

DeepSeek持续火爆 中国科技将重塑世界大模型市场格局

本报记者 曲忠芳 北京报道

凭借低成本、高性能在全球范围内“出圈”的DeepSeek持续火爆。据《中国经营报》记者粗略统计梳理,截至2月14日,已有数十家企业宣布接入DeepSeek R1模

型,涵盖云服务商、应用服务商、手机厂商、芯片厂商等,且覆盖不同行业领域。值得注意的是,同样作为AI大模型赛道重要“玩家”的科技大厂也纷纷加入其中,例如:科大讯飞开放平台上线DeepSeek全系大模型;百度文小言(原文心一

言)的最新版本接入DeepSeek R1模型,而且自4月1日起全面免费开放;腾讯元宝同步支持自家的混元模型,以及DeepSeek R1模型,字节跳动旗下飞书的相关功能也接入了DeepSeek R1模型等。

需要指出的是,相比科技大

厂的灵活多元选择,在DeepSeek的冲击波下,以AI“六小龙”为代表的大模型明星独角兽们却略显尴尬。市场格局正在重塑,AI“六小龙”亟须向资本、市场和用户回答“如何保持竞争优势”这一考题。

挑战“六小龙”

得益于幻方量化的资金积累,DeepSeek时至今日并没有进行外部融资,这意味着与“六小龙”相比,DeepSeek并不依赖外部投资“输血”。



资料来源:根据公开信息整理

刘洋/制图

数据机构Xsignal平台显示,DeepSeek的全球日活跃用户数截至2月8日已达到1.19亿。

“江山代有才人出。这对于大模型‘六小龙’来说,可能并不是个好消息。”知名经济学家、工业和信息化部通信经济专家委员会委员盘和林认为,DeepSeek免费开源,效果比肩OpenAI o1,因此DeepSeek的出圈火爆实际上打破了大模型原有的商业模式,简单来说,很多AI模型公司原本希望通过用户订阅付费来发展,在DeepSeek效应冲击下只能转向免费这条路;

而在B端,更多的企业通过开源模

型的本地部署来自我发展AI,这从很多应用厂商、云厂商开放接入DeepSeek也可见一斑。总之,DeepSeek极大地降低了AI赋能的门槛,“更加残酷的AI竞争可能已经到来”。

AI竞争的新格局正在重塑。关于DeepSeek对我国人工智能市场格局产生的影响,上海人工智能研究院数字经济研究中心资深咨询顾问于清扬发文解释,DeepSeek首次验证了强化学习用于模型训练的可行性。与目前常见的监督微调方式相比,强化学习通过设定奖励机制而非依赖人工

标注数据,使模型在不断试错中自我进化,逐渐发展反思、决策、长链推理等关键能力。与此同时,DeepSeek通过强化学习机制将模型的无效训练降低60%,对并行计算的需求较传统架构降低40%,使国产芯片在特定计算任务中的能效比可以达到英伟达GPU的75%。尽管仍面临算力依赖和泛化能力的挑战,但DeepSeek的创新模式已初步证明算法创新可突破算力瓶颈,伴随国产芯片在性能和能效方面的持续优化,未来或将大幅降低国内企业对进口芯片的依赖。

国产AI芯片厂商“抢配”DeepSeek 加速产业链突围

本报记者 谭伦 北京报道

DeepSeek引发的连锁效应下,国产AI芯片成为最先行动的产业圈层。

据最新不完全统计,截至2月13日,已有华为昇腾、沐曦、天数智芯、摩尔线程、海光信息等17家国产AI芯片企业相继宣布支持DeepSeek模型的快速部署和训练,

降低硬件适配门槛

在业内看来,众多国产AI芯片积极支持DeepSeek的重要原因,源于后者打破了以ChatGPT为代表的业界大模型对于超高性能算力芯片的适配门槛。

罗国昭向记者介绍,自ChatGPT诞生以来,其训练依赖于英伟达A100、H100等超高性能算力芯片,主要原因在于英伟达GPU在算力、生态、优化和规模化支持等方面的综合优势,尤其是在算力方面,据OpenAI测算,自2012年以来,AI模型训练算力需求每3—4个月就翻一番,每年训练AI模型所需算力增长幅度高达10倍,而英伟达的单卡

覆盖从训练到推理的全链条场景。

作为大模型训练的硬件基础,国产芯片产业已在过去两年的AI热潮推动下经历跨越式增长。据中商产业研究院测算,2023年,中国AI芯片市场规模达到1206亿元,同比增长41.9%,而到2025年,这一规模将增至1530亿元,年均复合增长率超过30%。

