

DeepSeek 惊艳世界，算力与应用将迎来结构性变化

核心观点

- DeepSeek 近期成为科技圈最炙手可热的明星企业：**DeepSeek 近期分别发布大模型 DeepSeek-v3 和推理模型 R1，前者在大模型主流榜单的开源模型中位列榜首，并与世界上最先进的闭源模型不分伯仲；后者性能对标 OpenAI 的 o1 正式版，在数学、代码和自然语言推理等任务上表现卓越。以上两个模型均通过算法和架构的创新，大幅降低了训练成本和推理成本，为 AI 技术的普及与创新作出了卓越的贡献。DeepSeek 相关模型自从推出后，立刻占据全球科技头条并引发巨大关注，我们认为，DeepSeek 呈现出的算法创新、性能表现、开源属性等多重因素，将对 AI 应用与算力行业产生深远的影响。
- DeepSeek 在模型训练与推理中采用了多项技术创新：**DeepSeek-V3 实现了多项工程技术上的创新，包括通过 FP8 精度训练、DualPipe 双向流水线等技术降低训练成本，通过优化 MoE 负载均衡、多头潜在注意力机制（MLA）来降低推理成本，并通过多 Token 预测（MTP）以及模型蒸馏来进一步提升模型性能，最终用极低的成本完成了训练过程，与此同时推理的成本也较其他模型有大幅下降。而 R1-Zero 的目标是验证纯 RL 能否激发模型的自主推理能力，探索无监督强化学习的潜力，而 R1 则通过数据引导+多阶段优化，平衡推理性能和实用价值，目标是打造更符合人类偏好的通用推理模型，两个模型均实现了较为理想的效果。
- DeepSeek 将对 AI 算力与应用产生结构性影响：**由于 DeepSeek-V3、R1 等模型通过算法与架构层面的多重创新大幅降低了训练端算力消耗，不过我们认为，在各模型公司仍致力于训练出性能更强大模型的目标指引下，庞大的训练集群将仍然被产业所追逐，训练算力长期看仍有前景和空间。而推理算力的需求空间则更为乐观，我们认为推理成本的大幅下降将带来需求更大幅度的增长，而在这样的过程中，算力需求结构可能将会改变，而美国如果进一步收紧 AI 芯片供应，则可能对国产芯片形成利好。而另一方面，此前 AI 应用的普及及其商业化还存在一定的困难，一方面在于模型性能仍然没法满足众多场景的需求，尤其是在推理能力和多模态环节，另一方面则是闭源模型性能领先但 API 调用的成本过高，影响了 AI 应用的大范围普及，我们认为，DeepSeek 的强推理能力、低算力成本、开源属性，使得其对 AI 应用的普及有望发挥重要的作用。

投资建议与投资标的

我们认为 AI 应用正逐步进入落地期，有望呈现较多的投资机遇，建议投资者关注行业整体性机会，而国产算力及推理算力服务企业，也将从 AI 应用繁荣中受益。

- B 端企业级应用，建议关注金山办公、鼎捷数智、泛微网络、汉得信息、软通动力、上海钢联、合合信息、新致软件、每日互动。
- 垂直行业应用领域有较强优势的企业，建议关注科大讯飞、焦点科技、润达医疗、同花顺、卫宁健康。
- AI 工具类应用，建议关注彩讯股份、虹软科技、万兴科技、福昕软件。
- 国产算力及算力服务企业，建议关注海光信息、寒武纪-U、中科曙光、云赛智联、安博通、首都在线、优刻得-W。

风险提示

行业评级

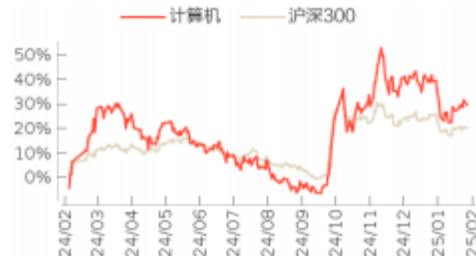
看好（维持）

国家/地区

中国

报告发布日期

计算机行业
2025 年 02 月 04 日



证券分析师

浦俊懿

021-63325888*6106

pujunyi@orientsec.com.cn

执业证书编号：S0860514050004

相关报告

字节加速 AI 落地、小米、理想"All in AI" 2024-12-29

AI 应用前景广阔、自主可控已是大趋势： 2024-12-27

计算机行业 2025 年度投资策略
AI 应用落地曙光已现

2024-11-18

技术落地不及预期；政策落地不及预期；美国进一步制裁风险；竞争激烈导致变现能力下降。

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

目录

引言	4
一、DeepSeek-V3、R1 模型推出后“火”遍全球	4
1.1 DeepSeek-V3 性能表现不输全球顶尖模型	4
1.2 R1 模型实现了比肩 OpenAI o1 的推理能力	5
1.3 DeepSeek-V3、R1 模型在产业中引发巨大反响	7
二、DeepSeek 模型训练与推理技术创新解析	8
2.1 DeepSeek-V3 模型采用了多项创新技术	8
2.2 R1-Zero 与 R1 模型尝试强化学习边界	10
三、DeepSeek 将对 AI 算力与应用产生结构性影响	11
3.1 DeepSeek 模型有望引发算力需求与市场结构变化	11
3.2 AI 应用有望在高性能、低成本模型支撑下繁荣	13
四、投资建议与投资标的	15
五、风险提示	16

