

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

持续掘金国产化、AI、智算三大赛道

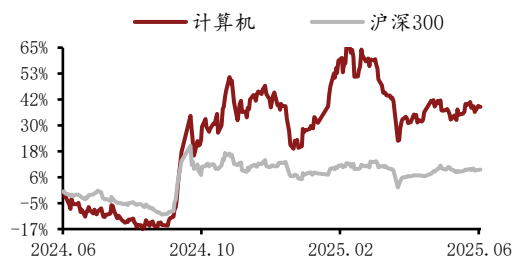
——计算机行业半年度策略

证券研究报告-行业半年度策略

强于大市(维持)

计算机相对沪深 300 指数表现

发布日期：2025 年 06 月 20 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：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

《计算机行业月报：中东加速算力建设，国产大模型或将面临更多生态围堵》

2025-05-19

《计算机行业月报：国产 AI 芯片实现重要突破，积极关注 5 月鸿蒙 PC 问世》

2025-04-17

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：

- 给予行业强于大市的投资评级。2025 年，在全球经济疲弱的大背景下，计算机行业收入、利润增速回升，但是毛利仍在下滑，行业继续呈现出内卷迹象，全年基本面改善或仍将来自于局部赛道。考虑到目前国际局势和科技产业正在迎来的 AI 技术变革，下半年我们继续看好国产替代、AI 技术的深入应用和智能算力需求的释放给企业带来的发展机会。

● 2025 年下半年可以重点关注的方向和机会主要包括：

(1) 国产替代：随着 H20 的受限，英伟达正在计划推出 B30 芯片，性能或将进一步下降。与此同时，国产 AI 芯片数据持续向好，昇腾 CloudMatrix384 推出后，通过规模化设计弥补了芯片的短板，实现了与英伟达超节点的对标。因而，我们认为 AI 芯片已经迎来了产业的拐点，国产芯片商用价值已经凸显。此外，鸿蒙电脑操作系统 5 月发布，是国产基础软件又一个重要突破。随着 EDA 成为美国对我们出口管控重点方向，EDA 的国产替代进程有望全面展开。推荐个股华大九天（301269）、中科曙光（603019）。

(2) 人工智能：2025 年 DeepSeek-R1 的发布作为里程碑式事件，让我国跻身于全球人工智能的第一梯队。随着竞争的加剧，大模型更新频率明显加快，但是 R1 的成本优势仍然明显。2025 年在推理能力提升、成本下降、模型开源三大因素的共同作用下，大模型应用需求呈现爆发性增长。下半年建议重点关注 GPT-5 和 R2 的发布以及相关创新所带来的投资机遇。推荐个股中控技术（688777）、金山办公（688111）、科大讯飞（002230）。

(3) 算力建设：在 AI 推动下，海外科技企业 Q1 的资本投入再创新高，云计算市场总体呈现加速增长趋势。2025 年，国内大规模智算中心或将呈现集中交付趋势，国内外最先进的超节点都应用了液冷技术，推动液冷步入规模化应用。我国算力互联网的积极推进，也将带动算力网络的全面升级。推荐个股润泽科技（300442）。

风险提示：国际局势的不确定性；下游企业削减开支。

内容目录

1. 2025 年行业增速回升，但仍呈现内卷迹象.....	4
1.1. 2025 年行业收入和利润都有改善迹象，但仍呈现内卷迹象.....	4
1.2. 行情回顾：总体好于沪深 300，概念板块多数呈现上涨行情.....	5
1.3. 估值：行业的估值位于历史均值之间.....	6
2. 国产化：鸿蒙、昇腾 CloudMatrix 双重突破，EDA 加速国产化进程.....	7
2.1. EDA：全面加快国产替代进程.....	7
2.2. 操作系统：鸿蒙 PC 在 5 月上市，适配应用数量快速增长.....	9
2.3. 华为芯片性能达到 H200 级别，推出业界最大超节点.....	10
2.4. 出口管制和国产芯片性能提升的双重作用下，AI 芯片迎来国产替代的拐点.....	12
3. AI：AI 应用加快落地，下半年密切关注 R2 和 GPT-5 的发布.....	14
3.1. DeepSeek：密切关注 R2 的发布.....	14
3.2. 大模型更新频率明显加快，GPT5 有望年内推出.....	16
3.3. 2025Q1 AI 应用需求呈现爆发性增长.....	18
3.4. 推理模型将 Scaling 向后训练和推理阶段转移，应用落地加大推理需求.....	19
4. 算力：大规模智算中心将集中交付，算力互联网推动算力网络升级.....	20
4.1. 海外科技企业资本投入再创新高.....	20
4.2. 云计算：市场总体呈现加速增长趋势.....	22
4.3. 数据中心：2025 年大规模智算中心或呈现集中交付趋势.....	22
4.4. 服务器：美国出口管制和国产替代带来行业分化.....	24
4.5. 液冷：国标即将发布，冷板液冷进入快速发展期.....	25
4.6. 算力互联互通积极推进中，将推动中国算力网络的全面升级.....	29
5. 投资建议.....	30
6. 风险提示.....	31

图表目录

图 1：2019-2025 年我国软件业务收入及增速.....	4
图 2：2019-2025 年我国软件业务利润总额及增速.....	4
图 3：2020-2025Q1 计算机行业收入增速对比（亿元）.....	4
图 4：2020-2025Q1 计算机行业净利润增速对比（亿元）.....	4
图 5：2020-2024 年计算机行业员工数量及增速（人）.....	5
图 6：2020-2025Q1 计算机行业毛利增速对比（亿元）.....	5
图 7：2025 年中信一级子行业涨跌幅（截止 2025.6.17）.....	5
图 8：近两年计算机行业相关概念年度涨跌幅（%，2025.6.18）.....	6
图 9：近 10 年中信计算机行业估值水平（截止 2025.6.17）.....	7
图 10：2021 年国内 EDA 市场格局.....	7
图 11：2024 财年新思科技不同区域收入分布（百万美元）.....	8
图 12：2024 年楷登电子不同区域收入分布（百万美元）.....	8
图 13：2018-2025 年 EDA 企业国内业务收入及增速（万元）.....	8
图 14：全球 EDA 产业市场格局.....	9
图 15：2022Q2-2024Q4 中国手机操作系统市场份额.....	9
图 16：2022Q2-2024Q4 全球手机操作系统市场份额.....	9
图 17：鸿蒙电脑应用适配情况.....	10
图 18：CloudMatrix 384 超节点结构.....	11

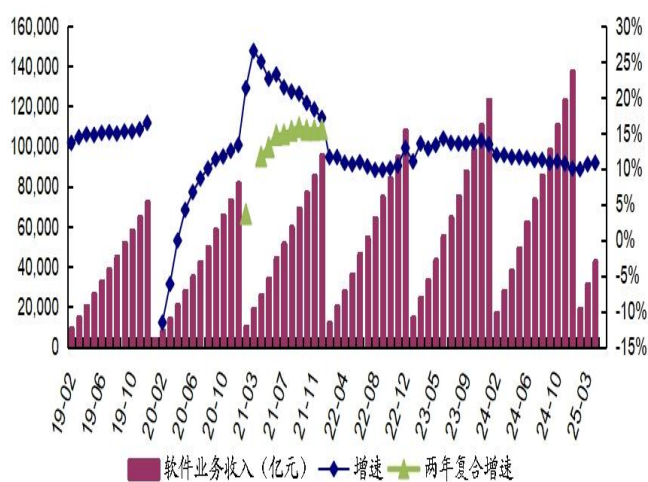
图 19: CloudMatrix 384 超节点与英伟达 GB200 NVL72 对比	11
图 20: CloudMatrix 384 超节点昇腾云服务与英伟达部署情况性能对比	12
图 21: 2022Q1 至 2025Q1 英伟达分区域收入（亿美元）及占比	13
图 22: 2023H1-2024H2 我国 AI 芯片出货及国产化率	14
图 23: 23Q1-25Q1 寒武纪、海光收入（亿元）及增速	14
图 24: DeepSeek-R1-0528 在基准测试中的评分情况	15
图 25: 头部大模型性能对比	17
图 26: 大模型调用价格（美元/百万 tokens）	17
图 27: 豆包大模型调用量最新情况	18
图 28: 2024-2025 年大模型单月份中标项目监测数据（万元）	18
图 29: 2025Q1 大模型订单行业分布（按数量）	19
图 30: 2025Q1 大模型订单行业分布（按金额）	19
图 31: 参数、数据和计算量与 LLM 性能的关系（在 WebText2 上的测试损失）	19
图 32: 模型 Scaling 的新范式	20
图 33: 阿里、腾讯、百度的资本开支（亿元）	21
图 34: 5 大科技厂商资本投入（百万美元）	21
图 35: 2020-2025 年阿里单季度资本开支及增速	21
图 36: 2020-2025 年腾讯单季度资本开支及增速（亿元）	21
图 37: 2019-2024 年我国公有云规模及增速（亿美元）	22
图 38: 2021-2024 年我国公用云细分市场增速	22
图 39: 2019Q4-2025Q1 阿里云收入及其增速	22
图 40: 2017Q1-2024Q4 金山云收入及其增速	22
图 41: 2024 年润泽科技收入分布	23
图 42: 2021-2024 年润泽科技分项业务收入（百万元）	23
图 43: 2020Q1-2025Q1 紫光和浪潮存货增长情况（百万元）	24
图 44: 2022-2029 年中国液冷服务器规模及增速（亿美元）	26
图 45: 2022-2024 年 智算服务细分市场（亿元）	27
图 46: 不同冷却技术与 PUE 相关性	27
图 47: 2024 年中国液冷市场份额	29
图 48: 2023 年中国液冷市场份额	29
图 49: 电力网络与算力网络建设方案对比	29
表 1: DeepSeek 的主要模型发布情况	15
表 2: OpenAI 大模型产品竞品推出情况	16
表 3: IDC 企业订单获取和运力投放情况	23
表 4: 2024 年企业服务器相关业务对比（亿元）	25
表 5: 液冷标准及参与起草单位	26
表 6: 液冷产业链公司业务动态	28
表 7: 重点推荐公司的估值情况（截止 2025.6.18）	31

1. 2025 年行业增速回升，但仍呈现内卷迹象

1.1. 2025 年行业收入和利润都有改善迹象，但仍呈现内卷迹象

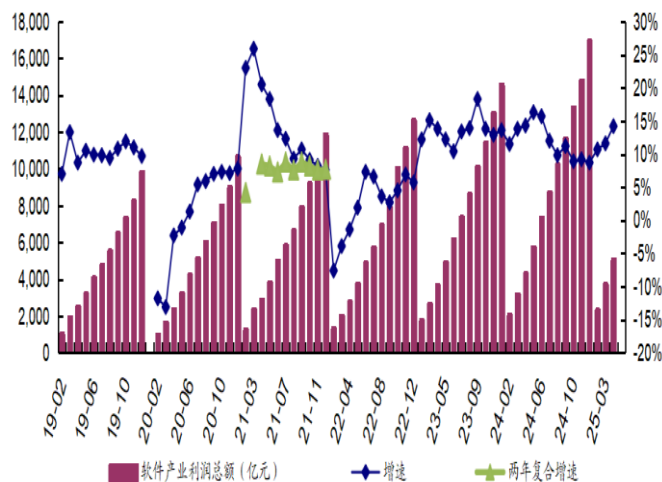
2025 年 1-4 月软件产业增速略有回升，盈利呈改善趋势。根据工信部数据，2025 年 1-4 月软件业务收入 4.26 万亿元，同比增长 10.8%，较 1-3 月回升了 0.2PCT。2025 年 1-4 月软件业务利润总额 5075 亿元，同比增长 14.2%，较 1-3 月回升了 2.6 PCT，高于收入增速 3.4 PCT。

