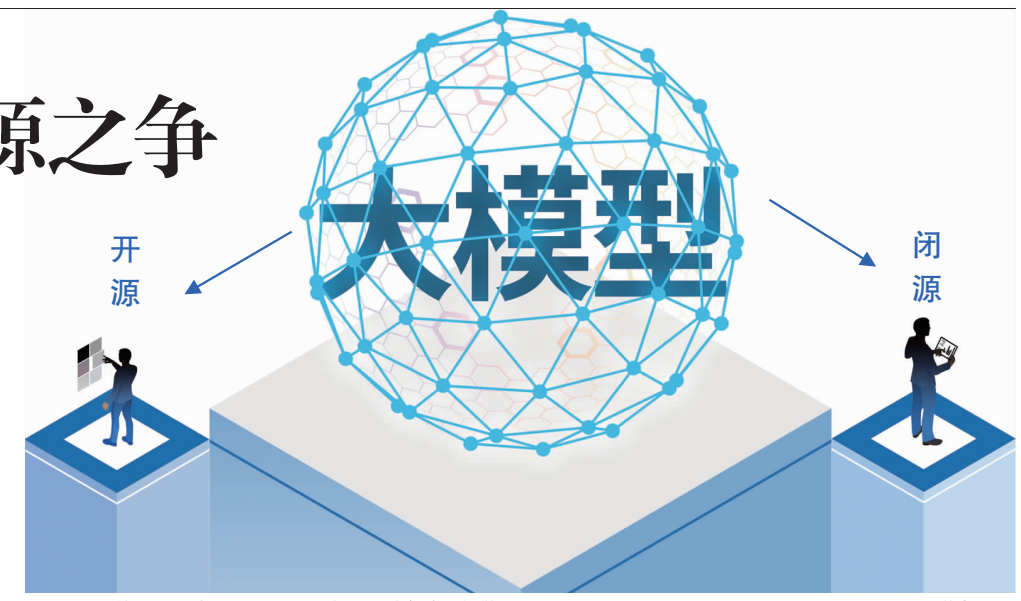
DeepSeek“开源周”点燃大模型开闭源之争

本报记者 秦枭 北京报道

DeepSeek近期举办的“开源周”，宛如一颗重磅炸弹，在全球AI领域激起千层浪。然而，DeepSeek“开源周”带来的影响远不止技术层面，它如同导火索，引发了大模型开源与闭源之争这一行业热议话题。在大模型领域，开源与闭源一直是两种不同的发展路径，各有拥趸，而DeepSeek的开源举动，让这场争论更加激烈。

需要指出的是，无论是开源还是闭源，其最终目标都是推动大模型技术的进步和应用落地。而开源、闭源之争，背后不仅关乎技术的发展路径，更涉及商业利益、数据安全、隐私保护等多个层面的博弈。《中国经营报》记者在采访中了解到，支持开源者认为开源能够促进技术共享与创新，吸引全球开发者共同参与，形成繁荣的生态系统；闭源的拥趸则担忧开源可能导致技术失控，企业难以实现商业变现，影响技术的持续投入与发展。



DeepSeek的开源举动，让大模型开源与闭源的争论更加激烈。 视觉中国/图

阵营

在AI发展的早期阶段，闭源模式凭借其对核心技术的严格把控，在行业中占据着主导地位。

在大模型的发展进程中，开源与闭源宛如两条截然不同的岔路，各自引领着独特的发展方向。这两种模式在技术创新、商业应用、生态构建等多个层面存在着显著差异，也都有着各自的特点与优势。

开源模式，简单来说，就是将软件的源代码公开，允许任何人使用、修改和分发。在大模型领域，开源模式的典型代表有DeepSeek以及Meta的Llama系列。

DeepSeek在“开源周”期间“火力全开”，连续开源五个代码库，涵盖训练、推理、通信等大模型开发的关键环节。从针对Hopper GPU优化的高效MLA解码内核FlashMLA，到首个用于MoE模型训练和推理的开源EP通信库DeepEP，再到支持稠密和MoE模型的FP8计算库DeepGEMM，以及优化并行策略DualPipe和EPLB，还有为应对人工智能训练和推理工作负载挑战而设计的3FS（Fire-Flyer File System）并行文件系统。

