控制方面，液冷落地有望加快。在国际局势的不确定下，国产软硬件成为了更优的选项。继续看好行业的市场表现，建议积极关注 EDA 企业华大九天（301269），有大规模智算中心交付计划的润泽科技（300442），国产算力下布局的大模型厂商科大讯飞（002230）。

风险提示：国际局势的不确定性；下游企业削减开支，半年报业绩的不确定性。

内容目录

1. 行业数据	5
1.1. 行业数据：2025 年 1-6 月软件行业收入增速继续回升	5
1.2. 2025 年 1-6 月高景气赛道为 IC 设计、基础软件、云+大数据服务	6
1.3. AI：GPT-5 未实现代际突破，继续关注 R2 进展	8
1.3.1. 头部大模型最新进展	8
1.3.2. OpenAI：GPT-5 发布，未实现代际突破	9
1.3.3. DeepSeek：继续关注 R2 的进展	15
1.3.4. 阿里：新模型位居开源模型第一	16
1.3.5. Meta 开启 AI 人才争夺战，科技企业加快业务调整步伐	17
1.3.6. 中标情况：AI 应用加快，行业端呈现百花齐放态势	17
1.3.7. AI 在广告、云业务等方向上给科技大商带来了较大的经济效益	19
1.4. 国产化：鸿蒙、CloudMatrix384 超节点双重突破，EDA 禁令加速国产进程	20
1.4.1. 国产算力芯片呈加速趋势，AI 芯片企业步入集中上市期	20
1.4.2. H20 供应仍存在较大不确定性，但是下游厂商在推理测已经有了更多选择	23
1.4.3. 华为开源 CANN，加速 AI 生态建设	25
1.4.4. EDA：禁令暂时解除，但国产替代仍是大势所趋	25
1.4.5. 鸿蒙电脑操作系统适配应用或超 2500 款，C 端销售情况良好	27
1.5. 算力：海外科技巨头加大资本投入，液冷加速落地中	28
1.5.1. 除英伟达以外的海外算力芯片发展格局	28
1.5.2. 2025 年大规模智算中心或将呈现较为明显的集中交付趋势	30
1.5.3. 海外科技巨头资本开支明显增大	32
1.5.4. 云计算：在 AI 业务的带动下，全球云计算业务整体呈现出加速趋势	35
1.5.5. 液冷：微软所有区域都可以支持液冷，超节点全面拥抱液冷，行业加速趋势明显	37
2. 河南计算机行业动态	39
2.1. 河南计算机行业要闻	39
2.2. 河南计算机行业数据跟踪	40
2.3. 河南上市公司行情回顾	40
3. 投资策略	41
3.1. 行情回顾：7 月计算机行业连续两个月上涨	41
3.2. 估值：行业的估值超过历史均值水平	41
3.3. 行业观点与投资建议	42
4. 风险提示	43

图表目录

图 1：2019-2025 年我国软件业务收入及增速（月度累计值）	5
图 2：2019-2025 年我国软件业务利润总额及增速（月度累计值）	5
图 3：2019-2025 年我国软件业务出口数据	6
图 4：2019-2025 集成电路设计服务收入及增速（亿元）	6
图 5：2020-2025 云服务大数据服务收入及增速（亿元）	6
图 6：2022-2025 基础软件收入及增速（亿元）	7
图 7：2019-2025 工业软件收入及增速（亿元）	7
图 8：2019-2025 信息安全收入及增速（亿元）	7

图 9: 2020-2025 嵌入式系统软件收入及增速 (亿元)	7
图 10: 2025 前 6 个月与上年同期软件子行业增速对比	8
图 11: 2025 前 6 个月我国软件业务收入结构	8
图 12: 2025 前 6 个月信息技术服务收入结构	8
图 13: 2025 前 6 个月软件产品收入结构	8
图 14: Artificial Analysis 基准测试	9
图 15: Artificial Analysis 基准测试及 GPT-5 的表现	11
图 16: GPT-5 在 AIME2025 中的成绩	11
图 17: Grok 4 在 AIME2025 中的成绩	11
图 18: GPT-5 在 SWE-bench Verified 中的成绩	12
图 19: Grok 4 在 Aider Polyglot 中的成绩	12
图 20: GPT-5 在 MMMU 中的成绩	12
图 21: GPT-5 在健康领域的成绩	12
图 22: GPT-5 在 GPQA 中的成绩	13
图 23: Grok 4 在 GPQA 中的成绩	13
图 24: GPT-5 在 HLE 中的成绩	13
图 25: Grok 4 在 HLE 中的成绩	13
图 26: GPT-5 的错误信息率	14
图 27: GPT-5 与 o3 在视觉推理方面的能效对比	14
图 28: GPT-5 与 o3 在智能体编程方面的能效对比	14
图 29: GPT-5 与 o3 在 GPQA 方面的能效对比	14
图 30: 主流模型 API 调用价格	15
图 31: 阿里 Qwen3-235B-A22B-Thinking-2507 各项模型能力测试得分	16
图 32: 2024-2025 年大模型单月份中标项目监测数据 (万元)	18
图 33: 2025 年一季度大模型项目类型分布 (按金额)	18
图 34: 2025 年 7 月大模型项目类型分布 (按金额)	18
图 35: 2025 年 7 月大模型订单行业分布 (按数量)	19
图 36: 2025 年 7 月大模型订单行业分布 (按金额)	19
图 37: 2025Q1 通用大模型厂商中标情况	19
图 38: 2025Q2 通用大模型厂商中标情况	19
图 39: 豆包大模型调用量最新情况	20
图 40: 2021-2025 年我国集成电路对进口依赖度和出口占比 (按数量)	21
图 41: 2022Q1 至 2025Q1 英伟达分区域收入 (亿美元) 及占比	21
图 42: 2023H1-2024H2 我国 AI 芯片出货情况及国产化比率	22
图 43: 2023Q1-2025Q2 寒武纪、海光收入 (亿元) 及增速对比	22
图 44: 2017-2024 国内主要上市 AI 芯片企业收入及增速对比	23
图 45: 2020Q1-2025Q1 紫光和浪潮存货增长情况 (百万元)	24
图 46: 2021Q1-2025Q2 美国 3 大芯片厂商数据中心业务收入及增速 (亿美元)	24
图 47: 2021 年国内 EDA 市场格局	26
图 48: 2018-2025 年 EDA 企业国内业务收入及增速 (万元)	26
图 49: 2022Q2-2024Q4 中国手机操作系统市场份额	27
图 50: 2022Q2-2024Q4 全球手机操作系统市场份额	27
图 51: 鸿蒙电脑应用适配情况	27
图 52: 2021Q1-2025Q2 AMD 数据中心业务收入及增速 (亿美元)	28
图 53: 2015-2023 数据中心加速器年度出货量 (片)	29
图 54: 2023Q4-2025Q3E 博通 AI 业务收入 (亿美元)	30

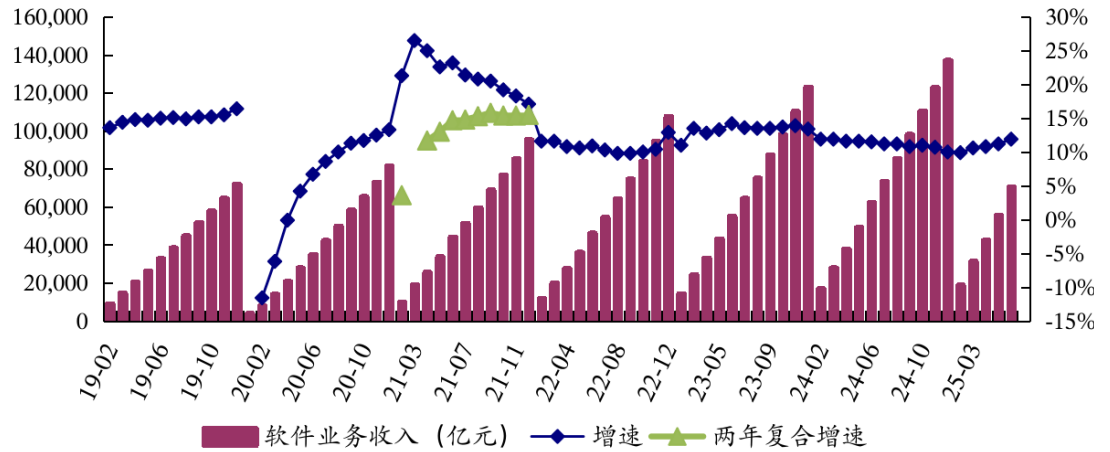
图 55: 23Q1-25Q1 Marvell 数据中心业务收入 (亿美元)	30
图 56: 2024 年润泽科技收入分布	30
图 57: 2021-2024 年润泽科技分项业务收入 (百万元)	30
图 58: 6 大科技厂商资本投入 (百万美元)	32
图 59: 6 大科技厂商资本投入增速	32
图 60: 2020-2025 年阿里单季度资本开支及增速	34
图 61: 2020-2025 年腾讯单季度资本开支及增速 (亿元)	34
图 62: 2016-2025 年我国及全球云基础设施服务的支出规模及增速 (亿美元)	35
图 63: 2025Q1 我国云基础设施服务市场结构	35
图 64: 2019Q4-2025Q1 阿里云收入及其增速	36
图 65: 2017Q1-2025Q1 金山云收入及其增速	36
图 66: 2019Q1-2025Q2 海外云巨头的相关业务收入及增速	36
图 67: 甲骨文 2023-2026 财年云基础设施业务收入规模 (亿美元)	37
图 68: 2022-2024 年 智算服务细分市场 (亿元)	38
图 69: 不同冷却技术与 PUE 相关性	38
图 70: 河南上市公司近期股价涨跌幅表现 (%)	40
图 71: 2025 年 7 月中信一级子行业涨跌幅	41
图 72: 近 10 年中信计算机行业估值水平 (截至 2025.8.14)	42
表 1: 2025 年主要大模型发布情况	8
表 2: OpenAI 大模型产品	10
表 3: DeepSeek 的主要模型发布情况	15
表 4: 阿里的主要模型发布情况	16
表 5: 主要科技厂商近期裁员情况	17
表 6: 2025 年 AI 芯片相关动态	23
表 7: 华为在 AI 训练各层的产品布局	25
表 8: 2025Q1 企业 IDC 相关业务对比 (亿元)	31
表 9: IDC 企业订单获取和运力投放情况	31
表 10: 大模型厂商算力相关投入	33
表 11: 液冷产业链公司业务动态	38

1. 行业数据

1.1. 行业数据：2025 年 1-6 月软件行业收入增速继续回升

2025 年 1-6 月软件产业增速持续上升。根据工信部数据，2025 年 1-6 月软件业务收入 7.06 万亿元，同比增长 11.9%，较 1-5 月回升了 0.7 PCT。

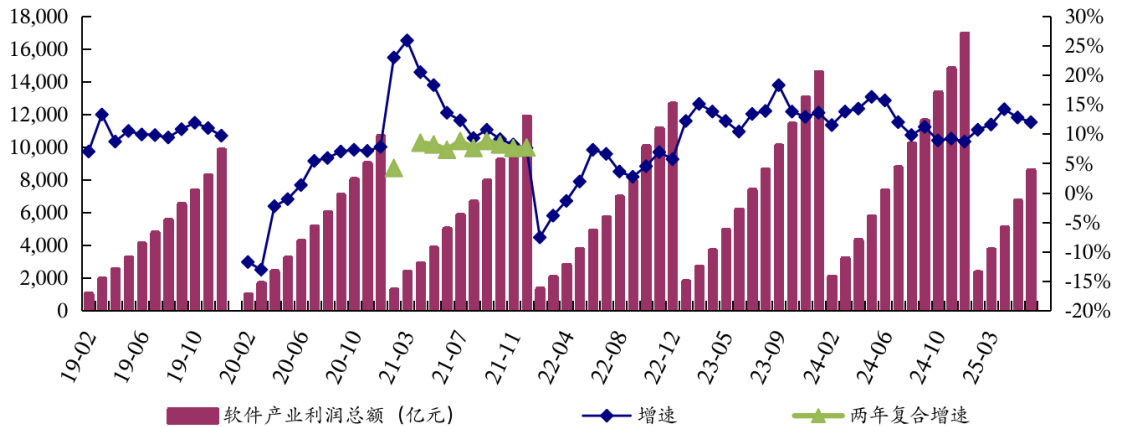
图 1：2019-2025 年我国软件业务收入及增速（月度累计值）



资料来源：工信部，中原证券研究所

利润方面，2025 年 1-6 月软件行业利润总额增速连续两个月放缓。2025 年 1-6 月软件业务利润总额 8581 亿元，同比增长 12.0%，较 1-5 月下降了 0.8 PCT，高于收入增速 0.1 PCT。

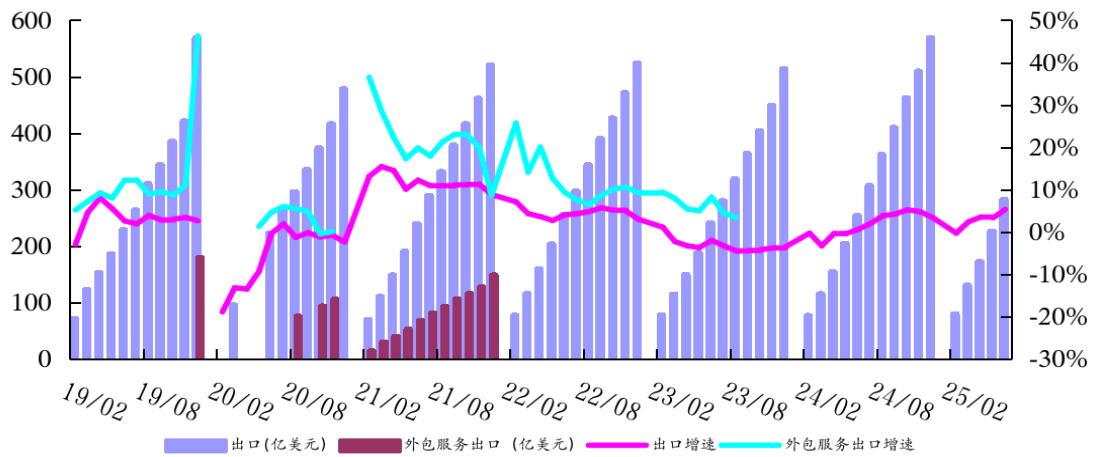
图 2：2019-2025 年我国软件业务利润总额及增速（月度累计值）



资料来源：工信部，中原证券研究所

2025 年 1-6 月软件出口金额增速呈现回升趋势。2025 年 1-6 月软件业务出口金额 283 亿美元，同比增长 5.3%，为 2022 年 1-9 月以来的最高值，较 2024 年增长 1.8 PCT，软件行业出口金额约占行业收入的 2.9%。

图 3：2019-2025 年我国软件业务出口数据



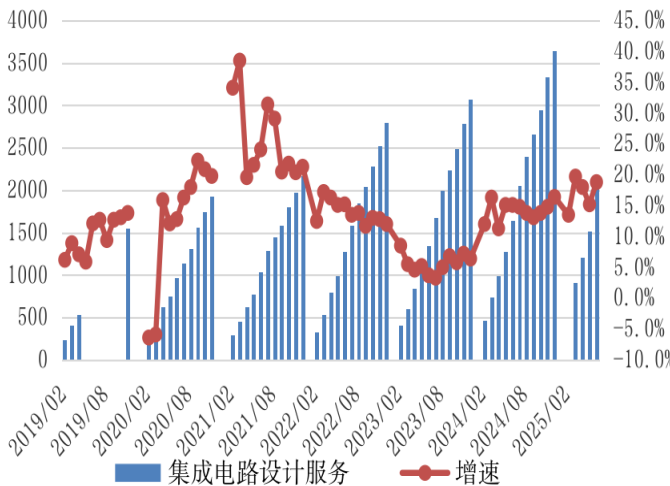
资料来源：工信部，中原证券研究所

1.2. 2025 年 1-6 月高景气赛道为 IC 设计、基础软件、云+大数据服务

IC 设计：作为 2025 年 1-6 月的软件行业景气度最高的子行业，同比增长 18.8%，较 2024 年上升了 2.4 PCT，较上年同期增长 3.7 PCT，高于软件行业整体增速 6.9 PCT。

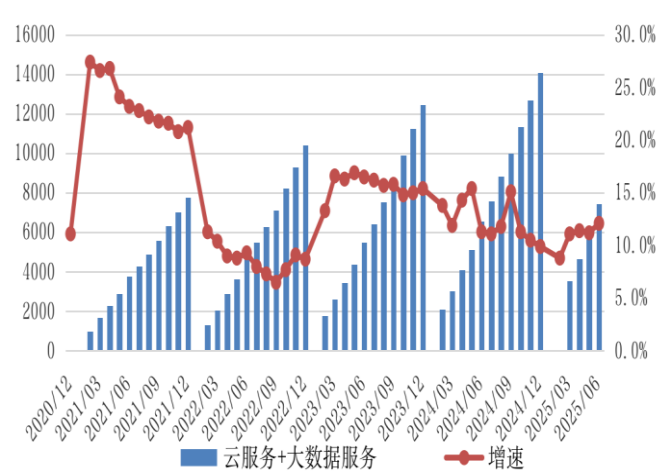
云服务和大数据服务：2025 年 1-6 月同比增长 12.1%，较 2024 年增长了 2.2 PCT，高于软件行业整体增速 0.2 PCT，或受到了 AI 应用的带动作用，呈现增长加快趋势。

图 4：2019-2025 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 5：2020-2025 云服务大数据服务收入及增速（亿元）

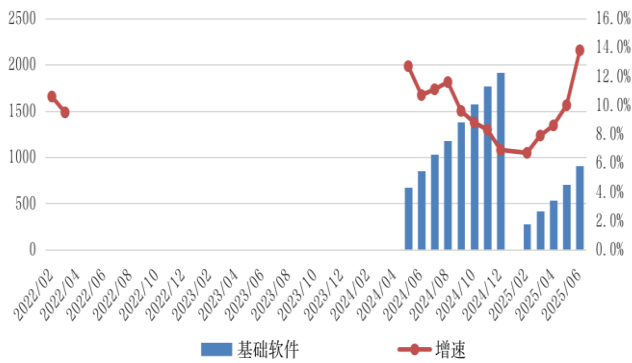


资料来源：工信部，中原证券研究所

基础软件：2025 年 1-6 月同比增长 13.8%，较 1-5 月增长 3.8 PCT，明显加快，连续 4 个月呈现回升趋势，高于软件行业整体增速 1.9 PCT，我们认为**基础软件下半年将继续受益于“十四五”收关之年的需求释放，增速有进一步加快的可能性。**

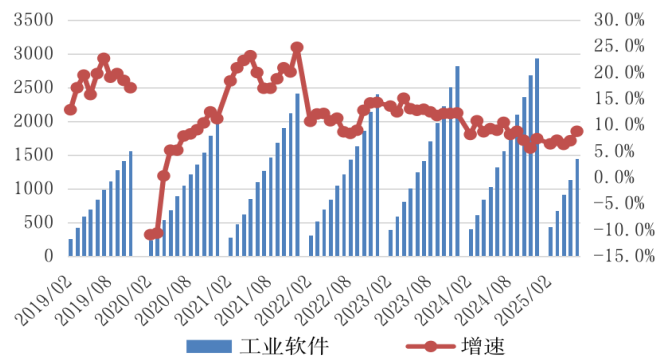
工业软件产品：2025 年 1-6 月同比增长 8.8%，较上年同期下滑了 0.2 PCT，景气度偏弱。

