【好文共赏】——

【编者按】

DeepSeek（深度求索）是一款由杭州深度求索人工智能基础技术研究有限公司研发的大型语言模型，拥有强大的自然语言处理能力，能够理解并回答问题，还能辅助写代码、整理资料和解决复杂的数学问题。在刚过去的春节，DeepSeek 不仅引爆了朋友圈，也在国外引发了美国股市与加密货币市场的震荡，一些外媒用“震动硅谷”来形容其带给美国的行业震撼。

许多业内人士认为，DeepSeek 已与 ChatGPT 比肩，低成本、高性能、开源成为海外媒体和业内人士谈论的高频词。它正以其多样化的模式、灵活的提示语技巧和广泛的应用场景，为人们带来前所未有的赋能体验。不仅能提高工作效率，更能够激发创新能力。现将《从 0 到 1 了解 DeepSeek》一文推荐给大家，供学习思考。DeepSeek 不仅仅是一个工具，更是一个能够与人类高效协作的伙伴。通过掌握一些基本的操作技巧和协作意识，可以更好地与 DeepSeek 进行互动，实现人机协同工作的最大化效益。作为与时俱进的市场监管人，让我们尝试拥抱 DeepSeek，激发更多的创新“火花”，共同探索 “AI+监管” 的无限可能，用智慧和勇气创造滁州市场监管更加美好的未来！

从 0 到 1 了解 DeepSeek

一、DeepSeek 公司成立背景与发展历程

DeepSeek,全称杭州深度求索人工智能基础技术研究有限公司,由幻方量化的联合创始人梁文峰创立。公司自 2023 年 7 月年成立以来，始终专注于大语言模型(LLM)及其相关技术的深度研发。公司坚持技术创新路线，开创性地提出多头潜在注意力机制(MLA)和 DeepSeekMoE 等创新架构。凭借这些创新成果，DeepSeek 的大模型在多项权威测评中展现出顶尖的性能表现。

DeepSeek 的团队成员大多来自清华大学、北京大学、中山大学、北京邮电大学等国内顶尖高校整体呈现出“年轻高学历、注重开源、重视创新”的特点。

根据彭博社报道，DeepSeek 的 AI 助手在 140 个市场中成为下载量最多的移动应用。根据 Appfigures 的数据,DeepSeek 的推理人工智能聊天机器人在 1 月 26 日登上苹果公司 AppStore 的榜首并保持全球第一，1 月 28 日起在美国的 AndroidPlayStore 中也位居榜首。根据 SensorTower 的数据，DeepSeek 在发布后的前 18 天内获得了 1600 万次下载，约为 OpenAI 的 ChatGPT 发布时 900 万下载量的两倍，印度贡献了所有平台下载量的 15.6%。

在用户体验方面，DeepSeek 表现不俗。用户普遍认为 DeepSeekR1 的性能出色，特别是在数学推理、编程能力和自然语言理解等领域。其推理速度和准确度在多个测试场景中达到业

界领先水平。此外，DeepSeekR1 的“聪明”特性使得用户无需复杂的提示词技巧，即可获得高质量的回答。在实际使用场景中，无论是游戏、视频播放还是日常工作的辅助，其流畅的操作体验都得到了用户的高度评价。用户反馈显示,DeepSeek 界面简洁直观、操作简单,在实时数据推送和内容推荐上十分出色,能够有效提升工作效率，减少用户在信息检索上的时间投入。

二、DeepSeek:从硬件竞赛到算法效率革命的技术颠覆

大模型在 AI 行业中占据核心地位，是推动技术创新、拓展应用场景及提升行业效率的关键因素。全球范围内的领军企业持续推动大模型性能的提升，随着模型规模的不断扩张，其性能也实现了显著提升。然而，这种规模的扩大也相应地带来了训练和部署成本的急剧增加，成为制约大模型广泛应用的瓶颈。