DeepSeek开源的一系列代码库，可以让全球的开发者都能够基于这些代码进行二次开发和创新。这种模式极大地促进了技术的创新，因为众多开发者可以共同参与到项目中，发挥各自的智慧和创造力，从不同角度对代码进行优化和改进。

闭源大模型则是由特定的组织或公司开发、拥有并维护其源代码、数据集和技术细节的不对外公开的模型。这种模型就像一座坚固的技术堡垒，保护着开发者的知识产权和商业利益。

在AI发展的早期阶段，闭源模式凭借其对核心技术的严格把控，在行业中占据着主导地位。以OpenAI为例，它通过投入大量的资金与顶尖人才，打造出如GPT系列这样的领先模型。这些模型的源代码被严格保密，仅在内部团队中进行开发与优化。OpenAI利用闭源模式，不仅实现了技术上的快速迭代与领先，还通过商业合作、API授权等方式，将其技术转化为巨大的商业利益。许多企业为了获得先进的自然语言处理能力，不得不向OpenAI购买API服务，这使得OpenAI在商业上取得了巨大的成功，也巩固了闭源模式在行业中的地位。闭源模式还能够保证技术的安全性和稳定性，企业可以对技术进行全面的测试与验证，避免因开源带来的潜在风险。

冲击

作为开源路线的坚持者，DeepSeek的成功被认为是开源模型的胜利。

DeepSeek“开源周”的成功，让行业内的巨头们不得不重新审视自己的开源闭源策略。百度作为国内AI领域的重要力量，此前一直是闭源路线的坚定支持者。百度创始人李彦宏曾多次强调闭源的优势，在“Create 2024百度AI开发者大会”上，他直言“开源模型会越来越落后”；在“2024世界人工智能大会”期间，李彦宏更是表示“开源其实是一种智商税”，他认为闭源模型比开源模型更强大，推理成本更低。

然而，DeepSeek的爆火出圈打破了这一局面。作为开源路线的坚持者，DeepSeek的成功被认为是开源模型的胜利。在这种形势下，百度宣布将在未来几个月中陆续推出文心大模型4.5系列，并于6月30日起正式开源，还宣布文心一言将于4月1日0时起全面免费。

李彦宏在公司2024年第四季度财报电话会上表示，生成式AI基础模型市场仍处于初期阶段，但发展速度非常快，DeepSeek的成功无疑会加快基础模型的应用速度，因为基础模型变得更易获取且成本更低。他认为将最为优秀的模型开源，能够极大地促进应用，当模型开源后，人们出于好奇自然会去尝试，这将扩大模型在更多场景中的影响力。

OpenAI同样受到了DeepSeek开源的冲击。OpenAI前不久推出全新推理模型o3-mini，并首次向免费用户开放推理模型。OpenAI首席执行官山姆·奥特曼在活动中罕见承认OpenAI过去在开源方面一直站在“历史错误的一边”，表示“需要想出一个不同的开源策略”。他称DeepSeek是“一个很好的模型”，并表示OpenAI将生产更好的模型，但与往年相比，领先优势更少。OpenAI首席产品官凯文·威尔也表示，正在努力展示比今天更多的内容，考虑是否开源较旧的AI模型，以适应市场变化并保持竞争力。

天使投资人、人工智能专家郭涛表示，DeepSeek开源后，闭源企业面临着更大的技术追赶压力。开源展示的先进技术使闭源企业原有技术优势不再凸显，它们需要投入资源搞懂开源代码原理并汲取长处，同时维持自身封闭体系下的特色功能。这导致闭源企业面临双重研发任务，时间紧迫。为了应对这一挑战，闭源企业可能需要改变策略，他们可能会加大基础研发投入，补齐短板，确保技术不落后。

