

# 算力平权，国产AI力量崛起

➤ **豆包+DeepSeek破局，国产大模型弯道超车。**豆包和DeepSeek分别在多模态和轻量化两方面加速了国产大模型的发展进程。2024年12月18日，豆包发布豆包视觉理解模型，一举成为国内最领先的多模态大模型之一；同为12月，DeepSeek发布DeepSeek-V3大模型，拥有671B总参数、37B激活参数，成本仅为557.6万美元，但在性能上却与GPT-4o和Claude-3.5-Sonnet同属世界顶尖水平。国内其他模型厂商也加速了追赶节奏，2025年以来，豆包、通义千问、百度、腾讯混元、阶跃星辰和Kimi等其他国产大模型加速了更新迭代，例如通义千问在3月连续发布QwQ-32B与Qwen2.5-Omni等。豆包对多模态的探索，以及DeepSeek对轻量化大模型的升级节奏，有望开创大模型开源、低成本、高性能的新时代，为应用侧公司探索创新AI应用提供巨大便利，**AI应用加速放量下推理侧需求有望提升。**

➤ **向“芯”而行，国产算力破局元年。**在国产大模型密集落地背景下，芯片厂商加速适配国产算力生态。中芯国际N+1工艺已逐步成熟，N+2持续推进，构建国产算力底座；昇腾910C量产落地，920系列研发加快，性能持续逼近国际主流水平；寒武纪、海光等在AI训推方向深度布局，硬件端多点突破，生态融合加快。云端ASIC正成为算力演进主流，谷歌、亚马逊持续加码自研芯片体系；国内翱捷科技、芯原等设计企业快速成长，覆盖多元应用场景，并与海内外头部厂商形成紧密合作，成长弹性充足。在软件层面，适配节奏同样加快。以DeepSeek为例，发布即获得17家芯片厂商支持适配，训推效率大幅提升，**助力算力生态向自主可控稳步迈进。**

➤ **算力基建加码，解决供给短板。**国内云计算厂商正加大算力储备及模型优化投入，AI计算基础设施建设布局逐步清晰，**相关资本开支进入新一轮扩张周期。**海量推理算力需求下行业供需关系显著改善，根据CloudPrice，AWS的云计算(H100)均价近一个月来上涨了76%，**云计算商业模式有望迎来“量价齐升”的局面。**算力基建作为云厂商资本开支落地的核心环节，华勤、浪潮等公司迅速推出适配国产算力的服务器，并针对卡间互联推出ETH-X超节点方案，而短期内，国产算力基建难以满足迅速增长的需求，算力租赁成为破局之道。处于数据隐私性及定制化考虑，以DeepSeek为代表的国产大模型的亦落地本地化部署，算力一体机具备软硬结合、开箱即用、灵活定制的特性，满足本地化部署需求。

➤ **电源及液冷创新，解决功率密度瓶颈。**AI浪潮下，单机柜功率密度提高，对电源及温控系统提出新的挑战。为保证数据中心稳定运行，三级供电/备电系统有所升级，**供电系统从UPS向HVDC演进，超级电容亦将成为全新增量**；温控方面，传统风冷难以满足高功率芯片散热需求，液冷逐步成为刚需，科智咨询预计23年到27年国内液冷市场复合增速达60%，**冷却液、CDU、快接头、manifold等上游零部件成为核心受益环节。**

➤ **投资建议：**国产大模型持续突破，DeepSeek、豆包、MCP等开始引领全球AI产业发展，AI应用落地加速。芯片厂商也在加速适配国产算力生态。以自主可控为基石，云厂商资本开支持续加速。带动服务器、电源、存储等上游产业链环节机遇。建议关注：1) **算力芯片：**中芯国际、海光信息、寒武纪、芯原股份、云天励飞、翱捷科技、沐曦技术(未上市)；2) **服务器：**华勤技术、联想集团、浪潮技术、弘信电子；3) **电源：**禾望电气、麦格米特、中恒电气、泰嘉股份、科泰电源、潍柴重机、科华数据、欧陆通、奥海科技；4) **温控：**申菱环境、英维克、溯联股份、川环科技；5) **其他供应链：**德明利、长光华芯、杰华特、华海诚科、华丰科技、兴森科技等。

➤ **风险提示：**AI应用落地不及预期，云商资本开支不及预期，宏观经济波动。

## 重点公司盈利预测、估值与评级

代码	简称	股价		EPS(元)		PE(倍)			评级
		(元)	2024A	2025E	2026E	2024A	2025E	2026E	
688981.SH	中芯国际	89.90	0.46	0.64	0.84	194	140	107	推荐
688041.SH	海光信息	150.01	0.83	1.32	1.91	181	114	79	推荐
688521.SH	芯原股份	91.79	-1.2	-0.18	0.07	/	/	1311	推荐
603063.SH	禾望电气	31.90	1.00	1.37	1.66	33	23	19	推荐
301018.SZ	申菱环境	40.60	0.43	1.03	1.44	93	39	28	推荐
301397.SZ	溯联股份	36.95	1.03	1.60	2.36	36	23	16	推荐
688048.SH	长光华芯	60.54	-0.57	0.02	0.41	/	3301	149	推荐
688343.SH	云天励飞	56.81	-1.63	/	/	/	/	/	/

资料来源：Wind，民生证券研究院预测（注：股价为2025年5月6日收盘价；未覆盖公司数据采用Wind一致预期）

## 推荐

维持评级



分析师 方竞

执业证书：S0100521120004  
邮箱：fangjing@mszq.com



分析师 邓永康

执业证书：S0100521100006  
邮箱：dengyongkang@mszq.com

分析师 宋晓东

执业证书：S0100523110001  
邮箱：songxiaodong@mszq.com

分析师 李萌

执业证书：S0100522080001  
邮箱：limeng@mszq.com  
研究助理 李伯语

执业证书：S0100123040030  
邮箱：liboyu@mszq.com

## 相关研究

1. 电子行业点评：384架构引领，910C蓄势待发-2025/04/29
2. 电子行业点评：并购整合加速，国产化ED A+IP平台冉冉升起-2025/04/16
3. 半导体行业点评：最新关税政策解读，坚定看好半导体自主可控-2025/04/06
4. 半导体行业点评：关注模拟IC投资机遇-2025/03/18
5. 存储行业动态报告：价格底部企稳，企业级需求构建增量-2025/03/14

# 目录

<b>1 豆包+DeepSeek 破局，国产大模型弯道超车 .....</b>	<b>3</b>
1.1 技术驱动引领，DeepSeek 加速布局 .....	4
1.2 国产模型竞速升级，轻量化浪潮提速 .....	8
1.3 基于 DeepSeek，探讨推理需求大爆发 .....	9
<b>2 向“芯”而行，国产算力破局元年 .....</b>	<b>12</b>
2.1 先进制程：中芯中流砥柱，筑基国产算力 .....	12
2.2 华为昇腾：384 架构引领，910C 蓄势待发 .....	13
2.3 算力双龙：攻坚突围、加速落地 .....	16
2.4 ASIC：云厂商的曙光，挑战 GPU 垄断 .....	20
2.5 算力协力模型，实现生态国产化 .....	23
<b>3 算力基建加码，解决供给短板 .....</b>	<b>26</b>
3.1 云厂商加大算力储备及模型投入，资本开支高增 .....	26
3.2 国产算力基建元年，构建算力云服务新生态 .....	28
3.3 服务器：算力之基座，全面适配国产算力 .....	31
3.4 算力租赁：短期内算力短缺破局之道 .....	32
3.5 本地部署需求旺盛，助推 AI 一体机发展 .....	35
<b>4 电源及液冷创新，解决功率密度瓶颈 .....</b>	<b>42</b>
4.1 全新电源架构，数据中心能源卫士 .....	42
4.2 液冷接棒传统风冷，温控系统革新升级 .....	46
<b>5 行业投资建议 .....</b>	<b>50</b>
5.1 长光华芯：国产光通信芯片新进者 .....	50
5.2 禾望电气：新能源与电气传动领跑者 .....	53
5.3 申菱环境：数据中心温控核心供应商 .....	55
5.4 液联股份：深耕快接头及管路，开拓数据中心液冷 .....	58
<b>6 风险提示 .....</b>	<b>62</b>
<b>插图目录 .....</b>	<b>67</b>
<b>表格目录 .....</b>	<b>68</b>

## 1 豆包+DeepSeek 破局，国产大模型弯道超车

总书记指出“人工智能是引领这一轮科技革命和产业变革的战略性技术”，从顶层建筑角度为国产 AI 发展指引了战略方向。而近期国产大模型也持续突破。豆包、Deepseek、MCP 等开始引领全球 AI 产业发展，为 AI 应用落地加速提供了技术底座。

**1) 豆包大模型：专注多模态融合。**推动了国产大模型在多模态领域的能力升级。豆包大模型始于 2023 年 8 月 17 日字节的 AI 对话产品“豆包”公测，2024 年，豆包逐步补全了语音、图像、代码等能力，并在 12 月 18 日发布豆包视觉理解模型，实现了更强的内容识别、理解和推理、以及视觉描述等能力，一举成为国内最领先的多模态大模型之一，引领国产化大模型在多模态领域的升级之路。

**2) DeepSeek：专注深度思考+轻量化。**相较于豆包在多模态方面的领先，DeepSeek 则更注重大模型的轻量化升级能力，在成本和性能方面做到极致。2024 年 12 月，DeepSeek 发布 DeepSeek-V3 大模型，拥有 671B 总参数、37B 激活参数，成本仅为 557.6 万美元，但在性能上却与 GPT-4o 和 Claude-3.5-Sonnet 同属世界顶尖水平；2025 年 1 月 20 日，DeepSeek 发布 DeepSeek-R1 推理大模型，性能追平 OpenAI-o1 正式版。**由于性能和成本端的优势，当前 DeepSeek 在国内外均得到了高度的认可：**国外方面：微软 CEO 纳德拉高度认可 DeepSeek 的新模型，并表明 Azure AI Foundry 和 GitHub 已经融入 DeepSeek AI，英伟达认为 DeepSeek-R1 是具有最先进推理能力的开放模型；国内方面：华为昇腾、沐曦、海光信息、壁仞科技、昆仑芯、云天励飞等 AI 芯片企业纷纷官宣适配 DeepSeek 大模型。我们认为，DeepSeek 开创了大模型开源、低成本的新时代，为应用侧公司探索创新 AI 应用提供了极大便利，AI 加速放量下推理侧算力需求有望增长。据电子工程专辑，**DeepSeek-R2 或将发布，该模型参数总量将达到 1.2 万亿，相较 R1 的 6710 亿提升一倍，并将采用更先进的混合专家模型 MoE 和门控网络层 Gating Network，进一步提升推理性能。**

**3) MCP：重新定义模型和数据的交互模式，推动 AI Agent 落地。**2024 年 11 月，Anthropic 推出 MCP (Model Context Protocol，模型上下文协议)，用于统一大语言模型与外部数据源和工具之间的通信协议。该协议为 AI 模型连接不同的数据源和工具提供了标准化的方式，就像是 AI 应用程序的 USB-C 接口，重新定义了模型和外部世界的互动模式，通过 MCP Client 和 MCP Server，提升了 AI Agent 从调用数据到做出决策的便利性，极大拓展了 AI Agent 的能力范围和应用场景。

**豆包、DeepSeek 和 MCP 的异军突起加快了国产大模型的迭代节奏。**2025 年以来，豆包、通义千问、百度、腾讯混元、阶跃星辰和 Kimi 等其他国产大模型加速了更新迭代，豆包推出 Doubao-1.5-pro，通义千问连续发布 QwQ-32B 与 Qwen2.5-Omni，各大模型厂商均加速了各自的能力迭代速度，提升了国产大模

型的综合性能。

**图1：豆包和 DeepSeek 突破后，国产大模型的迭代时点**



资料来源：各公司官网，民生证券研究院

## 1.1 技术驱动引领，DeepSeek 加速布局

### 1.1.1 量化巨头助力，DeepSeek 崛起之路

DeepSeek (杭州深度求索人工智能基础技术研究有限公司) 是 AI 领域的专业技术研发公司，致力于开发先进的大语言模型 (LLM) 和相关技术，由国内知名量化巨头幻方量化于 2023 年 7 月 17 日创立。幻方量化作为国内最早确立以 AI 为主要发展方向的公司之一，早在 ChatGPT 3.5 问世前，就拥有 1 万张 NVIDIA A100 显卡，在 GPU 储备上与商汤、百度、腾讯、字节和阿里位列国内同期第一梯队。得益于母公司在 AI 领域的前瞻性布局，DeepSeek 于 2023 年 11 月 2 日便开源了首款代码大模型 DeepSeek Coder，随后不断进行迭代，旗下于 2025 年 1 月 20 日发布的 DeepSeek-R1 推理模型，在性能方面对 OpenAI o1 正式版，表现卓越，被英伟达认为是具有最先进推理能力的开放模型。

**图2：DeepSeek 发展历程**



资料来源：DeepSeek 官网，GitHub，民生证券研究院

### 1.1.2 专注算法优化创新，DeepSeek 核心优势

DeepSeek 始终坚持以创新算法驱动大模型效率与能力的双重跃升，分别通过 1) 构建混合专家架构 DeepSeekMoE；2) 原创注意力机制 MLA；3) 多 token 预测技术 MTP 以及 4) 推理优化技术 GRPO 等创新技术为支点构建了其核心竞争力。

DeepSeek 大模型均采用 MoE 架构（混合专家模型），并进一步改进成

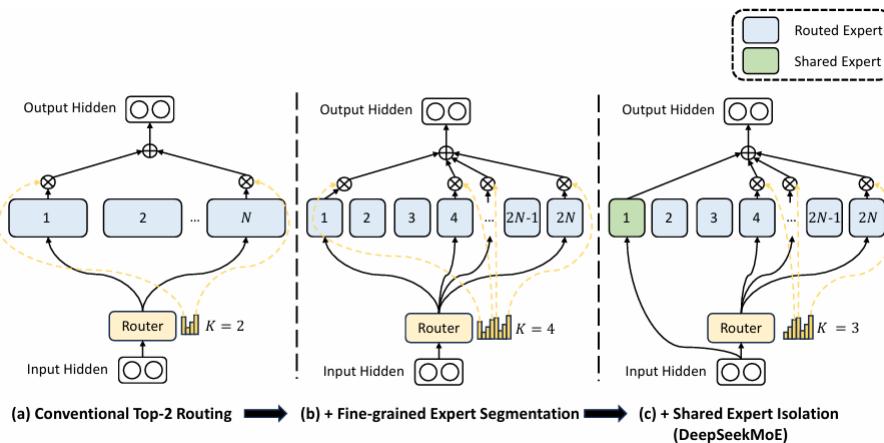
**DeepSeekMoE**。MoE 是一种基于 Transformer 架构的模型, 主要包括稀疏 MoE 层和门控网络, 其中稀疏 MoE 层中包含多个专家, 每一个专家是单独的神经网络, 而门控网络则决定了哪个 token 将被发送给哪个专家。DeepSeek-V3 采用 DeepSeekMoE 架构, 在 MoE 的基础上添加了细粒度专家分割和共享专家隔离, 并采用无辅助损失的负载均衡策略替代传统的专家平衡策略, 使得模型效率进一步提高:

**1) 细粒度专家分割**: 在保持专家总参数数量不变的前提下, 对专家进行更精细的划分, 使得专家获得更高的专业化水平, 在计算过程中等比例增加专家激活数量, 单个 token 能够被分配给更多的专家, 专家组合的多样性和灵活性也有所上升。

**2) 共享专家隔离**: 在所有专家中选取部分专家作为共享专家 (Shared Expert), 所有的 token 都将分配给共享专家进行计算, 从中获取通用知识, 并相应减少其他专家被激活的数量, 从而缓解专家间的参数冗余问题, 提升模型效率。

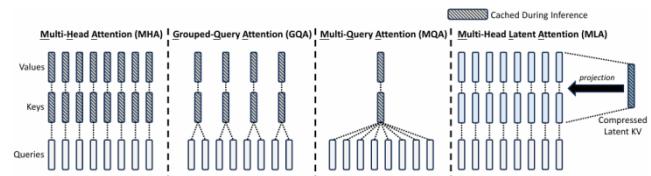
**3) 无辅助损失的负载均衡策略**: 通过动态偏差来实现均衡, 即对低负载专家增加偏差值以提高被选中的概率, 对高负载专家降低偏差值以降低被选中的概率, 并辅以部分辅助损失提高整体分配的稳定性。

**图3: DeepSeekMoE 示意图**



资料来源: 《DeepSeekMoE: Towards Ultimate Expert Specialization in Mixture-of-Experts Language Models》, Dama Dai 等, 民生证券研究院

**MLA (多头潜在注意力架构) 是 DeepSeek 原创的注意力机制, 能够实现性能和 KV 缓存的平衡。** KV 缓存是影响模型推理效率的关键因素, Transformer 架构下的 MHA 会产生大量的 KV 缓存, GQA 和 MQA 作为改进方案虽然减少了 KV 缓存, 但是在性能表现上受损。MLA 通过对 Key 和 Value 进行低秩联合压缩实现了内存的高效利用, 在降低 KV 缓存的同时获得了更强的性能。

**图4：不同注意力机制示意图**


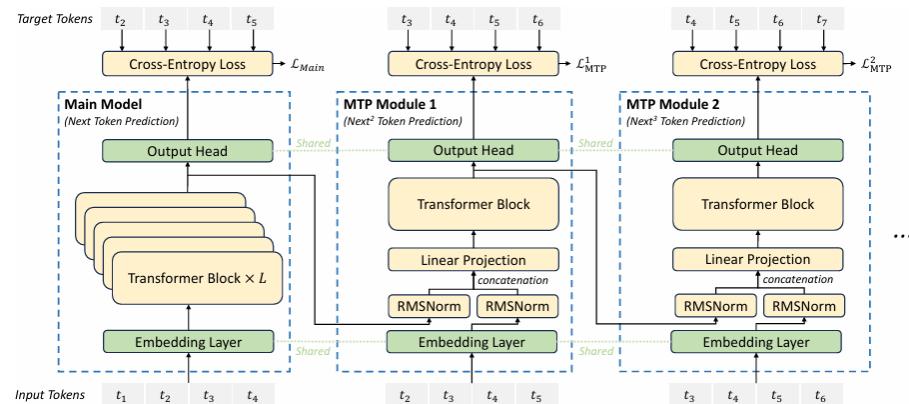
资料来源：DeepSeek-V2 技术报告，民生证券研究院

**图5：不同注意力机制 KV 缓存和性能对比**

Attention Mechanism	KV Cache per Token (# Element)	Capability
Multi-Head Attention (MHA)	$2n_h d_h l$	Strong
Grouped-Query Attention (GQA)	$2n_g d_h l$	Moderate
Multi-Query Attention (MQA)	$2d_h l$	Weak
MLA (Ours)	$(d_c + d_h^R) l \approx \frac{9}{2} d_h l$	Stronger

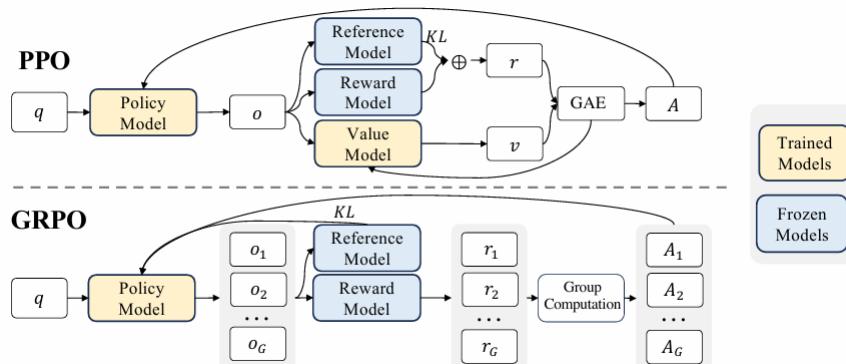
资料来源：DeepSeek-V2 技术报告，民生证券研究院

**MTP (多 token 预测技术) 通过预测多个后续 token 提升模型效率。**传统模型只根据前文预测 1 个 token，即根据  $t_1$  预测  $t_2$ ；而 MTP 技术则要求模型一次性生成多个 token，即在主模型中根据  $t_1$  预测  $t_2$ ，在 MTP Module 1 中根据  $t_1$  预测  $t_3$ ，在 MTP Module 2 中根据  $t_1$  预测  $t_4$ ，依此类推。这意味着通过一次训练便可以学习多个位置的 label，有效提升数据训练效率，从而降低训练成本，并提高模型对于长句的理解能力，一次性生成多个 token 也提高了推理速度。DeepSeek-V3 设定了 2 个 token 预测目标，且实验表明第二个 token 的接受率处于 85%-90%，高接受率保证了模型的解码速度能够达到 1.8 倍 TPS (每秒 token 数)。

**图6：MTP 技术示意图**


资料来源：DeepSeek-V3 技术报告，民生证券研究院

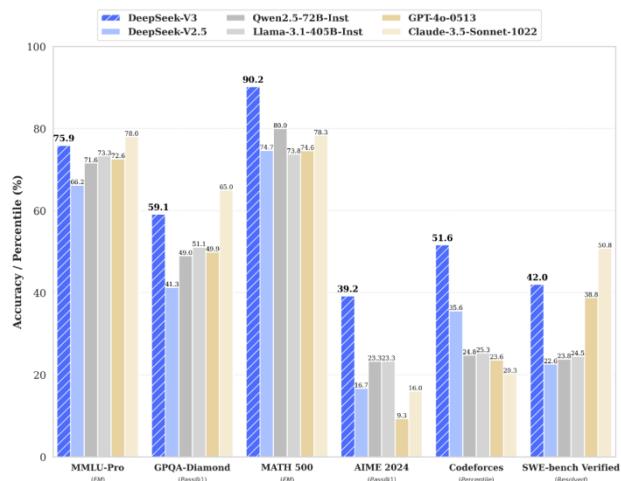
**GRPO (组相对策略优化) 通过取消价值模型提高了模型的推理能力。**传统强化学习方法为 PPO (近端策略优化)，需要同时训练价值模型和策略模型，在计算成本方面存在劣势，而 GRPO 则取消了价值模型，通过比较组内相对表现进行优势评估，显著降低了计算和内存消耗，同时简化了优势估计过程，组间相对得分相较于绝对得分更加适用于推理任务，提高了模型推理能力。

**图7：GRPO 技术示意图**


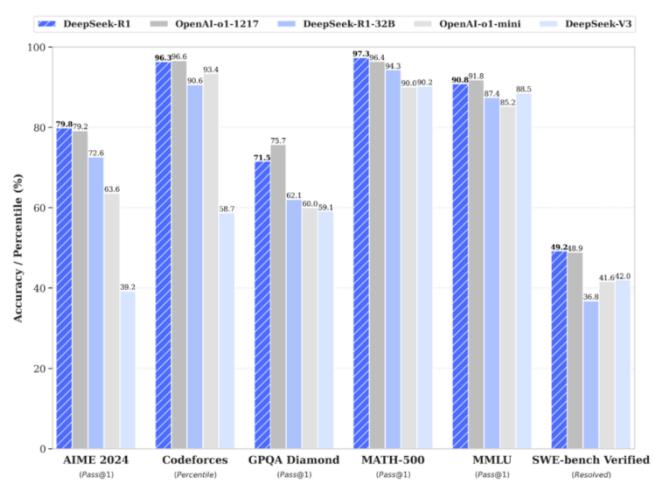
资料来源：《DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models》，Zhihong Shao 等，民生证券研究院

### 1.1.3 算法协同优化，DeepSeek 性能出色

通过对算法、框架和硬件进行协同优化，DeepSeek 大模型在性能上表现出色。DeepSeek 的技术团队擅于在软件层面进行技术优化创新，以此实现大模型的性能飞跃：1) 在 DeepSeek-V3 上，创新性使用无辅助损失的负载均衡策略、FP8 混合精度训练架构、DualPipe 跨节点通信、跨节点全对全通信内核等技术，使模型获得了能够媲美 GPT-4o 和 Claude-3.5-Sonnet 的性能表现，甚至在数学能力上性能大幅领先，达到世界领先水平；2) 在 DeepSeek-R1 上，探索通过大量纯强化学习（不通过 SFT 冷启动）和模型蒸馏提升模型推理能力的可能性，最终模型在性能上比肩 OpenAI o1 正式版。

**图8：DeepSeek-V3 与其他大模型性能对比**


资料来源：DeepSeek 官网，民生证券研究院

**图9：DeepSeek-R1 与其他大模型性能对比**


资料来源：DeepSeek 官网，民生证券研究院

## 1.2 国产模型竞速升级，轻量化浪潮提速

**DeepSeek 加速了国产模型的进展，引领了国产大模型的轻量化趋势。**自 DeepSeek-R1 模型发布以来短短几个月内，豆包、通义千问、百度、腾讯混元、阶跃星辰和 Kimi 等其他国产大模型均有重大更新，在模型轻量化方面取得了显著进展。以豆包和通义千问为例：

**豆包：**2025 年 1 月 22 日豆包发布 Doubao-1.5-pro，模型使用与 DeepSeek 相同的 MoE 架构，通过训练与推理一体化设计，在保持高性能的同时显著降低了推理成本。在相同的 9T tokens 训练数据下，激活参数仅为稠密模型 1/7 的 MoE 模型，超过了稠密模型的性能，实现了 7 倍的性能杠杆提升。同时，Doubao-1.5-pro 在视觉和语音多模态方面实现了全面升级。视觉方面，采用原生动态分辨率架构，支持任意分辨率和长宽比的图像输入，显著提升了模型在视觉推理、文档识别等任务中的表现；语音方面，提出了端到端的 Speech2Speech 框架，实现了语音理解与生成的一体化，提升了对话的自然性和情感表达能力。2025 年 3 月 11 日和 4 月 16 日，豆包还分别更新发布了图像生成大模型 Seedream 2.0 与 Seedream 3.0。

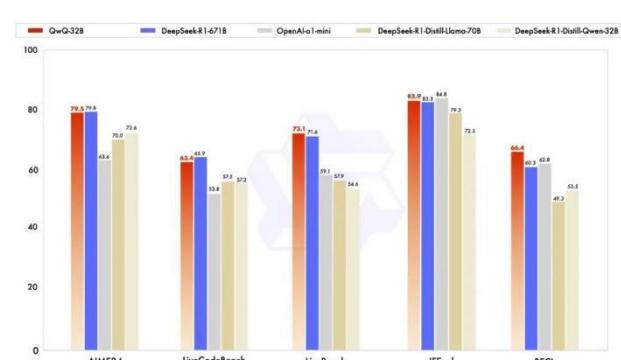
**通义千问：**2025 年 3 月 6 日，阿里巴巴发布并开源全新的推理模型通义千问 QwQ-32B，该模型通过大规模强化学习，在数学、代码及通用能力上实现质的飞跃，作为一款参数仅有 320 亿的轻量化模型，其性能可与具备 6710 亿参数（其中 370 亿被激活）的 DeepSeek-R1 媲美。同时，**QwQ-32B 大幅降低了部署使用成本，在消费级显卡上也能实现本地部署**，适合快速响应或对数据安全要求高的应用场景。在一系列权威基准测试中，QwQ-32B 模型表现出色，在测试数学能力的 AIME24 评测集上和评估代码能力的 LiveCodeBench 中，表现与 DeepSeek-R1 相当，远胜于 o1 - mini 及相同尺寸的 R1 蒸馏模型。2025 年 3 月 26 日，阿里巴巴还发布了多模态理解生成大模型 Qwen2.5-omni-7b。