图 1：2019-2025 年我国软件业务收入及增速



资料来源：工信部，中原证券研究所

图 2：2019-2025 年我国软件业务利润总额及增速

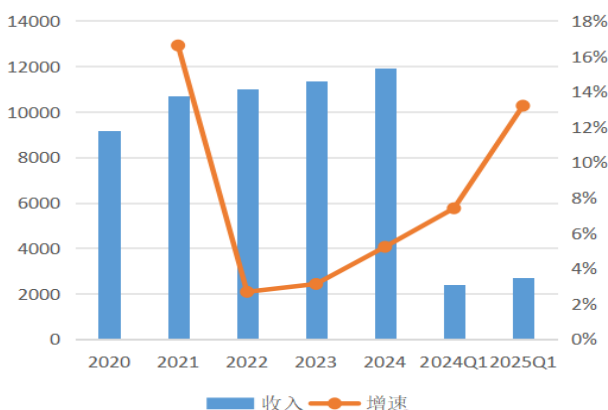


资料来源：工信部，中原证券研究所

财报数据来看，2024 年以来，行业收入增速整体呈现改善趋势。

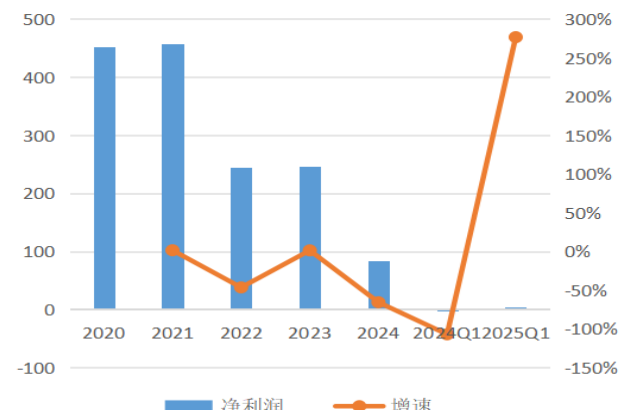
2024 年，中信计算机行业总计收入 1.80 万亿元。考虑到工业富联占比过高（分别占比 2024 年行业总收入和净利润的 34%和 73%，以下行业财报数据中均做剔除处理），将其剔除后行业总计收入 1.19 万亿元，同比增长 5.21%，较上年提升了 2.10 PCT。2025Q1 行业收入 2707 亿元，同比增长 13.20%，较上年提升了 5.80 PCT，继续呈现改善趋势。

图 3：2020-2025Q1 计算机行业收入增速对比（亿元）



资料来源：Wind，中原证券研究所（除工业富联）

图 4：2020-2025Q1 计算机行业净利润增速对比（亿元）



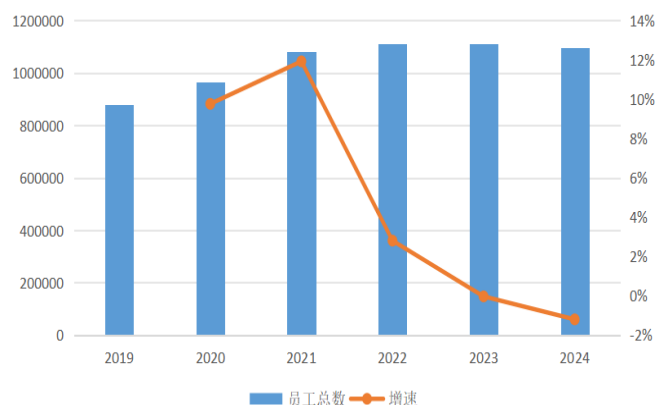
资料来源：Wind，中原证券研究所（除工业富联）

2024 年行业净利润 84.37 亿元，同比下滑了 39.87%，内卷迹象明显。2025Q1，行业净利润 4.91 亿元，实现了扭亏为盈，总体向好。

2024 年行业员工数量 109.74 万人，同比下滑 1.20%，近年来首次下滑，2025 年人员费用压力有望减少。其中有 137 家人员增长，2 家人数不变，224 家人员下滑，最高的人员缩减比例达到 50%。

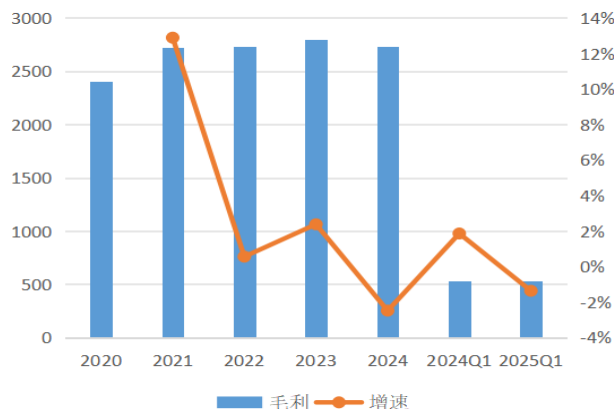
虽然 2025 年行业业绩向好，但是 2025Q1 毛利仍然同比下滑 1.37%，内卷迹象还在延续，企业也在积极关注海外市场拓展。

图 5：2020-2024 年计算机行业员工数量及增速（人）



资料来源：Wind，中原证券研究所（除工业富联）

图 6：2020-2025Q1 计算机行业毛利增速对比（亿元）

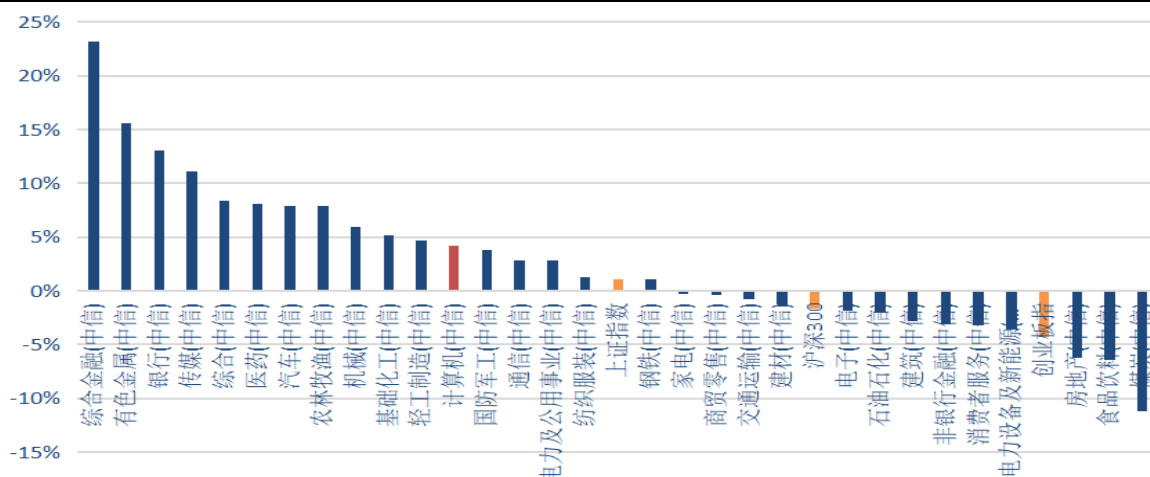


资料来源：Wind，中原证券研究所（除工业富联）

1.2. 行情回顾：总体好于沪深 300，概念板块多数呈现上涨行情

2025 年计算机行业略有上涨，总体表现好于沪深 300。截止 6 月 17 日，中信计算机指数上涨了 4.15%，跑赢上证 3.09 PCT，跑赢沪深 300 指数 5.79 PCT，跑赢创业板指数 8.43 PCT，在 30 个中信一级行业中排名第 12。

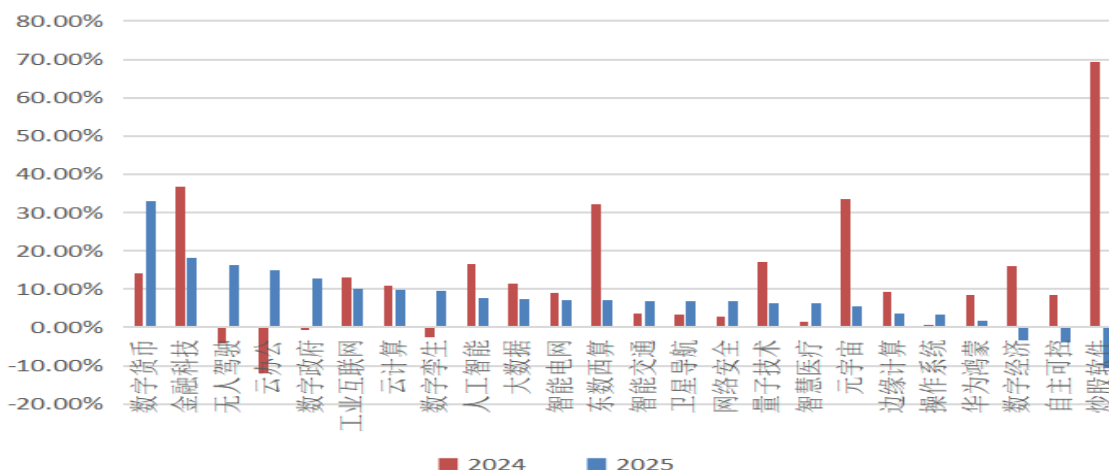
图 7：2025 年中信一级子行业涨跌幅（截止 2025.6.17）



资料来源：Wind，中原证券研究所

从概念指数表现来看，2025 年计算机行业多数概念指数均呈现上涨态势。数字货币受到稳定币利好影响上涨 33.02%，金融科技受到金融股上涨行情带动上涨 18.11%，2025 年比亚迪开启全民智驾带动了无人驾驶概念上涨 16.24%，云办公、数字政府、工业互联网、云计算、数字孪生、人工智能等概念都受到了人工智能应用的利好推动。表现较弱的概念中，2024 年涨幅第一的炒股软件概念下跌 10.75%，自主可控概念下跌 3.95%，数字经济下跌 3.33%。由于当前行业业绩支撑性不强，概念和个股的行情持续性也较弱。

图 8：近两年计算机行业相关概念年度涨跌幅（%，2025.6.18）

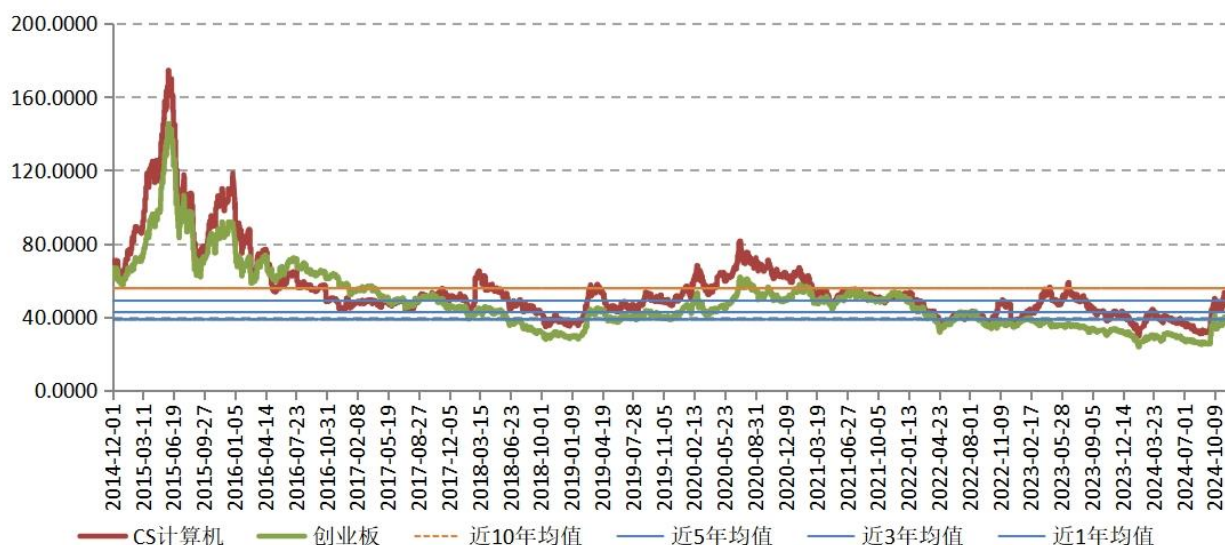


资料来源：Wind，中原证券研究所

1.3. 估值：行业的估值位于历史均值之间

行业的估值位于历史均值之间。根据 Wind 数据，2025 年 6 月 17 日中信计算机行业 TTM 整体法（剔除负值）估值为 46.22 倍，同期 A 股估值 15.27 倍，创业板估值 35.01 倍。行业近 1 年、3 年、5 年、10 年的平均估值分别为 44.17 倍、43.34 倍、47.91 倍、52.61 倍，当前估值仍然位于历史均值水平之间，近 3 年最高点 58.79 倍和最低点 29.69 倍。