争论

未来，开源与闭源模式可能会继续共存。

开源与闭源策略的选择，对大模型厂商的资金投入、技术发展方向以及外界关注的大模型商业化实施进程具有决定性影响。此外，该选择亦会对大模型市场的竞争格局产生深远影响，关乎未来数年的市场发展趋势。行业内的“大佬”也针对大模型的开源闭源展开唇枪舌剑。

360集团创始人周鸿祎，是一位坚定的开源倡导者，他以互联网的发展历程为证，强调没有开源就没有Linux，而没有Linux就没有如今蓬勃发展的互联网。在他看来，开源意味着打破一切界限，无论国家、种族、企业规模大小，只要对人工智能怀揣着浓厚的兴趣，都能投身于开源社区，共享智慧的结晶。这种开放性和包容性能够形成一种强大的虹吸效应，吸引全球的人才和资源汇聚于此。

他对DeepSeek的开源模式给予了高度评价，认为DeepSeek通过开源策略，成功建立了全球开发者生态联盟，成为行业的事实标准，奠定了AI“根技术”的地位。周鸿祎预测，开源模式将重构AI竞争格局，中国有望凭借开源生态的优势在AI领域保持长期领先。他还指出，开源模式将带来多赢的局面，中小企业能够以低成本获得顶尖的AI能力，云服务商可通过算力需求的激增获益，国产芯片厂商则有机会借推理算力优化实现弯道超车。

在红帽大中华区首席架构师张家驹看来，DeepSeek的开源模式的成功，印证了开源将加速创新，同时也有助于标准的形成，使AI更安全。

除此之外，阿里云CTO周靖人重申了阿里云开源开放的选择，他表示通义千问已经实现了真正意义上的全尺寸、全模态开源，拉平了开源、闭源模型之间的差距，通义千问开源模型下载量的增长和阿里云百炼服务客户数的大幅增加，证明了开源策略在阿里云的成功实践。

而月之暗面创始人杨植麟则认为，闭源会带来人才和资本的聚集，最终闭源会更具优势，他以海外基于开源扩散模型Stable Diffusion的应用为例，指出虽然有众多应用，但却没有一个能够脱颖而出。

萨摩耶云科技集团首席经济学家郑磊认为，大模型开源相比闭源，在技术创新速度上具有显著优势，能够通过社区协作和众包创新加速技术扩散和应用。同时，开源模式能够像DeepSeek一样，快速推动计算、通信、存储等多领域的协同创新。然而，开源模式也存在质量控制、安全风险等劣势。未来，开源与闭源模式可能会继续共存，企业会根据自身需求选择合适的策略。

DeepSeek重塑国产AI生态圈

本报记者 李玉洋 上海报道

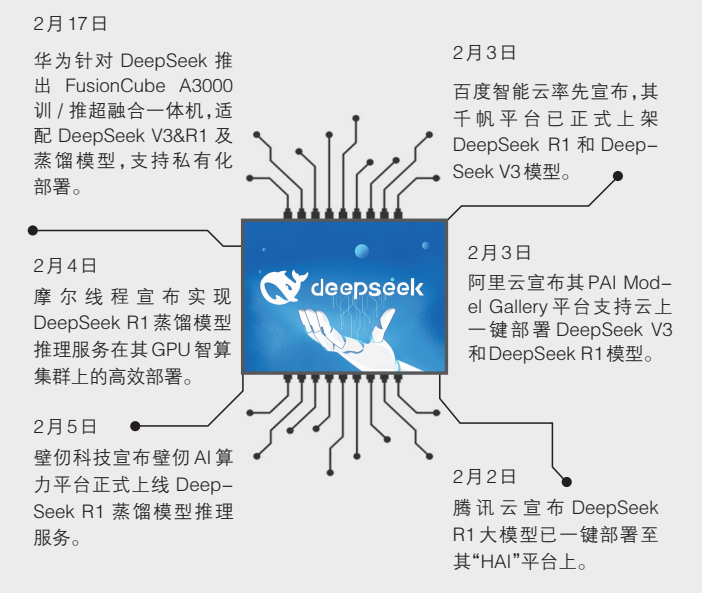
通过“开源周”以及公开发布V3/R1大模型的推理系统技术介绍，DeepSeek成为AI技术圈和开发者的“开源之神”，已经被昵称为DeepOpen。

