

贯通国产“七芯” 高效撬动算力

申城AI基础设施企业:让“满血版”DeepSeek更适配好用



本报记者 邵阳

DeepSeek 影响力持续“破圈”，可“甜蜜的烦恼”是，官网总是“服务器繁忙”。

好消息是，“满血版”DeepSeek-R1.V3正在得到越来越多第三方平台的支持。

11日，上海AI基础设施企业无问芯穹宣布获七家国产芯片企业鼎力支持，正打通DeepSeek-R1.V3在壁仞科技、海光信息、华为昇腾、摩尔线程、沐曦、燧原科技、天数智芯等7

个硬件平台的多芯片适配优化，现开发者已可通过无问芯穹Infini-AI异构云平台一键获取DeepSeek系列模型与多元异构国产算力服务。

自春节前夕爆火，DeepSeek已牵动国内超30家云服务商与近20家芯片企业宣布接入。随着无问芯穹宣布获7家国产芯片企业支持，并带着多芯片适配的DeepSeek-R1入场，DeepSeek为可控国产算力撬动的历史机遇，正变得愈发清晰。

“问题作答”到底谁更强？

DeepSeek是绝佳商业化试炼场，激发了大量推理算力需求。几乎主流云厂商均已宣布提供R1模型服务，但全网找到服务不繁忙的“满血版”还是不容易。

以前，国产AI芯片都号称支持大模型部署，奈何实际使用者并不多。现在，DeepSeek相当于给AI芯片公司出了同一道考题：你到底“来赛”吗？

短时间内，20家国产AI芯片企业争相适配或上线DeepSeek模型，以期抓住这千载难逢的增长机遇。这恰恰反映当下AI基础设施的特点：芯片和算力资源是有，但没被高效利用起来。

如何将大量不同架构的国产芯片变得能用、好用，并在使用过程中形成硬件与算法之间的正向循环？无问芯穹的策略是做“异构云”，即统筹模型和算力需求，把分散在不同地方的异构算力整合起来，实现不同模型和不同硬件之间的统一部署和联合优化，提供给需要算力和模型的客户。

去年7月，无问芯穹就在业内首次实现了千卡规模的4+2种芯（天数智芯、华为昇腾、沐曦、摩尔线程与AMD、英伟达）异构混训，集群算力利用率最高达到97.6%，其运营算力已经覆盖全国超13座城市。

“当下的生态环境，需要一个从算法、硬件到模型都深入理解的团队，无问芯穹便具备这样的全栈垂直优化能力。”无问芯穹联合创始人、CEO夏立雪表示。

加速全国产AI产业闭环

夏立雪说，DeepSeek的突破，激发了越来越多的下游应用创造力，未来行业日均tokens消耗量将达到百万亿级别，不仅将激发国产芯片的市场需求，也为打造全国产AI产业闭环，实现更可控的自主算力发展创造了有利条件。

记者了解到，在美国，模型、系统、芯片三个关键因素已经形成闭环生态；而国内多数大模型是通过国外芯片（如英伟达）训练得到的，与国内的系统、芯片难以形成闭环。

“DeepSeek作为开源模型，其之于AI 2.0时代，正如Android之于移动互联网革命，将重构整个产业生态，引发链式反应，加快上层应用发展和下层系统‘统一’增速，由此广泛调动起跨越软硬件和上下游的生态，一起加大投入‘国产模型-国产芯片-国产系统’的协同优化和垂直打通。”夏立雪说。

对此，无问芯穹提出“三步走”模式，来促进全国产AI产业闭环的打通——基于主

流芯片开展极致软硬件协同优化，以有限算力实现国产模型能力追赶国外模型；推动国产芯片开放底层生态，搭建“异构”AI系统解决算力缺口，实现模型能力赶超；构建国产“同构”系统，支持Scaling Law持续发展，打造“国产模型-国产芯片-国产系统”的全国产AI产业闭环，实现更可控的自主算力发展。

夏立雪表示，在劈开算力封锁的路上，DeepSeek踏出了成功的一步，接下来的“国产模型+国产算力+国产云服务+国产应用”全国产化AI之路，还要靠产业链上下游的生态伙伴合力闯出来。

软硬协同优化是必然路径

DeepSeek大语言模型共发布了3个大版本，参数规模从67B扩大到236B再到如今的671B，所用的训练算力并未与模型尺寸等比例地成倍增长，其核心便是利用软硬件协同设计的思想，不断提升系统开销——即模型的训练成本到模型尺寸、能力的转化率。

