

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

DeepSeek-R1 带来 AI 技术突破，持续关注 GPT-5 的推出进程

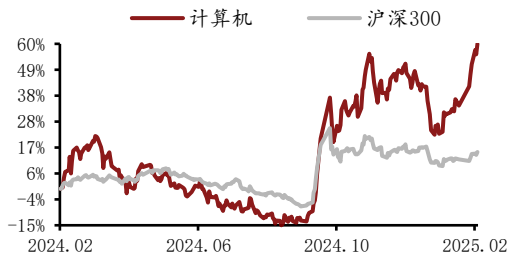
——计算机行业月报

证券研究报告-行业月报

同步大市(维持)

计算机相对沪深 300 指数表现

发布日期：2025 年 02 月 13 日



资料来源：中原证券研究所，聚源

相关报告

《计算机行业深度分析：美国制裁密集出台，全面围堵对我国的 AI 算力供应能力》
2025-01-24

《计算机行业月报：AI 模型成果密集发布，制裁或将全面升级》 2025-01-10

《计算机行业深度分析：经济工作会议聚焦人工智能+，地方发力产业建设》 2024-12-20

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：

- **2024 年我国软件产业增速呈现回落趋势。**2024 年软件业务收入 13.73 万亿元，同比增长 10.0%，较 2023 年 13.4% 的增速回落了 3.4PCT。2024 年软件业务利润总额 16953 亿元，同比增长 8.7%，较 2023 年 13.6% 的增速水平回落了 4.9 PCT。
- **重点关注子行业的主要数据和动态包括：**
 - (1) **国产化：**2024 年，我国集成电路在进口依赖度比例为 78%（即国产化占比 22%），较 1-11 月回落 2 PCT，下滑趋势明显加大，表明美国一系列禁令影响在持续释放。阿斯麦对中国出口量的快速上升，意味着未来国内芯片产量有望得到较大增长。华为 2024 年收入重回美国制裁前的水平，研发投入增长明显。
 - (2) **AI：**R1 在后训练阶段大规模使用了强化学习，来取代传统大模型中监督式微调的方法，不需要大量的标注数据，节省了算力，同时实现了与 o1 相媲美的性能。这意味着在 Scaling Law 失效的困扰下，人类在增加训练数据和算力以外，探索出了高效的架构与训练策略提升的新路径，这或将成为推动 OpenAI 的 GPT-5 问世的一个最重要的技术突破。
 - (3) **算力：**2024Q4，美国 5 大科技厂商的资本开支再创新高，总计达到 734.97 亿美元，同比增长 66%，较 Q3 继续提升了 7PCT。DeepSeek-R1 发布后，亚马逊、谷歌、Meta 仍然给出了 2025 年积极的资本投入计划，同时海内外云厂商都快速接入了 DeepSeek 的模型，国产芯片厂商也快速实现了适配。浙江省即将推出人工智能专项规划，力争 2025 年全省算力规模超过 100EFlops，加快趋势明显。
- **给予行业同步大市的投资评级。**2025 年 2 月 11 日中信计算机行业估值为 53.06 倍，经过节后一轮上涨，当前估值仍然位于历史均值水平之间，但是已经接近估值均值区间上限。由于 DeepSeek-R1 的技术突破，中国在大模型领域的的能力获得了全球的认可，也为后续大模型的发展探索出了一条新的提升路径。因为 DeepSeek-R1 采用开源策略，极大地加快国产芯片的适配，加大下游应用能力的释放。我们认为 DeepSeek-R1 虽然对算力需求降低，但是随着应用端的推理需求大量释放，全球算力需求还将持续快速增长。考虑到 GPT-5 即将发布和 AI 应用的持续落地，我们认为板块还将持续获得市场的关注。

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. 行业数据	4
1.1. 行业数据：全年收入利润增速略有回落，出口业务实现增长	4
1.2. 2024 年高景气赛道：IC 设计、嵌入式系统软件、电子商务平台服务	5
1.3. 国产化：国产化持续深入，美国出台新禁令	6
1.3.1. 拜登政府 1 月制裁措施	6
1.3.2. 芯片国产化比例加快提升，光刻机环节或有转机	9
1.3.3. 华为财报：2024 年收入重回美国制裁前的水平，研发投入增长明显	11
1.4. AI：DeepSeek-R1 出世，GPT-5 即将推出	12
1.4.1. OpenAI 即将推出 GPT-5	12
1.4.2. 国内外模型差距持续缩小	15
1.4.3. 针对 DeepSeek 的制裁和攻击	16
1.4.4. 近期主要大模型厂商进展	17
1.4.5. DeepSeek-R1 开发模式的创新	19
1.5. 算力：DeepSeek 发布加速芯片国产化，AI 应用将带来更多算力需求	20
1.5.1. DeepSeek-r1 发布以后，头部科技厂商仍然给出了 2025 年积极的投资计划	20
1.5.2. 行业加速接入 DeepSeek 大模型，算力供应的生态加速繁荣	21
1.5.3. 浙江算力规划也呈现加快趋势	23
1.5.4. 《国家数据基础设施建设指引》发布，当前仍处于起步建设阶段	26
1.5.5. 百度昆仑芯的万卡集群落地	27
2. 新闻与公告	29
3. 河南计算机行业动态	30
3.1. 河南计算机行业要闻	30
3.2. 河南计算机行业数据跟踪	31
3.3. 河南上市公司行情回顾	31
4. 投资策略	32
4.1. 行情回顾：1 月行业略跑赢沪深 300 指数	32
4.2. 估值：行业估值接近历史均值区间上限	33
4.3. 行业观点与投资建议	34
5. 风险提示	34

图表目录

图 1：2019-2024 年我国软件业务收入及增速（月度累计值）	4
图 2：2019-2024 年我国软件业务利润总额及增速（月度累计值）	4
图 3：2019-2024 年我国软件业务出口数据	5
图 4：2021 至 2024 年我国电子产品产量增速（累计同比）	5
图 5：2024 年和 2023 年我国软件业务子行业增速对比	6
图 6：美国芯片限制的区域划分	7
图 7：2022Q1-2024Q3 英伟达单季度收入的区域分布（亿美元）	8
图 8：2021-2024 年我国集成电路对进口依赖度和出口占比	10
图 9：2016-2024 年 ASML 收入的区域结构	10
图 10：2013-2024 年华为业绩（百万元）	11
图 11：2020Q1-2024Q4 年华为业绩（百万元）	12
图 12：o3-mini 与 o1 在代码补全 Codeforces 排名中的性能对比	13

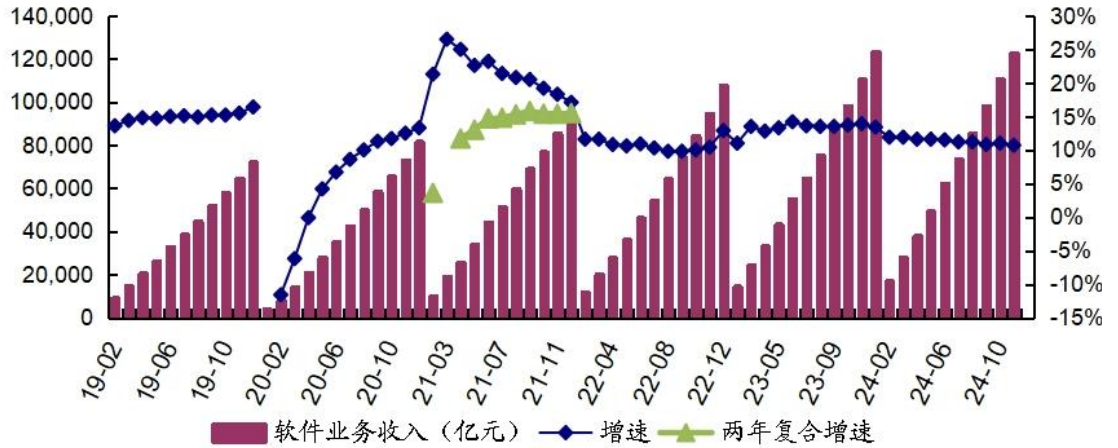
图 13: o3-mini 与 o1 在 AIME 竞赛数学的性能对比	13
图 14: o3-mini 与 deepseek-R1 在物理推理能力中性能对比 (多球碰撞问题)	14
图 15: 大语言模型企业客户的市场份额	14
图 16: 大语言模型企业客户的市场份额	16
图 17: DeepSeek-R1 与 OpenAI o1 模型的能力对比	19
图 18: 各地算力规划中关于总算力的相关目标 (EFLOPS)	26
图 19: 河南上市公司近期股价涨跌幅表现 (%)	32
图 20: 2025 年 1 月中信一级子行业涨跌幅	32
图 21: 2025 年 1 月计算机行业相关概念涨跌幅	33
图 22: 近 10 年中信计算机行业估值水平 (截止 2025.2.11)	33
表 1: OpenAI 大模型产品竞品推出情况	15
表 2: 大模型厂商新模型进展及发布计划	18
表 3: 大模型厂商算力相关投入	20
表 4: 各大厂商接入 DeepSeek 的情况	22
表 5: 硅基流动上 DeepSeek 主要模型情况及调用价格	22
表 6: 近期省级算力基础设施相关规划	23
表 7: 国家数据要素综合试验区批复情况	27
表 8: 国内主要万卡集群落地情况	27
表 9: 英伟达、AMD 人工智能芯片参数对比	28

1. 行业数据

1.1. 行业数据：全年收入利润增速略有回落，出口业务实现增长

2024 年我国软件产业增速呈现回落趋势。根据工信部数据，2024 年软件业务收入 13.73 万亿元，同比增长 10.0%，较 2023 年 13.4% 的增速回落了 3.4PCT，较 1-11 月回落了 0.7PCT。

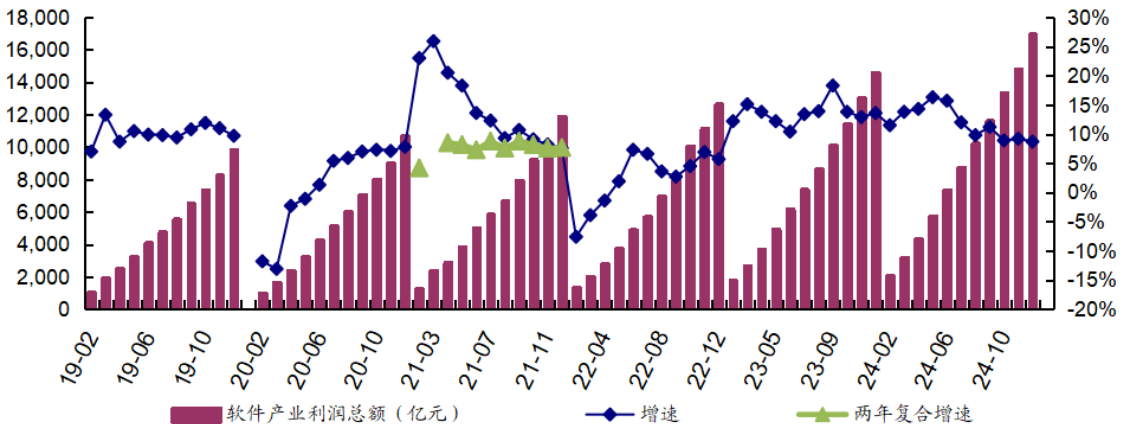
图 1：2019-2024 年我国软件业务收入及增速（月度累计值）



资料来源：工信部，中原证券研究所

利润方面，12 月软件行业利润总额增速回落，处于近两年低位水平。2024 年软件业务利润总额 16953 亿元，同比增长 8.7%，较 2023 年 13.6% 的增速水平回落了 4.9 PCT，较 1-11 月回落了 0.5 PCT。从行业上市公司财报数据来看，行业缩减人员规模并进行费用控制的迹象明显，一定程度上缓解了短期费用压力，但是毛利率下滑对盈利拖累较为明显。

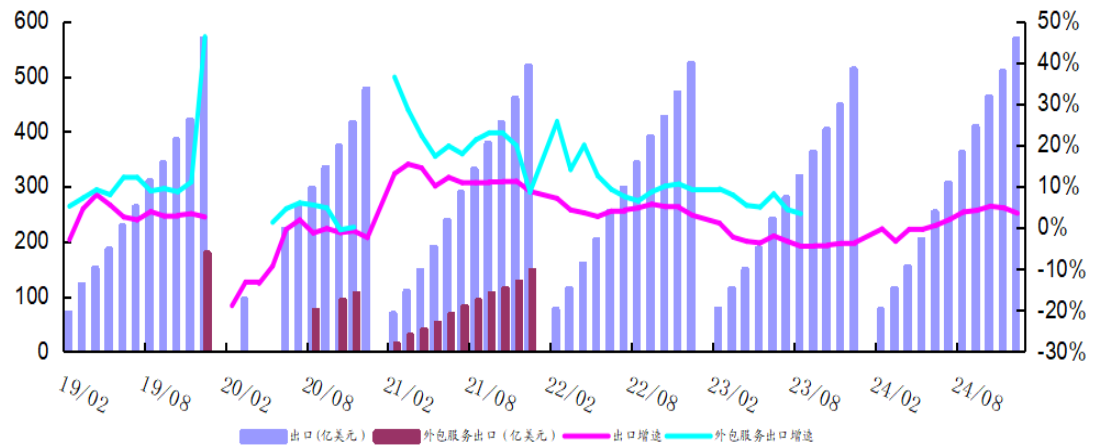
图 2：2019-2024 年我国软件业务利润总额及增速（月度累计值）



资料来源：工信部，中原证券研究所

2024 年出口数据总体向好。2024 年软件业务出口金额 569.5 亿美元，同比增长 3.5%，较 1-11 月回落 1.3 PCT，总体仍然呈现回升态势，较上年-3.6% 的增速回升了 7.1 PCT。

图 3：2019-2024 年我国软件业务出口数据

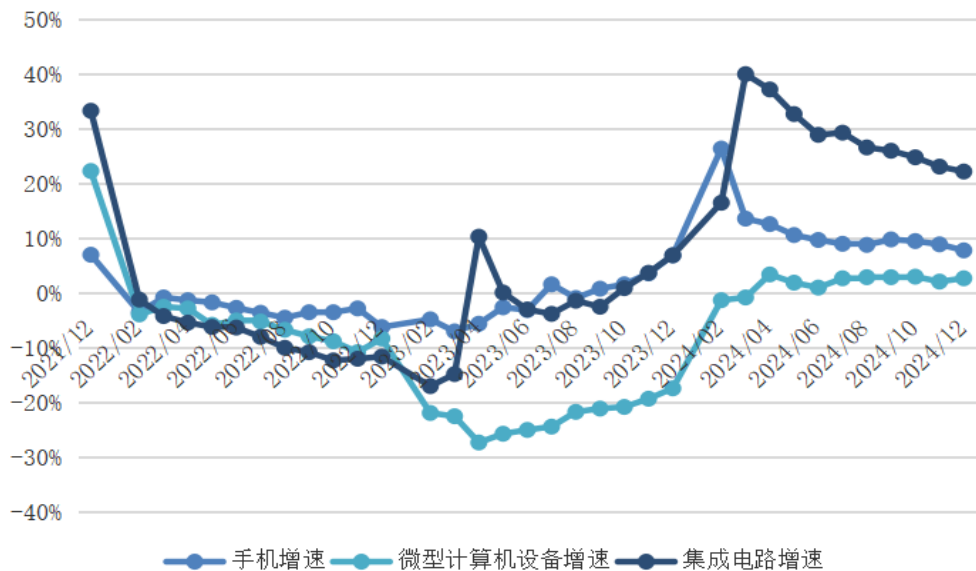


资料来源：工信部，中原证券研究所

1.2. 2024 年高景气赛道：IC 设计、嵌入式系统软件、电子商务平台服务

IC 设计：成为 2024 年的软件行业景气度最高的子行业，全年同比增长 16.4%，较 1-11 月回升 1.6 PCT，高于软件行业整体增速 6.4 PCT，较上年同期增长 10.0 PCT。由于电子产业的全面复苏，相关产业链的景气度也获得持续提升。