图 9：近 10 年中信计算机行业估值水平（截止 2025.6.17）



资料来源：Wind，中原证券研究所

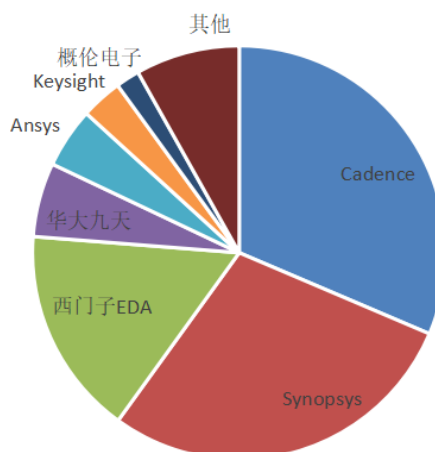
2. 国产化：鸿蒙、昇腾 CloudMatrix 双重突破，EDA 加速国产化进程

2.1. EDA：全面加快国产替代进程

当前，EDA 已经被划入在了美国出口管制的重点范畴。继 2024 年 12 月，华大九天被 BIS 列入实体清单以后，5 月 29 日，美国政府停止对华芯片设计软件（EDA）销售。

美国三大 EDA 巨头占据了 EDA 的绝对垄断地位，也占据了国内 EDA 市场约 8 成的市场份额。根据华经产业研究院的数据，2021 年，我国 EDA 市场中，海外一线产商仍然处于绝对的主导地位，其中楷登电子、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%。

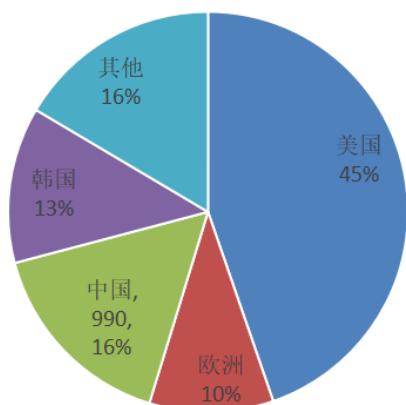
图 10：2021 年国内 EDA 市场格局



资料来源：华经产业研究院，中原证券研究所

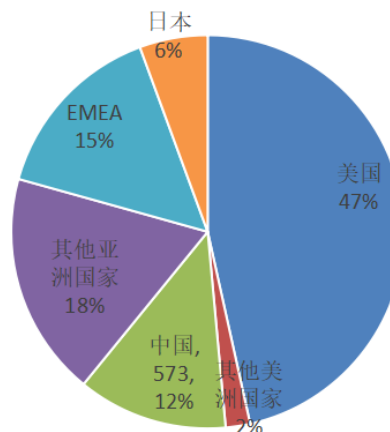
海外 EDA 企业有上百亿收入来自中国。根据相关公司的财报，2024 财年新思科技来自中国收入 9.9 亿美元，占其总收入的 16%；楷登电子 2024 年在中国的销售额为 5.7 亿美元，占其总收入的 12%。

图 11: 2024 财年新思科技不同区域收入分布(百万美元)



资料来源：新思科技财报，中原证券研究所

图 12: 2024 年楷登电子不同区域收入分布(百万美元)

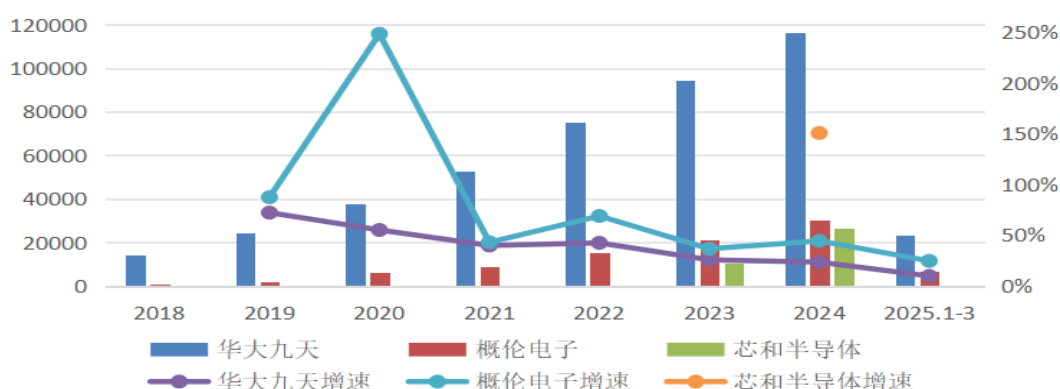


资料来源：楷登电子财报，中原证券研究所

在华为的带领下，国产 EDA 企业已经实现了对国外产品的基本替代。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。但是这一产品体系是建立在华为产品组合基础上的，以满足华为的对内服务为主，无法直接对外销售。

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 13: 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）

2025 年国内 EDA 企业接连开展大额收购，行业集中度持续提升。在华大九天 3 月 30 日发布芯和半导体的收购预案的同时，概伦电子也在 3 月 27 日发布公告，开启对锐成芯微的收购事宜。由于 EDA 产业链长、产品众多，目前国产 EDA 工具仍然不全，行业前十的公司基

本保持着错位竞争，通过收购可以快速进行产品补足，加速发展壮大，同时减少国内企业的重复投入，增加资金利用效率。此外，从 EDA 全球三大巨头的成长路径上来看，其都进行了一系列的收购，才形成了如今的垄断优势，因而对于 EDA 企业来说并购也成为了成长必经路径。

图 14：全球 EDA 产业市场格局

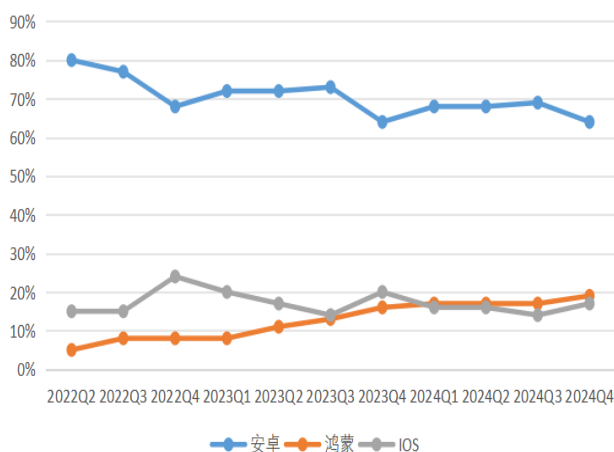


资料来源：赛迪，华大九天公告，中原证券研究所

2.2. 操作系统：鸿蒙 PC 在 5 月上市，适配应用数量快速增长

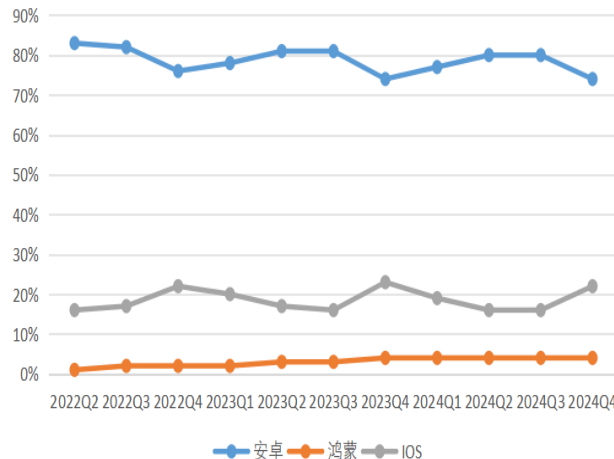
继 2024 年 10 月纯血鸿蒙 HarmonyOS NEXT 发布以后，鸿蒙在 2024 年 Q4 的份额有明显提升趋势。根据 Counterpoint 的数据，2024Q4 鸿蒙以 19% 的份额继续位居国内手机操作系统第二名，较 Q3 提升了 2 PCT，同时鸿蒙在全球继续占据 4% 的市场份额。根据新浪财经 4 月 15 日报道，华为 HarmonyOS NEXT 已经取消了版本回退功能，这预示着手机系统的成熟度已经达到了大规模的商用标准，为整体鸿蒙生态建设做出了重要的突破。

图 15：2022Q2-2024Q4 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 16：2022Q2-2024Q4 全球手机操作系统市场份额



资料来源：Counterpoint，中原证券研究所

继手机端的纯血鸿蒙推出以后，华为也在 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。

图 17：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。而鸿蒙电脑操作系统的正式发布，成为了国产操作系统又一次重要的突破，也意味着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。

2.3. 华为芯片性能达到 H200 级别，推出业界最大超节点

在 2025 年 4 月华为云生态大会 2025 上，华为发布了昇腾超节点 CloudMatrix 384，其的主要参数和特点为：

(1) **结构**：由 384 颗昇腾芯片、12 个计算柜、4 个总线柜组成，是业界最大的超节点，多个 CloudMatrix 384 可以组成十万卡级的 Atlas 900 SuperCluster 超节点集群；

图 18: CloudMatrix 384 超节点结构



资料来源: SemiAnalysis, 中原证券研究所

(2) **算力**: 可提供达 300PFLOPS 的密集 BF16 算力, 接近英伟达超节点 GB200 NVL72 的两倍, 总内存容量超出英伟达的 3.6 倍, 内存带宽 2.1 倍, 为大规模 AI 训练和推理提供了高效的硬件支持;

图 19: CloudMatrix 384 超节点与英伟达 GB200 NVL72 对比

Key Capabilities - Huawei Ascend 910C Cloud Matrix 384 vs Nvidia GB200 NVL72				
Chip and Package Level				
	Unit	GB200	Ascend 910C	Huawei vs Nvidia
BF16 dense TFLOPS	TFLOPS	2,500	780	0.3x
HBM capacity	GB	192	128	0.7x
HBM bandwidth	TB/s	8.0	3.2	0.4x
Scale Up Bandwidth	Gb/s uni-di	7,200	2,800	0.4x
Scale Out Bandwidth	Gb/s uni-di	400	400	1.0x
System Level				
	Unit	Nvidia GB200 NVL72	Cloud Matrix CM384	Huawei vs Nvidia
BF16 dense PFLOPS	PFLOPS	180	300	1.7x
HBM capacity	TB	13.8	49.2	3.6x
HBM bandwidth	TB/s	576	1,229	2.1x
Scale Up Bandwidth	Gb/s uni-di	518,400	1,075,200	2.1x
Scale Up Domain Size	GPUs	72	384	5.3x
Scale Out Bandwidth	Gb/s uni-di	28,800	153,600	5.3x
All-In System Power ¹	W	145,000	559,378	3.9x
All-in Power per BF16 dense FLOP	W/TFLOP	0.81	1.87	2.3x
All-in Power per memory bandwidth	W per TB/s	251.7	455.2	1.8x
All-in Power per memory capacity	kW/TB	10.5	11.4	1.1x

1. All-in System Power is total cluster power including scale-out networking, storage, etc.

资料来源: SemiAnalysis, 中原证券研究所

(3) **部署**: 已经在安徽、内蒙古、贵州部署;

(4) **创新性**: 应用了对等计算架构, 而非传统以 CPU 为中心的冯诺依曼架构。超节点内用**高速总线互联**替代传统以太, 通信带宽提升了 15 倍, 单跳通信时延也从 2 微秒做到 200 纳秒, 使得集群如同一台计算机般协同工作。

(5) 性能测试: CloudMatrix 384 最适合 MoE 模型。在密集型 LLama 3 模型时, 单卡性能能达到 132 TPS, 是传统集群的 2.5 倍。对于通信密集型 MoE 模型, Qwen2-VL 72B 单卡性能能达到 600 TPS, 类 DeepSeek 千亿模型单卡性能能达到 750 TPS, 都实现了传统节点 3 倍以上的提升。

硅基流动联合华为云推出基于 CloudMatrix 384 超节点昇腾云服务和高性能推理框架 SiliconLLM, 并基于此上线 DeepSeek-R1 昇腾云服务, 在保证单用户 20 TPS 水平前提下, 单卡 Decode 吞吐突破 1920 tokens/秒 (TPS), 部署性能比肩 H100。