记者了解到，软硬协同优化是连接硬件和算法的桥梁，区别于各种算法层面优化方法，和硬件本身迭代的方法，而是随着硬件变化，让软件也不断优化，或根据软件的迭代来定义未来的底层电路实现，以更好地支持AI算法的运行。

业内人士指出，DeepSeek通过软硬件协同降低算力成本，给国内的启示在于这套方法论可以打破现在的闭环生态瓶颈。

无问芯穹核心技术团队早在2018年就总结出了“深度学习算法—编译—芯片联合设计”路线，致力于人工智能场景下能效电路和系统设计方法的研究，是软硬协同路线上的早期开拓者之一。自成立以来，依托“多元异构、软硬协同”的核心技术优势，打造了连接“M种模型”和“N种芯片”的“M×N”AI基础设施新范式。

蛇年春节，DeepSeek和《哪吒之魔童闹海》两匹“黑马”，扭转了中国企业技不如人的成见，也证明了只要实力强，自有不分国界的国际掌声。前两天，本报报道了《国际权威科技商业化智库关注DeepSeek之外中国人工智能初创公司 四大AI新锐三家已入驻模速空间》，其中就包括了无问芯穹。有分析认为，DeepSeek正在让国际意识到，除了DeepSeek之外，中国还存在若干路径独特的团队，在用“力大砖飞”堆叠GPU算力储备、比拼模型精度的巨头竞赛之外，找到未被五角大楼战略家们关注到的“盲点(blind spot)”，跨越软硬件，撬动上下游，走出可能出奇制胜的技术路径。



新民随笔

有了“深度求索”更需“深度思考”

朱光

向“深度求索”(DeepSeek)提问:怎样才能超越你?它的回答有8点,概括起来就是持续学习,提升自己的专业能力;创新思维,因为创新是技术进步的关键;积累实践经验,这才是检验和提升能力的最好方式;掌握跨学科知识,能找到新的突破点;团队合作,集思广益的力量超过个人;具备伦理意识,确保技术进步是为了人类福祉;利用工具和资源,可以加速学习和开发过程;不断反思和改进,才能确保不断进步。结论:“通过以上方法,你可以在技术和知识上不断进步,最终超越现有的AI系统。”嗯,它“聪明”到在“AI系统”前加了一个时间限定“现有的”。一个人,看完这8点需要5分钟;而它生成这8点,才十几二十秒。反思一秒钟,就会发现它的答案逻辑严谨且充满善意。

“深度求索”问世后,似乎在挑战各行各业。有人问考研教师张雪峰:“考学的事情,DeepSeek都能帮我分析了,那么你会不会失业?”他答:“首先,人工智能不提供情绪价值;其次,它分析的都是基于公开信息,不是所有的信息都会同步公开。”张雪峰的回复有其合理之处,且基于其职业定位。

那么,什么工作一定会被替代呢?可以肯定的是,根据明确规则重复劳动的岗位被替代的可能性较高;其次,AI可以辅助分析大量数据,从而提高效率的岗位也堪忧。它自己也分析道:“虽然可以辅助教学或做出个性化诊断,但是医生和教师的角色依然重要,因为需要人文关怀和现场判断。”

“深度求索”的逻辑如此清晰,“倒逼”人的理性和感性必须同时提升——提升人的理性,有助于尽量与AI的理性同步;提升人的感性,有助于在与AI“竞争”时保持“领先”地位。因为,AI目前最缺乏的还是情感。它可以模仿雨果这样的浪漫主义作家的文风,但是它创造不出雨果这个人。在这个“倒逼”的过程中,其实“深度求索”逼出的是我们的“深度思考”(DeepThink)能力。

当机器变得几近“万能”,那么人更应该凸显人脑的“专能”——不可替代的,是创造力与想象力、价值观与伦理观,以及如何赋予意义。因而,艺术家、创意人一旦经过“深度思考”后结合了AI的高效工具,势必立于不败之地。

免费参观

上海 SJ-MART 2025

免费领珍珠胸针

上海国际时尚珠宝消费节

SHANGHAI INTERNATIONAL FASHION JEWELRY MART 2025

2025年02月14-17日

上海正大广场西区·9楼正大厅

扫码免费领取
电子入场券

先鉴定 后购买
展会检测机构:珠宝国检-NGTC

► 全品类珠宝展示 / 万千珠宝齐聚魔都!

展览地址 / 上海市浦东新区 陆家嘴西路 168号
主办单位/中禹展览(上海)有限公司

广告