

以 DeepSeek 视角，解读 DeepSeek 逆袭

春节期间，DeepSeek 成为全球 AI 行业关注焦点，从“无名之辈”化身“绝世高手”。我们围绕三个核心话题，包括最大亮点、技术创新和潜在影响等，借助 DeepSeek 视角，让 DeepSeek 来解读 DeepSeek，一起探讨这次“逆袭奇迹”。

DeepSeek 最大亮点是什么？“成本创新”。DeepSeek 整体表现与 OpenAI 等海外大模型第一梯队处于同一水平，不相上下，而成本控制方面具有显著优势，包括在训练、推理和商业化定价方面。整体成本仅为 OpenAI 的 5%-10%，商业化定价仅为 OpenAI 的 2%-5%。

DeepSeek 如何实现“成本创新”？DeepSeek 优化策略主要包括：1) 稀疏激活的混合专家 (MoE) 架构，仅激活任务相关参数，降低 90% 算力消耗；2) 本地化部署：支持消费级硬件，无需依赖昂贵的云服务等；3) 规则奖励系统：替代昂贵的“AI 监工”模型，减少额外训练；4) 开源生态：利用社区贡献降低研发成本，迭代速度提升 3 倍。

“成本创新”对 AI 产业链有何影响？DeepSeek 的低成本优势正在推动 AI 产业链从“算力竞赛”转向“算法优化”，其影响包括：1) 云计算：推理需求爆发倒逼硬件与架构创新，混合云与边缘计算成主流；2) 大模型：开源生态崛起加速技术民主化，垂直领域模型与 Agent 化应用成为新战场；3) 应用层：普惠化释放长尾市场潜力，端侧硬件与行业解决方案重塑商业逻辑。未来，AI 产业将更依赖算法密度与生态协同，而非单纯资源垄断。

投资建议：1) DeepSeek 打破了美国在大模型层面的垄断格局，或带动市场重新审视国产 AI 行业，带来价值重估；2) DeepSeek 的“成本创新”将使得大模型价格进一步平民化，AI 普惠将加速应用层的爆发；3) 虽然 DeepSeek 降低了单次算力成本，但随着未来大规模应用的指数级增长，长期来看将有助于推动云服务需求的上升。

投资风险：地缘政治风险，场景落地不及预期。

赵丹

首席互联网分析师
dan_zhao@spdbi.com
(852) 2808 6436

杨子超，CFA

互联网分析师
charles_yang@spdbi.com
(852) 2808 6409

2025 年 2 月 5 日



扫码关注浦银国际研究

附录：DeepSeek 问答文字版

2025 年 2 月 4 日，通过 DeepSeek 网页版提问。

Question 1: DeepSeek 和 OpenAI 相比，成本方面有何优势？

DeepSeek 在成本控制方面相较于 OpenAI 具有显著优势，具体体现在训练成本、推理成本和商业化定价等多个维度。以下是详细对比分析：

1. 训练成本对比

- DeepSeek:

- o DeepSeek-V3 的训练成本仅为 558 万美元，采用高效的混合专家（MoE）架构和稀疏激活策略，显著降低了算力消耗。
- o 其训练时间仅需 55 天，使用 2,048 块 NVIDIA H800 GPU，计算资源消耗远低于行业平均水平。

- OpenAI:

- o OpenAI 的 GPT-4 训练成本高达 10 亿美元，依赖大规模全参数激活和复杂的监督微调（SFT）流程。
- o 训练时间更长，且需要数百万 GPU 小时的支持，导致整体成本居高不下。

成本节省：DeepSeek 的训练成本仅为 OpenAI 的 5%-10%，实现了显著的效率提升。

2. 推理成本对比

- DeepSeek:

- o 输入费用：0.14 美元/百万 tokens（缓存未命中），缓存命中时低至 0.014 美元/百万 tokens。
- o 输出费用：0.28 美元/百万 tokens，显著低于 OpenAI 和 Claude 系列模型。
- o 采用稀疏激活策略，每次任务仅激活部分参数，进一步降低了推理成本。

- OpenAI:

- o GPT-4o 的输入费用为 2.5 美元/百万 tokens (缓存未命中), 输出费用高达 10 美元/百万 tokens。

- o 全参数激活模式导致算力消耗巨大, 推理成本居高不下。

成本节省: DeepSeek 的推理成本仅为 OpenAI 的 1/10 至 1/20, 性价比优势显著。

3. 商业化定价对比

- DeepSeek:

- o 提供开源模型和低成本 API, 每百万 tokens 的价格仅为 0.48 美元, 远低于 OpenAI 的 18 美元。

- o 本地化部署支持消费级硬件 (如 RTX 4090), 进一步降低企业使用门槛。

- OpenAI:

- o GPT-4 的 API 定价较高, 每百万 tokens 的价格为 18 美元, 且依赖昂贵的云服务资源。

- o 企业级部署需要大规模 GPU 集群, 硬件成本高昂。

成本节省: DeepSeek 的商业化定价仅为 OpenAI 的 2%-5%, 尤其适合预算敏感的中小企业和开发者。

总结

DeepSeek 在训练、推理和商业化定价方面的成本优势显著, 整体成本仅为 OpenAI 的 5%-10%。这种高效的成本控制不仅降低了企业使用门槛, 还推动了 AI 技术的普惠化发展。

Question 2: DeepSeek 如何实现成本创新？底层技术与 OpenAI

有何异同？

DeepSeek 通过以下创新策略实现成本优势：

- 稀疏激活的 MoE 架构：仅激活任务相关参数，降低 90% 的算力开销。
- 本地化部署：支持消费级硬件，无需依赖昂贵的云服务。
- 规则奖励系统：替代昂贵的“AI 监工”模型，减少额外训练开销。
- 开源生态：利用社区贡献降低研发成本，迭代速度提升 3 倍。

从技术层面来看，DeepSeek 和 OpenAI 在底层技术上既有相似之处，也存在显著差异。以下是两者的主要异同点分析：

一、相同点

1. 基础架构：Transformer 架构

- o 两者都基于 Transformer 架构，利用自注意力机制处理输入序列，能够高效处理长文本和复杂语言任务。
- o Transformer 架构的并行计算能力使得模型在处理大规模数据时表现出色。

2. 推理能力

- o 两者都强调推理能力的提升，通过不同的方式实现复杂问题的逐步分解和逻辑推理。
- o 例如，DeepSeek 和 OpenAI 都采用了多步骤推理技术（如思维链推理，Chain-of-Thought）来增强模型的逻辑处理能力。

3. 多阶段训练

- o 两者都采用了多阶段训练流程，通过不同阶段的优化提升模型的性能。例如，DeepSeek 的冷启动和多阶段强化学习，以及 OpenAI 的监督微调（SFT）和思维链推理。

二、 不同点

1. 训练方式

- o DeepSeek: 主要依赖强化学习 (RL), 尤其是其创新的 Group Relative Policy Optimization (GRPO) 算法。这种训练方式不依赖大量监督微调数据, 而是通过组内奖励对比优化策略, 实现推理能力的自主进化。
- o OpenAI: 高度依赖监督微调 (SFT), 通过大量人工标注的数据提升模型在特定任务上的表现。这种方式虽然能提高任务准确性, 但对数据质量和数量的依赖较大。

2. 架构设计

- o DeepSeek: 采用模块化框架和 Mixture-of-Experts (MoE) 架构。MoE 将模型划分为多个专家子模型, 每个子模型负责处理不同的输入任务, 从而提高模型的泛化能力和鲁棒性。
- o OpenAI: 基于 Transformer 架构, 通过优化层结构和注意力机制提升模型性能, 更注重通用性和多功能性。

3. 推理与生成能力

- o DeepSeek: 在需要深度推理的领域 (如数学、代码生成) 表现出色, 适合复杂任务的处理。例如, DeepSeek 在数学推理任务 (如 AIME 2024) 中表现优于 OpenAI。
- o OpenAI: 在自然语言处理 (NLP) 和文本生成方面表现更为突出, 适合广泛的应用场景, 如文本摘要、创意写作等。

4. 成本与效率

- o DeepSeek: 通过算法优化和模块化设计, 显著降低了训练和推理成本。例如, DeepSeek 的训练成本仅为 558 万美元, 远低于 OpenAI 的投入。
- o OpenAI: 虽然性能强大, 但其训练和推理成本较高, 主要依赖于大规模计算资源 and 高质量标注数据。

5. 应用场景

- o DeepSeek: 更适合需要深度推理和专业领域知识的场景, 如医疗数据分析、金融模式检测等。
- o OpenAI: 在通用任务 (如文本生成、客户服务) 中表现更为出色, 适合广泛的企业和个人用户。

Question 3: DeepSeek 低成本优势对全球 AI 产业链有何影响？

比如云计算、大模型和应用层等？

DeepSeek 的低成本优势正在深刻重塑全球 AI 产业链，其影响从底层算力到应用生态呈现出多维度的连锁反应。以下从云计算、大模型和应用层三个维度展开分析：

一、云计算：需求结构变革与降本压力传导

1. 算力需求从训练向推理倾斜

DeepSeek 的高效模型架构（如 MoE 动态激活、FP8 混合精度）显著降低训练成本（557.6 万美元，仅为 GPT-4 的 5%），导致训练端算力需求增速放缓。而推理端因模型普及化需求激增，预计 2026 年推理算力占 GenAI 总需求比例将突破 70%。