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

图表目录

图 1：DeepSeek-V3 在多项评测中成绩领先	4
图 2：DeepSeek-V3 在各类测试集上的表现	5
图 3：DeepSeek-V3 API 服务价格具备很强的吸引力	5
图 4：DeepSeek-R1 在推理能力上比肩 OpenAI o1	6
图 5：基于 R1 蒸馏的小模型性能超越 OpenAI o1-mini	6
图 6：DeepSeek-R1 API 服务定价继续大幅低于 OpenAI o1	7
图 7：DeepSeek-V3 模型训练仅需要 278.8 万 GPU 小时训练资源	7
图 8：DeepSeek 成为全球增速最快的 AI 应用	8
图 9：DeepSeek 发布后下载量增长迅猛	8
图 10：DeepSeek-V3 模型基本架构	8
图 11：DeepSeek-V3 采用 FP8 为主的混合精度训练方法	9
图 12：多 Token 预测（MTP）模块结构示意	10
图 13：R1-Zero 由基础模型直接进行强化学习训练而产生	10
图 14：DeepSeek-R1-Zero 训练期间 AIME 准确率	11
图 15：DeepSeek-R1-Zero 在 RL 过程中的平均响应长度（输出长度不断增加）	11
图 17：“杰文斯悖论”指出成本下降将刺激资源需求更大增长	12
图 18：硅基流动 x 华为云联合推出基于昇腾云的 DeepSeek R1 & V3 推理服务	13
图 19：AI 产品榜全球 Web 端访问前 15 情况（24 年 12 月）	14
图 20：AI 产品榜全球 APP 端访问前 15 情况（24 年 12 月）	14
图 21：2024 年全球 AI 产品付费用户规模	14
图 22：short-CoT 模式下，k1.5 模型多方面能力领先	15
图 23：long-CoT 模式下，k1.5 模型多方面能力比肩 o1	15
图 24：Qwen2.5-Max 测试结果领先各类指令模型	15
图 25：Qwen2.5-Max 表现优于各类开源模型	15

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

引言

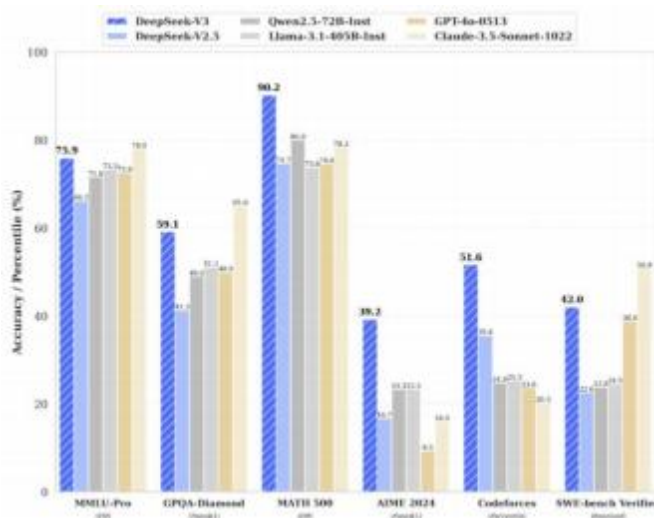
DeepSeek 近期分别发布大模型 DeepSeek-v3 和推理模型 R1，前者在大模型主流榜单的开源模型中位列榜首，并与世界上最先进的闭源模型不分伯仲；后者性能对标 OpenAI 的 o1 正式版，在数学、代码和自然语言推理等任务上表现卓越。以上两个模型均通过算法和架构的创新，大幅降低了训练成本和推理成本，为 AI 技术的普及与创新作出了卓越的贡献。我们认为，DeepSeek 呈现出的算法创新、性能表现、开源属性等多重因素，将对 AI 应用与算力行业产生深远的影响。

一、DeepSeek-V3、R1 模型推出后“火”遍全球

1.1 DeepSeek-V3 性能表现不输全球顶尖模型

2024 年 12 月 26 日，杭州深度求索（DeepSeek AI）发布 DeepSeek-V3 并同步开源，据介绍，DeepSeek-V3 多项评测成绩超越了 Qwen2.5-72B-Inst 和 Llama-3.1-405B-Inst 等其他开源模型，并在性能上和世界顶尖的闭源模型 GPT-4o 以及 Claude-3.5-Sonnet 不分伯仲。

图 1：DeepSeek-V3 在多项评测中成绩领先



数据来源：DeepSeek 微信公众号，东方证券研究所

具体而言，DeepSeek-V3 在知识类任务（MMLU, MMLU-Pro, GPQA, SimpleQA）上的水平相比前代 DeepSeek-V2.5 显著提升，接近当前表现最好的模型 Anthropic 公司于 2024 年 10 月发布的 Claude-3.5-Sonnet-1022；在长文本评测（DROP、FRAMES 和 LongBench v2）方面，V3 平均表现也超越其他模型。在算法类代码场景（Codeforces），V3 远远领先于市面上已有的全部非 o1 类模型，并在工程类代码场景（SWE-Bench Verified）逼近 Claude-3.5-Sonnet-1022。而在美国数学竞赛（AIME 2024, MATH）和全国高中数学联赛（CNMO 2024）上，DeepSeek-V3 大幅超过了其他所有开源闭源模型。

有关分析师的申明，见本报告最后部分。其他重要信息披露见分析师申明之后部分，或请与您的投资代表联系。并请阅读本证券研究报告最后一页的免责申明。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/325140240310012043>