图 20: CloudMatrix 384 超节点昇腾云服务与英伟达部署情况性能对比



资料来源: 硅基流动, 中原证券研究所

至此, 昇腾在国产 AI 芯片领域实现了重大的突破。5 月 30 日, 英伟达首次公开表示华为芯片技术已经达到 H200 级别, CloudMatrix 的拓展能力甚至超过了英伟达最新一代 Blackwell 架构。尽管单颗昇腾芯片性能约为 B200 的 1/3, 但华为通过规模化设计, 实现整体算力的跃升。在单颗芯片落后一代的情况下, 华为用数学补物理、非摩尔补摩尔、群计算补单芯片, 来满足实用性的要求。2025 年 6 月, 华为又发布了用于大模型训推的数字化风洞系统, 实现小时级预演万卡集群方案, 节省大模型训推过程的时间和算力的浪费。虽然 CloudMatrix 384 在制造成本和电力消耗上仍然还不及 GB200 NVL72, 但是已经完成国产算力的能力对标。由于我国电力供应位居全球第一, 在新能源和特高压技术上全球遥遥领先, 因而可以用电力长板来弥补硬件端的不足。

2.4. 出口管制和国产芯片性能提升的双重作用下, AI 芯片迎来国产替代的拐点

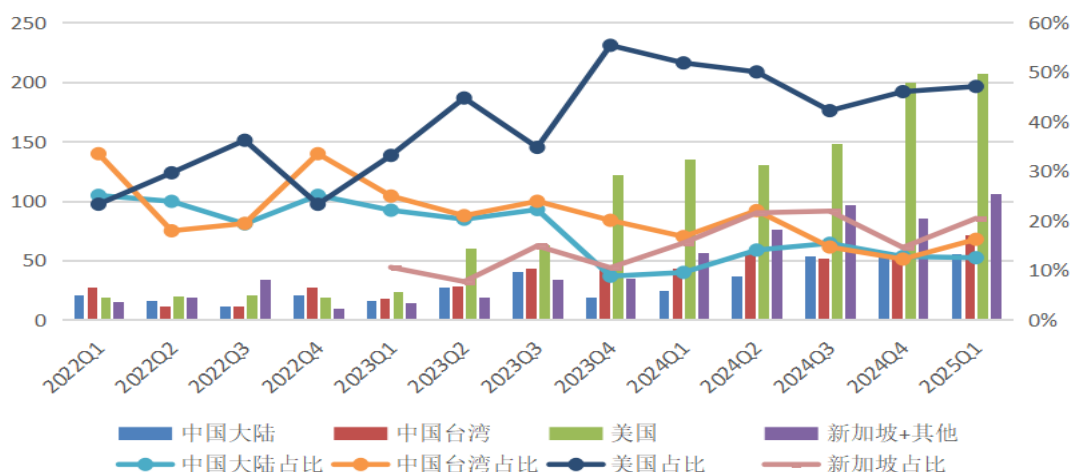
2025 年英伟达对中国供应 AI 芯片 H20 面临新的出口管制。4 月 16 日, 英伟达表示美国政府已经于 4 月 14 日通知公司, H20 芯片未来在出口至中国时需要“无限期”申请许可证。5 月 5 日, 路透社报道称, 一名美国议员计划在未来几周内提出法案, 强制英伟达芯片植入定

位系统，这一政策主要针对芯片走私到其他区域的问题。

面对最新禁令，英伟达又在计划推出新一代对华特供芯片。6月3日，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。

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。

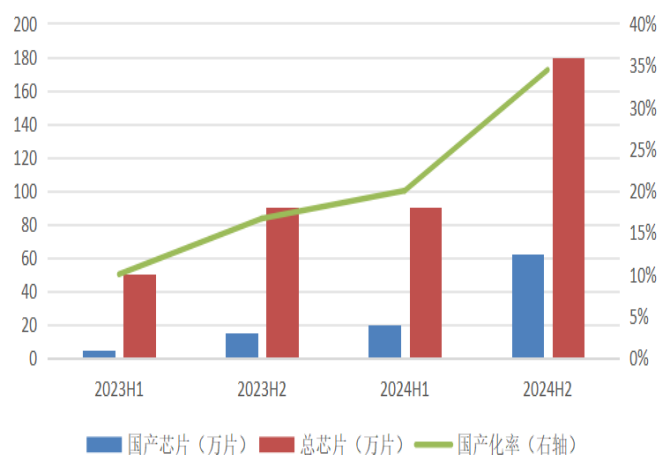
图 21：2022Q1 至 2025Q1 英伟达分区域收入（亿美元）及占比



资料来源：英伟达，中原证券研究所（英伟达财报时段较以上周期晚一个月）

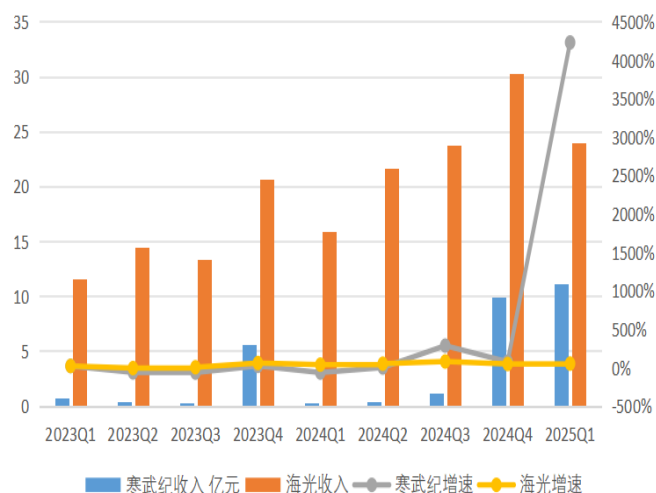
与此同时，国产 AI 芯片数据持续向好。根据 IDC 数据，2024 年下半年，我国 AI 芯片国产化比率从上半年的 20% 进一步提升到了 34%，提升趋势明显。2024 年寒武纪、海光的收入分别为 65.56% 和 52.40%，都实现了高增长。其中寒武纪 2025Q1 财报呈现业绩爆发趋势，收入同比增长 4230.22% 至 11.11 亿元，环比也实现了 12.36% 的增长，很好地印证了当前国产芯片和 AI 算力两大方向的算力需求。2025 年随着海光五号 CPU 和深算三号 DCU 的新品上市，海光四号芯片终端机的量产上市，海光在产品端迎来重要节点。华为超节点 CloudMatrix 384 完成了与英伟达 GB200 NVL72 的能力对标，通过规模化设计弥补了芯片的短板。

图 22：2023H1-2024H2 我国 AI 芯片出货及国产化率



资料来源：IDC，中原证券研究所

图 23：23Q1-25Q1 寒武纪、海光收入（亿元）及增速



资料来源：上市公司财报，中原证券研究所

在国产 AI 芯片能力持续提升的同时，英伟达供应中国芯片性能持续下降，并且我们认为当前 AI 芯片已经迎来了产业的拐点，国产芯片商业价值已经凸显。随着整体替代趋势的形成，国产芯片在生态建设上的短板也将得到弥补，并最终形成与英伟达进行全球市场竞争的实力。

3. AI：AI 应用加快落地，下半年密切关注 R2 和 GPT-5 的发布

3.1. DeepSeek：密切关注 R2 的发布

继年初发布了 R1 以后，DeepSeek 作为中国大模型的代表，追平了我们与全球第一梯队模型的差距，同时其低成本优势和开源策略对于大模型生态形成了较大的冲击，一举一动都成为了市场持续关注的焦点。

5 月 DeepSeek 并没有如期发布 R2，而是进行了 R1 的小版本升级，并实现了较为明显的性能提升。R1-0528 仍然是基于 2024 年 12 月的 DeepSeek V3 Base 模型作为基座，并在后训练中投入了更多算力来提升模型的推理和思维能力。DeepSeek 利用这样的更新方式，既实现了与第一梯队模型的性能对标，同时又可以保持模型生态的兼容性和稳定性。而从模型发展范式来看，R1-0528 也印证了后训练 Scaling 在模型能力提升方面的发展潜力。

图 24: DeepSeek-R1-0528 在基准测试中的评分情况

Benchmarks	DeepSeek-R1-0528	OpenAI-o3	Gemini-2.5-Pro-0506	Qwen3-235B	DeepSeek-R1
AIME 2024 数学竞赛 pass@1	91.4	91.6	90.8	85.7	79.8
AIME 2025 数学竞赛 pass@1	87.5	88.9	83.0	81.5	70.0
GPQA Diamond 科学测试 pass@1	81.0	83.3	83.0	71.1	71.5
LiveCodeBench 代码生成 pass@1	73.3	77.3	71.8	66.5	63.5
Aider 代码编辑 pass@1	71.6	79.6	76.9	65.0	57.0
Humanity's Last Exam 推理与百科知识 pass@1	17.7	20.6	18.4	11.75	8.5

资料来源：DeepSeek，中原证券研究所

后续来看，DeepSeek R2 的推出和多模态功能的实现都值得重点关注，市场对于 DeepSeek 下一代产品带来突破和创新都充满了期待。

4 月 7 日，DeepSeek 与清华大学研究团队联合发布题为《奖励模型的推理时 Scaling 方法及其在大规模语言模型中的应用》的论文，提出自我原则点评调优（SPCT）与元奖励模型（meta RM）两项核心技术，为提升大模型的推理能力提供了全新的方法论，或将为后续 R2 训练提供重要的技术支持。

据太平洋网消息，DeepSeek-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 512 PFLOPS 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%。同时 R2 成本预计将比 GPT-4 低 97%，成本优势明显。

在华为 CloudMatrix 384 推出以前，国产芯片主要在推理方面进行进口替代，R2 如果是在国产芯片上完成训练，将成为国产 AI 芯片发展史上一个里程碑性事件，为推动国产芯片生态建立起到积极的推动作用。

表 1: 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-o4

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

OpenAI 4 月在推出的 o3 和 o4-mini 中，首次将图像深度分析融入思维链，对比来看，DeepSeek-R1 在多模态能力方面仍有较大的提升空间。

3.2. 大模型更新频率明显加快，GPT5 有望年内推出

当前处于第一梯队的大模型分别为 OpenAI 的 o3 和 o4-mini 系列、谷歌的 Gemini 2.5 Pro Preview 0605 Thinking、Anthropic 的 Claude 4 和 DeepSeek 的 DeepSeek-R1-0528。

表 2：OpenAI 大模型产品竞品推出情况

对标模型	时间	说明	厂商	产品	推出时间	备注
o1	2024.9.13	推理模型	Anthropic	升级版 Claude 3.5 Sonnet	2024.10.23	
				Claude 3.7 Sonnet	2025.2.24	首款混合推理模型
			谷歌	Gemini 2.0 Flash Thinking	2024.12.20	
			DeepSeek	DeepSeek-R1-Lite-Preview	2024.11.21	对标 o1-preview
				DeepSeek-R1	2025.1.20	开源
			阿里	QwQ-32B-Preview	2024.11.28	对标 o1-mini 和 o1-preview 开源
				QwQ-32B	2025.3.6	开源
				Qwen3-235B-A22B	2025.4.29	开源
			智谱	GLM-Zero-Preview	2024.12.31	对标 o1-preview
			科大讯飞	讯飞星火 X1	2025.1.15	130B，对标 o1-preview
				讯飞星火 X1 升级	2025.4.20	70B
			xAI	Grok-3 Reasoning Beta	2025.2.18	
				Grok-3 mini Reasoning	2025.2.18	
			字节	Seed - Thinking - v1.5	2025.4.17	200B，支持多模态
			小米	MiMo	2025.4.30	7B，对标 o1-mini，开源
o3	2025.4.17	实现用图像思考	谷歌	Gemini 2.5 Pro Exp	2025.3.25	
				Gemini 2.5 Pro Preview	2025.5.6	
				Gemini 2.5 Pro Preview 0605 Thinking	2025.6.5	
			Anthropic	Claude 4	2025.5.23	
			DeepSeek	DeepSeek-R1-0528	2025.5.28	R1 更新

资料来源：OpenAI，界面新闻，机器之心 Pro，腾讯科技，中原证券研究所

大模型迭代周期缩短，同时基准测试榜单第一名也经常在发生变化。继 5 月 6 日更新以后，谷歌在 6 月 5 日又进行了模型更新，并且在综合能力上位居行业第一。OpenAI 表示将在未来几个月将发布 GPT-5，将对 ChatGPT 免费用户无限量开放，或将成为行业新的标杆。

