

超配（维持）

半导体行业事件点评

deepseek 大模型实现训练降本，有望加快 AI 应用场景落地

2025 年 2 月 4 日

刘梦麟（SAC 执业证书编号：S0340521070002）

电话：0769-22110619 邮箱：liumenglin@dgzq.com.cn

陈伟光（SAC 执业证书编号：S0340520060001）

电话：0769-22110619 邮箱：chenweiguang@dgzq.com.cn

事件：

1月20日，DeepSeek发布一系列DeepSeek-R1模型，包括DeepSeek-R1、DeepSeek-R1-Zero和DeepSeek-R1-Distill系列。作为DeepSeek目前最先进的模型，R1系列模型推理能力优异，AME2024测试表现略高于OpenAI-o1-1217，MATH-500成绩表现与OpenAI-o1-1217相当，但API服务定价相比OpenAI同类产品具有明显优势，显示出明显的成本优势和商业化落地潜力。

点评：

DeepSeek性能比肩OpenAI-o1正式版，但训练、推理成本大幅下降。根据第三方基准测试，DeepSeek-R1实际性能出色，表现优于OpenAI、Anthropic和Meta等海外领先AI企业，在数学、代码编写和自然语言推理等任务中表现优异。其中，在AIME2024数学基准测试中，DeepSeek R1得分为79.8%，超越OpenAI的o1推理模型；在标准化编码测试中，DeepSeek在Codeforces上获得2029Elo的评级，超越96.3%人类竞争对手。成本方面，DeepSeek训练成本仅为OpenAI同类模型的十分之一，仅花费557.6万美元与2048块英伟达H800 GPU便完成了性能对标GPT-4o的模型训练；推理方面，DeepSeek的推理成本低至每百万Token 0.14美元，而OpenAI推理成本为每百万Token 7.5美元。

DeepSeek通过大幅降本引领AI成本革命，为突破海外高端GPU“卡脖子”提供思路。根据Scaling Law，在AI领域，模型性能随着模型规模（如参数数量）、训练数据量及计算资源量的增加而提升，且这种提升遵循可预测的幂律关系，即：模型性能会随着模型规模的指数级增加而实现线性提升。此前发布的OpenAI的GPT系列、谷歌的Gemini系列以及Anthropic的Claude系列，最新版本规模均已突破千亿参数大关。相比美国的领先大模型，DeepSeek-R系列在技术路线上实现突破性创新，成为首次摒弃监督微调环节并完全依赖强化学习训练的大语言模型，对于小模型则通过知识蒸馏（Knowledge Distillation）技术成功获取大模型的推理能力，实现小模型在推理任务上的显著提升。得益于技术路线上的创新，DeepSeek-R1 API服务定价为每百万输入tokens 1元（缓存命中）/4元（缓存未命中），每百万输入tokens 16元，定价远低于OpenAI等可比海外大模型API服务。我们认为，在硅谷传统“堆算力、拼资本”的AI发展路径下，DeepSeek通过技术创新实现降本增效，不仅引领人工智能成本革命，也为应对美国高端GPU出口管制提供了思路。

DeepSeek实现大模型成本大幅优化，有望加快AI应用场景落地，长期有望增加算力总消耗量。我们认为，相比可比大模型，DeepSeek通过算法优化、模型压缩等技术大幅降低单位算力需求，该模式可能被其他企业借鉴，企业通过软件/架构优化而非硬件堆叠来满足需求，可能导致短期算力芯片采购节奏放缓；但从长期来看，DeepSeek引领大模型成本缩减，算力效率提升有望降低行业技术门槛，加快AI应用场景落地，而更广泛的用户与应用场景可能引发对更大参数与更复杂的大模型的迭代需求，如自动驾驶等场景对算力需求呈现指数级增长。因此，从长远来看，大模型降本增效可能刺激更大规模的应用部署，算力总消耗量有望增加；国产替代方面，目前美国限制高端GPU出口，国产算力自主可控需求迫切，若DeepSeek的降本增效技术与国产算力芯片结合，可能加速国产算力的国产替代进程，