图 6：2022-2025 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 7：2019-2025 工业软件收入及增速（亿元）

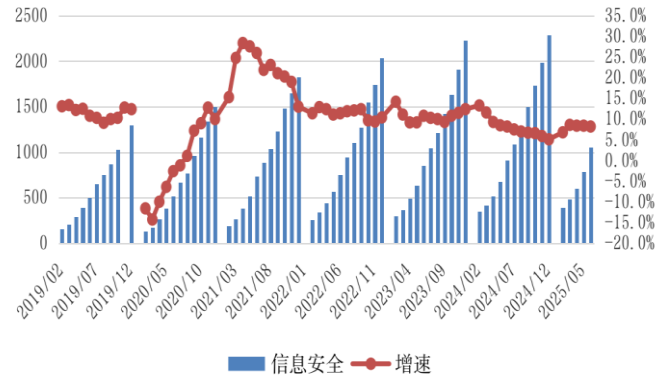


资料来源：工信部，中原证券研究所

信息安全：2025 年 1-6 月同比增长 8.2%，较 2024 年回升 3.1 PCT，景气度 2024 年见底后略有回升。

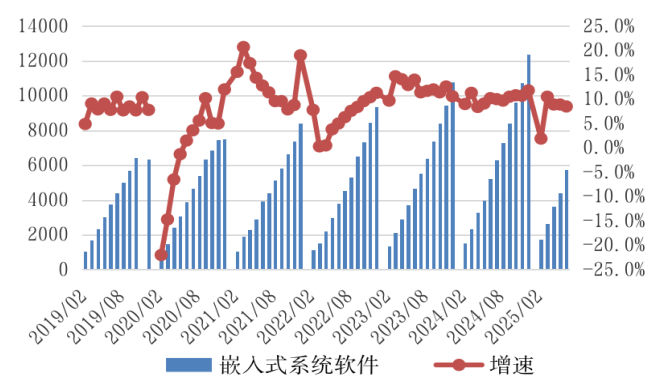
嵌入式系统软件：2025 年 1-6 月同比增长 8.5%，较 2024 年下降 3.3 PCT，处于相对平稳的增速水平。

图 8：2019-2025 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

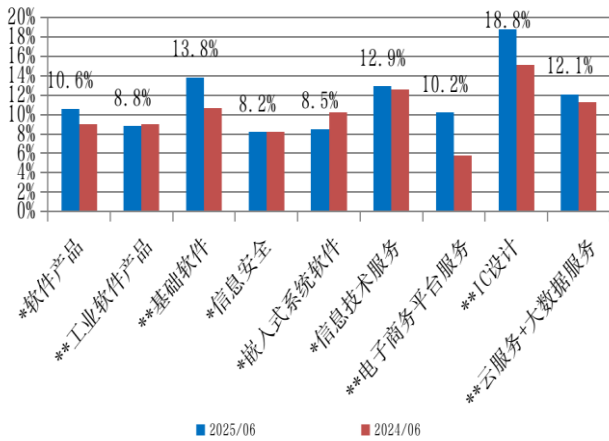
图 9：2020-2025 嵌入式系统软件收入及增速（亿元）



资料来源：工信部，中原证券研究所

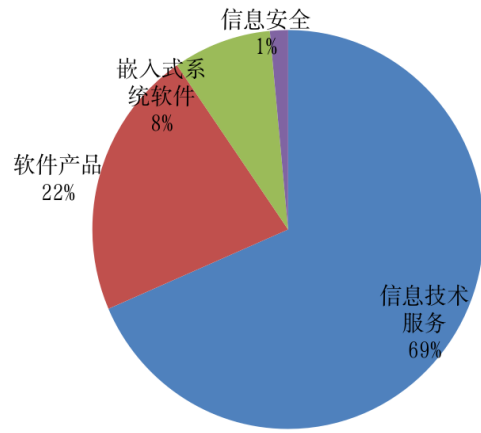
软件业务中，信息技术服务收入占比持续提升。2025 年 1-6 月信息技术服务收入增速 12.9%，高于软件业务整体增速 1.0 PCT，占比软件业务整体收入比重进一步提升到了 68.5%，较 2024 年提升了 1.2 PCT。

图 10：2025 前 6 个月与上年同期软件子行业增速对比



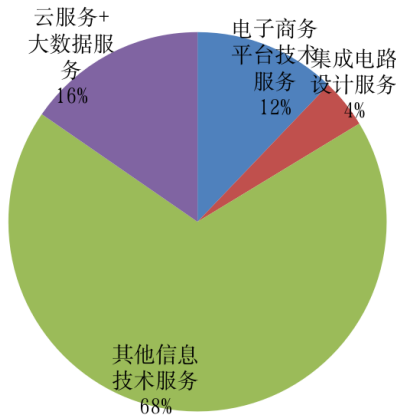
资料来源：工信部，中原证券研究所（*为一级子行业，**为二级子行业）

图 11：2025 前 6 个月我国软件业务收入结构



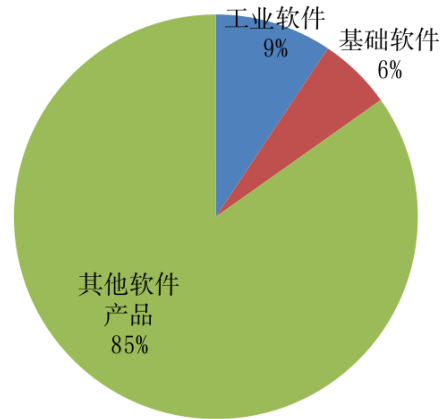
资料来源：工信部，中原证券研究所

图 12：2025 前 6 个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 13：2025 前 6 个月软件产品收入结构



资料来源：工信部，中原证券研究所

1.3. AI: GPT-5 未实现代际突破，继续关注 R2 进展

1.3.1. 头部大模型最新进展

当前处于第一梯队的大模型分别为美国的 OpenAI 的 GPT-5 系列、xAI 的 Grok 4、谷歌的 Gemini 2.5 Pro、Anthropic 的 Claude Opus 4.1，中国的 DeepSeek 的 DeepSeek-R1-0528、阿里的 Qwen3-235B、月之暗面的 Kimi-K2。

表 1：2025 年主要大模型发布情况

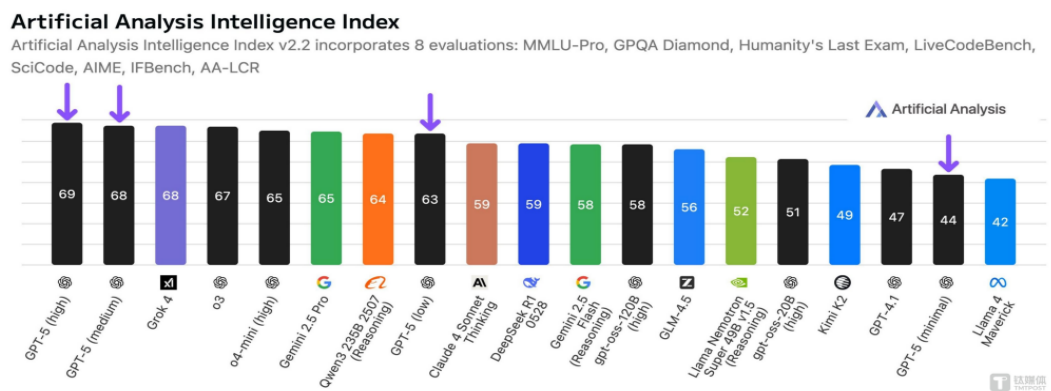
时间	模型	厂商	架构	参数	模型类型	备注
2025. 1. 20	DeepSeek-R1	DeepSeek	MoE	6710 亿	推理	开源，性能比肩 OpenAI o1 正式版，价格约为其 1/30，并开放思维链输出功能，64K 上下文
2025. 1. 31	o3-mini	OpenAI			推理	o1 的下一代产品
2025. 2. 18	Grok-3	xAI				

2025. 2. 24	Claude 3.7 Sonnet	Anthropic				
2025. 2. 28	GPT-4.5	OpenAI			语言	较 4o 有更高情商。2025 年 7 月 API 服务下线（成本过高）
2025. 3. 6	QwQ-32B	阿里		320 亿		开源
2025. 3. 24	DeepSeek-V3-0324	DeepSeek	MoE	6600 亿	语言	开源，性能超过 GPT-4.5
2025. 3. 25	Gemini 2.5 Pro Exp	谷歌				
2025. 4. 17	o3	OpenAI			推理	实现用图像思考
2025. 4. 29	Qwen3-235B-A22B	阿里		2350 亿		开源
2025. 5. 6	Gemini 2.5 Pro Preview	谷歌				
2025. 5. 23	Claude 4	Anthropic				
2025. 5. 28	DeepSeek-R1-0528	DeepSeek	MoE	6850 亿	推理	开源，128K 上下文
2025. 6. 5	Gemini 2.5 Pro Preview 0605 Thinking	谷歌				
2025. 7. 10	Grok-4	xAI				
2025. 7. 11	Kimi-K2	月之暗面		1 万亿		开源
2025. 7. 25	Qwen3-235B-A22B-Thinking-2507	阿里		2350 亿	推理	开源，256k 上下文
2025. 8. 6	Claude Opus 4.1	Anthropic				Opus 4 的一次小升级
2025. 8. 8	GPT-5	OpenAI				命名不再区分基础及推理模型

资料来源：DeepSeek, OpenAI, 阿里, 谷歌, Anthropic, 月之暗面, 中原证券研究所

可以看到美国的大模型企业主要以闭源路线为主，而中国的大模型厂商在积极拥抱开源。闭源模型虽然暂时领先，但是总体差距有限。在 Artificial Analysis 基准测试中，GPT-5 (high) 达到了 69 分，位列第一；GPT-5 (medium) 68 分，与 Grok 4 并列第二；谷歌 Gemini 2.5 Pro 65 分，阿里 Qwen 2.5 Pro 64 分，DeepSeek R1 和 Anthropic 的 Claude 4 Sonnet thinking 都是 59 分。

图 14: Artificial Analysis 基准测试



资料来源：钛媒体, 中原证券研究所

1.3.2. OpenAI: GPT-5 发布，未实现代际突破

2025 年 8 月 6 日，OpenAI 开源了 GPT-oss-120b 和 GPT-oss-20b 两款模型，是两款轻量模型，其核心产品仍然聚焦在闭源模型端。

2025年8月8日，OpenAI的GPT-5正式发布。原本OpenAI在开发一款内部代号为“Orion”的模型，将作为GPT-5发布，但Orion未能实现预期性能，最终在2025年2月作为GPT-4.5发布，同时由于成本过高，其API服务在2025年7月下线。整体来看，8月发布的GPT-5性能虽然有较大提升，价格也下降，但是并不是期待中有“代差”的产品，整体性能不及预期。

表 2: OpenAI 大模型产品

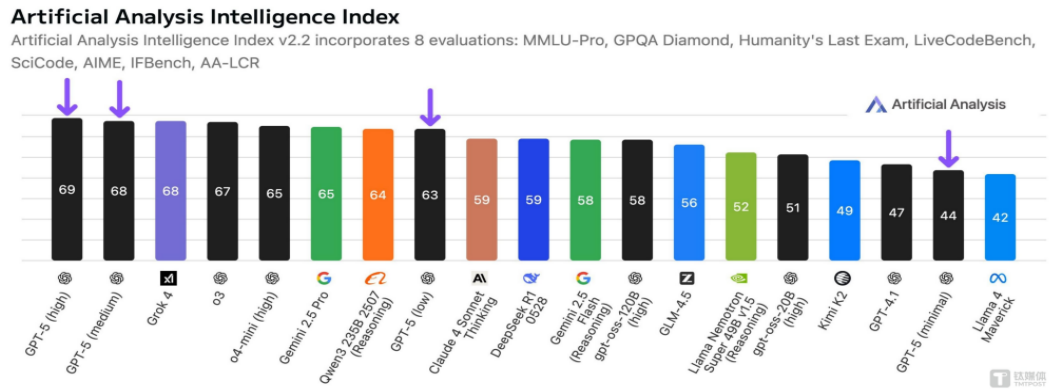
模型	时间	参数	上下文限制	训练数据截止时间	公开的程度	备注
GPT-1	2018.6	1.17 亿			开源（技术细节）	
GPT-2	2019.2	15 亿			开源（源代码）	
GPT-3	2020.5	1750 亿		2019.10	论文	
GPT-3.5	2022.3	-	4K/16K	2021.6		2022.11 ChatGPT 基于 GPT 3.5 推出
GPT-4	2023.3	1.8 万亿	8K/32K	2021.9	测试结果	2023.7 向所有开发者开放使用权限
GPT-4 Turbo	2023.11	-	128K	2023.4		速度是 GPT4 的 5 倍，价格降低 2/3
Sora	2024.2					首个文生视频模型
GPT-4o	2024.5					具有了强大的多模态交互能力
o1	2024.9					推理模型，更擅长解决科学、编码、数学和类似领域的复杂问题
o3-mini	2025.1.31					推理模型，o1 的下一代产品
GPT-4.5	2025.2.28					较 4o 有更高情商。2025 年 7 月 API 服务下线（成本过高）。
o3	2025.4.17					推理模型，实现用图像思考
GPT-5	2025.8.8					

资料来源：OpenAI，中原证券研究所

本次推出的 GPT-5 包括了 GPT-5-nano、GPT-5-mini、GPT-5 和 GPT-5-pro 4 个版本，其中 GPT-5-pro 仅供 Pro 订阅用户使用，免费用户、Plus 用户可以使用 GPT-5-mini、GPT-5，API 可以调用 GPT-5、GPT-5 mini、GPT-5 nano 3 个新模型。与此前模型命名方式不同，GPT-5 开始 OpenAI 将不再单独区分基础模型和推理模型，o 系列和 GPT 系列的突破被统一在了 GPT-5 中。

GPT-5 目前总体领先竞争对手，但是评分差距有限。在 Artificial Analysis 基准测试中，GPT-5 (high) 达到了 69 分，位列第一；GPT-5 (medium) 68 分，与 Grok 4 并列第二；谷歌 Gemini 2.5 Pro 65 分，阿里 Qwen 2.5Pro 64 分，DeepSeek R1 和 Anthropic 的 Claude 4 Sonnet thinking 都是 59 分。

图 15: Artificial Analysis 基准测试及 GPT-5 的表现

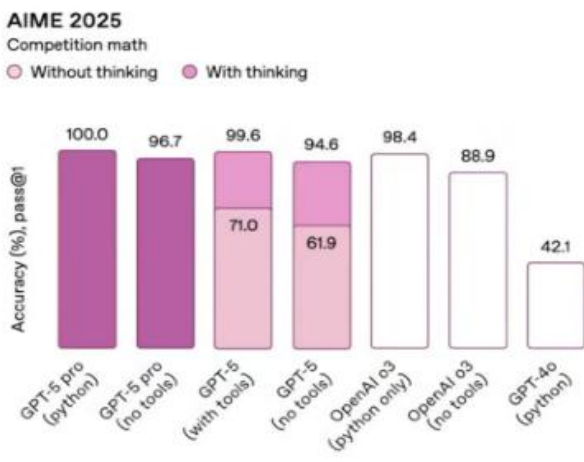


资料来源：钛媒体，中原证券研究所

从模型的各项基准测试成绩来看，GPT-5 的表现主要包括：

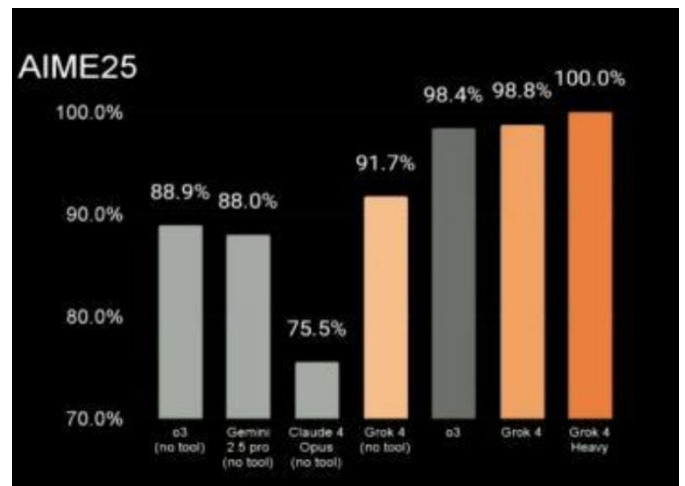
(1) 数学能力: GPT-5 Pro 在推理模式和调用工具的情况下，在美国数学邀请赛 AIME2025 上拿下了满分成绩，而 o3 同样在调用工具的基础上可以获得 98.4 的成绩，对比来看此前 Grok-4 Heavy 在该项目中同样取得了 100 的成绩，表面该指标在区分头部模型能力方面已经逐步失效，同时两者在 AIME 2025 中处于同一水平级别。

图 16: GPT-5 在 AIME2025 中的成绩



资料来源：OpenAI，量子位，中原证券研究所

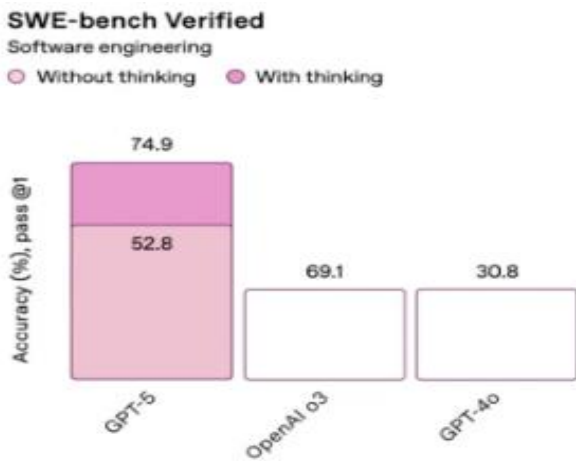
图 17: Grok 4 在 AIME2025 中的成绩



资料来源：xAI，中原证券研究所

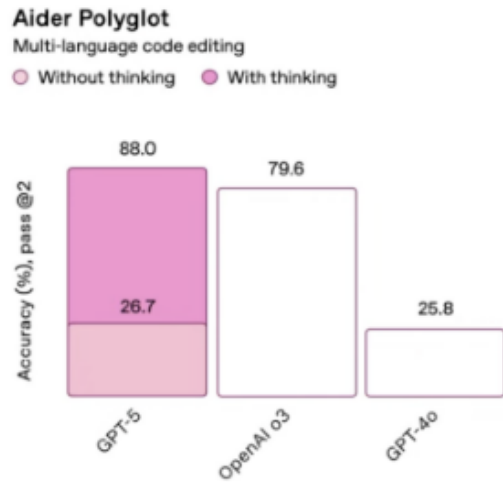
(2) 在编程能力方面: GPT-5 在 SWE-bench Verified 和 Aider Polyglot 中分别得到 74.9% 和 88% 的成绩，分别较 o3 69.1% 和 79.6% 有了一定的提升，较 Opus 4.1 在 SWE-bench Verified 中 74.5% 的成绩来说，优势有限。