图 4：2021 至 2024 年我国电子产品产量增速（累计同比）



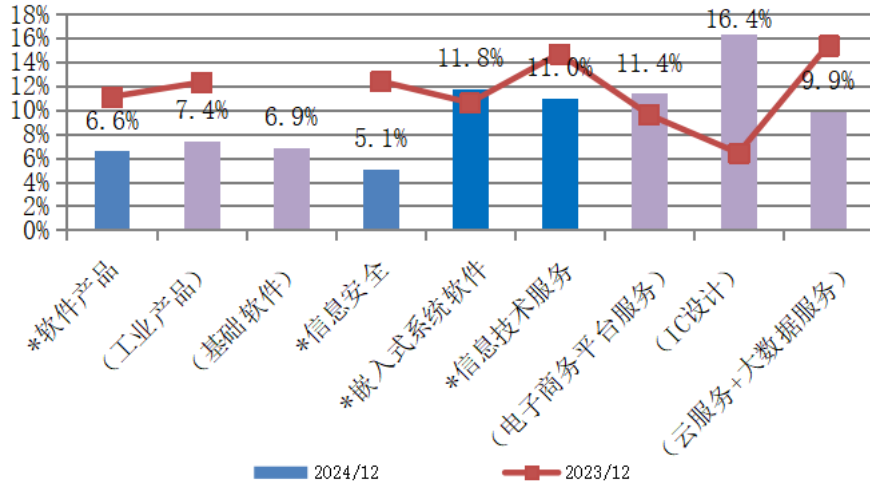
资料来源：工信部，中原证券研究所

嵌入式系统软件：2024 年同比增长 11.8%，较上年增长 1.2 PCT，同时高于软件行业增速 1.8 PCT，较 1-11 月提升了 1.1 PCT。我们认为随着 AI 技术的持续发展，终端层面的软件需求也有望随着 AI 应用提升。

电子商务平台服务：2024 年同比增长 11.4%，较上年增长 1.8 PCT，较 1-11 月增长了

2.7 PCT，高于软件行业整体增速 1.4 PCT。

图 5：2024 年和 2023 年我国软件业务子行业增速对比



资料来源：工信部，中原证券研究所（蓝色柱体为一级子行业，紫色柱体为二级子行业）

基础软件：2024 年同比增长 6.9%，虽然不及软件行业整体增速，但是高于软件产品整体 6.6% 的增速水平，表明目前的国产化软件仍在加快替代的过程中。

1.3. 国产化：国产化持续深入，美国出台新禁令

1.3.1. 拜登政府 1 月制裁措施

当地时间 2025 年 1 月 6 日，美国国防部发布最新版本“中国涉军企业”（CCMC）将新一批中国公司列入黑名单，其中新增腾讯控股、宁德时代、移远通信(603236.SS)、长鑫存储、商汤科技、无人机制造商道通科技 Autel Robotics (688208.SS)等。同时五角大楼取消了 6 家据称不再符合指定要求的公司，包括人工智能公司旷视科技、中国铁建、深圳市科思科技、中国海防、中国建设和中国电信。

腾讯首次进入美国制裁名单，表明接下来美国制裁将不仅仅局限于限制我国军事领域、关键技术环节等领域。

2025 年 1 月 13 日，BIS 正式发布了全球 AI 管控新规，首次全面规范了 AI 芯片和模型在全球范围内的管控，对出口限制进行了全面的升级。

新规将全球国家和地区分为 3 个等级进行管控，包括：

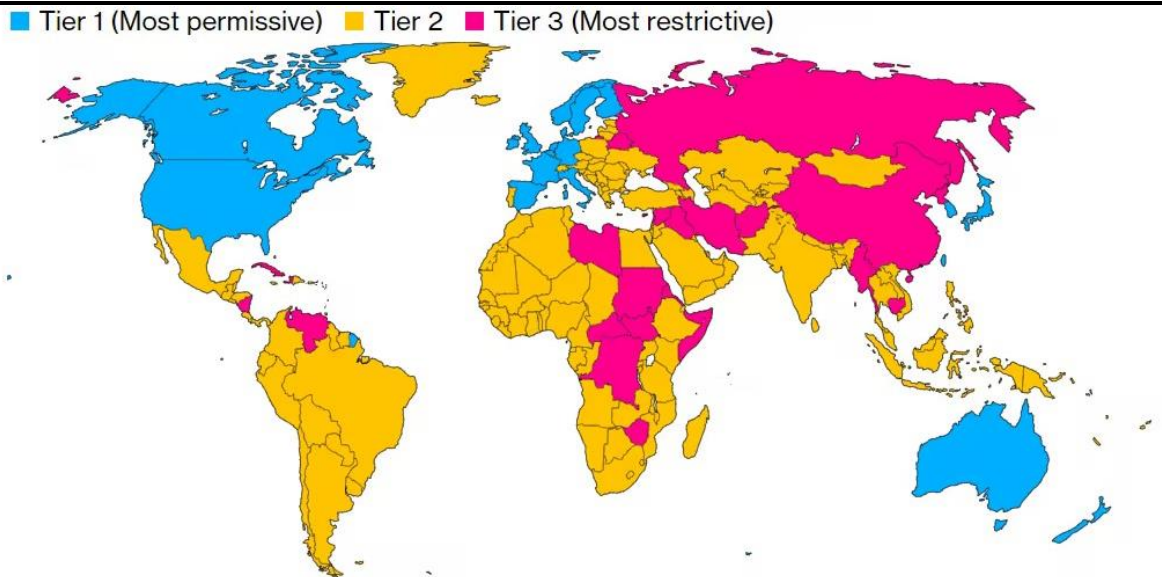
Tier 1 的美国及其少数盟友（包括美国及德国、荷兰、日本、韩国等 18 个同盟），可以不受限制获取美国芯片，总部在该区域的企业在 Tier1 地区以外的算力不得超过总和的 25%，在单一 Tier 2 区域的算力不超过 7%，总部在美国的企业需保证 50%的算力在美国境

内。

Tier 2 的国家和地区（包括了全球多数国家和地区），面临以国为单位的总算力限制，2025-2027 年期间每个国家只能获得 5 万块 H100 等效 GPU（这些国家通过申请 NVEU 资格可以获批到最高 32 万块 H100 的等效算力），以保证这些国家在训练模型是所用的计算集群落后美国 12 个月或一代。企业的高性能闭源模型部署在 Tier 2 区域的部署需要获得美国政府许可。

Tier 3 的 24 个国家和地区（包括中国、俄罗斯、白俄罗斯、伊朗、中国澳门等美国武器禁运的国家和地区），数据中心将被全面禁止进口芯片，企业不能在这些区域部署高性能的闭源模型。

图 6：美国芯片限制的区域划分



Source: Bloomberg reporting

Note: Mapped data show level of restrictions on chip shipments for distinct markets

资料来源：彭博，新智元，中原证券研究所

如果卖家订购不超过 1700 个 H100 将不用计入上限，只需向政府进行通报，无需获得许可，这意味着大学、医疗机构、研究组织等低风险出货不受到以上限制。

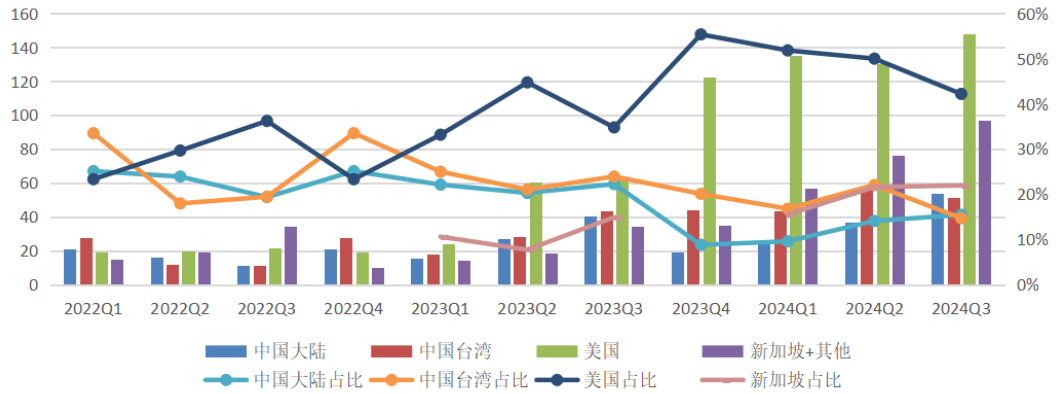
新规将在 1 年后生效，将对全球 AI 算力格局产生深远影响，将由新一届政府决定是否继续执行。

从英伟达最新财报来看，2024Q3 其收入中，42%来自美国，22%来自新加坡，来自中国大陆和中国台湾收入占比都分别为 15%。其中新加坡是 Tier 2 区域，如果新规实施，我们认为出口到新加坡的算力芯片将受到较大的影响。

结合此前新闻，2024 年 12 月 20 日消息，美国商务部近期要求英伟达调查过去一年该公司产品是如何通过非正规渠道流入中国。因而我们认为在新规的全面管控下，通过非合法渠道满足的国内算力需求也将同样受到影响。

考虑到全球大部分地区都处于受限区域，这意味着这些区域将释放出大量的 AI 算力需求。我国国产 AI 芯片厂商如果能解决代工问题，将在全球范围内迎来一个广阔的市场空间。

图 7：2022Q1-2024Q3 英伟达单季度收入的区域分布（亿美元）



资料来源：英伟达公告，Wind，中原证券研究所（英伟达财报较正常季度延后 1 个月）

当地时间 2025 年 1 月 15 日，美国商务部工业和安全局（BIS）发布两项新规。

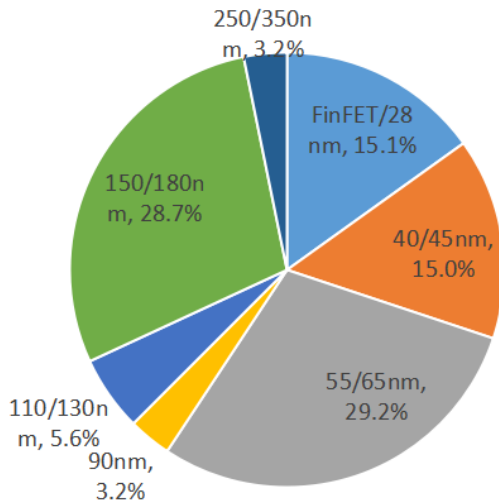
(1) 新规将对 16nm 及以下工艺或采用非平面晶体管结构生产的逻辑电路进行出口管制，对 DRAM 会实施更加严格的管制，而对于晶体管数量少于 300 亿个并由一家值得信赖的公司封装的先进芯片不会受到限制。

(2) 14 家中国实体和 2 家新加坡实体列入实体清单，其中中国实体为 13 家智算芯片企业算能科技及其子公司和封测企业渠梁电子。同时，BIS 将 11 家中国实体列入实体清单，理由是 10 家实体（智谱及其子公司）开发和整合了先进的 AI 研究，哈勃投资的科益虹源参与了中国先进节点制造设施的光刻技术开发。本次实体清单企业涉及到了大模型、芯片、封测、光刻机等产业环节。

本次新规对先进制程 AI 芯片的代工环节限制进一步加大，美国还将加强对台积电、三星、英特尔等晶圆代工厂的尽职调查，并防止其流向中国。

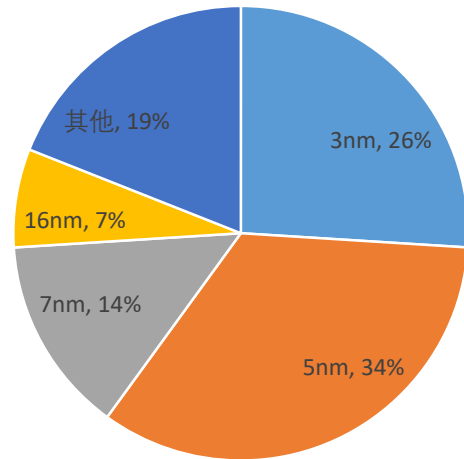
从 2024Q4 台积电财报来看，其收入中有 74% 来自 7nm 及以下制程，另有 16nm 制程收入占比 7%，对应本次新增的制程范围。对于国内代工能力来说，根据中芯国际财报，其在 2021 年 14nm 和 28nm 制程的收入占比 15.1%，具有 14nm 代工能力，但是更多产能集中在成熟支撑领域（中芯国际后续不在年报中披露各制程收入比例数据，此处仅作为参考）。

图 29：2021 年中芯国际不同制程的收入结构



资料来源：中芯国际公告，中原证券研究所

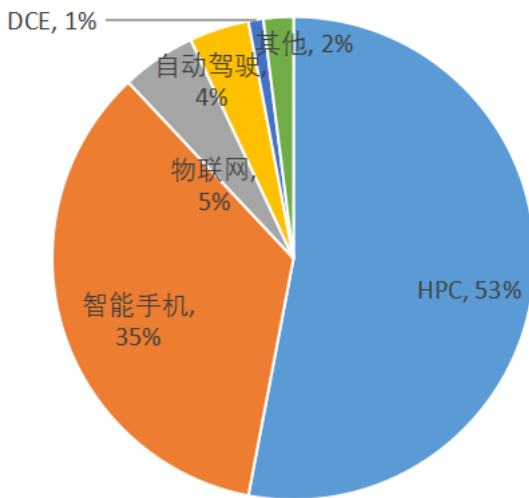
图 30：2024Q4 台积电不同制程的收入占比



资料来源：台积电，中原证券研究所

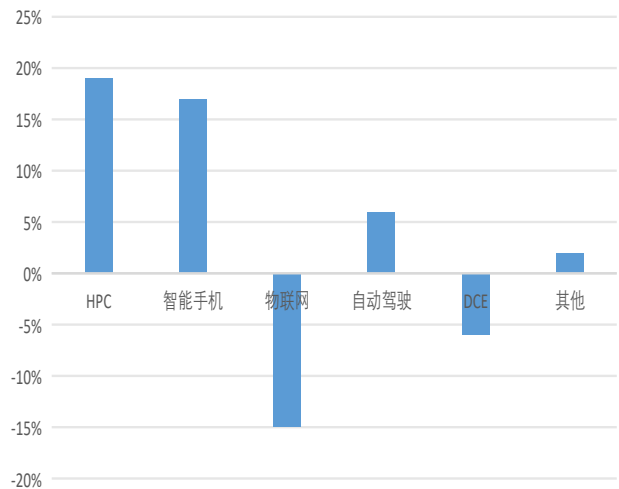
按平台来计算，2024Q4 台积电来自 HPC（高性能计算）的收入占比已经达到 53%，这部分收入在 2024Q4 同比增长了 19%，需求仍然非常强劲。

