

科技论坛

DeepSeek 技术及人工智能发展趋势研究

杨知晓¹ 孙樱校² 何斌¹

1.沈阳联勤保障中心药品仪器监督检验站 辽宁沈阳 110026;

2.沈阳市和平区南京街第九小学 辽宁沈阳 110026

【摘要】DeepSeek, 一种被认为是新一代人工智能典型“代表”的国产大语言模型, 自正式发布以来受到广泛关注。它的出现预示着我国在人工智能领域迈出一大步, 表明了人工智能将迎来新的发展机遇。基于此, 在概括性介绍DeepSeek的基础上, 重点介绍其创新应用, 特别是DeepSeek-R1和DeepSeek-V3在推理技术和大语言模型领域的创新应用; 深入探讨和分析其背后所潜藏的人工智能发展趋势, 旨在进一步加深DeepSeek的认知以及对人工智能未来发展态势的了解, 从而更好地抓住发展机遇, 推动人工智能实现质的飞跃。

【关键词】DeepSeek; 人工智能; 发展趋势

Research on DeepSeek Technology and the Development Trend of Artificial Intelligence

Yang Zhixiao¹ Sun Yingxiao² He Bin¹

1.Drug and Instrument Supervision and Inspection Station of Shenyang Joint Logistics Support Center, Shenyang, Liaoning Province 110026; 2.Nanjing Street Ninth Primary School, Heping District, Shenyang, Liaoning, China 110026

【Abstract】DeepSeek, a domestically produced large language model considered a typical "representative" of the new generation of artificial intelligence, has received widespread attention since its official release. Its emergence heralds a big step forward for China in the field of artificial intelligence, indicating that artificial intelligence will usher in new development opportunities. Based on this, on the basis of a general introduction to DeepSeek, the focus is on its innovative applications, especially the innovative applications of DeepSeek-R1 and DeepSeek-V3 in the fields of inference technology and large language models; In depth exploration and analysis of the hidden trends of artificial intelligence behind it aims to further deepen DeepSeek's understanding and knowledge of the future development trend of artificial intelligence, in order to better seize development opportunities and promote a qualitative leap in artificial intelligence.

【Key words】DeepSeek; artificial intelligence; development trend

引言:

2025 年年初, DeepSeek 大模型全面落地, 它的出现推动了人工智能行业从“工具辅助”向“认知协同”的范式转变。现如今, 人工智能行业正面临算法效率颠覆性突破、行业应用边界极速扩张的发展现状, 换言之, DeepSeek 驱动下的技术变革正在悄然进行。在这种局势下, 认识和掌握DeepSeek、解读其背后所隐藏的趋势特征, 将成为各企业抢占市场先机的重要抓手。

1. DeepSeek 技术概述

1.1 基本概念

从本质上讲, DeepSeek 是一种具备强大的自然语言处理能力、高效推理效率和多模态交互功能的智能搜索与分析系统, 其核心目标在于以海量数据的深度分析为支撑为用户提供高精度的决策支撑。相较于传统的搜索引擎, DeepSeek 部署了多种搜索与分析逻辑, 包括传统的关键词匹配以及基于深度学习模式的数据上下文语义分析, 正因如此, 它有着更加智能化的搜索和分析功能。

1.2 核心技术

DeepSeek 的核心技术包括自然语言处理 (NLP)、混合专家模型 (MoE)、多头潜在注意力 (MLA)、强化学习 (RL)

与监督微调 (SFT)、模型蒸馏与性能提升、多模态交互、训练框架与并行训练^[1]。

1.3 基本分支

DeepSeek 模型家族共有两大分支, 一个是强调通用性和自然语言覆盖广度的 V 系列大模型, 如 DeepSeek-V3; 一个是强调推理与思维链的 R 系列大模型, 如 DeepSeek-R1 等。

1.4 典型应用场景

DeepSeek 的独特技术性使其得以被应用到多个场景中。

第一, 企业智能决策。DeepSeek 通过深度分析销售数据、市场趋势等企业内部数据, 帮助企业识别和规避潜在的市场风险, 抓住市场机遇。第二, 金融风控与投资分析。DeepSeek 通过分析历史交易数据, 预测金融产品的未来走向, 以此帮助投资者规避金融投资风险。第三, 辅助医疗诊断。DeepSeek 通过分析大量医学文献和病例资料, 增强自身的“疾病智能诊断能力”, 为医生制定治疗方案提供建议。