图 18: GPT-5 在 SWE-bench Verified 中的成绩



资料来源: OpenAI, 量子位, 中原证券研究所

图 19: Grok 4 在 Aider Polyglot 中的成绩

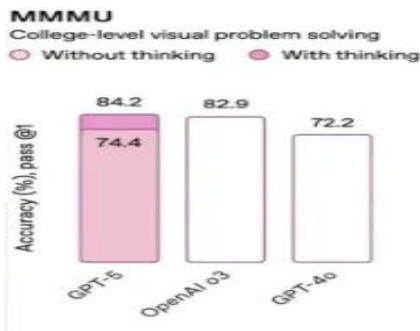


资料来源: OpenAI, 中原证券研究所

(3) 在多模态理解 (图像、视频、图标) 方面: GPT-5 在 MMMU 中得分 84.2%, 与 o3 的 82.9% 的成绩差距不大。

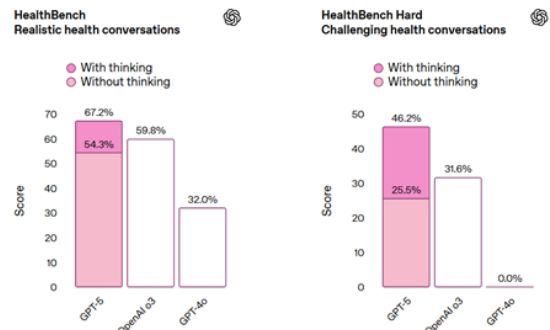
(4) 在健康领域方面: GPT-5 在 HealthBench Hard 得分 46.2%, 较 🤖 o3 的 31.6% 提升明显, OpenAI 也称 GPT-5 是最好的健康模型。

图 20: GPT-5 在 MMMU 中的成绩



资料来源: OpenAI, 中原证券研究所

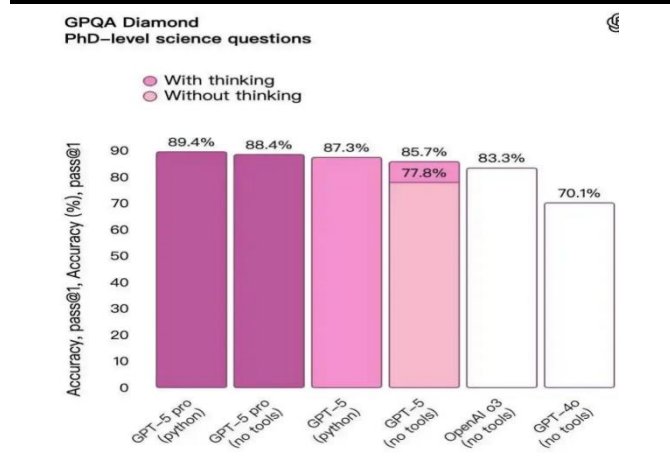
图 21: GPT-5 在健康领域的成绩



资料来源: OpenAI, 中原证券研究所

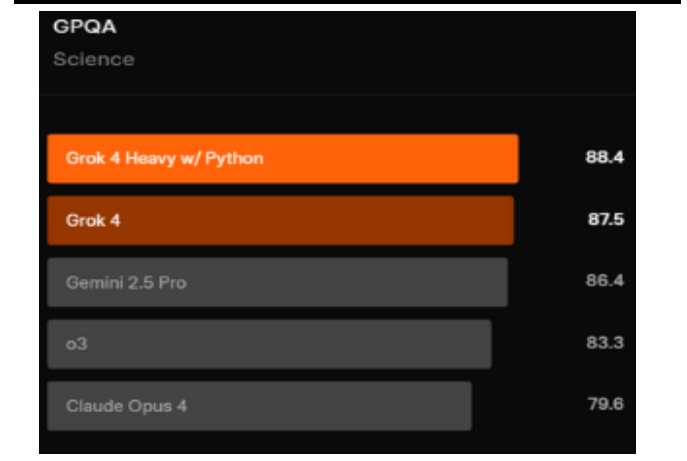
(5) 推理与百科知识方面: GPT-5 pro 在不使用工具时, GPQA 得分 88.4%, 创造了新的 SOTA, 但优势并不明显, 较 o3 83.3% 的得分提升也有限。

图 22: GPT-5 在 GPQA 中的成绩



资料来源: OpenAI, 中原证券研究所

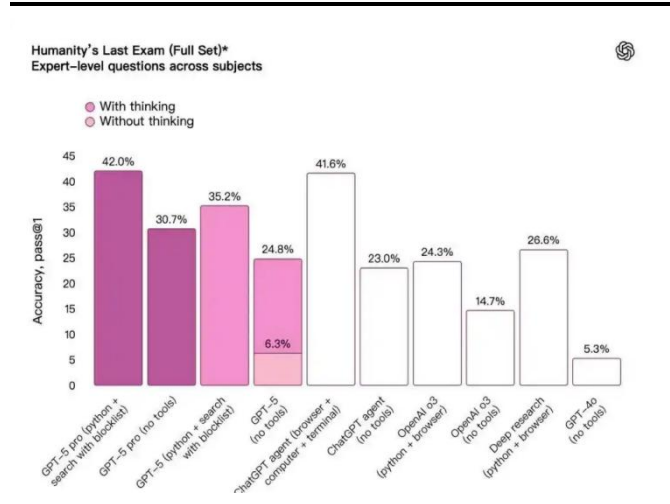
图 23: Grok 4 在 GPQA 中的成绩



资料来源: xAI, 中原证券研究所

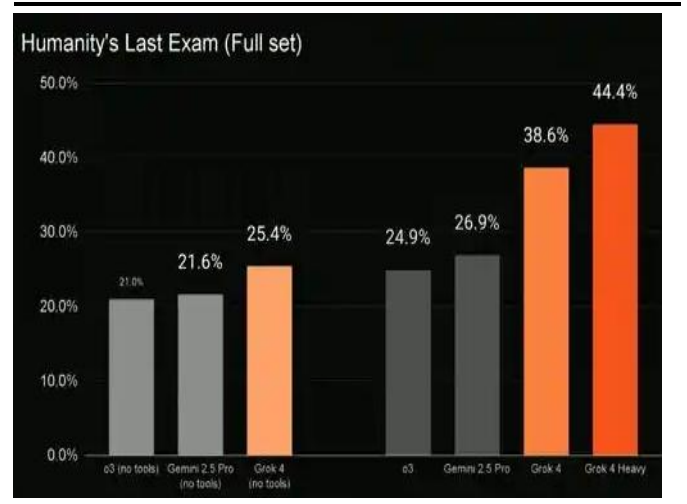
GPT-5 Pro 在 Humanity's Last Exam (HLE) 中得分最高得分 42%，较 o3 24.3% 的成绩提升明显，但是对比来看，Grok 4 Heavy 在该测试中得分 44.4%，表现较 GPT-5 更优。

图 24: GPT-5 在 HLE 中的成绩



资料来源: OpenAI, 中原证券研究所

图 25: Grok 4 在 HLE 中的成绩

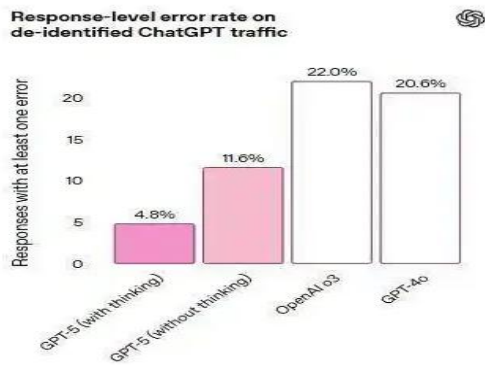


资料来源: xAI, 中原证券研究所

在其他模型优势方面:

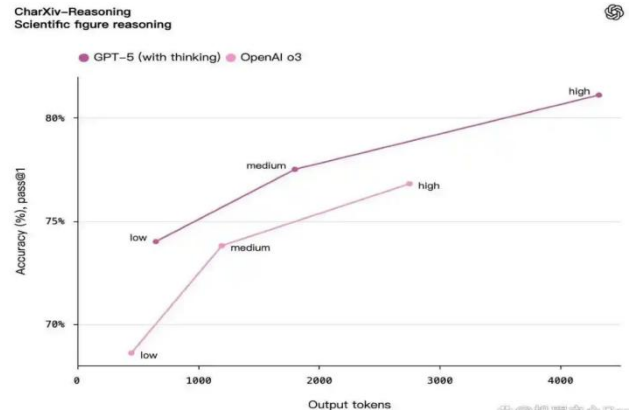
(1) **GPT-5 的幻觉可能性显著降低。** GPT-5 (非思考模式) 较 GPT-4o 降低约 45% 至 11.6%，GPT-5 (思考模式) 对比 o3 的错误性降低 80% 至 4.8%。

图 26: GPT-5 的错误信息率



资料来源: OpenAI, 中原证券研究所

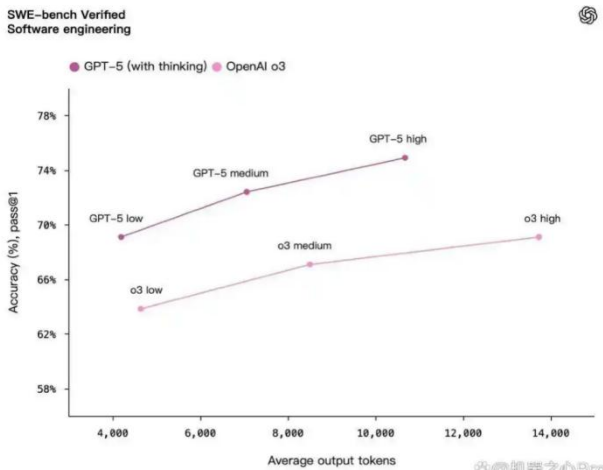
图 27: GPT-5 与 o3 在视觉推理方面的能效对比



资料来源: OpenAI, 中原证券研究所

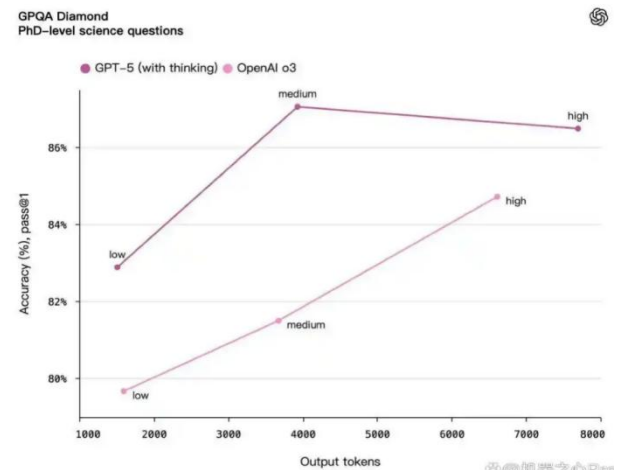
(2) 需要更少的思考时间。GPT5 (思考模式) 在视觉推理、智能体编程、GPQA 场景中, 较 o3 性能表现更优, 同时输出的 token 量减少 50-80%。

图 28: GPT-5 与 o3 在智能体编程方面的能效对比



资料来源: OpenAI, 中原证券研究所

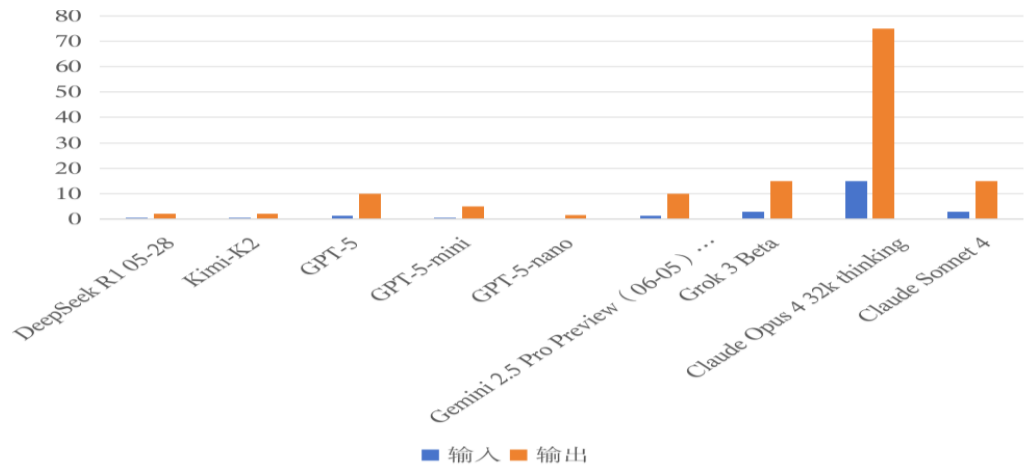
图 29: GPT-5 与 o3 在 GPQA 方面的能效对比



资料来源: OpenAI, 中原证券研究所

从目前主流模型调用价格上来看, R1 和 Kimi-K2 位于同一价位水平, 仍然是行业最低的, 输入价格 0.55 美元/百万 tokens, 输出价格 2.19 美元/百万 tokens; GPT-5 与 Gemini 2.5 Pro 的调用价格相同, 输入输出价格分别较 R1 提升了 127%和 357%; Grok 3 输入输出价格分别较 R1 提升了 445%和 585%, Grok 4 未开发 API 调用; Claude Opus4 价格最高, 输入输出价格分别较 R1 提升了 2627%和 3325%。

图 30: 主流模型 API 调用价格



资料来源: OpenAI, DeepSeek, Kimi, 谷歌, xAI, Anthropic, 中原证券研究所

1.3.3. DeepSeek: 继续关注 R2 的进展

根据 4 月太平洋网相关新闻, DeepSeek 计划于 5 月发布下一代 AI 模型 DeepSeek-R2。R2 总参数达到 1.2 万亿, 基于华为昇腾 910B 进行训练, 在 FP16 精度下达到 512PFLOPS 计算力, 芯片使用效率高达 82%, 相当于 A100 的 91%。在推理方面, R2 速度达到 320 tokens 每秒, 对于复杂逻辑推理任务上的准确率提升了 83%。

表 3: DeepSeek 的主要模型发布情况

时间	模型	架构	参数	模型类型	备注
2023. 11. 29	DeepSeek LLM 67B		670 亿	语言模型	开源, 对标 LLaMA2 70B
2024. 1. 11	DeepSeek-MoE	MoE	1450 亿		开源, 国内首个 MoE 大模型, 有 2B、16B、145B 三个尺寸,
2024. 5. 6	DeepSeek-V2	MoE	2360 亿	语言模型	开源, 性能比肩 GPT-4 Turbo, 价格为其 1/70
2024. 9. 5	DeepSeek V2. 5			语言模型	开源, 由 DeepSeek Coder V2 和 DeepSeek V2 Chat 两个模型合并升级而来
2024. 11. 20	DeepSeek-R1-Lite			推理模型	发布了对标 o1-preview 的预览版, 并开放思维链输出功能
2024. 12. 26	DeepSeek-V3	MoE	6710 亿	语言模型	开源, 性能比肩 GPT-4o
2025. 1. 20	DeepSeek-R1	MoE	6710 亿	推理模型	开源, 性能比肩 OpenAI o1 正式版, 价格约为其 1/30, 并开放思维链输出功能, 64K 上下文
2025. 3. 24	DeepSeek-V3-0324	MoE	6600 亿	语言模型	开源, 性能超过 GPT-4. 5
2025. 5. 28	DeepSeek-R1-0528	MoE	6850 亿	推理模型	开源, 128K 上下文

资料来源: DeepSeek, 中原证券研究所

在 7 月的 ACL 2025 颁奖礼上, DeepSeek 梁文锋参与撰写的论文《Native Sparse Attention: Hardware-Aligned and Natively Trainable Sparse Attention》获得了最佳论文奖 (今年总计选出 4 篇最佳论文)。该论文发布于 2025 年 2 月, 提出了新的注意力机制 NSA, 通过算法和硬件的协同优化, 把长文本处理速度提升了 11 倍, 并实现了与传统的全注意力模型相当或更优的

性能。颁奖礼上，文章作者表示该技术可以把模型上下文拓展到百万 tokens，并将应用到下一个前沿模型中。

应用了 DeepSeek-V3 的架构的 Kimi-K2, 在 7 月 11 日发布后, 成功位列开源大模型第一, 再次证明了 DeepSeek 架构的先进性。

后续重点关注 DeepSeek 在 R2 和模型多模态能力构建方面的进展。考虑到 DeepSeek 目前没有融资需求, 其模型和研究成果的释放节奏可控性更高, 不受市场预期影响和束缚。

1.3.4. 阿里：新模型位居开源模型第一

2025 年 7 月 25 日, 阿里开源了 Qwen3-235B-A22B 思考模型升级版 Qwen3-235B-A22B-Thinking-2507, 成为了全球最强大的开源模型, 整体实力居全球大模型第一梯队。

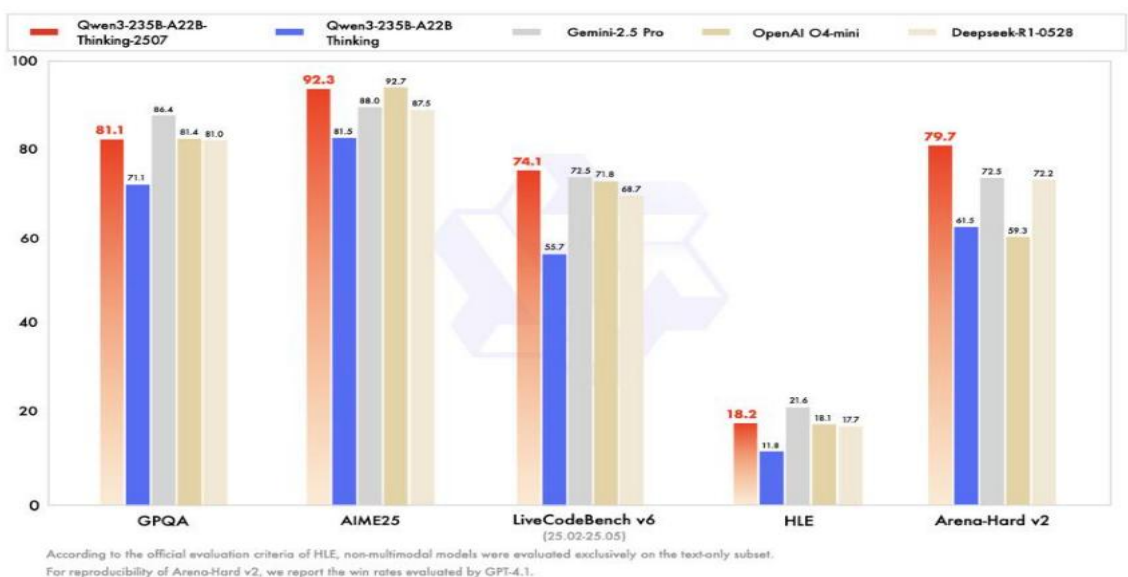
表 4: 阿里的主要模型发布情况

时间	模型	架构	参数	模型类型	备注
2025. 3. 6	QwQ-32B		32B		开源
2025. 4. 29	Qwen3-235B-A22B		235B		开源
2025. 7. 22	Qwen3-235B-A22B-Instruct	MOE	235B	基础	开源, 256k 上下文
2025. 7. 23	Qwen3-Coder	MOE	480B	编程	开源
2025. 7. 25	Qwen3-235B-A22B-Thinking-2507	MOE	235B	推理	开源, 256k 上下文, 最强开源模型

资料来源：阿里，中原证券研究所

相较于 2025 年 4 月发布的 Qwen3-235B-A22B Thinking 而言, Qwen3-235B-A22B-Thinking-2507 在推理性能和通用能力上都实现了较大的进步, 上下文长度也有显著提升。

