

帕克探测器创人造物体“触摸”太阳距离新纪录

科技日报北京12月30日电 (记者刘震)据美国国家航空航天局(NASA)官网28日报道,该机构的帕

克太阳探测器24日来到距离太阳表面608万千米的位置,创下人造物体接近太阳表面新纪录。27日,探测

器传回的信号证实,它经受住了太阳表面极端环境的考验,目前运行正常。

美国约翰斯·霍普金斯应用物理实验室(APL)负责帕克太阳探测器的建造和运营工作。该实验室表示,此次探测器与太阳如此近距离接触,使其能对太阳进行更精确的科学测量。这将加深科学家对这颗恒星的理

解,并帮助预测可能影响地球生命的空间天气事件。据悉,当时帕克太阳探测器以约69万公里/小时的速度穿过太阳大气层,超越了此前任何人造天体的移动速度。这一速度足以在一分钟内从美国华盛顿飞抵日本东京。

自2018年发射以来,帕克太阳探测器先后7次借助金星的引力进行加速,从而得以更近距离靠近并观测太阳。在靠近太阳的过程中,探测器依

靠碳泡沫隔热罩来保护自身免受日冕(太阳大气的最外层)极端高温的侵袭。

APL团队表示,帕克太阳探测器飞越日冕,有望帮助科学家解决有关太阳的一些谜团,包括太阳风的起源,日冕为何比位于其下的太阳表面更热,以及日冕物质抛射的形成机制等。

帕克太阳探测器此前提供的数据已极大丰富了科学家对太阳的认知。2021年,探测器首次进入太阳大气层,发现了日冕外缘布满了尖峰和低谷,而非像此前认为的是一个光滑球形。此外,它收集的数据还确定了太阳风中锯齿状结构的起源。

此次飞行是帕克太阳探测器计划中的3次创纪录“触摸”太阳的首次尝试,接下来的两次飞行将发生在2025年3月22日和2025年6月19日。



帕克太阳探测器(艺术图)。

图片来源:NASA/APL

AI发展:训练数据即将遭遇瓶颈

科技创新世界潮 378

◎本报记者 张佳欣

得益于神经网络规模的扩大以及海量数据的训练,人工智能(AI)在过去10年间突飞猛进。“做大做强”的策略,在构建大型语言模型(LLM)上取得了显著成果,ChatGPT就是一个典型的例子。

然而,《自然》《麻省理工科技评论》等多家杂志网站指出,AI扩展正逼近极限。一方面,AI“吞噬”着越来越多的能源;另一方面,滋养无数模型成长的传统数据集,正被LLM开发人员过度开垦。

训练数据即将遭遇的瓶颈已悄然浮现。有研究机构预测,到2028年左右,用于训练AI模型的数据集典型规模将达到公共在线文本总估计量的规模。换句话说,AI可能会在大约4年内耗尽训练数据。与此同时,数据所有者(如报纸出版商)开始打击对其内容的滥用行为,进一步收紧了访问权限,这将引发“数据共享”规模上的危机。为此,开发人员必须寻找变通之道。

数据集供需失衡

过去10年间,LLM的发展显示出了对数据的巨大需求。自2020年以来,用于训练LLM的“标记”(或单词)数量已增长100倍,从数百亿增加到数万亿。一个常见的数据集RedPajama,包含数万亿个单词。这些数据会被一些公司或研究人员抓取和清洗,成为训练LLM的定制数据集。

然而,可用互联网内容的增长速度出乎意料的缓慢。据估计,其年增长率不到10%,而AI训练数据集的大小每年增长超过一倍。预测显示,这两条曲线将在2028年左右交汇。

与此同时,内容供应商越来越多地

加入软件代码或修改条款,阻止爬虫及AI抓取其数据。在这些内容中,被明确标记为限制爬虫访问的数量,从2023年的不足3%猛增到了2024年的20%至33%之间。

当前,围绕AI训练中数据使用的合法性,试图为数据提供者争取应有赔偿的多起诉讼正在进行。2023年12月,《纽约时报》向OpenAI及其合作伙伴微软提起了诉讼,指控其侵犯了版权;今年4月,纽约市Alden全球资本旗下的8家报纸联合发起了一起类似的诉讼。对此,OpenAI表示,《纽约时报》的诉讼“毫无根据”。

