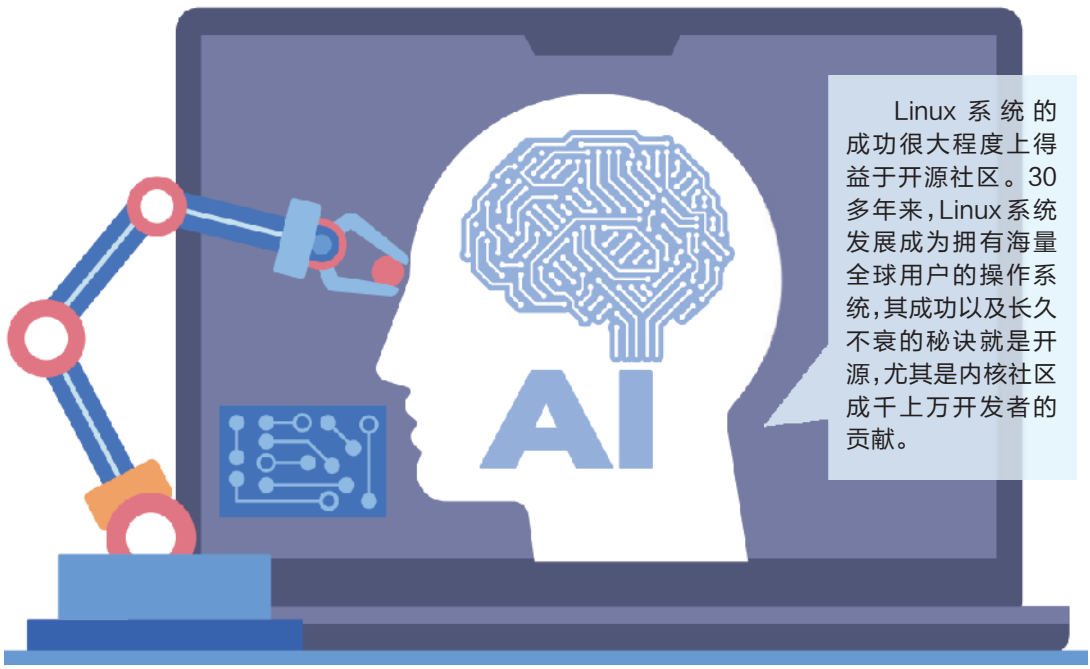


数据开源为 AI 发展“推波助澜”

大模型是 AI 未来发展的重要方向,其研究和应用将逐步成为 AI 发展的关键方向,并有望形成新一波 AI 推广浪潮,而 AI 数据开源将进一步促进大模型的发展。

在日前举行的 2023 北京智源大会上,“AI 数据开源”引发广泛关注。AI 数据为什么要开源? AI 数据开源面临哪些挑战? 它会是未来 AI 发展的重要趋势吗? 记者带着这些问题采访了相关专家。



Linux 系统的成功很大程度上得益于开源社区。30 多年来, Linux 系统发展成为拥有海量全球用户的操作系统,其成功以及长久不衰的秘诀就是开源,尤其是内核社区成千上万开发者的贡献。

ITMT 快报

我国首次高压 纯氢管道试验成功

6月25日,我国首次高压纯氢管道试验取得成功。

在哈密国家管网集团管道断裂控制试验场,工作人员完成了6.3兆帕管道充氢测试和9.45兆帕管道爆破测试,各项结果均达到预期。

本次试验是国内首次对输氢非金属管道进行的高压在线测试,也是国内首次对非金属管道进行高压纯氢爆破试验。此次试验依托的国家管网集团管道断裂控制试验场,是继英国、意大利之后全球第三个管道断裂控制试验场。

国家管网集团西部管道公司科技信息服务中心工艺所所长杨明称:“我们将加快推进氢气、二氧化碳等多种介质储运前沿技术攻关,推动传统油气管输业务与新能源储运业务融合发展,助力氢能‘制储输用’全链条产业建设。”

氢能具有零碳、高效、可储存、安全可控等显著优势,据《氢能产业发展中长期规划(2021-2035年)》,氢能是国家能源体系的重要组成部分,是用能终端实现绿色低碳的重要载体。目前,已有多个省市制定并发布氢能发展政策。