**图10：豆包 Doubao-1.5-pro 与其他大模型性能对比**

		Doubao-1.5-pro	Llama3-405B	GPT4c-0806	Gemini-exp-1205	Claude-3.5-Sonnet-latest	Qwen2.5	DeepseekV3
Knowledge	IMLU	88.6	88.6	<b>89.7</b>	86.8	88.5	85.6	88.5
	IMLU-PRO	<b>80.1</b>	73.3	74.9	76.4	78.0	71.1	75.9
	GMOA	<b>65.0</b>	51*	53.1	62.1	<b>65.0</b>	49.0	59.1
MATH	Math	88.6	73.8	75.9	<b>89.7</b>	78.3	83.1	87.8
	Calculus	59.8	37.1	40.7	<b>64.7</b>	43.5	50.0	59.1
	MBPP-1	78.0	72.8	78.3	78.6	76.5	76.9	<b>79.3</b>
Code	MBPP-1	78.0	72.8	78.3	78.6	76.5	76.9	<b>79.3</b>
	Math	<b>70.2</b>	58.7	68.2	67.0	68.2	61.7	59.4
	TensorNetwork	<b>65.1</b>	53.6	61.8	62.6	60.3	56.9	63.3
Reasoning	IRB	91.6	89.2	91.7	<b>92.6</b>	<b>92.6</b>	90.3	92.3
	PAR	<b>93.0</b>	91.2	79.8	80.7	88.3	87.4	91.6
Instruction Following	IFEval	69.5	86.0	85.7	<b>89.8</b>	89.3	84.7	86.1
	SyBench	67.6	58.9	62.2	<b>69.0</b>	<b>69.0</b>	47.2	66.5
Chinese	CNHI-J	<b>90.9</b>	75.4	77.3	84.3	81.2	81.3	83.5
	C-Eval	<b>91.8</b>	72.7	76.0	85.9	90.0	84.7	86.5

资料来源：豆包官网，民生证券研究院

**图11：通义千问 QwQ-32B 与其他大模型性能对比**



资料来源：通义千问官网，民生证券研究院

## 1.3 基于 DeepSeek, 探讨推理需求大爆发

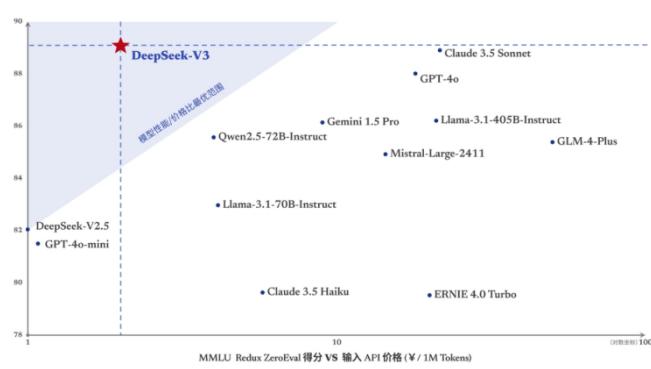
前文我们详细分析了 DeepSeek 在模型结构优化上的突破所带来的广泛影响。目前国产大模型呈现出百花齐放、快速迭代的发展态势，尤其是在模型轻量化方面取得显著进展。事实上，今年以来推出的多个代表性模型，如 Doubao-1.5-pro、通义千问 QwQ-32B 等，均在轻量精简上下足了功夫，力求在保证性能的基础上，大幅降低训练与推理成本，实现 AI 平权的愿景，但随之而来，市场也开始关注算力需求的持续性。

在此，我们认为，大模型轻量化是以更低门槛激发出更广泛的推理场景与需求。轻量化降低了 AI 模型的使用成本，加速了 AI 在消费级市场和垂直行业的渗透，催生出更丰富的 AI 原生应用，进而推动推理算力需求持续上升。下文中，我们将以 DeepSeek 为例，解读大模型轻量化趋势是如何拉动推理需求的。

### 1.3.1 DeepSeek 开创了低成本大模型的新范式

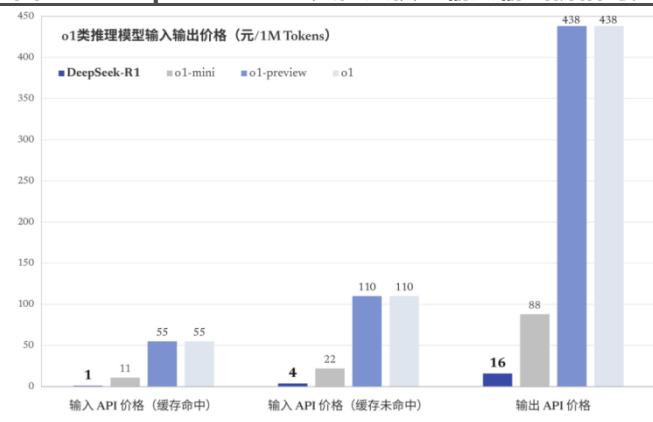
通过技术优化，DeepSeek 极大程度降低了大模型训练和使用成本。得益于技术路线优化，DeepSeek-V3 仅使用 2048 块 NVIDIA H800 GPU，累计训练 278.8 万个小时，耗费 557.6 万美元（假设单位训练成本为 2 美元/小时）。作为对比，GPT-4o 训练成本约为 1 亿美元，Llama 3.1-405B 训练时长达到 3080 万 GPU 小时，DeepSeek 超高的模型训练效率打破了大模型领域对于算力堆叠的过度依赖。凭借低训练成本的优势，DeepSeek 大模型的使用成本也大幅降低，DeepSeek-V3 的 API 服务价格较 GPT-4o 减少 90% 以上，DeepSeek-R1 的 API 服务价格较 o1 正式版减少 96% 以上。

图12：DeepSeek-V3 与其他大模型性能/价格对比



资料来源：DeepSeek 官网，民生证券研究院

图13：DeepSeek-R1 与其他大模型输入输出价格对比



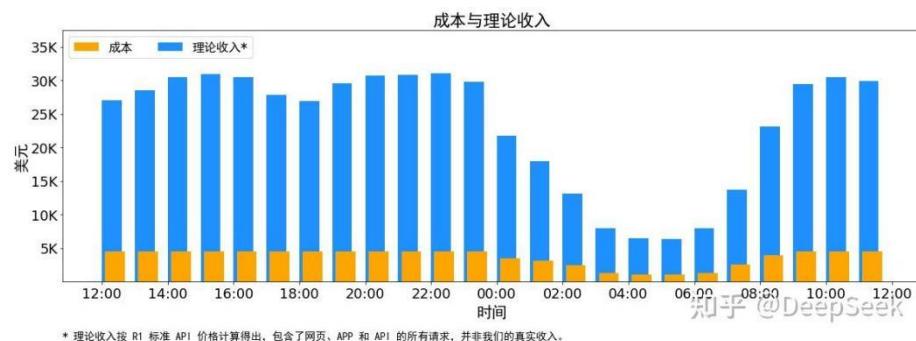
资料来源：DeepSeek 官网，民生证券研究院

得益于超低的成本，DeepSeek 理论成本利润率(收入/成本-1)可高达 545%。

DeepSeek-V3 和 DeepSeek-R1 的线上系统服务器均采用 NVIDIA H800 GPU，并在白天将所有节点用于推理，晚上用于训练，日均成本为 8.7 万美元（假设单位租赁成本为 2 美元/小时），对应理论日均收入为 56.2 万美元（假设所有 token 按照 DeepSeek-R1 API 定价），理论成本利润率高达 545%。实际上，由于

DeepSeek-V3 API 价格远低于 DeepSeek-R1、夜间时段 API 折扣、API 付费率远低于 100% 等原因，实际成本利润率将低于理论水平，但成本效益依旧可观。

**图14：DeepSeek 大模型理论收入与成本**



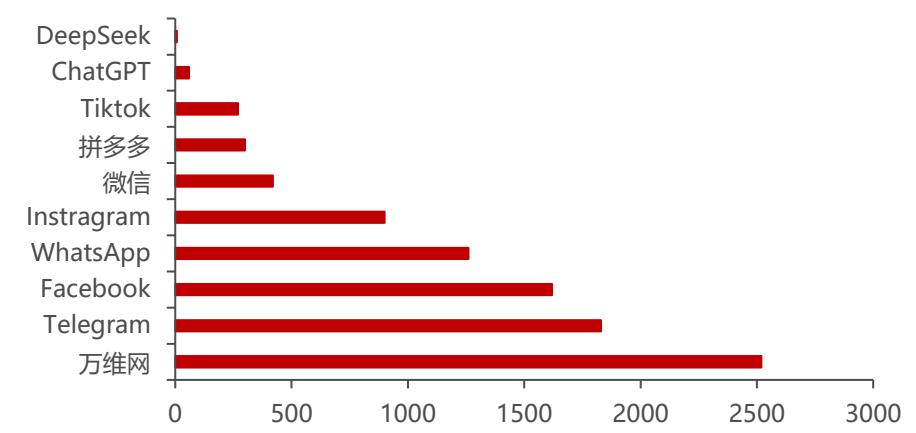
资料来源：DeepSeek 知乎认证账号，民生证券研究院

### 1.3.1 低成本的轻量化大模型将极大带动推理需求增长

**DeepSeek-R1** 作为开源推理大模型，将带动大模型推理能力提升，推动推理需求增长。DeepSeek-R1 在执行推理任务时与人类思考相似，并且完全展示其思考过程，DeepSeek 选择同步开源模型权重，向行业展示大模型在处理推理任务上的潜力。大模型推理能力成为未来 AI 落地的关键因素，DeepSeek 的开源举措将加速大模型推理能力进步，有望拓展全新应用场景，促进推理需求增加。

**DeepSeek** 推出后实现了用户的高速增长，彰显出大众对低成本轻量化大模型的强烈需求。在无广告投入的前提下，仅用短短 7 天便实现用户量突破 1 亿的里程碑式增长，充分展现了 DeepSeek 大模型产品的吸引力与市场渗透能力。

**图15：各产品增长 1 亿用户所用时间 (天)**



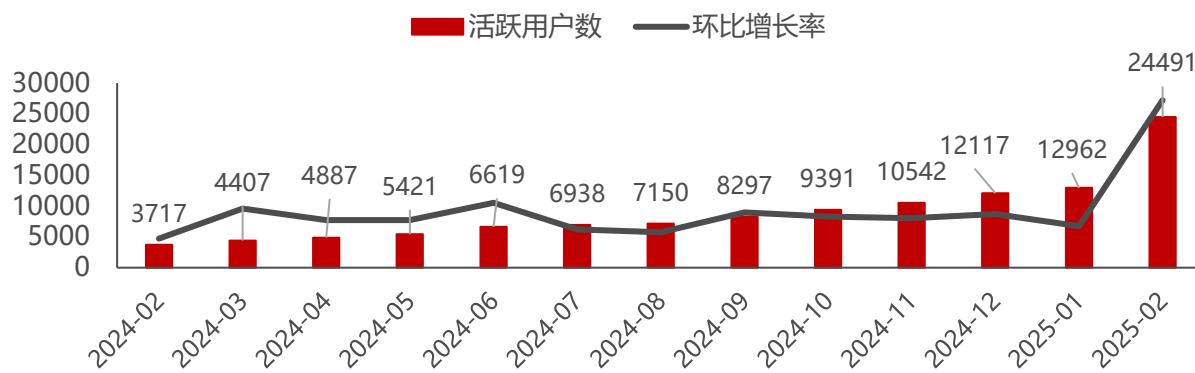
资料来源：AI 产品榜，民生证券研究院

注：DeepSeek 包含网站 Web/应用 App 累加不去重复，Tiktok 不包含国内版抖音

**DeepSeek-R1** 模型在 2025 年 1 月份发布后迅速引爆 AI 大模型需求。在 DeepSeek-R1 发布后的一个月内，国内原生 App 行业规模几近翻倍，截至 2025

年2月底，AI原生APP用户规模达到2.4亿，相比1月份增长88.9%。这一增长速度体现出DeepSeek-R1在大模型市场上的爆发效应，进一步印证了轻量化、高性能的大模型具备强大的用户吸引力，推动了大模型的需求扩张。

图16：AI原生App月活跃用户规模趋势（万人）



资料来源：QuestMobile，民生证券研究院

## 2 向“芯”而行，国产算力破局元年

**在国产大模型密集落地背景下，芯片厂商加速适配国产算力生态。**中芯国际 N+1 工艺已逐步成熟，N+2 持续推进，构建国产算力底座；昇腾 910C 量产落地，920 系列研发加快，性能持续逼近国际主流水平；寒武纪、海光等在 AI 训推方向深度布局，硬件端多点突破，生态融合加快。云端 ASIC 正成为算力演进主流，谷歌、亚马逊持续加码自研芯片体系；国内翱捷科技、芯原等设计企业快速成长，覆盖多元应用场景，并与海内外头部厂商形成紧密合作，成长弹性充足。在软件层面，适配节奏同样加快。以 DeepSeek 为例，发布即获得 17 家芯片厂商支持适配，训推效率大幅提升，助力算力生态向自主可控稳步迈进。

### 2.1 先进制程：中芯中流砥柱，筑基国产算力

**中芯国际是全球领先的集成电路晶圆代工企业之一，也是中国大陆技术最先进、规模最大、配套最完善的专业晶圆代工厂。**随着先进制程稀缺性和其在国家战略中的重要性日益凸显，公司在半导体行业周期性波动背景下，仍保持坚定的扩产节奏。伴随核心设备陆续到位，国产设备的配套能力也在不断提升。中芯国际作为国内晶圆制造龙头，近年来在算力芯片制程适配和先进工艺升级方面持续加码，正加快构建与国产算力需求相匹配的工艺体系，加速推动 AI 芯片国产化进程。

**在成熟制程方面，公司已在 28nm 实现技术突破，对原 40nm 客户需求进行了平台级的承接与优化。**在起步相对较晚的背景下，公司以海外主流厂商进行对标，在工艺平台完整性、品质控制与 IP 生态方面进行针对性迭代，逐步在 28nm/40nm 工艺节点上形成以模型计算与射频通信为导向的多平台演化，推动工艺走向市场化、多样化需求覆盖。在应用端，公司从消费电子切入，逐步扩展至工业级，再到车规级制程要求，工艺精度及可靠性水平持续提升。

**在先进制程方面，中芯国际围绕 FinFET 路线持续演进，已逐步形成较为完整的代工能力体系。**2019 年公司率先实现 14nm FinFET 的量产，并主要应用于 AI、与高性能低功耗计算等领域。为适配大模型时代的高算力场景需求，公司进一步推出 N+1 工艺。作为中芯国际下一代 FinFET 先进制程节点，N+1 工艺制程于 2020 年完成测试及流片，第一次采用四重图形处理 (SAQP) 形成 Fin 架构，并由自对准双重成像技术 (SADP) 形成 Dummy Gate，以达到更小 pitch 的制程需求。工艺功耗较上一代 14nm 下降 57%，性能提升 20%，该工艺在性能、逻辑密度上具有显著进步。而公司最新的 N+2 制程也在加速突破中，制程有望等同于 7nm 工艺。尽管中芯国际在技术水平上仍与全球最大晶圆代工厂台积电存在一定差距，但通过对成熟制程的持续优化与重构，中芯正在加速缩小差距，逐步推进先进工艺的国产替代。

**从财务端来看，中芯国际在 AI 浪潮带动下表现亮眼。**2024 年公司实现营业收入 571.01 亿元，同比增长 +28.1%；归母净利润为 36.99 亿元，同比下降 -23.3%。

公司 2024 年晶圆代工业务实现收入 532.46 亿元，同比增长+30.3%，维持稳健增长态势。全年产能利用率为 85.6%，销售晶圆总量折合 8 英寸为 802.1 万片，24 年年末月产能折合 8 英寸为 94.8 万片。随着地缘政治博弈加剧、AI 芯片相关禁令持续推进，以及国产替代化浪潮不断深化，公司给出 2025Q1 指引，销售收入预计环比增长 6%~8%，毛利率预计在 19%~21% 之间，在外部与内需双重带动下，中芯国际营收表现有望保持稳健增长。

## 2.2 华为昇腾：384 架构引领，910C 蓄势待发

**通过自研的 AI 推理引擎，昇腾 910B 与高端 GPU 之间的性能鸿沟有望逐步弥补。**潞晨通过自研的国产推理引擎，成功实现了昇腾 910B 与 DeepSeekR1 系列模型的推理适配优化。得益于 910B 的 CANN 软件栈提供的软硬协同优化，以及对 PyTorch 等主流框架的全场景兼容，昇腾 910B 在性能表现上与高端 GPU 相当。比如在采用梯度积累与现存复用技术后，昇腾 910B 可承载的模型数量提升至 1.7 倍。同时，在相同算力条件下，DeepSeek-R1 系列模型在昇腾 910B 上的推理延迟较 H100GPU 仅相差 5%，吞吐量差距缩小至 8% 内。这证明了在硬件架构和工艺技术仍存差距的情况下，国产 AI 推理芯片通过 DeepSeek 兼容适配，具备平替高端 GPU 的潜力，并有望缩小与国际高端竞品的硬件差距。

**表1：昇腾 910B 与英伟达 B200 的性能差异**

	昇腾 910B	英伟达 B200
产品方向	加速机器学习等任务专为数据中心设计的高性能 ai 处理器	为需要最高级别计算性能和内存带宽的应用提供支持，如科研等
框架	DaVinci 架构，面向人工智能应用的高效计算架构。	Blackwell 架构，专为 AI 计算架构
晶体管数量	496 亿	2080 亿
制程工艺	7 纳米制程工艺	4 纳米制程工艺
算力	256Tera-FLOPS(FP16);512Tera-OPS(INT8)	40 Peta-FLOPS(FP4);4.5/9Peta-OPS(INT8);
功耗	310W	1000W
内存支持	64GBHBM 内存,400GB/s 内存带宽	192GB 内存，达到了 8TB/s 内存带宽
灵活性	支持多种 AI 框架的灵活扩展如 MindSpore	具备出色的灵活性，适用于多种高性能计算和 AI 应用
数据中心应用	支持高效应用，包括高吞吐量、低延迟和高可靠性场景	应用于数据中心，支持云计算、大数据处理等场景
智能网卡特性	高性能、低延迟的智能网卡特性，用于内部的数据传输	低延迟，适用于数据中心内部的数据传输

资料来源：IT 技术订阅，芯智讯，民生证券研究院

B200 依托 Blackwell 架构，AI 算力达 20Peta FLOPS(FP4)，内存带宽 8TB/s，并支持 NVLink，适用于大规模分布式计算，主要面向高端算力需求场景。昇腾 910B 在性能上暂时难以与 B200 直接竞争。但随着 DeepSeek 模型优化降低算力

需求，昇腾系列在本地化生态和成本控制上的优势放大，有望在 AI 推理市场打破国际 GPU 市场的垄断，为国产 AI 产业提供更具竞争力的选择。

**昇腾 910C 在 910B 性能短板上进行了优化。**昇腾 910C 发布实现了算力的显著提升 (640 TFLOPS/FP16)，接近 NVIDIA H100 (680 TFLOPS/FP16)。910C 的推出不仅强化了在云计算、AI 训练等高性能计算场景的应用能力，也表明其在 AI 芯片市场的竞争力进一步增强，加速缩小与英伟达在高端算力市场的差距。同时，昇腾 910C 通过与 DeepSeek 模型的深度优化，实现训练效率提升 40%，推理速度突破每秒 3000 次，进一步缩小与国际算力表现的差距。昇腾对于“芯片·模型·场景”全栈闭环的生态兼容性，也大幅降低了企业切换技术路线的成本，增强 AI 产业范围的需求性并构建自主可控的国产 AI 算力生态。

**同时，华为云也在超节点上实现突破。**4 月 10 日华为云宣布 CloudMartix 384 推出，产品使用 384 颗 910c 芯片堆叠制程，并在系统上采用 Scale Up 架构，通过单层网络实现 GPU 全互联，同时配备 6912 个 400G 光收发器，有效保障了带宽功能指标的达成。而在性能端，通过五倍芯片堆叠配置下，产品算力指标优势明显，BF16 性能达到 300 PFLOPS，约为 GB200 NVL72 的 1.7 倍；HBM 总容量达 49.2TB，是 GB200 的 3.6 倍；但在功耗上，CloudMatrix 384 总功率上达到 559.4kW，是 GB200 NVL72 的近四倍功耗。能耗设计上的取舍，换来了在大规模训练及推理场景下显著的性能释放。本次昇腾超节点的推出，标志着国产 AI 算力在实际应用场景中的渗透步伐进一步加快，国产算力迎来破局元年。

**920 研发加速有望进一步放大昇腾硬件优势。**作为昇腾第三代 AI 处理器，昇腾 920 采用中芯国际 6nm (N+3) 工艺制造，晶体管密度达到 120 亿个，较上一代 910 提升约 50%。芯片基于 DaVinci 架构，集成 64 个 Core，每个 Core 支持 32 个矩阵计算单元。在 BERT-large 模型训练场景下，千卡集群吞吐量提升可达 30%。产品在性能层面实现显著突破，峰值算力达到 1800TOPS (INT8) /900TFLOPS (BF16)，内存带宽达 4000GB/s。与英伟达 H20 对比，920C 在 BF16 算力上提升约 37%，内存带宽增加 33%，能效比 (TOPS/W) 优化约 42%。软件生态方面，华为推出 CUADtoCANN 转换器，实现约 92% 的 CUDA 代码自动迁移效率，显著降低开发者适配与迁徙成本。在实际应用场景中，如金融行业的实时风控系统，基于昇腾 920 构建的平台可支持每秒百万级交易请求处理，并通过动态稀疏计算优化，将模型更新延迟从小时级压缩至分钟级，进一步强化数据处理能力与系统响应效率。

**昇腾与 DeepSeek 的适配加速了昇腾算力在 AI 计算市场的渗透。**DeepSeek-V3 发布即原生支持昇腾硬件，推动大模型在本地算力上的高效部署。2025 年 2 月，70% 的国内客户基于昇腾部署 DeepSeek，覆盖金融、政务、教育等 30 余个行业。据 IDC 预测，2025 年中国 AI 算力市场国产化比例将从 2022 年的 28% 提升至 65%，其中昇腾生态预计占据 40% 份额。随着本地化 AI 计算需求增长，昇腾依托一体机方案在推理吞吐、并发能力、强化学习微调等方面优化算力利用效率，

降低整体部署成本，进一步巩固昇腾系列在 AI 产业链的核心地位。

**昇腾依托成熟的国产供应链体系，在提升硬件性能表现的同时，逐步缩小与国际龙头竞品的技术差距。**通过长期业务合作、战略投资及股权合作，昇腾在复杂供应链体系下依然能够维持稳定的出货节奏与产品供应管控。这一供应链韧性不仅得益于与核心供应商的深度协同，也依托上游环节的成熟技术支撑，为昇腾系列在 AI 算力市场的可持续发展提供长期保障。

**泰嘉股份：**公司主营电源业务，涵盖消费电源、数据电源和站点电源三大领域，其中数据电子业务重点布局服务器电源模块，服务于行业头部客户。自 2023 年投产以来，公司服务器电源模块业务产能逐步释放，并根据市场需求优化生产布局，以适应 AI 服务器、电源管理方案的增长需求。此外，公司以参股方式与罗定雅达形成战略合作，就消费电子领域的业务协同、技术研发、资源共享、资本运营等多个领域深入合作，而雅达电子也是华为在适配器及供电设备领域的重要供应商。

**华丰科技：**公司深耕光、电连接器及线缆组件供应链服务，产品覆盖防务、通讯、工业三大领域。其中，通讯板块作为核心业务，公司为华为、中兴等主流设备供应商，专注于电源连接器及背板连接器，后者广泛应用于高端服务器及 AI 计算基础设施。随着 AI 智能产业发展，公司作为上游核心部件供应商，受益于算力需求增长。华为在公司通讯类业务销售占比高达 70%，公司已成为其关键供应伙伴。

**杰华特：**公司凭借完善的产品矩阵与自主研发能力确立了虚拟 IDM 运营模式在模拟集成电路领域的竞争优势，并依托自研国际先进工艺平台，持续推动高端产品迭代。公司加速拓展服务器、AI 等高算力应用市场，已构建 DrMOS+多相供电方案的完整产品体系，强化在 AI 计算与数据中心的竞争力。同时，华为哈勃投资持股约 5%，为公司在高性能电源管理领域的协同发展提供支撑。

**兴森科技：**公司在 AI 领域的主营业务围绕 PCB 和半导体两大主线展开，并提供 PCB 印制电路板等核心产品。在 PCB 领域，公司通过宜兴硅谷基地的管理架构优化，集中资源于服务器等高端赛道，并在产线升级改造后，新增加速卡等 AI 服务器类载板的产能。同时，公司与华为在 PCB 和半导体业务领域保持长期稳定合作，是其重要合作伙伴，进一步巩固了公司在 AI 产业链中的市场地位。

**华海诚科：**公司专注于半导体封装材料的研发及产业化，主要产品涵盖环氧塑封料与电子胶黏剂，已发展成为国内规模领先、具备持续创新能力的环氧塑封料厂商。在半导体芯片材料领域，公司已构建完整的自主研发与生产体系，并拥有自主知识产权。目前，公司在 GMC 生产工艺方面取得突破，已建立全套量产工艺流程，相关设备通过客户考核并持续优化。值得关注的是，华为哈勃投资为公司上市前的原始股东，进一步强化其在产业链中的战略价值。

**飞龙股份：**公司自 1952 年成立以来，已深耕汽车水泵领域 73 年。随着数字时代发展，数据中心与算力需求同步增长，单机柜散热需求持续攀升，液冷技术凭借高效散热性能，在“双碳”目标推动下加速渗透。公司自 2019 年起布局 IDC 液

冷赛道，已成功开发μ（微型泵）、S（小型泵）、M（中型泵）、L（大型泵）四大产品平台，覆盖不同功率散热需求。其中，HP22K 作为 L 型泵旗舰型号，填补了国产高性能液冷泵市场空白，推动国内液冷技术实现突破。2024 年 11 月，公司与华为云正式启动深度合作，围绕数字化转型、技术应用等领域展开协同，进一步加速数字化赋能进程。

**鸿日达：**公司在精密连接器领域深耕多年，近年来亦积极布局半导体散热领域，华为为公司重要客户之一。目前，公司半导体金属散热片产品可广泛应用于算力 CPU 处理器、GPU 及 AI 算力等领域。现阶段，公司已建成一条半导体金属散热片生产线，并规划根据终端客户需求、产能稼动率及利用率，逐步扩产至年产 1090 万片的规模。产品方面，公司已完成对多家核心终端客户的送样并获得部分供应商代码，标志挂拍卖行是着在半导体散热领域迈出实质性步伐。