若法院最终站在内容提供商一方,支持其获得经济赔偿,那么对于AI开发人员,尤其是那些资金紧张的学者而言,获取所需数据无疑将变得更加艰难。

新方法有待印证

数据匮乏对AI的传统扩展策略构成了潜在挑战。

寻找更多数据的一个途径是收集非公开数据,如社交媒体消息或视频文

字记录。然而,这种做法的合法性尚存争议。

一些公司选择使用自己的数据来训练AI模型,如Meta利用虚拟现实头显收集的音频和图像进行训练。但各公司政策不同,包括Zoom在内的一些公司则明确表示不会使用客户内容训练AI。

另一种选择可能是专注于快速增长的专业数据集,如天文学或基因组学数据,但其对训练LLM的可用性和实用性尚不清楚。

如果AI接受除文本之外的多种类型的数据训练,可能会为丰富数据的涌入打开闸门。Meta首席AI科学家勒丘恩强调,人类通过观察物体而“吸收”的数据远超过于训练LLM的数据量,机器人形态的AI系统或许能从中获取经验。

此外,制造数据也是解决之道。一些AI公司付费让人们生成训练内容,或使用AI生成的合成数据来训练AI。这已成为一个潜在的巨大数据源。然而,合成数据也存在问题,如递归循环可能巩固错误、放大误解,并降低学习质量。

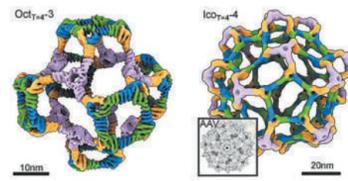
小模型更专更精

另一种策略是摒弃模型“越大越好”的开发观念。一些开发者已在追求更高效、专注于单一任务的小型语言模型。这些模型需要更精细、更专业的数据以及更好的训练技术。

12月5日,OpenAI发布了新的OpenAI o1模型。尽管该公司未透露模型的规模或训练数据集大小,但o1采用了新方法:在强化学习上投入更多时间,让模型对每个回答进行更深入的思考。这标志着一种转变,即从依赖大规模数据集进行预训练,转向更注重训练和推理。

当前,LLM可能已饱览互联网大部分内容,或许无需更多数据即可变得更智能。美国斯坦福大学一项研究表明,模型从多次读取给定数据集中学到的内容,与从相同数量的唯一数据中学习到的一样丰富。

合成数据、专门数据集、多次读取和自我反思等因素的结合,或将共同推动AI的进一步飞跃。



对AI设计的新生蛋白质“纳米笼”进行低温电子显微镜分析。

图片来源:韩国浦项科技大学

科技日报北京12月30日电(记者张梦然)韩国浦项科技大学研究团队利用人工智能(AI)技术,设计出一种“纳米笼”,成功模拟出病毒的复杂结构。它可递送治疗基因,进而成为一种医疗创新平台。这项研究展示了AI在生物医学领域的巨大潜力,特别是在改善基因治疗载体方面。该研究发表在最新一期《自然》杂志上。

病毒具有特殊结构,可以将遗传物质封装在一个球形蛋白质外壳中,这使得它们能够复制并侵入宿主细胞,进而引发疾病。受到这种结构的启发,科学家试图模仿病毒行为,设计出能够递送治疗基因到特定细胞的人造蛋白质“纳米笼”。

