

大模型创新的另一种可能

□ 陈杰

热点观察

自ChatGPT横空出世,算力便被视作大模型产业创新的核心驱动力。这一趋势使得GPU(图形加速芯片)的地位急速攀升,甚至成为AI产业发展的关键制约因素,也就是我们常常听到的“卡脖子”。

然而,这种趋势却被我国一家成立仅一年半的人工智能创业公司深度求索(下称DeepSeek)终结。近日,DeepSeek仅凭借2048张英伟达GPU加速卡,在短短两个月内成功训练出拥有6710亿参数的开源大模型DeepSeek-V3,着实让刚高呼着进入“万卡”时代(1万张以上的加速卡组成的算力系统)的AI大模型业界惊掉了一地的“下巴”。

从测评数据来看,相对于全球公认的顶级大模型GPT-4o,DeepSeek-V3的训练成本仅为其二十分之一、输入+输出的售价也只有GPT-4o的十分之一,性能已然赶超GPT-4o。

还处在研发阶段的GPT-5,或许更强。但OpenAI立项GPT-5已经一年半有余,目前也只做了两轮训练,且每轮训练仅计算成本就接近5亿美元。这也意味着,人们所期待的新一代通用大模型训练成本已经达到十数亿美元,甚至还会更高。产业界估计,按照这种传统的训练路径,未来3年内大模型的训练成本将上升至100亿美元,甚至上不封顶。

这种大参数、大算力、大投入的研发模式,对于实力稍弱的AI企业来说,无疑是一道难以逾越的门槛。即便能“组局”成功,这种成本结构下产出的大模型产品,又有多少人能用得起呢?

再来看DeepSeek-V3,它证明了即使在硬件资源有限的情况下,通过数据与算法层面的优化创新,依然能够高效利用算力,实现一流的大模型性能,同时还能大幅降低应用成本。其实,早在去年5月发布的DeepSeek-V2,就成功以“价格屠夫”的方式杀入市场,让一众大模型产品不得不跟进降价。“AI界拼多多”的称号,

绝非浪得虚名。

性价比方面,DeepSeek已经甩了OpenAI等业界大佬好几条街了。在AI领域,一直有这样的说法:硅谷企业擅长从0到1的创新,而中国企业则在从1到10的应用拓展上更具优势,因为中国拥有广阔的应用市场。

对于大模型产业而言,性价比已成为阻碍产业规模化应用的关键因素,DeepSeek的出现为解决这一难题提供了新思路。

当前,AI大模型仍处于技术创新的爆发阶段,距离应用的全面爆发还有一段路要走,这也意味着产业发展充满变数与机遇。不可否认,今天的英伟达市值逆天、GPU产品也供不应求,但如果DeepSeek的成功可以复制,产业界不再执着于万卡集成的研发路径时,GPU的主导地位还能持续多久呢?

大模型创新的另一种可能,不仅出现在AI产业,通信、芯片、量子计算等前沿科技领域也正悄然发生着改变。科技创新的魅力在于,它永远蕴含着无限的可能性,并不断突破边界。

资讯品读

TikTok在美续存现转机

1月20日,TikTok(海外版抖音)在美国停服仅半天,就已开始正常提供服务了。

这得益于美国当选总统特朗普宣布延长了TikTok“不卖就禁”法律生效前的时间,以便能够达成保护国家安全的协议,该命令发布前,任何帮助TikTok继续运营的公司将不承担任何责任。同时,特朗普计划设立合资公司,美国在其中拥有50%的股份,这样既确保美国可以掌握TikTok,也能令其继续在美运营。

不过,有消息人士称,TikTok母公司字节跳动并未与美方达成任何协议。此次恢复运营,只是因为美方决定暂不执行法案。

点评:由此次停服引发的“TikTok难民”迁徙潮中,中国社交App小红书无疑是最大赢家,仅两天就获得了70万美国人注册。目前,小红书已经上线一键翻译功能来承接这波天的流量了。

今年春晚AI含量高

中央广播电视总台近日宣布,2025年春节联欢晚会将首次进行视障版和听障版无障碍转播。

据了解,在春晚视障版和听障版无障碍转播中,手语演员将通过增强现实技术进行手语“解说”,辅以AI语音技术制作字幕,满足听障人士的观看需求;同时,还将通过音频对春晚节目进行创作解读与背景拓展,让视障人士更好地“听”春晚。

此外,今年总台春晚将首次采用全国产8K装备完成“8K超高清视频+三维菁彩声音频”制作;首次在春晚演播大厅顶部部署由30台相机组成的云上多视角拍摄系统,以“演播厅天空环绕视角”摄制,利用立体视觉AI算法对多机位信号进行实时三维点云渲染,并采用智能虚拟运镜手法,呈现出摄像机难以拍摄的影像画面。