当前掣肘氢能产业化发展的核心问题是成本问题,其中包括氢制取、氢储运、氢加注三个环节的成本。在储运环节,远距离运输是氢能产业目前主要攻坚的关键问题。相关报道指出,随着我国首次高压纯氢管道试验取得成功,为我国今后实现大规模、低成本的远距离纯氢运输提供技术支撑。

制取环节的成本也在进一步良性下降。据中信建投最新研报指出,2023年以来,由于采购范围变动、集成方案优化等一系列因素,制氢设备价格降低,目前稳定在700万元/套-800万元/套的合理价格区间。

具体来看,采购范围方面,剥离电气设备采购制氢设备;集成方案方面,制氢规模推动系统集成方案从一拖一优化为多拖一,后处理设备共用带来成本均摊。机构还进一步指出,预计制氢规模进一步提升集成化水平,采购成本将继续良性下降。

据科创板日报

特斯拉“超级计算机” 或在7月正式投产

近日,特斯拉在推特上上线了新账号“Tesla AI”,并发布了人工智能、机器人相关的多条消息。

值得注意的是,该账号还披露了特斯拉首台“超级计算机”Dojo的进展,其将于7月正式投入生产,到2024年初,将成为全球最先进的5台超级计算机之一。

据悉,Dojo是一个强大的算力中心,有能力处理大量的AI任务,加速特斯拉的Autopilot和完全自动驾驶(FSD)系统的迭代,同时也将为特斯拉的人形机器人Optimus提供算力支持。

业界普遍认为,特斯拉此举的目的就是为了摆脱对英伟达GPU的依赖,提升自身话语权。因为目前人工智能训练芯片和计算平台均掌握在英伟达手中,这也促成了其股价的飙升。

综合

今年全球 OLED 面板 营收预计同比降 7%

根据市场研究机构 DSCC 最新报告显示,2023 年预计全球 OLED 面板出货量同比增长 1%,营收同比下降 7%至 389 亿美元(折合人民币约 2793 亿元)规模。

按 OLED 不同应用领域,DSCC 预计 2023 年 OLED 智能手机面板出货量将同比增长 4%,营收同比下降 5%;OLED 电视面板的出货量将同比下降 29%,营收也同比下降 29%;OLED 笔记本电脑面板的出货量将同比下降 15%,营收同比下降 20%。

另外,该机构此前预测 OLED 在车用显示器销量中的占比将从 2023 年的 6%增至 2025 年的 11%和 2027 年的 17%,包括液晶显示器(LCD)和 OLED 在内的汽车显示器年销售额预计将从 2020 年的约 70 亿美元(折合人民币约 502.5 亿元)增长到 2027 年的约 130 亿美元(折合人民币约 933 亿元)。

值得一提的是,根据市场研究机构 Stone Partners 此前发布的数据,今年第二季度,中国柔性 OLED 的出货量占比预计将超过 50%,首次超过韩国。数据显示,三星显示柔性 OLED 出货量预计较去年同期的 4650 万片减至 3890 万片,市场份额预计将从 59.5%下降至 41.2%,而京东方、天马、华星、维信诺等中国柔性 OLED 面板的市场份额预计为 50.2%。

综合

合法地从公开或可公开获得的资源中收集的数据。人们可以在开源协议允许的范围内,以 AI 大模型训练、AI 算法开发为目的,对数据进行访问、修改和使用。部分数据可能要求使用过程中遵守更严格的协议。”林咏华表示。

此外,今天的基础 AI 大模型不只具备理解能力,还具有生成能力,它能够对外进行认知输出、价值观输出等,可能给社会带来巨大影响。“我们在训练基础大模型的时候,所使用的预训练数据会对 AI 生成内容质量起到很大程度的决定性作用。因此,开源数据的质量十分重要。”

林咏华指出,由于高质量的数据(如文章、图片、视频等)通常有版权,由于版权或商业因素导致的闭源以及数据孤岛等挑战会制约 AI 的发展,所以需要多方推动构建更多高质量的开源数据集,尤其是用于训练基础 AI 大模型的开源数据集。

LF AI & DATA 基金会董事主席堵俊平对此也深有感触:“AI 大模型就像一个贪吃的‘怪兽’,始终需要研究人员投喂更多的、质量更好的数据。”他说,当前数据几乎都是从“在网络上主动收集”“从第三方购买”“利用公开数据集”这三个渠道得来。在堵俊平看来,从第一个渠道得到的数据局限性较强,由于版权问题,很多公司只能从其私域获得数据;从第二个渠道获取的数据面临数据定价、数据质量等问题;而从第三个渠道获取的数据往往只能作为研究使用,在商用或者其他方面有很多限制。