现有的“纳米笼”存在一些问题,比如:由于尺寸小,其携带的遗传物质有限;其简单的结构无法完全再现天然病毒蛋白的多功能性。

为了解决这些问题,研究团队转而使用AI驱动的设计方法。他们不仅捕捉到了大多数病毒所共有的对称性,还重现了那些细微的不对称特征。

基于此,团队首次成功设计出四面体、八面体和二十面体形状的新型“纳米笼”。这些纳米结构由4种不同的人工蛋白质组成,形成一个包含6种独特蛋白质-蛋白质界面的复杂架构。

特别值得一提的是,其中一种直径达到75纳米的二十面体结构,可比传统基因传递载体多装3倍的遗传物质。这意味着,在基因治疗领域,可能会取得显著进展。

为了验证效果,团队使用电子显微镜检查了“纳米笼”的结构,并进行了功能测试。结果显示,“纳米笼”确实按照预期构建了精确的对称结构,并且高效地将治疗有效载荷递送到目标细胞中。这一进展为未来医疗应用奠定了坚实的基础,也为研发更先进、更高效的治疗方法开辟了新途径。

AI再次在生物医学领域显示出它的独特价值。此次,科研人员利用人工智能分析病毒的细微特征,成功设计出各种新型“纳米笼”。其中,一种直径达75纳米的二十面体“纳米笼”,一举打破“纳米笼”的尺寸限制,容纳的遗传物质质量达到传统载体的3倍。这种纳米级的结构,可以将治疗基因导入靶细胞,从而治愈疾病。在感慨AI力量的同时,我们也更清楚地意识到,用好AI工具,能破除枷锁,打开新局面。

一种基因突变可抑制肿瘤恶化进程

科技日报北京12月30日电(记者刘震)美国得克萨斯大学西南医学中心科学家在针对小鼠开展的试验中,鉴别出一种名为H2-Aa的基因突变。这种基因突变能够利用免疫系统自己的力量,减缓黑色素瘤等癌症的恶化进程。最新研究有望改进现有癌症免疫疗法效果。相关论文发表于新一期《实验医学杂志》。

为了找到具有肿瘤耐受性的基因,包括2011年诺贝尔生理学或医学奖得主布鲁斯·博伊特勒教授在内的研究团队,创建了携带各种基因突变的小鼠模型。经过深入研究,他们发现了H2-Aa。

结果显示,携带H2-Aa基因两个突变拷贝的小鼠体内完全没有H2-Aa蛋白,在接触黑色素瘤细胞后

并未出现肿瘤生长。与携带原始H2-Aa基因的小鼠相比,携带一个突变拷贝的小鼠体内肿瘤生长明显降低。团队认为,H2-Aa负责产生一种免疫蛋白,从而有助于免疫系统区分自身蛋白和非自身蛋白,为攻击潜在的入侵者做好准备。

随后,团队开发出一种靶向H2-Aa蛋白的单克隆抗体。该抗体在用于治疗罹患黑色素瘤的小鼠时,展现出了显著的抗癌效果。而且,当这些小鼠同时接受检查点抑制剂药物(一种免疫疗法)治疗时,抗癌效果大大增强。

团队认为,这种抗体既可以单独作为一种可行的癌症疗法,也可以作为免疫疗法的“帮手”。他们计划开展临床试验验证这一点。

总编辑 卷点
环球科技24小时
24 Hours of Global Science and Technology

分子在光“牵线”下实现特殊配对

科技日报北京12月30日电(记者张佳欣)意大利博洛尼亚大学团队借助光的力量完成了一项有趣的研究。他们成功地把丝状偶氮苯分子插入环糊精分子的空腔中,形成了一种分子特殊配对。这种组合在自然状态下是无法实现的,但在光的“牵线搭桥”下变成了可能。相关研究发表于最新一期《化学》杂志。