本报告的风险等级为中风险。

本报告的信息均来自已公开信息，关于信息的准确性与完整性，建议投资者谨慎判断，据此入市，风险自担。

请务必阅读末页声明。

实现“弯道超车”。

投资建议：DeepSeek性能比肩海外领先大模型，且推理、训练成本实现大幅降低，展现出明显的商业化落地潜力，有望加快AI应用场景落地，长期有望增加算力总消耗量，并加快算力芯片的国产替代。建议关注AI算力、存储、终端SoC等环节。

风险提示：国产替代进程不及预期、AI应用场景落地不及预期、行业资本开支不及预期等。

东莞证券研究报告评级体系：

公司投资评级	
买入	预计未来 6 个月内，股价表现强于市场指数 15%以上
增持	预计未来 6 个月内，股价表现强于市场指数 5%-15%之间
持有	预计未来 6 个月内，股价表现介于市场指数±5%之间
减持	预计未来 6 个月内，股价表现弱于市场指数 5%以上
无评级	因无法获取必要的资料，或者公司面临无法预见结果的重大不确定性事件，或者其他原因，导致无法给出明确的投资评级；股票不在常规研究覆盖范围之内
行业投资评级	
超配	预计未来 6 个月内，行业指数表现强于市场指数 10%以上
标配	预计未来 6 个月内，行业指数表现介于市场指数±10%之间
低配	预计未来 6 个月内，行业指数表现弱于市场指数 10%以上

说明：本评级体系的“市场指数”，A股参照标的为沪深 300 指数；新三板参照标的为三板成指。

证券研究报告风险等级及适当性匹配关系	
低风险	宏观经济及政策、财经资讯、国债等方面的研究报告
中低风险	债券、货币市场基金、债券基金等方面的研究报告
中风险	主板股票及基金、可转债等方面的研究报告，市场策略研究报告
中高风险	创业板、科创板、北京证券交易所、新三板（含退市整理期）等板块的股票、基金、可转债等方面的研究报告，港股股票、基金研究报告以及非上市公司的研究报告
高风险	期货、期权等衍生品方面的研究报告

投资者与证券研究报告的适当性匹配关系：“保守型”投资者仅适合使用“低风险”级别的研报，“谨慎型”投资者仅适合使用风险级别不高于“中低风险”的研报，“稳健型”投资者仅适合使用风险级别不高于“中风险”的研报，“积极型”投资者仅适合使用风险级别不高于“中高风险”的研报，“激进型”投资者适合使用我司各类风险级别的研报。

证券分析师承诺：

本人具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地在所知情的范围内出具本报告。本报告清晰准确地反映了本人的研究观点，不受本公司相关业务部门、证券发行人、上市公司、基金管理公司、资产管理公司等利益相关者的干涉和影响。本人保证与本报告所指的证券或投资标的无任何利害关系，没有利用发布本报告为自身及其利益相关者谋取不当利益，或者在发布证券研究报告前泄露证券研究报告的内容和观点。

声明：

东莞证券股份有限公司为全国性综合类证券公司，具备证券投资咨询业务资格。

本报告仅供东莞证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因接收人收到本报告而视其为客户。本报告所载资料及观点均为合规合法来源且被本公司认为可靠，但本公司对这些信息的准确性及完整性不作任何保证。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，可随时更改。本报告所指的证券或投资标的的价格、价值及投资收入可跌可升。本公司可发出其它与本报告所载资料不一致及有不同结论的报告，亦可因使用不同假设和标准、采用不同观点和分析方法而与本公司其他业务部门或单位所给出的意见不同或者相反。在任何情况下，本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并不构成对任何人的投资建议。投资者需自主作出投资决策并自行承担投资风险，据此报告做出的任何投资决策与本公司和作者无关。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所引致的任何损失负任何责任，任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。本公司及其所属关联机构在法律许可的情况下可能会持有本报告中提及公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行、经纪、资产管理等服务。本报告版权归东莞证券股份有限公司及相关内容提供方所有，未经本公司事先书面许可，任何人不得以任何形式翻版、复制、刊登。如引用、刊发，需注明本报告的机构来源、作者和发布日期，并提示使用本报告的风险，不得对本报告进行有悖原意的引用、删节和修改。未经授权刊载或者转发本证券研究报告的，应当承担相应的法律责任。

东莞证券股份有限公司研究所

广东省东莞市可园南路 1 号金源中心 24 楼

邮政编码：523000

电话：（0769）22115843

网址：www.dgzq.com.cn