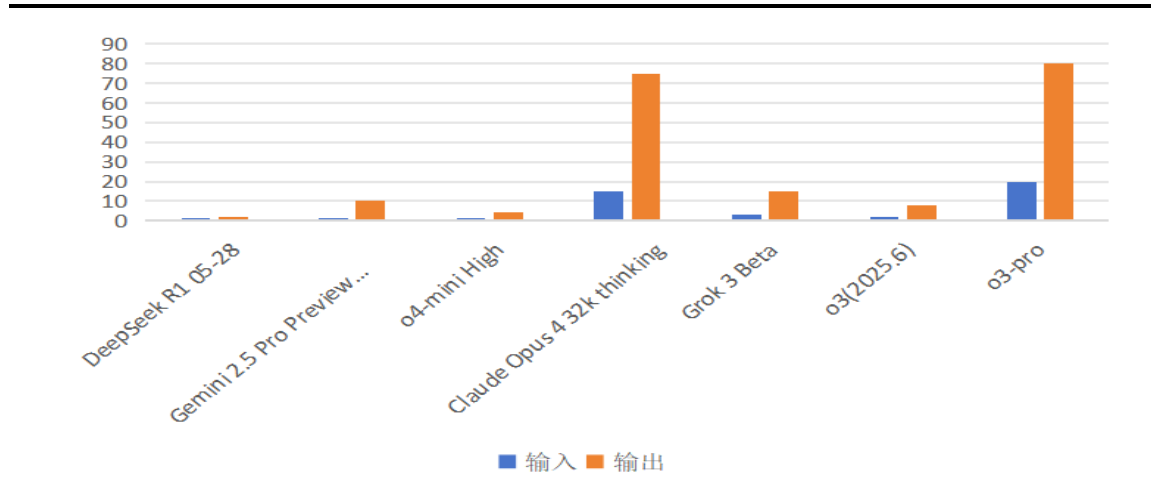
图 25：头部大模型性能对比

Benchmark		Gemini 2.5 Pro Preview 06-05 Thinking	OpenAI o3	OpenAI o4-mini	Claude Opus 4	Grok 3 Beta	DeepSeek R1
Input price	\$/1M tokens no caching	\$1.25 \$2.50 > 200k tokens	\$10.00	\$1.10	\$15.00	\$3.00	\$0.55
Output price	\$/1M tokens	\$10.00 \$5.00 > 200k tokens	\$40.00	\$4.40	\$75.00	\$15.00	\$2.19
Reasoning & knowledge Humanity's Last Exam (no tools)		21.6% 🧐	20.3%	14.3%	10.7%	—	14.0%*
Science GPQA diamond	single attempt multiple attempts	86.4% 🧐 —	83.3% —	81.4% —	79.6% 83.3%	80.2% 84.6%	81.0% —
Mathematics AIME 2025	single attempt multiple attempts	88.0% 🧐 —	88.9% —	92.7% —	75.5% 90.0%	77.3% 93.3%	87.5% —
Code generation LiveCodeBench (UI: 1/1/2025-5/1/2025)	single attempt	69.0% 🧐	72.0%	75.8%	51.1%	—	70.5%
Code editing Aider Polyglot		82.2% 🧐 diff-fenced	79.6% diff	72.0% diff	72.0% diff	53.3% diff	71.6%
Agentic coding SWE-bench Verified	single attempt multiple attempts	59.6% ❤️ 67.2%	— 49.4%	— 68.1%	72.5% 79.4%	— —	— 57.6%
Factuality SimpleQA		54.0% 🧐	48.6%	19.3%	—	43.6%	27.8%
Factuality FACTS Grounding		87.8% 🧐	69.6%	62.1%	77.7%	74.8%	—
Visual reasoning MMMU	single attempt multiple attempts	82.0% 🧐 —	82.9% —	81.6% —	76.5% —	76.0% 78.0%	no MM support no MM support
Image understanding Vibe-Eval (Reka)		67.2% ❤️	—	—	—	—	no MM support
Video understanding VideoMMMU		83.6% ❤️	—	—	—	—	no MM support
Long context MRCR v2 (8-needle)	128k (average) 1M (pointwise)	58.0% 🧐 16.4%	57.1% no support	36.3% no support	— no support	34.0% no support	— no support
Multilingual performance Global MMLU (Lite)		89.2% ❤️	—	—	—	—	—

资料来源：谷歌，澎湃，中原证券研究所

从价格上来看，DeepSeek 总体的调用成本优势仍然明显，输入价格约为 Gemini 2.5 Pro Preview 0605 Thinking 和 o4-mini 的 1/2，输出价格约分别为两者的 1/5 和 1/2，约为 6 月降价以后 o3 输入输出价格的 1/4。

图 26：大模型调用价格（美元/百万 tokens）

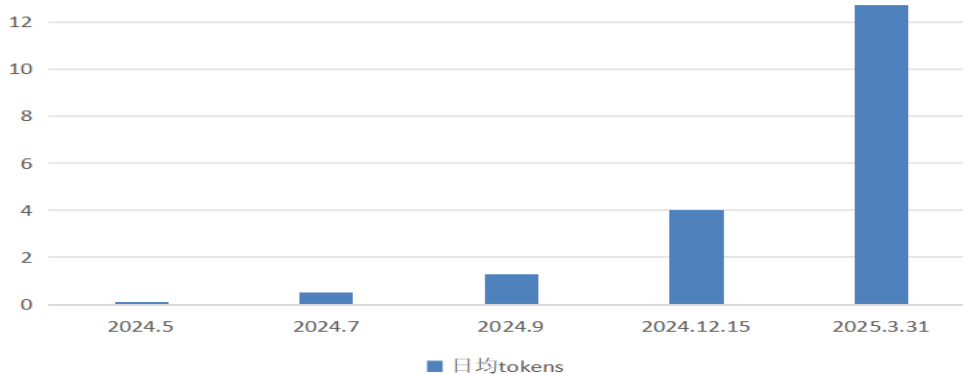


资料来源：谷歌，澎湃，中原证券研究所

3.3. 2025Q1 AI 应用需求呈现爆发性增长

截止 2025 年 3 月底,豆包大模型日均 tokens 调用量已达到 12.7 万亿,较发布之初增长超 106 倍,是 2024 年 12 月 15 日 4 万亿的 3.18 倍。

图 27: 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

根据智能超参数统计的数据,2025Q1 大模型相关中标项目数量为 505 个,其中 121 个项目未披露中标金额,其余 384 个中标项目披露的金额约为 24.67 亿元,中标项目数量和中标金额同比都有 9 倍左右的增长。

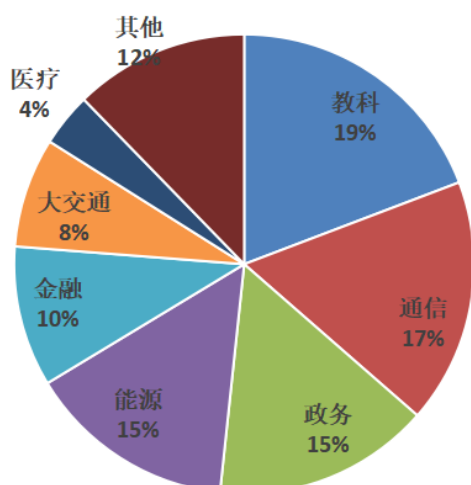
图 28: 2024-2025 年大模型单月份中标项目监测数据 (万元)



资料来源: 智能超参数, 中原证券研究所

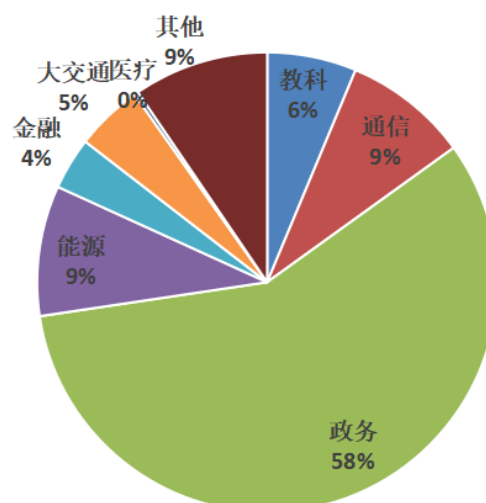
2025Q1 大模型在教科、通信、政务、能源领域都有较多的订单,但是相对而言政务行业订单金额却明显领先于其他行业。

图 29：2025Q1 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 30：2025Q1 大模型订单行业分布（按金额）

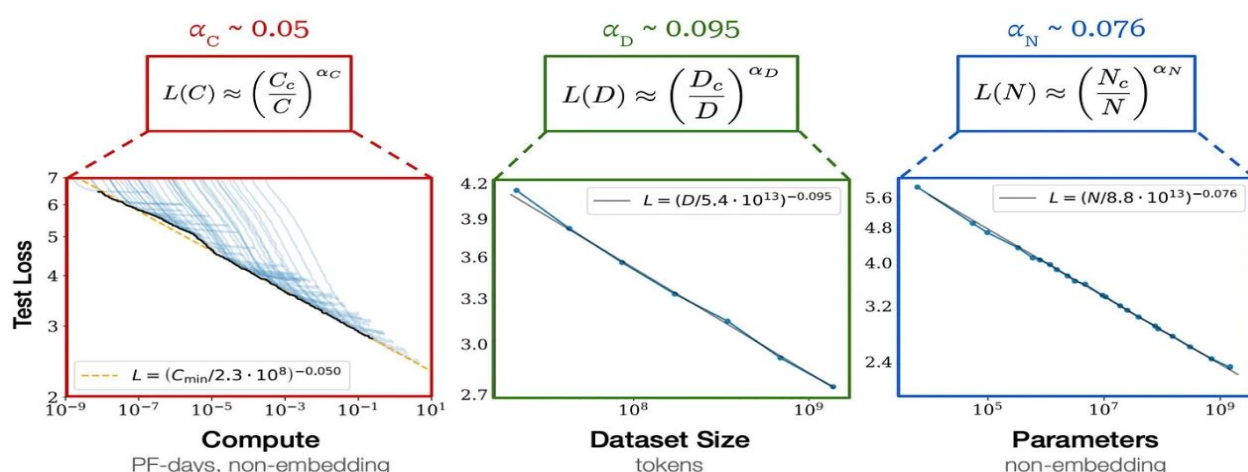


资料来源：智能超参数，中原证券研究所

3.4. 推理模型将 Scaling 向后训练和推理阶段转移，应用落地加大推理需求

预训练的 Scaling Law 带来了大模型的繁荣，但也迎来了发展瓶颈。OpenAI 的成功源于其利用了 Transformer 的所遵循的 Scaling Law（规模效应），在增大模型参数数量、训练数据、用于训练是计算量的同时，带来系统能力稳步的增长。甚至在模型参数达到一定量级，模型能力会得到显著提升，表明较大模型有涌现能力。随着模型持续增大，我们看到 Scaling Law 失效的风险。一方面，高质量数据已经作为训练数据，面临增长瓶颈，另一方面，头部模型厂商在更大参数的新一代模型的推进方面进度不及预期。