图 29：2024Q4 台积电来自不同平台的收入占比



资料来源：台积电，中原证券研究所

图 30：2024Q4 台积电来自不同平台的收入增速



资料来源：台积电，中原证券研究所

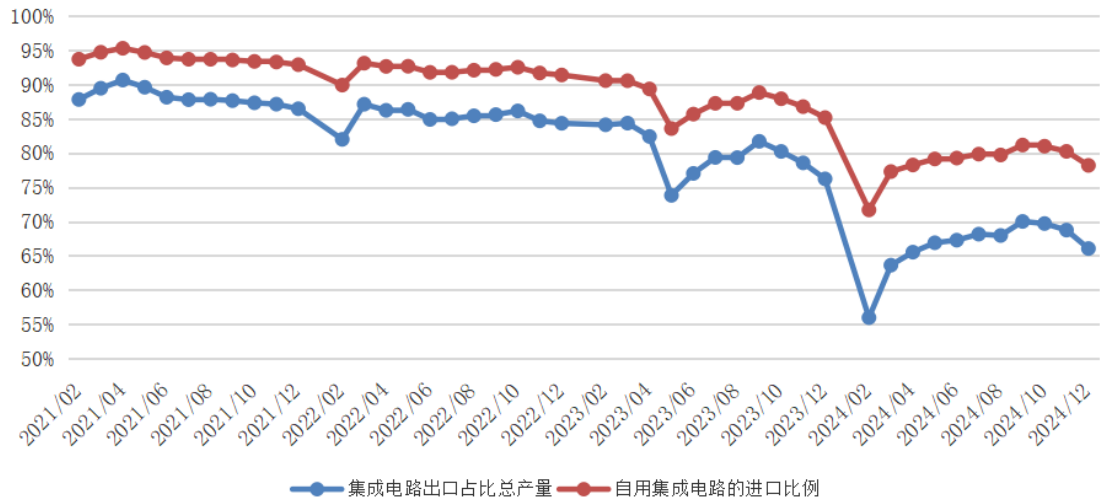
结合全球 1 月 13 日的全球 AI 管控新规来看，美国一方面在进行全球的算力供应方面的管控，另一方面通过代工环节限制竞争对手的供应。我们的突破取决于禁令执行层面的严格程度和上游芯片代工、光刻机等卡脖子环节的技术突破能力。

1.3.2. 芯片国产化比例加快提升，光刻机环节或有转机

2024 年，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 78%（即国产化占比 22%）和 66%，分别较 1-11 月回落 2 PCT 和 3 PCT，下滑趋势明显加大，这也体现出美国一系列禁令影响在持续释放。考虑到 2025 年 1 月禁令集中出台，我们预计在接下来几个

月，这两大指标还将持续下滑。

图 8：2021-2024 年我国集成电路对进口依赖度和出口占比

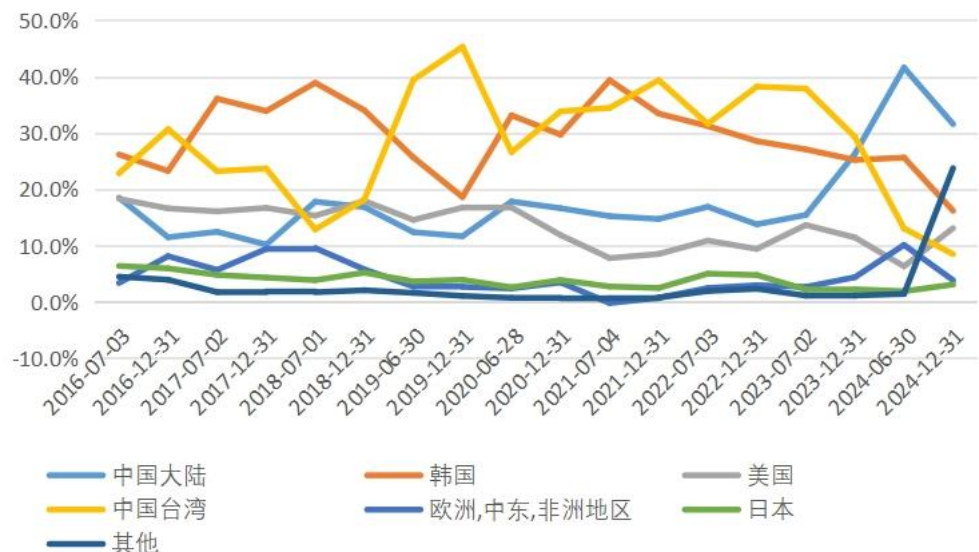


资料来源：工信部，海关总署，中原证券研究所

光刻机方面，荷兰的对华出口政策或有松动。1月17日，荷兰政府决定将ASML对华销售情况排除在敏感商品出口信息的披露范围之外。1月23日，荷兰首相斯霍夫在采访中表示，在荷兰光刻机巨头ASML对华出口的问题上，荷兰政府希望自行决定实施什么样的政策。

阿斯麦对中国出口量的快速上升，意味着未来国内芯片产量有望得到较大增长。根据Wind数据，自2023年下半年以来，ASML来自中国大陆的收入增速获得了较快的增长，2024年中国大陆成为ASML最大收入来源区域，占比达到了31.6%，表明中国大陆芯片产能还将在后续持续释放。而与此同时，2024年下半年，ASML来自其他区域的收入也快速增长，比重从2024H1的1.4%增长到2024年全年占比23.7%，而其来自韩国、中国台湾的收入占比持续下滑。

图 9：2016-2024 年 ASML 收入的区域结构

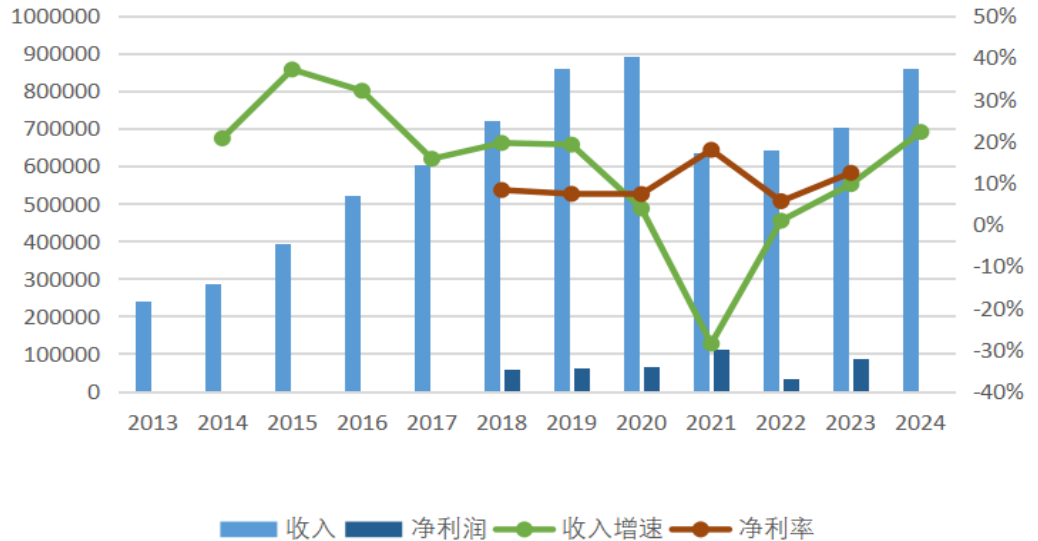


资料来源：Wind，ASML 财报，中原证券研究所

1.3.3. 华为财报：2024 年收入重回美国制裁前的水平，研发投入增长明显

在广东省高质量发展大会期间，华为披露了 2024 年年报相关数据。2024 年其收入达到 8600 亿元，这标志着其收入重回美国制裁前的水平，达到历史第二高水平。这期间，华为还剥离了荣耀、超聚变等资产，因而我们认为这是一份非常优秀的成绩单。

图 10：2013-2024 年华为业绩（百万元）

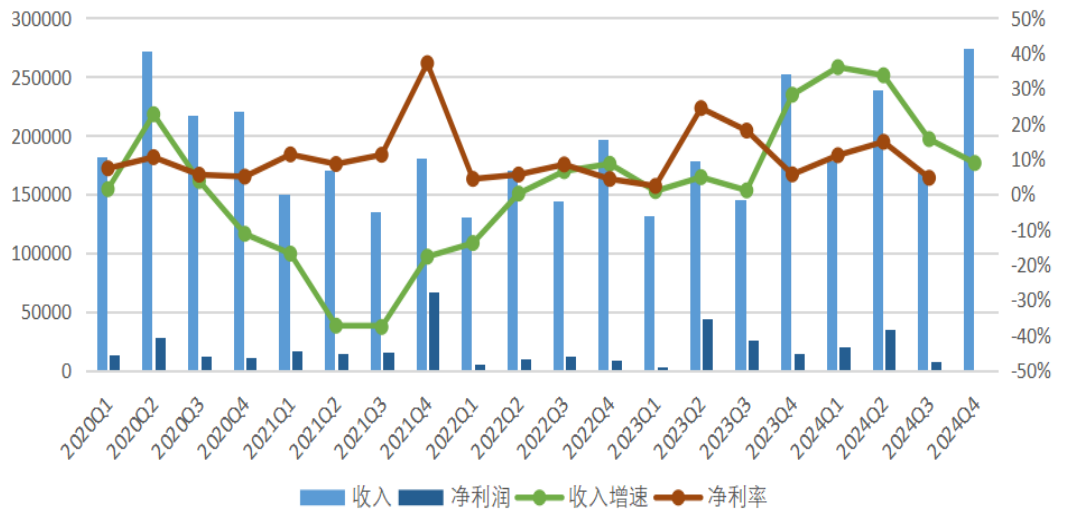


资料来源：华为，中原证券研究所

同时，从单季度数据来看，华为收入增速在 2024Q1 见顶后逐季回落，2024Q4 约回落到 8.8%，但是考虑到 2024 年 4 季度华为增速较高，我们认为在考虑基数因素的基础上，2024Q4 仍然取得了不错的成绩。

需要注意的是华为净利率在 2024Q2 见顶后回落较为明显，2024Q3 仅为 4.6%。同时 2024 年前三季度，华为研发费用同比增长 10.8%，较 2023 年 2% 和 2024 上半年 7.6% 的增速有较大提升，让我们看到了华为积极加大研发投入的趋势。

图 11: 2020Q1-2024Q4 年华为业绩 (百万元)



资料来源: 华为, 中原证券研究所

1.4. AI: DeepSeek-R1 出世, GPT-5 即将推出

1.4.1. OpenAI 即将推出 GPT-5

2024 年 12 月 21 日, OpenAI 宣布了 o1 的下一代模型 o3 (o2 代号因为重名问题被跳过), 包括了 o3 完整版和精简版 o3-mini。OpenAI 将在 1 月底正式推出 o3-mini, 之后推出 o3 完整版。

2025 年 1 月 31 日, OpenAI 如期推出了 o3-mini, 支持 low、medium 和 high 低中高三种推理难度, 成为了 OpenAI 推理系列中最具成本效益的模型, 同时 ChatGPT 首次向免费用户提供的推理模型。

表 1: 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 对用户免费, API 速度较 GPT-4 Turbo 提高 2 倍, 价格降低 50%

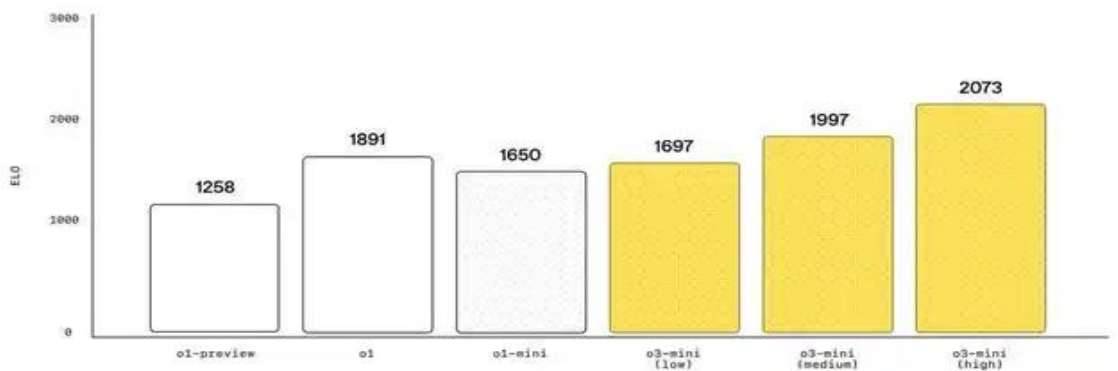
o1	2024.9					更擅长解决科学、编码、数学和类似领域的复杂问题
o3	2025.1					o1 的下一代产品

资料来源: OpenAI, 中原证券研究所

在模型性能方面, o3-mini 针对 STEM 进行了优化, 在数学代码等基准测试中, 均获得了最高成绩, 在物理模拟等高难度问题上也明显超过 R1, 再次将大模型能力推向了新高度。

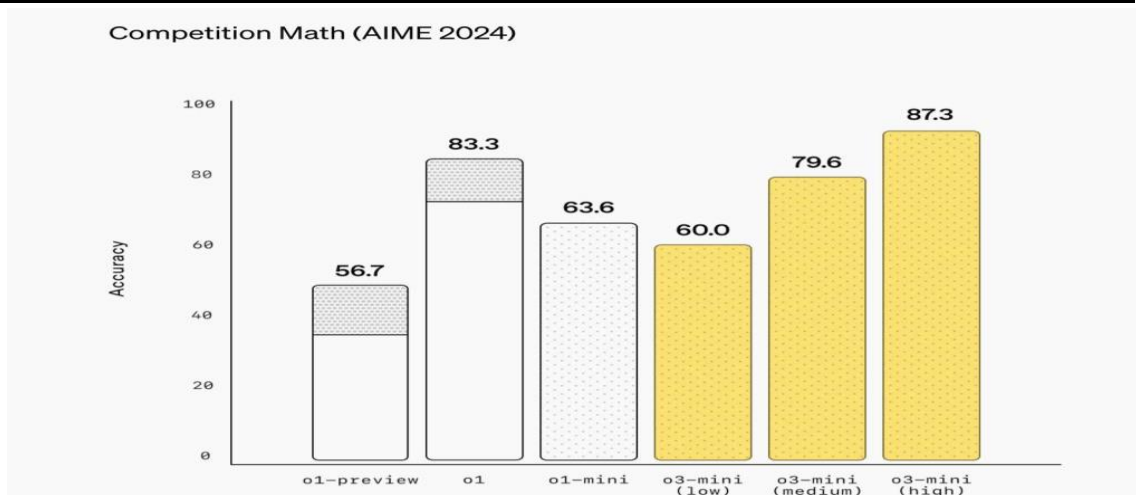
图 12: o3-mini 与 o1 在代码补全 Codeforces 排名中的性能对比

