

科技评论

DeepSeek 引发的 AI 发展路径思考

李国杰

摘要 阐述了因 DeepSeek 横空出世引发的关于 AI 发展路径的思考。首先解释了为什么 DeepSeek 会引起全球性的科技震撼,接着讨论了“规模法则”(Scaling Law)是否已遇到天花板、发展“通用人工智能”(AGI)应选择什么道路、发展人工智能应该追求高算力还是高算效(高能效)、“开源”为什么有这么大的威力等问题。最后对中国在人工智能领域的实力提升和如何实现人工智能自立自强提出建议。

关键词 DeepSeek;通用人工智能;规模定律;开源大模型;低成本

1 为什么 DeepSeek 会引起全球性的科技震撼

DeepSeek 的横空出世是人工智能(AI)发展史上新的标志性事件。7天之内 DeepSeek 的用户增长超过1亿,创造了用户增长速度新的世界纪录。与此同时,芯片巨头公司英伟达(NVIDIA)的股价单日暴跌17%,市值缩水5890亿美元,创下美国上市公司单日最大损失纪录。DeepSeek 的崛起,打破了“高算力和高投入是发展人工智能唯一途径”和“集成电路制程优势=人工智能技术霸权”的迷信,引领人工智能行业进入以算法和模型架构优化为主,同时高度重视数据质量与规模、

理性提高算力的新时期。同时,DeepSeek 的崛起也标志着中国科技公司从“追赶者”变为“规则改写者”,在全球最关注的人工智能领域,以颠覆性的创新开始挑战西方在 AI 领域的霸权。

全球人工智能龙头企业纷纷拥抱 DeepSeek,凸显其不可抗拒的影响力。微软最早宣布将 DeepSeek R1 模型添加到其云平台 Azure AI Foundry,开发者可用于构建基于云的应用程序和服务。亚马逊云科技(AWS)、英伟达、超威半导体(AMD)等公司相继宣布在其 AI 服务平台上部署 DeepSeek V3 和 R1 模型。不管某些国家政府如何抵制,少数媒体如何恶意中伤,上亿用户和众多大公司根据 DeepSeek 的性价比和亲身体验做出选择,主动融入 DeepSeek 生态。DeepSeek 推出的高效率、低

成本的推理模型和开源商业模式,将引领人工智能行业新潮流。

DeepSeek 的 V3 和 R1 模型广受欢迎,首先是因为其在模型算法和系统软件层次都有重大创新。DeepSeek-V3 的模型参数量高达6710亿,但由于采用了自主研发的混合专家模型(MoE)架构,每一层有256个细分领域的路由专家和1个共享专家,每次调用只激活约370亿个参数,显著降低了训练计算成本。DeepSeek 改进的多头潜在注意力机制(MLA),减少了键值缓存开销,把显存占用降到了其他大模型的5%~13%,极大提升了模型运行效率^[1]。DeepSeek-R1 模型摒弃了传统的监督微调(SFT),开创性地提出群组相对策略优化(GRPO),直接通过强化学习从基础模型中激发推理能力,大幅降低了数据

中国科学院计算技术研究所,北京
100190

收稿日期:2025-02-03;修回日期:2025-02-08

作者简介:李国杰,研究员,中国工程院院士,研究方向为计算机体系结构、并行算法、人工智能、大数据、计算机网络、信息技术发展战略等,电子信箱:lig@ict.ac.cn

引用格式:李国杰. DeepSeek 引发的 AI 发展路径思考[J]. 科技导报, 2025, 43(3): 14-19; doi:10.3981/j.issn.1000-7857.2025.02.00183

标注成本,简化了训练流程^[2]。DeepSeek 揭示了一个真相,即推理模型的开发比想象中更为简单,各行各业都可以做。DeepSeek 的这些发明并非首次提出来的原始创新,但 DeepSeek 通过艰苦的努力把技术做到极致,在前人公开成果基础上,登上新的技术高峰。