图 31：参数、数据和计算量与 LLM 性能的关系（在 WebText2 上的测试损失）



资料来源：Scaling laws for neural language models，机器之心，中原证券研究所

o1 为代表的推理模型的出现，探索出了大模型能力提升的新路径，将算力开销进一步向后训练和推理阶段转移，主要增长因素包括：

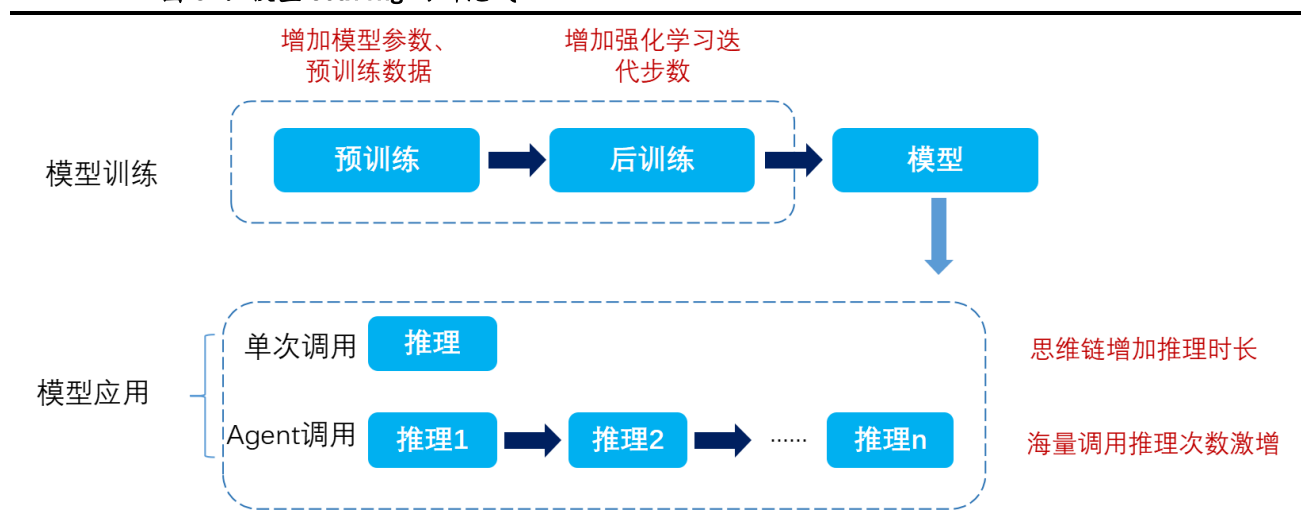
（1）后训练：以 DeepSeek-R1 的训练过程来看，DeepSeek 使用 DeepSeek-V3-Base

作为基础模型，通过强化学习（RL）训练出了 DeepSeek-R1-Zero，并由此产生 DeepSeek-R1。在 DeepSeek-R1-Zero 的训练中，通过增加强化学习的训练迭代步数，其思考的时间在持续增加，AIME 精度也在持续提升，这意味着后训练的 Scaling 成为了提升模型能力的有效路径

（2）**单次推理时长**：推理模型使用思维链，思维链的生成也导致了模型在推理过程中的算力大幅增长。

（3）**推理次数**：DeepSeek - R1 发布后，大模型逻辑推理能力的提升、大模型成本显著下降、开源阵营在性能上实现超越，AI 应用集齐全面落地的三大关键要素。随着 AI 应用进入爆发期，应用端的推理需求将成为下一阶段拉动算力增长的主要驱动力。

图 32：模型 Scaling 的新范式



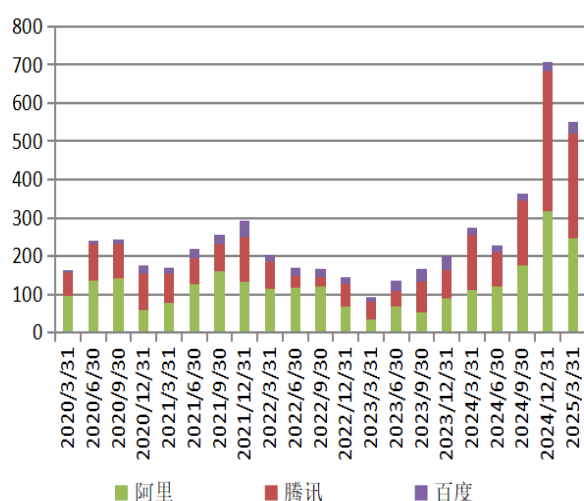
资料来源：中原证券研究所

4. 算力：大规模智算中心将集中交付，算力互联网推动算力网络升级

4.1. 海外科技企业资本投入再创新高

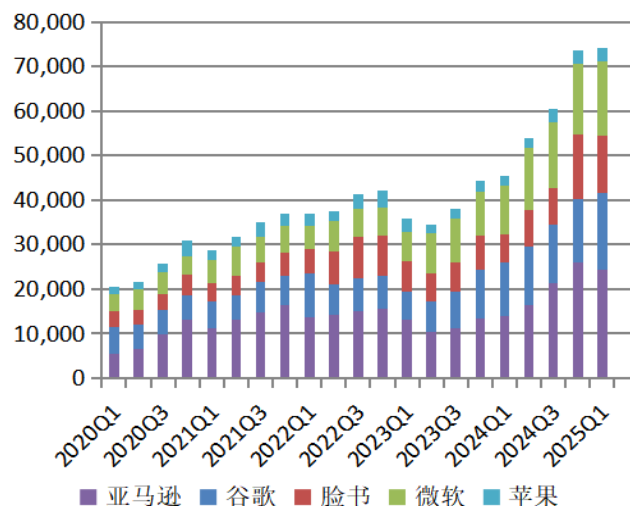
从海外市场来看，2025Q1，美国 5 大科技厂商的资本开支再创新高，总计达到 742.09 亿美元，同比增长 64%，较 2024Q4 略有下滑了 2PCT。其中 Meta 以 102% 的增速领涨，或主要源于 DeepSeek 开源的冲击；亚马逊以 74% 的增速位居第二，苹果以 54% 的增速名列第三，微软 53% 的增速，较上季度下滑了 9 PCT；谷歌增长 43%，表现最弱。

图 33：阿里、腾讯、百度的资本开支（亿元）



资料来源：上市公司财报，中原证券研究所

图 34：5 大科技厂商资本投入（百万美元）

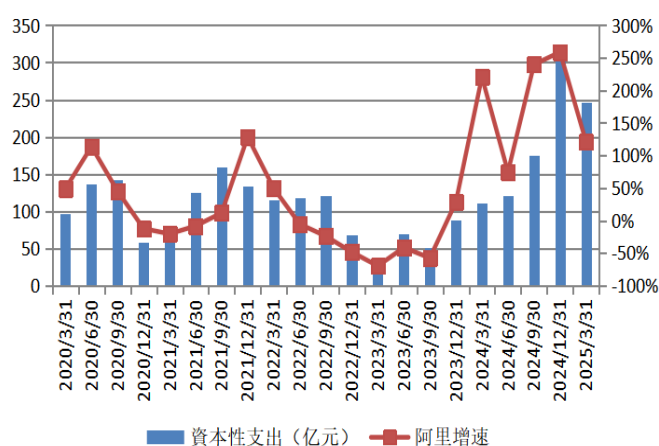


资料来源：Wind，中原证券研究所

Meta 增加了全年 AI 业务的资本开支计划，微软和谷歌资本维持较强的投入计划。微软 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。谷歌仍预计投资 750 亿美元。Meta 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 亿至 650 亿美元之间上调到 640 亿美元至 720 亿美元之间。

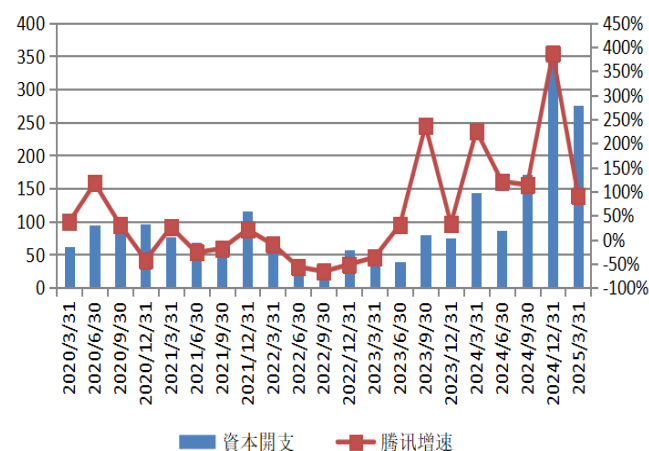
从国内市场来看，Q1 阿里、腾讯的资本开支均呈现环比明显下滑，但是对比 2024Q3 仍有明显增长。腾讯 Q1 资本开支 274.76 亿元，同比增长 91.4%，环比下滑 24.9%，整体位于较高水平。阿里 Q1 资本开支 246.12 亿元，同比增长 120.7%，环比下滑 22.5%。百度 Q1 资本开支 28.69 亿元（不含爱奇艺），同比增长 42.3%，环比增长 24.09%，还没有超过 2023Q4 36.41 亿元的峰值水平，投入相对保守。

图 35：2020-2025 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 36：2020-2025 年腾讯单季度资本开支及增速（亿元）

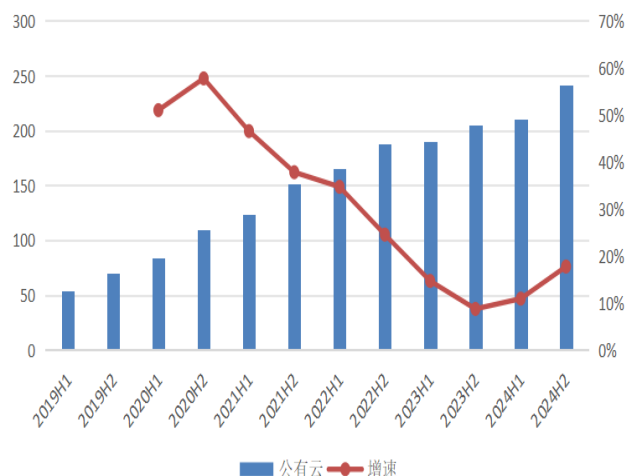


资料来源：腾讯财报，中原证券研究所

4.2. 云计算：市场总体呈现加速增长趋势

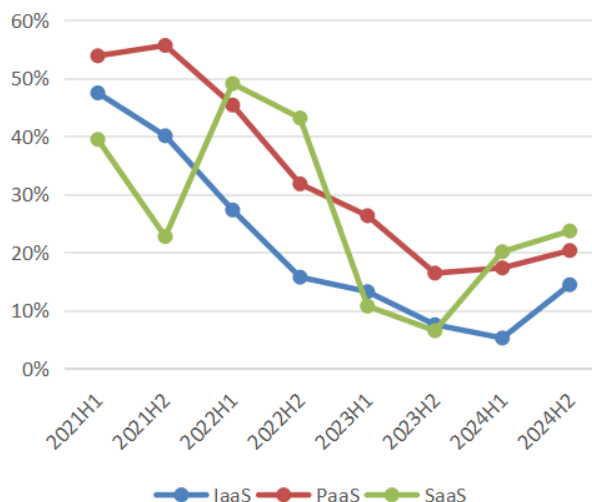
2024 年下半年，我国云计算业务见底回升。根据 IDC 数据，2024H2 我国公有云业务收入 241.1 亿美元，同比增长 17.72%。继 PaaS、SaaS 的增速在 2023H2 见最低点以后，IaaS 增速也在 2024H1 出现阶段性低点 5.2%，2024H2 重回两位数增速至 14.4%。

图 37：2019-2024 年我国公有云规模及增速（亿美元）



资料来源：IDC，中原证券研究所

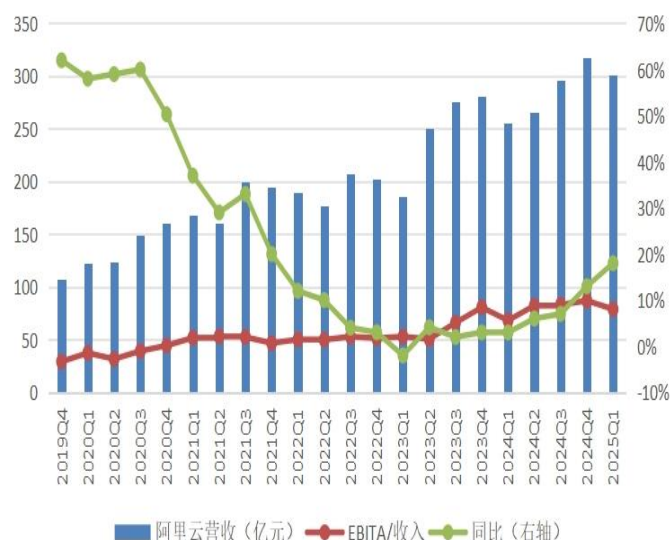
图 38：2021-2024 年我国公用云细分市场增速



资料来源：IDC，中原证券研究所

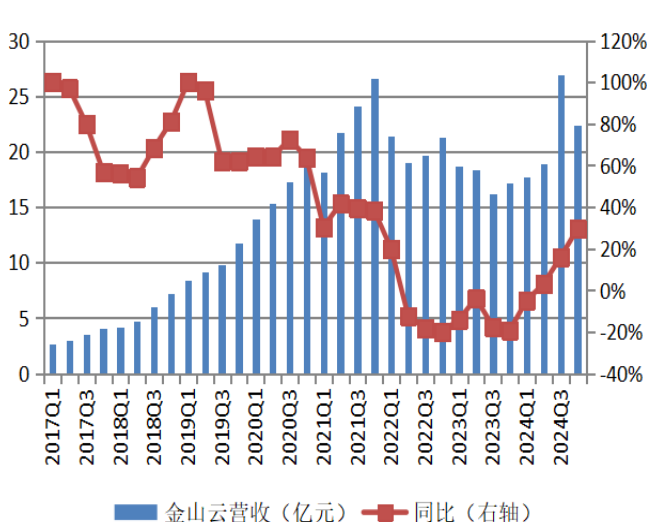
从厂商业绩来看，阿里云和金山云收入增速都从 2024Q1 开始有了明显的改善趋势。其中 2025Q1 阿里云收入增速达到了 18%，较上个季度提升了 5PCT，继续呈现明显加快趋势。