开源渐成 AI 发展重要趋势

记者了解到,智源对 2023 年 1 月到 5 月底发布的、具有影响力的语言模型进行过统计。统计结果表明,国外发布的开源语言模型有 39 个,国内发布的开源语言模型有 11 个。

“开源是推动 AI 技术进步的重要力量, AI 开源开放生态及平台建设也日益受到重视。开源开放毫无疑问已经成为重要的 AI 发展趋势之一。”林咏华表示,“开源能够促进 AI 大模型科研创新,推动和降低 AI 大模型落地乃至整个 AI 产业落地的门槛。”

然而,通往开源的道路并非一帆风顺,在数

据之外,算力也是开源路上的一只“拦路虎”。AI 大模型训练依赖庞大的数据、算力。训练参数量级的增长使得算力需求也随之增长,算力集群正变得愈发庞大。

然而算力成本却是小型开发者的“不可承受之重”。拿到 AI 大模型开源数据后,往往需要对其进行微调 and 二次开发。但现实的情况是,对一些小型开发者来说,仅仅是做推理都很困难,就更别提对 AI 大模型做微调、二次开发。以 ChatGPT 为例,仅就算力而言,Open AI 为了训练它,就构建了由近 3 万张英伟达 V100 显卡组成的庞大算力集群。有消息称,Open AI 公司发布的新一代语言模型 GPT-4 甚至达到了 100 万亿的参数规模,其对应的算力需求同比大幅增加。

目前,有一些研究机构希望用技术的革新抵消巨大的算力成本。最直接的手段是通过训练技术的革新加快 AI 大模型推理速度、降低算力成本、减少能耗,以此来提高 AI 大模型的易用性,让开源数据更好地发挥价值,但这只能从工程上对算力资源的约束起到缓解作用,并非终极方案。

有业内专家表示,解决算力问题最终还是要回到 AI 大模型自身寻找突破点,一个十分被看好的方向便是稀疏大模型。稀疏大模型的特点是容量很大,但只有用于给定任务、样本或标记时,模型的部分功能才会被激活。也就是说,这种稀疏大模型的动态结构能够让 AI 大模型在参数量上再跃升几个层级,同时又不必付出巨大的算力代价,一举两得。

此外,开源社区的作用同样不容忽视。开源社区是推动开源发展的重要基石,开源的最初发源地,就是来自于社区开发者的贡献。“Linux 系统的成功很大程度上得益于开源社区。30 多年来, Linux 系统发展成为拥有海量全球用户的操作系统,其成功以及长久不衰的秘诀就是开源,尤其是内核社区成千上万开发者的贡献。”林咏华举例说。

“开源开放可以使得我们站在前人的肩膀上前行。”林咏华总结道,“这些年 AI 领域取得的成果大多受益于开源,如果没有开源, AI 不会发展到今天。”

据《科技日报》

背后挑战不容忽视

开源固然会为 AI 发展带来诸多好处,但其背后的挑战也不容忽视。其中之一,便是开源安全与合规挑战。林咏华认为,对传统的商业软件而言,开源中的安全、合规、许可证和代码质量风险等是使用开源组件必须面临的挑战。然而在 AI 大模型时代,更大的挑战则在开源数据集管理方面。

因此, AI 数据开源应在协议许可的范围内进行。“用于 AI 大模型训练的开源数据必须是

高级别自动驾驶提速 “车路云”协同发展成趋势

L3级自动驾驶正渐行渐近。

日前,工业和信息化部副部长辛国斌在国务院新闻办公室举行的政策例行吹风会上表示,将启动智能网联汽车准入和上路通行试点,组织开展城市级“车路云一体化”示范应用,支持有条件的自动驾驶。

“这里面讲的是 L3 级及更高级别的自动驾驶功能商业化应用。我们已经启动了这项工作,地方也在积极响应。相信在不久的将来,更多有智能化水平的产品会跑在路上。”辛国斌表示。

积极推进自动驾驶商业化落地

不少地方政府已经在为 L3 级自动驾驶技术落地而积极推进配套的基础设施建设。比如,深圳市坪山区正在加快推进深圳智能网联交通测试示范区、全域路口网联化改造等基础设施建设,其中环境园封闭测试区将于今年下半年建设完成,届时可全面支撑智能网联汽车 L3 级、L4 级产品准入测试。