第3波人工智能兴起后,美国政府、AI 龙头企业和投资界形成一个基本信念:发展人工智能需要高算力,而目前实现人工智能计算性能最高的芯片就是英伟达的 GPU。因此,美国认为只要控制 GPU 的销售,就能在人工智能领域独霸全球。特朗普在就职典礼第2天就签署法案,启动星际之门计划,投资 5000 亿美元,打造人工智能的基础设施。可见,美国政府是把夯实人工智能的算力基础当成维持其全球领导地位的关键。以英伟达为代表的美国 AI 龙头企业,有真老虎的一面,也有纸老虎的另一面。初生牛犊不怕虎的年轻的中国科技工作者,在纸老虎上戳了一个洞,让世人看清了原来这只老虎也没有那么可怕。DeepSeek 震撼全球就是揭示真相的威力。

2 “规模法则”(Scaling Law)是否已遇天花板

2020 年 1 月,OpenAI 发表论文《神经语言模型的规模法则》(《Scaling Laws for Neural Language Models》)^[3],提出规模法则:“通过增加模型规模、数据量和计算资源,可以显著提升模型性

能。”在 AI 领域,规模法则被一些人认为是“公理”,俗称“大力出奇迹”,OpenAI 等龙头企业和美国的 AI 投资界把它当成制胜法宝。

但规模法则不是像牛顿定律一样经过无数次验证的科学定律,而是 OpenAI 等公司近几年研制大模型的经验归纳。从科学研究的角度看,属于一种对技术发展趋势的猜想;从投资的角度看,属于对某种技术路线的押注。人工智能是对未来技术的探索。技术路线存在多种可能,人工智能本身也存在多元化的目标,探索的道路上有很多高山需要去攀登,攀登一座高山的路径也不止一条。把一种信仰或猜想当成科学公理,不是科学的态度。近几年大模型训练的实际效果表明,要获得大模型性能的线性增长,必须在模型规模、数据量和算力投入上高指数性的增长,几个月就翻一倍。从 GPT-3 到 GPT-4,参数规模增加约 10 倍,用于训练的 GPU 数量增加了近 24 倍,总计算量增加了近 70 倍。任何投入的高指数性增加都不可能是长久的,民航大飞机的速度提高、集成电路的主频提高等都是到适当的时候就停止了,大模型也应该不会破例。

鼓吹“Scaling Law”的人,常以强化学习之父理查德·萨顿(Richard S. Sutton)的文章“苦涩的教训”作为追求高算力的依据:“研究人员曾一次又一次试图通过精巧的工程设计来提升性能,但最终都败给了简单粗暴的‘加大算力’方案,历史证明,通用方法总是在 AI 领域胜出。”^[4]但是,

萨顿本人这两年对“Scaling Law”做了深刻反思。他指出,虽然 Scaling Law 在提升模型性能方面确实有效,但它并不是解决所有问题的万能钥匙。AI 系统不仅需要具备强大的计算能力,还需要具备持续学习、适应环境、理解复杂情境等能力,这些能力往往难以通过简单地增加算力来实现。

但现在就说规模法则已经走到尽头,也没有根据。与人脑的神经连接复杂性相比,现在的人工神经网络至少还有上百倍的差距。继续扩大神经网络的规模和增加训练的数据量,是否还能取得与投入相称的回报,要看今后的实际效果。但 GPT-5 迟迟不能问世,可能说明规模扩张的效果已经不太明显。图灵奖得主杨立昆(Yann LeCun)和 OpenAI 前首席科学家伊利亚·苏茨克维(Ilya Sutskever)等直言,规模法则已触及天花板。

DeepSeek 的出现,逼迫 AI 界严肃地思考这一技术发展路线问题:是继续投入巨资追求高算力,还是另辟蹊径,在算法优化上下更多功夫? DeepSeek 的问世标志着人工智能训练模式从“大力出奇迹”的外延式发展阶段,转向集约化系统优化的内涵式发展阶段。DeepSeek 的成功并没有否定算力在人工智能发展中的重要作用。实际上,由于用于推理的设备比训练设备多得多,推理所需要的算力将来会成为主要需求。但绿色发展是必须遵循的大原则,降低人工智能所需要的能耗一定是科技界的重要目标。