- o 硬件厂商转型：传统 GPU 厂商（如英伟达）面临算力过剩风险，加速向专用 AI 芯片（如 ASIC）及液冷技术研发转型。
- o 云服务商策略调整：阿里云、AWS 等需优化推理集群架构（如铜连接方案、CPO 光通信技术），并探索“按需付费”定价模式以应对成本压力。

2. 边缘计算与混合云崛起

DeepSeek 支持消费级硬件（如 RTX 4090）和端侧部署，推动算力从集中式云向边缘侧迁移。例如，硅基流动已基于华为昇腾算力支持 DeepSeek API，海光 DCU 适配其模型，预示混合云架构将成为主流。

二、大模型：开源生态与技术路径重构

1. 打破“算力垄断”，推动技术民主化

DeepSeek 通过算法优化（如动态注意力、多令牌预测）和工程创新（如 MoE 负载均衡），以 1/20 的训练成本实现与 GPT-4o 相当的推理性能，验证了“轻量级颠覆”路径的可行性。

- o 开源生态繁荣：R1 模型开源后 6 天下载量达 110 万次，吸引 Meta 等巨头加码开源社区，形成“技术共享-生态繁荣-反哺创新”的正循环。
- o Scaling Law 转向：模型开发重心从预训练转向强化学习与推理优化，催生新 Scaling Law（如 GRPO 算法），加速复杂任务（数学、代码）的 Agent 化落地。

2. 垂直领域模型爆发

低成本模型降低企业私有化部署门槛，推动 LLM 向金融、医疗、教育等垂直领域渗透。例如，某券商使用 DeepSeek 生成合规研报，效率提升 60%。

三、应用层：普惠化与场景革命

1. 中小开发者与长尾市场激活

DeepSeek 的 API 定价仅为 OpenAI 的 1/10（输入 0.5 元/百万 tokens），且支持设备端运行，使中小开发者能够低成本开发 AI 应用。例如，教育机构通过 AI 教师方案覆盖三四线城市，潜在市场规模超千亿。

2. 商业模式创新

传统云计算厂商（如腾讯云）转向“AI+行业解决方案”高毛利模式，衍生数据服务（如金融舆情 API）成为新增长点。

总结：价值分配重构与长期趋势

DeepSeek 的低成本优势正在推动 AI 产业链从“算力竞赛”转向“算法优化”，其核心影响包括：

- 云计算：推理需求爆发倒逼硬件与架构创新，混合云与边缘计算成主流。
- 大模型：开源生态崛起加速技术民主化，垂直领域模型与 Agent 化应用成为新战场。
- 应用层：普惠化释放长尾市场潜力，端侧硬件与行业解决方案重塑商业逻辑。

未来，AI 产业将更依赖算法密度与生态协同，而非单纯资源垄断。中国企业若能抓住开源化、设备端化趋势，有望在 AI 2.0 时代占据先机，但需警惕技术跃进伴随的伦理与战略风险。

免责声明

本报告之收取者透过接受本报告(包括任何有关的附件),表示及保证其根据下述的条件下有权获得本报告,且同意受此中包含的限制条件所约束。任何没有遵循这些限制的情况可能构成法律之违反。

本报告是由从事证券及期货条例(香港法例第 571 章)中第一类(证券交易)及第四类(就证券提供意见)受规管活动之持牌法国-浦银国际证券有限公司(统称“浦银国际证券”)利用集团信息及其他公开信息编制而成。所有资料均搜集自被认为是可靠的来源,但并不保证数据之准确性、可信性及完整性,亦不会因资料引致的任何损失承担任何责任。报告中的资料来源除非另有说明,否则信息均来自本集团。本报告的内容涉及到保密数据,所以仅供阁下为其自身利益而使用。除了阁下以及受聘向阁下提供咨询意见的人士(其同意将本材料保密并受本免责声明中所述限制约束)之外,本报告分发给任何人均属未经授权的行为。

任何人不得将本报告内任何信息用于其他目的。本报告仅是为提供信息而准备的,不得被解释为是一项关于购买或者出售任何证券或相关金融工具的要约邀请或者要约。阁下不应将本报告内容解释为法律、税务、会计或投资事项的专业意见或为任何推荐,阁下应当就本报告所述的任何交易涉及的法律及相关事项咨询其自己的法律顾问和财务顾问的意见。本报告内的信息及意见乃于文件注明日期作出,日后作修改而不另通知,亦不一定会更新以反映文件日期之后发生的进展。本报告并未包含公司可能要求的所有信息,阁下不应仅仅依据本报告中的信息而作出投资、撤资或其他财务方面的任何决策或行动。除关于历史数据的陈述外,本报告可能包含前瞻性的陈述,牵涉多种风险和不确定性,该等前瞻性陈述可基于一些假设,受限于重大风险和不确定性。