2. DeepSeek 的创新应用

2.1 DeepSeek-V3: 基于新的模型架构和算法设计, 优化模型训练方法

DeepSeek-V3 的问世, 为行业提供了低成本、高效率的模型训练范式。

早期出现的 GPT 系列、LLaMA 系列都面临图形处理器

(GPU)资源消耗巨大这一技术难题。而 DeepSeek-V3 模型, 凭借 MLA+MoE+MTP (多 token 预测机制) 这一创新的设计理念, 不仅解决了资源利用的问题, 还促进了模型推理效率的提高。第一, 基于低秩联合压缩技术的 MLA 通过减少推理过程中的键值 (KV) 缓存需求, 达到减少内存占用、提高系统整体效率的目的。第二, MoE 与动态调节偏置机制、辅助损失的有机结合, 有效提升了负载分配的平衡性。第三, MTP 通过一次性预测多个后续 token 的方式, 让模型拥有更高的预测效率以及对数据上下文的理解能力。需要注意的是, 这种技术在训练阶段具有较高的实用性, 即可以增强模型学习能力, 但是并不适用于推理阶段。综上, 在技术的创新应用支撑下, DeepSeek-V3 模型同时兼具高性能、高推理效率、低资源消耗等属性。

FP8 混合精度训练框架的应用是 DeepSeek 的一大创新

表 1 多模型训练成本对比表

模型名称	Gemini Ultra	Llama-3	Claude 3	DeepSeek-V3	GPT-4o	讯飞星火 X1-13B
训练成本/美元	约 1.9 亿	约 9000 万	约 1 亿	约 557 万	约 1 亿	约 780 万

2.2 DeepSeek-R1: 基于 RL 与 SFT 的有机结合, 增强模型的推理能力

判定一个模型的智能水平, 首先要关注的是模型的认知

体现。与传统的 FP16/FP32 方案不同, FP8 格式在 8 位浮点计算位宽的支持下, 仅占用少量显存, 且不会出现较大的带宽压力。在该框架的支撑下, DeepSeek 得以高度协同优化大模型算法与智能硬件, 而这意味着我国基于国产软硬件的协同优化设计取得重大突破。

在训练成本方面, DeepSeek-V3 模型具有显著优势, 如表 1 所示。相较于 Gemini Ultra、Claude 3、GPT-4o 等动辄需支付上亿美元训练费的模型, DeepSeek-V3 模型的训练成本仅约为它们的 1/2 ~ 1/3。虽然, 讯飞星火 X1-13B、Llama-3 两大模型的训练成本并未达到或是突破 1 亿美元, 但是, 相较于 DeepSeek-V3 模型, 其训练成本仍是很高^[2]。从这组数据可以看出, DeepSeek-V3 模型在 671B (6710 亿) 参数量训练领域是一种更为低廉、高效的训练范式, 而这也无形中彰显了 DeepSeek 的技术实力。