3 发展“通用人工智能”(AGI) 应选择什么道路

“通用人工智能”是一个模糊的没有形成广泛共识的术语。OpenAI 公司追求的通用人工智能 (artificial general intelligence, AGI) 是其中的一种,指的是 AI 在多个领域以人类水平处理复杂问题的能力。人工智能界有一个莫拉维克悖论:“复杂的问题是易解的,简单的问题反而是难解的。”从这个角度看,能解复杂问题的人工智能不一定就是通用人工智能。许多人认为,能够应对设计者预料之外的情况,才叫“通用”。因此,人工智能学术界更关注智能系统持续学习,自我改进的能力。人工智能的通用性不仅表现在对语言的处理上,还包括像人一样基于常识和日常经验与外部客观世界互动的能力。

人工智能是对人类智能某一个方面的再现和超越。在科学技术领域,所谓“通用”一定是相对的,有一定的条件或范围。我们要认识人工智能的局限性,不能盲目追求能够解决所有问题的人工智能。重点还是要根据实际需求,将相对通用的人工智能技术落地到各个行业,让一定范围内的人工智能技术见到实效。实现通用智能是一个渐进过程,不会因某项技术的发明就突然到来。人工智能的通用性已经比前两波有明显的提高,但在某些应用中通过图灵测试只是阶段性成果,目前的技术离真正的通用智能还有较大差距。

究竟如何才能实现通用人工

智能,现在还没有结论。DeepSeek 和 OpenAI 都以发展“通用人工智能”为目标,但走的路径不一样。OpenAI 公司相信 Scaling Law 正在尽量扩大模型规模,希望先做出通用的基础模型,再“蒸馏”出各行业可使用的行业垂直模型,走的是“由通到专”的路。除降低通用大模型的训练成本外,如何在保持泛化能力的同时,提高在特定领域或任务上的性能和效率,仍是需要解决的问题。与之相反,DeepSeek 走的是“由专到通”的人工智能发展之路,试图在模型算法和工程优化方面进行系统级创新,为在受限资源下探索通用人工智能开辟新的道路。所谓“混合专家模型”就是集小智为大智,集专智为通智。“小而精”的模型将人工智能的重点发展方向从面向企业的 to B 引向更贴近消费者的 to C,从广泛覆盖的“横向通吃”引向深入钻研的“纵向做精”,让更多的中小企业参与,可能会创造更大市场空间。但整合多个专用模型为通用模型也需要解决诸多技术和工程问题,如模型间的接口、数据格式的统一、训练时的负载均衡等。

通用与专用的竞争是技术发展的普遍现象,集成电路领域就有“通用”与“专用”10 年交替演化的“牧本周期”。“由通到专”和“由专到通”的人工智能,究竟哪条路能走通,还要等历史下结论。也许最后的结果是通专融合,专用多模型的“垂直做精”与通用大模型的“横向扩展”形成互补,共同构建起智能时代的产业新生态。

4 发展人工智能应该追求高算力还是高算效(高能效)

科技界公认图灵是人工智能的奠基人,因为他提出了可以用计算来模拟人类智能的科学假说。他的论文暗示,计算等价于智能。迄今为止,人工智能取得的成果几乎都离不开计算,大模型的出现将算力的作用抬高到前所未有的高度。我们需要认真地想一想,高算力是不是人工智能的本质要求?