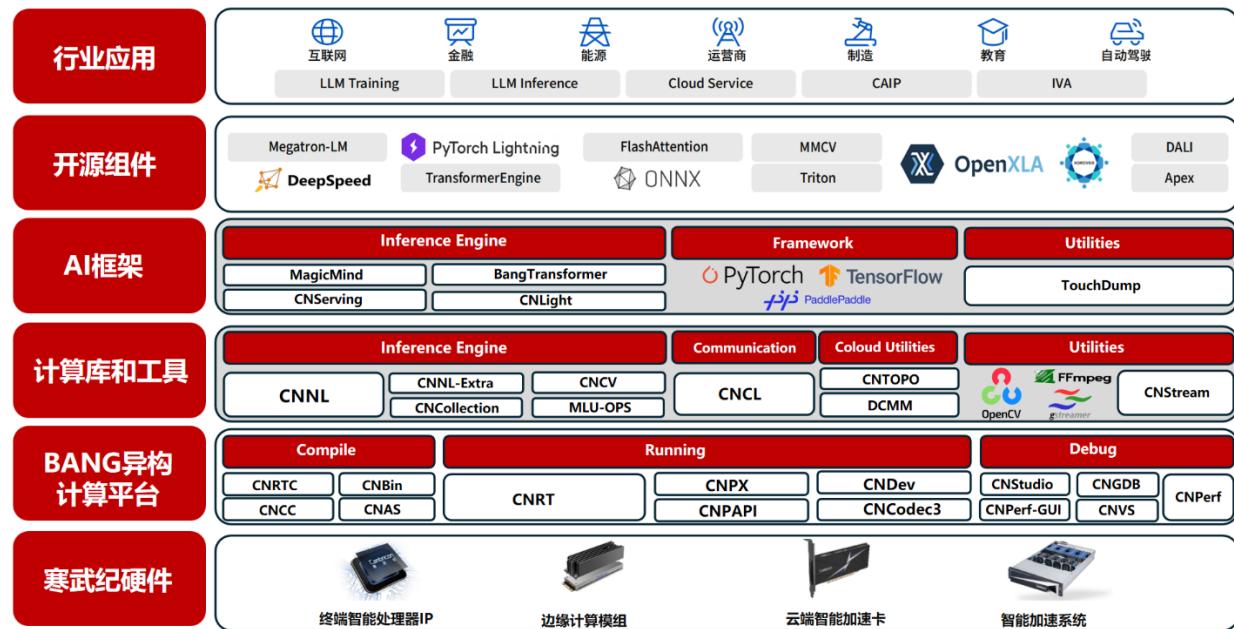
## 2.3 算力双龙：攻坚突围、加速落地

### 2.3.1 寒武纪：软硬协同发力，推进高性能 AI 芯片国产化

**寒武纪是智能芯片领域全球知名的新兴公司，持续聚焦 AI 芯片领域的技术突破与商业化落地。**公司具备从云到边、从硬件到系统软件的完整能力，已形成云边端一体化、软硬协同、训推融合、生态统一的平台化布局。自 2016 年 3 月成立以来，公司不断推进核心技术的产业化应用，旗下思元系列云端智能加速卡产品已广泛部署于多家服务器厂商场景中，思元 IP 产品也已集成至超过 1 亿台智能终端，产业落地能力持续验证。

**当前，公司研发的全新一代云端 AI 训练芯片为思元 590。**该产品采用 ASIC 架构设计，具备低功耗、高效率等特点，针对大模型训练与高复杂 AI 任务进行深度支持，单卡算力达 512 TOPS，在性能上深度对标 A100。在存储架构方面，思元 590 的 MTP 与 DDR/HBM 间的带宽大幅优化，支持 L2 Cache 读写缓存，MTP Cluster 的访存带宽相较上代提升达 4 倍，为大规模数据处理提供强力支持。思元 590 已部署在智能安防、自动驾驶等多元应用场景，有效满足多种边缘 AI 应用的需求。

**软件生态方面，寒武纪持续完善大模型训练平台。**公司围绕 Megatron、Transformer 等主流组件持续优化，全面适配 PyTorch 系列训练框架，并通过改善热点算子性能、支持通算融合等优化策略，针对 DeepSeek、Llama 等主流模型进行适配，进一步提升训练效率并已达到业内主流水平。在推理侧，公司已对主流文生文模型及多模态模型如 Flux 等实现深化支持，覆盖范围持续拓展，夯实了训推一体的平台底座。值得关注的是，搭载思元 590 的阿里云智能计算集群在 Llama-3 训练任务中已实现训练成本下降 40%，进一步验证其商业应用潜力。

**图17：寒武纪基础软件系统平台**


资料来源：寒武纪 2024 年社会责任报告，民生证券研究院

**在业绩层面，公司积极把握大模型齐开花背景下国产算力需求快速释放所带来的市场机遇。**2025 年 Q1 公司实现营收 11.11 亿元，同比增长 4230.22%，大幅提速。细分行业方面，公司在互联网及运营商市场中产品性能与稳定性获得广泛认可，在金融行业亦已实现从传统 AI 任务支持向大模型适配优化的升级，持续推动 AI 算力在银行、保险、基金等业务场景中的深度落地。2025 年 Q1 公司归母净利润为 3.55 亿元，实现扭亏为盈，盈利能力显著改善。2025 年 Q1 公司在存货端 27.55 亿元，较去年年末大增 9.81 亿，侧面验证下游订单动能延续，未来放量节奏具备良好支撑。

### 2.3.2 海光信息：架构开放兼容，DCU 加速国产 AI 模型落地

**海光信息是国内领先的高端计算芯片设计企业，围绕 CPU 和 DCU 产品进行双重布局。**公司成立于 2014 年，专注于高性能计算芯片和服务器的研发，面向 AI 产品端，其高端协处理器 (DCU) 为主力产品。海光 DCU 产品基于 GPGPU 架构，兼容类 CUDA 通用环境，主要应用于数据中心与云计算场景，旨在解决传统架构在大规模并发、低延迟和高带宽需求下的性能瓶颈。随着 AI 算力时代的到来，DCU 产品展现出广阔的发展前景。

**海光 DCU 产品以深算系列命名，当前运行于 ROCm 平台，兼容 CUDA 生态。**与国内竞品相比，海光 DCU 具备显著的算力优势，依托大规模并行计算架构，在双精度和整型计算方面表现优异。公司目前的主力产品为深算二号，相较于前代产品，在浮点数据精度和主流数据类型的计算能力方面大幅提升，整体性能提升超过 100%，并针对深度学习模型的训练与推理场景进行深入优化，显著提升 AI 算力效率。海光最新 DCU 产品深算三号正在积极研发中。

**表2：英伟达、AMD、海光同期 GPU 参数对比**

	英伟达	AMD	海光信息
产品型号	A100	MI100	深算一号
生产工艺	7nmFinFET	7nmFinFET	7nmFinFET
核心数量	2560CUDA processors 640Tensorprocessors	120CUs	4096 (64CUs)
内核频率	Upto1.53Ghz	Upto1.5Ghz (FP64) Upto1.7Ghz (FP32)	Upto1.5Ghz (FP64) Upto1.7Ghz (FP32)
显存容量	80GBHBM2e	32GBHBM2	32GBHBM2
显存位宽	5120bit	4096bit	4096bit
显存频率	3.2GHz	2.4GHz	2.0GHz
显存带宽	2039GB/s	1228GB/s	1024GB/s
TDP	400W	300W	350W
CPUtoGPU 互联	PCIeGen4x16	PCIeGen4x16	PCIeGen4x16
GPUtoCPU 互联	NVLink Upto600GB/s	InfinityFabricx3 Upto276GB/s	xGMIx2 Upto184GB/s

资料来源：海光信息招股说明书，民生证券研究院

**在模型适配方面，海光也在积极推进。**海光 DCU 已构建自主开发的完整软件栈，包括 DTK、开发工具链等，支持 TensorFlow、PyTorch 等主流深度学习框架。在开放生态基础上，公司搭建了完善的层次化软件栈和统一的底层硬件驱动平台，能够适配多种 API 接口和编译器，并支持常见 AI 算法与框架。在 AI 快速发展的背景下，海光 DCU 支持全精度模型训练，已实现 LLaMA、GPT、子东太初等大模型的全面适配，并与 Qwen、DeepSeek 等国产大模型完成全面对接，持续提升产品渗透率。以 DeepSeek 为例，海光于 2025 年 2 月完成适配结合，其 MTP 优化机制，有效提升了文本数据处理效率，进一步增强芯片在大模型场景下的算力表现和性能稳定性。

**图18：海光完整软件栈系统**


资料来源：海光信息官网，民生证券研究院

**在最新公布的 2025 年第一季度财报中，公司实现营业收入 24.00 亿元，同比增长 50.76%；归母净利润为 5.06 亿元，同比增长 75.33%。**受益于地缘政治推动下的国产化进程及公司产品在国产大模型中的快速适配，整体需求加速拉动营收高增。同时，公司一季度研发投入达 7.64 亿元，同比增长 16.26%，体现出公司在技术迭代和性能突破上的持续投入，通过构筑产品护城河夯实长期竞争力。

在未来地缘环境仍不稳定、国产大模型全面开花的背景下，下游客户需求有望持续释放，推动海光信息维持稳健增长态势。公司 25Q1 合同负债 32.37 亿元，较去年末大增 23.33 亿元，订单需求持续旺盛，后续业绩释放具备较强增长动能。

### 2.3.3 云天励飞：算法芯片化加速推进，边缘 AI 应用深化突破

云天励飞成立于 2014 年 8 月，长期深耕算法与芯片双平台架构，围绕算法芯片化和端云协同的技术路径，已在数字城市等核心应用场景中实现落地突破。公司顺应大模型时代发展契机，自 2022 年 ChatGPT 发布以来，持续加码大模型训练及推理方向，在软硬协同能力、产品落地广度及算力平台演进方面取得积极进展。

在硬件端，公司自 2016 年推出首款自研 AI 芯片 DeepEye 100 后，逐步实现产品迭代，陆续推出 DeepEdge 10 系列 SoC 芯片，并不断提升芯片性能与适配能力。最新产品 DeepEdge 10 Max 基于公司自研的 NNP400T 神经网络处理器打造，相较前代产品在浮点与整数计算性能上显著提升，适配主流深度学习网络结构。公司现阶段研发的可重定义 AI 芯片，面向嵌入式前端及边缘推理需求，支持灵活适配多种算法框架并兼顾性能与成本效率。同时公司新一代产品采用自主可控的国产工艺平台，提出算力积木式架构，覆盖 Transformer、BEV 等主流模型，面向多样化算力场景，满足不同行业在边缘部署中的高效推理需求。

表3：云天励飞 DeepEdge 10 系列性能参数

性能	DeepEdge 10	DeepEdge 10 C	DeepEdge 10 Max
<b>MCU 核</b>	单核 MCU 处理器, 450 DMIPS	单核 MCU 处理器, 450 DMIPS	4 个单核 MCU 处理器, 1.8K DMIPS;
<b>异构计算平台</b>	NNP400T 8 TOPS (INT16), 2 TFLOPS (FP16) 的	NNP400T 4 TOPS (INT16)/ 1.5 TFLOPS (FP16)	NNP400T; 32 TOPS (INT16), 8 TFLOPS (FP16)
<b>图形处理器</b>	OpenGL ES 3.1/3.0/2.0/1.1	支持 3D 图形处理单元	OpenGL ES 3.1/3.0/2.0/1.1
<b>存储器接口</b>	LPDDR4@3733Mbps/ LPDDR4X@3733Mbps/ DDR4@3200Mbps	LPDDR4@3200Mbps / LPDDR4X@3200Mbps	LPDDR4@3733Mbps / LPDDR4X@3733Mbps;
<b>接口</b>	RGMII、RMII	PCIe 3.0	PCIe 3.0
<b>功耗</b>	10W	-	40W
<b>封装</b>	FCBGA 封装: 23nm * 23nm	FCCSP 封装: 15mm x 15mm	FCBGA 封装: 40mm * 40mm/0.8mm 间隙

资料来源：云天励飞官网，民生证券研究院

注：未标注的数据为没有在公开渠道披露的信息

在模型端，公司推出自研大模型体系“云天天书”，具备行业领先的图文理解与问答能力，形成通用、行业、场景三类大模型分层体系。模型可针对政务、企业办公等典型场景实现快速定制化部署。同时公司也与华为联合打造云天天书大模型训推一体机，基于昇腾 AI 平台构建边缘智算中心，推动训练与推理一体化部署。该一体机近期已实现与 DeepSeek R1 全尺寸模型（6710 亿参数）FP16 精度的本地部署，并完成面向客户的场景化集成测试，有效支撑下游应用的智能升级。

**图19：云飞励天下游场景解决方案**

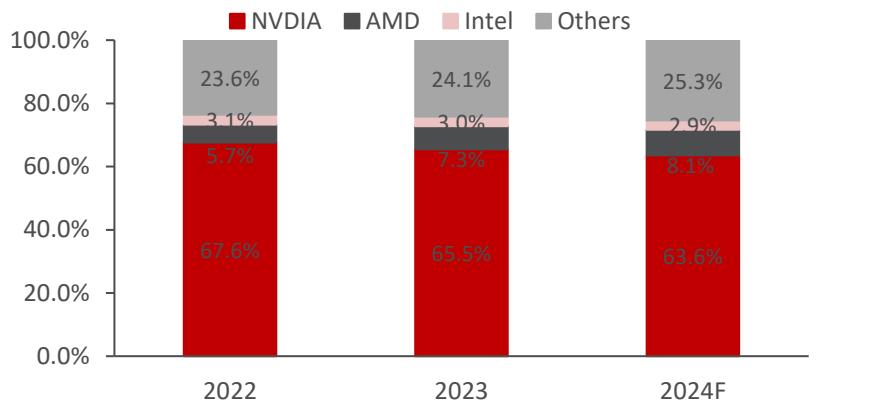

资料来源：云飞励天 2023 年年报，民生证券研究院

**财务端来看公司有望释放进一步的业绩潜力。**公司 2025 年 Q1 实现营业收入 2.64 亿元，同比+168.23%；归母净利润为-0.86 亿元，仍处于亏损状态，但亏损幅度同环比均有显著改善。随着大模型场景化需求持续释放，公司在训练与推理一体化的产品链条将逐步完善。

## 2.4 ASIC：云厂商的曙光，挑战 GPU 垄断

**目前英伟达在 AI 加速卡领域仍然占据主要供应商的地位，但云商自研 ASIC 的比例正在逐步提升。**AI 加速卡的竞争格局中主要有三类“玩家”：1) 全球龙头英伟达；2) 其他商业 AI 加速卡公司，如 AMD、Intel、昇腾等；3) 云厂商自研 ASIC。据 Trendforce 统计，2022 年全球 AI 芯片市占率来看，英伟达独占 67.6%，2024 年英伟达市占率将有所下降，而 AMD 以及其他云商自研 ASIC 比例有望提升。

**我们认为，云厂商自研 ASIC 将成为未来 AI 芯片增量最核心的来源。**一方面，云商自研加速卡在成本方面显著优于向英伟达等商业公司外采，4Q24 英伟达毛利率已达到 73.0%，采用自研加速卡的方式，将帮助云商在有限的资本开支下获得更多的 AI 算力。另一方面，云商自研 ASIC 更加灵活，云厂商可以根据自身的模型训练和推理需求，进行 AI 芯片和服务器架构的设计，从而实现更好的训练和推理效果。伴随着云厂商自研 ASIC 产品的逐步成熟，未来云商在 AI 算力的布局中自研的比例有望逐步提升。

**图20：2022-2024 年全球 AI 芯片竞争格局**


资料来源：Trendforce，民生证券研究院

**谷歌采取自研加速卡为主，同时采购部分英伟达加速卡的策略。**谷歌研发 TPU 的时间始于 2013 年，相较于其他云厂商有近 10 年的时间优势。由于谷歌在加速卡领域布局早，产品完善度高，谷歌或为 2024 年北美四大云厂商中采购英伟达加速卡最少的厂商。2023 年 12 月，谷歌推出面向云端的 AI 加速卡 TPU v5p，相较于 TPU V4，TPU v5p 提供了 1.7 倍的浮点运算能力和 3 倍内存带宽提升；集群方面，TPU v5p pod 由 8960 颗芯片组成，使用最高带宽的芯片间连接（每芯片 4,800 Gbps）进行互连；从训练效果来看，相较于上一代产品，TPU v5p 训练大型 LLM 的速度提升了 2.8 倍。

**Meta 自 2021 年以来便将企业发展的重点放在元宇宙和 AI 领域，并且修改了公司名称。**2023 年，Meta 宣布自研 MTIA v1 芯片。2024 年 4 月，Meta 发布最新版本 MTIA v2 加速卡，新一代 MTIA 加速卡在算力、内存容量、内存带宽等方面更加平衡，采用台积电 5nm 工艺制造，Int 8 稀疏算力可以达到 708TOPS，LPDDR5 内存容量达到 128GB。目前 Meta 仍主要采购英伟达等厂商的加速卡用于 Llama 等模型的训练。

**微软 Azure 服务的世界 500 强企业数量超过 300 家，是目前采购英伟达加速卡最为激进的云厂商，但考虑到采购英伟达加速卡的高昂成本，微软也宣布了自研加速卡的计划。**微软的 Maia 100 在 2023 年推出，专为 Azure 云服务设计。Maia 100 采用台积电 5nm 工艺，单芯片拥有 1050 亿个晶体管，FP 8 算力可以达到 1600TFLOPS，同时支持 FP 4 运算，算力达到 3200TFLOPS。除了微软自身以外，OpenAI 也在尝试采用微软的加速卡，强大的下游客户支持有望为微软自研加速卡的进步带来重要动力。

**亚马逊同样在自研加速卡方面加大投入，并且已经完善了训练和推理的两方面布局。**2023 年，亚马逊推出了用于训练的 Trainium2 加速卡，以及用于推理的 Graviton4 加速卡，补全了亚马逊在训练和推理领域加速卡的布局。亚马逊的 Trainium2 加速卡 Int 8 算力达到 861TOPS，相较上一代产品性能提升 1.3 倍，在云厂商自研加速卡中表现优秀。同时公司的产品可以在新一代 EC2

UltraClusters 中扩展多达 10 万颗 Trainium2 加速卡，与 Amazon EFA 网络互联，提供高达 65 EFlops 算力，为云端大模型的训练和推理提供强大的算力保障。

表4：云商自研 ASIC 的布局情况

厂商	大类	型号	发布 时间	制程 nm	峰值算力 TOPS/TFLOPS			内存信息 类型	容量 GB	互联带宽 GB/s
					INT8/FP8 Dense/Sparse	BF16/FP16 Dense/Sparse	TF32/FP32 Dense/Sparse			
谷歌	训练	TPUv5E	2023	-	394	197	-	HBM2	16	400
		TPUv5P	2023	-	918	459	-	HBM2	95	800
Meta	推理	MTIA v2	2024	5	354/708	177/354	2.76	LPDDR5	128	-
微软	训练	Maia 100	2023	5	1600	800	-	HBM3	64	1200
亚马逊	训练	Trainium2	2023	4	861	431	215	-	96	-
	推理	Graviton4	2023	-	-	-	-	DDR5	-	-
TESLA	训练	D1	2021	7	362	362	22.6	-	32	-

资料来源：各公司官网，Semianalysis 等，民生证券研究院

注：未标注的数据为没有在公开渠道披露的信息

**芯原股份是国内领先的半导体 IP 供应商，拥有包括 GPU IP、NPU IP 在内的多款处理器 IP，具备为客户提供一站式芯片定制服务的能力。**公司芯片设计能力覆盖 14nm/10nm/7nm/5nm FinFET 和 28nm/22nm FD-SOI，能够为提供从硬件到软件的系统解决方案。2023 年 3 月，蓝洋智能发布与芯原股份合作打造的基于 Chiplet 架构的高性能 AI 芯片，其中内置芯原 GPGPU IP 的芯片算力可达 8 TFLOPS，内置芯原 NPU IP 的芯片算力可达 240 TOPS。目前，内置芯原 GPU IP 的芯片在全球范围内出货近 20 亿颗，芯原 NPU IP 已被 72 家客户用于其 128 款人工智能芯片中，在全球范围内出货超过 1 亿颗。公司正积极研发面向数据中心和 GPU AI 计算的高性能图形处理器技术，将支持 16-32 TFLOPs FP32 算力、128-1536 Texel/cycle 纹理处理能力、32-384 Pixel/cycle 像素填充能力。

**翱捷科技是国内少数同时完成“5G+AI”技术突破的企业之一，专注于无线通信芯片领域，具备提供超大规模高速 SoC 芯片定制服务能力。**在无线通信领域，公司掌握 2G/3G/4G/5G 全制式蜂窝基带芯片及多协议非蜂窝物联网芯片设计能力，相关客户涵盖国家大型电网、中兴通讯、美的集团、Hitachi、360、TP-Link 等知名企业。同时，基于公司成熟的超大规模芯片设计能力，公司还为人工智能、互联网、智能手机等不同行业内的头部企业提供芯片定制服务。目前，公司累计芯片出货量超过 3 亿颗，累计量产全新芯片超过 30 颗。

**芯原股份和翱捷科技作为国内顶尖 ASIC 设计公司，在各方面都体现了强劲的 ASIC 研发实力：**1) 在研发实力上，截至 2024 年，芯原股份研发人员规模达到 1800 人，团队成员来自 Broadcom、Marvell 等，翱捷科技研发人员规模为 1138 人，团队成员来自 Marvell、智擎信息等；2) 在 IP 能力上，芯原股份已全面布局 GPU、NPU、VPU 等芯片 IP，服务于多家全球知名企业，翱捷科技则更加专注无线通信芯片领域，2024 年，芯原股份 IP 授权收入为 7.36 亿元，翱捷科技则为 0.35 亿元；3) 在 ASIC 能力上，芯原股份具备从先进 5nm FinFET、22nm FD-

SOI 到传统 250nm CMOS 制程的设计能力，服务于英特尔、亚马逊、谷歌、微软等知名企业和企业，翱捷科技则拥有超大规模芯片量产经验，单颗芯片晶体管数量达到 177 亿，2024 年，芯原股份芯片定制收入为 15.81 亿元，翱捷科技为 3.36 亿元。

**表5：芯原股份和翱捷科技 ASIC 研发实力一览**

	芯原股份		翱捷科技	
	2023年	2024年	2023年	2024年
IP授权收入（亿元）	7.65	7.36	1.23	0.35
ASIC/芯片定制收入（亿元）	15.64	15.81	2.26	3.36
研发人员规模（人）	1662	1800	1135	1138
研发团队背景	Broadcom、Marvell等		Marvell、智擎信息等	
IP能力	拥有GPU、NPU、VPU等全面的芯片IP布局，AI储备第一梯队，服务于多家全球知名企业		专注无线通信芯片领域，ISP、高速通信接口IP、射频IP等能力领先	
ASIC能力	多工艺节点芯片设计能力，多家客户量产经验。为新能源车企提供5nm的自动驾驶芯片；为蓝洋智能设计基于 Chiplet的AI芯片		超大规模芯片量产经验，单颗芯片晶体管数量高达177亿	

资料来源：各公司 2023 年年报、各公司 2024 年年报，民生证券研究院

**中兴微电子是中兴通讯控股的子公司，拥有复杂 SoC 芯片前后端全流程设计能力。**公司在芯片架构实现、设计自动化、接口标准化，设计周期达到业界先进水平，并在低功耗技术、先进工艺和封装测试技术上具有平台优势。截止 2024 年底，公司拥有专利超过 4800 项，由公司自主研发并成功商用的芯片超过 120 种，每年芯片发货量达到 2 亿颗。

## 2.5 算力协力模型，实现生态国产化

**在国产大模型加速涌现、百花齐放的背景下，国产芯片厂商纷纷加快适配进程。**当前，主流芯片企业正加速与国产大模型的适配联动，持续推动国内算力体系的生态完善。在夯实 AI 产业链自主可控的基础上，国产大模型逐步弥合国内硬件参数与生态适配之间的差距，构建起从模型到底层硬件的高效迁移桥梁。同时，大模型的快速落地进一步加速了 AI 技术在各类产业与场景中的广泛渗透，拉动了国产高性能算力的结构性需求。

**以 DeepSeek 为例，其适配工作的推进为国产芯片提供了实际应用场景落地与测试反馈，加速了国产算力平台的成熟与技术优化。**根据与非网报道，近期，华为昇腾、海光、沐曦、天数智芯、摩尔线程、壁仞、燧原、昆仑芯、云天励飞、灵汐科技、鲲云等多家国产芯片厂商纷纷宣布完成对 DeepSeek 系列模型的适配，涵盖从 1.5B 到 70B 的多参数版本，实现了推理服务的高效部署。

**从技术层面来看，DeepSeek 的蒸馏模型和高效算法优化，使得国产芯片能够在推理任务中实现与高端 GPU 相当的性能表现。**例如，海光信息、沐曦等国产 GPU 已成功完成 DeepSeek-V3 和 R1 模型的适配，并在实际应用中展现了优异

的推理效率。这种技术突破不仅降低了算力成本，也为国产芯片在边缘计算、智能终端等场景的普及奠定了基础。

**从生态层面来看，DeepSeek 的开源属性和低成本特性，吸引了大量开发者和企业加入国产 AI 生态。**优刻得、华为云、腾讯云等云计算厂商纷纷支持 DeepSeek 模型的部署，形成了从底层硬件到上层应用的完整技术栈。这种“国产算力+国产模型”的闭环生态，不仅增强了国产 AI 产业链的自主可控能力，也提供了更具性价比的解决方案。

**国产芯片适配 DeepSeek 多点开花。**随着 DeepSeek 平台上线，各大芯片厂商对其较低算力需求下的高效能表现反响积极，加速适配。自昇腾 2 月 2 日在华为云与硅基流动推出 DeepSeekR1/V3 推理服务以来，国内厂商如沐曦、海光信息等陆续优化 DeepSeek 相关蒸馏模型，推动本土芯片生态多元化发展。

**表6：DeepSeek 发布，各平台适配情况**

芯片厂商	日期	主要事件
华为昇腾	2.1-2.6	2 月 1 日，华为云与硅基流动推出基于昇腾云的 DeepSeekR1/V3 推理服务；2 月 4 日，DeepSeek 模型上线昇腾社区支持一键部署；潞晨科技与华为昇腾联合发布基于昇腾算力的 DeepSeekR1 系列推理 API 及云镜像服务；2 月 6 日，百信信息通过自主研发的恒山 326TA 服务器实现昇腾算力与 DeepSeek 模型的推理适配优化。
沐曦	2.2-2.5	2 月 2 日，开源平台 giteeai 上线 1.5 等规模的 DeepseekR1，并部署在沐曦曦云 gpu；2 月 5 日，确认 DeepSeekv3(671b)可在沐曦训推一体 gpu 上寻用，并且 DeepSeek 为基础由联想与沐曦联合发布国产一体机方案。
天数智芯	2.4	完成对 DeepSeekr1 的适配，并上线 1.5b 等模型；此前已完成 DeepSeekV3671B 模型的适配精度对标论文标准。
摩尔线程	2.4	宣布实现基于夸娥 gpu 智算集群上对于 DeepSeekR1 蒸馏模型的高效部署，也开放基于 MTT80 等显卡的推理部署。
海光信息	2.4	完成 DeepSeekV3 和 R1 模型与海关 DCU 的适配上线。
壁仞科技	2.5	宣布上线自家 ai 算力平台对于 DeepSeekR1 蒸馏模型推理适配，覆盖多场景应用并全面支持 1.5B 至 70B 全系列模型
太初元碁	2.5	宣布再 T100 加速卡上完成对于 DeepSeekR1 系列模型的适配，并上线多款大模型服务
云天励飞	2.5	DeepEdge10 芯片平台完成对于 DeepSeek-R11.5B 至 8B 等模型适配，DeepSeekV3/R1671B 等模型正在适配中
燧原科技	2.6	完成 DeepSeek-R1/V3671B 原生模型等全量模型的高效适配
昆仑芯	2.6	宣布支持 DeepSeek 模型，帮助其在智能计算机等领域应用
灵汐科技	2.6	联合脑技术社区完成基于 KA200 类脑芯片对于 DeepSeek-R1 系列模型适配
鲲云科技	2.6	CAISA430 芯片成功适配 DeepSeekR1 蒸馏模型推理并达成动态优化性能，支持 Qwen 和 Llama 模型
希姆计算	2.6	完成 RISC-V 推理加速卡与 DeepSeekR1 全系列蒸馏模型的适配并落地多个智算中心
算能	2.7	自家 BM1684 处理器完成 DeepSeekR17B/1.5B 蒸馏模型适配