Competition Code (Codeforces)



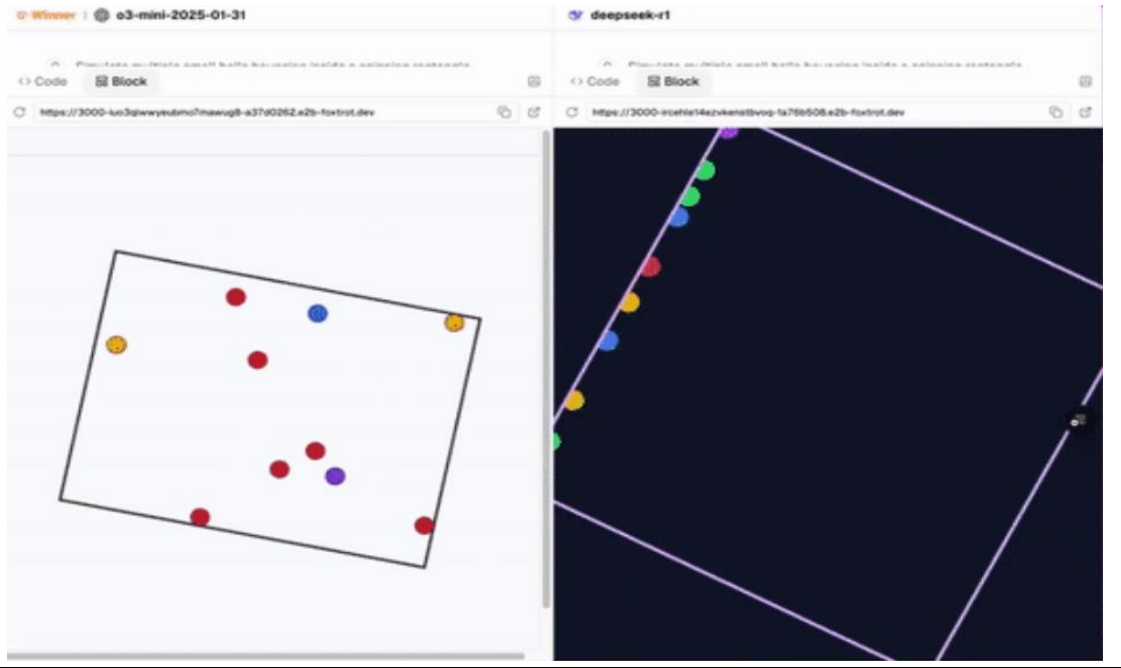
资料来源: OpenAI, 中原证券研究所

图 13: o3-mini 与 o1 在 AIME 竞赛数学的性能对比



资料来源: OpenAI, 中原证券研究所

图 14: o3-mini 与 deepseek-R1 在物理推理能力中性能对比 (多球碰撞问题)

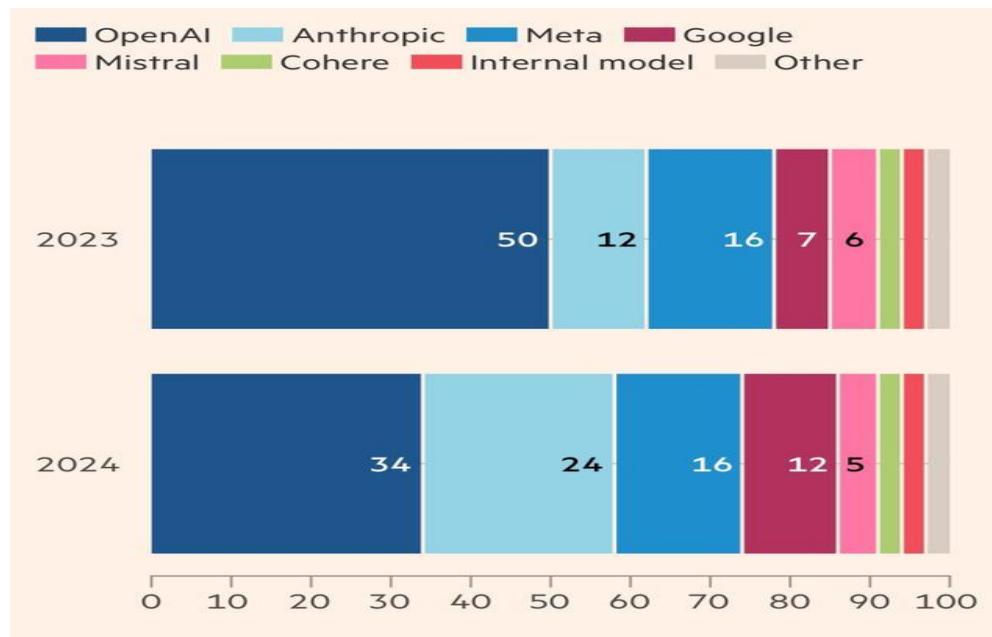


资料来源: 新智元, 中原证券研究所

根据 Information 的报道, 2024 年 ChatGPT 付费用户数达到了 1550 万, 企业模型采用率增长了 7 倍, 新推出的 200 美元/月 Pro 订阅计划年收入达到 3 亿美元。

但与此同时, OpenAI 在企业客户的份额从 2023 年的 50% 下降到了 2024 年的 34%, 而 Anthropic 在此过程中进步最快, 实现了份额的翻倍, 对 OpenAI 形成了较大的挑战。

图 15: 大语言模型企业客户的市场份额



资料来源: Menlo Ventures, 中原证券研究所

1 月 31 日, OpenAI 表示需要想出一个不同的开源策略, 2 月 6 日 OpenAI 宣布向所有用户开放 ChatGPT 的搜索功能, 无需注册。2 月 13 日, OpenAI 表示 GPT-4.5/5 将很快陆续发

布，免费版 ChatGPT 将在标准智能设置下无限制使用 GPT-5 进行对话。

此外，根据《金融时报》报道，OpenAI 正以 3000 亿美元的估值寻求新一轮 400 亿美元的融资，这将为 OpenAI 后续模型研发提供持续的资金支持。

1.4.2. 国内外模型差距持续缩小

从目前人工智能的发展来看，我们与 OpenAI、谷歌、Anthropic 等全球头部大模型公司差距基本控制在半年以内。相比 2023 年来看，国内大模型在模型能力、产品价格、算力需求等方面都取得了都亮点十足，同时国内的头部模型还在积极选择开源方式，积极推动大模型技术加快发展。

表 1: OpenAI 大模型产品竞品推出情况

对标模型	时间	参数	厂商	产品	推出时间	备注
GPT-4o	2024.5.14	多模态交互模型	谷歌	Project Astra	2024.5.15	
				Gemini Live	2024.8.14	
			Anthropic	Claude 3.5 Sonnet	2024.6.21	
			智谱	GLM-4-Plus	2024.8.29	
			阿里	Qwen-Max 升级版	2024.9.19	
				Qwen2.5 Max	2025.1.29	
			字节	豆包通用模型 pro	2024.12.18	价格为 GPT-4o 的 1/8
				豆包大模型 1.5 Pro	2024.1.22	
幻方	DeepSeek-V3	2024.12.26	开源			
o1	2024.9.13	推理模型	Anthropic	升级版 Claude 3.5 Sonnet	2024.10.23	
			谷歌	Gemini 2.0 Flash Thinking	2024.12.20	
			幻方	DeepSeek-R1-Lite-Preview	2024.11.21	对标 o1-preview
				DeepSeek-R1	2025.1.20	开源
			智谱	GLM-Zero-Preview	2024.12.31	对标 o1-preview
o3	2025.1.31	o1 的下一代产品				

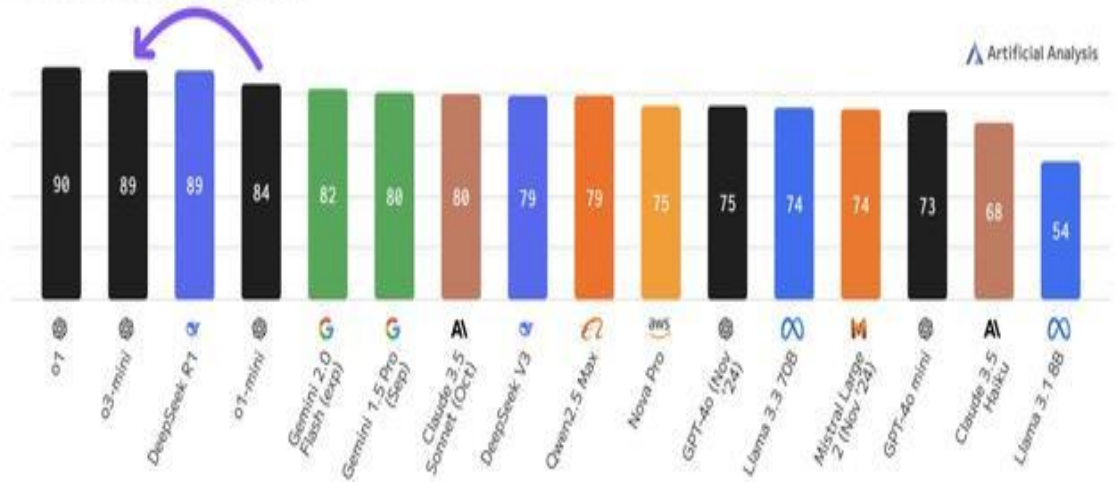
资料来源：OpenAI，界面新闻，机器之心 Pro，腾讯科技，中原证券研究所

在性能方面，根据人工智能分析公司 Artificial Analysis 的人工智能模型质量指数，DeepSeek-R1 以 89 分匹敌 o3-mini，同样位于前十的国内模型还包括 DeepSeek-V3 和阿里的 Qwen2.5 Max。

图 16: 大语言模型企业客户的市场份额

Artificial Analysis Quality Index

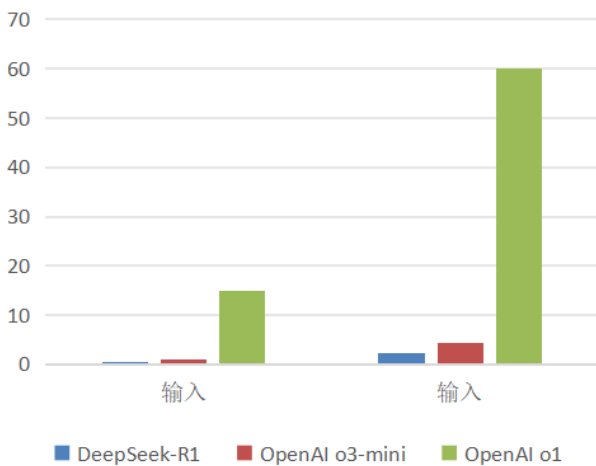
Artificial Analysis Quality Index includes MMLU, GPQA Diamond, MATH-500 and HumanEval; Evaluation results measured independently by Artificial Analysis; Higher is better



资料来源: Artificial Analysis, 中原证券研究所

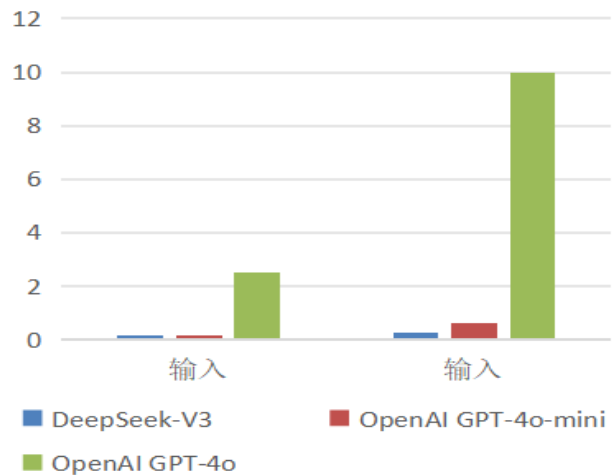
在价格方面, 虽然 OpenAI 刚推出的 o3-mini 价格较 o1 更便宜, 但是仍然约为 DeepSeek-R1 两倍水平。

图 29: DeepSeek 与 OpenAI 推理模型价格对比



资料来源: OpenAI, DeepSeek, 智东西, 中原证券研究所

图 30: DeepSeek 与 OpenAI 多模态模型价格对比



资料来源: OpenAI, DeepSeek, 智东西, 中原证券研究所

1.4.3. 针对 DeepSeek 的制裁和攻击

由于 DeepSeek-R1 以较低成本达到了 o1 的模型性能, 同时实现了开源, 对闭源模型厂商竞争优势形成了较大的冲击, 也从根本上改变了中美大模型竞争格局, 因而带来美国科技界的焦虑恐慌, 也成为了近期美国重点打压对象。

(1) 在政府部门禁用方面：美国消费者新闻与商业频道 (CNBC) 报道称，1 月 24 日美国海军发送警告信，要求其工作人员不得使用 DeepSeek 模型。美国国防信息系统局 (DISA) 于 1 月 28 日晚采取措施，五角大楼封锁了对中国初创公司 DeepSeek 网站的访问。1 月 30 日，美国国会众议院首席行政事务官向国会办公室发出通知警告国会办公室不要使用 DeepSeek 的服务。1 月 31 日，CNBC 获悉：因「安全和隐私问题」，美国宇航局 (NASA) 禁止员工使用 DeepSeek 并封锁该平台访问。2 月 1 日，美国得克萨斯州州长格雷格·阿博特宣布，禁止该州官员在政府设备上使用中国人工智能服务 DeepSeek 及社交平台小红书。

(2) 影响深远的法案正在孕育：1 月 31 日，美国参议员 Josh Hawley 提出了一项激进的法案《2025 年美国人工智能能力与中国脱钩法案》(Decoupling America's Artificial Intelligence Capabilities from China Act)，要求美国对中国 AI 技术进行严格限制，甚至把下载和使用中国 AI 模型 (如 DeepSeek) 定为非法行为——相关违法者将面临最高 20 年监禁及最高 1 亿美元的罚款。

(3) 网络攻击：2025 年 1 月 28 日，美国多名官员发声称 DeepSeek 剽窃，正对其影响开展国家安全调查。同日，DeepSeek 官网显示，其线上服务受到大规模恶意攻击，所受网络攻击的 IP 地址都在美国。

(4) 企业屏蔽：根据 1 月 31 日彭博报道，已经有数百家公司 (尤其是政府相关的企业) 屏蔽了 DeepSeek，原因是担忧 AI 模型可能会将数据泄露给中国。

(5) 指控 DeepSeek “偷窃数据”：当地时间 1 月 28 日，特朗普提名的 AI 事务主管 David Sacks 声称，有大量证据表明 DeepSeek 利用了 OpenAI 模型的输出数据来开发自身技术。然后微软的研究人员就跟着站出来表示，与 DeepSeek 相关的个人，曾在去年秋天使用 OpenAI 的 API 大规模提取数据。

(6) 进一步的芯片禁令：2025 年 1 月 30 日，彭博社报道，特朗普政府官员正在考虑加大对英伟达芯片的出口限制，将 H20 芯片加入限制范围。1 月 30 日，Anthropic 的首席执行官 Dario Amodei 发表博文《关于 DeepSeek 和出口控制》，提出需要堵住芯片出口漏洞，阻止我国获得数百万芯片，从而增加美国领先的单极世界的可能性。