发展人工智能的初始动机是模拟人脑,自然界进化了数百万年的人脑是一个计算效率和能效极高的计算装置,功耗只有 20 W 左右。人脑的极低功耗是因为采取了分布式的模拟计算。目前计算机的高能耗是因为采用软硬件分离的数字计算。深度学习的奠基人辛顿(Hinton)教授最近提出“凡人计算”的新研究方向,采用与人脑一样的存算一体模拟计算方式,颠覆了硬件与软件分离的传统计算模式。这类研究追求的是计算的高算效和高能效,从长远来讲,是发展人工智能的正确方向^[5]。

DeepSeek 发布以后,斯坦福大学华裔科学家李飞飞指导的团队,以阿里通义千问(Qwen)模型为基础,通过“蒸馏”谷歌的 AI 推理模型 Gemini 2.0 Flash Thinking Experimental,并结合 SFT 技术,使用 16 个英伟达 H100 GPU 进行了 26 min 的训练,以不到 50 美元的云计算费用成功训练出 s1 模型,其性能超过了 Open AI 的 o1-preview 模型^[6]。这种低训练成

本的模型的通用性可能不如大公司的模型,但令人吃惊的低成本可在某些应用上做到与高端模型相媲美的性能,显示出人工智能的低成本化还有巨大的提升空间。

低成本是技术普及的基本要求,蒸汽机、电力和计算机的普及都是其成本降低到大众可以接受时才做到的,人工智能肯定也会走这条路。目前,盲目地追求高算力导致人工智能的成本居高不下,阻碍了人工智能技术的大规模普及。DeepSeek 不仅是技术突破者,更是规则重构者,开辟了一条低成本发展人工智能的可行之路。DeepSeek 的崛起说明 AI 不再局限于简单堆砌算力,而是进入了以追求高算效和高能效为主的新阶段。

5 “开源”为什么有这么大的威力

过去几年,开源大模型的性能始终与龙头企业的闭源大模型有一代以上的差距,这一次 DeepSeek 的性能追上了闭源模型,大大增强了开源社区的信心。图灵奖得主杨立昆认为,“对 DeepSeek 崛起的正确解读,应是开源模型正在超越闭源模型”。这一评价十分中肯,因为改变 AI 发展模式比单项技术的突破更重要。

虽然 OpenAI 为代表的生成式人工智能红红火火,但多数企业不敢将本公司的数据交给私有 AI 平台生成自己的垂直模型,因为害怕提交的数据泄露自己的技术机密。人工智能之所以在各个

行业难以落地,这可能是一个根本性的原因。DeepSeek 的彻底开源模式破解了这一难题,现在全世界的企业和用户都可以将 DeepSeek 提供的小而精的模型下载到本地,即使断网也可以“蒸馏”出高效率的垂直模型,真正实现了技术的民主化。

长期以来,美国的 AI 龙头公司夸大开源 AI 的安全风险,力图通过监管抑制开源 AI。事实上,开源模型对于全球 AI 供应链至关重要,发展中国家尤其需要开源 AI 技术。若美国继续在这一领域设置障碍,中国就有望在开源 AI 全球供应链中占据核心地位,从而使更多企业转向中国企业而非美国企业的技术方案。

真正的 AI 竞争,不仅仅是技术和模型的竞争,更是生态系统、商业模式,以及价值观的竞争。开源模型让每个开发者都能轻松调用强大 AI 工具,不再受大公司的约束,AI 的进化速度将会明显提升。DeepSeek 的开源战略将向历史证明:在这场 AI 竞赛中,谁拥抱开源,谁就能赢得未来。

6 中国是否已具有在人工智能上引领全球的实力

有人说 ChatGPT 是 0 到 1 的突破,而 DeepSeek 只是 1 到 N 的扩展,这种看法不符合人工智能发展的历史轨迹。人工智能是一个没有严格定义的研究领域,没有智能和不智能的 0 到 1 界限,只有智能化水平不断提高的发展过程。长期以来,中国人工智能领域的高技术企业大多重视应用创

新和商业模式创新,追求的目标是快速盈利,很少参与核心技术创新。随着经济的发展和技术的积累,中国的企业已经开始具备原创的能力。DeepSeek 的一鸣惊人可能是一个分水岭,标志着中国 AI 产业开始从“技术跟跑”向“技术并跑和领跑”迈进。