在机器学习领域，尤其是大型语言模型(LLMS)的应用场景中，模型性能的提升与模型规模、数据集的大小以及计算资源之间存在着紧密的关联，这一关系通常被描述为“规模定律”(ScalingLaw)。根据规模定律，模型的性能会随着模型规模的指数级增加而实现线性提升。目前，国际上主流的大模型，诸如 OpenAl 的 GPT 系列、Anthropic 的 Claude 以及谷歌的 Gemini 等，其最新版本的规模均已突破千亿参数大关。尽管这些模型在性能上展现出了卓越的表现，但对于众多公司和开发者而言，其高昂的硬件资源使用成本、计算时间等依然构成了巨大的挑战。

长期以来，大算力训练一直是基座模型厂商用于融资与构建竞争壁垒的重要手段。

从技术层面来看，GPU 等硬件设施效率的提升以及算法的优化等方式，均有望带动大模型成本的显著下降。在全球 GPU 短缺以及美国限制政策的双重压力下，我国的人工智能公司 DeepSeek 通过算法优化的创新路径，进一步降低了训练成本，为大模型的大规模应用提供了前所未有的可能性。DeepSeek 在 1 月 20 日正式发布了其 R1 模型，并同步开源了模型权重。在第三方的基准测试中，DeepSeek-R1 的表现优于 OpenAI、Meta 和 Anthropic 等美国领先的人工智能公司。在 AIME2024 数学基准测试中，DeepSeek-R1 的成功率高达 79.8%,成功超越了 OpenAl 的 o1 推理模型。在标准化编码测试中，DeepSeek-R1 更是展现出了“专家级”的性能，在 Codeforces 上获得了 2029Elo 的评级，并超越了 96.3%的人类竞争对手。同时，DeepSeek-R1 真正令人瞩目的地方并不仅仅在于其卓越的性能，而在于其极低的成本。它打破了硅谷传统的“堆算力、拼资本”的发展路径,仅用 557.6 万美元和 2048 块英伟达 H800GPU 便完成了性能对标 GPT-4o 的模型训练,成本仅为 OpenAI 同类模型的十分之一，推理成本更是低至每百万 Token0.14 美元，而 OpenAI 的推理成本则为 7.5 美元每百万 Token。

与专有模型不同，DeepSeek-R1 的代码和训练方法均在 MIT 许可下完全开源，这意味着任何人都可以无限制地获取、使用和

修改该模型。全球开发者对 DeepSeek-R1 的贡献代码使其推理效率每小时提升 0.3%，这一开放性的举措极大地激发了业界的创新活力。DeepSeek-R1 在芯片资源利用、算法复杂性和推理速度上实现了重大突破，为 AI 行业的发展树立了新的标杆。

三、DeepSeek:技术特点和优势

高性能与低成本： DeepSeek 在性能上表现出色，其最新发布的 DeepSeekV3 模型在多项基准测试中优于 GPT-4 等主流闭源模型。这得益于 DeepSeek 在算法和工程上的优化，使得模型在保持高精度的同时，显著降低了计算负担。例如，DeepSeek-V3 的训练成本仅为 560 万美元，远低于同类产品的数亿美元投入。

双语处理能力： DeepSeek 在 2 万亿个中英文 token 的数据集上进行预训练，展现出强大的双语处理能力，使得其在处理中英文混合内容或跨语言任务时具有更高的准确性和效率。

混合专家（MoE）架构： DeepSeek 大模型采用混合专家（MoE）网络结构，这种设计使得模型能够智能地选择不同的专家模型进行计算，针对不同的任务激活相应的网络分支，从而实现了计算资源的高效利用。

多模态处理能力： DeepSeek 不仅限于文本处理，还进军了文生图领域。其发布的 Janus-Pro 多模态大模型在 GenEval 和 DPG-Bench 基准测试中击败了 Stable Diffusion 和 OpenAI 的 DALL-E3，展现了其在多模态处理能力上的卓越表现。

四、DeepSeek:创新的 MOE 架构

DeepSeek 的 MOE 架构可以理解为一种“**智能分工协作**”的深度学习模型设计，就像组建了一个高效的项目团队。用日常场景做个比喻来解释:

假设你开了一家广告设计公司，传统方式（普通神经网络）是：1.每个员工都要从头到尾处理每个设计需求。2.即使遇到简单需求（比如改个 LOGO 颜色），也要所有员工都参与。3.处理复杂需求时（比如全套品牌设计），大家又都挤在一起工作。