推理能力。如图 1 所示, DeepSeek-R1 有着创新的推理技术和架构设计作为支撑, 正因如此, DeepSeek-R1 具备更加强大的深度推理能力。

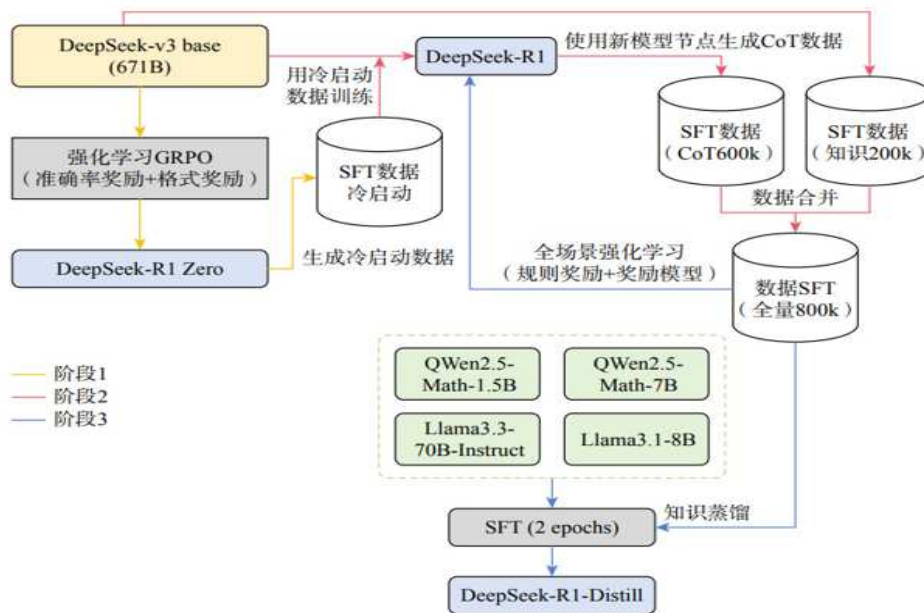


图 1 DeepSeek 强化学习训练流程示意图

DeepSeek 研究团队基于“纯强化学习过程可以提高大模型推理能力”这一发现, 通过采用准确度奖励+格式奖励这一奖励机制, 研发出了 DeepSeek R1-Zero 模型, 这一可以具备推理强化训练功能的大模型。虽相较于其他模型, 该模型仍存在推理表现不理想的问题, 但受益于生成中间思维步骤, 仍具备可行的推理能力。而这也恰恰证明了“纯强化学习过程可以提高大模型推理能力”这一结论。此外, DeepSeek-R1 推理性能的优化是借助于监督微调与强化学习的有机结合, 其原理是以生成未经初始训练阶段的“冷

启动”监督微调数据为起点, 以指令微调训练为手段, 提升模型推理能力。具体训练方法如表 2 所示。同时, 为避免出现语言混乱等问题, 引入准确度、格式一致性和语言一致性奖励机制。这一认知推理性能优化逻辑与 OpenAI o1 有着明显不同。OpenAI o1 推理性能的优化是基于复杂任务分解与依次解决逻辑, 这种优化方式高度依赖于高质量的人工标注数据, 并且很难判断中间步骤是否存在错误等^[3]。所以, 相较于 OpenAI o1, DeepSeek-R1 在复杂推理任务中有着更加优越的性能表现。

表 2 DeepSeek-R1 模型的训练方法

训练阶段	具体表现
预训练	首先, 模型在包含数十亿词汇的多样化语料库上进行预训练, 学习通用的语言表示; 其次, DeepSeek-R1 采用掩码语言模型任务, 随机掩码输入序列中的部分词汇, 并让模型预测被掩码的词汇, 其目的是让模型学习词汇之间的上下文关系; 最后, 为增强模型对句子间关系的理解, DeepSeek-R1 采用了下一句预测任务, 判断两个句子是否连续。

首先，在预训练完成后，DeepSeek-R1 微调特定任务（如文本分类、机器翻译、问答系统等）的数据集，以提升其在具体应用场景中的适用性；其次，模型依托部署的多任务学习框架，同时优化多个任务的损失函数，以提升模型的泛化能力。

3.人工智能发展趋势分析

3.1 逐步打破数据、算力与算法困境

在整个 AI 产业链体系中，涵盖数据、算力和算法三大核心要素的 AI 基础层是整个链条得以正常运行的基础和前提。

现阶段，我国在这一领域仅取得了“规模庞大”这一成就，在质量方面，仍有很多路要走，换言之，我国在 AI 基础层研发领域尚处于国际跟跑阶段，不过，现已出现局部突破的迹象。

首先，在数据方面，我国取得了“拥有全球最大的数据资源池”这一成就，但是，存在“数据孤岛”、数据价值密度不高、数据标注质量有待提高等问题。DeepSeek 以“四两拨千斤”之势让数据训练效率有所提高，即通过总结、分类、选择性处理外加模型训练的方式，提高数据价值密度、共享性与标注质量等。

其次，在算力方面，我国取得了“算力规模已达 230EFLOPS（每秒百亿亿次浮点运算）”的成就，不过，结构性问题相对突出，表现为高端芯片国产化缺口较大、超算中心能耗与算力资源调度效率居高不下。对于这一问题，DeepSeek 以智能化管理低配版 GPU 集群、多源异构算力+协同调度 CPU、GPU、边缘计算设备这一方案，攻克算力封锁难关。