图 31: 阿里 Qwen3-235B-A22B-Thinking-2507 各项模型能力测试得分



资料来源：阿里，中原证券研究所

1.3.5. Meta 开启 AI 人才争夺战，科技企业加快业务调整步伐

在 Llama 4 遇挫和多名核心员工离职以后，Meta 在近期成立新部门“超级智能实验室”，并重点围绕高级推理、多模态系统、大语言模型基础设施、AI 智能体等重点方向开启了 AI 人才争夺战。6 月 26 日，Meta 证实从 OpenAI 苏黎世研发中心挖走拜耶、科列斯尼科夫、翟晓华三位核心研究员，薪酬达到“亿元美金级”。

于此同时，科技企业裁员趋势明显。2025 年微软和谷歌都已经进行了 3 次裁员，微软裁员人数约 1.6 万，Meta 也在年初提出裁员 5%。一方面，企业在 AI 等技术的加持下，企业编程效率提升，对普通员工需求减少，另一方面，经济景气度偏弱和科技发展加快的背景下，企业有更强烈的业务调整和成本控制的需求，资源不断向 AI 等领域聚集。

表 5：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
Meta	2025.1	3600	计划裁员约 5%，对象是“表现不佳的员工”
惠普	2025.2	2000	削减多至 2000 个职位
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		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
英特尔	2022.10	约 12000	应对 PC 市场下滑，大幅收缩业务，约裁员 5%
	2023.5		为聚焦 AI 和先进制程，实施结构性裁员，CCG 和 DCAI 部门各裁员约 20%，预算削减 10%
	2024.8	约 15000	因盈利未达预期，英特尔为削减开支，将裁员比例提至 15%以上
	2025.7	约 34800	计划在年底前将员工总数从约 10.98 万人削减至 7.5 万人，目前员工数量减少约 15%。
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。

资料来源：光明网，彭博社，极目新闻，中原证券研究所

1.3.6. 中标情况：AI 应用加快，行业端呈现百花齐放态势

根据智能超参数统计的数据，2025 年 7 月大模型相关中标项目数量为 574 个，中标项目

披露的金额约为 13.35 亿元，中标项目数量和中标金额同比分别增长 422%和 540%。

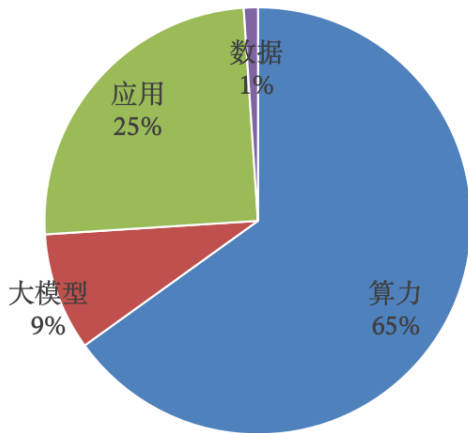
图 32：2024-2025 年大模型单月份中标项目监测数据（万元）



资料来源：智能超参数，中原证券研究所

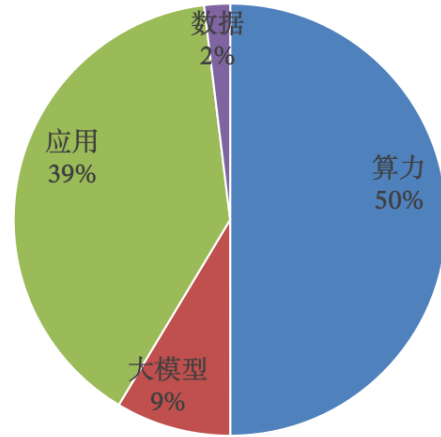
从项目类型来看，不同于 Q1 算力项目中标金额占 65%的比例，7 月算力项目占比降低到了 50%，仍然是建设主力。同时，应用型项目从 Q1 的 25%提升到了 39%，表明大模型的应用落地已经在加快过程中。

图 33：2025 年一季度大模型项目类型分布（按金额）



资料来源：智能超参数，中原证券研究所

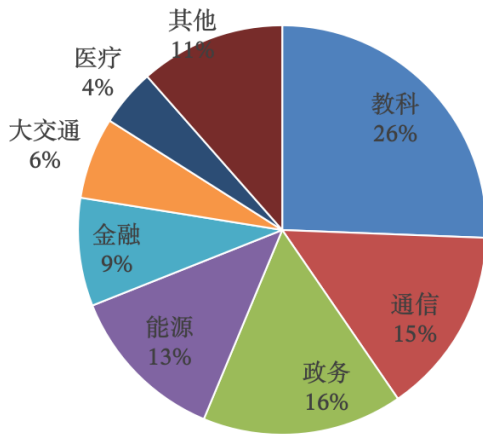
图 34：2025 年 7 月大模型项目类型分布（按金额）



资料来源：智能超参数，中原证券研究所

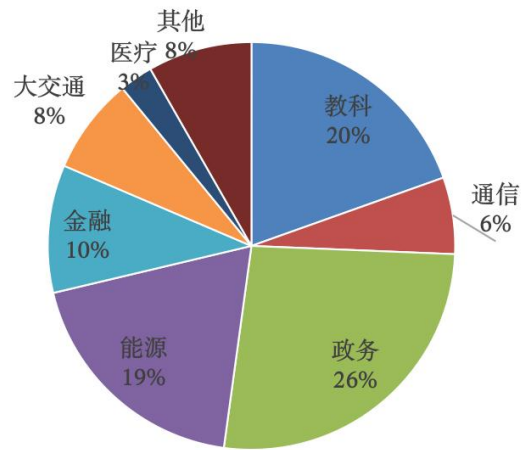
2025 年 7 月，大模型在教科、政务、通信、能源领域都有较多的订单，政务行业订单金额依然领先于其他行业，但是占比从 Q1 的 58%下降到了 26%，呈现出百花齐放的态势。

图 35：2025 年 7 月大模型订单行业分布（按数量）



资料来源：智能超参数，中原证券研究所

图 36：2025 年 7 月大模型订单行业分布（按金额）



资料来源：智能超参数，中原证券研究所

从大模型厂商中标情况来看，Q2 科大讯飞在中标数量和金额方面都全面超过百度，位列行业第一，火山引擎和阿里中标项目金额（已披露）都超过 2 亿。

图 37：2025Q1 通用大模型厂商中标情况

排行	厂商	中标数量	中标金额 (万元)
01	百度	19	45152
02	科大讯飞	16	2321
03	阿里云	13	8460
04	火山引擎	8	12266
05	智谱	7	1545
06	腾讯云	5	2799

注：排行以中标项目数量为优先指标，随后是中标金额。

资料来源：智能超参数，中原证券研究所

图 38：2025Q2 通用大模型厂商中标情况

排行	厂商	中标项目数量	披露金额 (万元)
01	科大讯飞	32	35160
02	百度	26	5208
03	火山引擎	23	23955
04	智谱	13	14248
05	腾讯云	12	4228
06	阿里云	11	20105

注：排行以中标项目数量为优先基准，其次是中标项目的披露金额。各家厂商都有一定数量项目未披露中标金额。

资料来源：智能超参数，中原证券研究所

1.3.7. AI 在广告、云业务等方向上给科技大商带来了较大的经济效益

腾讯：直接向用户收费的方式在中国较难执行，腾讯将 AI 用于为公司其他业务增长提供动能，未来腾讯还将在模型、多模态能力、代码、智能体等 AI 投入方向发力。

(1) 游戏：利用 AI 工具提升成熟游戏的内容生成速度和规模，提升通过提供更加逼真的非玩家角色提升游戏体验，AI 还提升了游戏营销的精准度。

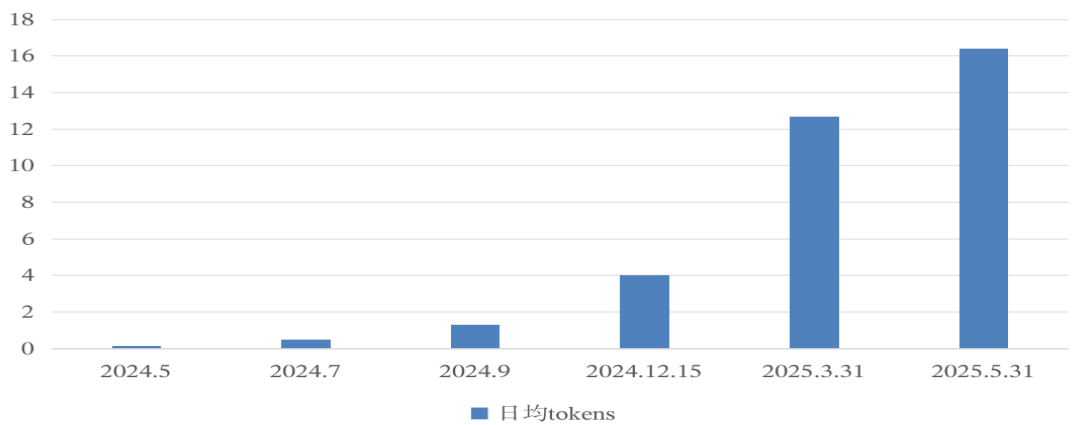
(2) 广告：基于 AI 的广告素材创作、智能推荐和效果分析，实现了 AI 技术赋能。

(3) 云服务：为客户 AI 需求提供 GPU 和 API 调用。

(4) AI 应用：元宝，AI 增强搜索，腾讯会议的纪要总结，腾讯文档的写作助手等。元宝已经可以接入并深度解读视频号内容。

字节：截至 2025 年 5 月底，豆包大模型日均 tokens 调用量已达到 16.4 万亿，较发布时增长超 137 倍，较 2025 年 3 月底增长了 3.7 万亿。

图 39：豆包大模型调用量最新情况



资料来源：字节跳动，中原证券研究所

Meta：Meta 在 7 月的业绩发布会上表示，AI 技术显著提升了广告系统效率，从而使得 Instagram 广告转化率提升约 5%，FaceBook 提升约 3%。生成式 AI 广告创意工具对于小广告主尤为为重要，也贡献了一定比例的广告收入。Meta AI 的约活跃用户超 10 亿。

微软：AI 带来了云业务 Azure 的 AI 增长，同时 AI 应用 Copilot 表现超预期。AI 功能月活跃用户已超过 8 亿，Copilot 应用商业和消费用户月活跃用户突破 1 亿。

谷歌：得益于 AI 技术的广泛应用，谷歌在搜索、云服务和 AI 功能上均实现显著增长，月处理 Token 数量飙升至 980 万亿，AI 概览功能覆盖全球 200 多个国家/地区，月活跃用户突破 20 亿。

OpenAI：2025 年 7 月 OpenAI 周活跃用户数达到 7 亿，高于 3 月的 5 亿，同时实现同比超 4 倍的增长。

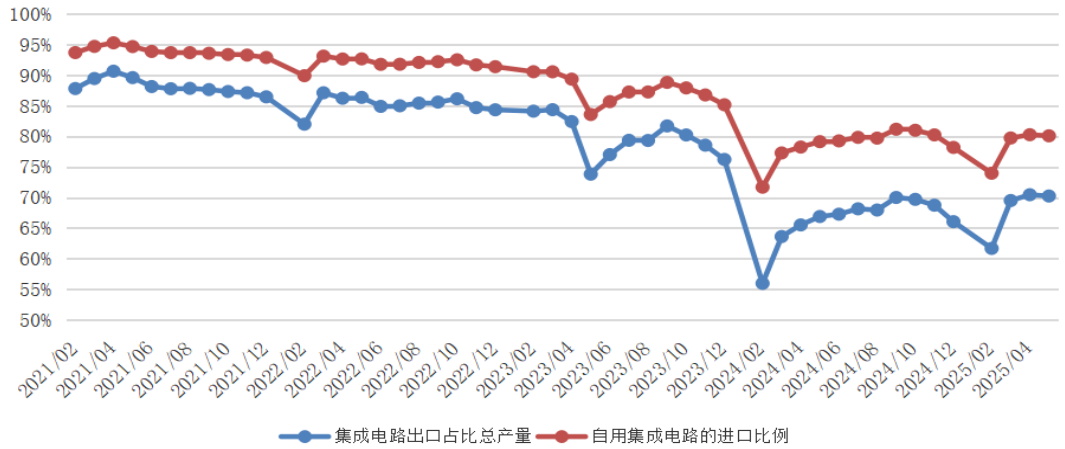
1.4. 国产化：鸿蒙、CloudMatrix384 超节点双重突破，EDA 禁令加速国产进程

1.4.1. 国产算力芯片呈加速趋势，AI 芯片企业步入集中上市期

2025 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 80%（即

国产化占比 20%) 和 70%，都与 1-4 月持平。

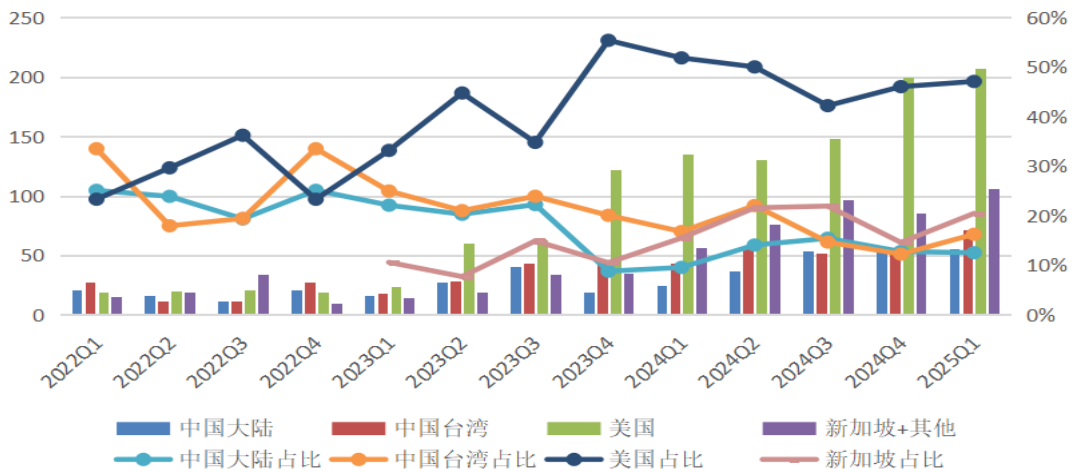
图 40：2021-2025 年我国集成电路对进口依赖度和出口占比（按数量）



资料来源：工信部，海关总署，中原证券研究所

2025Q1 中国大陆已经退居英伟达第四大出口区域。其来自美国的收入占比为 47%，较前一个季度提升了 1 PCT；其第二大收入区域新加坡的收入占比从上个季度的 15% 回升到 20%，需求还在快速释放；而与此同时，中国台湾超过中国大陆，成为了英伟达第三大收入来源区域，占比达到 16%；其来自中国大陆的收入已经连续 3 个季度保持了 55 亿美元左右的规模水平，没有看到因为 AI 需求较为明显的释放效应，我们认为这其中仍有较大的需求受到了压制。

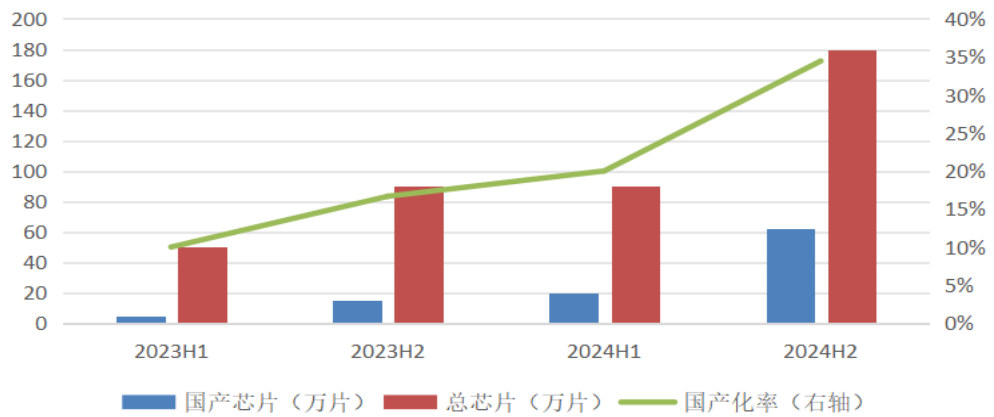
图 41：2022Q1 至 2025Q1 英伟达分区域收入（亿美元）及占比



资料来源：英伟达，中原证券研究所

2024 年 AI 芯片国产替代趋势明显。根据 IDC 数据，2024 年下半年，我国 AI 芯片国产化比率从上半年的 20% 进一步提升到了 34%，提升趋势明显。

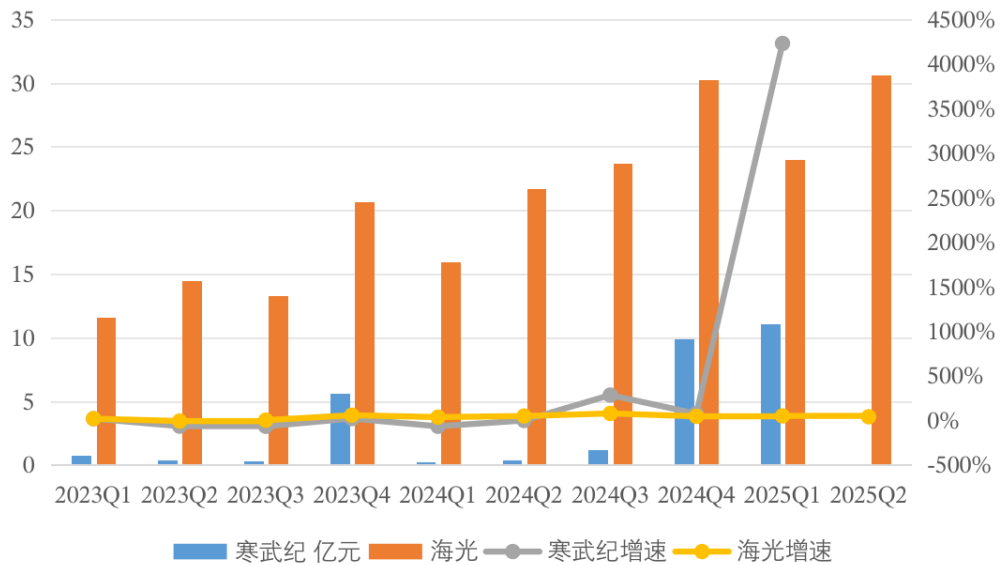
图 42：2023H1-2024H2 我国 AI 芯片出货情况及国产化比率



资料来源：IDC，中原证券研究所

国内数据中心芯片厂商中，寒武纪、海光继 2024 年的收入高增长以后，2025 年业务都呈现加速迹象。其中寒武纪 2025Q1 财报呈现业绩爆发趋势，收入同比增长 4230.22%至 11.11 亿元，环比也实现了 12.36%的增长，很好地印证了当前国产芯片和 AI 算力两大方向的算力需求。截至 2025 年上半年，海光合同负债达到 30.91 亿元，较上年同期的 9.03 亿元同比增长 242.1%，主要来自客户预付的合同贷款，将对后续收入高增长带了有力的支撑，同时其 DCU 已在智算中心、人工智能等多个领域实现规模化应用。