美国各界的一系列反应，进一步佐证了 R1 的模型能力和其对科技界的影响力。

1.4.4. 近期主要大模型厂商进展

随着 1 月 DeepSeek-R1 的问世，我们看见大模型发展呈现加速趋势，海外 OpenAI 发布了全新的 o3-mini 同时首次向免费用户开放推理模型，还表示将考虑开源，而后宣布向所有用户开放 ChatGPT 的搜索功能，紧接着 OpenAI 又宣布即将发布 GPT-5，并免费无限使用。百度也同样宣布将在 4 月 1 日起，全面免费 (PC 端和 APP 端)。

表 2: 大模型厂商新模型进展及发布计划

公司	模型	时间	详情
xAI	Grok 3	2024.7.2	Grok 3 训练过程中使用了 10 万块英伟达 H100, 预计于年底发布。
		2025.1.4	马斯克在社交平台 X 上表示, Grok 3 已经完成预训练, 计算量比 Grok 2 高 10 倍。
OpenAI	O3 (O2)	2024.11.4	Sam Altman 泄露新模型 o2
		2024.12.21	OpenAI 发布了新模型 O3, 能够进行长时间思考, ARC 基准测试超越人类的智商。OpenAI 将在 1 月底正式推出 o3-mini, 之后推出 o3 完整版。
		2025.1.31	正式推出 o3-mini。
		2025.2.13	O3 不会作为独立模型推出
	GPT-4.5 (Orion)	2024.11.14	Orion 的编程表现不理想, 部分原因是缺乏足够的编程数据进行训练。Orion 仍未达到 OpenAI 希望的可向用户发布的水平, 公司也不太可能在明年年初之前推出该系统。
		2024.12.23	Orion 在开发过程中遭遇重重困难, 导致项目进度严重滞后。自项目启动以来, 已历时 18 个月, 但至今未能达到预期目标。
	GPT-5	2025.1.13	Orion 将很快发布, 是最后一个非思维链模型。 会整合 GPT 和 O 系列两大模型打造一个全新的系统, 能自动选择思考和非思考功能, 适用于多种任务。在 ChatGPT 和 API 中, 将发布 GPT-5, 这是一个整合了包括 o3 模型在内诸多系统。免费版 ChatGPT 将在标准智能设置下无限制使用 GPT-5 进行对话。
谷歌	Gemini 2.0	2024.10.26	Gemini 2.0 计划在 12 月正式发布。
		2024.11.14	即将推出的 Gemini 未达内部预期。
		2024.12.11	发布 Gemini 2.0
		2024.12.20	发布首个推理模型 Gemini 2.0 Flash Thinking。
		2025.2.5	发布发布新一代 Gemini 系列大模型, 包括 Gemini 2.0 Flash、Gemini 2.0 Flash-Lite 及旗舰模型 Gemini 2.0 Pro 实验版本, 此次升级重点强化了多模态处理能力
Meta	Llama 4	2024.8.5	Llama 4 训练已开启, 预计将在 2025 年发布, 将用 Llama 3 的十倍计算量
Anthropic	Claude 3.5 Opus	2024.11.14	时间表推迟
		2024.12.12	semianalysis 分析师爆料, Claude 3.5 超大杯被藏起来, 只用于内部数据合成以及强化学习奖励建模。
幻方	DeepSeek-V3	2024.12.26	模型 DeepSeek-V3 首个版本上线并同步开源。
	DeepSeek-R1	2025.1.20	发布并开源 DeepSeek-R1, 在数学、编程和推理等多个任务上达到了与 OpenAI o1 相当的表现水平, 同时将应用程序编程接口调用成本降低了 90-95%。
月之暗面	k0-math	2024.11.16	推出新一代数学推理模型 k0-math。
	k1	2024.12.16	Kimi 发布视觉思考模型 k1, 该模型基于强化学习技术打造, 原生支持端到端图像理解和思维链技术, 标志着 Kimi 正式跨入打造多模态功能的新阶段。
阿里	QwQ-32B-Preview	2024.11.28	发布全新 AI 推理模型 QwQ-32B-Preview。
	Qwen2.5 Max	2025.1.29	发布全新的通义千问 Qwen 2.5-Max 超大规模 MoE 模型
MiniMax	MiniMax-01 系列	2025.1.15	发布并开源新一代 01 全新系列模型, 该系列模型包含基础语言大模型 MiniMax-Text-01 和视觉多模态大模型 MiniMax-VL-01。

字节	豆包大模型 1.5	2025.1.22	发布豆包大模型 1.5Pro
----	-----------	-----------	----------------

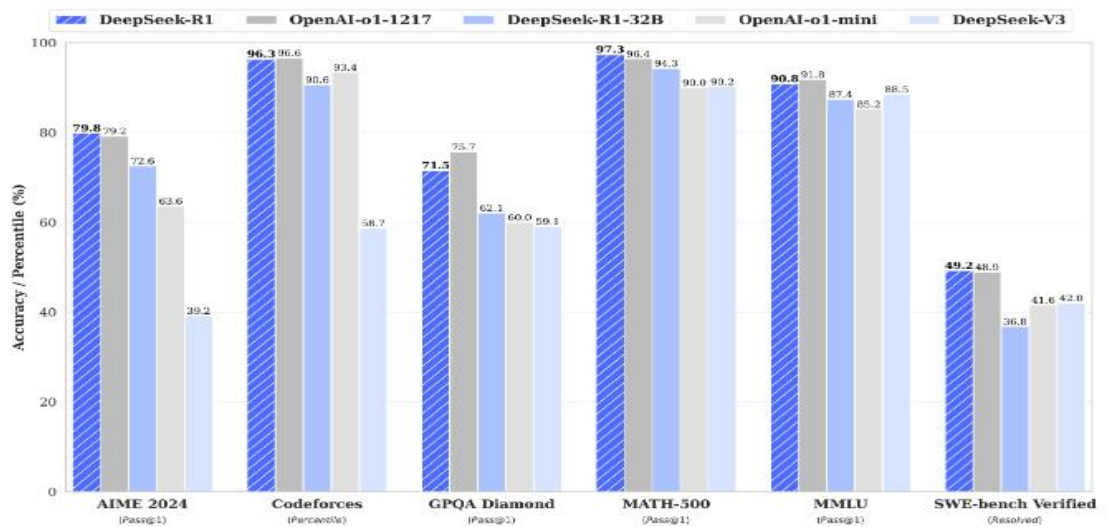
资料来源：澎湃，快科技，华尔街见闻，钛媒体，新智元，智通财经网，新浪财经，太平洋科技，量子位，中原证券研究所

同时，国内阿里 Qwen 2.5-Max 在发布后也取得了较好的成绩，成为了继 Deepseek 之外国内又一个竞争实力突出的模型。2月11日，根据 The Information 报道，苹果公司正与阿里巴巴公司合作，开发和推出面向中国 iPhone 用户的人工智能。这是对阿里在模型能力和服务能力等多方面综合能力的肯定。

1.4.5. DeepSeek-R1 开发模式的创新

R1 在后训练阶段大规模使用了强化学习，来取代传统大模型中监督式微调的方法，不需要大量的标注数据，节省了算力，同时实现了与 o1 相媲美的性能。这意味着在 Scaling Law 失效的困扰下，人类在增加训练数据和算力以外，探索出了高效的架构与训练策略提升的新路径，这或将成为推动 OpenAI 的 GPT-5 问世的一个最重要的技术突破。

图 17: DeepSeek-R1 与 OpenAI o1 模型的能力对比



资料来源：DeepSeek, Github, 中原证券研究所

监督式微调是建立在人类已有知识体系的训练策略，而强化学习是机器自信进行的路径探索。AlphaGo 正是借助了强化学习的方法，获得了人类此前并不获知的围棋棋谱，实现了对人类智能的超越。因而强化学习路径下，大模型的思维能力将被进一步释放。

表 1: DeepSeek 的主要模型发布情况

时间	模型	架构	参数	模型类型	
2023.11.29	DeepSeek LLM 67B		670 亿	语言模型	开源，对标 LLaMA2 70B
2024.5.	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
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
2024.1.20	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

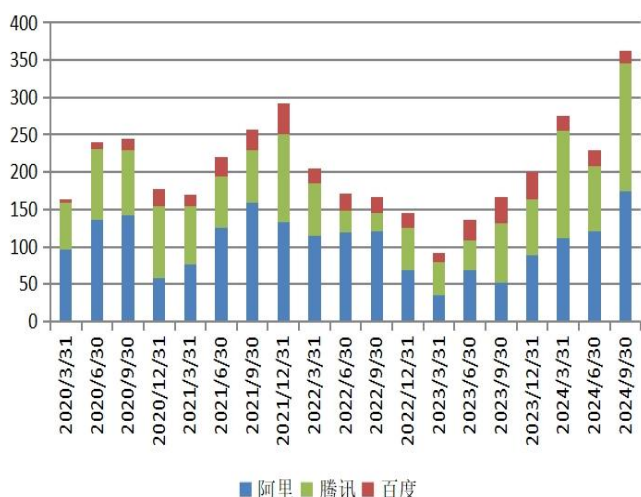
资料来源：OpenAI，中原证券研究所

1.5. 算力：DeepSeek 发布加速芯片国产化，AI 应用将带来更多算力需求

1.5.1. DeepSeek-r1 发布以后，头部科技厂商仍然给出了 2025 年积极的投资计划

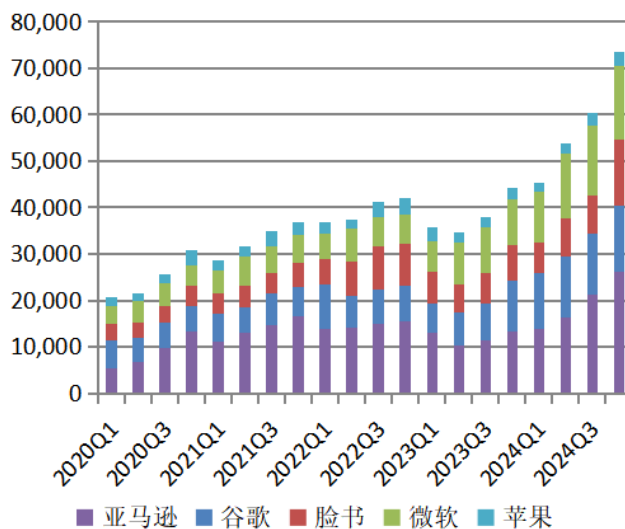
2024Q4，美国 5 大科技厂商的资本开支再创新高，总计达到 734.97 亿美元，同比增长 66%，较 Q3 继续提升了 7PCT。其中亚马逊以 95% 的增速继续领涨，脸书增长 88% 呈现明显加速迹象，微软以 62% 的增速名列第三，而谷歌仅增长 30%，增速下滑明显；苹果 23% 的增速，表现最弱。

图 29：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 30：5 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

在 2025 年资本开支方面，亚马逊计划投资 1000 亿美元，谷歌为 750 亿美元，Meta 为 600-650 亿美元，都较上一年有了较为明显的提升，微软也提出了 2025 财年（6 月结束）800 亿的数据中心投资目标。考虑到除了微软以外，其他科技厂商都是在 DeepSeek-R1 发布以后才公布的投资目标，这表明 2025 年全球在 AI 领域的投入仍将呈现快速增长趋势。

同时，在 1 月美国公布了 5000 亿美元的“星际之门”投资计划以后，法国也在 2 月公布了 1090 亿欧元的 AI 投资计划。

表 3：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。
OpenAI	2025.1.21	美国总统特朗普在白宫宣布，由美国甲骨文公司、OpenAI 和软银共同出资投建，三家公司组成的联合企业将投资 5000 亿美元 ，在美建设名为“星际之门”的 AI 基础设施。该项目计划在美国建设数据中心，以支持人工智能技术发展。
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 。
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 。
Meta	2025.1.24	计划在 2025 年投入 600 亿至 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 。
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。
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，The Information，同花顺，中原证券研究所

1.5.2. 行业加速接入 DeepSeek 大模型，算力供应的生态加速繁荣

由于 Deep-R1 和 DeepSeek-V3 采用开源策略，在其发布以后，海内外云厂商都快速接入了 DeepSeek 的模型，省去了开发者买卡、装驱动、配网络、配存储、装环境、装框架、下载模型等繁琐步骤。同时，包括昇腾、昆仑芯、寒武纪、飞腾、海光在内的国产芯片也都对 DeepSeek 进行了适配，加速了国产替代进程。下游的软件厂商也都快速接入了 DeepSeek，加速了在大模型的应用落地进程，也可以很好地满足更多私有化的行业需求。

表 4: 各大厂商接入 DeepSeek 的情况

时间	公司	内容
1.31	英伟达	NVIDIA NIM 已经可以使用 DeepSeek-R1 模型
1.31	亚马逊	DeepSeek-R1 模型可以在 Amazon Web Services 上使用
1.31	微软	DeepSeek-R1 正式纳入 Azure AI Foundry, 成为该企业级 AI 服务平台的一部分
2.1	华为云、硅基流动	经过硅基流动和华为云团队连日攻坚, 现在, 双方联合首发并上线基于华为云昇腾云服务的 DeepSeek R1/V3 推理服务。得益于自研推理加速引擎加持, 硅基流动和华为云昇腾云服务支持部署的 DeepSeek 模型可获得持平全球高端 GPU 部署模型的效果。
2.1	中国电信	通过“息壤”智算平台, 提供 DeepSeek-R1 模型的推理、训练及微调支持, 并在 GPU 云主机、科研助手、AI 云电脑等业务场景中部署应用。
2.2	腾讯云	DeepSeek-R1 大模型一键部署至腾讯云「HAI」上, 开发者仅需 3 分钟就能接入调用。
2.3	中国联通	借助“星罗”平台, 实现 DeepSeek-R1 在编程助手、云桌面等产品中的集成, 并在全国 270 多个骨干云池预部署, 以优化推理效率和数据安全能力。
2.3	百度智能云	百度智能云千帆平台已正式上架 DeepSeek-R1 和 DeepSeek-V3 模型
2.3	阿里云	阿里云 PAI Model Gallery 支持云上一键部署 DeepSeek-V3、DeepSeek-R1。
2.4	海光	成功完成 DeepSeekV3 和 R1 模型与海光 AI 加速卡 DCU 的适配并正式上线
2.5	华为	基于原生鸿蒙操作系统的小艺助手 App 已经接入 DeepSeek, DeepSeek 系列新模型正式上线昇腾社区
2.5	中国移动	支持 DeepSeek 全版本、全尺寸模型, 并通过智算中心、云计算平台进行深度适配, 同时结合自研的 COCA 算力平台, 实现更灵活的模型调用与部署。
2.6	国家超算互联网	DeepSeek-R1、V3、Coder 等系列模型已登陆该平台, 此次上线的模型涵盖了 1.5B 到 14B 的蒸馏模型, 以及 7B 到 32B 的 DeepSeek Chatbot 可视化界面, 近期还会陆续上线 32B、70B 等更多版本。
2.6	寒武纪	南京智算中心与寒武纪、苏宁科技合作, 成功上线全国产算力版 DeepSeek, 将助力苏宁灵思大模型, 优化零售场景应用能力
2.8	金山云	在公有云场景和国资云/政务云场景已支持 DeepSeek-R1/V3。
2.8	商汤	企业客户和开发者可在商汤大装置万象平台中, 快速部署 DeepSeek-V3、DeepSeek-R1 等模型, 可享受 3 个月内 1000 万 tokens 免费使用权益。
2.9	阿里云	百炼平台宣布全面上线 DeepSeek-V3、DeepSeek-R1、DeepSeek-R1-Distill-Qwen-32B 等 6 款全尺寸模型
2.9	飞腾	飞腾腾云 S5000C、腾锐 D3000 处理器已成功实现对 DeepSeek 全系列大模型的端到端支持, 覆盖数据中心和终端场景
2.10	寒武纪	中国移动通信集团青海有限公司基于寒武纪 MLU 590 国产化算力资源池, 已全面部署上线 DeepSeek 国产化大模型服务