最后，在算法方面，我国对 Transformer 架构、MoE（混合专家模型）、Diffusion Model（扩散模型）等源自海外的核心框架的依赖于仍是很高。DeepSeek 的问世虽为我国的算法突破提供了“二次创新”和“集中创新”的解决方案，但是并未取得真正意义上的颠覆性基础理论突破。也就是说，人工智能的未来研究仍需关注算法基础理论领域的创新与突破。

3.2 技术架构创新

DeepSeek 摆脱了“参数竞赛”误区，它以“重新定义技术路线、基于群智协同的系统级工程优化”这一新思路，为人工智能行业的未来发展提供了新的方向指引。正如上文所述，DeepSeek 通过创新性应用 MoE、FP8 混合精度训练框架、MLA、RL 与 SFT，实现了对系统的协同优化，如依托 MoE 架构拆分模型，使其以多个“子专家”按需调用的方式动态配置资源，从而提高模型推理效率；依托 RL 与 SFT，打造模仿“先思考再行动”的认知模式，然后，借助决策树模拟，进一步减少无效计算等^[4]。在未来，人工智能领域的研究势必要向着技术架构创新这一发现转型，以有效

提高模型算力。另外，DeepSeek 研发团队以“量子计算与动态数学建模融合”为目标开展下一阶段的研究。通过分解复杂问题，使其成为量子比特态空间，可有望再次提高模型的推理速度。若将其应用到药物研发领域，这种量子化推理架构有望提高分子模拟效率 10-6 倍，或是将 12 年的新药研发周期缩短至 3 年以内。因此，量子化推理架构的构建与应用研究有可能会成为人工智能领域的新发展方向。需要特别注意的一点是量子化推理架构对算力有着极高的需求，所以，要想让量子化推理架构落地，可能还需要先攻克模型算力难关。

3.3 构建人机共生生态

DeepSeek 的突破意味着中国人工智能创新进入新阶段。当技术演进迈入认知革命阶段时，人才成为人工智能领域持续破局的关键。因此，构建人机共生生态成为人工智能领域未来发展的另一大趋势。

一是重构人才战略，破解人才瓶颈。围绕数据洞察、伦理决策、人机协同与敏捷创造四个维度，着力培养复合型人才。二是革新组织形态。推进人工智能未来发展，需要行业企业优化组织配置，通过组建细胞型团队，提高人工智能研发效率，同时，还要依托云平台等数字化手段连接人工智能开发者，以实现资源的跨组织调度。三是落实产业范式。人工智能的研发以应用为落脚点，因此，在研发期间，行业还要关注人工智能产品的商业化落地。以 DeepSeek 为例，它正持续不断地向着垂直行业渗透，致力于从通用性向着专用型转变，如金融领域的国泰君安、中泰证券、兴业证券等多家券商陆续完成 DeepSeek-R1 模型的本地化部署；医疗领域的恒瑞医药等多家医疗公司也相继引入 DeepSeek，致力于打造 AI 医疗模式等。鼓励全球开发者二次开发和创新是 DeepSeek 研发团队发出的新提倡，这种做法，可以推动大模型应用多样化转型与发展，对延续大模型的升级与活力等具有促进意义。所以，在未来，商业化落地可能会成为人工智能领域的新发展趋势。

结语：

综上，DeepSeek 的突破证明了我国人工智能创新进入新阶段，同时，也为人工智能领域的发展提供了新的启示。但是，DeepSeek 的突破并不意味着我国在人工智能领域完全成为“领跑者”，所以，仍需持续纵深研究人工智能创新，不断补齐技术短板。

参考文献

- [1]蔡天琪,蔡恒进.DeepSeek 的技术创新与生成式 AI 的能力上限[J/OL].新疆师范大学学报(哲学社会科学版),2025,(04):1-8[2025-04-25].
 - [2]吴文峻,廖星创,赵金琨.DeepSeek 技术创新与通用人工智能发展趋势[J].科技导报,2025,43(06):14-20.
 - [3]张锦程,李瑞.中国 AI 大模型自主创新形成机制——来自 DeepSeek 的启示[J/OL].西安财经大学学报,1-14[2025-04-25].
 - [4]刘海军,温赞玲.深度求索 DeepSeek: 人工智能、技术创新与新质生产力[J/OL].当代经济管理,1-13[2025-04-25].
- 作者简介:杨知晓(1993.10),男,蒙古族,辽宁沈阳人,初级职称,大学本科,研究方向信息通信技术。