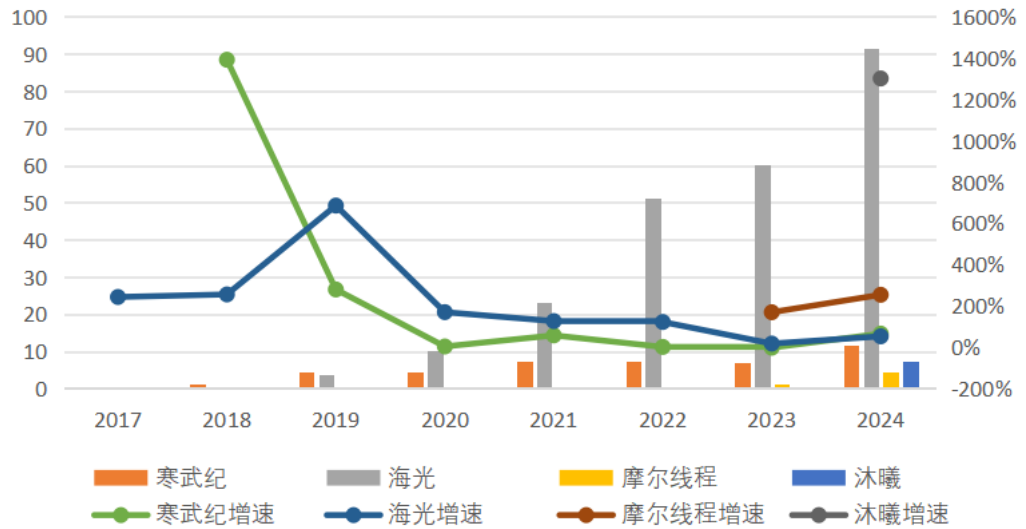
图 43：2023Q1-2025Q2 寒武纪、海光收入（亿元）及增速对比



资料来源：上市公司财报，中原证券研究所

同时，国内 AI 芯片企业也步入集中上市阶段。6 月 30 日，沐曦与摩尔线程科创板 IPO 申请获得受理，将分别募集 39 亿元与 80 亿元。两家公司都还处于发展早期，研发投入带来较高的亏损。2024 年沐曦营收 7.43 亿元，净亏损 14.09 亿元；摩尔线程营业收入 4.38 亿元，净亏损 14.92 亿元。此外，燧原科技在 2024 年 8 月发布了上市辅导备案报告，壁仞科技拟 2025Q3 提交港股上市申请。

图 44：2017-2024 国内主要上市 AI 芯片企业收入及增速对比



资料来源：上市公司财报，中原证券研究所

1.4.2. H20 供应仍存在较大不确定性，但是下游厂商在推理测已经有了更多选择

虽然在 7 月中旬，GPU 禁令松动，但是整体出货情况仍然面临较多不确定性。继为期 3 个月的禁令期以后，7 月 15 日英伟达宣布美国已批准 H20 芯片销往中国。根据 IT 之家的报道，2024 年 H20 销量 100 万片，其现有库存 60-70 万片，在强劲需求下英伟达又紧急向台积电订购了 30 万片。7 月 31 日，国家网信办因为 H20 芯片漏洞后面安全风险问题约谈了英伟达，主要涉及英伟达算力芯片的“追踪定位”“远程关闭”功能，对此英伟达回应芯片中不存在后门、终止开关和监控软件，但路透报道美国在人工智能芯片中设置跟踪器，来发现转运至中国的情况。但与此同时，美国方面也因为主管部门瘫痪而没有进行 H20 许可发放，而后提出征收营收的 15% 作为许可费。

表 6：2025 年 AI 芯片相关动态

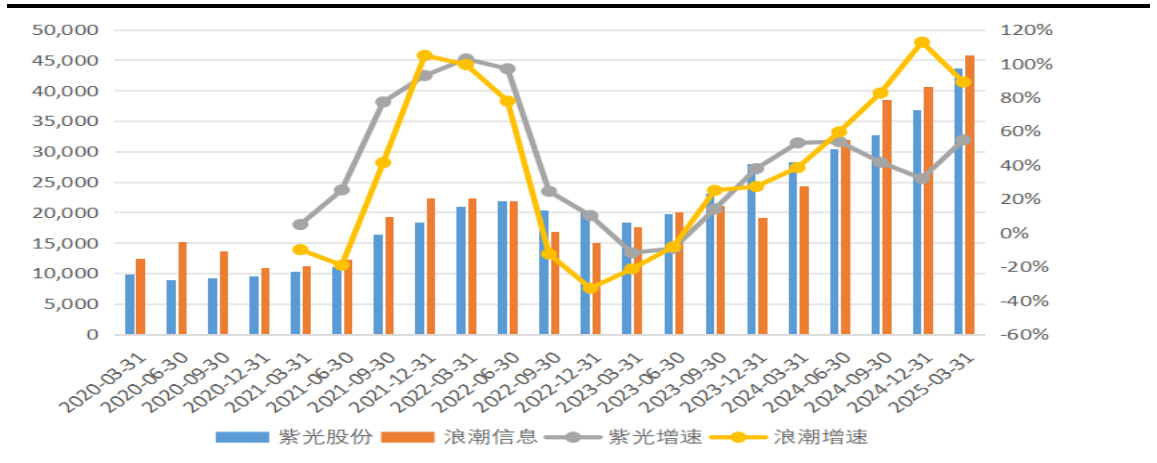
厂商	内容
2025. 3. 25	美国商务部将 54 家中国实体列入清单，包括了浪潮集团旗下 6 家公司、宁畅信息、中科可控旗下的服务器品牌 Suma
2025. 4. 16	英伟达表示美国政府已经于 4 月 14 日通知公司，H20 芯片未来在出口至中国时需要“无限期”申请许可证
2025. 6. 3	报道称英伟达将为中国研发 B30 芯片，其采用 Blackwell 架构，GDDR7 显存，而非 HBM，也不会采用台积电的先进封装技术，首度支持多 GPU 扩展，允许用户通过连接多组芯片来打造更高性能的计算集群。采用 GB20X 芯片，也就是 RTX 50 系列的芯片，其售价预计在 6500 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 算力芯片后面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”。

资料来源：和讯网，中原证券研究所

当前服务器厂商销售仍然需要依靠存货来维持。2023Q3 开始，浪潮存货增长速度持续加快，从 2023Q3 的 25% 增长到 2024Q4 的 113%，2025Q1 又继续增长了 89%。与此同时，紫光股份存货也保持了较快增长，2024Q3 和 2024Q4 或受到收购新华三股份带来财务压力，增速放缓，但是 2025Q1 又快速增长了 55%。

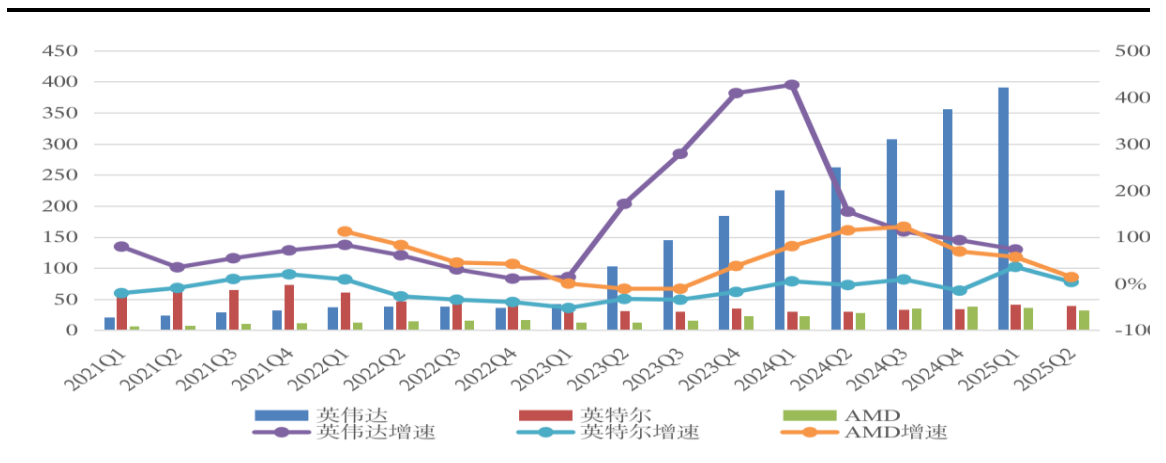
图 45：2020Q1-2025Q1 紫光和浪潮存货增长情况（百万元）



资料来源：上市公司公告，中原证券研究所

禁令同样对海外的 AI 芯片企业带来了较为明显的负面影响。2025 年 4 月，英伟达表示因为对 H20 的出口限制，英伟达二季度将记录约 55 亿费用。AMD 也在 2025 年 5 月表示，其 Instinct MI308X 也因为对中国的出口管制将在二季度产生 8 亿美元库存减记。同时，由于出口限制，AMD Q2 数据中心业务收入 32 亿美元，同比增长 14%，环比下跌了 12%。

图 46：2021Q1-2025Q2 美国 3 大芯片厂商数据中心业务收入及增速（亿美元）



资料来源：上市公司公告，中原证券研究所

国产算力的持续提升，软件的优化升级，给了国产云计算和大模型厂商更多的选择和腾挪空间。在腾讯 2 季报电话会议中，其高管表示目前 GPU 供应充足，未来将进一步提升云计算中的 GPU 租赁。同时腾讯已经拥有足够的芯片来进行基础模型的训练和持续迭代，同时在推理芯片的供应侧具备多种选择。而且腾讯还在通过软件优化和升级来提升推理效率，从而在同等芯片数量下承载更高的负载。

1.4.3. 华为开源 CANN，加速 AI 生态建设

8 月 5 日，华为在昇腾计算机产业发展峰会上宣布 CANN Mind 系列应用套件及工具链全面开源。

CANN（Compute Architecture for Neural Networks）是华为主导的神经网络异构计算架构。从 AI 训练的产品架构来看，最底层是 AI 芯片层，然后进入到 AI 计算架构层，CANN 就是华为适配昇腾芯片打造的计算架构，接着是 AI 训练框架层，主流的有 TensorFlow、PyTorch 以及华为的 MindSpore 等，其借助计算架构调用 AI 芯片，最后是上层的大模型及 AI 应用层。

表 7：华为在 AI 训练各层的产品布局

	华为产品	对标产品	产品性质	华为产品开源情况
AI 芯片	昇腾	英伟达 A100、H100、B200 等	硬件	
计算架构	CANN	英伟达 CUDA、AMD 的 ROCm、摩尔线程 MUSA、壁仞科技 BIRENSUPA、寒武纪 Neware、海光信息 DTK	软件	2025.8 开源
AI 框架	MindSpore	PyTorch、TensorFlow	软件	2020.3 开源
AI 大模型	盘古大模型	OpenAI GPT-5、DeepSeek R1、谷歌 Gemini 2.5 Pro、Anthropic 的 Claude 4，阿里 Qwen3	软件	
AI 应用	小艺等		软件	

资料来源：中原证券研究所

2018 年 9 月，华为发布了 CANN 1.0，此后又在持续进行产品更新，2024 年 9 月升级到 CANN8.0。截至 2025 年 8 月，CANN 已支持 PyTorch、MindSpore、TensorFlow、飞桨、ONNX、计图、OpenCV 和 OpenMMLab 等深度学习框架与第三方库。

CANN 对标英伟达 CUDA。英伟达 CUDA 与 GPU 和 NVlink 构成英伟达的业务护城河，用户想要进行英伟达 GPU 迁移，需要进行大量代码重写，失去完善的代码库和庞大技术社区的支持。在国内芯片替代的过程中，CANN 持续完善。华为在推出昇腾 384 以后，又通过这次 CANN 开源的方式，加快补齐短板，加速赶超英伟达。

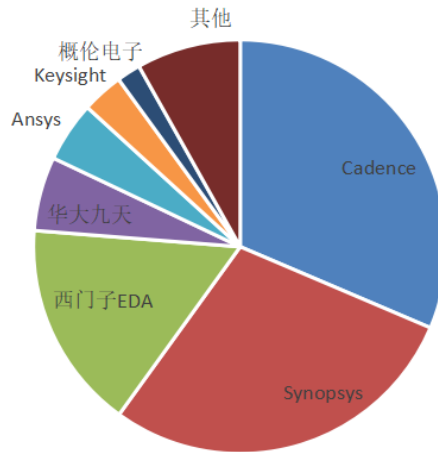
1.4.4. EDA：禁令暂时解除，但国产替代仍是大势所趋

美国对华 EDA 禁令解除，但美国方面在出口管制方面保持强硬态度。5 月 29 日《金融时报》报道，美国政府停止对华芯片设计软件（EDA）销售。7 月 2 日，彭博社报道称，西门子发布声明，称在中国开展业务不再需要获得“政府许可”。7 月底，楷登电子因为向国防科大

销售芯片设计软件，违反了美国出口管制，而受到了美国 1.4 亿美元罚金。

根据 IDC 数据，2024 年我国 EDA 市场规模 105.2 亿元，楷登电子、新思科技、西门子 EDA 合计占据 70% 以上的份额。根据华经产业研究院的数据，2021 年，我国 EDA 市场中楷登电子、新思科技、西门子 EDA 占比分别为 32%、29%、17%，国产厂商中华大九天位居第四，占比 6%，概伦电子位居第 7，占比 2%。

图 47：2021 年国内 EDA 市场格局

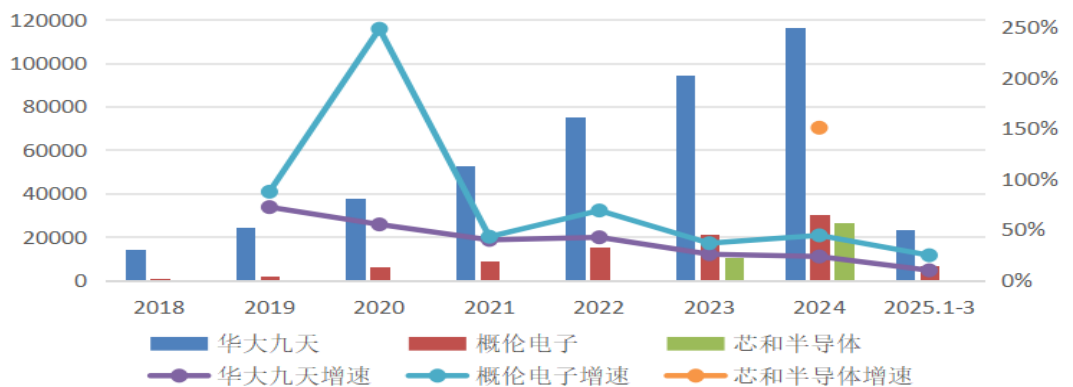


资料来源：华经产业研究院，中原证券研究所

早在 2023 年，华为轮值董事长徐直军就表示华为芯片设计 EDA 工具团队联合国内 EDA 企业，共同打造了 14nm 以上工艺所需 EDA 工具，基本实现了 14nm 以上 EDA 工具国产化，2023 年将完成对其全面验证。

在芯片加快国产化的趋势下，国内 EDA 企业也受到国产化趋势拉动，维持较高增速。2024 年华大九天、概伦电子的国内业务收入分别增长了 23.4% 和 44.27%，芯和半导体也获得了 150.38% 的收入增长；2025Q1，华大九天收入增长 9.77%，概伦电子国内业务收入增长 24.61%。

图 48：2018-2025 年 EDA 企业国内业务收入及增速（万元）



资料来源：上市公司公告，中原证券研究所（华大九天 2025Q1 和芯和半导体取整体收入）

虽然 EDA 禁令解除，但是考虑到当前的国际局势和政策的持续变化，我们认为进行国产 EDA 替换仍是大势所趋。

1.4.5. 鸿蒙电脑操作系统适配应用或超 2500 款，C 端销售情况良好

继 2024 年 10 月纯血鸿蒙 HarmonyOS NEXT 发布以后，我们看到鸿蒙在 2024 年 Q4 的份额也有明显提升趋势。根据 Counterpoint 的数据，2024Q4 鸿蒙以 19% 的份额继续位居国内手机操作系统第二名，较 Q3 提升了 2 PCT，同时鸿蒙在全球继续占据 4% 的市场份额。

图 49: 2022Q2-2024Q4 中国手机操作系统市场份额

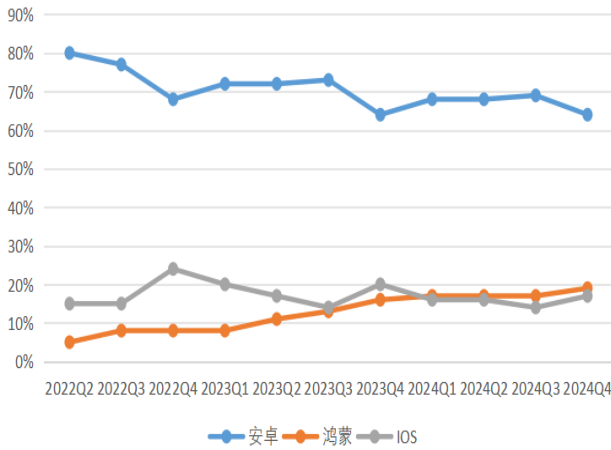
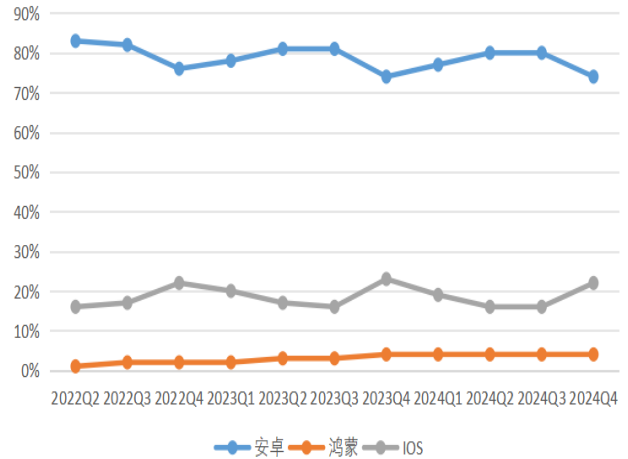


图 50: 2022Q2-2024Q4 全球手机操作系统市场份额



资料来源: Counterpoint, 中原证券研究所

资料来源: Counterpoint, 中原证券研究所

继手机端的纯血鸿蒙推出以后，华为在 5 月正式发布了鸿蒙电脑操作系统。5 月 8 日鸿蒙电脑操作系统首次亮相，5 月 19 日首款鸿蒙电脑搭载鸿蒙 5.0 操作系统正式发布。在生态系统建设方面，5 月 8 日鸿蒙电脑有 300 多款应用完成适配，而到了 5 月 19 日这一数量提升到了 1000 多款，6 月 6 日人民日报报道鸿蒙电脑生态应用数量已经突破 2000 款，预计到 6 月底突破 2500 款（根据 6 月 20 日华为开发者大会 HDC2025）。

图 51: 鸿蒙电脑应用适配情况



资料来源: 华为 2025.5.19 发布会, 中原证券研究所

由于微软对于华为的 Windows 授权许可在 2025 年 3 月到期，2025 年 4 月开始华为笔记本预装 Linux 操作系统，但是用户可以自行安装 Windows 操作系统。随着 2 款鸿蒙电脑的开

售，华为笔记本也如同苹果 Macbook 一样走上了软硬一体的发展道路，用户开始正式使用鸿蒙电脑系统，部分没有适配的软件可以通过铯大师虚拟机切换到 Windows 场景。