团队发现,向水溶液中施加光能,可以让分子排列偏离自然状态下的“最

稳定结构”,从而生成新型分子组合。这种现象在生物体中很常见,但在人工分子中却少有研究。

此次实验中使用了两种主要分子,即环糊精和偶氮苯。环糊精的中间有空腔分子,外形像一个倒扣的圆锥。偶氮苯则是一个可以在光的作用下改变形状的分

子。通常情况下,这两种分子在水里相遇时,偶氮苯会插进环糊精的空腔里,形成两种可能的配对方式,即复合物A

和复合物B。

复合物A比复合物B更稳定,但复合物B形成得更快。在没有光的情况下,溶液最终只会形成稳定的复合物A。但当用光照射时,偶氮苯分子会改变形状,使复合物解离并重新组装。也就是说,复合物A和复合物B会不断地“分手”再配对。由于复合物B形成得更快,持续光照下,溶液中主要是复合物B。一旦停止光照,偶氮苯分子会恢复原状,溶液最终又回到只含复合物A

的状态。

这种用光操控分子自组装的方法为人工系统模拟生物体内的动态过程提供了新思路。实验表明,光不仅是一种清洁、可持续的能源,还能够通过简单的方式,改变分子的排列和反应机制。

这一技术有望在智能药物、动态材料、纳米马达等领域得到广泛应用,甚至有望开发出对外界刺激有反应的智能材料。

最令人振奋的是,这种从患者自身

组织中生成成熟心肌细胞的技术,未来有可能用于修复因心脏病发作或其他心血管疾病造成的损伤。如果这项技术能够成功应用于临床治疗,未来将能提供一种用于心脏组织再生的个性化解决方案,从而在治疗心血管疾病方面迈出重要一步。

成纤维细胞“变身”诱导心肌细胞

科技日报北京12月30日电(记者张梦然)韩国高丽大学研究团队开发了一种创新技术,能够直接将人体内的成纤维细胞(一种常见的结缔组织细胞),转化为成熟且功能完整的诱导心肌细胞。该成果发表在最新《实验与分子医学》杂志上,为再生医学领域带来了新希望。

传统的直接心脏重编程方法旨在跳过中间的干细胞阶段,直接将成纤维细胞转换为心肌细胞。尽管这种方法潜力巨大,但一直难以生成足够成熟和功能完善的心肌细胞。此次,团队通过激活JAK2-STAT3信号通路克服了这一难题,这项新技术使用了成纤维细胞生长因子4与维生素C的组合,不仅加速了细胞的成熟过程,还增强了它们的功能。

为了验证这种方法的效果,团队运用了RNA测序、荧光成像和电生理测试等先进手段。结果显示,转化后的心肌细胞表现出更好的结构特征,包括清晰的肌节和T小管边界,以及更强的电活动和离子通道功能。这表明,新方法可以更高效率地生产出与天然心肌细胞相似的细胞。

创新连线·俄罗斯

俄将建无人机系统区域协调基础设施

俄罗斯技术联盟计划在2030年之前建立一个覆盖全俄的无人机系统区域协调基础设施,包括通用无人机港、高精度着陆系统等。

俄罗斯国家技术倡议项目支持基金会总经理瓦季姆·梅德韦杰夫表示,该项目的主要任务之一是简化无人机系统的飞行程序,增加飞行次数、减少人为因素、提高自动化水平。其核心目标是确保无人机系统可用于修复因心脏病发作或其他心血管疾病造成的损伤。如果这项技术能够成功应用于临床治疗,未来将能提供一种用于心脏组织再生的个性化解决方案,从而在治疗心血管疾病方面迈出重要一步。

参与该项目的除了俄罗斯无人机港公司,还包括“沉默之翼”公司、因诺波利斯大学和飞翔无人机公司等,俄罗斯国家技术倡议项目支持基金会对该项目提供支持并协助。俄罗斯国家技术倡议是一个由企业和专家群体代表组成的联盟,旨在推动俄罗斯前沿技术市场和产业发展。

(本栏目稿件来源:俄罗斯卫星通讯社 编辑整理:本报记者张浩)