图 39：2019Q4-2025Q1 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 40：2017Q1-2024Q4 金山云收入及其增速



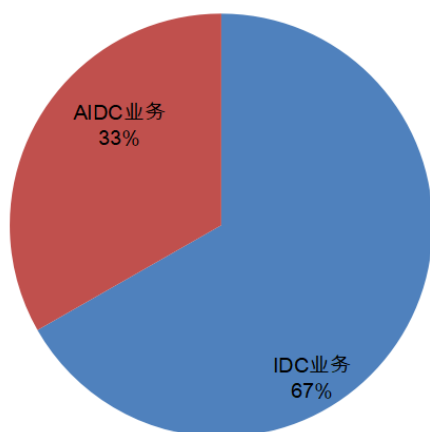
资料来源：金山云财报，中原证券研究所

4.3. 数据中心：2025 年大规模智算中心或呈现集中交付趋势

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

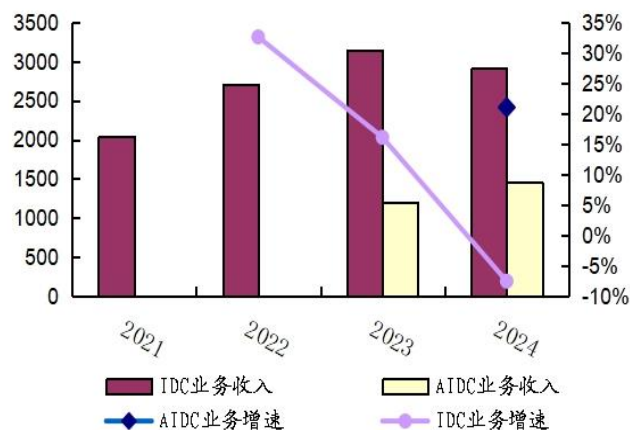
架非常快，2023 年上架当年就实现了 12 亿收入。2024 年由于进行了 AIDC 业务计入方式的调整（从总额法变为净额法），AICD 业务仍然实现了 21%的收入增长（2024H1 公司 AIDC 业务收入在调整后从 20.54 亿元减少到 6.44 亿元，2023 年不做调整）。商汤科技的 AIDC 业务在 2024 年增长了 50%达到了 16.5 亿元，总运营算力达到 23EFLOPS，较上年 12EFLOPS 接近翻倍。

图 41：2024 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 42：2021-2024 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 业务竞争加剧。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。

2025 年随着 AI 应用的普及和大规模智算中心的交付，AIDC 有望迎来业绩加速期，加速数据中心行业的分化。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 年将有较为明显的大规模智算中心集中交付趋势。

表 3：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 年全年预签约容量超过 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

在 AIDC 业务快速上量和 IDC 业务竞争加剧、需求减弱的过程中，数据中心行业或将产生进一步的分化。对于区位、资金、能耗指标、电力供应等优势明显，AIDC 业务转型成功，同时有大规模智算中心的交付的企业来说，2025 年有望迎来较好的增长趋势。

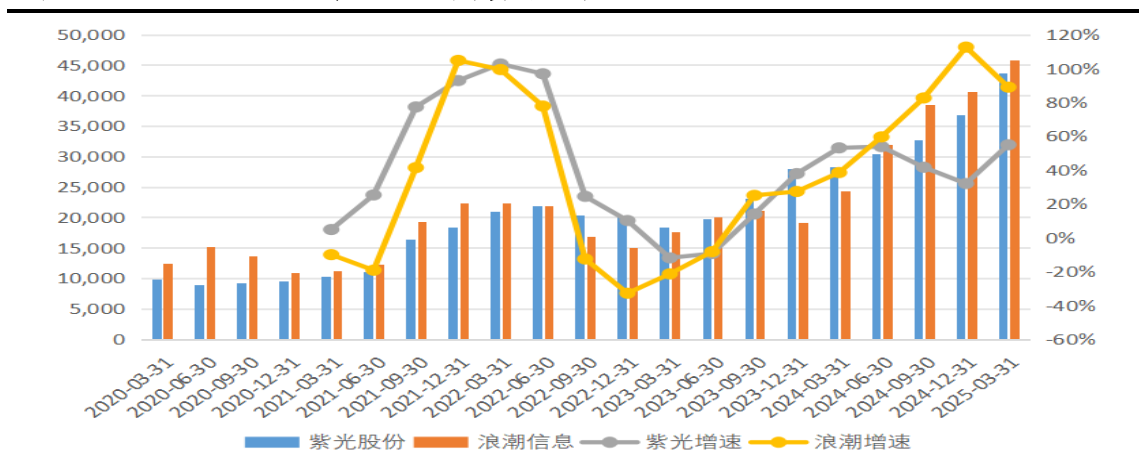
4.4. 服务器：美国出口管制和国产替代带来行业分化

由于 2025 年我国进入到 AI 芯片产业的拐点，我们认为行业分化将加剧显现。

针对国内 AI 产业和服务器行业的禁令再次出台。3 月 25 日，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，加速国内服务器企业的分化。4 月英伟达表示 H20 芯片未来在出口至中国时需要“无限期”申请许可证，6 月报道称英伟达将为中国研发性能或进一步下滑的 B30 芯片，这意味着英伟达芯片对国内厂商的吸引力将进一步下滑。

服务器厂商加大芯片备货力度，禁令对短期业务冲击有限，但长期仍有较大不确定性。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 43：2020Q1-2025Q1 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

从 2024 年服务器行业业绩来看，不同阵营的企业之间的分化也更加明显。受益于海外科技大厂 2024 年资本开支增长快于国内，对标海外需求的工业富联增长最为明显。浪潮作为国内互联网厂商服务器的最大供应商，2024 年服务器业务增幅明显，但是毛利率显著下滑。国产服务器供应商分化明显，超聚变作为第一大国产服务器供应商，2024 年增速明显提升，而神州数码的自有品牌业务下半年销售额不及上半年，增速不及预期。

2025Q1 浪潮继续获得了 165.31% 的高增速，但是毛利率进一步下滑到了 3.45%，成为历史新低，显示出 Q1 国内算力需求放量的迹象，同时服务器端竞争进一步加剧。随着 H20 步入存货消耗阶段，Q2 行业盈利情况有望改善。

表 4：2024 年企业服务器相关业务对比（亿元）

公司	相关业务	2024 年							2023 年
		收入	增速	净利润	增速	毛利率	毛利额	人员增速	收入
工业富联	云计算	3194	64%						1943
	AI 服务器		150%						
浪潮信息	-	1148	74%	23	29%	6.9%	78.6	-3%	659
联想集团	基础设施解决方案	\$129	51%	-1.69	-17%				\$86
超聚变	-	400	41%						284
新华三	计算存储								328
中科曙光	IT 设备业务	117	-8%			27.3%	32.0		127
宝德		100	8%	2.35	17%				92.7
中兴通讯	服务器及存储								100*
清华同方	（软通动力持股）								46*
软通动力	计算产品与智能电子	129.2							
软通计算机	（软通动力持股，原同方）								46*
华鲲振宇									34*
神州数码	自有品牌产品	45.8	20%	0.71		11.5%	5.3		38
拓维信息	计算机、通信和其他电子设备制造业	20.7	17%			6.8%	1.4		17.8
湘江鲲鹏	（拓维信息持股）	15.8	2%	0.014	162%				15.6

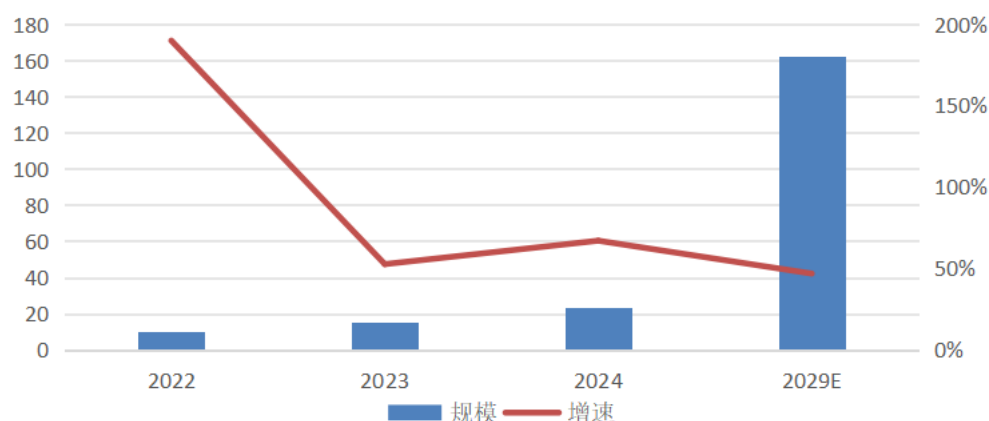
资料来源：上市公司公告，Wind，大河网，中原证券研究所（*为 2022 年数值）

由于 CloudMatrix 384 的发布和国产芯片性能的持续提升，2025 年国产 AI 芯片进口替代的商业意义更加突出，将推动下游服务器行业的生态转型。5 月海光信息公告了拟吸收合并中科曙光的事宜，这意味着在芯片国产化的大背景下，芯片和服务器等上下游深度绑定和推广应用的趋势更加明显。

4.5. 液冷：国标即将发布，冷板液冷进入快速发展期

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。根据 IDC 数据，2024 年中国液冷服务器市场规模达到 23.7 亿美元，同比增长 67.0%，冷板式解决方案市场占有率进一步提高。IDC 预计 2024-2029 年，中国液冷服务器市场年复合增长率将达到 46.8%，2029 年市场规模将达到 162 亿美元。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。