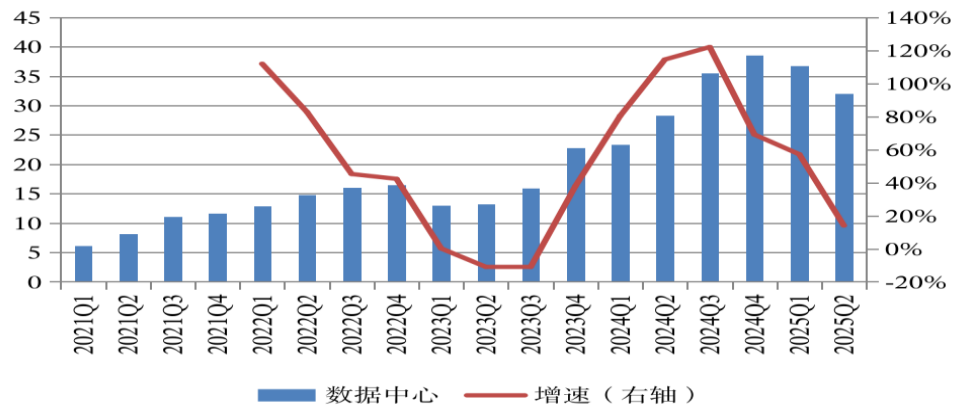
由于鸿蒙电脑可以较好地满足轻办公需求，同时新品的轻薄体验较优，门店出现了明显缺货的情况，这意味着华为已经可以较好地实现 C 端的国产系统替代，加速推动 PC 端国产化比例的提升。

1.5. 算力：海外科技巨头加大资本投入，液冷加速落地中

1.5.1. 除英伟达以外的海外算力芯片发展格局

AMD 的芯片渗透率持续提升，新芯片已经开始量产。十大模型厂商中，已经有 7 家采用了 AMD 的 AI 芯片，标志着 AMD 在生态方面取得了重要的进步。AMD 的 MI350 系列（MI355/MI355X）在 2025 年 6 月已开始提前启动量产，其在关键的训练和推理工作负载中达到或超过了 B200，在核心工作负载中性能与 GB200 相当，但成本和复杂度显著降低，已经在甲骨文数据中心进行了大规模部署。2025 年下半年，MI350 即将在多家客户端进行大规模部署，有望带动产品的渗透率进一步提升。同时，MI400 系列将在 2026 年上市，将达到 40PFLOPS 的 FP4 性能。

图 52：2021Q1-2025Q2 AMD 数据中心业务收入及增速（亿美元）



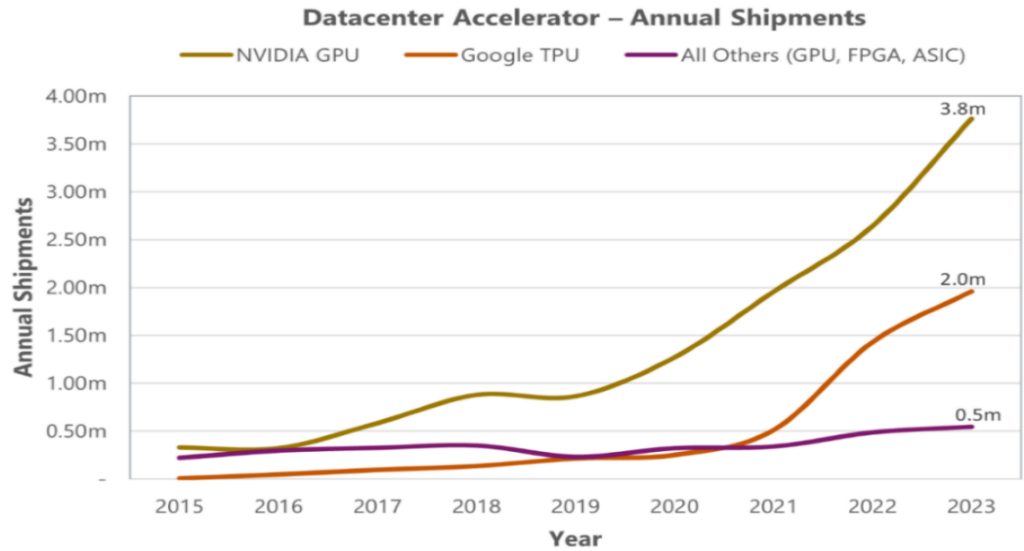
资料来源：AMD 公告，中原证券研究所

为了降低在芯片端的支出和能耗需求，同时减少对英伟达的依赖，大型云厂商都在陆续开启芯片自研，并有望在 2025 年较好地实现对英伟达 GPU 部分需求的替代：

(1) **谷歌**：从 2013 年就开始自主研发云端 AI 加速芯片，并在 2015 年推出了自研的 TPU（张量处理器）。2025 年 4 月 9 日，谷歌发布了其第七代 TPU Ironwood，是谷歌首款专为推理设计的 TPU，其峰值算力 4614TFLOPs，内存 192GB，内存带宽 7.2Tbps，互联带宽 1.2Tbps，可扩展至 9216 片芯片集群，性能是第六代 TPU Trillium 的 2 倍。同时，作为云厂商中自有 AI 芯片占比最高的厂商，谷歌 2023 年自用的 TPU 芯片量已经突破了 200 万颗，已经成为了仅次于英伟达的全球第二大数据中心 AI 芯片厂商。2025 年谷歌还宣布计划进入数据中心 CPU 市场，

将实现微处理器和加速器芯片皆自研覆盖的发展格局。

图 53：2015-2023 数据中心加速器年度出货量（片）



资料来源：TechInsights，中原证券研究所

(2) **亚马逊**: 在 2013 年提出了开发定制硬件策略, 2018 年推出了基于 ARM 架构的 CPU Graviton, 2019 年发布首代推理芯片 Inferentia, 2020 年发布了训练芯片 Trainium。2023 年末谷歌发布了 Trainium 2, 主要对标英伟达旗下 A100/H100 等旗舰 AI 训练芯片。谷歌将在 2025 年下半年发布 Trainium 3, 将采用 3nm 工艺, 计算性能较 Trainium 2 提升一倍。目前, Trainium2 性价比比其他 GPU 供应商高出 30% 到 40%, 并已成为 Anthropic 等公司新一代模型训练的支柱。

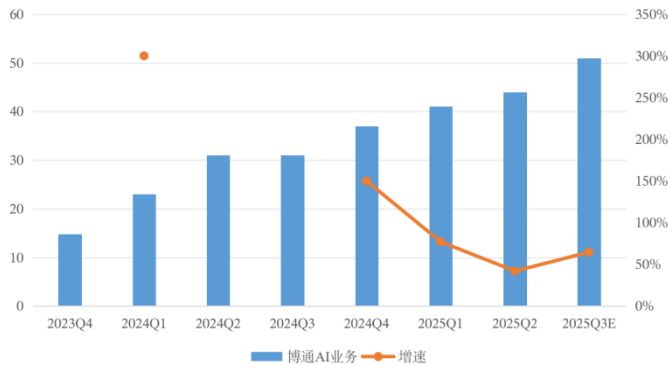
(3) **Meta**: 在 2023 年 5 月推出了第一代的 AI 芯片 MTIA v1, 采用了 7nm 工艺, 内存 128MB。2024 年 4 月, Meta 推出了 MTIA v2 (Artemis), 采用 5nm 工艺, INT8 精度下的稠密算力达到上一代的近 3.5 倍, 稀疏算力达到上一代的近 7 倍, 达到 708TFLOPS, 内存 256MB。META 与博通联合开发的 MTIA T-V1 预计最早在 2025Q4 推出, 被认为能效将较英伟达下一代 GPU Rubin 提升 20%, 预计出货量达 30-40 万片; 计划在 2026 年中推出的 MTIA T-V1.5, 计算密度接近 GB200, 出货量计划 50-80 万片; 2027 年计划推出 MTIA T-V2, 目标可进行万亿参数的模型训练。

(4) **微软**: 2023 年 11 月, 微软推出了自研 AI 芯片 Maia 100 和 CPU 芯片 Cobalt 100, Maia 100 采用 5nm 工艺, 1050 亿颗晶体管。微软原计划在 2025 年推出的推理芯片 Maia 200 (Braga) 芯片, 目前已经推迟到了 2026 年, 同时微软还计划在 2027 年发布 Maia 280 芯片, 将两个 Braga 芯片连接组成, 性能功耗有望比英伟达同年产品高出 30%。微软计划在 2028 年量产的推理芯片 Maia 400 (Braga-R), 将采用更先进的连接技术。微软计划随着每一代新芯片的推出而逐步提高产量, 最终目标是年产数十万颗自研 AI 芯片。

受到客户自研芯片业务的带动, 2025Q2 ASIC 厂商博通的 AI 业务实现收入 44 亿美元, 预计 2025Q3 将进一步增长到 51 亿美元, 其客户包括了谷歌、Meta、字节跳动、OpenAI、软银

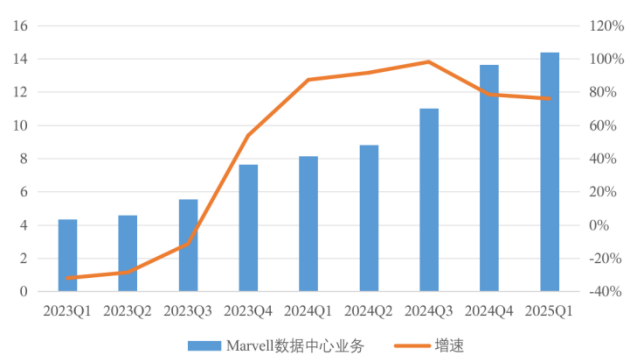
等重要厂商。而与此同时，第二大 ASIC 厂商 Marvell 在亚马逊 Trainium 2 等需求带动下，2025 年第一个季度数据中心业务收入 14.406 亿美元，同比增长 76.46%。

图 54：2023Q4-2025Q3E 博通 AI 业务收入（亿美元）



资料来源：博通财报，中原证券研究所

图 55：23Q1-25Q1 Marvell 数据中心业务收入（亿美元）

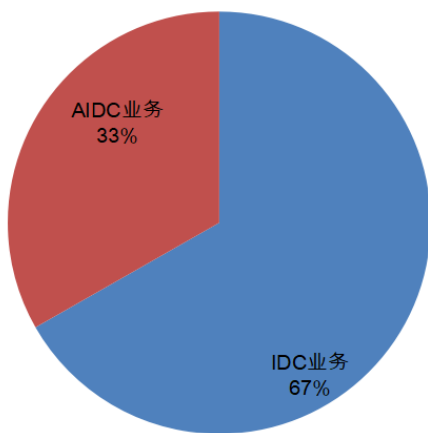


资料来源：Wind，中原证券研究所（Marvell 财报时段较以上周期晚一个月）

1.5.2. 2025 年大规模智算中心或将呈现较为明显的集中交付趋势

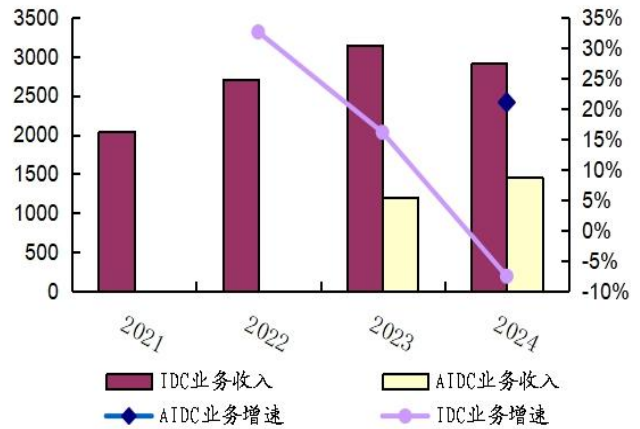
AIDC 业务需求快速增长，新进入者众多。根据润泽科技财报，可以看到其 AIDC 业务上架非常快，2023 年上架当年就实现了 12 亿收入。2024 年 IDC 业务面临了较大的竞争压力，同时公司进行了旧数据中心的改造，导致 IDC 规模有所下滑，而 AIDC 实现了 21% 的收入增长。由于公司 2024 年进行了部分 AIDC 业务计入方式的调整，从总额法变为净额法，因而实际 AIDC 业务增长应明显超过 21%，同时按照金额法计入收入以后，公司 AIDC 业务毛利率 51.75%，明显高于 IDC 业务同期 47.53% 的毛利率水平。商汤科技的 AIDC 业务在 2024 年增长了 50% 达到了 16.5 亿元，总运营算力达到 23EFLOPS，较上年 12EFLOPS 接近翻倍。

图 56：2024 年润泽科技收入分布



资料来源：润泽科技公告，中原证券研究所

图 57：2021-2024 年润泽科技分项业务收入（百万元）



资料来源：润泽科技公告，中原证券研究所

传统 IDC 竞争加剧。科智咨询数据显示，2024 年，中国传统 IDC 业务市场规模为 1,583 亿元，同比增长 8.7%，增速缓慢，传统低功率数据中心需求持续白热化。

2025Q1 来看，3 家 IDC 企业中，润泽科技和奥飞数据都保持了较好的增长，数据港净利润增速回升明显。

表 8：2025Q1 企业 IDC 相关业务对比（亿元）

公司	相关业务	2025Q1					2024 年			
		收入	增速	净利润	增速	毛利率	收入	毛利率	算力规模	人员增速
中国联通	数据中心	72.2	8.8%				259			
万国数据		27.2	4%	7.6	320%	24%	100	22%	61.36 万平米	-3%
世纪互联		22.5	18%	-2.4	-72%	25%	74	22%	5.21 万架+ 486MW	-15%
秦淮数据*								42%	12.26 万架	10%
润泽科技		12	21%	4.3	-9%	50%	44	49%	8.2 万架	9.5%
宝信软件	服务外包						35	45%		
奥飞数据		5.4	41%	0.52	2%	32%	13	26%	4.3 万架	1%
光环新网	IDC 及其增值服务						22	34%	5.19 万架	
数据港		4.0	4%	0.44	23%	29%	15	31%	7.42 万架	49%
科华数据	IDC 服务						13	23%	3 万架+	

资料来源：上市公司公告，Wind，中原证券研究所（*秦淮退市，故取 2022 年数据，秦淮数据、数据港、光环新网为 5KW 标准机架折算，科华数据为 2024H1 数据）

2025H1，中国联通 AIDC 业务签约金额同比增长 60%，同时其数据中心业务同比增长 9.4%，较 Q1 8.8% 的增速略有加速趋势。而于此同时，中国电信 AIDC 业务仅增长 7.4%。

在新一代智算建设和需求方面，2025 年润泽科技和奥飞数据都有相对于自身较大的算力投放计划，万国数据也在 2025Q1 获得了单笔高达 152MW 的大规模订单。由于超大规模客户通常需要在短时间内交付至少 50MW 的 IT 容量，预计 2025 年将会有较为明显的大规模智算中心集中交付趋势。

表 9：IDC 企业订单获取和运力投放情况

厂商	发布时间	内容
万国数据	2025.3.19	2025Q1 获得来自现有超大规模客户的订单，与 AI 需求相关，面积约 40,000 平方米，相当于 152MW IT 容量，分布于廊坊和常熟的两座数据中心，预计在 6 个月内交付，该订单是公司在国内有史以来最大的单笔订单。
奥飞数据	2025.4.21	廊坊固安数据中心预计将于 2025 年内交付 3 栋数据中心，天津武清数据中心二期预计将于 2025 年内交付，河北怀来数据中心的土建环节预计 2025 年中开始分期完工并于 2025 年底或 2026 年初交付首期项目，河北定兴数据中心预计 2025 年上半年交付。
润泽科技	2025.4.27	廊坊 A 区剩余算力中心近期即将交付，廊坊 B1 的 200MW 新一代智算中心，正在紧张建设中；平湖 B1 单体 100MW 智算中心，建设基本完成，近期即将交付；佛山园区第 1 栋已经上架，第 2 栋正在筹备交付，两栋楼合计 80MW 预计今年年底趋向成熟；惠州园区 A1 正在筹备交付，单栋 40MW 预计今年年底趋向成熟。
DayOne	2025.5.20	2025Q1 新增数据中心签约容量 70MW，总签约容量 530MW。公司在泰国数据中心与客户达成明确签约意向，另一方面成功进军欧洲市场，在芬兰落子首个数据中心项目，并于客户签约，两个项目总容量预计超过 220MW，预计将使公司 2025 年全年签约容量超过 750MW。

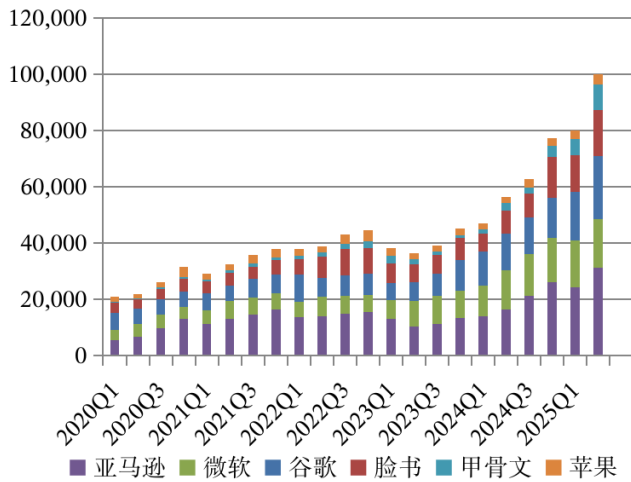
世纪互联 | 2025.5.28 | 未来12个月公司计划交付377MW。

资料来源：上市公司公告，中原证券研究所

1.5.3. 海外科技巨头资本开支明显增大

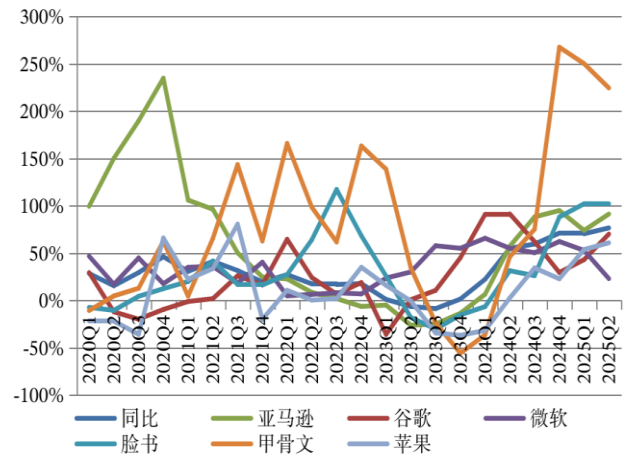
从海外市场来看，2025Q2，美国6大科技厂商（亚马逊、微软、谷歌、脸书、甲骨文、苹果）的资本开支再创新高，总计达到999.73亿美元，同比增长77%，增速再创新高。其中甲骨文自2024Q4以来资本开支增速都在200%以上，投入规模持续加大；Meta以102%的增速加快投入，自2024Q4以来都保持了积极的投入策略；亚马逊以91%的增速位居第三，总规模仍然牢居第1；谷歌和苹果近2个季度都呈现加速迹象，增速分别为70%和61%；微软连续两个季度增速下滑，增速23%，资本投入规模有即将被Meta超越的趋势。

图 58：6 大科技厂商资本投入（百万美元）



资料来源：Wind，中原证券研究所

图 59：6 大科技厂商资本投入增速



资料来源：Wind，中原证券研究所

海外大型科技厂商的资本开支计划呈现上调趋势。随着海外科技巨头接连公布了 2025Q2 业绩，我们看到了资本开支都呈现出了上调趋势，也预示着 AI 需求整体超预期。其中，微软 2025Q3 预计资本开支 300 亿，同比将增长 50%，环比增长 24%；亚马逊 2025 年原计划投入 1000 亿美元，按照下半年与 Q2 同一开支水平来计算，2025 年实际投入有望超过 1200 亿美元；Meta 在 2025 年两次提高了投入目标，从 1 月的 600-650 亿美元，到 5 月的 640-720 亿美元，到 7 月的 660-720 亿美元；谷歌将年度资本开支从 750 亿美元提高到了 850 亿美元；甲骨文预计 2026 财年资本开支超过 250 亿美元。此外，2025Q2，甲骨文 RPO 同比大增 41% 至 1380 亿美元，亚马逊订单积压同比增长 25% 至 1950 亿美元，总体仍然呈现出供给瓶颈，有望带动资本开支的持续增大。