资料来源: 快科技, 金融界, 南方都市报, 每日经济新闻, 中原证券研究所

从 DeepSeek 发布的模型来看, 包括了 6710 亿的 R1 和 V3 模型, 也包括了从 700 亿到 15 亿参数不等的蒸馏小模型, 可以满足不同体量用户的模型需求。

表 5: 硅基流动上 DeepSeek 主要模型情况及调用价格

模型	模型类型	架构	参数	上下文	输入 (百万 token)	输出 (百万 token)
DeepSeek-R1	推理模型	MoE	6710 亿	64k	4	16
DeepSeek-V3	语言模型	MoE	6710 亿	64k	2	8
DeepSeek-R1-Distill-Llama-70B	推理模型		700 亿	32k	4.13	4.13

DeepSeek-R1-Distill-Qwen-32B	推理模型		320 亿	32k	1.26	1.26
DeepSeek-R1-Distill-Qwen-14B	推理模型		140 亿	32k	0.7	0.7
DeepSeek-R1-Distill-Llama-8B	推理模型		80 亿	32k	免费	免费
DeepSeek-R1-Distill-Qwen-7B	推理模型		70 亿	32k	免费	免费
DeepSeek-R1-Distill-Qwen-1.5B	推理模型		15 亿	32k	免费	免费

资料来源：硅基流动，中原证券研究所

1.5.3. 浙江算力规划也呈现加快趋势

1月23日在于杭州举行的浙江省2025年经济政策新闻发布会上，浙江省表示近期将出台《浙江省“人工智能+”行动计划(2025-2027年)》。浙江省将聚焦“算力强省”，进一步推动存量算力设施改造升级，集约化布局建设5个万卡智算集群，力争2025年全省算力规模超过100EFlops，还将出台智能券专项政策，对人工智能企业使用算力、语料、模型成本提供一定比例补贴。

表 6: 近期省级算力基础设施相关规划

时间	区域	政策	规划
2023.1.10	湖南	《湖南省算力支撑能力提升行动方案(2022—2025年)》	到2025年底，全省总算力从目前3EFLOPS达到10EFLOPS，其中先进算力达2EFLOPS，数据中心算力达8EFLOPS；数据中心总存储能力达50EB；数据中心机架规模年均增速保持在15%左右，数据中心平均利用率不低于60%；数据中心电能利用效率指标控制在1.5以内，新建大型和超大型数据中心电能利用效率(PUE)控制在1.3以内；省内网间访问时延降低80%以上，丢包率降低50%以上；打造特色鲜明、泛在高效、规模适度的算力支撑体系，努力将湖南建设成为全国先进绿色算力枢纽和国际领先的算法创新中心。2022年人工智能算力达200PFLOPS，2025年达到 1500PFLOPS 。
2023.12.15	安徽	《安徽省数字基础设施建设发展三年行动方案(2023-2025年)》	预期2025年数据中心机架规模不少于45万架，数据中心平均上架率不少于65%，算力规模14.6EFLOPS，智能算力规模超过 12EFLOPS 。同时要求2025年新建大型及以上数据中心电能利用率低于1.3。
2024.1.6	山西	《山西省算力基础设施高质量发展实施方案》	到2025年，全省算力规模超过9EFLOPS，智能算力占比达到 35% ，建成超过4个智能计算中心。存储总量超过36EB，先进存储容量占比达到30%以上，重点行业核心数据、重要数据灾备覆盖率达到100%。重点应用场所光传送网(OTN)覆盖率达到80%，算力全光调度OXC节点占比达到80%，骨干网、城域网全面支持IPv6、SRv6等创新技术使用占比达到40%，实现省内城市内1ms、城市间3ms、到京津冀枢纽节点5ms、到其它枢纽节点20ms的运力网络。
2024.2.2	贵州	《贵州省算力基础设施高质量发展行动计划(2024-2025年)》	到2025年，计算供给均衡合理，总算力规模提升到80EFLOPS，超算算力与智算算力占比达到 35% 。存储保障安全可靠，存储总量超过60EB，全省先进存储容量占比达到30%以上，重点行业的核心数据、重要数据的灾备存储覆盖率达到100%。算力设施绿色集约，加强对存量数据中心的改造提升，推动数据中心采用高效节能冷却系统，全省新建大型以上数据中心PUE值低于1.2。
2024.2.5	青海	《青海省绿色算力基地建设方案》	到2025年，立足西部、服务全国的青海省绿色算力基地初步建成。全省数据中心标准机架规模超过10万架，上架率达到65%以上。算力总规模超过2.06EFLOPS，其中智能算力占比超过 35% 。省内数据中心间时延低于2ms，至国家算力枢纽节点时延不高于20ms，重点场所算力网络(OTN)通达率超过50%，1ms全光城市数量达到5个。存力总规模超过10.7EB，先进存储容量占比达到30%以上，数据中心核心数据灾备存储

			达到 100%。新建(改扩建)大型及以上数据中心 PUE 严格控制在 1.2 以下,并积极申请获得国家级绿色数据中心认证;清洁能源利用率达到 100%,实现清洁能源可追溯。
2024.3.22	上海	《上海市智能算力基础设施高质量发展“算力浦江”智算行动实施方案(2024-2025年)》	到 2025 年,本市智能算力规模超过 30EFlops,占比达到总算力的 50%以上。算力网络节点间单向网络时延控制在 1 毫秒以内。智算中心内先进存储容量占比达到 50%以上。本市新建智算中心国产算力芯片使用占比超过 50%,国产存储使用占比超过 50%,服务具有国际影响力的通用及垂直行业大模型设计应用企业超过 10 家。本市新建智算中心 PUE 值达到 1.25 以下,存量改造智算中心 PUE 值达到 1.4 以下。智算中心内绿色能源使用占比超过 20%,液冷机柜数量占比超过 50%。
2024.3.22	广东	《广东省算力基础设施高质量发展行动暨“粤算”行动计划(2024-2025年)》	到 2025 年,在算力方面,算力规模达到 38EFLOPS,智能算力占比达到 50%。建成智能计算中心 10 个。在存储力方面,存储总量超过 260EB,先进存储容量占比达到 30%以上,重点行业核心数据、重要数据灾备覆盖率达到 100%。力争到 2025 年底,新增国产化算力占比达到 70%。
2024.4.23	江苏	《江苏省算力基础设施发展专项规划》	算力基础设施方面,到 2030 年,全省数据中心机架规模达 120 万标准机架,全省在用总算力超过 50EFLOPS,智能算力占比进一步提升,占比超过 45%,智算中心数量突破 20 个,算力算效水平显著提高;存储总量超过 500EB,进一步提升先进存储容量占比至 40%以上。到 2030 年,全国一体化算力网络长三角国家枢纽节点数据中心集群平均 PUE(电能利用效率)低于 1.25,太阳能、风能、冷能等可再生能源利用率达到 25%以上。基础软硬件和云操作系统实现自主可控,新建算力中心国产算力芯片使用占比达 70%以上、国产存储使用占比达 90%以上。
2024.4.24	北京	《北京市算力基础设施建设实施方案(2024—2027年)》	改变智算建设“小、散”局面,集中建设一批智算单一大数据集群,到 2025 年,本市智算供给规模达到 45EFLOPS。到 2027 年,实现智算基础设施软硬件产品全栈自主可控,整体性能达到国内领先水平,具备 100%自主可控智算中心建设能力。本市新建和改扩建智算中心 PUE 值一般不超过 1.25,年能耗超过 3 万吨标煤的大规模先进智算中心 PUE 值一般不超过 1.15。推进本市存量数据中心升级改造,到规划期末所有存量数据中心 PUE 值均不高于 1.35。
2024.4.30	甘肃	《甘肃算力基础设施高质量发展三年行动计划(2024-2026年)》	到 2026 年,甘肃算力枢纽节点“东数西算”工程取得阶段性成果,一批面向全国提供算力服务的应用示范案例落地。全省算力规模超过 30EFLOPS,智能算力占比达到 50%,提升算力基础设施国产化率,力争在 2026 年国产化算力占比达到 70%以上,存储总量超过 65EB,先进存储容量占比达到 32%以上,推动人工智能算力赋能新型工业化,着力在工业领域打造 20 个可复制推广的标杆人工智能应用场景。
2024.5.23	河北	《河北省人民政府办公厅关于进一步优化算力布局推动人工智能产业创新发展的意见》	原则上平均机架设计功率不低于 8KW,PUE 不高于 1.5,上架率不低于 70%。到 2025 年,算力基础设施布局进一步优化完善,智能算力供给能力显著提升,全省算力规模达到 35EFlops 以上,智能算力占比达到 35%左右,新增算力基础软硬件设施自主可控比例 60%以上。人工智能应用行业进一步拓展,在智能制造、医疗健康等优势领域孵化一批行业应用大模型,培育典型应用场景 30 个。到 2027 年,多元算力泛在融合,智能算力规模全国领先,数据要素高质量供给,大模型孵化应用走在全国前列。
2024.6.3	陕西	《陕西省加快推动人工智能产业发展实施方案(2024-2026年)》	2026 年建设运营智能算力达到 3000P 以上,可统筹的公共智能算力达到西部领先水平。
2024.6.6	山东	《山东省算力基础设施高质量发展行动方案》	到 2025 年,全省数据中心在用标准机架总数达到 45 万个,存力规模达到 65EB,先进存储占比达到 35%以上。总算力达到 12.5EFLOPS,智能算力占比达到 35%。全省各类数据中心总出口带宽达到 25 万 Gbps,10G-PON 及以上端口占比达到 65%以上,枢纽间网络延迟降至毫秒级,总体水平进入全国前列,打造绿色低碳高质量发展先行区算力底座。
2024.6.29	西藏	《“算力珠峰”高质	到 2026 年,全区算力规模达到 100PFLOPS,智算算力占比达到 60%;存储总量超

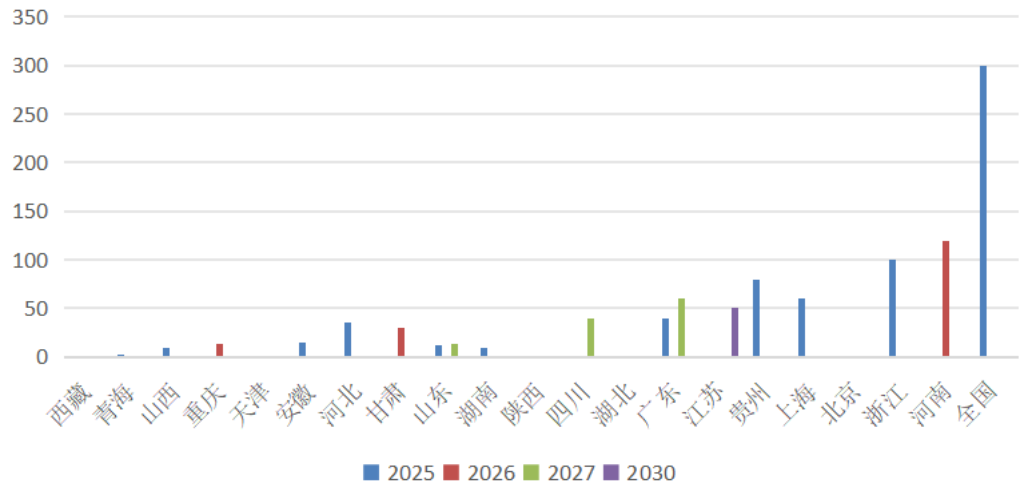
		量发展行动计划 (2024—2026)》	过 350PB, 先进存储占比达到 40%以上; 重点应用场所光传送网覆盖率达到 90%, 分段路由等创新技术使用占比达到 50%; 在能源、工业、教育、医疗、旅游等领域打造 5 个以上应用标杆。
2024.7.18	天津	《天津市算力产业发展实施方案(2024—2026年)》	到 2026 年, 全市算力中心国产算力芯片使用占比超过 60%, 力争智能算力规模达到 10EFLOPS 以上, 建成 5 个以上赋能成效显著的大规模智算中心, 全市超级算力应用率达到 70%以上, 既有大型及以上数据中心能源利用效率(PUE)值力争降至 1.4 以下, 新建大型及以上数据中心 PUE 值降至 1.25 以下, 算力赋能标杆应用场景超过 30 个, 先进存储容量占比达到 30%以上, 实现数据中心核心数据 100%容灾备份。
2024.9.5	湖南	《湖南省人工智能产业发展三年行动计划(2024—2026年)》	到 2026 年将全省的人工智能产业规模突破 1000 亿元, 智能算力达到 3600PFlops
2024.11.6	重庆	《重庆市算力高质量发展三年行动计划(2024—2026年)》	到 2026 年, 重庆总纳管算力规模达到 35EFLOPS, 全市算力规模达到 14EFLOPS, 智算规模占比达到 50%, 绿色算力占比达到 60%, 省际算力直连城市数达到 35 个, 算力使用成本下降 50%, 算力核心产业规模达到 1500 亿元, 优秀行业大模型产品数达到 100 个, 赋能中小微企业数字化转型 2000 家。
2024.11.14	河南	《河南省算力基础设施发展规划(2024—2026年)》	到 2026 年年底, 形成布局合理、绿色低碳、高效集约、安全可靠的算力基础设施发展格局, 全省算力基础设施标准机架数达到 35 万架, 平均利用率达到 70%以上, 算力规模超过 120EFlops, 智算、超算等高性能算力占比超过 90%, 形成 10 个以上算力规模达到 E 级的大型算力中心, 算力中心平均利用率 70%, 算力中心平均 PUE 小于 1.3, 算力中心国产芯片使用占比 60%。力争到 2026 年, 全省智算规模达到 100EFlops 以上。
2024.11.21	四川	《四川省算力基础设施高质量发展行动方案(2024-2027年)》	到 2027 年, 计算力方面, 算力总规模达到 40EFLOPS, 其中智能算力占比不低于 60%, 大规模智能计算中心达到 15 个左右, 数据中心平均 PUE 降低到 1.3 以下。运载力方面, 实现省内各城市之间网络时延低于 3 毫秒, 天府集群到重庆集群网络时延低于 3 毫秒, 到京津冀、长三角、粤港澳大湾区等枢纽节点网络时延低于 18 毫秒。5G 基站总数达 32 万个, 其中 5G-A 基站总数达 4 万个。存储力方面, 存储总量超过 100EB, 先进存储容量占比达到 35%以上, 重点行业核心数据、重要数据灾备覆盖率达到 100%。
2024.12.23	山东	《山东省人工智能产业科技创新行动计划(2025—2027年)》	力争到 2027 年, 全省算力规模达到 14EFlops (每秒百亿亿次浮点运算), 智能算力占比达到 40%。
2025.1.17	云南	《云南省算力基础设施高质量发展实施方案》	到 2025 年, 算力规模超过 3000PFLOPS, 智能算力占比达到 35%。力争存储总量超过 2000PB, 先进存储占比达到 30%以上, 重点行业核心数据、重要数据灾备覆盖率达 100%。
待发布	浙江	《浙江省“人工智能+”行动计划(2025-2027年)》	将聚焦“算力强省”, 进一步推动存量算力设施改造升级, 集约化布局建设 5 个万卡智算集群, 力争 2025 年全省算力规模超过 100EFlops; 出台智能券专项政策, 对人工智能企业使用算力、语料、模型成本提供一定比例补贴, 显著降低浙江人工智能企业发展成本
待发布	内蒙古	《内蒙古算力枢纽节点建设方案》	