清微智能 2.7 RPU 芯片完成 DeepSeekR1 系列模型的适配和部署，并支持推一机和无交换机自组网调度

芯动力 2.7 AzureBladeK340LM.2 加速卡完成 DeepSeekR1 大模型的适配并支持在 AIPC 等设备上运行

龙芯中科 2.7 搭载龙芯 3 号 CPU 的设备成功运行 DeepSeekR17B 模型从而实现本地化部署

资料来源：泛半导体之家，民生证券研究院

**在各大国产主流芯片厂商加速适配的背景下，国产算力市场的产业活力持续凸显。**AI 技术在各类应用场景中的落地进程正显著拉动上游芯片厂商的供给响应，推动硬件产品线不断升级。在 AI 与生活日益融合的趋势下，算力硬件的部署需求与数据处理能力的要求愈发突出，持续带动上游算力基础设施的投入与扩展。关注未来，AI 软硬件的协同演进与系统化落地将成为国产算力生态构建的关键支撑。  
**在全球技术体系重构及供应链安全博弈加剧的背景下，推动全链条的自主可控与国产化替代，将成为中国智能计算产业高质量发展的核心路径。**

### 3 算力基建加码，解决供给短板

#### 3.1 云厂商加大算力储备及模型投入，资本开支高增

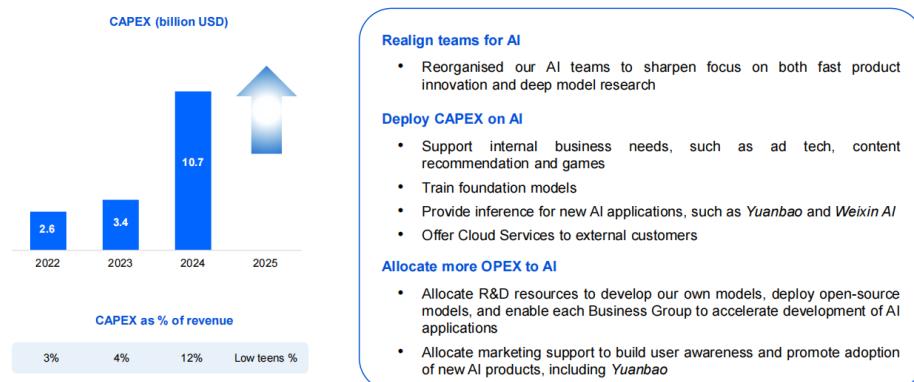
随着各大国产模型发布，AI 领域的用户规模持续扩大，对大模型训练及推理的 Token 计算需求持续攀升。截至 2024 年底，全国算力总规模已突破 230 EFLOPS，但供需缺口仍达 35%。为应对市场需求增长，国内云计算厂商正加大算力储备及模型优化投入，AI 计算基础设施建设布局逐步清晰，相关资本开支进入新一轮扩张周期。

以腾讯为例，腾讯在 24 年资本支出达到 768 亿元，较 23 年增长 221%，创下历史新高，占全年总收入 6603 亿元的约 12%。而 24 年更多资本投入主要由于购买更多的 GPU 以满足推理需求，仅 2024 年全年腾讯向英伟达订购 AI 芯片约为 23 万，名列全球第二。此外，腾讯云持续推进全球基础设施布局，国际业务连续三年保持两位数增长，截至 2024 年底，基础设施已覆盖全球五大洲 21 个国家和地区，运营 58 个可用区，进一步提升全球算力调度能力。同时，腾讯云在海外市场布局加速，近期宣布将在沙特建设首个中东数据中心，并计划在未来数年内投入超 1.5 亿美元，用于基础设施、算力资源相关投资，以扩展弹性计算、存储、安全、AI 云服务的全球供应能力。随着国内外 AI 产业链的持续升级，腾讯在云计算及 AI 训练推理领域的战略投资预计将持续深化，并开始对外界用户使用腾讯云服务。

同时，腾讯也在持续强化其自研基础大模型混元的开发。该模型已在 LLM 方向完成多轮迭代与架构衍生，最新版本融合了 Mamba-Transformer-MoE 的混合架构。在性能表现方面，T1 版本在深度推理能力与 CoT 方面已达到业内顶尖水平。同时，混元在视频生成方向同样表现亮眼，于 2024 年 12 月在 Hugging Face 榜单中排名第一，进一步验证了其在多模态领域的技术实力。

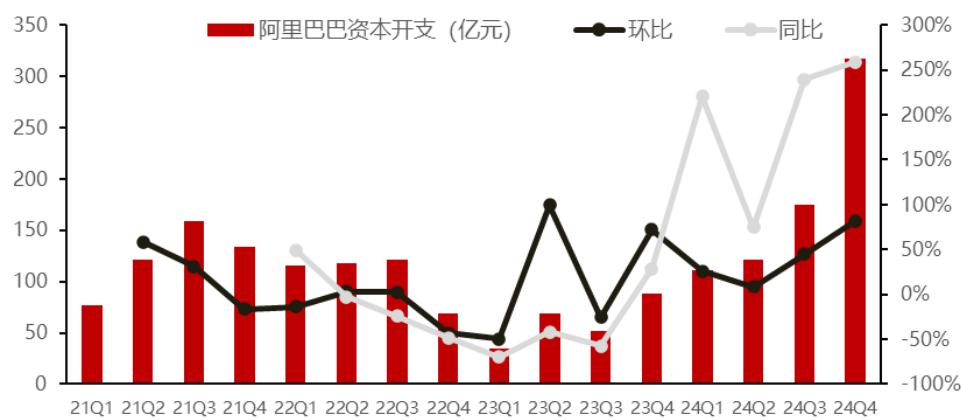
在模型策略方面，腾讯采取多模型产品布局，以更好适配不同用户需求。在 C 端，公司推出面向普通用户的元宝应用，通过支持不同推理深度的输出路径，实现从网络侧获取综合性答案，满足用户在语言、视频等多模态交互中的多样化指令需求。在 B 端，混元模型已广泛部署于腾讯会议中，支持实时语音转文字与会议摘要生成等功能，助力腾讯会议 MAUs 同比实现翻倍增长。此外，公司还推出 HCC 与 HAI 两大企业端 AI 服务平台，提供低延迟、高精度的训练与推理能力，并结合可视化编排功能，显著提升企业客户在本地化部署与调用大模型时的效率。

展望未来，2025 年腾讯资本开支将继续扩大，预计维持在全年收入的低两位数占比区间。公司将继续加大对开源模型与自研通用大模型的研发投入，推动 AI 训练技术的前沿探索。同时，基于自身庞大的业务体系，公司也将在各大商业板块中深化 AI 应用，进一步促进混元模型在 C 端与 B 端的广泛渗透与实际落地，提升产品生态竞争力。

**图21：腾讯 AI 投资展望**


资料来源：腾讯官网，民生证券研究院

**阿里巴巴也在 AI 领域投资上持续加码。**在最新业绩会上，阿里巴巴宣布 2024 年第四季度资本开支达 318 亿元，环比大增 80%，进一步加大对 AI 领域的战略投入。公司管理层指出，未来三年将是 AI 产业扩张的关键窗口期，并将在这一时期投入超过 3800 亿元人民币，这一金额超过过去十年总和。阿里云将全面深化 AI 计算基础设施布局，以应对 AI 技术增长和国内产业生态的蓬勃发展。相关的投资将重点集中于大模型研发与开源，AI 硬件设施升级以及生态协同为行业提供普惠、高效的技术支持。此次大规模投资不仅奠定了阿里在 AI 计算领域的核心地位，也标志着国内民营企业在云计算与 AI 基础设施建设上的最大规模资本投入。

**图22：2021Q1-2024Q4 阿里巴巴资本开支**


资料来源：阿里巴巴季报，民生证券研究院

**阿里巴巴的资本投资主要集中在云计算基础设施建设和 AI 算力优化领域。**在 AI 和云计算基础设施方面，阿里巴巴正加速提升数据中心（IDC）算力规模，以满足日益增长的 AI 计算需求。目前，阿里云已在全球 29 个地域，87 个可用区部署云基础设施，并计划进一步拓展数据中心布局，以支撑 AI 计算的规模化扩展。随着算力需求的激增，阿里巴巴也在加大对高性能 AI 加速卡的采购，同时积极推进国产芯片自研与部署。旗下平头哥半导体已推出首款云原生处理器芯片——倚天

710，并已在阿里云数据中心实现大规模部署，算力性价比提升超30%。此外，阿里还推出了高性能RISC-V处理器如玄铁910等产品，推动全球RISC-V生态发展，同时增强其在云计算及AI推理领域的技术竞争力。

**阿里巴巴在数据中心功率节能领域也取得显著进展。**阿里云的仁和数据中心已部署国内最大规模的液冷计算集群，成为国内首座5A级绿色数据中心。该中心采用单相浸没式液冷技术，可实现15kW至100kW机柜功率密度的灵活扩展，有效降低数据中心对空调等传统散热设备的依赖，将PUE优化至1.09，显著降低AI服务器的运行能耗。此外，阿里推出的磐久ImmersionDC1000系列液冷解决方案，在提升散热能力的同时，实现50%设备空间优化，36%能耗节省，为AI计算基础设施提供长期可持续的发展路径。

**阿里巴巴的资本投资聚焦于云计算基础设施和AI算力优化，尤其在Qwen系列模型上持续加码，巩固其AI领先地位。**最新发布的Qwen2.5-Max在高难度问答、知识理解、通用推理等方面表现突出，并采用MoE架构，智能调用子网络，优化算力资源，提高推理效率，同时降低计算成本，增强商业化落地能力。此外，Qwen开源生态迅速扩展，截至1月，基于HuggingFace平台的衍生模型超9万个，成为全球最活跃的AI模型之一。其开源许可允许开发者零成本本地部署QwQ-32B，并通过蒸馏技术轻量化定制，广泛适用于各类行业场景，进一步强化Qwen系列的全模态、全场景AI生态布局。

**阿里巴巴正围绕AI算力、国产芯片、数据中心升级及AI赋能业务进行全面布局，资本开支持续向云计算基础设施、AI训练模型、绿色数据中心倾斜。**未来，随着Qwen系列大模型生态进一步扩展，国产AI芯片及液冷数据中心规模化落地，阿里在AI产业链的竞争力有望进一步强化。

## 3.2 国产算力基建元年，构建算力云服务新生态

云厂商资本开支的上修，我们认为不仅仅是资金的投入，更是生态的全面完善和构建。

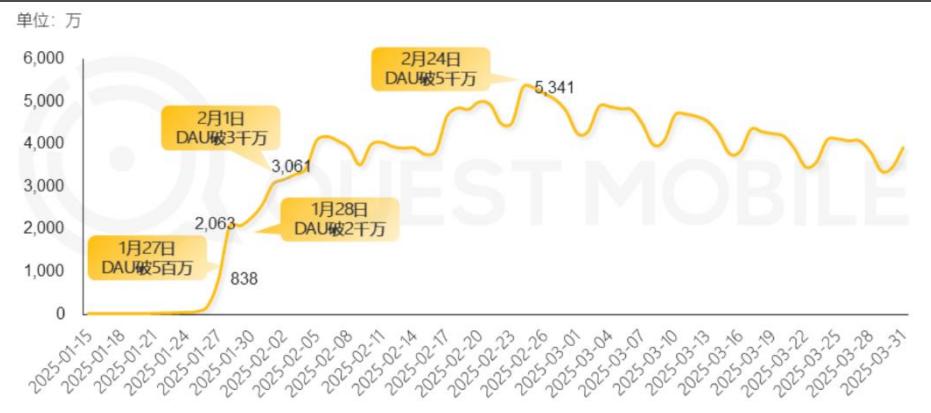
### 3.2.1 国产大模型蓬勃发展，催生AI推理需求

国产大模型迎来蓬勃发展，Kimi、文心一言等大模型改变了国内搜索习惯，而豆包凭借其优异的推理能力及抖音、头条等自有超级流量入口引流，在24年下半年迅速出圈，24年6月-11月豆包月活用户由2751万增长至5600万，实现翻倍增长。而25年初DeepSeek以其强大的算法和框架优化能力开创了大模型“高性能+低成本”的时代，平价AI推理算力给AI应用的迅速落地创造土壤，海量推理算力需求重新定义AI云计算业态。

据统计机构QuestMobile，DeepSeek日活在2月24日突破5000万，目

前已成为国内活跃用户数量最大,全球活跃用户数量第二大的AI APP。DeepSeek目前成为了历史上活跃用户增长最快的APP,短时间内海量用户的接入引起云端算力需求短缺,大量服务器繁忙问题影响了短期用户体验,也反馈出云端AI算力大幅增长的态势。

**图23: 1Q25 DeepSeek 日活用户规模趋势**



资料来源: QuestMobile, 民生证券研究院

**云厂商纷纷推出 DeepSeek 云上部署服务。**在 DeepSeek 大模型爆火后,得益于其开源策略,国内各大云厂商纷纷推出 DeepSeek 全系模型云上部署服务,成为用户在 DeepSeek 官方服务器繁忙时的全新选择:

**1) 火山引擎:** 2月4日宣布支持 DeepSeek 开源模型,可在火山引擎机器学习平台 veMLP 部署和火山方舟调用模型,并提供全网最高的 500 万 TPM 限流。

**2) 腾讯云:** 2月2日,宣布高性能应用服务 HAI 支持一键部署 DeepSeek-R1 大模型,且随后在腾讯地图、腾讯云 AI 代码助手、腾讯元器、腾讯云 TI 平台等十余款腾讯云产品接入 DeepSeek 模型,提供 671B、1.5B、7B、32B、70B 等多种尺寸。

**3) 阿里云:** 2月3日,宣布 PAI Model Gallery 支持开发者在云上一键部署 DeepSeek-R1 和 DeepSeek-V3,以“零代码”实现从训练到部署再到推理的全过程;2月9日,旗下百炼大模型服务平台全面上线 DeepSeek 模型,用户享有 100 万 token 的试用额度。

**4) 华为云:** 2月1日,宣布与硅基流动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,能够获得持平全球高端 GPU 部署的效果;2月12日,宣布 ModelArts Studio(MaaS)平台支持昇腾适配版的 DeepSeek-R1/V3(满血版)体验,并且提供 200 万 token 支持开发者免费使用;

**5) 百度云:** 2月3日,宣布千帆 ModelBuilder 平台支持 DeepSeek-V3/R1,API 价格极低;2月8日,支持在百舸 AI 异构计算平台部署满血版 DeepSeek-V3/R1,并同步在客悦、曦灵、一见、甄知大模型产品应用上接入 DeepSeek 大模型。

**表7：各云厂商 DeepSeek 系列模型 API 定价**

云厂商	模型版本	输入价格 (元/百万token)	输出价格 (元/百万token)	免费token额度 (万)
火山引擎	DeepSeek-V3	2	8	50
	DeepSeek-R1	4	16	50
腾讯云	DeepSeek-V3	2	8	限时免费至2025年
	DeepSeek-R1	4	16	2月25日
阿里云	DeepSeek-V3	2	8	100
	DeepSeek-R1	4	16	100
华为云	DeepSeek-V3	-	-	200
	DeepSeek-R1	-	-	200
百度云	DeepSeek-V3	0.8	1.6	限时免费至2025年
	DeepSeek-R1	2	8	2月18日

资料来源：各公司官网，民生证券研究院

注：API 定价均为非优惠价格

### 3.2.2 AI 部署成本降低，云服务新生态助力“低成本”转化“高收入”

**DeepSeek 从性能与成本上改变了行业格局，将带动云厂商构建算力云服务新生态。** DeepSeek 大幅降低了企业和个人的 AI 部署成本，提升了开源可触达的大模型性能，大幅提升了客户部署意愿，提升了云计算市场规模，而全新的业态也对云厂商提出了新的要求。一方面，云厂商的本质是提供算力和服务，在 AI 大模型全民部署的情况下，云厂商需要向上兼容更多的大模型，向下提升用户触达的深度和广度。另一方面，但是仅依靠部署是不够的，云厂商需要构建算力云服务新生态来将“低成本”转化为“高收入”：

**1) 软硬件协同优化：**目前国内头部云厂商均具备 AI 硬件生产能力，例如阿里含光、百度昆仑和华为昇腾，其中昇腾与昆仑芯已经宣布适配 DeepSeek 系列模型。云厂商有望形成“硬件-软件-大模型”的正向循环，通过硬件兼容加快大模型部署，并从软件层面进行优化协调，例如昇腾自研推理引擎 MindIE，与大模型实现深度适配、全栈整合，提升大模型在云平台运行的性能和效率。

**2) 以产品构建生态：**海外已有云厂商通过产品构建服务生态，微软开发了 copilot 并植入 Windows、Office、Bing 等软件中，通过产品庞大的用户基础实现对云服务的深度绑定。国内方面，金山云已经部署 DeepSeek，有望效仿微软，以 WPS Office 作为切入点构建服务生态。另外，百度、阿里、腾讯等企业旗下亦有细分领域尖端产品，有望实现“算力-AI 大模型-产品”生态闭环。

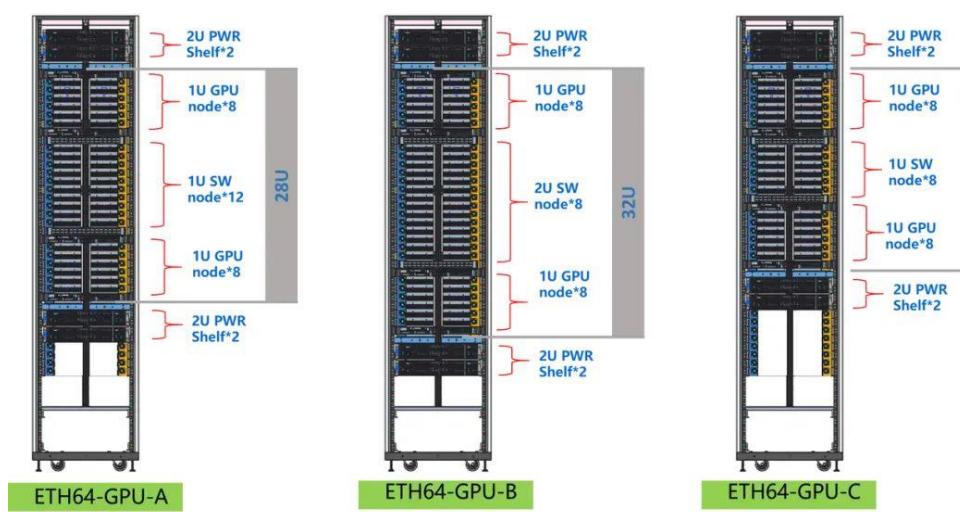
**3) 以服务构建生态：** DeepSeek 大模型的超高性价比必将带来 AI 应用快速增长，垂直领域专业 AI 应用将应运而生，云厂商有望通过提升与开发者配合的应用服务能力获得竞争力，目前昇腾已经具备为客户提供从服务化到引擎，再到底层算子的全栈解决方案的能力。

### 3.3 服务器：算力之基座，全面适配国产算力

大模型蓬勃发展催生算力需求，云厂商资本开支高增，AI 服务器作为算力之基，率先受益于国内算力基建。传统大型企业倾向于采购**品牌服务器厂商**提供的标准化服务器。2024Q1 戴尔、浪潮、联想、惠普、超微及其他品牌厂商以强大的品牌影响力、全面的服务和支持，占据 64% 的市场份额。**白牌服务器厂商**通过提供定制化的服务器解决方案，能满足云计算厂商的特定需求。如广达、纬颖等 ODM 厂商，2024Q1 占据了服务器市场 36% 的市场份额，华勤技术则为内资白牌服务器 ODM 龙头，成为云厂商 AI 服务器核心供应商。

美国对英伟达 AI 算力卡实行出口限制，国内对于自主可控的追求必将引起国产 AI 算力芯片浪潮，除华为昇腾、寒武纪、海光信息等算力芯片收到云厂商及智算中心批量采购外，沐曦、燧原、摩尔线程、壁仞等国产算力芯片亦发展迅速，**国内服务器厂商已推出适配国产算力芯片的服务器，为国产算力爆发做好准备**。由开放数据中心委员会 (ODCC) 主导、中国信通院与腾讯牵头设计，联合华勤技术、立讯技术等 30 余家产学研机构共同研发的 ETH-X 开放超节点项目，在华勤技术东莞智能制造基地下线。ETH-X 以超大带宽、超低时延、开放互联技术为核心构建业界首个弹性超节点系统，探索在单芯片算力发展受限情况下突破算力瓶颈的新途径，各子系统可以由专业计算、交换、互联厂家按照 ETH-X 硬件规范独立研发生产并完成多厂家互联互通，同时可以支持多种不同厂家 GPU 芯片、交换芯片的组合使用。

图24：ETH-X 超节点方案



资料来源：开放数据中心委员会，民生证券研究院

**浪潮信息：**公司全球领先的 IT 基础设施产品、方案和服务提供商，为客户提供云计算、大数据、人工智能等各类创新 IT 产品和解决方案。根据 IDC 数据，

2024 年公司服务器市占率达全球第二，中国第一；公司发布了业界首个仅靠 4 颗 CPU 运行千亿参数大模型的 AI 通用服务器 NF8260G7，能够满足基于大模型的 AI 应用及云计算、数据库等通用场景；并发布了元脑服务器第八代新品，其基于开放架构设计，在业界率先实现“一机多芯”，具备更全面的智能能力和更高能效，能够更好支撑 AI 大模型开发与应用创新。2024 年，浪潮信息还与 Intel 联合发布了 AI 通用服务器，业界首次实现服务器基于通用处理器支持千亿参数大模型的运行，能够灵活满足基于大模型的 AI 应用及云计算、数据库等通用场景。

**华勤技术：**公司为 ODM 平台型公司，为全球安卓系 ODM 龙头，产品涵盖手机、平板、PC、可穿戴、AIOT、数据产品、车载产品等。公司 2017 年进入服务器市场，超前布局高性能计算和 AI 驱动的数据处理需求，能够提供通用服务器、存储服务器、AI 服务器、交换机的全栈式产品组合，成为多个国内知名互联网厂商的核心供应商。**2024 年公司数据业务收入突破 200 亿元，AI 新芯片平台行业率先发货，用于客户人工智能大模型训练及行业推理场景，25 年公司数据业务目标实现 300 亿营收。**

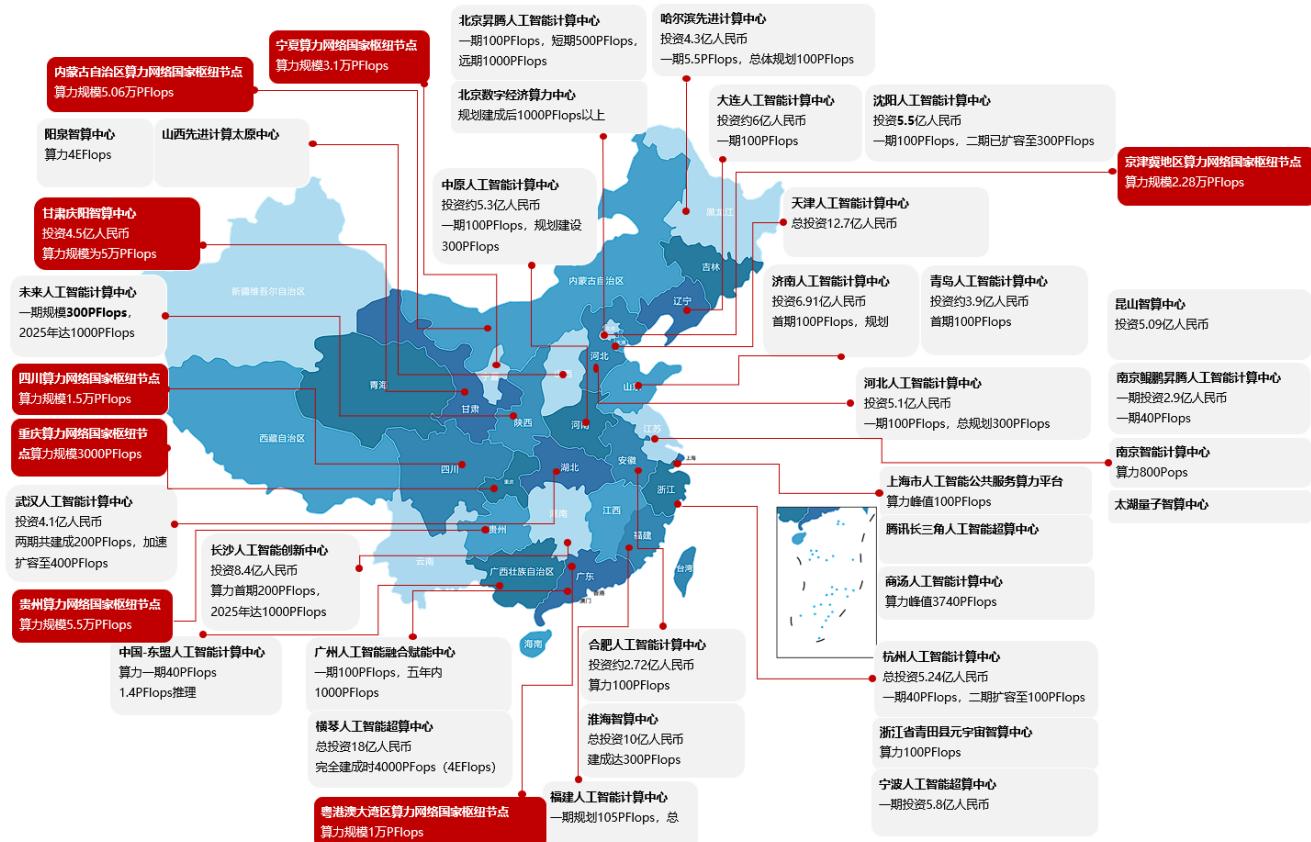
### 3.4 算力租赁：短期内算力短缺破局之道

**算力租赁助力解决短期内算力缺口问题。**算力租赁，即算力资源方通过自建或外购算力的方式，形成一定规模的自有算力资源，然后以租赁收费的模式，为 AI 公司、科研院所、云厂商等提供算力支持。算力租赁厂商一般拥有 AI 算力卡拿卡渠道，以及有充沛的在手现金可以满足重资产的算力投资，能够为没有足够现金自建机房的中小机构提供算力支持，同时也能够为云厂商提供 AI 算力补充。

DeepSeek 推出以来，各大 AI 公司、开发者甚至制造业纷纷尝试开发相关垂类应用，同时云厂商也积极接入 DeepSeek，如腾讯“元宝”接入 DeepSeek 后 APP 下载量登顶排行榜第一，催生广泛算力需求，算力租赁价格随之上涨，**根据人民财讯 3 月 5 日报道，闲鱼平台近三个月高端显卡算力租赁搜索量增长 300%，RTX 3090、4090 等型号显卡成了“电子硬通货”。根据 CloudPrice，AWS 的云计算 (H100, us-west-1) 均价近一个月来上涨了 76%。**

**算力租赁公司成为各地智算中心的重要算力提供商。**除大模型厂商、云厂商直接对算力租赁公司下单，各地智算中心算力建设如火如荼，亦成为算力租赁公司重要客户。智算中心一般由当地政府主导规划，算力提供商（算力建设出资方）包含资金充沛的央国企、CSP 以及算力租赁公司等，其中算力租赁公司为智算中心算力建设的重要一环。根据通信产业网，**甘肃、宁夏、内蒙等八大算力网络国家枢纽节点 2024 年算力规模之和超过 20 万 PFlops，算力建设需求旺盛，算力租赁公司深度受益。**