Meta 已宣布了两大重要 AI “泰坦” 集群建设计划：（1）位于俄亥俄州的 Prometheus 集群，预计 2026 年上线，将成为首批计算能力达 1 千兆瓦的 AI 超级集群；（2）位于路易斯安那州的 Hyperion 集群，规模将与曼哈顿相当，并有望在数年内扩展至 5 千兆瓦的计算能力。

表 10：大模型厂商算力相关投入

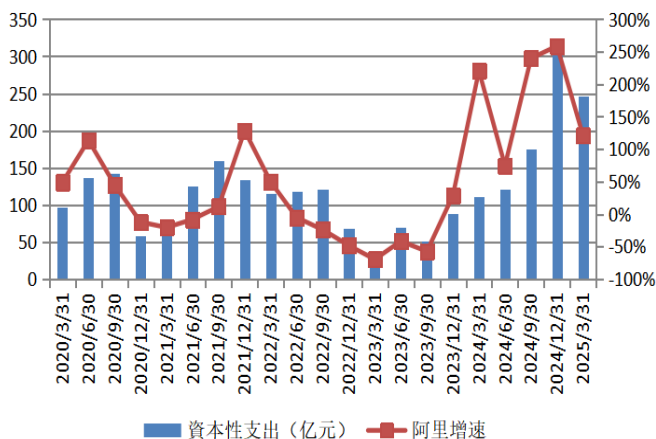
公司	时间	详情
微软	2024. 3. 31	与 OpenAI 正计划建设一个名为“星际之门”的特殊数据中心，为 OpenAI 提供强大的计算支持，计划最快在 2028 年启动该项目，并且 2030 年之前会进一步扩建，所需要的电力可能高达 5 千兆瓦。“星级之门”超算会配备百万专用的服务器芯片，项目成本预计高达 1150 亿美元。
	2025. 1. 4	将在本财年（6 月结束）投资 800 多亿美元建设人工智能负载数据中心，以训练人工智能模型，并部署人工智能和基于云的应用程序。
	2025. 5. 1	Q2 的资本投入绝对额会比本季有所提升，2026 财年资本支出会继续增长，只是同比增速会放缓。
	2025. 7. 31	考虑到对云和人工智能产品的强劲需求信号以及大量合同积压，公司将继续在资本支出和运营费用方面进行投资，预计 2026 财年第一季度资本支出将超过 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。
OpenAI 甲骨文	2025. 1. 21	美国总统特朗普在白宫宣布，由美国甲骨文公司、OpenAI 和软银共同出资建设，三家公司组成的联合企业将投资 5000 亿美元，在美建设名为“星际之门”的 AI 基础设施。该项目计划在美国建设数据中心，以支持人工智能技术发展。
	2025. 5. 27	甲骨文计划斥资 400 亿美元采购英伟达的 GB200 芯片，以支持其在德克萨斯州阿比林市建设的“星际之门”项目中首个启动的数据中心，该中心预计将在 2026 年年中全面投入运营。
	2025. 7. 22	OpenAI 与甲骨文达成协议，将共同在美国开发 4.5GW 新增“星际之门”数据中心算力。据介绍，结合德州阿比林市在建的首个“星际之门”“基地，此次与甲骨文的合作将使 OpenAI 开发中的 AI 数据中心总容量突破 5GW，可支持超 200 万枚芯片运行。
甲骨文	2025. 6. 12	预计 2026 财年资本开支超过 250 亿美元。
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
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。
Meta	2025. 1. 24	计划在 2025 年投入 600-650 亿美元，用于 AI 相关的资本支出，计划在 2025 年上线 1GW 的算力，预期到年底时能够拥有 130 万块 GPU。
	2025. 5. 1	下调了 2025 年全年总支出的预估范围，从 1140 亿美元至 1190 亿美元之间下调到 1130

		亿美元至 1180 亿美元之间，其中用于生成式 AI 以及核心业务方面的投资范围从 600-650 亿美元之间上调到 640-720 亿美元之间。
	2025. 7. 31	将大力推进人工智能基础设施建设，计划 2025 年相关资本支出（含融资租赁本金支付）达 660-720 亿美元，其中值较上年同比增长约 300 亿美元。这一积极的投资增长态势将延续至 2026 年，以持续提升在线产能，满足人工智能研发及业务运营的需求。
xAI	2024. 9. 3	xAI 仅用 122 天就建成了 10 万张 H100 的 Colossus 集群，未来还会扩展到 15 万张 H100 和 5 万张 H200。
	2024. 12. 5	xAI 承诺将其超级计算机项目 Colossus 的算力扩大十倍，未来将搭载 100 万块 GPU，从而对抗谷歌、OpenAI 和 Anthropic 等竞争对手。
小米	2024. 12. 26	小米集团正在积极构建自己的 GPU 万卡集群，以加大对 AI 大模型的投入。
字节	2024. 3. 1	截止 2023 年 9 月，字节跳动已经建立超过一万张的英伟达 Ampere 架构 GPU 集群，目前正在建设 Hopper 架构的集群。
	2024. 12. 31	Information 报道，字节跳动计划在 2025 年斥资 70 亿美元购买英伟达芯片，该消息随后被字节跳动发言人否定。
	2025. 1. 22	据财联社报道，已为 2025 年在中国购买人工智能芯片编列了 400 亿人民币（55 亿美元）的预算，是去年支出的两倍。公司还计划在海外投资约 68 亿美元，采用英伟达芯片加强基础模型训练能力。
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，总额超过过去十年的总和。
法国	2025. 2. 9	未来几年，法国将投资超 1090 亿欧元用于建设 AI。

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

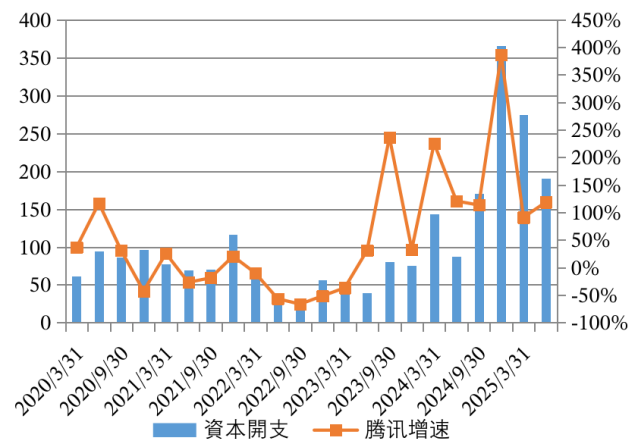
从国内市场来看，阿里、腾讯的资本开支均呈现环比下滑趋势，下半年或有改善预期。腾讯 Q1 资本开支 191.07 亿元，同比增长 118.9%，环比下滑 30%，整体位于较高水平。阿里 Q1 资本开支 246.12 亿元，同比增长 120.7%，环比下滑 22.5%。百度 Q1 资本开支 28.69 亿元（不含爱奇艺），同比增长 42.3%，环比增长 24.09%，还没有超过 2023Q4 36.41 亿元的峰值水平，投入相对保守。H20 的禁令解除后，阿里、腾讯、字节或有资本开支和芯片采购的改善。

图 60：2020-2025 年阿里单季度资本开支及增速



资料来源：阿里财报，中原证券研究所

图 61：2020-2025 年腾讯单季度资本开支及增速（亿元）



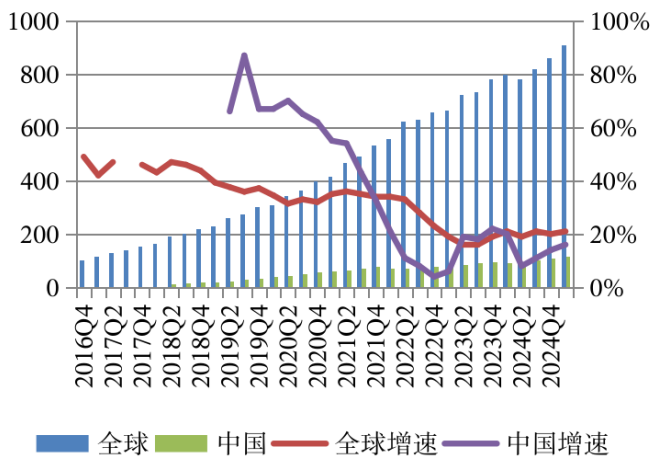
资料来源：腾讯财报，中原证券研究所

1.5.4. 云计算：在 AI 业务的带动下，全球云计算业务整体呈现出加速趋势

国内云业务连续 3 个季度提速。根据 Canalys 数据，2025Q1 我国云基础设施服务上的支出总计达到 116 亿美元，同比增长 16%，较上个季度增长 2PCT，连续 3 个季度增速回升，但仍然低于全球 21% 的增速水平。我们预计随着 DeepSeek 的开源和国内大模型的发展，国内云业务的增速有望持续提升。

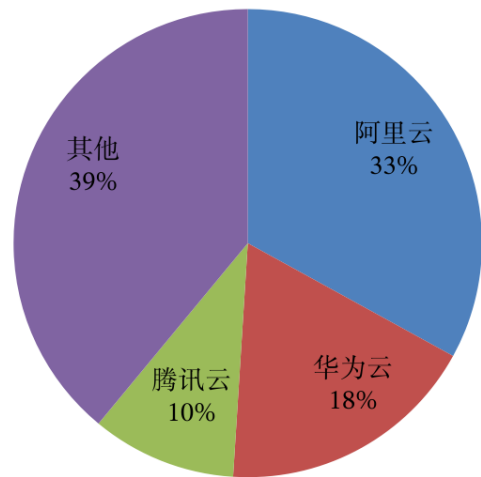
从云基础设施服务的厂商布局来看，阿里仍然牢居第一，占比 33%；华为云在生态发展壮大的趋势下，份额达到 18%，增速高于行业 2PCT，预计随着昇腾 384 的部署，其云业务还有进一步提升的趋势；腾讯云以 10% 的份额位居第 3，由于优先对内供给，GPU 供给紧张，对市场份额产生负面影响，在这个过程中，我们也看到腾讯更加积极的布局 AI 领域。

图 62：2016-2025 年我国及全球云基础设施服务的支出规模及增速（亿美元）



资料来源：Canalys，中原证券研究所

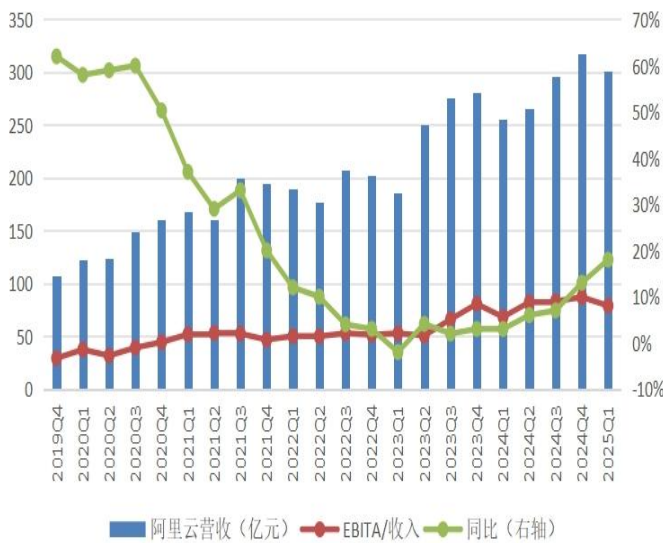
图 63：2025Q1 我国云基础设施服务市场结构



资料来源：Canalys，中原证券研究所

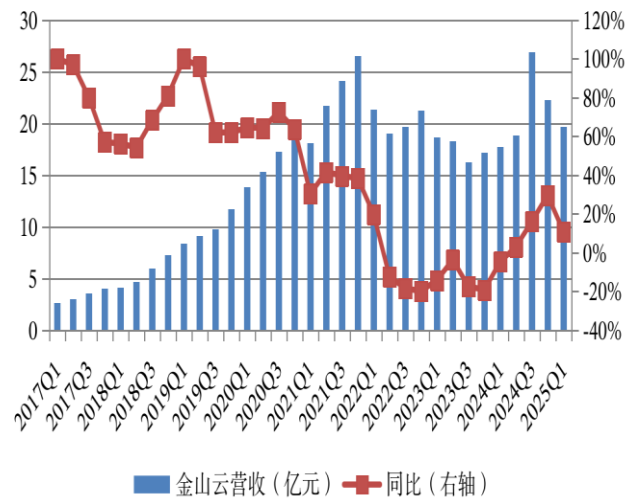
从厂商业绩来看，阿里云和金山云收入增速都从 2024Q1 开始有了明显的改善趋势。其中 2025Q1 阿里云收入增速达到了 18%，较上个季度提升了 5PCT，继续呈现明显加快趋势，其中 AI 相关收入连续 7 个季度实现同比三位数增长。2025Q1 金山云同比增长 11%，仍处于增长趋势中，其中人工智能业务总收入同比增长 228%，至 5.25 亿元。

图 64：2019Q4-2025Q1 阿里云收入及其增速



资料来源：阿里财报，中原证券研究所

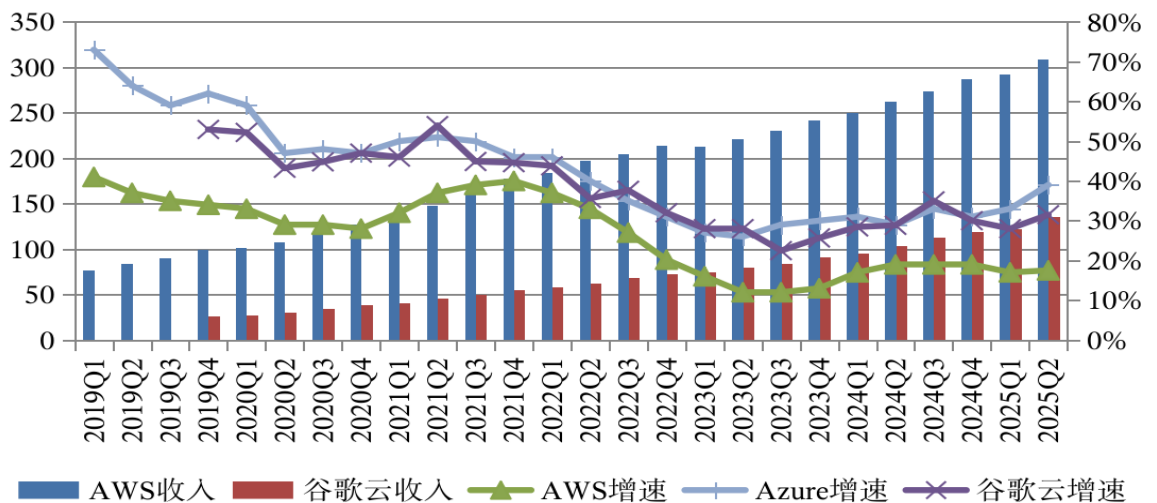
图 65：2017Q1-2025Q1 金山云收入及其增速



资料来源：金山云财报，中原证券研究所

Q2 全球云巨头都呈现了业务加速趋势。企业的根据 2025Q2 相关公司财报，亚马逊的 AWS 实现收入 309 亿元，同比增长 17.5%，较上个季度提升了 0.5 PCT。在芯片、数据中心、电力方面的资本投入，导致 AWS 的折旧费用大幅上升，导致 Q2 营业利润率从 Q1 的 39.5% 下降到了 Q2 32.9%，同时 AI 需求超过了亚马逊的供给能力，特别是在电力方面面临较大的供给瓶颈。而于此同时，微软 Azure 和谷歌云分别增长了 39% 和 31.5%，分别较上个季度增长了 6 和 3.5 PCT，增速绝对值和增速提升幅度都明显快于 AWS。

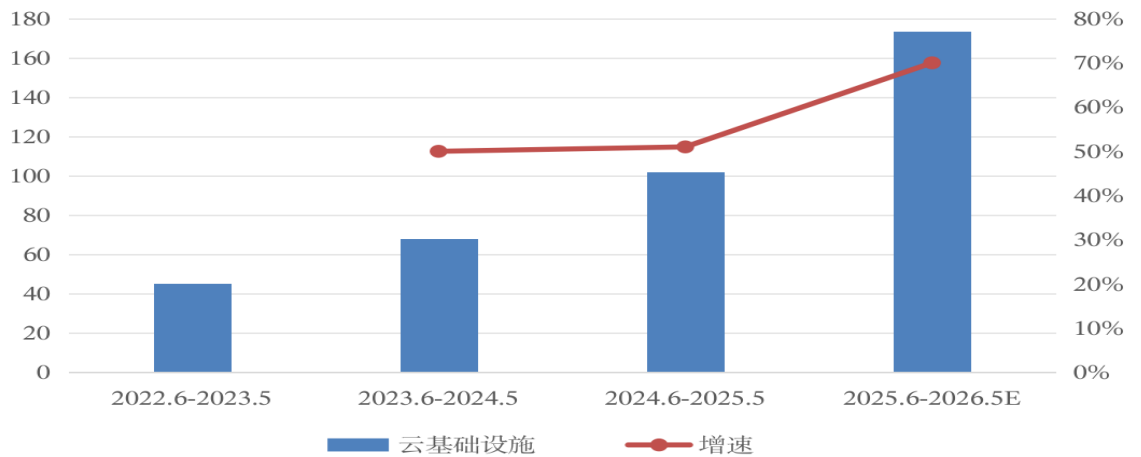
图 66：2019Q1-2025Q2 海外云巨头的相关业务收入及增速



资料来源：上市公司财报，中原证券研究所

此外，正在加速进行云投入的甲骨文，即将成为全球第四大云基础设施服务厂商。2025 财年 4 季度，甲骨文云基础设施营收 30 亿美元，同比增长 52%，预计 2026 财年营收增长超过 70%，已经与 OpenAI、xAI、Meta、英伟达和 AMD 签署了云协议，同时还在 6 月 30 日披露签署了 2028 财年开始贡献超过 300 亿美元年收入的大客户。

图 67：甲骨文 2023-2026 财年云基础设施业务收入规模（亿美元）



资料来源：甲骨文财报，中原证券研究所

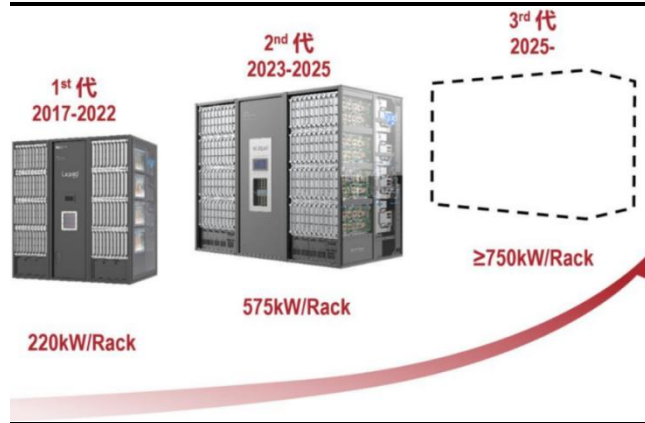
1.5.5. 液冷：微软所有区域都可以支持液冷，超节点全面拥抱液冷，行业加速趋势明显