就在 2023 未来汽车先行者大会暨全球智能网联汽车商业化创新大会期间,安途智驾、鹏电集团和萝卜运力联合体(萝卜快跑)获颁智能网联汽车无人商业化试点牌照,美团、新石器、白犀牛也获颁无人小车测试牌照。

“萝卜快跑获得深圳市坪山区颁发的首批智能网联汽车无人商业化试点通知书,反映出国内自动驾驶商业化落地进程在加速。然而,这也暴露出一些问题,如政策环境、技术标准和道路基础设施等方面的不足。此外,虽然自动驾驶技术取得了显著进展,但其商业化仍面临一定挑战,包括成本、安全性、法规合规性等问题。不过,这

一消息对自动驾驶行业来说是利好消息,因为它意味着中国正在积极推动自动驾驶技术的商业化应用。”商务部研究院电商所副研究员洪勇表示。

此前,华为常务董事、终端 BG CEO、智能汽车解决方案 BU CEO 余承东在 2023 重庆车展期间曾表示,中国 L3 级自动驾驶标准预计在 6 月底出炉。

对此,洪勇认为:“如果中国 L3 级自动驾驶标准能在 6 月底出炉,将对当前的智能驾驶市场产生深远影响。首先,这将有助于提高自动驾驶技术的标准化程度,从而推动整个行业的健康发展。其次,这将有助于消除不同地区和制造商之间的互操作性问题,使自动驾驶车辆能够在全球范围内无缝运行。最后,这将为消费者提供更多的选择,并可能降低自动驾驶技术的使用成本。总体来说,标准的出台,将进一步推动自动驾驶技术的发展和普及。”

实际上,不只是中国,当前全球不少地区也在积极推进自动驾驶商业化落地。例如,日前,英国启动了首个无人驾驶租赁项目,提供方为初创公司 Imperium Drive,该公司也是欧洲首家提供此类服务的公司,计划在接下来的 18 个月里进行进一步测试,争取实现真正的全自动无人驾驶,届时将不再需要安全员的监控6月8日,美国加州机动车辆管理局(DMV)批准了梅赛德斯-奔驰的 L3 级自动驾驶系统的上路行驶申请,意味着装备该系统的奔驰车型可以在指定公路上开启自动驾驶功能,奔驰也成为首家获批在加州公路使用 L3 级自动驾驶的厂商。今年 5 月,谷歌母公司 Alphabet 旗下的自动驾驶公司 Waymo 计划在旧金山和凤凰城扩大无人驾驶出租车的运营规模。

“车路云”融合协同是未来发展重点

毫无疑问,“车、路、云”的融合协同发展成为了自动驾驶接下来发展的重点。除了芯片、高精度传感器、操作系统这类车辆本身的技术之外,未来国家将推进加快 C-V2X、路侧感知、边缘计算等基础设施建设;而云端则是基于边缘云、区域云和中心云三级架构的云控基础平台。

具体而言,我国将首先支持重点大企业牵头,大中小企业参与,开展跨行业跨领域协同创新;加快关键芯片、高精度传感器、操作系统等新技术新产品的研发和推广应用,进一步提升产业发展内生动力。

其次是进一步完善网联基础设施。加快 C-V2X、路侧感知、边缘计算等基础设施建设,建立基于边缘云、区域云和中心云三级架构的云控基础平台,形成统一的接口、数据和通信标准,进一步提升网络感知、云端计算能力。

最后,要深化测试示范应用。辛国斌表示将启动智能网联汽车准入和上路通行试点,组织开展城市级“车路云一体化”示范应用,支持有条件的自动驾驶。

在此方面,我国已经采取地方先行先试的模式。目前,已逐步开放公路(高速公路)测试、无人化(远程)测试、载人载物示范运营,包括北京、上海、深圳等多个城市已发布相关的道路测试与示范应用管理法规,其他城市相关细则也持续推进。截至 2023 年 5 月,全国各地加快测试示范部署与基础设施建设,目前累计开放超过 15000 公里测试道路,累计测试总里程超过 6000 万公里,发放测试牌照超过 2500 张。这为我国自动驾驶汽车的发展奠定了坚实基础。

综合《每日经济新闻》《中国科技信息》