图25：中国智算中心布局



资料来源：黑龙江新闻网、金普新闻、华为官网、天津数据发展中心、济南公共资源交易中心、山东省人民政府官网等，民生证券研究院

表8：部分算力租赁业务梳理

企业名称	简介
弘信电子	公司战略性布局 AI 算力服务器硬件研发和生产制造与 AI 智算中心算力服务两大业务。公司算力业务的战略为将公司打造为算力硬件及整体解决方案提供商，既涉及到与合作伙伴共同进行算力卡及算力服务器等算力硬件的研发生产与销售，亦会根据大模型客户、垂直应用客户及其他各种类型客户的需求开展算力服务业务。公司目前在甘肃天水设有 AI 服务器智能制造工厂。
海南华铁	算力租赁是华铁应急推进围绕经营租赁主业开展的多品类扩张战略落地的举措之一，旨在打造新的盈利增长点，增强公司抗风险能力，因此海南控股将大力支持华铁应急创新性租赁业务全面发展。
宏景科技	公司与中国移动宁夏公司于签署了《中国移动宁夏公司 2025 年算力服务引入采购框架协议》，为中国移动宁夏公司提供算力服务，本框架协议含税总金额上限为人民币 3.12 亿元。
首都在线	公司深度践行“一体两翼”战略规划，以“一云多模”、“一云多芯”、“一云多池”为切入点，全力打造基于“M 种大模型”与“N 种芯片”的首都在线智算云平台。公司目前在全球布局了近 100 个数据中心节点，还规划在美国达拉斯、海南、张家口怀来、芜湖建设 4 个自有数据中心。

资料来源：各公司公告，民生证券研究院

弘信电子专注于提供灵活高效、安全易用、智能运维的算力租赁服务。公司于2023年开展AI算力租赁服务，致力于为客户提供量身定制混合异构算力资源池适配服务及全方位的算力运维解决方案，拥有智能运维、高效稳定和快速响应优势。

公司资源池集成了市场上主流的英伟达合规算力资源，如 4090D、L40S、A800、H800、L20、H20 等型号，以及燧原的 I20、T20、S60 等型号，可根据客户需求灵活配置，重点服务于互联网大厂、三大运营商、头部大模型公司及自动驾驶头部企业等大公司。

**图26：弘信电子算力租赁方案优势**

智能运维	高效稳定	快速响应
● <b>自动化监控</b> 运用先进的监控技术，对算力资源进行全天候的实时监控，自动检测性能指标和潜在问题，确保服务的稳定运行。	● <b>高性能硬件</b> 部署业界领先的高性能硬件设备，确保算力服务的快速响应和高效率，满足客户的算力需求。	● <b>实时报警系统</b> 建立完善的实时报警系统，能够快速识别问题并通知运维团队，确保问题得到及时处理。
● <b>预测性维护</b> 借助大数据分析，预测系统故障，提前进行维护和干预，有效减少停机时间，保障服务的连续性。	● <b>冗余设计</b> 通过冗余设计，确保关键组件的可靠性，降低单点故障的风险，为客户提供稳定可靠的算力环境。	● <b>快速故障恢复</b> 制定高效的故障恢复流程，能够在最短时间内恢复服务，减少客户损失。
● <b>智能调度</b> 根据工作负载和需求的变化，智能调度系统能够自动调整资源分配，优化资源利用率，确保客户获得最佳的计算体验。	● <b>负载均衡</b> 采用先进负载均衡技术，合理分配任务，避免系统过载，保持服务的稳定和高可用性。	● <b>敏捷部署</b> 具备敏捷的部署能力，能够迅速部署及更新服务，以响应市场和客户需求的变化。
	● <b>持续性优化</b> 定期对系统进行性能评估和优化，确保服务的长期稳定运行，为客户提供持续高效的服务体验。	● <b>客户支持</b> 提供快速、专业的客户支持服务，确保客户问题能够及时得到解决，提升客户满意度。

资料来源：弘信电子官网，民生证券研究院

**润泽科技是全国领先的综合算力中心领域的科技公司。**公司采用“自投、自建、自持、自运维”高等级综合算力中心集群的模式保证服务的长期性、稳定性和可靠性，已在京津冀·廊坊、长三角·平湖、大湾区·佛山和惠州、成渝经济圈·重庆、甘肃·兰州和海南·儋州等全国 6 大区域建成了 7 个 AIDC 智算基础设施集群，基本完成了全国“一体化算力中心体系”框架布局，全国合计规划 61 栋智算中心、约 32 万架机柜，服务于字节跳动、华为、京东等知名企业。

**表9：润泽国际信息港项目部分数据中心规格参数**

数据中心	机房面积 (平方米)	机房数量 (个)	可容纳机柜数量
A-1数据中心	11200	27	6000个电信标准机柜
A-2数据中心	16200	28	6000个电信标准机柜
A-3数据中心	11200	28	7000个电信标准机柜
A-5数据中心	11200	27	6000个电信标准机柜
A-6数据中心	11200	28	7000个电信标准机柜
A-18数据中心	-	-	6500个电信标准机柜

资料来源：润泽科技官网，民生证券研究院

**云赛智联是一家以云计算与大数据、行业解决方案及智能化产品为核心业务的专业化信息技术服务企业。**公司拥有宝山云计算中心、宝山大数据中心、松江大数据中心等机柜资源，为政府和大中小型企业提供公有云、私有云托管和混合云服务，下属公司南洋万邦和信诺时代均为微软金牌合作伙伴、均获得全国最高级别云托管服务商认证 AEMSP、AIASP 等高级别认证。

**图27：云赛智联主要合作伙伴**


资料来源：云赛智联官网，民生证券研究院

## 3.5 本地部署需求旺盛，助推 AI 一体机发展

### 3.5.1 何为 AI 大模型一体机

大模型一体机是专为企业和政府设计的“一站式 AI 工具箱”，它将高性能算力硬件（如 GPU 芯片）、大模型软件（如 DeepSeek）和安全管理功能集成在一个机柜中。用户无需连接互联网，即可在本地完成数据训练、模型部署等任务，确保敏感数据不外泄。其核心价值在于降低算力使用门槛，提升 AI 训练与推理效率，尤其适用于大模型训练、自动驾驶、智能制造等高算力需求场景。

**图28：息壤智算一体机训推一体场景**


资料来源：天翼云官网，民生证券研究院

**图29：息壤智算一体机模型推理场景**


资料来源：天翼云官网，民生证券研究院

相比于云部署模式，采用一体机进行私有化部署具有以下优点：

- 1) **数据安全**：本地化部署，数据不脱离内网。相对于互联网客户，**政府、金融、电信**等行业客户具备大量敏感数据，处于对数据安全、隐私安全的考虑，上述客户对于使用大模型，以及将数据上传至云端进行训练持保守态度。

2) **简化部署**: 企业在大模型开发全流程中, 常常面临切换硬件集群、适配底层硬件以及微调优化等复杂问题。采用 AI 算力一体机可以实现在**同一个集群内无缝切换大模型的训练和推理过程**, 从而为客户提供高效的大模型训练和推理部署流程。同时, 解决方案内置了大模型推理引擎、调优工具和算子加速库, 支撑客户**开箱即用**, 无需进行二次硬件适配, 即可让客户能够直接使用大型预训练模型, 从而降低了技术门槛;此外, 通过芯片层面的算法调优, 能够充分释放硬件性能, 从而**提升模型训练和推理的效率**。

3) **成本更低**: 短期看通过云服务使用大模型无需一次性硬件投入, 但长期使用公有云 API 按 token 付费成本较高, 通过一体机私有化部署有助于**降低总体成本**并更好地掌控预算。

**表10: 传统云服务和大模型一体机技术对比**

比较项目	传统云服务	大模型一体机
安全性	数据需上传至云端, 存在泄露风险	数据完全本地处理, 物理隔离
安装便捷程度	部署周期长达数周	3小时开机即用
成本	按流量/时长付费, 成本不可控	一次性采购, 5年运维成本降低40%
兼容性	依赖国外芯片 (如英伟达)	支持国产芯片 (华为昇腾等)

资料来源: CSDN、民生证券研究院

### 3.5.2 DeepSeek 赋能 AI 一体机发展

DeepSeek 通过其卓越的模型性能与极低的推理成本, 为 AI 一体机的快速发展提供了核心技术支撑, 推动国产 AI 生态的全面落地。

技术层面, DeepSeek 的 V3 和 R1 系列大模型采用创新的 FP8 混合精度训练框架与 DualPipe 算法, 不仅性能媲美 GPT-4 等国际顶尖模型, 还将训练成本降低 90%以上, “**高性价比+开源**”特性使其成为 AI 一体机的理想基座。

除 671B 的“满血版本”, DeepSeek 还使用知识蒸馏技术, 将 R1 大模型的复杂知识及思维链能力蒸馏至 Qwen/Llama 的开源小模型中, 实现模型的轻量化, 使用 DeepSeek-R1 进行蒸馏后的小模型推理能力显著提高, 甚至能够超越 o1-mini, 表明了大模型的推理能力向小模型迁移的可能。**蒸馏后的小模型参数量在 1.5B-70B 不等, 适合利用价格相对便宜、配置相对较低的一体机进行本地部署, 当客户有成本控制需求, 且对性能没有过高追求时, 低成本一体机。**

**图30：蒸馏模型性能对比**

	AIME 2024 pass@1	AIME 2024 cons@64	MATH-500 pass@1	GPQA Diamond pass@1	LiveCodeBench pass@1	CodeForces rating
GPT-4o-0513	9.3	13.4	74.6	49.9	32.9	759.0
Claude-3.5-Sonnet-1022	16.0	26.7	78.3	65.0	38.9	717.0
o1-mini	63.6	80.0	90.0	60.0	53.8	<b>1820.0</b>
QwQ-32B	<b>44.0</b>	60.0	90.6	<b>54.5</b>	<b>41.9</b>	1316.0
DeepSeek-R1-Distill-Qwen-1.5B	28.9	52.7	83.9	33.8	16.9	954.0
DeepSeek-R1-Distill-Qwen-7B	55.5	83.3	92.8	<b>49.1</b>	37.6	1189.0
DeepSeek-R1-Distill-Qwen-14B	69.7	80.0	93.9	<b>59.1</b>	53.1	1481.0
DeepSeek-R1-Distill-Qwen-32B	<b>72.6</b>	83.3	94.3	62.1	<b>57.2</b>	1691.0
DeepSeek-R1-Distill-Llama-8B	50.4	80.0	89.1	49.0	39.6	1205.0
DeepSeek-R1-Distill-Llama-70B	70.0	<b>86.7</b>	<b>94.5</b>	<b>65.2</b>	<b>57.5</b>	1633.0

资料来源：DeepSeek 官网，民生证券研究院

目前，DeepSeek 已赋能金融、医疗、汽车等 20 余个行业。这种“模型+算力+场景”的深度融合，使得 AI 一体机能够以开箱即用、灵活定制的特性，降低企业智能化转型门槛，加速“AI+垂类应用”的全方位落地。政务或教育、医疗等行业有不同的专属数据，利用特定数据对大模型投喂而开发的垂类应用，不仅具备隐私性、快速响应的优势，更重要的是客户可以利用个性化的行业数据对模型进行调优，生成更符合行业特点和使用习惯的模型。此前的端侧模型不具备思维链能力，DS 可清晰展示模型思考过程，客户亦可对其思考过程进行反馈，使模型的思考方式与企业自身所需趋同，有利于生成更加个性化、专业化的模型。

**表11：DeepSeek 一体机生态图谱**

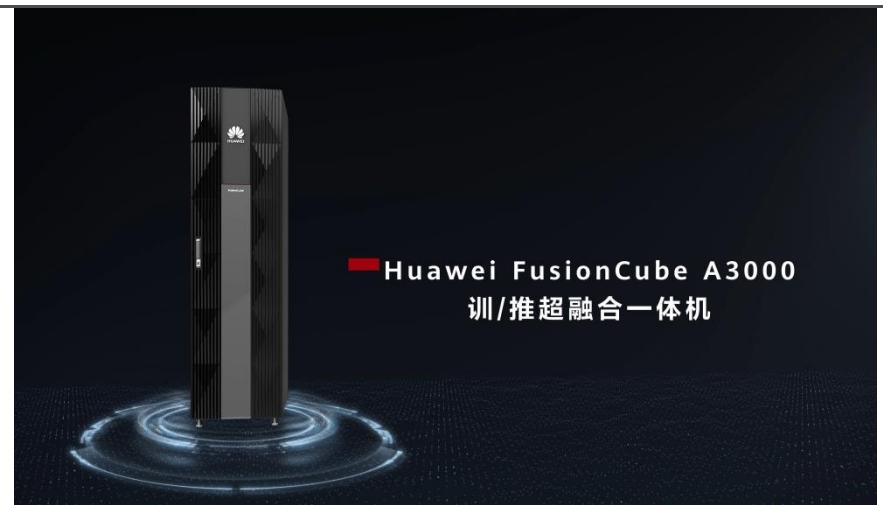
领域	相关案例
教育	智海AI教育一体机依托DeepSeek模型构建全栈国产化方案，支持从单机到集群的灵活配置，将AI推理资源消耗降低60%，使教育机构能以低成本获得高性能AI教学体验
政务	科华数据推出的全国产算力政务智能体一体机搭载DeepSeek-R1蒸馏模型，通过自研RISC-V加速卡实现1.2PFLOPS算力，24小时智能咨询服务准确率达95%，推动政务服务智能化转型
制造	LED显示企业通过模型生成个性化内容，使传统显示屏升级为智能交互终端
医疗	天津医科大学总医院联合中国移动天津公司完成DeepSeek“智算一体机”部署
生态建设	天翼云息壤智算一体机实现从芯片到模型的全栈国产化，紫光股份新华三的灵犀使能平台则提供一站式模型管理服务，支持多行业快速构建智能应用
智慧城市	大华股份基于DeepSeek-R1开发的安防一体机，可实现实时视频分析与行为预测

资料来源：公司官网、腾讯网等、民生证券研究院

### 3.5.3 各大厂商竞相推出 DeepSeek 一体机产品

昇腾推出从服务器、推理卡、到加速模组等丰富的一体机产品形态，全面适配 DeepSeek V3/R1 满血版 / 蒸馏版本全系列模型，涵盖智能对话、智能编程、文档分析、开发板等全场景应用，能够满足各类用户在不同场景下的 AI 需求。

图31：华为推出 FusionCube A3000 训/推超融合一体机



资料来源：华为官网，民生证券研究院

**浪潮创新技术服务器 NF5468H7** 深度融合 DeepSeek 全系列模型与国内 AI 加速卡，打造“开箱即用”的智算解决方案。全面兼容 DeepSeek V3 (通用底座)、R1 (推理优化)、Janus Pro (多模态) 等系列模型，支持智能客服与对话、文本分析、工业质检、医疗影像识别等场景快速适配。

图32：浪潮云帆超融合一体机 (智算型-7000 G3)



资料来源：光明网，民生证券研究院

**华勤推出 DeepSeek 一体机方案**，兼容 NV 和国产算力，并可根据模型大模型参数提供不同配置。**H9236 服务器**采用 8 卡 H20 模组的方案；**H8230 服务器**采用 PCIe 插卡方案，可通过 4-8 台小集群组网灵活配置。实现 DeepSeek-R1/V3 671B 的本地部署；**P6236 服务器**采用单机部署的方案，GPU 可配置 1-2 块提升推理性价比。

**图33：华勤技术 H9236 服务器**


资料来源：华勤数据中心业务微信公众号，民生证券研究院

**联想**全栈自研的 AI 工作站和 AI 一体机正式推出。这一系列产品基于联想“全栈 AI”战略，深度融合了“软硬服一体”协同的 AI 能力，预装合作伙伴 DeepSeek 模型(包括 DeepSeek-R1 不同版本)，用户可根据业务场景灵活选择从 7B 到 671B 满血版本的不同容量参数模型，快速完成部署，无需额外配置。

此次发布的 AI 工作站与 AI 一体机家族以“**高效部署、场景覆盖、智能协同**”为核心设计理念，全面适配包含 DeepSeek 在内的多款大模型。通过 AI 工作站和 AI 一体机，用户无需复杂的部署与配置，即可**快速启动智能体的开发与应用**，实现从模型训练到应用落地的全过程管理，进一步简化了 AI 技术的应用流程，助力企业用户轻松接入并快速运行。

**图34：联想 AI 工作站和 AI 一体机方案**


资料来源：联想官方公众号，民生证券研究院

**弘信电子**双擎布局一体机。公司正式推出搭载国产高性能 GPU 算力卡燧原 i20/S60 的燧弘一体机，进一步推动国产 AI 算力生态的升级。此外，集团子公司

安联想作为英伟达中国区双料 (GPU+网络) Elite 级别合作伙伴, 也宣布推出搭载 NVIDIA 高性能 GPU L20 及 H20 的 **Anlink 一体机**, 满足不同场景的应用需求。两款一体机均适配 DeepSeek-671B 满血版、DeepSeek-R1-Distill-Qwen-32B、DeepSeek-R1-Distill-Qwen-70B 等高效蒸馏小模型, 支持开箱即用、私有化部署及灵活配置。

图35: 弘信电子燧弘一体机和 Anlink 一体机



资料来源: 厦门电子官方公众号, 民生证券研究院

表12: 国内 AI 一体机梳理

供应商	一体机名称	详细信息
华为昇腾	FusionCube A3000 训/推超融合一体机	推出从服务器、推理卡、到加速模组等丰富的一体机产品形态, 全面适配 DeepSeek V3/R1 满血版 / 蒸馏版本全系列模型
联想集团	AI 工作站和 AI 一体机	深度融合“软硬服一体”协同的 AI 能力, 预装合作伙伴 DeepSeek 模型 (包括 DeepSeek-R1 不同版本), 用户可根据业务场景灵活选择从 7B 到 671B 满血版本的不同容量参数模型
云天励飞	云天书大模型训推一体机	该一体机由云天励飞和华为联合推出, 已成功适配 DeepSeek, 可以部署在华为昇腾服务器上, 支持私有化部署, 实现开箱即用, 满足本地化、专属化的业务需求。
浪潮信息	浪潮云帆超融合一体机	深度融合 DeepSeek 全系列模型与国内 AI 加速卡, 打造“开箱即用”的智算解决方案
中科曙光	DS 人工智能一体机	内置曙光自研 AI 管理平台 SothisAI3.0, 支持从 10 亿级参数模型推理到 1000 亿级参数模型训练的 AI 全场景需求
云从科技	从容大模型训推一体机	集成大模型一体化生产平台, 具备模型训练、推理、管理全流程等功能, 支持语言、视觉、多模态三大类基础模型, 可灵活适配十亿至千亿级参数规格
软通动力	软通训推一体化平台	推出结合昇腾 AI 芯片的“软通训推一体化平台”, 并发布基于昇腾底座的软通天璇 2.0 Maas 平台, AI 一体机取得华为及鲲鹏技术认证, 进入华为云商城
中国电信	息壤智算一体机 - DeepSeek 版	深度融合 DeepSeek-R1/V3 系列大模型, 天翼云全力加码, 在“息壤”智算平台、科研助手等多个平台上线 DeepSeek-R1 和 V3, 提供私有云、混合云和一体机等多元化部署方案
中国移动	移动云智算一体机 - DeepSeek 版	支持 DeepSeek 预置模型的一键推理、微调 (蒸馏模型) 及智能体构建, 预置的 DeepSeek 模型助力企业快速构建智能体, 实现复杂任务的自动化处理

中国联通	DeepSeek 一体机	基于“星罗”算力平台预置多尺寸模型，适智一体化算力网支撑
京东云	京东云 DeepSeek 大模型一体机	以“国产算力全适配”和“开箱即用”为核心优势，内置智能体与知识库双引擎，预置 100 多个行业解决方案模板及千种插件，支持本地知识库快速构建
百度	百舸 DeepSeek 一体机	基于百度百刚平台打造，提供纯国产的算力组合，支持昆仑芯 P800 单机 8 卡部署满血版 DeepSeek R1/V3，支持 8bit 推理。
龙芯	龙芯 DeepSeek 推理一体机	政务金融安全之选，支持 100 人并发推理，单卡 FP16 算力达 128 TFLOPS
昆仑技术	昆仑 DeepSeek 本地化部署一体机	基于昆仑技术的 KunLun AI 服务器强劲算力和中科加索的 SigInfer 模型推理引擎，能够实现卓越的性能表现，为用户提供稳定可靠的生产级 DeepSeek-R1 和 V3 服务
科大讯飞	星火+DeepSeek 双引擎一体机	讯飞星火 DeepSeek 塔式一体机支持办公室活动部署，讯飞星火 DeepSeek 一体机适配机房部署，覆盖多种部署场景，助力开发者和科创企业加速业务创新
拓维信息	拓维信息智能数据标注一体机	基于“异腾+鲲鹏”打造的“兆箭”AI 算力硬件，可本地化轻松调用 671B 满血版及蒸馏版 DeepSeek 模型，完成了软硬件对 DeepSeek 满血版 671B 与蒸馏版 70B、32B、8B、7B、1.5B 等不同参数版本的适配工作

资料来源：联想集团官方公众号，北京通信信息协会，云从科技官网等，民生证券研究院整理

## 4 电源及液冷创新，解决功率密度瓶颈

### 4.1 全新电源架构，数据中心能源卫士

国内算力基建如火如荼，截至 2024 年末，全国在用算力中心标准机架数超过 880 万，算力总规模较上年末增长 16.5%；云厂商亦投入大量资本开支于算力基建，如阿里巴巴宣布未来三年将投入超过 3800 亿元用于建设云和 AI 硬件基础设施，这一金额超过过去十年总和。我们认为，2025 年 AI 产业的核心矛盾之一为功率密度。传统数据中心以 CPU 云计算为主，芯片功耗较低，而 AI 浪潮下，算力芯片单芯片功耗 (TDP) 快速提高，同时高密度计算要求采用机柜架构，单机柜功率密度提高，对电源系统提出新的挑战。例如，传统数据中心单机柜功率仅为 3 - 8 千瓦，而 AI 数据中心功率密度则跃升至 20 - 100 千瓦，对供电系统提出更高挑战；此外为保证数据中心稳定运行，UPS、柴发、超容等后备电源系统亦有所升级。

#### 4.1.1 解读 AIDC 的三级供电和三级备电

下图为典型的 AIDC 电气设备结构。可分为三级供电，以及三级备电。

##### 三级供电：

1) **UPS/HVDC**: UPS 即不间断电源。市电输入 UPS 后，UPS 需要首先进行 ACDC 转换，用于给蓄电池直流充电，然后进行 DCDC，把电压等级转换到合适后续设备工作的电压，再进行 DCAC 逆变到 230V 或 400V 交流电，用于后续的交流设备供电。

而 HVDC 则是高压直流输电，过去主要长距离、大容量的电力传输。高压直流可以大幅减少输电损耗。并支持更高电压+功率等级。所以目前已有部分 AIDC 机房开始采用 HVDC 方案作为第一级供电。如阿里巴巴张北云计算庙滩数据中心。

2) **ACDC**: 即机柜内部的 Powershelf，将交流电转换为算力硬件所需的稳定直流电，一般输出电压为 48V。主要供应商包括台达、光宝、麦格米特。

3) **DCDC**: 即 DCDC 电源模块，将 48V 电压进一步下变压至适用于算力芯片的工作电压，如 0.8V, 1.3V。其中具体供电策略由 Drmos 与多相电源控制器芯片决定。主要供应商包括：MPS、瑞萨、AOSL、晶丰明源。

##### 三级备电：

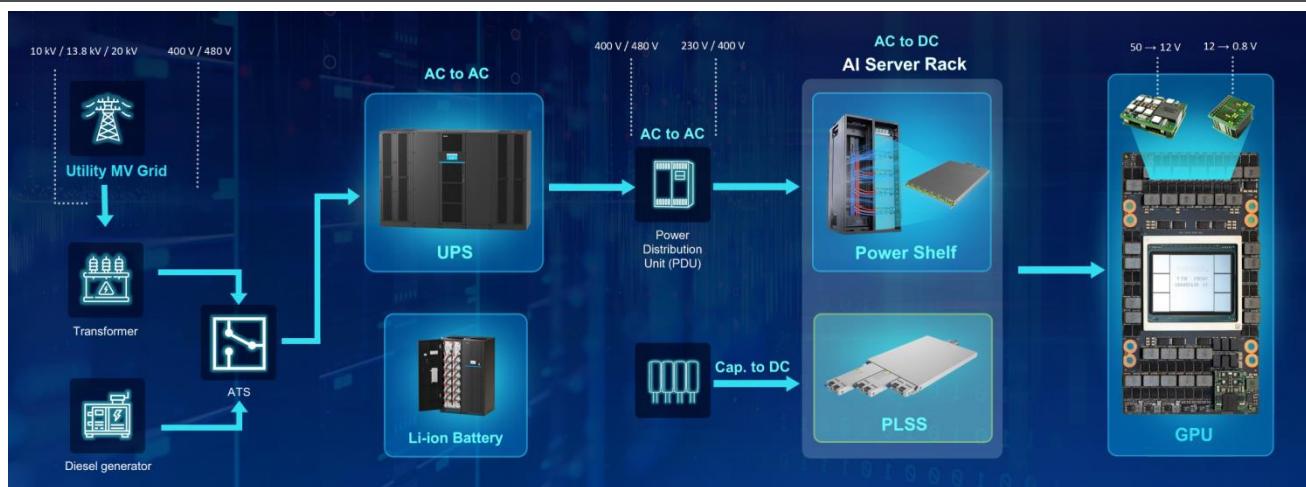
1) **超级电容**: 当遇到市电短暂中断的情况，超级电容可以在极短时间内释放存储的电能，为服务器的关键组件（如内存、缓存等）提供足够的电力，确保数据不会因为突然断电而丢失；同时，超级电容还可抑制电压波动，通过吸收和释放电能来稳定电压，消纳电网浪涌。

2) **UPS 蓄电池**: 在断电情况下，UPS 可以切换到电池供电模式，电池的直流

电会立即通过简单的逆变器电路为设备提供临时电源。

**3) 柴油发电机：**与 超级电容和 UPS 不同的是，柴油发电机只要有充足的柴油供应，就可以持续供电数小时甚至数天。这为在市电恢复之前维持服务器的运行提供了可靠的保障，同时也为采取其他措施（如抢修电网、调配移动发电设备等）争取了足够的时间。在字节 2025 年 1 月 17 日披露的火山引擎算力中心中，柴油发电机设备也作为重点项目进行了公示。

图36：AIDC 电气设备结构

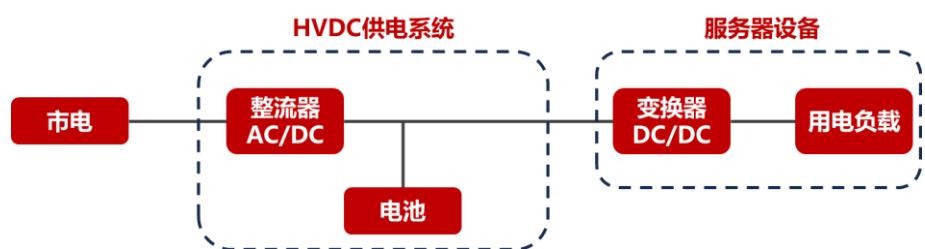


资料来源：CSDN，民生证券研究院

#### 4.1.2 从 UPS 向 HVDC 演进，高效电源未来式

传统数据中心供配电系统以 UPS 为主，目前已有数据中心升级为 HVDC（高压直流输电）系统。国内方面，最早由中国电信推进使用 HVDC，目前已成功扩展至多个领域，主要应用有阿里巴巴的巴拿马电源、腾讯的第三代数据中心 T-MDC 和第四代数据中心 T-block 等；国外方面，Intel、微软、Meta 等公司也已采用 400V 直流供电系统。相较于 UPS，HVDC 取消了逆变器环节，供电线路更加简单。当市电正常时，市电经由整流器转换为直流电后直接向服务器设备供电，同时为电池充电；当市电发生故障时，由电池直接向服务器设备供电。