DeepSeek的横空出世,无疑将

这一进程加速。中信证券研报指出,DeepSeek的低成本特性降低了企业准入门槛,推理端需求大幅增长有望推动国产AI芯片在推理场景的应用,一定程度上减少对海外硬件的依赖。银河证券则指出,看好国产算力需求爆发和半导体制造产业链“天花板”抬升。

受此带动,国产半导体也在资本市场获得高关注度。2月12日,

国内目前唯一跟踪中证全指半导体指数的半导体ETF(512480)强势收涨3.39%,成交额突破20亿元大关,累计达20.11亿元。

“DeepSeek的意义,在于能让国产AI芯片获得更多发展的空间,激活基于大模型带来的国产AI生态的壮大。”CHIP全球测试中心中国实验室主任罗国昭向《中国经营报》记者表示。

算力及并行计算能力能够最大程度满足其高需求。

但是,以OpenAI为代表的头部企业长期依赖英伟达GPU的局面,也使得产业链的众多追随者效仿,从而在训练大模型形成了技术和供应链的深度绑定。据机构测算,2023年英伟达GPU在全球AI训练市场的份额超过90%。

而DeepSeek于2024年12月推出DeepSeek V3模型后,于2025年1月20日相继发布了参数规模达6710亿的DeepSeek R1、DeepSeek R1 Zero以及参数范围覆盖15亿至700亿的DeepSeek R1 Distill系列

模型。2025年1月27日,DeepSeek又新增了基于视觉的Janus Pro 7B模型。

据官方测算,其成本效益比同类模型高出90%—95%。这意味着,DeepSeek最新版本的性能比肩OpenAI的顶尖模型o1,但成本却是GPT-4o的十分之一,API调用成本更是只有OpenAI的三十分之一。官方表示,该系列模型通过强化学习一系列创新训练方法,在推理能力方面具有显著优势。

东方证券分析师浦俊懿认为,此前当算力需求主要集中在训练阶段时,英伟达系列芯片以强算

力、高通信带宽和大HBM容量,成为行业的主导性厂商。而在推理阶段,尤其是对私有化部署的AI算力而言,其对计算性能和集群通信速度的追求将低于训练阶段,而DeepSeek模型的算力消耗较低,蒸馏后的小模型需要的闪存容量较小,也与国产AI芯片的情况更加匹配。

因此,其认为,国产算力的性能短板有望在推理需求阶段得到一定缓解,而较低的售价则对用户而言更具吸引力,国产算力芯片以及基于国产芯片的服务器厂商有望迎来更多的需求增长。

增强软硬协同能力

降低适配门槛的同时,国产AI芯片厂商集体接入DeepSeek,也将极大增强中国AI芯片的软硬件协同能力。

在罗国昭看来,业界对于英伟达的依赖,除了其算力性能的领先,还在于其构建的以CUDA为代表的软件“护城河”,而软硬件协同正是实现算法层面优化、更好支持AI算法运行的关键能力。其表示,DeepSeek通过软硬件协同降低算

力成本,从而有助于国产AI芯片实现突围。

“由于国产大模型的竞争此前非常激烈,而适配不同大模型的国产AI芯片也有着不同架构和相应工具链,这使得国产AI芯片的发展显得有些分散且混乱。”罗国昭表示,DeepSeek的出现,则为解决这一挑战提供了窗口和机会。

国信证券分析师胡剑认为,

通过与国产大模型的适配,国内全功能GPU对复杂AI任务的支持能力将得到验证和优化,有助于国内AI全产业链发展。

以国产GPU代表摩尔线程为例。2月4日,摩尔线程便宣布了快速实现DeepSeek R1蒸馏模型推理服务在其GPU智算集群上的高效部署。摩尔线程相关负责人告诉记者,DeepSeek R1 Distill系列模型展示了从较大模型中提取

推理模式到较小模型的有效性,并在各种基准测试中取得了令人瞩目的成绩。这些模型包括DeepSeek R1 Distill Qwen系列,已在摩尔线程的KUAE集群上成功部署,进一步验证了其在实际应用中的实用性和可扩展性。

未来,通过DeepSeek提供的蒸馏模型,能够将大模型的能力迁移至更小、更高效的版本,在国产GPU上实现高性能推理。

200亿估值“光环”下

AI“六小龙”因为大模型投入成本高,盈利前景不明朗,要达到资本预期的投资回报率还有很大的距离,更为重要的是,AI“六小龙”对外部融资的依赖性较高。

曾在2024年获资本青睐的AI“六小龙”将如何应对DeepSeek冲击波?