液冷作为数据中心新兴的散热解决方案，虽然在 AI 带来的高算力需求下蓬勃发展，但是在数据中心渗透率仍不足 10%，且以冷板液冷为主。相对于浸没式以及喷淋式液冷技术，冷板式液冷对原有风冷机房以及风冷服务器的改动较小，因而成为当前应用最广泛的液冷技术。

在冷板液冷方面，由于行业内尚没有统一的国家级标准，目前仍以定制化解决方案为主。定制化方案带来验证成本高，原料差异化程度高，无法形成规模效应，导致了冷板液冷市场毛利率呈下行趋势（叠加上下游企业数据中心散热市场）。当前，液冷算力基础设施领域首批国家标准《数据中心冷板式液冷系统技术规范》、《数据中心冷板式液冷机柜通用规范》即将发布，国内首个液冷部件国标《信息技术 服务器及存储设备用液冷部件技术规范》也在起草中，有助于推动行业的标准化和规模化。

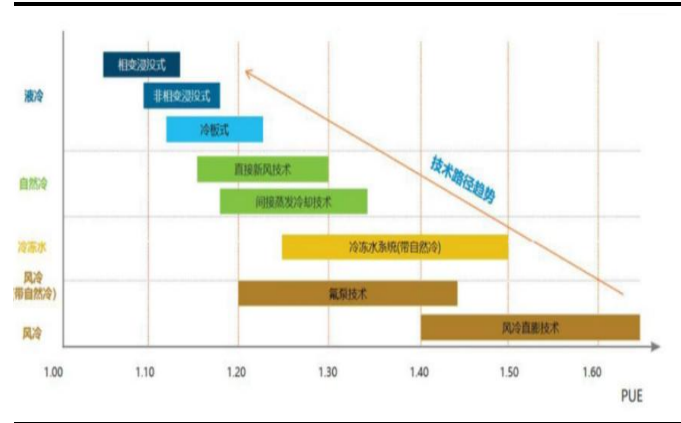
在浸没式液冷方面，目前主要用于超算中心，具有高定制化特征。曙光数创作为国内唯一实现浸没相变液冷技术大规模商业化部署的企业，在 2024 年推出了第三代 C8000 浸没液冷解决方案，单机柜功率密度突破至 750kW 以上，为未来算力芯片性能的持续提升和数据中心更高的节能需求提供了技术储备。

图 68：2022-2024 年 智算服务细分市场（亿元）



资料来源：曙光数创公告，中原证券研究所

图 69：不同冷却技术与 PUE 相关性



资料来源：《中兴通讯液冷技术白皮书》，曙光数创公告，中原证券研究所

随着产业链发展逐渐成熟，在一定规模情况下，冷板液冷初始投资已经低于风冷。只考虑初始投资情况下，在单机柜功率不超过 10kW 时，采用风冷相对经济，一般认为单机柜 20kW 是风冷可解的散热极限；单机柜在 10-100kW 的数据中心，采用冷板液冷较为合适；单机柜功率超过 100kW 的数据中心采用相变浸没液冷更经济。

英伟达 GB200 NVL72 和华为昇腾 384 都应用了液冷技术，这意味着国内外最先进的超节点都步入了液冷技术的范畴，随着计算密度的进一步密度提升，浸没液冷方案也将成为未来的发展方向。此外，微软也在 7 月 Q2 电话会议中表示，其所有区域现在都可以支持液冷技术，液冷加速趋势明显。

表 11：液冷产业链公司业务动态

公司	年份	概况
曙光数创	2011	2011 年底在业内首先提出液冷技术路线并启动相关技术研究。
	2023	累计建设超过 260MW 液冷数据中心，涉及科研、能源、政府、金融、互联网、运营商等多个行业。根据赛迪顾问《2023 中国液冷应用市场研究报告》显示，2021 年至 2023 年上半年，公司以平均 58.8% 的市场份额，位列中国液冷数据中心基础设施市场部署规模第一，多年稳居行业头名。
	2024	累计建设超过 400MW 液冷数据中心，涉及科研、能源、政府、金融、互联网、运营商等多个行业。据中国电子技术标准化研究院主编的《液冷数据中心白皮书》，公司在 2023 年度液冷基础设施市场份额高达 61.30%，连续三年蝉联行业榜首。
英维克	2022	得益于液冷需求的快速增长及英维克在液冷的“全链条”平台优势，来自数据中心机房及算力设备的液冷技术相关营业收入约为上一年度的 5 倍左右。
	2023	得益于液冷需求的快速增长及英维克在液冷的“全链条”平台优势，来自数据中心机房及算力设备的液冷技术相关营业收入约为上一年度的 4 倍左右。
	2024	得益于液冷需求的快速增长及英维克在液冷的“全链条”平台优势，来自数据中心机房及算力设备的液冷技术相关营业收入约 3 亿元。 2024 年 3 月第一名中标中国电信弹性 DC 舱 2024-2025 年集采，产品应用于“临港智算谷”万卡算力集群等项目实施中。作为机电总承包商，为广州新一代液冷算力中心项目提供了从系统设计到设备集成，从工程交付到调试验收的整体交钥匙服务
	2025	截至 3 月公司在液冷链条的累计交付已达 1.2GW。
申菱环境	2024	液冷产品营收约为去年同期的 2.9 倍。

浪潮信息	2024	持续践行“ All in 液冷 ”战略，推进数据中心绿色降碳。在产品层面，发布全栈液冷产品，实现全栈服务器产品均支持冷板式液冷，并从部件、整机到数据中心，持续进行液冷产品创新，包含全液冷冷板服务器、全液冷机柜、机柜式冷量分配单元、高密液冷智算算力仓等，数据中心产品体系不断完善，液冷服务器连续 3年 蝉联中国市场第一。
润泽科技	2021	年底开始交付实验性液冷机柜。
	2023	2023年1月初开始批量交付液冷机柜，2023年7月交付了单机柜功率超 20KW 的业内首例整栋纯液冷智算中心。
	2024	将自研的新一代冷板式液冷制冷技术全面应用于全国多园区新一代智算中心项目中，实现单机柜 40kW 以上的高密度、绿色化部署。
	2025E	廊坊B区单栋智算中心承载运行 16384 台 12kW 以内高性能服务器，支持 13万卡 以上超大规模算力场景。

资料来源：上市公司公告，中原证券研究所

2. 河南计算机行业动态

2.1. 河南计算机行业要闻

7月1日，宇树科技智能机器人河南首个体验店正式开业。（消息来源：河南日报）

7月3日18时，下班晚高峰。在鹤壁市淇滨区兴鹤大街，搭载着雷达传感器的智慧合杆正将实时车流数据传向云端，红绿灯配时将会根据通行需求实时调整，实现了从“车看灯”到“灯看车”的转变。5.5万个智能终端就像城市的‘数字神经元’，实时数据融合**33类AI**算法，可实现全方位感知城市状态，形成物联、视联、数联、智联‘四联’的城市数据体系，推动城市治理向智能化的‘智能预判、主动响应’转变。（消息来源：河南日报）

7月9日，省长王凯深入郑州市和航空港区部分企业，实地调研人工智能产业发展情况。（消息来源：大象新闻）

7月17日下午，“探元计划2024”龙门石窟场景共创营开放日活动在洛阳龙门石窟举办，现场发布了聚焦龙门石窟浅浮雕高精度建模与数字化纹饰提取成果。（消息来源：河南日报）

7月17日消息，省发展改革委、省财政厅、省自然资源厅近日联合印发《河南省深化智慧城市发展推进城市全域数字化转型实施方案（2025—2027年）》（以下简称《方案》），聚焦数字经济高质量发展和智慧城市高效能治理，谋划部署**21项**重点任务，系统构建了我省城市全域数字化转型工作的基础框架体系。（消息来源：河南日报）

上半年，郑州市机电产品、高新技术产品出口占据主导地位。其中，计算机与通信技术产品、手机和汽车出口仍居顶流。此外，电动汽车、音视频设备及其零件等出口增长幅度较大。（消息来源：河南日报）

8月9日消息，省政府近日印发《河南省支持人工智能产业生态发展若干政策措施》（简称《措施》），推动人工智能科技创新与产业创新深度融合，打造人工智能发展新高地。河南

省将设立总规模 30 亿元的人工智能产业基金，壮大耐心资本，探索投补联动、投贷联动等支持机制，满足人工智能企业不同生命周期阶段的融资需求。（消息来源：河南日报）

8 月 10 日，盾构/TBM 大数据挖掘共同体在郑州成立，吸引了来自国内 160 余家单位的 400 余名代表到场见证。活动现场还同步发布了国内首个隧道与地下空间领域垂直大模型——“先锋·隧道大模型”。（消息来源：河南日报）

2.2. 河南计算机行业数据跟踪

电子信息产业延续快速增长态势。在超聚变、昆仑技术等服务器企业带动下，上半年，全省电子信息产业增加值增长 11.1%、比一季度加快 0.8 个百分点。

上半年，全省限额以上单位可穿戴智能设备、通讯器材、计算机及其配套产品类商品零售额分别增长 95.3%、92.3%、87.0%，分别比一季度加快 0.2、16.1、44.8 个百分点。

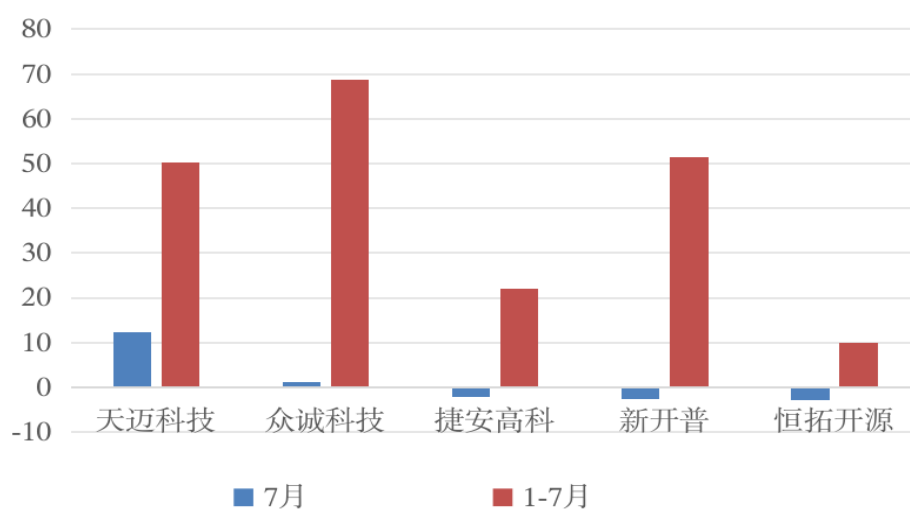
上半年，计算机、通信和其他电子设备制造业固定资产投资增长 15.8%，信息传输业投资下降 18.2%。

（数据来源：河南省统计局）

2.3. 河南上市公司行情回顾

7 月，5 家河南计算机行业上市公司涨跌互现。7 月河南 5 家计算机公司中 2 家上涨，其中天迈科技上涨 12.43%，上月领涨的众诚科技上涨 1.21%，捷安高科、新开普、恒拓开源分别下跌 2.15%、2.67%、2.86%。

图 70：河南上市公司近期股价涨跌幅表现（%）



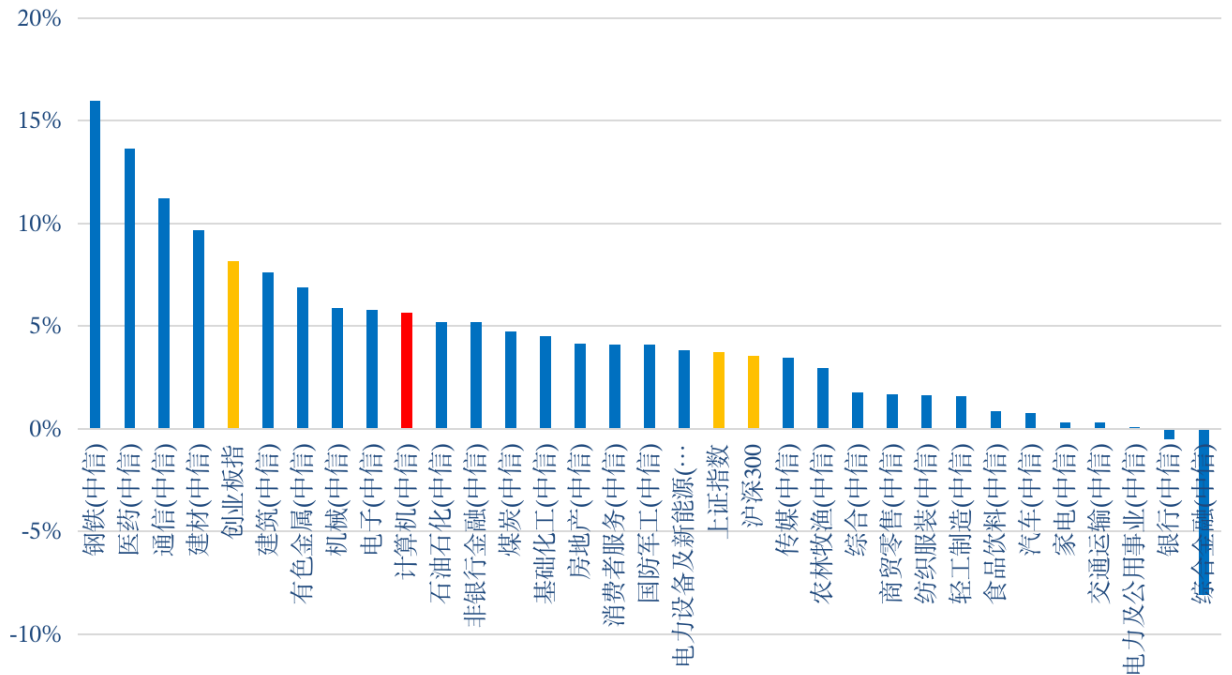
资料来源：Wind，中原证券研究所

3. 投资策略

3.1. 行情回顾：7 月计算机行业连续两个月上涨

7 月计算机行业结束连续两个月上涨，在行业指数中表现居前。7 月中信计算机指数上涨了 5.64%，跑赢上证指数 1.90 PCT，跑输创业板指数 2.49 PCT，跑赢沪深 300 指数 2.10 PCT，在 30 个中信一级行业中排名第 9。

图 71：2025 年 7 月中信一级子行业涨跌幅

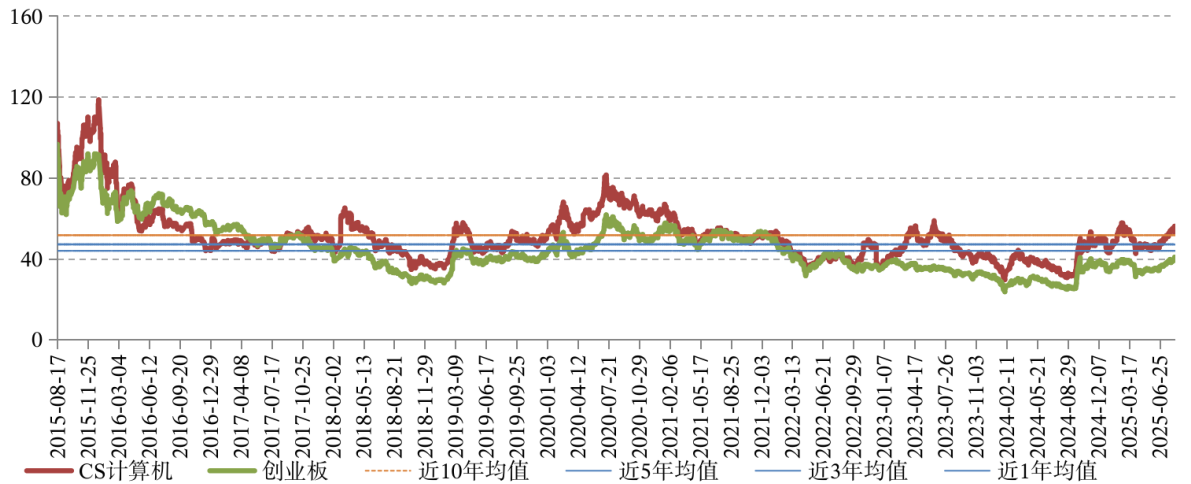


资料来源：Wind，中原证券研究所

3.2. 估值：行业的估值超过历史均值水平

行业的估值超过历史均值水平。根据 Wind 数据，2025 年 8 月 14 日中信计算机行业 TTM 整体法（剔除负值）估值为 55.65 倍，创业板估值 40.15 倍。行业近 1 年、3 年、5 年、10 年的平均估值分别为 47.00 倍、43.90 倍、47.18 倍、51.62 倍，当前估值已经超过历史均值水平之间，近 3 年最高点 58.79 倍和最低点 29.69 倍。

图 72：近 10 年中信计算机行业估值水平（截至 2025.8.14）



资料来源：Wind，中原证券研究所

3.3. 行业观点与投资建议

2025 年 1-6 月软件产业增速持续上升。2025 年 1-6 月软件业务收入 7.06 万亿元，同比增长 11.9%，较 1-5 月回升了 0.7 PCT。行业利润总额 8581 亿元，同比增长 12.0%，较 1-5 月下降了 0.8 PCT，高于收入增速 0.1 PCT。

重点关注子行业的主要数据和动态包括：

(1) AI：8 月发布的 GPT-5 性能虽然有较大提升，价格也下降，但是并不是期待中有“代差”的产品，总体领先竞争对手，但是评分差距有限。7 月阿里开源了升级版模型 Qwen3-235B-A22B-Thinking-2507，成为了全球最强大的开源模型，整体实力居全球大模型第一梯队。2025 年 7 月大模型相关中标项目数量为 574 个，中标项目披露的金额约为 13.35 亿元，同比分别增长 422%和 540%。Q2 财报来看，AI 在广告、云业务等方向上给科技大商带来了较大的经济效益。随着模型能力提升面临发展瓶颈，我们预计后续大厂更多注意力聚焦在 AI 应用落地和算力成本控制方面。

(2) 国产化：2025 年 1-6 月基础软件增长 13.8%，增速连续 4 个月呈现回升趋势，高于软件行业整体增速 1.9 PCT，下半年或将继续受益于“十四五”收官之年的需求释放，增速有进一步加快的可能性。H20 解禁后，仍面临了较多政策不确定性，于此同时，国产 AI 芯片厂商 2025 上半年整体呈现出加速发展趋势，CANN 的开源有助于加快国产 AI 芯片的生态建设。

(3) 算力：2025Q2，美国 6 大科技厂商（亚马逊、微软、谷歌、脸书、甲骨文、苹果）的资本开支再创新高，总计达到 999.73 亿美元，同比增长 77%，增速再创新高，同时海外大型科技厂商的资本开支计划呈现上调趋势。在 AI 业务的带动下，Q2 全球云巨头都呈现了业务加速趋势，甲骨文云基础设施业务预计 2026 财年营收增长超过 70%，即将借力 AI 成

为全球第四大云基础设施服务厂商。

给予行业强于大市的投资评级。在 AI 实现业绩助力的大背景下，海外 AI 投入加大趋势明显。随着头部大模型性能提升面临瓶颈，后续大厂有望投入跟多关注度在 AI 应用和算力成本控制方面，液冷落地有望加快。国产芯片在国际局势的不确定下成为了更优的选项。仍然继续看好近期行业市场表现，建议积极关注 EDA 企业华大九天（301269），有大规模智算中心交付计划的润泽科技（300442），国产算力下布局的大模型厂商科大讯飞（002230）。

4. 风险提示

国际局势的不确定性；下游企业削减开支；半年报业绩的不确定性。

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。