点评:有了AI技术和8K超高清装备的加持,已经42岁的央视春晚晚会正在焕发新生。

三星联手OpenAI推AI电视

近日,有消息称三星将与OpenAI建立“开放式合作伙伴关系”,合作开发AI电视。

三星会借助OpenAI的文本生成功能ChatGPT、图像创建功能DALL-E、语音识别功能Whisper,以及其他OpenAI的AI模型,有望开发出功能创新的AI电视,包括提供个性化内容推荐、对话式助手、实时翻译、定制娱乐等功能。比如,用户可以通过与电视语音交互,获取正在观看的演员或电影的详细信息,以及类似节目的推荐,获得更加个性化和便捷的观看体验。

点评:从文本到图像生成,AI已经通过无数的软件和硬件迅速进入我们的日常生活中。近半年,AI生成视频持续刷屏,作为视频最佳载体之一的电视,理所当然会加入这场席卷时代的技术变革中。

AI为地震预测研究带来新契机

□ 石玉若

近来,全球范围内地震活动频繁,让地震预测预警相关话题的热度飙升。不少人疑惑:地震为什么不能像天气一样实现精准预报呢?

地震,这一自古以来便伴随着人类生活的自然灾害,对其的预测至今仍是一个世界性的科学难题。尽管困难重重,但科学家从未停止过探索的脚步。近年来,随着AI技术的兴起与应用,地震预测研究也迎来了新的发展契机。

AI技术带来了什么

虽同为自然现象,但地震和天气所依据的预测原理截然不同。

地震是地壳快速释放能量过程中造成的振动,它的产生与地球内部的地质构造活动密切相关。其孕育过程复杂且大多发生在地下深处,监测难度极大。相比之下,基于大气的运动、温度、湿度等相对容易观测的参数,再加上大数据和AI技术的助力,天气情况如今已经能较为精准地预测了。

不过,AI为天气预报的精准性提供了技术支撑,也为地震预测研究带来了新的可能。

AI在地震预测中的应用,就像是给科学家们配备了一个“超级大脑”。它能够快速处理和分析海量的地震波数据,效率远超人工。比如,在2022年四川甘孜州泸定县6.8级地震中,AI自动检测到的余震数量是人工目录的3.3倍。这不仅提高了监测效率,还让科学家们能从更多数据中寻找地震的蛛丝马迹。

AI还能在时间和空间上“大显身手”。它能整合不同监测站点的数据,评估一个区域的地震风险,还能根据信号变化趋势推测地震可能发



我国已建成全球规模最大地震预警网,能提供数秒至数十秒的地震预警时间。
视觉中国供图

生的时间。全球首个亿级参数地震波大模型“谛听”,已经在地震信号识别和余震监测中取得了突破,未来有望进一步提升地震预测的精度。

此外,AI技术还可以用于地震危险性分析、抗震设防优化等方面。通过对历史地震数据的深入挖掘和分析,科学家们可以更好地了解地震活动的规律和特点,为防震减灾工作提供科学依据。

“地震预报”能成真吗

尽管AI在地震预测研究中展现出了巨大的潜力,但我们也必须清醒地认识到,地震预测仍然是一个世界性的科学难题。以目前的技术手段,还无法实现在地震发生前准确地预测其具体时间、地点和强度。

但是,AI技术在地震发生时,能为人们争取宝贵的逃生时间。基于AI加持的手机地震预警,就是一项已经开始广泛应用的重要技术手段。

地震发生时会产生P波和S波。P波是纵波,传播速度较快,对人类影响较小;S波是横波,传播速度较

慢,但破坏性更强。AI能够快速识别并分析这些复杂的地震波信号,在地震波传播的短暂时间内,通过强大的运算能力,能迅速判断地震的震级、震源等关键信息,并及时将预警信息发送到手机端。在一些地震频发地区,基于AI的手机地震预警系统已经多次在地震发生时,提前几秒到几十秒向民众发出警报,为人们提供宝贵的时间窗口,减少地震带来的损失。

AI技术加持之下的手机地震预警,只是地震预测领域阶段性的成果,并非真正意义上的地震预测。对于网络上时不时“精准预测地震”的传言,我们要保持警惕,避免上当受骗。

当前,人类对地震孕育机制的认识正在逐步加深。随着科技的不断进步,如果人类能够进一步弄清楚地震发生前各种物理化学变化的规律,并结合多种监测手段和先进的AI大数据分析,未来一定程度上的“地震预报”或许真有“梦想成真”的可能。

(作者系中国地质科学院地质研究所研究员)