当记者联系百川智能、智谱AI、阶跃星辰、月之暗面等AI大模型明星公司询问这段时间的产品数据变化以及如何应对DeepSeek引发的行业变化时,多家企业的负责人婉拒了采访,表示不方便回应。其中,月之暗面强调了DeepSeek对于AI行业所带来的整体带动。该公

司联合创始人张予彤通过社交媒体披露,Kimi新版本在春节期间的用户规模创下新高,呈现陡增状态,并感慨AI行业“每逢春节发展速度远超想象”。不过,对于Kimi具体的用户规模及增长数据,月之暗面方面则并未透露。

“DeepSeek的出现,对于大模型初创公司来说既有压力,也是动力。”快思慢想研究院院长、《田丰说》策划人田丰指出,以DeepSeek为代表的开源模型阶段性领先时,会启发、加速闭源模型的研发速度,包括MoE创新性架构、RL自强化学习训练方法、算力网络成本节约思路等。当前DeepSeek具有很强的复杂任务推理能力,又具备联网搜索能力,搜索深度能达到50个相关网页,所以免费版DeepSeek不仅在抢其他to C大模型App的用户,还在抢搜索引擎的用户与广告流量。

DeepSeek开源模型的领先性,改变了全球领先模型都是闭环甚至不公开状态,所以类似DeepSeek R1蒸馏数据给V3模型(或其他模型)会越来越普遍,推动全球大模型水平的快速提升,部分缓解数据瓶颈问题。DeepSeek开源模型的领先性,改变了全球领先模型都是闭环甚至不公开状态,所以类似DeepSeek R1蒸馏数据给V3模型(或其他模型)会越来越普遍,推动全球大模型水平的快速提升,部分缓解数据瓶颈问题。

记者使用DeepSeek统计梳理并进行人工比对核实时注意到,AI“六小龙”在2024年均有融资事件发生,其中智谱AI分别在6月、9月、12月完成了三次融资,最新一轮是30亿元人民币融资;月之暗面则分别在2月、8月各完成一轮融资;2024年12月,阶跃星辰完成B轮融资数亿美元融资,核心投资方包括上海国有资本投资有限公司及

顶着明星光环的AI“六小龙”,谁能在DeepSeek冲击波中率先实现突围,稳住市场地位、吸引更多用户,尚有待时间的验证,本报记者将持续关注报道。

2月5日

DeepSeek R1、DeepSeek V3、DeepSeek V2、Janus Pro正式上线昇腾社区,支持一键获取DeepSeek系列模型,支持昇腾硬件平台上开箱即用,推理快速部署。

2月4日

摩尔线程宣布实现DeepSeek R1蒸馏模型推理服务在其GPU智算集群上的高效部署。

2月3日

海光信息技术团队成功完成DeepSeek V3和R1模型与海光DCU(深度计算单元)国产化适配,并正式上线。

2月3日晚

百度智能云率先宣布,其千帆平台已正式上线DeepSeek R1和DeepSeek V3模型。

2月3日

阿里云也宣布其PAI Model Gallery平台支持云上一键部署DeepSeek V3和DeepSeek R1模型。

2月2日

腾讯云宣布DeepSeek R1大模型已一键部署至其“HAI”平台上。

刘洋/制图

向“生态主导者”转型

此外,值得注意的是,在国产AI芯片厂商纷纷接入DeepSeek的同时,华为云、腾讯云、阿里云等10家国内云计算巨头和至少12家独立云及智算企业,均宣布对DeepSeek的支持。业内广泛认为,这标志着中国AI产业从硬件到软件构建完整生态迈出关键一步。

而生态链的构建,正是中国AI产业突围的关键。正如中国工程院院士、清华大学计算机系教授郑纬民日前指出,近两三年国内企业的AI芯片虽然取得很大进步,但软件生态不成熟是制约我国AI算力产业发展的重要原因。

国科微AI产品总监刘峰向记者表示,从产业角度观察,DeepSeek的出现是好事,在加速技术发展的同时,结束了当前“群魔混战”的局面,这将利于国产AI芯片厂商收敛技术路线,让产业参与者更有方向感。

与主心骨,对产品生态发展更有利。

国内AI相关统计数据显示出,上线仅20天后,DeepSeek应用的日活跃用户数(DAU)便突破2000万大关,达2215万。业内预期,DeepSeek的出现,将使得2025年有望成为国产AI应用爆发元年。咨询机构Omdia预计,未来五年中国生成式AI的市场规模将增长5.5倍,到2029年将达到98亿美元。

中信证券研报指出,DeepSeek通过技术突破和生态开放,有望带动国内AI全产业链的快速发展。DeepSeek为国产AI芯片厂商提供了技术验证和商业化机会,还将帮助云计算公司及运营商实现服务升级与收入增长。其低成本、高性能的特性正在重塑全球AI产业链格局,推动中国从“算力追随者”向“生态主导者”转型。