应当承认,中国在人工智能的基础研究和核心技术上与美国仍然有差距。尽管在人工智能领域,中国的论文发表总数和专利授权数量超过美国,但是引用最多的源头性论文大多出自美国,美国也是顶尖 AI 模型的主要来源国。斯坦福大学发布的《2024 年人工智能指数报告》^[7]显示,2023 年,美国有 61 个较著名的 AI 模型,中国只有 15 个。近几年中国在 AI 领域快速追赶,进步速度喜人。根据《日本经济新闻》对 2020—2024 年神经信息处理系统大会(NIPS)等 3 个机器学习顶级会议的统计,在 3 万多篇发表的论文中,中国作者有 8491 人(美国 14766 人),过去 4 年中国作者增长了 8 倍。

人工智能不同于资本密集型和经验积累型的集成电路产业,不仅要“烧钱”,更要“烧脑”,本质上是拼人的智力的新兴产业。因此人工智能产业具有明显的不对称性,一个具有 100 多个聪明头脑的小企业就可以挑战市值上万亿的龙头企业。DeepSeek 只是中国具有潜力的人工智能企业之一。近日,《麻省理工科技评论》刊发了一篇题为《关注 DeepSeek 之外的 4 家中国人工智能初创公司》的报道,指出阶跃星辰(Step-

fun)、面壁智能(Model Best)、智谱 AI(Zhipu)、无问芯穹(Infinigence AI)4家企业同样展现出不逊于 DeepSeek 的技术实力与全球竞争力。DeepSeek 脱颖而出以后,人们盛传“杭州 6 小龙”的故事(深度求索、宇树科技、游戏科学、云深处、群核科技、强脑科技 6 家初创 AI 公司)。迄今为止,中国共有 52 家人工智能领域的独角兽企业,约占全球人工智能独角兽企业的 18%。这说明在 AI 领域,中国已经有一批创新型小企业进入世界前列,开始展现引领全球的实力。

DeepSeek 的成功说明,在人工智能的发展中,算法优化和系统级工程优化不可或缺,优秀工程师发挥着至关重要的作用。良好的工程教育体系,庞大的工程师队伍是中国的一大优势,我们要充分发挥这一优势。进入并跑状态以后,就不要过于在意与美国相差几个月,可以你做你的、我做我的,比谁能发现正确的研究方向。年轻一代正在成为科研的主力,我们要有信心在人工智能的研究和应用上走在美国前面。

7 中国实现人工智能自立自强要如何发力

实现人工智能自立自强,不仅要靠国家的顶层规划和充分的资金支持,更要做好人才的使用培养和产业生态的构建,克服重重困难的前提是要有自信心。DeepSeek 成功的前提是其创始人梁文锋的自信。他在接受采访时说:

“中国的 AI 不能永远做跟随者,必须有人走在前沿。OpenAI 并非神一般的存在,他们不可能永远领先。”80 后、90 后的中国年轻人已经开始平视美西方国家,具有“敢为天下先”的勇气和自信,他们是中国科技自立自强的希望。

DeepSeek 的人才聘用模式打破了传统,梁文锋选择了一条与众不同的用人策略。他拒绝经验丰富的,而选择初出茅庐的年轻人。招聘时工作经验超过 8 年者,直接拒聘;超过 5 年,需要特别出色才能入选。DeepSeek 团队成员几乎全是国内顶尖高校的应届毕业生或博士实习生。真正的创新往往来自那些没有包袱的人,DeepSeek 重用极致热情、富有好奇心的年轻人,而不是习惯用经验找答案的人,这种用人理念为公司带来了令人惊奇的创新动力,也对传统的中国教育模式和人才聘用模式提出了警示。