当DeepSeek的开源代码如蒲公英种子飘向世界，国内AI芯片行业是否能借此东风，迎来属于自己的春天？

“（DeepSeek的开源）对整个AI行业都有一定程度的推动。”行业研究机构Omdia人工智能首席分析师苏廉节告诉《中国经营报》记者，AI芯片厂商通过这些开源代码更加了解DeepSeek大模型的架构和特点，进而做出相对应的优化，特别是算力的配置、训练推理的架构、存储资源的需求等。

摩尔线程和壁仞科技这两家国内AI芯片的代表厂商都在接受采访时表示，DeepSeek这种低算力需求的大模型，对国产AI芯片的发展是一个重要机遇。

记者还注意到，随着DeepSeek的出圈，国产算力迎来火爆行情，一大批一体机密集上线，由此出现“2025是一体机元年”的观点。“一体机今年火起来主要是因为DeepSeek，很多政企客户都想把DeepSeek用起来。”容联云大模型产品负责人唐兴才表示，市面上目前满血版DeepSeek大模型一体机售价约为200万元。



吴双/制图

盘活国产AI生态

在“开源周”上，FlashMLA是DeepSeek专为英伟达Hopper架构GPU（如英伟达H100/H800）优化的注意力解码内核，已投入生产，现在被视为提升显卡潜力的“加速器”。DeepEP则是首个用于MoE模型训练和推理的开源EP通信库，可以直接调用Hopper GPU的TMA张量内存加速器，被称为大模型训练的“通信管家”。而DeepGEMM是一个优化矩阵乘法的工具，实现FP8低精度下的1350+TFLOPS算力，代码仅300行，被称为矩阵计算的“省电小能手”。DualPipe主要用于解决流水线并行中的“等待时间”问题；比如，多任务步骤速度不一时，其能双向调度，减少空闲时间。EPLB则用于自动平衡GPU负载，当某些AI专家模型任务过重时，会复制任务到空闲显卡，避免“忙的忙死，闲的闲死”。最后的是3FS，被称为数据处理的“极速组合”，采用了分布式文件系统，利用高速存储和网络技术（如SSD、RDMA），让数据读取速度达到每秒6.6TB。

值得注意的是，DeepSeek在包括上述开源项目中直接调用比英伟达CUDA更底层的指令PTX（Parallel Thread Execution，一种底层硬件指令集，用于直接与GPU驱动函数进行交互，实现更为精细的硬件操作、优化TMA加速器等），显示出DeepSeek对于GPU微架构的深度了解。这种能力通常为芯片设计团队所独有。

苏廉节也表示，DeepSeek团队对GPU硬件底层技术的理解力很强，这在大模型行业并不多见。甚至有消息传出，DeepSeek在寻找芯片设计人才，想要做自己的芯片。对此，苏廉节认为，目前140人的DeepSeek团队要做芯片设计很困难，但它背后的幻方量化所在的金融领域确实有定制化芯片的需求。

目前，摩尔线程已实现对DeepSeek开源周“全家桶”的支持，涵盖FlashMLA、DeepEP、DeepGEMM、DualPipe以及Fire-Flyer文件系统（3FS）；壁仞科技在“开源周”之前就已经实现对FlashMLA、DeepGEMM、DeepEP等核心模块类似功能和优化技术。

事实上，春节期间已有多家国产芯片企业陆续宣布对DeepSeek模型的适配或者上架服务，包括华为昇腾、沐曦、天数智芯、摩尔线程、海光信息、壁仞科技、云天励飞、燧原科技、昆仑芯等。

“通过‘开源周’，更多人尤其是开发者看到了DeepSeek的优势和如何去进行调优和适配。”苏廉节认为，国内AI芯片厂商可以从DeepSeek的开源代码库中看到和进一步了解底层的哪些代码对未来的适配性有帮助。