图37：HVDC 全链路供电架构



资料来源：邓健威等《数据中心供配电系统节能技术探讨》，民生证券研究院

**新一代数据中心 HVDC 系统将输出更高电压。**目前普遍采用的 HVDC 系统的输出电压为 240V/336V，但是随着 AI 芯片性能不断攀升、数量不断增加，单台服务器功耗激增，以英伟达 GB200 NVL 72 为例，机架功率便达到 120kw，已经远远超过传统服务器，叠加超大算力集群建设，数据中心整体功耗水平急剧上升，传统方案难以保证效率。在这种情况下，数据中心供电系统的高电压输出成为趋势。百度 2024 年推出的“瀚海”直流电源配备了 750V 输出柜，支持单机柜 100kw，效率提升 2%-4%；台达最新推出的 10kV 供电系统 DPSST 系列最高输出电压更是达到 1000V，整机效率达到 98% 以上。

图38：百度“瀚海”一体化直流电源



资料来源：2024 数据中心标准大会，民生证券研究院

#### 4.1.3 超级电容护航 AI 服务器

超级电容器作为一种新型储能技术，具有快速充电、绿色环保、耐温范围广、输出功率高等优异性能，被广泛应用于新能源、轨道交通、工业、电网和消费电子等领域。

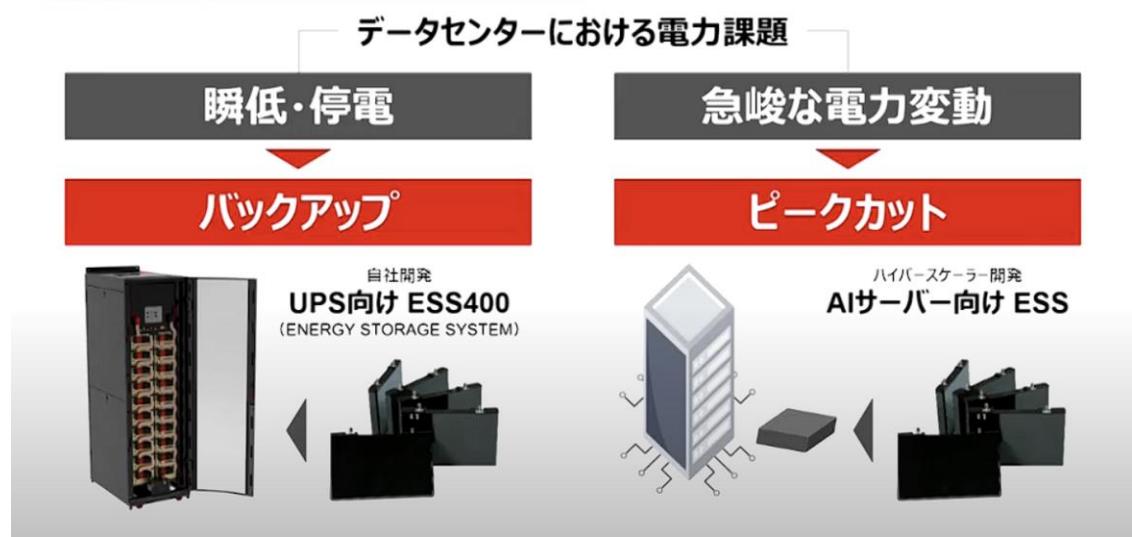
据北美领先的 AI 基础设施提供商伟创力 (Flex) 披露，其基于超容的储能系统(CESS)将于 1H25 开始生产，并将于 3Q25 开始商业化供应，**超容或将成为英伟达 GB300 系列标配**。超容在 AI 数据中心的主要作用包括：**提供紧急供电、平滑电路波动等。**

**紧急供电：**当数据中心发生停电时，**超级电容作为一级备电，会率先为服务器提供临时电源。因其响应速度极快，可在毫秒级时间内完成充电和放电。**最大程度防止服务器因供电延迟而导致的数据丢失的情况。而后则由不间断电源 (UPS) 的蓄电池以及柴油发电机接续供电。

**平滑波动：**随着 AI 芯片功耗增大，在训练或推理的过程中功率波动会更加陡峭，对电路造成损害并影响芯片性能。**超级电容器的高功率密度快速响应特性可以较好应对电源波动，能够在电力负载突然变化时立即提供瞬时功率补偿，实现削峰填谷，平抑负载浪涌，确保电压的稳定及电路和芯片的正常运行，不仅提供了高功率支持，而且优化了空间和能效。**

图39：超级电容在数据中心起到紧急供电及平滑电路波动的作用

## DX Issue & Solution



资料来源：武藏官网，民生证券研究院

### 4.1.4 相关公司

**江海股份：深耕超级电容器，已成为国内超容龙头。**作为国内电容器行业龙头企业，江海股份经过多年积累，超容产品已形成多方面的竞争优势，广泛应用领域于智能三表、风电变桨、智能电网改造、轨道交通、油改电项目、港口机械等领域。其中**锂离子超级电容器**在发电侧、用户侧调频、石油勘探、油改电等诸多应用领域开始应用，在线运营效果良好；由于很多应用都在首次采购中选择了江海的锂离子超容产品，公司在后续维护、更新及方案推进上更占据主动权，有利于公司超容产品加速渗透。

**禾望电气：拥有完整的大功率电力电子装置及监控系统的自主开发及研发实力与测试平台。**公司专注于新能源和电气传动领域，提供 0.75-22400kW 低压变频、8-136MVA 中压变频传动解决方案，可广泛应用于冶金、石油石化、矿山机械、港口起重、分布式能源发电、大型试验测试平台、海洋装备、纺织、化工、水泥、市政及其它各种工业应用场合。

**麦格米特：**公司深耕电源领域多年，是少数具备高功率高效率服务器电源技术与海外全流程生产供应能力的综合型电源供应商。2024 年公司成为英伟达 Blackwell GB200 系统 40 余家供应商之一。目前公司产品能力已从 Powershelf 拓展至一站式配电解决方案，包括 HVDC、Powershelf、BBU 和超级电容。2025 年英伟达 GTC 大会上公司展出 800V HVDC Rack 方案，电源采用 Side-car 方案，单柜功率为 570kW，内含 19 个 800V 输出的 PSU 和 19 个 30kW 整流器。

**中恒电气：深耕直流方案，在行业中享有较高地位。**作为数据中心 HVDC 技

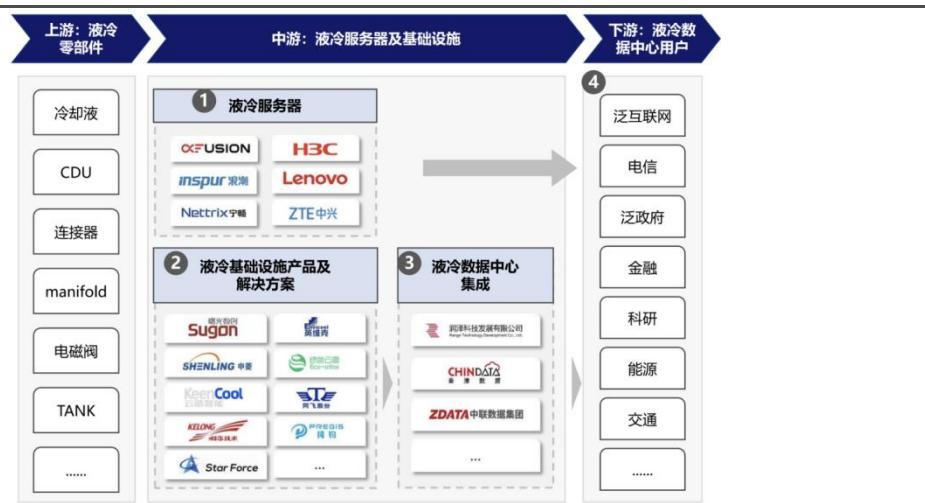
术方案的先行者，公司牵头制定了《信息通信用 240V/336V 直流供电系统技术要求和试验方法》国家标准，拥有超过 10000MW 的装机容量，能够支持单机柜 600kW 的负载，正持续探索高可靠性、集中式架构下的 750V 直流解决方案与产品，以满足未来数据中心更高效、更环保的需求。公司凭借优秀技术水平为各大数据中心提供定制 HVDC 直流配电系统，服务于阿里巴巴、腾讯、字节跳动等知名企业，先后签订“百度云计算中心项目之高压直流供应工程项目”、“2021 年腾讯高压直流及综合配电柜采购项目以及 T-Trian 电源集采项目”和“山东移动数据中心直流不间断电源项目”等。

**泰嘉股份：**公司电源业务主要涵盖消费电源、数据电源和站点电源三大领域，其中数据电子业务重点布局服务器电源模块，服务于行业头部客户。自 2023 年投产以来，公司服务器电源模块业务产能逐步释放，并根据市场需求优化生产布局，以适应 AI 服务器、电源管理方案的增长需求。此外，公司通过直接参股雅达电子，就消费电子领域的业务协同、技术研发、资源共享、资本运营等多个领域展开深入合作。同时雅达电子也成为华为主要供应商，进一步推动双方在产业链中的深度合作。

## 4.2 液冷接棒传统风冷，温控系统革新升级

液冷数据中心产业链由**上游液冷零部件**、**中游液冷服务器及基础设施**和**下游液冷数据中心用户**构成。上游液冷零部件主要包括冷却液、CDU、快接头、manifold、电磁阀、TANK 等；中游主要包括浪潮、联想等服务器制造商，以及曙光、浪潮、联想等液冷解决方案提供商；下游主要客户为互联网厂商、三大运营商、政府、科研院所等机构。

图40：中国液冷数据中心产业链

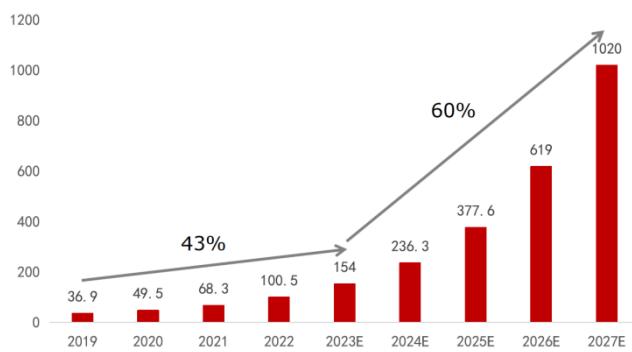


资料来源：科智咨询《中国液冷数据中心市场深度研究报告》，民生证券研究院

近年来，中国液冷数据中心市场快速发展，市场规模增速**超过 40%**；预计

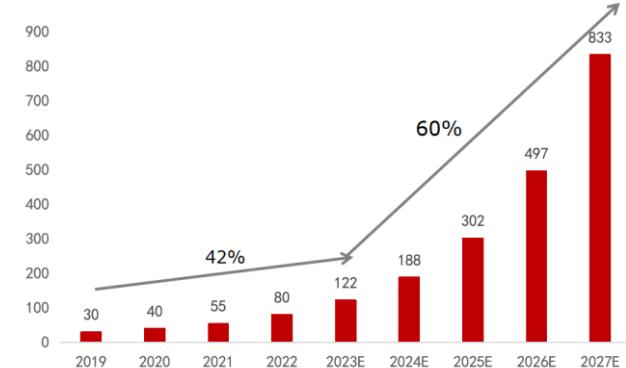
到 2027 年，液冷市场将以 60% 的复合增长率，突破千亿规模。

图41：2019-2027 年中国液冷数据中心规模（亿元）



资料来源：科智咨询《中国液冷数据中心市场深度研究报告》，民生证券研究院

图42：2019-2027 年中国液冷服务器市场规模（亿元）



资料来源：科智咨询《中国液冷数据中心市场深度研究报告》，民生证券研究院

#### 4.2.1 冷板式为短期内主流，重视 CDU、快接头等产业链相关环节

目前数据中心液冷技术主要分为冷板式液冷、浸没式液冷及喷淋式液冷三种路线。

**冷板式液冷**是通过冷板将发热元器件的热量间接传递给封闭在循环管路中的冷却液体，通过冷却液体将热量带走的一种实现方式。冷板式液冷具有**极高的散热性能、更好的材料兼容性和简便的改装方式**的优势，目前行业成熟度最高。

**浸没式液冷**是最近几年备受业界关注的新型散热技术：它主要采用**特定的冷却液**作为散热介质，将 IT 设备直接浸没在冷却液中，通过冷却循环带走 IT 设备运行过程中产生的热量。同时，冷却液通过循环过程与外部冷源进行热交换，将热量释放到环境中去。由于架构特殊，浸没式液冷具有**散热效率高、运行温度变化小、支持更高功率 IT 部署**的优势。

图43：液冷系统通用架构原理图

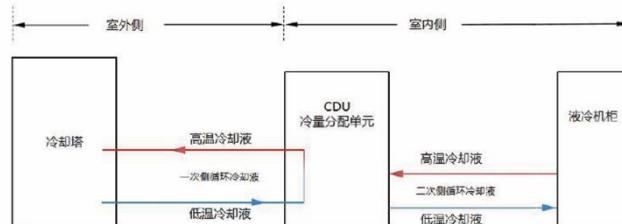
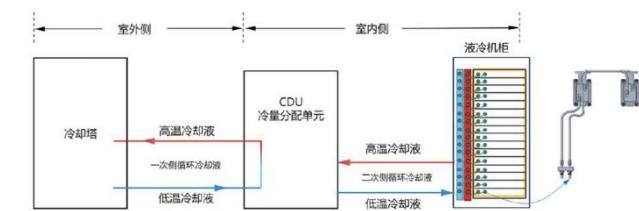


图44：冷板式液冷系统原理图



资料来源：《中兴通讯液冷技术白皮书》，民生证券研究院

资料来源：《中兴通讯液冷技术白皮书》，民生证券研究院

**喷淋式液冷**通过在服务器内部部署喷淋模块，使用一种对人、IT 设备和环境

无害无腐蚀的绝缘液体，有针对性的对发热器件进行喷淋降温的解决方案。喷淋液冷具有**器件集成度高、散热效率强、高效节能**和**静音**的特点，但是需要保证冷却液按需分配，运行维护复杂，**应用局限**。

图45：单相浸没液冷系统原理图

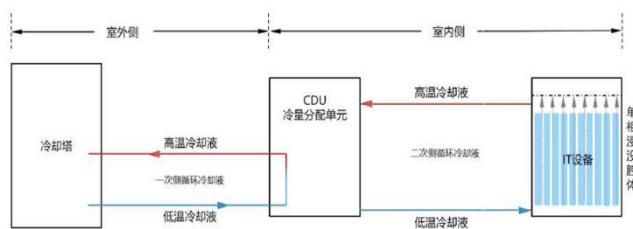
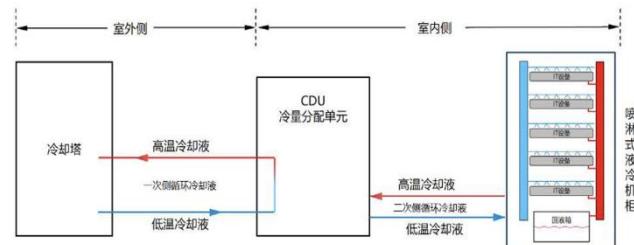


图46：喷淋式液冷系统原理图



资料来源：《中兴通讯液冷技术白皮书》，民生证券研究院

资料来源：《中兴通讯液冷技术白皮书》，民生证券研究院

相比冷板式液冷，浸没式和喷淋式液冷可完全去除散热风扇，散热及节能效果更好，数据中心**PUE 值可降至 1.1 及以下**，但相对对服务器机柜及机房配套设施的投入及改造成本更高，运维难度更大。

表13：冷板式、浸没式、喷淋式液冷技术对比

比较项目	冷板式	浸没式	喷淋式
成本	冷板要求的规格多，大多需要单独定制，成本较高	冷却液用量较多，与冷板式相比成本居中	通过改造旧式服务器和机柜，增加必须的装置，成本较小
可维护性	较简单	复杂	复杂
PUE	1.1-1.2	小于 1.09	小于 1.1
兼容性	未与主板和芯片模块进行直接的接触，材料兼容性较强	直接接触，材料兼容性较差	直接接触，材料兼容性较差
安装便捷程度	不改变服务器主板原有的形态，保留现有服务器主板，安装便捷	改变服务器主板原有结构，需重新安装	不改变服务器主板原有的形态，安装便捷
可循环	采用双路环状循环对冷冻液实现二次利用，降低运营成本	通过室外冷却装置进行循环，降低运营成本	采用循环泵，实现资源的再利用，降低运营成本

资料来源：兰洋科技官网，《联想液冷服务器白皮书》、民生证券研究院

## 4.2.2 相关公司

**英维克**：2022 年，公司紧跟 AI 等新应用场景带来的液冷技术应用需求增长的机遇，持续推进产品研发，23 年 10 月，为高效解决 5G 基站设备散热难题，满足 5G 应用加速规模化落地的需要，公司全新推出 3D-TVC 零功耗相变液冷方案。公司于 24 年 8 月推出 Coolinside 全链条液冷解决方案，包括冷板式液冷方案和浸没式液冷方案，能够根据不同应用场景灵活配置，对应于大型、中型、小型等各类数据中心应用场景，截至 2023 年 3 月，公司已经累计交付超过 500MW 液冷项目。

**申菱环境**：公司是全球领先的数据中心温控散热解决方案服务商和专业特种空调系统解决方案服务商，布局**数据中心、特种工业、轨道交通**等多个领域。历年

来公司数据服务产品已广泛服务于**华为、腾讯、字节、秦淮数据、浪潮、润泽科技、超聚变等众多知名客户**，并作为联合发布厂商与秦淮共同发布“智算中心全栈解决方案 2.0”。

公司深度参与**液冷技术的研究与应用**。推出基于智算、超算等多种数据中心发展需求的**风液同源大液冷系统**。**数据服务板块占比提升，液冷产品收入成倍增长**。截至 2024 上半年，公司数据服务板块表现突出，**营收同比增长约 61.35%**，新增订单接近 2023 年全年总量，尤其在高能效创新产品方面表现显著，蒸发冷却产品营收为去年同期的**3.6 倍**，液冷产品为**6.5 倍**。

**高澜股份**：公司在 IDC 液冷领域已经拥有**冷板式和浸没式液冷服务器热管理解决方案**、集装箱液冷数据中心解决方案，业务涵盖了从液冷板、CDU、Tank、集装箱等部件到数据中心设计、设备集成、系统调试、设备运维的系统集成。子公司高澜创新目前已积累了**字节跳动、阿里巴巴、腾讯、浪潮信息、万国数据等优质客户**，且在境外北美、欧洲等地区成功部署了热管理产品。

**溯联股份**：公司研发、生产的新能源汽车冷却管路及零部件专门针对新能源汽车液冷技术路线，截至 2022 年末，公司相关产品已广泛布局**30 多家**新能源产业链客户，遍布**整车、动力电池、热管理系统、储能**等领域，形成了较为完善的新能

源客户结构。

公司的储能液冷产品已于 2022 年开始量产。公司的流体管路及注塑零部件产品大量应用在新能源汽车热管理的液冷领域，是公司**增长最快的业务板块之一**。目前公司在电池储能方面的在手项目较多，公司正积极往**液冷充电桩和服务器**等领域拓展相关业务。**数据中心液冷方面**，管路系统主要采用 FEP 或聚四氟乙烯等氟塑料管路，**亦配置快速连接器**，公司具备相关产品能力，当前正完善相关工艺和技术储备。

**川环科技**：公司自主研发的用于液冷领域的**高性能液冷管路系统**，可以**取代现有市场普遍使用的聚四氟乙烯结构**，该液冷管路系统在耐高温、耐腐蚀和耐老化方面表现优异，相较于聚四氟乙烯结构，它能更好地应对服务器工作过程中产生的高温，提高冷却效果，能够更好地抵御冷却液体中可能存在的腐蚀物质，延长冷却管路的使用寿命，减少液冷管路的更换频率，从而降低维护成本，有助于**提高数据中心的冷却效率和稳定性**，满足不断增长的数据处理需求。

## 5 行业投资建议

国产大模型持续突破，DeepSeek、豆包、MCP等开始引领全球AI产业发展，AI应用落地加速。芯片厂商也在加速适配国产算力生态。以自主可控为基石，云厂商资本开支持续加速。带动服务器、电源、存储等上游产业链环节机遇。**建议关注：1) 算力芯片：中芯国际、海光信息、寒武纪、芯原股份、云天励飞、翱捷科技、沐曦技术（未上市）；2) 服务器：华勤技术、联想集团、浪潮技术、弘信电子；3) 电源：禾望电气、麦格米特、中恒电气、泰嘉股份、科泰电源、潍柴重机、科华数据、欧陆通、奥海科技；4) 温控：申菱环境、英维克、溯联股份、川环科技；5) 其他供应链：德明利、长光华芯、杰华特、华海诚科、华丰科技、兴森科技等。**

表14：重点关注个股

证券代码	证券简称	股价(元)	EPS				PE				评级
			2024A	2025E	2026E	2027E	2024A	2025E	2026E	2027E	
688981.SH	中芯国际	89.90	0.46	0.64	0.84	1.00	194	140	107	90	推荐
688041.SH	海光信息	150.01	0.83	1.32	1.91	2.64	181	114	79	57	推荐
688521.SH	芯原股份	91.79	-1.2	-0.18	0.07	0.61	/	/	1311	150	推荐
603063.SH	禾望电气	31.90	1.00	1.37	1.66	2.01	33	23	19	16	推荐
301018.SZ	申菱环境	40.60	0.43	1.03	1.44	1.80	93	39	28	23	推荐
301397.SZ	溯联股份	36.95	1.03	1.60	2.36	3.19	36	23	16	12	推荐
688048.SH	长光华芯	60.54	-0.57	0.02	0.41	0.66	/	3301	149	92	推荐
688343.SH	云天励飞	56.81	-1.63	/	/	/	/	/	/	/	/

资料来源：Wind，民生证券研究院预测

注：股价为 2025 年 5 月 6 日收盘价；未覆盖公司数据采用 Wind 一致预期。

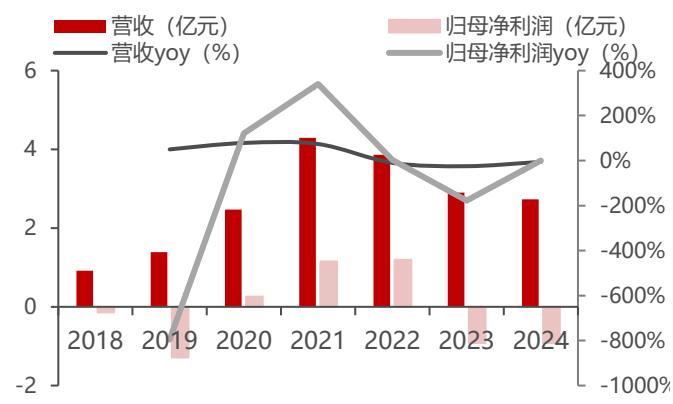
### 5.1 长光华芯：国产光通信芯片新进者

**深耕高功率激光芯片十载，国产光通信新进者。**公司 2012 年成立之初，便建成包括芯片设计、封装测试、光纤耦合等工艺产线；2018 年成立 VCSEL 事业部与激光系统事业部，产线和事业部构建完善。目前建立了半导体激光芯片 GaAs (砷化镓) 和 InP (磷化铟) 两大材料体系，构建了先进的边发射和面发射两大芯片结构工艺技术和制造平台，公司产品中高功率半导体激光芯片和光通信芯片属于边发射激光芯片，而 VCSEL 芯片属于面发射激光芯片。下游应用场景涉及工业、3D 传感、激光雷达、数据通信等。实现快速从 1 到 N 的产品矩阵扩展。

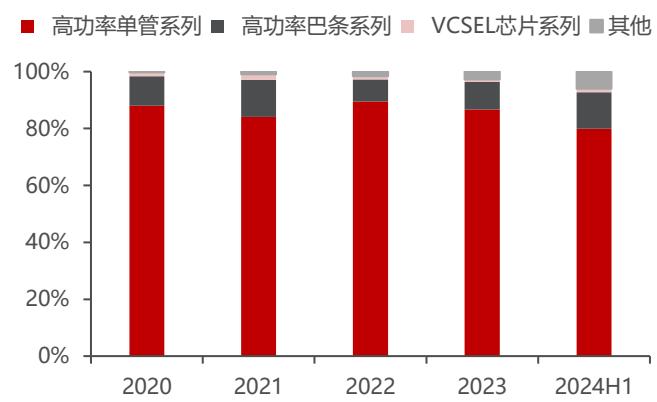
**图47：长光华芯产品系列布局**


资料来源：长光华芯招股书，民生证券研究院

2019-2022 年，随着公司收入规模的快速增长，以及公司 18W 及以上单管芯片等高端产品放量带动盈利能力提升，公司在 2020 年迅速扭亏为盈，并于 2021 年达到 1.15 亿元净利润，同比增长 340.49%。2023 年由于工业市场的持续疲软，公司利润端承压，归母净利润实现亏损 0.92 亿元。2024 年公司亏损与 23 年同比基本持平，后续新能源汽车、光通信等新型消费电子领域不断增长的需求，公司营收及利润有望迎来修复。

**图48：2018-2024 年长光华芯营收和利润情况**


资料来源：ifind，民生证券研究院。

**图49：2020-2024H1 长光华芯营收结构**


资料来源：ifind，民生证券研究院。

### 分业务收入预测：

**高功率单管系列业务：**公司作为高功率半导体激光芯片领域头部企业，得益于

智能设备、消费电子、新能源等领域对激光器需求的持续上升，以及新兴应用领域增量市场的不断开拓，公司 2022 年之前维持高速发展时期。23 年受宏观经济环境等因素的影响，行业竞争加剧，公司于 23 年初对价格进行调整，使得主业的毛利水平下降。根据公司数据，2024 年上半年工业市场缓慢恢复，并在未来几年维持此态势。工业市场面向激光制造和材料加工领域，下游市场增长量进入相对平稳时期；此外，公司在光通讯的技术更新迅速，200G EML 芯片研发进展良好，预计 26 年通过小批量交付和样品验证，在数据中心和高速网络的市场份额有望增长。我们预计 2025-2027 年高功率单管系列业务营收增速为-8.0%/5.0%/5.0%。毛利率方面，后续随成本端压力缓解，历史存货减值逐步消化完成，公司收入+利润有望逐步回暖到前期水平，预计 2024-2026 年毛利率分别为 29.0%/35.0%/35.0%。

**高功率巴条系列业务：**公司巴条器件产品主要使用方为客户 A2，产品最终用于其科研，整体销量受客户 A2 具体科研项目进度影响，进而过去几年该板块营收一直保持波动态势。24 年上半年，特殊领域的出货主要来自于 Q2 的贡献，Q3 及 Q4 继续交付现有订单。根据公司数据，2024 年上半年为订单旺季，下半年需求趋于平稳，预计全年增速放缓，特殊领域在 2025 年市场需求将逐步进入放量阶段，多家客户订单需求旺盛，行业目前仍处于高速增长阶段，预计公司该板块业务会有较好的营收增长。故我们预计 2025-2027 年高功率巴条系列业务营收增速为 160.0%/90%/50%。由于该产品主要用于科研攻关，更关注产品性能表现，客户对产品价格不敏感，预计批量量产出货后整体毛利率稳中有升，预计 25-27 年毛利率为 65.0%/68.0%/70.0%。