本报告之观点、推荐、建议和意见均不一定反映浦银国际证券的立场。浦银国际控股有限公司及其附属公司、关联公司(统称“浦银国际”)及/或其董事及/或雇员,可能持有在本报告内所述或有关公司之证券、并可能不时进行买卖。浦银国际或其任何董事及/或雇员对投资者因使用本报告或依赖其所载信息而引起的一切可能损失,概不承担任何法律责任。

浦银国际证券建议投资者应独立地评估本报告内的资料,考虑其本身的特定投资目标、财务状况及需要,在参与有关报告中所述公司之证券的交易前,委任其认为必须的法律、商业、财务、税务或其它方面的专业顾问。惟报告内所述的公司之证券未必能在所有司法管辖区或国家或供所有类别的投资者买卖。对部分的司法管辖区或国家而言,分发、发行或使用本报告会抵触当地法律、法则、规定、或其它注册或发牌的规例。本报告不是旨在向该等司法管辖区或国家的任何人或实体分发或由其使用。

美国

浦银国际不是美国注册经纪商和美国金融业监管局(FINRA)的注册会员。浦银国际证券的分析师不具有美国金融监管局(FINRA)分析师的注册资格。因此,浦银国际证券不受美国就有关研究报告准备和分析师独立性规则的约束。

本报告仅提供给美国 1934 年证券交易法规则 15a-6 定义的“主要机构投资者”,不得提供给其他任何个人。接收本报告之行为即表明同意接受协议不得将本报告分发或提供给任何其他人士。接收本报告的美国收件人如想根据本报告中提供的信息进行任何买卖证券交易,都应仅通过美国注册的经纪交易商来进行交易。

英国

本报告并非由英国 2000 年金融服务与市场法(经修订)(「FSMA」)第 21 条所界定之认可人士发布,而本报告亦未经其批准。因此,本报告不会向英国公众人士派发,亦不得向公众人士传递。本报告仅提供给合格投资者(按照金融服务及市场法的涵义),即(i)按照 2000 年金融服务及市场法 2005 年(金融推广)命令(「命令」)第 19(5)条定义在投资方面拥有专业经验之投资专业人士或(ii)属于命令第 49(2)(a)至(d)条范围之高净值实体或(iii)其他可能合法与之沟通的人士(所有该等人士统称为「有关人士」)。不属于有关人士的任何机构和个人不得遵照或倚赖本报告或其任何内容行事。

本报告的版权仅为浦银国际证券所有,未经书面许可任何机构和个人不得以任何形式转发、翻版、复制、刊登、发表或引用,浦银国际证券对任何第三方的该等行为保留追述权利,并且对第三方未经授权行为不承担任何责任。

权益披露

- 1) 浦银国际并没有持有本报告所述公司逾 1%的财务权益。
- 2) 浦银国际跟本报告所述公司在过去 12 个月内并没有任何投资银行业务的关系。
- 3) 浦银国际并没有跟本报告所述公司为其证券进行庄家活动。

评级定义

证券评级定义:

“买入”: 未来 12 个月, 预期个股表现超过同期其所属的行业指数

“持有”: 未来 12 个月, 预期个股表现与同期所属的行业指数持平

“卖出”: 未来 12 个月, 预期个股表现逊于同期其所属的行业指数

行业评级定义 (相对于 MSCI 中国指数):

“超配”: 未来 12 个月优于 MSCI 中国 10%或以上

“标配”: 未来 12 个月优于/劣于 MSCI 中国少于 10%

“低配”: 未来 12 个月劣于 MSCI 中国超过 10%

分析师证明

本报告作者谨此声明: (i) 本报告发表的所有观点均正确地反映作者有关任何及所有提及的证券或发行人的个人观点, 并以独立方式撰写; (ii) 其报酬没有任何部分曾经, 是或将会直接或间接与本报告发表的特定建议或观点有关; (iii) 该等作者没有获得与所提及的证券或发行人相关且可能影响该等建议的内幕信息 / 非公开的价格敏感数据。

本报告作者进一步确定 (i) 他们或其各自的关联人士 (定义见证券及期货事务监察委员会持牌人或注册人操守准则) 没有在本报告发行日期之前的 30 个历日内曾买卖或交易过本报告所提述的股票, 或在本报告发布后 3 个工作日 (定义见《证券及期货条例》(香港法例第 571 章)) 内将买卖或交易本文所提述的股票; (ii) 他们或其各自的关联人士并非本报告提述的任何公司的雇员; 及 (iii) 他们或其各自的关联人士没有拥有本报告提述的证券的任何金融利益。

浦银国际证券机构销售团队

杨增希

essie_yang@spdbi.com

852-2808 6469

浦银国际证券财富管理团队

王玥

emily_wang@spdbi.com

852-2808 6468

浦银国际证券有限公司

SPDB International Securities Limited

网站: www.spdbi.com

地址: 香港轩尼诗道 1 号浦发银行大厦 33 楼