图 44：2022-2029 年中国液冷服务器规模及增速（亿美元）



资料来源：IDC，中原证券研究所

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。

表 5：液冷标准及参与起草单位

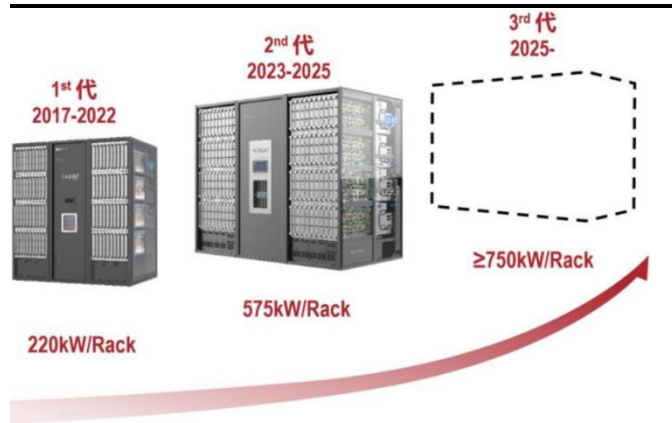
时间	标准	简介	起草单位
2022.4	《数据中心液冷服务器系统总体技术要求和测试方法》（标准编号 YD/T 4024-2022）	行业标准，适用于数据中心液冷系统的设计、施工、部署、运维等环节的技术指导。	信通院、百度、腾讯、阿里巴巴、三快科技、中国电信、中国移动、3M、绿色云图、华为、联想、浪潮信息（000977）、中兴通讯（000063）、IBM、新华三（紫光股份（000938）控股）、施耐德、曙光节能、合一新材料研究院
2024.10	《冷板式液冷整机柜服务器技术要求和测试方法》（标准编号 YD/T 6049-2024）	首个液冷整机柜服务器领域的行业标准，规定了冷板式液冷整机柜服务器的机柜、供液组件、供配电组件、计算节点、交换节点、整机柜内组网、机柜管理模块技术方面的技术要点。	由信通院牵头制定，超聚变、中国移动（600941）、中国电信（601728）、中国联通（600050）、华为、中兴通讯（000063）、浪潮信息（000977）、曙光数创（872808）、新华三（紫光股份（000938）控股）、中移设计研究院、英特尔、中天宽带、宁畅、英维克（002837）、昆仑技术（超聚变子公司）、兰洋科技、云边低碳技术、中通服咨询、秦淮数据、华鲲振宇、公诚咨询、南方电网大数据服务参编
征求意见中	《数据中心冷板式液冷系统技术规范》	我国液冷算力基础设施领域首批 国家标准 ，规定了数据中心冷板式液冷系统冷却工质、液冷机柜子系统、冷却环路、冷量分配单元、液冷智能管理子系统、室外冷源设备等技术要求及测试方法。适用于数据中心冷板式液冷系统的设计、部署和运维。可填补国内冷板式液冷系统的国家标准空白，推动冷板式液冷系统解耦交付，降低冷	主要起草单位曙光数创（872808），中国电子技术标准化研究院、清华大学、中科赛能参编

		板式液冷系统生产、部署成本，提升冷板式液冷系统的稳定性、可靠性，并可为液冷算力基础设施领域高质量发展提供标准化路径参考。	
征求意见稿	《数据中心冷板式液冷机柜通用规范》	我国液冷算力基础设施领域首批 国家标准 ，适用于数据中心单相冷板式液冷机柜的设计、制造、部署和运维。	主要起草单位中国电子技术标准化研究院，移动设计院(600941)、北京工业大学、曙光数创(872808)参编
起草中	《信息技术 服务器及存储设备用液冷部件技术规范》	国内首个液冷部件 国标 ，适用于服务器及存储设备用液冷系统中液冷部件的设计、制造和应用。旨在推动冷板组件、浸没组件、冷量分配单元(CDU)、连接系统等液冷核心部件在加工工艺、规格尺寸、性能及可靠性等层面达到统一标准	主要起草单位浪潮信息(000977)，中国电子技术标准化研究院、英特尔、联想、申菱环境(301018)、珂垓冷却、比赫电气、英维克(002837)参编
起草中	《液冷数据中心快速接头通用规范》	行业标准	曙光数创(872808)、中国电子技术标准化研究院、清华大学
起草中	《液冷数据中心冷板式冷量分配系统通用规范》	行业标准	移动设计院(600941)、中国电子技术标准化研究院、清华大学

资料来源：中国电子技术标准化研究院，全国标准信息公共服务平台，工信部，中原证券研究所

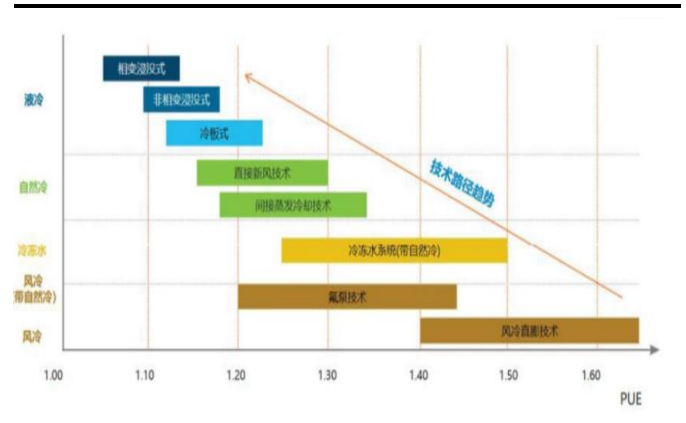
在浸没式液冷方面，目前主要用于超算中心，具有高定制化特征。曙光数创作为国内唯一实现浸没相变液冷技术大规模商业化部署的企业，在 2024 年推出了第三代 C8000 浸没液冷解决方案，单机柜功率密度突破至 750kW 以上，为未来算力芯片性能的持续提升和数据中心更高的节能需求提供了技术储备。

图 45：2022-2024 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 46：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。和英伟达 GB200 NVL72 一样，华为在 CloudMatrix 384 中也应用了液冷技术，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。

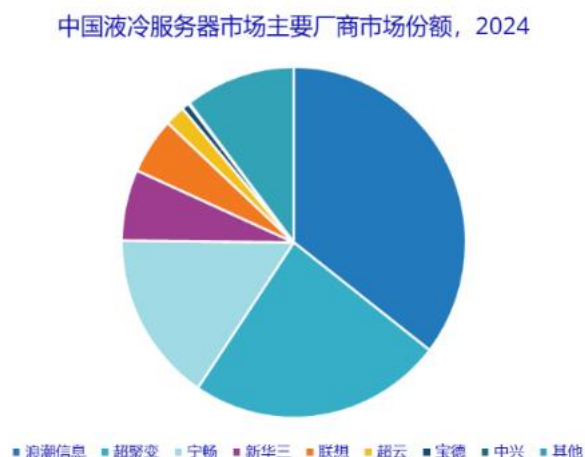
表 6：液冷产业链公司业务动态

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	截止 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，实现全线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

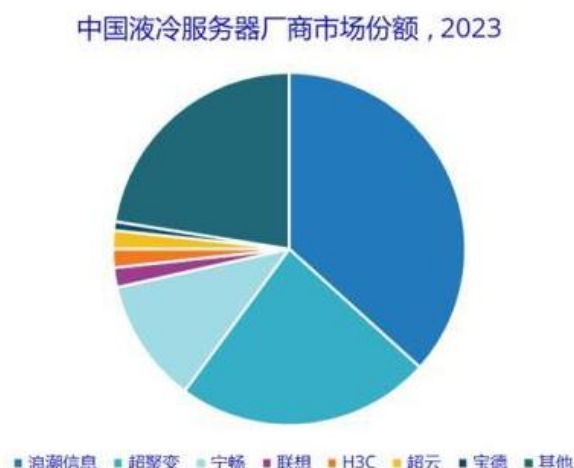
2024 年液冷服务器市场集中化趋势更加明显。前三大厂商分别是浪潮信息、超聚变和宁畅，占据了七成左右的市场份额。对比 2023 年来看主要变化表现为，新华三超过联想成为行业第 4，前五大厂商中宁畅、新华三、联想的份额增幅都较为明显。

图 47：2024 年中国液冷市场份额



资料来源：IDC，中原证券研究所

图 48：2023 年中国液冷市场份额



资料来源：IDC，中原证券研究所

4.6. 算力互联互通积极推进中，将推动中国算力网络的全面升级

在东数西算实现算力节点和数据枢纽资源合理配置的基础上，中国算力网的建设也在持续展开，意图将西部的算力通过网络高效输送到东部，将分散在各地的算力资源进行连接，实现跨区域调度和使用，更好地解决当前各地算力资源利用不均衡的问题，提升使用效率、降低使用成本，最终建立全国一体化算力网，让用户使用算力时像用电一样方便。

图 49：电力网络与算力网络建设方案对比



资料来源：人工智能与金融未来主题峰会，中原证券研究所

鹏城实验室牵头了“中国算力网”研究计划，2022 年 6 月，一期工程智算中心正式上线，跨领域纳管了 20 多家公共智算中心，汇聚算力 5EFLOPS。2023 年 10 月，由韶关市和鹏城实验室共建的中国算力网粤港澳大湾区调度中心上线，是我国算力网项目的首个区域级资源调度中心。

5 月 17 日，信通院与三大运营商联合启动“算力互联网试验网”的建设，并发布了《算力

互联网架构 1.0》。

5月30日，工信部印发了《算力互联网互通行行动计划》，提出要以“算力互联互通为主线，推动建设算力互联网，逐步培育形成全国统一算力服务市场”。根据行动计划，到2026年，我国将建立较为完备的算力互联互通标准、标识和规则体系；推广新型高性能传输协议，提升算力节点间网络互联互通水平；建成国家、区域、行业算力互联互通平台；推动算、存、网多种业务互通，实现跨主体、跨架构、跨地域算力供需调度；开展算力互联网试验网试点，赋能产业普惠用算。到2028年，基本实现全国公共算力标准化互联，逐步形成具备智能感知、实时发现、按需获取的算力互联网。

在中国算力网发展过程中的关键技术挑战主要包括：（1）核心算力的供给；（2）高通量、大带宽、低延迟的网络，预计实现东数西算所有集群节点100Tbps低延时直连通信；（3）算力调度挑战，分布式实现超算、智算、云计算等多种算力在云网的调度，最终实现异构算力混合跨地域共享；（4）零排放算力。

5. 投资建议

给予行业强于大市的投资评级。2025年，在全球经济疲弱的大背景下，计算机行业收入、利润增速回升，但是毛利仍在下滑，行业继续呈现出内卷迹象，全年基本面改善或仍将来自于局部赛道。考虑到目前国际局势和科技产业正在迎来的AI技术变革，下半年我们继续看好国产替代、AI技术的深入应用和智能算力需求的释放给企业带来的发展机会。

2025年下半年可以重点关注的方向和机会主要包括：

（1）国产替代：随着H20的受限，英伟达正在计划推出B30芯片，性能或将进一步下降。与此同时，国产AI芯片数据持续向好，昇腾CloudMatrix384推出后，通过规模化设计弥补了芯片的短板，实现了与英伟达超节点的对标。因而，我们认为AI芯片已经迎来了产业的拐点，国产芯片商用价值已经凸显。此外，鸿蒙电脑操作系统5月发布，是国产基础软件又一个重要突破。随着EDA成为美国对我们出口管控重点方向，EDA的国产替代进程有望全面展开。**推荐个股华大九天（301269）、中科曙光（603019）。**

（2）人工智能：2025年DeepSeek-R1的发布作为里程碑式事件，让我国跻身于全球人工智能的第一梯队。随着竞争的加剧，大模型更新频率明显加快，但是R1的成本优势仍然明显。2025年在推理能力提升、成本下降、模型开源三大因素的共同作用下，大模型应用需求呈现爆发性增长。下半年建议重点关注GPT-5和R2的发布以及相关创新所带来的投资机遇。**推荐个股中控技术（688777）、金山办公（688111）、科大讯飞（002230）。**

（3）算力建设：在AI推动下，海外科技企业Q1的资本投入再创新高，云计算市场总体呈现加速增长趋势。2025年，国内大规模智算中心或将呈现集中交付趋势，国内外最先进的超节点都应用了液冷技术，推动液冷步入规模化应用。我国算力互联网的积极推进，也将带动算

力网络的全面升级。推荐个股润泽科技（300442）。

表 7：重点推荐公司的估值情况（截止 2025.6.18）

代码	简称	股价	总市值 (亿元)	EPS		PE		PEG		PB	评级
				2024E	2025E	2024E	2025E	2024E	2025E		
301269	华大九天	117.69	639	0.40	0.57	297.42	207.02	3.09	4.74	12.82	买入
688206	概伦电子	26.05	113	-0.02	0.07	-1223.0	394.70	-13.54	0.96	5.72	未评级
603019	中科曙光	67.40	986	1.68	2.05	40.08	32.89	1.39	1.50	4.98	增持
688777	中控技术	43.15	341	1.65	1.96	26.08	22.05	1.52	1.20	3.46	买入
688111	金山办公	273.91	1,269	4.15	5.05	65.92	54.29	3.88	2.53	10.72	增持
002230	科大讯飞	47.69	1,102	0.41	0.59	115.95	80.93	1.66	1.87	6.17	增持
300442	润泽科技	44.15	760	1.56	1.93	28.38	22.84	0.57	0.94	7.32	买入
300738	奥飞数据	19.13	188	0.22	0.34	86.84	55.79	1.16	1.00	5.51	未评级

资料来源：Wind 一致预期，中原证券研究所

6. 风险提示

国际局势的不确定性；下游企业削减开支

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。