“比如DeepEP是一个专门为混合专家模型开发的并行通信技术，需要芯片厂商支援。”苏廉节表示，芯片厂商因此会开发相对应的工具，让开发者能更顺畅地进行代码转移和应用支撑。

摩尔线程方面则认为，DeepSeek的开源模式为国产AI芯片厂商提供了与软件开发者合作的机会。“通过与DeepSeek为代表的开源模型的合作，国内AI芯片厂商可以更好地理解AI应用的需求，进行针对性优化；国产模型+国产芯片可以形成完整的AI闭环，加速国产AI生态的发展进程。”

“短期内，国产GPU厂商应保持训练芯片的持续迭代，比如最好支持FP8，确保技术不脱节，同时通过推理芯片快速切入商业化场景。”摩尔线程方面还表示，长期来看应该瞄准“训（练）推（理）一体”架构，通过统一计算平台降低客户切换成本，最终在自主生态中实现训练与推理的协同增长。

激活一体机市场

摩尔线程方面还提到，DeepSeek大幅降低AI成本，让AI更加普及，反过来又会提升行业对算力规模的需求。

中信证券研报指出，算力算法联合优化带来的降本让人们看到AI应用落地的更多可能，同时杰文斯悖论有望支撑长期推理算力需求。杰文斯悖论指的是，当技术进步提高了使用资源的效率，但成本降低导致需求增加，底层资源的消耗量反而提升。

根据《DeepSeek-V3/R1推理系统概览》一文，DeepSeek算了一笔账：“假定GPU租赁成本为2美元/小时，总成本为87072美元/天。如果所有Tokens全部按照DeepSeek R1的定价计算，理论上一天的总收入为562027美元，利润率545%。”

如此高的成本利润率，让中小厂商在技术平权之下迎来降本机遇。

记者注意到，当多地政府宣布政务系统接入DeepSeek、一大批AI公务员上岗时，DeepSeek一体机也颇为火爆。据不完全统计，至少已有华为昇腾、中科曙光、浪潮、新华三等60余家厂商，在加速部署一体机。

“一体机一直都在的，只是DeepSeek非常适合本地化部署。”苏廉节指出，一体机并不是新产品，就是一个结合算力、存储和网络的小型数据中心，“主要由几个小型服务器构成，用于边缘侧小规模的商用场景”。

在唐兴才看来，大模型一体机是把大模型和硬件（如CPU、GPU、存储设备等）结合，封装为一体化设备。“一体机客户目前来看主要是国央企、政府、金融机构这些对隐私安全要求比较高的客户。”唐兴才说。

据唐兴才观察，大模型一体机市场玩家主要可分为系统集成商、应用厂商、模型厂商和GPU资源厂商，具体有华为、联想、阿里巴巴、百度、浪潮、新华三、中科曙光等。

“我们主要是大模型应用厂商，会和硬件厂商一起做一体机。因为客户想要的是模型+应用场景。”唐兴才表示。

“相比传统的云方案，大模型一体机具备私有部署、交付便捷、算力门槛低和稳定性的优势，能够满足金融、能源、政务、医疗等数据敏感型行业对于安全和隐私的要求。”摩尔线程方面表示。

据市场反馈，DeepSeek一体机的价格从几十万元到数百万元不等。有创业公司表示，“满血一体机”价格在150万—200万元。

唐兴才表示，200万元一般能跑满血版DeepSeek一体机。而另有大模型公司人士表示，一体机售价通常包括硬件+软件，硬件毛利率约为15%，软件毛利率在40%左右。

据浙商证券测算，随着DeepSeek快速部署需求的增加，一体机的市场需求有望显著增长，预计2025—2027年，一体机需求量将分别达到15万台、39万台和72万台，未来三年DeepSeek一体机市场空间有望达到1236亿元、2937亿元和5208亿元。

唐兴才所在公司刚开始推一体机，市场反响还可以。“现在还看不清，等几个月看看吧。”他对市场前景谨慎乐观。