要实现人工智能自立自强,最困难的是构建自主可控的产业生态。英伟达公司的“护城河”不是 GPU 芯片本身,而是统一计算设备架构(compute unified device architecture, CUDA)软件生态。DeepSeek 冲击了 CUDA 生态,但没有完全绕过 CUDA,其生态壁垒仍然存在。从长远来讲,需要开发一套比 CUDA 更优秀的自主可控的 AI 软件工具系统,重构 AI 软件生态。实现这个目标需要周密的规划和长期努力,相关部门应该下决心组织全国的开发力量,充分调动上下游企业的积极性,完成这件大事。

资金投入不是决定 AI 成败的唯一因素,但最近几年中国投资市场规模急剧萎缩值得警醒。CB Insights 数据显示,2023 年,美国的 AI 投资达到 672 亿美元,是中国 AI 投资的 8.7 倍。这一年美国的 AI 投资实现了 22.1% 的增长,而中国 AI 私人投资下降了 44.2%。其中,在生成式 AI 私人投资上,2023 年美国总投资额达到 224.6 亿美元,中国仅为 6.5 亿美元。风险投资和私募基金对于支撑科创产业极为重要,他们为创新起到了资金池和担保的作用。中美科创市场曾经并驾齐驱,但到 2023 年,中国科创投资额仅相当于美国的 8%。尽管美国投资界追捧“大算力”,存在一定的泡沫,但正常的金融支持是发展 AI 的必要条件。政府和资本界要合力构建一个健康的科创金融生态,为创新提供必备的动力,这样才会有更多 DeepSeek 出现。

产业生态的形成也要靠市场牵引。国家要通过政策引导,鼓励在 PC、手机和物端设备上推广 AI 应用,增强国产 GPU、CPU 和国产软件的市场占有率。要高度重视芯片设计和大模型的开源战略,争取中国在全球人工智能的开源系统中起到主导作用。我们需要在有限算力条件下,通过算法和软件的协同创新,充分发挥硬件的极致性能,挖掘所有可能的优化空间。国内的人工智能模型与美国的水平十分接近,我们要做好算力资源与人工智能平台的优化适配,力争中国的人工智能科研和应用走在世界前列。

参考文献 (References)

- [1] DeepSeek-AI, Liu A X, Feng B, et al. DeepSeek-V3 technical report[J]. Computer Science, 2024, doi: 10.48550/arXiv.2412.19437.
- [2] DeepSeek-AI, Guo D Y, Yang D J, et al. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning[J]. Computer Science, 2025, doi: 10.48550/arXiv.2501.12948.
- [3] Kaplan J, McCandlish S, Henighan T, et al. Scaling laws for neural language models[J]. Computer Science, 2020, doi: 10.48550/arXiv.2001.08361.
- [4] Sutton R. The bitter lesson[EB/OL]. (2019-03-13) [2025-02-06]. <http://www.incompleteideas.net/IncIdeas/BitterLesson.html>.
- [5] Hinton. 放弃永生的凡人计算[EB/OL]. (2023-06-11) [2025-02-06]. <https://cloud.tencent.com/developer/news/1099853>.
- [6] Muennighoff N, Yang Z T, Shi W J, et al. S1: Simple test-time scaling[J]. Computer Science, 2025, doi: 10.48550/arXiv.2501.19393.
- [7] Stanford HAI. Artificial Intelligence Index Report 2024[R/OL]. [2025-02-06]. <https://aiindex.stanford.edu/report>.

Thoughts on the DeepSeek triggered path of AI development

LI Guojie

Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Abstract This article expounds the author's reflections on the development path of AI triggered by the emergence of DeepSeek. Firstly, it explains why DeepSeek has caused a global technological shock. Then, it discusses whether the scaling law has reached its ceiling, what path should be chosen for the development of artificial general intelligence (AGI), whether the development of AI should pursue high computing power or high computing efficiency (high energy efficiency), and why open source has such great power. Finally, it presents several views on China's strength in the field of AI and how to achieve self-reliance and self-improvement in AI.

Keywords DeepSeek; artificial general intelligence(AGI); scaling law; open-source large model; low-cost ●



(责任编辑 王丽娜)