**VCSEL 芯片系列业务：**得益于智能驾驶赛道突飞猛进对激光雷达市场的需求开拓及 3D 传感器技术的不断迭代升级，作为其中核心零部件的 VcSEL 将迎来广阔市场空间。根据公司数据，在高性能车载激光雷达芯片，其 24Q4 已取得部分成绩，25 年会进入量产阶段；公司还在其他场景中进行技术储备，例如机器人的“眼睛”、消费电子领域等，后续市场需求有望持续增加。此外，公司在光通讯领域的 200G VCSEL 芯片研发进展良好。故我们预计 2025-2027 年 VCSEL 芯片系列业务营收增速为 40.0%/40.0%/40.0%。公司产品高端化后产品毛利率趋势平稳，预计 24-26 年毛利率为 50.0%/48.0%/47.0%。

**表15：长光华芯分业务收入预测（百万元）**

业务	项目名称	2022A	2023A	2024A	2025E	2026E	2027E
高功率单管系 列	销售额 (百万元)	345.24	251.70	213.45	230.53	242.05	254.15
	YOY	-4.36%	-27.09%	-15.20%	8.00%	5.00%	5.00%
	毛利率	48.80%	29.66%	10.93%	29.00%	35.00%	35.00%
高功率巴条系 列	毛利润	168.48	74.65	23.33	66.85	84.72	88.95
	销售额 (百万元)	29.98	28.08	36.84	95.78	181.99	272.98
	YOY	-46.14%	-6.34%	31.20%	160.00%	90.00%	50.00%
	毛利率	79.92%	65.19%	64.07%	65.00%	68.00%	70.00%
	毛利润	23.96	18.31	23.60	62.26	123.75	191.09

	销售额 (百万元)	3.94	2.47	3.62	5.07	7.10	9.93
VCSEL芯片系	YOY	-51.95%	-37.31%	46.56%	40.00%	40.00%	40.00%
列	毛利率	26.75%	11.42%	50.31%	50.00%	48.00%	47.00%
	毛利润	1.05	0.28	1.82	2.53	3.41	4.67
	销售额 (百万元)	6.44	7.96	18.72	21.53	24.76	28.47
其他业务	YOY	51.21%	23.59%	135.08%	15.00%	15.00%	15.00%
	毛利率	83.04%	51.48%	86.88%	50.00%	50.00%	50.00%
	毛利润	5.35	4.10	16.26	10.76	12.38	14.24
	销售额 (百万元)	385.60	290.21	272.63	352.91	455.89	565.54
合计	YOY	8.69%	-11.45%	-6.06%	29.45%	29.18%	24.05%
	毛利率	51.57%	33.54%	23.85%	40.35%	49.19%	52.86%
	毛利润	198.84	97.34	65.02	142.41	224.26	298.95

资料来: Wind, 民生证券研究院预测

**投资建议:** 我们预计公司 2025-2027 年将实现营收 3.53/4.56/5.66 亿元, 归母净利润 0.03/0.72/1.17 亿元, 对应现价 2025-2027 年 PS 为 30/23/19 倍, 25 年基本实现盈亏平衡、26 年实现盈利。我们看好公司产品结构改善, 激光雷达+光通信+特殊领域进入放量阶段, 首次覆盖, 给予“推荐”评级。

**风险提示:** 产品创新不及预期; 上游原料价格波动; 客户拓展不及预期

**表16: 长光华芯盈利预测与财务指标**

项目/年度	2024A	2025E	2026E	2027E
营业收入 (百万元)	273	353	456	566
增长率 (%)	-6.1	29.4	29.2	24.1
归属母公司股东净利润 (百万元)	-100	3	72	117
增长率 (%)	-8.5	103.2	2121.7	63.1
每股收益 (元)	-0.57	0.02	0.41	0.66
PE	/	3301	149	91
PS	39.1	30.2	23.4	18.9

资料来源: Wind, 民生证券研究院预测; (注: 股价为 2025 年 5 月 6 日收盘价)

## 5.2 禾望电气: 新能源与电气传动领跑者

**国内外市场双轮驱动, 新能源电控业务领跑市场。** 禾望电气是一家致力于新能源、电气传动和电源领域的高新技术企业, 专注于产品研发、生产、销售及服务。公司自成立以来, 一直秉承“科技创新, 人才强企”的理念, 通过持续的技术创新和优质服务, 不断为客户创造价值。禾望电气在风电、光伏、储能、氢能及电气传动等领域取得了显著成就, 成为国内新能源及电气传动领域最具竞争力的企业之一。在新能源板块, 公司凭借在风力发电产品及控制系统的关键技术突破, 全方位支持陆上风电、海上风电、集中式光储、分布式光储及氢能产业, 双馈变流器、全功率变流器、风电变桨系统在市场上表现优异, 有力推动了新能源行业的高质量、

规模化发展。在电气传动领域,禾望电气持续深化细分市场,助力高端装备制造业转型升级,进一步稳固了公司在行业内的领先地位。

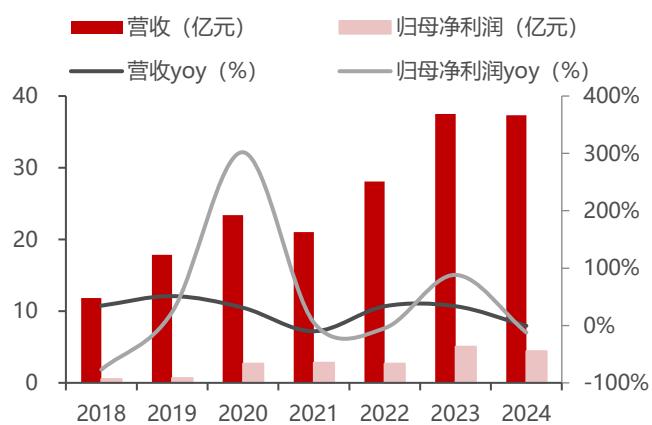
图50: 禾望电气产品系列布局



资料来源: 禾望电气官网, 民生证券研究院

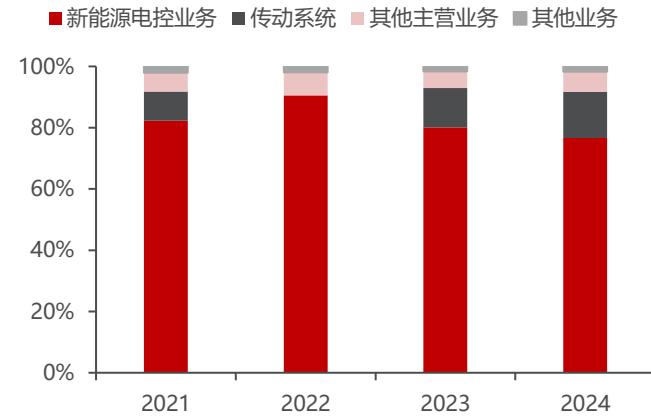
2018-2024年,禾望电气营收与利润呈现“阶梯式增长—短期承压”特征。营业收入从2018年的11.81亿元增长至2023年的37.52亿元,新能源电控业务的快速放量是核心驱动力,尤其在2021年后加速扩张,2024年营收同比微降0.5%至37.33亿元。归母净利润波动较大,2018-2021年处于低位,缓慢爬升,2022年受风电抢装潮退坡影响利润同比下滑4.7%至2.67亿元,2023年随光伏、储能业务放量利润翻倍至5.02亿元(同比+88.2%),创历史新高。但2024年利润回落至4.41亿元(同比-12.3%),原因可能在于新能源电控业务收入下滑4.79%,同时费用端压力增大,拖累盈利水平。

图51: 2018-2024年禾望电气营收和利润情况



资料来源: ifind, 民生证券研究院

图52: 2021-2024年禾望电气营收结构



资料来源: ifind, 民生证券研究院

**投资建议:** 我们预计公司2025-2027年将实现营收47.35/56.90/68.42亿元,归

母净利润 6.19/7.47/9.04 亿元, 对应现价 2025-2027 年 PE 为 23/19/16 倍, 在国内外市场双轮驱动下, 新能源电控业务稳健增长, 同时公司数据中心业务有望快速放量, 我们看好公司业务前景, 维持“推荐”评级。

**风险提示:** 行业需求波动; 海外市场拓展受阻的风险; 产品创新不及预期的风险。

**表17: 禾望电气盈利预测与财务指标**

项目/年度	2024A	2025E	2026E	2027E
营业收入 (百万元)	3,733	4,735	5,690	6,842
增长率 (%)	-0.5	26.8	20.2	20.3
归属母公司股东净利润 (百万元)	441	619	747	904
增长率 (%)	-12.3	40.6	20.7	21.0
每股收益 (元)	0.98	1.37	1.66	2.01
PE	33	23	19	16
PB	3.4	3.0	2.6	2.3

资料来源: Wind, 民生证券研究院预测; (注: 股价为 2025 年 5 月 6 日收盘价)

## 5.3 申菱环境: 数据中心温控核心供应商

公司 2000 年成立, 于 2021 年上市, 主要围绕空气环境调节设备开展业务、致力于液冷和风冷技术, 聚焦数据中心和电力能源包括核电相关的下游。公司主要产品类别包括数据服务、工业工艺、专业特种、高端公建等。在数据服务板块, 受益于 AI 技术, 公司数据中心与配套产品不断升级迭代, 提供更高能效产品; 在工业板块, 公司重点聚焦国内新能源产业链等优势产业海外设厂等业务; 在特种板块, 公司重点聚焦飞机地面空调等高技术含量设备的业务拓展, 机场、核电站、医院将是特种板块重要增量来源。

**图53: 申菱环境主要产品布局**

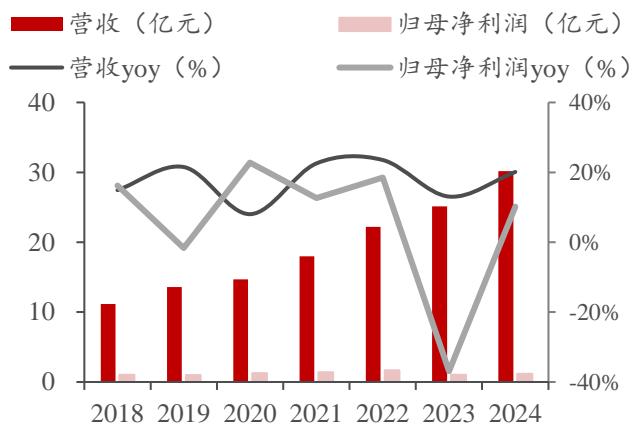


资料来源: 申菱环境官网, 民生证券研究院

公司 2024 年实现营收 30.16 亿元，同比增长 20.11%，受益于与华为、运营商、互联网企业及第三方数据中心运营商的业务拓展，公司 2018-2024 年营收实现了较快增长，期间 CAGR 达 17.97%。公司的净利润在 2022 年达到高点 1.66 亿元，随后于 2023 年同比下降了 36.9%，主要由于 2023 年公司加大了研发、扩能市场的投入，此外公司重建智能温控设备智能制造项目损失了约 2200 万元，对公司的净利润也造成了影响。2024 年公司实现归母净利润 1.16 亿元，同比增长 10.17%，重启正增长。

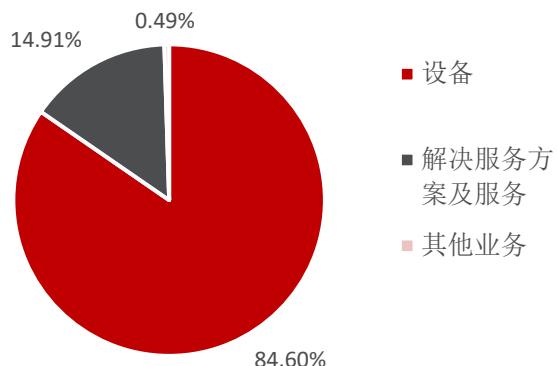
设备营收占公司 2024 年营收的 84.60%，解决服务方案及服务的营收占总营收的 14.91%，而其他业务营收占 2024 总营收的 0.49%。分下游应用看，24 数据服务板块、特种板块、工业板块、公建及商用行业、其他行业占公司营收比重分别为 51.31%/23.33%/20.26%/4.61%/0.49%。未来能源和数字化是公司重点业务方向，预计在将来几年数据中心需求仍会保持较好的增长，数据中心和核电的两个新基地将陆续投产。当前液冷业务呈现快速增长态势，预期未来液冷占比会持续提升并成为智算中心主要散热手段，公司将重点加大在液冷业务的研发和产能投入。

图54：2018-2024 年申菱环境营收和利润情况



资料来源：ifind，民生证券研究院

图55：2024 年申菱环境营收结构



资料来源：ifind，民生证券研究院

### 业务收入预测：

**设备：**公司是全球领先的数据中心温控散热解决方案服务商和专业特种空调系统解决方案服务商，布局**数据中心、特种工业、轨道交通**等多个领域。历年来公司数据服务产品已广泛服务于华为、腾讯、字节、秦淮数据、浪潮、润泽科技、超聚变等众多知名客户，并作为联合发布厂商与秦淮共同发布“智算中心全栈解决方案 2.0”。公司深度参与**液冷技术的研究与应用**。推出基于智算、超算等多种数据中心发展需求的**风液同源大液冷系统**。2024 年，公司数据服务板块表现突出，**营收同比增长约 75.40%**，新增订单同比增长约 95%，尤其在高能效创新产品方面表现显著，蒸发冷却产品营收为去年同期的**2.3 倍**，液冷产品为**2.9 倍**。

云厂商资本开支高增。腾讯 2024 年度资本开支达 767 亿美元，环比 221%，

**主要用于 GPU 采购，以支撑 AI 训练与推理需求；阿里巴巴宣布 2024 年第四季度资本开支达 318 亿元，环比达 80%，并进一步加大对 AI 领域的战略投入，在未来三年投入超过 3800 亿元人民币，超过过去十年总和。公司设备业务主要客户为云厂商，受益于腾讯、阿里等云厂商资本开支上修，公司应用于温控领域的空调、液冷等设备预计放量，我们预计 2025-2027 年公司设备业务营收增速为 33.0/28.0/20.0%。毛利率方面，随着液冷相关的 CDU、管路等产品放量，规模效应逐渐体现，该业务毛利率预计将持续上行，预计 2025-2027 年该业务毛利率分别为 27.0/28.0/28.5%。**

**解决方案及增值服务业务、其他业务：**公司其他业务主要包括非设备类的运维服务以及废料销售等，预计在 2025-2027 年受主业拉动呈现小幅增长，我们预计 2025-2027 年公司其他业务和解决方案及增值服务业务的增速为 10%/10%/10%。预计这两个业务的毛利率保持相对稳定，参考 2024 年，预计 2025-2027 年该业务毛利率分别为 32.0/32.0/32.0%。

**表18：申菱环境分业务收入预测 (百万元)**

业务	项目名称	2023A	2024A	2025E	2026E	2027E
设备	营收 (百万元)	2201.43	2551.73	3393.80	4344.06	5212.87
	YOY	21.50%	15.91%	33.00%	28.00%	20.00%
	毛利率	27.03%	22.93%	27.00%	28.00%	28.50%
	毛利润	594.96	585.22	916.33	1216.34	1485.67
解决方案及增值服务业务、其他业务	营收 (百万元)	309.76	464.46	510.90	561.99	618.19
	YOY	-24.31%	49.94%	10.00%	10.00%	10.00%
	毛利率	32.03%	32.00%	32.00%	32.00%	32.00%
	毛利润	99.22	132.65	163.49	179.84	197.82
合计	营收 (百万元)	2,511.19	3,016.18	3,904.70	4,906.05	5,831.07
	YOY	13.06%	20.11%	29.46%	25.64%	18.85%
	毛利率	27.64%	23.80%	27.65%	28.46%	28.87%
	毛利润	694.18	717.87	1,079.81	1,396.17	1,683.49

资料来：Wind，民生证券研究院预测

**投资建议：**我们预计公司 2025-2027 年将实现营收 39.05/49.06/58.31 亿元，归母净利润 2.74/3.84/4.79 亿元，对应现价 2025-2027 年 PE 为 39/28/23 倍。我们看好公司液冷产品在数据中心客户中的导入及放量，首次覆盖，给予“推荐”评级。

**风险提示：**AI 数据中心需求不及预期；行业竞争加剧；客户拓展不及预期。

**表19：申菱环境盈利预测与财务指标**

项目/年度	2024A	2025E	2026E	2027E
营业收入 (百万元)	3,016	3,905	4,906	5,831
增长率 (%)	20.1	29.5	25.6	18.9
归属母公司股东净利润 (百万元)	116	274	384	479
增长率 (%)	10.2	137.2	40.1	24.7
每股收益 (元)	0.43	1.03	1.44	1.80
PE	93	39	28	23
PB	4.3	3.9	3.5	3.2

资料来源：Wind，民生证券研究院预测；(注：股价为 2025 年 5 月 6 日收盘价)

## 5.4 液联股份：深耕快接头及管路，开拓数据中心液冷

公司是国内领先的老牌汽车用塑料流体管路产品及其零部件的供应商，公司的主营产品为塑料流体管路，具备前期技术沉淀与先发优势，自制率高；公司产品可细分为管路、真空管、阀体、快接插头、塑料管夹，其具有耐高温、耐高压、耐腐蚀、防渗透、密封性好等特点，可同时适用于燃油车和新能源车。2003 年，公司首推车用尼龙管快速接头并获专利；2018 年，公司研发的首款动力电池包专用水管接头，填补国内细分市场空白。公司在连接件、阀类、传感器等流体控制件方面的定制化优势，缩短了客户开发周期，契合了新能源汽车快速迭代需求。

**图56：液联股份主要产品布局**

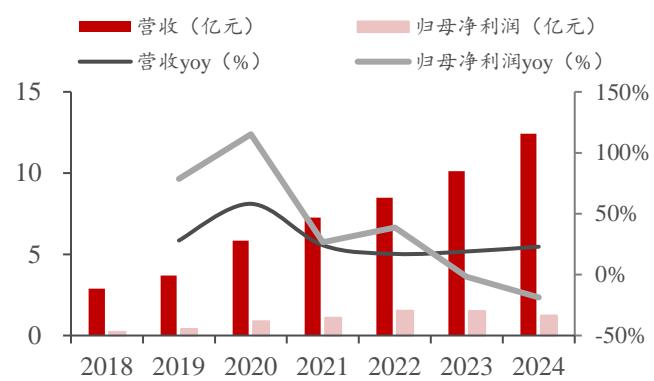

资料来源：液联股份官网，民生证券研究院

公司 2024 年实现营收 12.42 亿元，同比增长 22.89%，主要受益于汽车客户增加、新产品持续上量，2024 年公司新增 1000 余项产品定点，结合前期在开发项目共 2000 余项；2018-2024 年公司的营收 CAGR 达到 27.5%。22 年公司归母净利润达到高点 1.53 亿元，随后连续两年下滑至 24 年的 1.23 亿元，主要归因

于产品降价，以及期间费用有所升高。

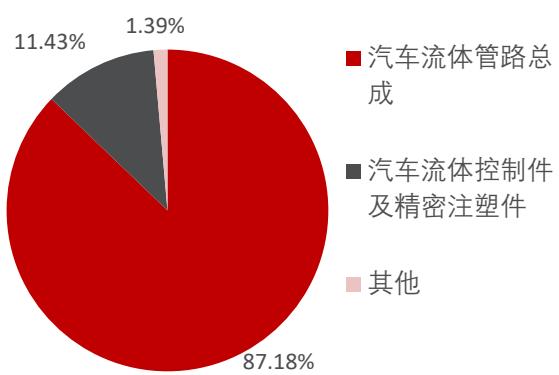
公司 2024 年实现营收 12.42 亿元，87.18% 的营收占比为汽车流体管路及总成，11.43% 的营收占比为汽车流体控制及精密注塑件，剩下 1.39% 的营收由其他产品构成。公司未来将持续向数据中心液冷领域拓展业务：数据中心液冷管路系统主要采用 FEP 或聚四氟乙烯等氟塑料管路，亦配置快速连接器，公司按照 OCP(开放计算项目)技术规范和标准开发了应用于数据中心的相关产品，**其中 UQD (应用于数据中心液冷的通用快速断开接头) 产品已经开发成功**。2024 年 11 月 22 日公司成立子公司溯联智控，独立发展储能及算力热管理液冷产品业务，积极为公司开拓第二增长曲线。

**图57：2018-2024 年溯联股份营收和利润情况**



资料来源：ifind，民生证券研究院

**图58：2024 年溯联股份营收结构**



资料来源：ifind，民生证券研究院

#### 分业务收入预测：

**汽车流体管路总成业务：**公司研发、生产的新能源汽车冷却管路及零部件专门针对新能源汽车液冷技术路线，形成了较为完善的新能源客户结构，2024 年公司新能源车型收入约 7.61 亿元，同比增长超过 75%，成为公司业务收入重要支柱，实现了燃油+电动车业务的“双轮驱动”。公司的流体管路及注塑零部件产品大量应用在新能源汽车热管理的液冷领域，**并向储能领域液冷热管理延伸**，应用于储能液冷产品已于 2022 年开始量产，储能业务客户既包括国内领先的电池制造企业，也包括知名的储能系统集成商。**数据中心液冷方面，管路系统主要采用 FEP 或聚四氟乙烯等氟塑料管路，亦配置快速连接器**，公司具备相关产品能力，当前正完善相关工艺和技术储备。随着 AI 芯片功率提高，液冷管路需求迅速增长，溯联作为汽车尼龙管路龙头，车规级产品防震动、耐久、插拔、可靠性等方面优于工规级产品，产品有望实现向数据中心液冷横向拓展，我们预计 2025-2027 年汽车流体管路总成业务营收增速为 22.0/25.0/20.0%。毛利率方面，数据中心产品竞争激烈度相比汽车产品较小，毛利率有望持续增长，价格相对不敏感，预计 2025 年该业务毛利率将明显修复，2025-2027 年毛利率分别为 25.0/26.0/27.0%。

**汽车流体控制件及精密注塑件：**该业务主要包含尼龙快速接头、流体控制阀类、消音器、塑料紧固件及其他各类应用于机舱、车身、内饰、地盘等位置的覆盖见、

结构件。相比金属快接头采用 CNC 工艺，公司尼龙快接头采用注塑加工，生产效率高，成本低，且竞争格局相对较好，溯联股份有望凭借尼龙快接头导入数据中心液冷市场，25-26 年该业务有望保持加速增长态势。据 IDC 数据，**中国液冷服务器市场 2024 上半年同比大幅增长 98.3%，市场规模达到 12.6 亿美元，出货量同比增长 81.8%**；IDC 预计 2023-2028 年，中国液冷服务器年复合增长率将达 47.6%，市场规模有望在 2028 年达到 102 亿美元。预计公司液冷快接头产品将伴随 AI 服务器市场放量，2025 年处于导入及放量初期，并在 2026 年加速放量，并在 27 年保持较快增长。我们预计 2025-2027 年汽车流体控制件及精密注塑件业务营收增速为 50.0/100.0/50.0%。毛利率方面，数据中心产品竞争激烈度相比汽车产品较小，毛利率有望持续增长，价格相对不敏感，且尼龙快接头具备创新性，竞争格局较好，预计 2025-2027 年毛利率为 27.0%/28.0%/29.0%。

**其他：**公司其他业务主要包括工装模具、设备、技术服务及其他生产性物料的出售，为项目性质业务收入，波动较大，占公司整体营收比重较低；我们预计 2025-2027 年公司其他业务增速为 10%/10%/10%。预计该业务毛利率保持相对稳定，参考 2023 及 2024 年，预计 2025-2027 年该业务毛利率分别为 60.0/60.0/60.0%。

**表20：溯联股份分业务收入预测 (百万元)**

业务	项目名称	2023A	2024A	2025E	2026E	2027E
汽车流体管路	营收 (百万元)	880.67	1082.59	1320.76	1650.95	1981.14
	YOY	16.09%	22.93%	22.00%	25.00%	20.00%
	毛利率	27.95%	21.81%	25.00%	26.00%	27.00%
总成业务	毛利润	246.16	236.07	330.19	429.25	534.91
	营收 (百万元)	107.62	141.98	212.97	425.94	638.91
	YOY	58.15%	31.93%	50.00%	100.00%	50.00%
汽车流体控制件及精密注塑件	毛利率	22.64%	26.02%	27.00%	28.00%	29.00%
	毛利润	24.36	36.95	57.50	119.26	185.28
	营收 (百万元)	22.23	17.25	18.98	20.87	22.96
其他	YOY	-1.20%	-22.40%	10.00%	10.00%	10.00%
	毛利率	60.68%	63.21%	60.00%	60.00%	60.00%
	毛利润	13.49	10.90	11.39	12.52	13.78
合计	营收 (百万元)	1010.52	1241.82	1,552.70	2097.76	2643.01
	YOY	19.00%	22.89%	25.03%	35.10%	25.99%
	毛利率	28.11%	22.86%	25.70%	26.74%	27.77%
	毛利润	284.01	283.92	399.08	561.03	733.97

资料来：Wind，民生证券研究院预测

**投资建议：**我们预计公司 2025-2027 年将实现营收 15.53/20.98/26.43 亿元，归母净利润 1.92/2.84/3.82 亿元，对应现价 2025-2027 年 PE 为 23/16/12 倍。我们看好公司由汽车管路及快接头向数据中心横向拓展的能力，首次覆盖，给予“推荐”评级。

**风险提示：**液冷设计方案变化；行业竞争激烈；数据中心客户拓展不及预期

**表21：溯联股份盈利预测与财务指标**

项目/年度	2024A	2025E	2026E	2027E
营业收入 (百万元)	1,242	1,553	2,098	2,643
增长率 (%)	22.9	25.0	35.1	26.0
归属母公司股东净利润 (百万元)	123	192	284	382
增长率 (%)	-18.7	56.3	48.0	34.9
每股收益 (元)	1.02	1.60	2.36	3.19
PE	36	23	16	12
PB	2.3	2.2	2.0	1.9

资料来源：Wind，民生证券研究院预测；(注：股价为2025年5月6日收盘价)

## 6 风险提示

**1) AI 应用落地不及预期的风险。**技术成熟度不足可能导致 AI 模型在端侧应用等场景中表现不佳，尤其是在面对复杂、多变的环境时，模型的泛化能力和稳定性可能无法满足需求。此外，AI 应用场景与大模型的匹配度不高以及应用场景的复杂性，可能导致解决方案无法有效解决实际问题，进而导致 AI 应用落地不及预期。

**2) 云厂商资本开支不及预期的风险。**作为 AI 大模型训练和部署的核心基础设施提供者，云厂商的资本开支主要用于数据中心建设、算力扩容以及硬件升级，是支撑大模型研发和规模化应用的硬件基础。若资本开支不及预期，可能导致算力资源供给不足，限制大模型的训练规模和迭代速度，进而延缓技术突破和端侧等 AI 应用商业化落地。

**3) 宏观经济波动的风险。**受到全球宏观经济的波动、行业景气度等因素影响，集成电路行业存在一定的周期性，与宏观经济整体发展亦密切相关。如果宏观经济波动较大或长期处于低谷，集成电路行业的市场需求也将随之受到影响。另外，下游市场需求的波动和低迷亦会导致集成电路产品的需求下降，或由于半导体行业出现投资过热、重复建设的情况进而导致产能供应在景气度较低时超过市场需求，进而影响集成电路晶圆代工企业的盈利能力，将可能对行业内公司的经营业绩造成一定的影响。