而 MOE 架构相当于引入了智能分工机制:

动态任务分配（MoE 机制）：有个项目经理（路由机制）先评估每个设计需求的难度: 简单任务只分配给初级设计师（浅层网络），复杂任务才会调用资深设计师团队（深层网络）。

灵活资源调配（条件计算）：遇到需要插画的任务，自动调用插画师团队（特定专家模块）；需要 3D 建模时，则启用建模师小组（动态激活不同子网络）；不同专业团队之间共享基础工具（参数共享），经验传承体系（知识蒸馏）；资深设计师会定期给新人做培训（教师模型指导学生模型）。

把复杂案例的处理经验提炼成标准化流程（知识压缩）。

这种架构的优势

省时省力：简单需求不用惊动整个团队，节省 60-70%人力成本（计算资源）。

专业高效：每个任务都由最合适的专家处理，质量更高（模型效果）。

灵活扩展：新业务来时只需增加对应专家，不影响原有团队（模块化扩展）。

相当于把传统的“全员坐班”模式，升级成了“智能调度+自由职业者平台”的工作方式，特别适合处理多样化、复杂度不均的任务（如自然语言处理中的长文本/多轮对话等场景）。

蒸馏模型

据 DeepSeek-V3 的技术文档，该模型使用数据蒸馏技术生成的高质量数据提升了训练效率。通过已有的高质量模型来合成少量高质量数据，作为新模型的训练数据，从而达到接近于在原始数据上训练的效果 DeepSeek 发布了从 15 亿到 700 亿参数的 R1 蒸馏版本。这些模型基于 Qwen 和 Llama 等架构，表明复杂的推理能力可以被封装在更小、更高效的模型中。蒸馏过程包括使用由完整 DeepSeek-R1 生成的合成推理数据对这些较小的模型进行微调，从而在降低计算成本的同时保持高性能。让规模更大的模型先学到高水平推理模式，再把这些成果移植给更小的模型

五、DeepSeek:深度思考与联网搜索

深度思考:深度思考是一种基于内部知识和经验，通过逻辑推理、分析综合等方法来深入探究问题本质的思考方式。它更侧重于对信息的深度加工和理解，而不是简单地获取信息。在写作、学习和创意生成等场景中，深度思考能够发挥重要作用。

写作：在写作过程中，深度思考有助于作者挖掘主题内涵、构思文章结构、锤炼语言表达等。通过深度思考，作者可以更加清晰地表达自己的观点，使文章更具说服力和感染力。

学习：在学习过程中，深度思考有助于学习者深入理解知识点、建立知识体系、解决疑难问题等。通过深度思考，学习者可以更加透彻地掌握所学知识，提高学习效果和成绩。

创意生成：在创意生成过程中，深度思考有助于激发创新思维、拓展想象空间、寻找灵感等。通过深度思考，人们可以打破常规思维束缚，创造出更具新颖性和独特性的作品或想法。

联网搜索:联网搜索是一种基于互联网的信息检索方式，通过搜索引擎等工具可以快速查找和获取各种信息。在查新闻、找资料、实时数据等场景中，联网搜索具有显著优势。

查新闻：联网搜索可以实时获取最新的新闻报道和资讯信息，帮助人们及时了解时事动态和社会热点。

找资料：联网搜索可以方便地查找各种学术资料、技术文档、历史资料等，为学习和研究提供丰富的信息来源。

实时数据：联网搜索可以迅速获取各种实时数据，如股票价格、天气预报、交通状况等，为人们的决策和行动提供及时的数据支持。

深度思考和联网搜索各有其优势和适用场景。在实际应用中，我们可以根据具体需求和场景选择合适的方式来获取信息和处理问题。目前，DeepSeek 的预训练数据已经更新到 2024 年 7

月。但对于之后的新闻或技术动态，DeepSeek 的联网搜索模式就显得尤为重要，它能够根据网络实时获取最新信息，弥补知识库的空白。

（史学冬、赵艳洁整理）