资料来源: 地方政府网站, 中原证券研究所

对比此前出台的算力规划来看, 我们看到目前各省在算力加码的趋势明显, 继河南省算力规划创新高(2026 年 120 EFlops)以后, 浙江省算力建设规模规划也呈现出赶超趋势(不

是相同年份，无法直接对比）。配合着 DeepSeek、宇树科技等科技厂商的快速崛起，浙江的算力规划也将起到重要的配套作用。

图 18：各地算力规划中关于总算力的相关目标（EFLOPS）



资料来源：政府网站，中原证券研究所

同时，我们看到随着大模型加速发展，万卡集群建设也成为建设趋势。

1.5.4. 《国家数据基础设施建设指引》发布，当前仍处于起步建设阶段

2024 年 12 月 31 日，国家发展改革委、国家数据局、工业和信息化部联合发布了《国家数据基础设施建设指引》，提出到 2029 年的发展目标为：基本建成国家数据基础设施主体结构，初步形成横向联通、纵向贯通、协调有力的国家数据基础设施基本格局，构建协同联动、规模流通、高效利用、规范可信的数据流通利用体系，协同构筑数据基础设施技术和产业良好生态，国家数据基础设施建设和运营体制机制基本建立。

同时值得关注的是，该政策认为目前“我国数据基础设施处于起步建设阶段”，在未来规划方面基本按照三步走的战略：2024—2026 年的 2—3 年左右时间做试点工作、统一数据登记标准规范、完成顶层设计，2027—2028 年在大中型城市各种跨区域、跨部门的规模化数据流通，并在 2029 年实现“国家数据基础设施建设和运营体制机制基本建立”。

2025 年 1 月，国家数据局批复了山东和贵州两个国家数据要素综合试验区 2022 年 12 月 2 日，中共中央、国务院发布了《关于构建数据基础制度更好发挥数据要素作用的意见》（以下简称“数据二十条”），提出了构建数据基础制度体系。其中浙江省成为了“数据二十条”唯一明确的先行先试地区，提出创建国家数据要素综合试验区总体思路。国家数据要素综合试验区的建设，将对推动数据基础制度体系建设、培育数据要素市场、发挥示范引领作用具有重要意义。

表 7: 国家数据要素综合试验区批复情况

批复时间	区域	该区域获批的其他数据相关国家级试验区
2023.2.24	浙江	国家数字经济创新发展试验区
2025.1.7	山东	
2025.1.14	贵州	国家大数据综合试验区、数字经济发展创新区

资料来源：国家数据局，中原证券研究所

1.5.5. 百度昆仑芯的万卡集群落地

根据百度智能云公众号 2 月 5 日消息，百度智能云点亮昆仑芯三代万卡集群，成为国内首个正式点亮的自研万卡集群，同时将进一步点亮 3 万卡集群。

表 8: 国内主要万卡集群落地情况

时间	公司	概况
2023.7	华为	华为昇腾 AI 集群规模从最初的 4000 卡集群扩展至 16000 卡，是国内首个万卡 AI 集群。
2023.10.24	科大讯飞、华为	科大讯飞、华为联合发布万卡集群算力平台“飞星一号”
2024.3.22	中国电信	天翼云上海港国产万卡算力池启用，这是国内首个投入正式运营的国产单池万卡液冷算力集群。
2024.4.28	中国移动	宣布 2024 年将商用哈尔滨、呼和浩特、贵阳三个自主可控万卡集群，总规模近 6 万张 GPU 卡
2024.6.28	中国移动	中国移动智算中心（呼和浩特）全部集群正式上线。该智算中心是全球运营商最大单体液冷智算中心，定位于服务全国的万卡训练算力池，建设近 2 万张 AI 加速卡，可提供 6.7EFlops（FP16）智能算力。项目采用冷板式液冷、智能母线等先进节能技术，PUE 可低至 1.15
2024.7.6	中国联通	中国联通首个全液冷国产超万卡智算集群——上海临港智算中心（机架数 1.5 万架）正式发布
2024.7.19	中国联通	宣布打造上海、呼和万卡智算集群，全网智算算力超 15EFLOPS
2024.8.8	中国电信	目前上海、北京两个万卡集群已经投产。
2024.10.18	中国移动	中国移动黑龙江公司打造基于国产生态体系的运营商单集群规模最大超万卡智算集群，在哈尔滨智算中心共计部署 1.8 万张 AI 加速卡。该集群具有单集群算力规模最大，国产化网络设备组网规模最大，国内智能融合分级存储规模最大等特点，可实现集群故障的分钟级定界定位，提供算力 6.9EFLOPS。
2024.12.17	中国联通	上海联通携手加佳科技、沐曦共同打造曦源一号 SADA 万卡集群算力项目，其第一期千卡国产算力集群于 11 月 29 日在上海联通临港机房正式落地。该项目于今年 8 月份启动，以国产自主可控的沐曦 GPU 芯片技术产品为核心
2024.12.26	小米	据界面新闻等报道，小米正着手搭建自己的 GPU 万卡集群，将对 AI 大模型大力投入。小米大模型团队在成立时已有 6500 张 GPU 资源。
2025.1.16	中国联通、阿里云	中国联通与阿里云的合作项目——万卡绿色算力项目盛大揭幕，是国内首个基于云计算调度的大规模智算集群，项目总投资 20 亿元，项目整体规模为 16000 卡。第一期 8000 卡先行启动，其中 2025 年中实现 4000 卡投产运营，快速响应市场对算力的迫切需求；第二期同样规划建设 8000 卡。
2025.2.5	百度	百度智能云点亮昆仑芯三代万卡集群，成为国内首个正式点亮的自研万卡集群，同时将进一步点亮 3 万卡集群。

资料来源：澎湃，新浪财经，雷锋网，西宁晚报，人民日报，中原证券研究所

英伟达计划在 2025 年 3 月 17 日-20 日的 GTC 大会上发布下一代 GB300 AI 服务器平台，由功率为 1400W 的 B300 组成。

表 9: 英伟达、AMD 人工智能芯片参数对比

	H100 SXM	H200 SXM	B200	B300	MI300X	MI325X	MI355X
厂商	英伟达	英伟达	英伟达	英伟达	AMD	AMD	AMD
发布时间	2022.3	2023.11	2024.3	2025.3	2023.6	2024.6	2025
架构	Hopper	Hopper	Blackwell	Blackwell Ultra	CDNA 3	CDNA 3	CDNA 4
工艺	4nm	4nm	4nm		5nm	5nm	3nm
晶体管数量	800 亿	800 亿	2080 亿		1530 亿	1530 亿	
最大功率	700W	700W	1000W	1400W	750W	1000W	
显存容量	80GB	141GB	192GB	288GB	192GB	256GB	288GB
显存带宽	3.35TB/s	4.8TB/s	8TB/s		5.3TB/s	6TB/s	8TB/s
显存颗粒	HBM3	HBM3e	HBM3e 8hi	HBM3e 12hi	HBM3	HBM3e	HBM3e 12hi
传输速率	900GB/s	900GB/s	1800GB/s		896GB/s		
FP64	34 TFLOPS	34 TFLOPS					
FP64 Tensor Core	67 TFLOPS	67 TFLOPS					
FP32	67 TFLOPS	67 TFLOPS					
TF32					653.7 TFLOPS		
TF32 Tensor Core	989 TFLOPS	989 TFLOPS	2.2 PFLOPS				
BFLOAT16	989.4 TFLOPS				1307.4 TFLOPS		
BFLOAT16 Tensor Core	1979 TFLOPS	1979 TFLOPS					
FP16	989.4 TFLOPS				1307.4 TFLOPS	1.3 PFLOPS	2.3 PFLOPS
FP16 Tensor Core	1979 TFLOPS	1979 TFLOPS	4.5 PFLOPS				
FP8	1978.9 TFLOPS				2614.9 TFLOPS	2.6 PFLOPS	4.6 PFLOPS
FP8 Tensor Core	3958 TFLOPS	3958 TFLOPS	9 PFLOPS				
INT8	1978.92TOPS				2614.92 TOPS		
INT8 Tensor Core	3958 TOPS	3958 TOPS	9 PFLOPS				
FP6	-	-	9 FLOPS		-		9.2 PFLOPs
FP4	-	-	18 PFLOPS	27 PFLOPS	-		9.2 PFLOPs

资料来源：英伟达，AMD，中原证券研究所

2. 新闻与公告

美国财政部在 2025 年 1 月 3 日发布声明称，美国财政部已对永信至诚科技集团股份有限公司（Integrity Technology Group）实施制裁，将公司列入 SDN 清单（特别指定国民清单），这家中国公司被控为国家网络攻击提供便利。声明称，股票在上海证券交易所交易的该公司被指控操控与黑客组织“亚麻台风”（Flax Typhoon）有关的僵尸网络。美国联邦调查局（FBI）称，该组织从事网络间谍活动，并对全球连接互联网的计算机和工业设备进行网络攻击。（消息来源：俄罗斯卫星通讯社）

2025 年 1 月 14 日，商务部发布公告，将参与对台湾地区军售的 7 家美国企业列入不可靠实体清单。（消息来源：界面新闻）

2025 年 1 月 14 日，美国政府通过白宫官网正式公布了针对网联汽车的最终限制规定，以规避中国和俄罗斯利用美国网联汽车供应链所带来的美国国家安全风险。最终规定禁止美国销售和进口中国产和俄罗斯产网联汽车软硬件系统及网联汽车，包括车辆互联系统（VCS），或通过蓝牙、蜂窝网络、卫星连接和 Wi-Fi 模块将车辆与外部世界连接起来的系统和零部件，以及自动驾驶系统（ADS），使用 VCS 和 ADS 软件的网联汽车，以及 VCS 硬件设备。对软件和网联汽车的限制规定将于 2027 车型年起生效，对硬件的限制规定将于 2030 车型年起生效。上述最终规定适用于乘用车，而针对商业网联车辆或总重量超过 10000 磅的车辆供应，美国商务部也在计划制定中。（消息来源：盖世汽车）

2025 年 1 月 16 日，商务部就国内有关芯片产业反映自美进口成熟制程芯片低价冲击国内市场事答记者问。商务部称，国内有关芯片产业反映，一段时间以来，拜登政府对芯片行业给予了大量补贴，美企业因此获得了不公平竞争优势，并对华低价出口相关成熟制程芯片产品，损害了中国国内产业的合法权益。对此商务部表示，调查机关将按照中国相关法律法规，遵循世贸组织规则进行审查，并将依法启动调查。（消息来源：新快报）

2025 年 1 月 20 日，DeepSeek 发布了全新的开源推理大模型 DeepSeek-R1，在数学、编程和推理等多个任务上达到了与 OpenAI o1 相当的表现水平，同时将应用程序编程接口（API，Application Programming Interface）调用成本降低了 90-95%。（消息来源：DeepTech 深科技）

据新华社报道，中共中央政治局常委、国务院总理李强 2025 年 1 月 20 日下午主持召开专家、企业家和教科文卫体等领域代表座谈会，听取对《政府工作报告（征求意见稿）》的意见建议。座谈会上，张辉、任少波、刘璐、梁文锋、魏洪兴、陈学东、陈红彦、杜斌、邹敬园等先后发言。其中，梁文锋正是头部量化私募幻方量化创始人、DeepSeek 的创始人。（消息来源：澎湃）

当地时间 2025 年 1 月 21 日，美国总统特朗普在白宫宣布 5000 亿美元项目，加大美国数据中心的投资。OpenAI、软银和美国最大的数据中心运营商之一甲骨文（Oracle），以及

总部位于阿布扎比的投资公司 MGX 成立新合资企业星际之门 (Stargate)，特朗普称该公司是“历史上最大的人工智能基础设施项目”。(消息来源：澎湃)

财联社 2025 年 1 月 22 日电，字节跳动计划今年在人工智能基础设施上投入逾 120 亿美元。字节跳动已为 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。(消息来源：财联社、英国金融时报)

当地时间 2025 年 1 月 31 日，人工智能 (AI) 巨头 OpenAI 正式发布推理模型 o3-mini，支持 low、medium 和 high 低中高三种推理难度。与 o1 类似，o3-mini 针对 STEM (科学、技术、工程、数学) 推理进行了优化。o3-mini-medium 的能力在数学、编码和科学方面与 o1 表现相当，响应速度更快。专家测试人员的评估表明，与 o1-mini 相比，o3-mini 的答案更准确、更清晰，推理能力更强。(消息来源：澎湃)

当地时间 2025 年 1 月 24 日，针对半导体设备巨头阿斯麦对华出口问题，荷兰首相斯霍夫表示，中国是荷兰“极其重要”的贸易伙伴，荷兰并不认为特朗普上台后会改变半导体设备出口政策，但荷兰会自行决定出口政策。(消息来源：观察者网)

2025 年 2 月 6 日，OpenAI 宣布向所有用户开放 ChatGPT 搜索功能，无需注册。(消息来源：光明网)

3. 河南计算机行业动态

3.1. 河南计算机行业要闻

经省政府同意，省财政厅、省市场监管局日前联合印发的《河南省知识产权领域省与市县财政事权和支出责任划分改革方案》(以下简称《方案》)，已于 2025 年 1 月 1 日起实施。《方案》从知识产权宏观管理、授权确权、运用促进、知识产权保护、公共服务、涉外工作和其他事项等七个方面，划分知识产权领域的省与市县财政事权和支出责任。(消息来源：河南日报)

2024 年 12 月 31 日上午，河南省自然资源厅在郑州市举行河南省卫星导航定位基准服务系统“北斗三号”差分信号启用仪式，标志着河南省卫星导航定位服务系统步入“北斗三号”新时代。(消息来源：河南日报)

1 月 3 日，安阳市在全省范围内首次开展山区无人机邮政快递航线测试飞行。(消息来源：河南日报)