**中芯国际财务报表数据预测汇总**

利润表 (百万元)	2024A	2025E	2026E	2027E
营业总收入	57,796	69,248	78,315	89,396
营业成本	47,051	55,633	61,955	70,567
营业税金及附加	276	291	316	358
销售费用	282	346	392	349
管理费用	3,835	4,501	5,012	5,453
研发费用	5,447	6,232	6,892	7,688
EBIT	3,238	5,702	7,813	9,533
财务费用	-1,833	-1,462	-1,344	-1,189
资产减值损失	-524	-1,113	-1,170	-1,333
投资收益	1,100	692	861	983
营业利润	6,299	6,743	8,849	10,399
营业外收支	-7	3	3	3
利润总额	6,292	6,746	8,852	10,402
所得税	919	337	434	458
净利润	5,373	6,409	8,418	9,945
归属于母公司净利润	3,699	5,127	6,692	8,015
EBITDA	26,394	32,467	41,260	48,435

资产负债表 (百万元)	2024A	2025E	2026E	2027E
货币资金	48,029	42,166	34,404	37,254
应收账款及票据	3,288	6,534	7,390	8,436
预付款项	405	1,113	1,239	1,411
存货	21,267	26,704	28,086	31,991
其他流动资产	33,289	33,377	33,420	33,472
流动资产合计	106,279	109,893	104,539	112,563
长期股权投资	9,004	9,004	9,004	9,004
固定资产	113,545	149,637	180,365	204,183
无形资产	3,225	3,225	3,225	3,225
非流动资产合计	247,137	267,137	287,137	297,137
资产合计	353,415	377,030	391,675	409,700
短期借款	1,070	1,070	1,070	1,070
应付账款及票据	5,658	7,727	8,605	9,801
其他流动负债	54,816	69,953	75,303	82,186
流动负债合计	61,544	78,750	84,978	93,058
长期借款	57,785	57,785	57,785	57,785
其他长期负债	4,978	4,978	4,978	4,978
非流动负债合计	62,763	62,763	62,763	62,763
负债合计	124,308	141,513	147,741	155,821
股本	226	226	226	226
少数股东权益	80,917	82,199	83,925	85,854
股东权益合计	229,108	235,517	243,935	253,879
负债和股东权益合计	353,415	377,030	391,675	409,700

资料来源：中芯国际公告、民生证券研究院预测

主要财务指标	2024A	2025E	2026E	2027E
<b>成长能力 (%)</b>				
营业收入增长率	27.72	19.82	13.09	14.15
EBIT 增长率	28.17	76.11	37.03	22.01
净利润增长率	-23.31	38.62	30.52	19.77
<b>盈利能力 (%)</b>				
毛利率	18.59	19.66	20.89	21.06
净利润率	6.40	7.40	8.55	8.97
总资产收益率 ROA	1.05	1.36	1.71	1.96
净资产收益率 ROE	2.50	3.34	4.18	4.77
<b>偿债能力</b>				
流动比率	1.73	1.40	1.23	1.21
速动比率	0.87	0.65	0.52	0.52
现金比率	0.78	0.54	0.40	0.40
资产负债率 (%)	35.17	37.53	37.72	38.03
<b>经营效率</b>				
应收账款周转天数	20.01	22.58	28.24	28.11
存货周转天数	155.49	155.21	159.18	153.24
总资产周转率	0.17	0.19	0.20	0.22
<b>每股指标 (元)</b>				
每股收益	0.46	0.64	0.84	1.00
每股净资产	18.56	19.20	20.04	21.04
每股经营现金流	2.84	5.47	6.05	6.79
每股股利	0.00	0.00	0.00	0.00
<b>估值分析</b>				
PE	190	137	105	88
PB	4.7	4.6	4.4	4.2
EV/EBITDA	27.96	22.73	17.89	15.24
股息收益率 (%)	0.00	0.00	0.00	0.00

现金流量表 (百万元)	2024A	2025E	2026E	2027E
净利润	5,373	6,409	8,418	9,945
折旧和摊销	23,156	26,765	33,447	38,902
营运资金变动	-4,129	6,609	2,643	1,565
经营活动现金流	22,659	43,705	48,320	54,238
资本开支	-54,347	-46,535	-53,217	-48,645
投资	21,401	0	0	0
投资活动现金流	-30,669	-45,843	-52,356	-47,662
股权募资	2,384	0	0	0
债务募资	10,986	0	0	0
筹资活动现金流	9,999	-3,726	-3,726	-3,726
现金净流量	1,732	-5,864	-7,762	2,850

**长光华芯财务报表数据预测汇总**

利润表 (百万元)	2024A	2025E	2026E	2027E
<b>营业总收入</b>	<b>273</b>	<b>353</b>	<b>456</b>	<b>566</b>
营业成本	208	210	232	267
营业税金及附加	4	3	4	5
销售费用	10	12	16	19
管理费用	41	48	55	62
研发费用	127	138	160	187
EBIT	<b>-151</b>	<b>-31</b>	35	83
财务费用	<b>-8</b>	<b>-10</b>	<b>-12</b>	<b>-15</b>
资产减值损失	<b>-49</b>	<b>-5</b>	<b>-5</b>	<b>-5</b>
投资收益	24	25	23	17
<b>营业利润</b>	<b>-133</b>	<b>-1</b>	<b>66</b>	<b>109</b>
营业外收支	2	4	5	6
<b>利润总额</b>	<b>-131</b>	<b>3</b>	<b>70</b>	<b>115</b>
所得税	<b>-33</b>	0	<b>-1</b>	<b>-2</b>
净利润	<b>-98</b>	3	72	117
<b>归属于母公司净利润</b>	<b>-100</b>	<b>3</b>	<b>72</b>	<b>117</b>
EBITDA	<b>-67</b>	59	137	194

资产负债表 (百万元)	2024A	2025E	2026E	2027E
货币资金	592	675	794	963
应收账款及票据	154	166	225	279
预付款项	7	11	12	13
存货	169	147	156	172
其他流动资产	712	495	297	197
<b>流动资产合计</b>	<b>1,633</b>	<b>1,494</b>	<b>1,483</b>	<b>1,625</b>
长期股权投资	267	267	267	267
固定资产	760	861	961	1,028
无形资产	15	15	15	15
<b>非流动资产合计</b>	<b>1,669</b>	<b>1,772</b>	<b>1,871</b>	<b>1,869</b>
<b>资产合计</b>	<b>3,302</b>	<b>3,266</b>	<b>3,354</b>	<b>3,494</b>
短期借款	31	31	31	31
应付账款及票据	182	140	151	170
其他流动负债	47	50	55	60
<b>流动负债合计</b>	<b>260</b>	<b>222</b>	<b>238</b>	<b>261</b>
长期借款	0	0	0	0
其他长期负债	50	49	49	49
<b>非流动负债合计</b>	<b>50</b>	<b>49</b>	<b>49</b>	<b>49</b>
<b>负债合计</b>	<b>310</b>	<b>270</b>	<b>286</b>	<b>310</b>
股本	176	176	176	176
少数股东权益	6	6	6	6
<b>股东权益合计</b>	<b>2,992</b>	<b>2,996</b>	<b>3,067</b>	<b>3,185</b>
<b>负债和股东权益合计</b>	<b>3,302</b>	<b>3,266</b>	<b>3,354</b>	<b>3,494</b>

资料来源：长光华芯公告、民生证券研究院预测

主要财务指标	2024A	2025E	2026E	2027E
<b>成长能力 (%)</b>				
营业收入增长率	-6.05	29.44	29.18	24.05
EBIT 增长率	-40.69	79.25	213.31	133.60
净利润增长率	-8.47	103.24	2121.71	63.09
<b>盈利能力 (%)</b>				
毛利率	23.85	40.35	49.19	52.86
净利润率	-36.58	0.92	15.76	20.71
总资产收益率 ROA	-3.02	0.10	2.14	3.35
净资产收益率 ROE	-3.34	0.11	2.35	3.69
<b>偿债能力</b>				
流动比率	6.28	6.74	6.24	6.22
速动比率	5.46	5.94	5.45	5.43
现金比率	2.27	3.04	3.34	3.69
资产负债率 (%)	9.38	8.28	8.54	8.87
<b>经营效率</b>				
应收账款周转天数	174.72	123.38	125.15	130.40
存货周转天数	319.69	270.57	235.84	221.75
总资产周转率	0.08	0.11	0.14	0.17
<b>每股指标 (元)</b>				
每股收益	-0.57	0.02	0.41	0.66
每股净资产	16.94	16.96	17.36	18.03
每股经营现金流	-0.38	0.13	0.53	0.90
每股股利	0.00	0.00	0.00	0.00
<b>估值分析</b>				
PE	/	3301	149	91
PB	3.3	3.6	3.5	3.4
EV/EBITDA	/	176.21	75.48	53.21
股息收益率 (%)	0.00	0.00	0.00	0.00
<b>现金流量表 (百万元)</b>				
<b>净利润</b>	<b>-98</b>	<b>3</b>	<b>72</b>	<b>117</b>
折旧和摊销	84	90	101	111
营运资金变动	<b>-78</b>	<b>-57</b>	<b>-69</b>	<b>-66</b>
<b>经营活动现金流</b>	<b>-66</b>	24	93	158
资本开支	<b>-79</b>	<b>-185</b>	<b>-195</b>	<b>-104</b>
投资	<b>-88</b>	200	200	100
<b>投资活动现金流</b>	<b>-141</b>	62	28	13
股权募资	0	0	0	0
债务募资	15	0	0	0
<b>筹资活动现金流</b>	<b>31</b>	<b>-2</b>	<b>-1</b>	<b>-1</b>
<b>现金净流量</b>	<b>-175</b>	83	119	170

**申菱环境财务报表数据预测汇总**

利润表 (百万元)	2024A	2025E	2026E	2027E
<b>营业总收入</b>	<b>3,016</b>	<b>3,905</b>	<b>4,906</b>	<b>5,831</b>
营业成本	2,298	2,825	3,510	4,148
营业税金及附加	20	27	34	41
销售费用	216	262	329	391
管理费用	159	207	260	309
研发费用	171	226	285	338
EBIT	131	335	461	572
财务费用	14	31	40	49
资产减值损失	-8	-11	-13	-16
投资收益	1	0	0	0
<b>营业利润</b>	<b>118</b>	<b>293</b>	<b>407</b>	<b>507</b>
营业外收支	-5	-5	-3	-3
<b>利润总额</b>	<b>113</b>	<b>288</b>	<b>404</b>	<b>504</b>
所得税	2	12	16	20
净利润	111	277	388	484
<b>归属于母公司净利润</b>	<b>116</b>	<b>274</b>	<b>384</b>	<b>479</b>
EBITDA	215	425	561	682

资产负债表 (百万元)	2024A	2025E	2026E	2027E
货币资金	507	466	455	528
应收账款及票据	1,756	2,169	2,726	3,239
预付款项	39	56	70	83
存货	863	1,049	1,303	1,540
其他流动资产	360	448	530	602
<b>流动资产合计</b>	<b>3,526</b>	<b>4,188</b>	<b>5,084</b>	<b>5,992</b>
长期股权投资	0	0	0	0
固定资产	1,176	1,405	1,608	1,793
无形资产	159	159	159	159
<b>非流动资产合计</b>	<b>1,834</b>	<b>2,057</b>	<b>2,207</b>	<b>2,357</b>
<b>资产合计</b>	<b>5,360</b>	<b>6,245</b>	<b>7,290</b>	<b>8,348</b>
短期借款	305	375	445	515
应付账款及票据	1,289	1,569	1,950	2,304
其他流动负债	660	815	973	1,115
<b>流动负债合计</b>	<b>2,255</b>	<b>2,760</b>	<b>3,368</b>	<b>3,934</b>
长期借款	349	499	649	799
其他长期负债	209	205	205	205
<b>非流动负债合计</b>	<b>558</b>	<b>704</b>	<b>854</b>	<b>1,004</b>
<b>负债合计</b>	<b>2,813</b>	<b>3,464</b>	<b>4,222</b>	<b>4,938</b>
股本	266	266	266	266
少数股东权益	13	16	20	25
<b>股东权益合计</b>	<b>2,547</b>	<b>2,781</b>	<b>3,068</b>	<b>3,410</b>
<b>负债和股东权益合计</b>	<b>5,360</b>	<b>6,245</b>	<b>7,290</b>	<b>8,348</b>

资料来源：申菱环境公告、民生证券研究院预测

主要财务指标	2024A	2025E	2026E	2027E
<b>成长能力 (%)</b>				
营业收入增长率	20.11	29.46	25.64	18.85
EBIT 增长率	-0.28	155.41	37.41	24.08
净利润增长率	10.17	137.21	40.10	24.67
<b>盈利能力 (%)</b>				
毛利率	23.80	27.65	28.46	28.87
净利润率	3.83	7.02	7.83	8.21
总资产收益率 ROA	2.16	4.39	5.27	5.74
净资产收益率 ROE	4.56	9.91	12.60	14.14
<b>偿债能力</b>				
流动比率	1.56	1.52	1.51	1.52
速动比率	1.06	1.00	0.98	0.99
现金比率	0.22	0.17	0.14	0.13
资产负债率 (%)	52.48	55.47	57.91	59.15
<b>经营效率</b>				
应收账款周转天数	180.97	175.74	174.20	178.61
存货周转天数	122.11	121.84	120.61	123.38
总资产周转率	0.60	0.67	0.72	0.75

每股指标 (元)	2024A	2025E	2026E	2027E
每股收益	0.43	1.03	1.44	1.80
每股净资产	9.52	10.39	11.46	12.73
每股经营现金流	0.51	0.47	0.56	1.14
每股股利	0.16	0.38	0.53	0.66
<b>估值分析</b>				
PE	86	36	26	23
PB	3.9	3.6	3.3	3.2
EV/EBITDA	48.47	24.52	18.56	16.53
股息收益率 (%)	0.43	1.02	1.42	1.63

现金流量表 (百万元)	2024A	2025E	2026E	2027E
净利润	111	277	388	484
折旧和摊销	84	90	101	110
营运资金变动	-136	-364	-483	-462
<b>经营活动现金流</b>	<b>135</b>	<b>125</b>	<b>149</b>	<b>302</b>
资本开支	-374	-286	-245	-255
投资	258	20	20	10
<b>投资活动现金流</b>	<b>-110</b>	<b>-294</b>	<b>-225</b>	<b>-245</b>
股权募资	0	0	0	0
债务募资	21	220	220	220
<b>筹资活动现金流</b>	<b>-50</b>	<b>128</b>	<b>64</b>	<b>15</b>
<b>现金净流量</b>	<b>-24</b>	<b>-41</b>	<b>-11</b>	<b>73</b>

**溯联股份财务报表数据预测汇总**

利润表 (百万元)	2024A	2025E	2026E	2027E
<b>营业总收入</b>	<b>1,242</b>	<b>1,553</b>	<b>2,098</b>	<b>2,643</b>
营业成本	958	1,154	1,537	1,909
营业税金及附加	6	9	13	16
销售费用	32	40	52	66
管理费用	60	71	94	119
研发费用	66	82	109	137
EBIT	123	202	302	408
财务费用	-7	1	4	5
资产减值损失	-5	-5	-6	-7
投资收益	13	16	21	26
<b>营业利润</b>	<b>138</b>	<b>213</b>	<b>315</b>	<b>425</b>
营业外收支	0	0	0	0
<b>利润总额</b>	<b>138</b>	<b>213</b>	<b>315</b>	<b>425</b>
所得税	16	21	32	42
净利润	123	192	284	382
<b>归属于母公司净利润</b>	<b>123</b>	<b>192</b>	<b>284</b>	<b>382</b>
EBITDA	189	274	386	505

资产负债表 (百万元)	2024A	2025E	2026E	2027E
货币资金	413	341	329	308
应收账款及票据	383	435	530	632
预付款项	3	6	8	10
存货	381	412	506	603
其他流动资产	587	545	589	626
<b>流动资产合计</b>	<b>1,767</b>	<b>1,738</b>	<b>1,962</b>	<b>2,179</b>
长期股权投资	0	0	0	0
固定资产	384	475	572	654
无形资产	98	118	138	158
<b>非流动资产合计</b>	<b>761</b>	<b>994</b>	<b>1,127</b>	<b>1,219</b>
<b>资产合计</b>	<b>2,528</b>	<b>2,732</b>	<b>3,089</b>	<b>3,398</b>
短期借款	134	182	237	237
应付账款及票据	386	433	576	716
其他流动负债	53	60	78	80
<b>流动负债合计</b>	<b>572</b>	<b>674</b>	<b>891</b>	<b>1,033</b>
长期借款	0	0	0	0
其他长期负债	7	13	18	23
<b>非流动负债合计</b>	<b>7</b>	<b>13</b>	<b>18</b>	<b>23</b>
<b>负债合计</b>	<b>579</b>	<b>687</b>	<b>909</b>	<b>1,056</b>
股本	120	120	120	120
少数股东权益	0	0	0	0
<b>股东权益合计</b>	<b>1,949</b>	<b>2,045</b>	<b>2,180</b>	<b>2,342</b>
<b>负债和股东权益合计</b>	<b>2,528</b>	<b>2,732</b>	<b>3,089</b>	<b>3,398</b>

资料来源：溯联股份公告、民生证券研究院预测

主要财务指标	2024A	2025E	2026E	2027E
<b>成长能力 (%)</b>				
营业收入增长率	22.89	25.03	35.10	25.99
EBIT 增长率	-23.72	64.57	49.70	35.24
净利润增长率	-18.72	56.31	48.05	34.89
<b>盈利能力 (%)</b>				
毛利率	22.86	25.70	26.74	27.77
净利润率	9.87	12.33	13.52	14.47
总资产收益率 ROA	4.85	7.01	9.18	11.25
净资产收益率 ROE	6.29	9.37	13.01	16.33
<b>偿债能力</b>				
流动比率	3.09	2.58	2.20	2.11
速动比率	2.04	1.55	1.21	1.08
现金比率	0.72	0.51	0.37	0.30
资产负债率 (%)	22.92	25.15	29.43	31.08
<b>经营效率</b>				
应收账款周转天数	89.43	93.16	80.20	76.50
存货周转天数	110.84	123.63	107.49	104.53
总资产周转率	0.51	0.59	0.72	0.81
<b>每股指标 (元)</b>				
每股收益	1.02	1.60	2.36	3.19
每股净资产	16.25	17.05	18.17	19.53
每股经营现金流	-0.82	1.38	1.90	2.61
每股股利	0.80	1.24	1.83	2.47
<b>估值分析</b>				
PE	35	23	15	11
PB	2.2	2.1	2.0	1.9
EV/EBITDA	22.75	15.71	11.14	8.53
股息收益率 (%)	2.21	3.42	5.07	6.83

现金流量表 (百万元)	2024A	2025E	2026E	2027E
净利润	123	192	284	382
折旧和摊销	67	73	84	96
营运资金变动	-288	-104	-142	-166
<b>经营活动现金流</b>	<b>-99</b>	<b>166</b>	<b>228</b>	<b>314</b>
资本开支	-220	-220	-210	-181
投资	-334	100	50	50
<b>投资活动现金流</b>	<b>-390</b>	<b>-184</b>	<b>-139</b>	<b>-105</b>
股权募资	0	0	0	0
债务募资	63	48	55	0
<b>筹资活动现金流</b>	<b>-76</b>	<b>-53</b>	<b>-101</b>	<b>-229</b>
<b>现金净流量</b>	<b>-565</b>	<b>-71</b>	<b>-13</b>	<b>-20</b>

## 插图目录

图 1: 豆包和 DeepSeek 突破后, 国产大模型的迭代时点	4
图 2: DeepSeek 发展历程	4
图 3: DeepSeekMoE 示意图	5
图 4: 不同注意力机制示意图	6
图 5: 不同注意力机制 KV 缓存和性能对比	6
图 6: MTP 技术示意图	6
图 7: GRPO 技术示意图	7
图 8: DeepSeek-V3 与其他大模型性能对比	7
图 9: DeepSeek-R1 与其他大模型性能对比	7
图 10: 豆包 Doubao-1.5-pro 与其他大模型性能对比	8
图 11: 通义千问 QwQ-32B 与其他大模型性能对比	8
图 12: DeepSeek-V3 与其他大模型性能/价格对比	9
图 13: DeepSeek-R1 与其他大模型输入输出价格对比	9
图 14: DeepSeek 大模型理论收入与成本	10
图 15: 各产品增长 1 亿用户所用时间 (天)	10
图 16: AI 原生 App 月活跃用户规模趋势 (万人)	11
图 17: 寒武纪基础软件系统平台	17
图 18: 海光完整软件栈系统	18
图 19: 云飞励天下游场景解决方案	20
图 20: 2022-2024 年全球 AI 芯片竞争格局	21
图 21: 腾讯 AI 投资展望	27
图 22: 2021Q1-2024Q4 阿里巴巴资本开支	27
图 23: 1Q25 DeepSeek 日活用户规模趋势	29
图 24: ETH-X 超节点方案	31
图 25: 中国智算中心布局	33
图 26: 弘信电子算力租赁方案优势	34
图 27: 云赛智联主要合作伙伴	35
图 28: 息壤智算一体机训推一体场景	35
图 29: 息壤智算一体机模型推理场景	35
图 30: 蒸馏模型性能对比	37
图 31: 华为推出 FusionCube A3000 训/推超融合一体机	38
图 32: 浪潮云帆超融合一体机 (智算型-7000 G3)	38
图 33: 华勤技术 H9236 服务器	39
图 34: 联想 AI 工作站和 AI 一体机方案	39
图 35: 弘信电子燧弘一体机和 Anlink 一体机	40
图 36: AIDC 电气设备结构	43
图 37: HVDC 全链路供电架构	43
图 38: 百度 “瀚海” 一体化直流电源	44
图 39: 超级电容在数据中心起到紧急供电及平滑电路波动的作用	45
图 40: 中国液冷数据中心产业链	46
图 41: 2019-2027 年中国液冷数据中心规模 (亿元)	47
图 42: 2019-2027 年中国液冷服务器市场规模 (亿元)	47
图 43: 液冷系统通用架构原理图	47
图 44: 冷板式液冷系统原理图	47
图 45: 单相浸没液冷系统原理图	48
图 46: 喷淋式液冷系统原理图	48
图 47: 长光华芯产品系列布局	51
图 48: 2018-2024 年长光华芯营收和利润情况	51
图 49: 2020-2024H1 长光华芯营收结构	51
图 50: 禾望电气产品系列布局	54
图 51: 2018-2024 年禾望电气营收和利润情况	54
图 52: 2021-2024 年禾望电气营收结构	54
图 53: 申菱环境主要产品布局	55
图 54: 2018-2024 年申菱环境营收和利润情况	56

图 55: 2024 年申菱环境营收结构 .....	56
图 56: 涣联股份主要产品布局 .....	58
图 57: 2018-2024 年涣联股份营收和利润情况 .....	59
图 58: 2024 年涣联股份营收结构 .....	59

## 表格目录

重点公司盈利预测、估值与评级 .....	1
表 1: 昇腾 910B 与英伟达 B200 的性能差异 .....	13
表 2: 英伟达、AMD、海光同期 GPU 参数对比 .....	18
表 3: 云天励飞 DeepEdge 10 系列性能参数 .....	19
表 4: 云商自研 ASIC 的布局情况 .....	22
表 5: 芯原股份和翱捷科技 ASIC 研发实力一览 .....	23
表 6: DeepSeek 发布, 各平台适配情况 .....	24
表 7: 各云厂商 DeepSeek 系列模型 API 定价 .....	30
表 8: 部分算力租赁业务梳理 .....	33
表 9: 润泽国际信息港项目部分数据中心规格参数 .....	34
表 10: 传统云服务和大模型一体机技术对比 .....	36
表 11: DeepSeek 一体机生态图谱 .....	37
表 12: 国内 AI 一体机梳理 .....	40
表 13: 冷板式、浸没式、喷淋式液冷技术对比 .....	48
表 14: 重点关注个股 .....	50
表 15: 长光华芯分业务收入预测 (百万元) .....	52
表 16: 长光华芯盈利预测与财务指标 .....	53
表 17: 禾望电气盈利预测与财务指标 .....	55
表 18: 申菱环境分业务收入预测 (百万元) .....	57
表 19: 申菱环境盈利预测与财务指标 .....	58
表 20: 涣联股份分业务收入预测 (百万元) .....	60
表 21: 涣联股份盈利预测与财务指标 .....	61
中芯国际财务报表数据预测汇总 .....	63
长光华芯财务报表数据预测汇总 .....	64
申菱环境财务报表数据预测汇总 .....	65
涣联股份财务报表数据预测汇总 .....	66

## 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投资咨询执业资格并登记为注册分析师，基于认真审慎的工作态度、专业严谨的研究方法与分析逻辑得出研究结论，独立、客观地出具本报告，并对本报告的内容和观点负责。本报告清晰准确地反映了研究人员的研究观点，结论不受任何第三方的授意、影响，研究人员不曾因、不因、也将不会因本报告中的具体推荐意见或观点而直接或间接受到任何形式的补偿。

## 评级说明

投资建议评级标准	评级	说明
以报告发布日后的 12 个月内公司股价（或行业指数）相对同期基准指数的涨跌幅为基准。其中：A 股以沪深 300 指数为基准；新三板以三板成指或三板做市指数为基准；港股以恒生指数为基准；美股以纳斯达克综合指数或标普 500 指数为基准。	推荐	相对基准指数涨幅 15%以上
	谨慎推荐	相对基准指数涨幅 5%~15%之间
	中性	相对基准指数涨幅-5%~5%之间
	回避	相对基准指数跌幅 5%以上
行业评级	推荐	相对基准指数涨幅 5%以上
	中性	相对基准指数涨幅-5%~5%之间
	回避	相对基准指数跌幅 5%以上

## 免责声明

民生证券股份有限公司（以下简称“本公司”）具有中国证监会许可的证券投资咨询业务资格。

本报告仅供本公司境内客户使用。本公司不会因接收人收到本报告而视其为客户。本报告仅为参考之用，并不构成对客户的投资建议，不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告所包含的观点及建议并未考虑获取本报告的机构及个人的具体投资目的、财务状况、特殊状况、目标或需要，客户应当充分考虑自身特定状况，进行独立评估，并应同时考量自身的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专家的意见，不应单纯依靠本报告所载的内容而取代自身的独立判断。在任何情况下，本公司不对任何人因使用本报告中的任何内容而导致的任何可能的损失负任何责任。

本报告是基于已公开信息撰写，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、意见及预测仅反映本公司于发布本报告当日的判断，且预测方法及结果存在一定程度局限性。在不同时期，本公司可发出与本报告所刊载的意见、预测不一致的报告，但本公司没有义务和责任及时更新本报告所涉及的内容并通知客户。

在法律允许的情况下，本公司及其附属机构可能持有报告中提及的公司所发行证券的头寸并进行交易，也可能为这些公司提供或正在争取提供投资银行、财务顾问、咨询服务等相关服务，本公司的员工可能担任本报告所提及的公司的董事。客户应充分考虑可能存在的利益冲突，勿将本报告作为投资决策的唯一参考依据。

若本公司以外的金融机构发送本报告，则由该金融机构独自为此发送行为负责。该机构的客户应联系该机构以交易本报告提及的证券或要求获悉更详细的信息。本报告不构成本公司向发送本报告金融机构之客户提供的投资建议。本公司不会因任何机构或个人从其他机构获得本报告而将其视为本公司客户。

本报告的版权仅归本公司所有，未经书面许可，任何机构或个人不得以任何形式、任何目的进行翻版、转载、发表、篡改或引用。所有在本报告中使用的商标、服务标识及标记，除非另有说明，均为本公司的商标、服务标识及标记。本公司版权所有并保留一切权利。

## 民生证券研究院：

上海：上海市浦东新区浦明路 8 号财富金融广场 1 幢 5F； 200120

北京：北京市东城区建国门内大街 28 号民生金融中心 A 座 18 层； 100005

深圳：深圳市福田区中心四路 1 号嘉里建设广场 1 座 10 层 01 室； 518048