1 月 19 日，记者从省发展改革委获悉，《关于印发 2025 年河南省重点建设项目名单的通知》于近日正式印发，明确河南省 2025 年省重点项目 1037 个、总投资约 3.1 万亿元，力争年度完成投资 1 万亿元左右。项目涵盖了创新驱动能力提升、基础设施建设、新型基础设

施建设、产业转型发展、绿色低碳转型、民生社会事业改善等六大领域。其中新型基础设施建设领域遴选项目 18 个、总投资 454 亿元、年度计划投资 131 亿元，主要建设国家超算互联网核心节点、国家交通大数据产业园、中国移动 5G 建设、中国电信河南通信枢纽等项目。（信息来源：大河网）

1 月 16 日至 21 日，2024 世界机器人大赛总决赛激情开战，来自国内外超 4 万名机器人爱好者、机器人行业产学研代表汇聚郑州。

2 月 7 日获悉，今年春节期间，位于鹤壁市城乡一体化示范区的京东云（中原）算力中心上线了 DeepSeek 模型部署和推理服务，成为河南省第一个提供该服务的城市级算力中心。（信息来源：河南日报）

2 月 8 日，河南空港智算中心已同时接入全量级 DeepSeek-R1 和多模态 DeepSeek-Janus-Pro 模型。（信息来源：河南日报）

3.2. 河南计算机行业数据跟踪

2024 年河南先进计算产业链、新能源汽车产业链增加值分别同比增长 71.3%、46.1%。

2024 年，全省限额以上单位新能源汽车、家用电器和音像器材等商品零售额同比分别增长 30.8%、15.7%。出口电动汽车 88.5 亿元，同比增长 91.4%。

2024 年，全省高技术制造业、工业战略性新兴产业增加值分别同比增长 12.0%、9.2%，分别高于全省规上工业增加值增速 3.9、1.1 个百分点；新能源汽车产业、新一代信息技术产业、新材料产业增加值分别同比增长 30.3%、11.8%、11.6%。全省集成电路、锂离子电池、服务器、光电子器件、电子计算机整机、工业机器人等新产品产量分别同比增长 2.5 倍、84.3%、32.2%、32.0%、31.6%和 15.6%。

在中原科技城、超聚变研发中心及总部基地、国家超算中心等项目带动下，2024 年，全省高技术制造业投资同比增长 20.9%，其中航空航天器及设备制造业、计算机及办公设备制造业投资分别同比增长 2.6 倍、1.7 倍。

全省研究与试验发展（R&D）经费投入预计连续 4 年超千亿元。

工业战略性新兴产业、高技术制造业增加值占规上工业增加值的比重分别达 24.8%、14.0%。

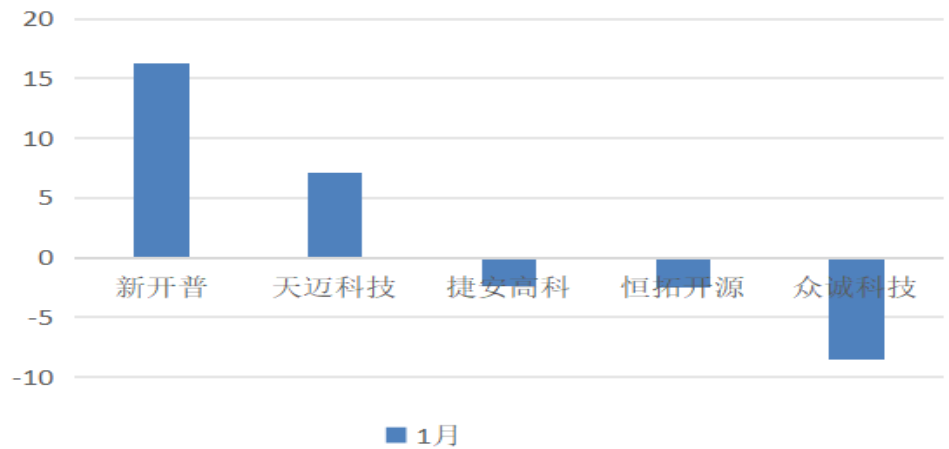
高技术制造业投资占固定资产投资的比重为 5.7%、比 2023 年提高 0.7 个百分点。

（数据来源：河南省统计局）

3.3. 河南上市公司行情回顾

1月，5家河南计算机行业上市公司表现分化。其中新开普以16.31%领涨，天迈科技上涨7.09%，捷安高科、恒拓开源、众诚科技分别下跌2.34%、2.50%、8.56%。

图 19：河南上市公司近期股价涨跌幅表现（%）



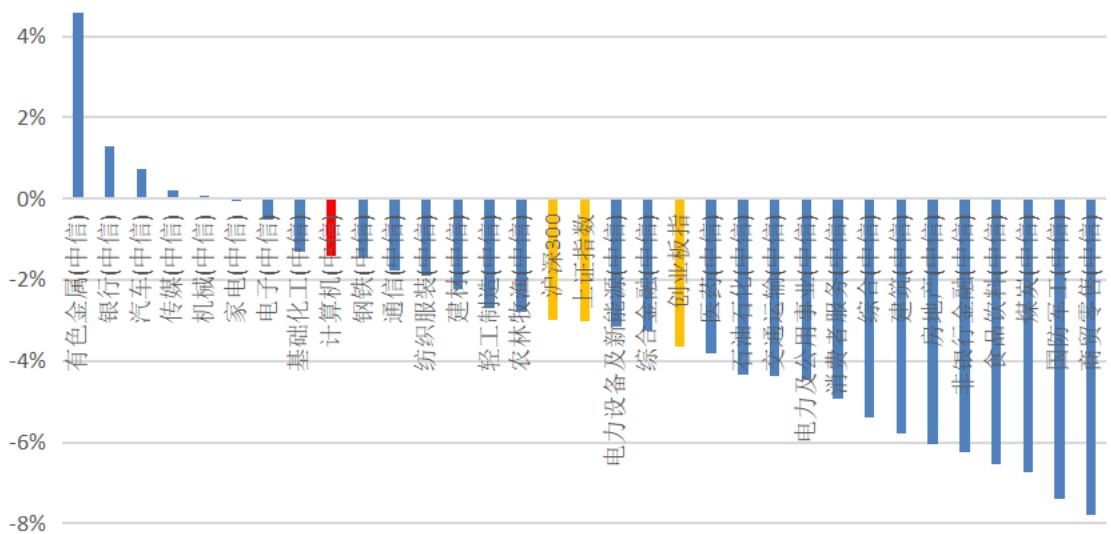
资料来源：Wind，中原证券研究所

4. 投资策略

4.1. 行情回顾：1月行业略跑赢沪深300指数

1月计算机行业步略跑赢沪深300指数。1月中信计算机指数下跌了1.40%，跑赢上证指数1.62PCT，跑赢创业板指数2.23PCT，跑赢沪深300指数1.60PCT，在30个中信一级行业中排名倒数第9。

图 20：2025年1月中信一级子行业涨跌幅

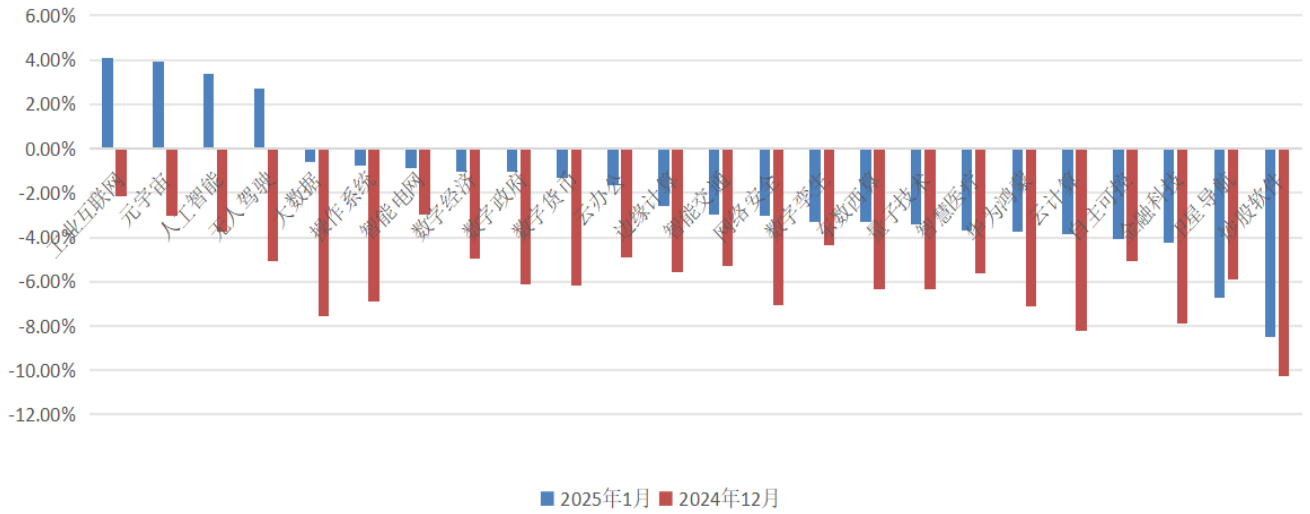


资料来源：Wind，中原证券研究所

从概念指数表现来看，1月计算机相关概念以下跌为主。其中工业互联网、元宇宙、人工

智能、无人驾驶皆上涨，涨幅分别为 4.10%、3.91%、3.37%、2.69%，领跌概念为炒股软件、卫星导航、金融科技、自主可控，分别下跌 8.51%、6.74%、4.26%、4.09%。

图 21: 2025 年 1 月计算机行业相关概念涨跌幅

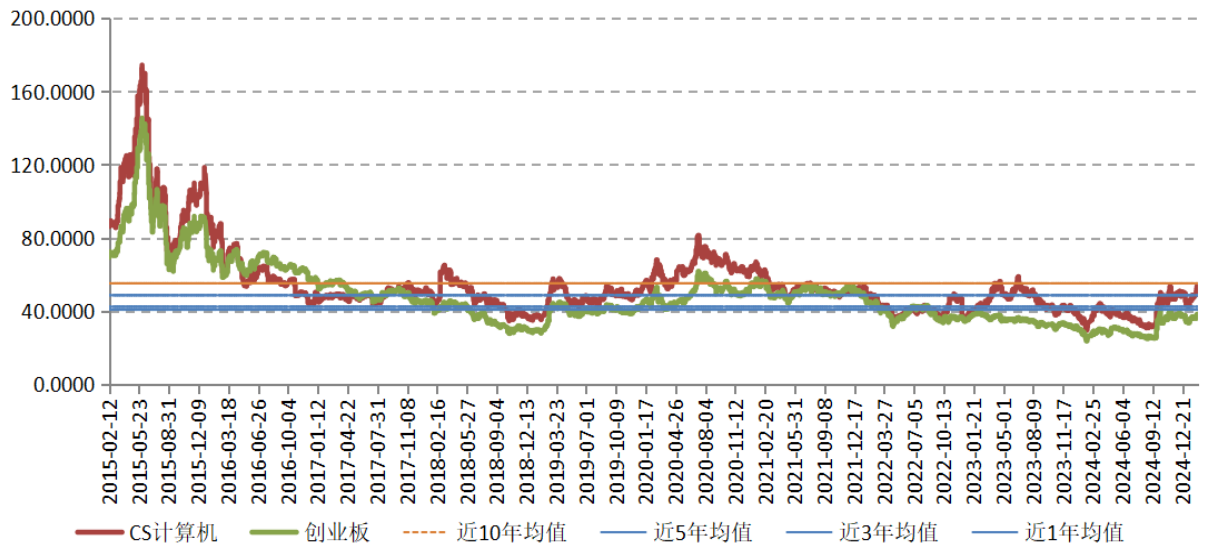


资料来源: Wind, 中原证券研究所

4.2. 估值: 行业估值接近历史均值区间上限

行业的估值位于历史均值之间。根据 Wind 数据, 2025 年 2 月 11 日中信计算机行业 TTM 整体法 (剔除负值) 估值为 53.06 倍, 同期 A 股估值 15.10 倍, 创业板估值 38.06 倍。行业近 1 年、3 年、5 年、10 年的平均估值分别为 40.73 倍、42.36 倍、48.66 倍、55.18 倍, 当前估值仍然位于历史均值水平之间, 但是已经接近估值均值区间上限, 近 3 年最高点 58.79 倍和最低点 29.69 倍。

图 22: 近 10 年中信计算机行业估值水平 (截止 2025.2.11)



资料来源: Wind, 中原证券研究所

4.3. 行业观点与投资建议

2024 年我国软件产业增速呈现回落趋势。根据工信部数据，2024 年软件业务收入 13.73 万亿元，同比增长 10.0%，较 2023 年 13.4% 的增速回落了 3.4PCT，较 1-11 月回落了 0.7PCT。2024 年软件业务利润总额 16953 亿元，同比增长 8.7%，较 2023 年 13.6% 的增速水平回落了 4.9 PCT，较 1-11 月回落了 0.5 PCT。

重点关注子行业的主要数据和动态包括：

(1) 国产化：2024 年，我国集成电路在进口依赖度比例为 78%（即国产化占比 22%），较 1-11 月回落 2 PCT，下滑趋势明显加大，表明美国一系列禁令影响在持续释放。阿斯麦对中国出口量的快速上升，意味着未来国内芯片产量有望得到较大增长。华为 2024 年收入重回美国制裁前的水平，研发投入增长明显。

(2) AI：R1 在后训练阶段大规模使用了强化学习，来取代传统大模型中监督式微调的方法，不需要大量的标注数据，节省了算力，同时实现了与 o1 相媲美的性能。这意味着在 Scaling Law 失效的困扰下，人类在增加训练数据和算力以外，探索出了高效的架构与训练策略提升的新路径，这或将成为推动 OpenAI 的 GPT-5 问世的一个最重要的技术突破。

(3) 算力：2024Q4，美国 5 大科技厂商的资本开支再创新高，总计达到 734.97 亿美元，同比增长 66%，较 Q3 继续提升了 7PCT。DeepSeek-R1 发布后，亚马逊、谷歌、Meta 仍然给出了 2025 年积极的资本投入计划，同时海内外云厂商都快速接入了 DeepSeek 的模型，国产芯片厂商也快速实现了适配。浙江省即将推出人工智能专项规划，力争 2025 年全省算力规模超过 100EFlops，加快趋势明显。

(4) 电子：IC 设计成为 2024 年的软件行业景气度最高的子行业，全年同比增长 16.4%，较 1-11 月回升 1.6 PCT，高于软件行业整体增速 6.4 PCT，较上年同期增长 10.0 PCT。

给予行业同步大市的投资评级。2025 年 2 月 11 日中信计算机行业估值为 53.06 倍，经过节后一轮上涨，当前估值仍然位于历史均值水平之间，但是已经接近估值均值区间上限。由于 DeepSeek-R1 的技术突破，中国在大模型领域的的能力获得了全球的认可，也为后续大模型的发展探索出了一条新的提升路径。因为 DeepSeek-R1 采用开源策略，极大地加快国产芯片的适配，加大下游应用能力的释放。我们认为 DeepSeek-R1 虽然对算力需求降低，但是随着应用端的推理需求大量释放，全球算力需求还将持续快速增长。考虑到 GPT-5 即将发布和 AI 应用的持续落地，我们认为板块还将持续获得市场的关注。

5. 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。