

与业内权威“较真”

上海“独角兽”研究员“硬杠”苹果团队论文

阶跃星辰员工测试评论帖发布后,引来海内外数十万网友关注

上海大模型“独角兽”的研究员,为何要“硬杠”一篇大厂论文,还获得了一片叫好?这几天,来自上海的大模型“独角兽”阶跃星辰的研究员杨磊,在国内外机器学习社区火了;他“硬杠”大厂一篇“漏洞百出”的论文,引来数十万计的“吃瓜群众”,更不乏行业大V为他

站台。AI 初创公司、进入大语言模型赛道不过一年多的程序员……这样的背景,怎么也不像是“哪吒闹海”的主人公!可恰恰是一份简单而纯粹、对科研较真的态度,让业界看到并认可了来自速空间的实力。

反复测试 论文“坑”多

上个月初,有着多模态“卷王”之称的阶跃星辰研究员杨磊,被同事“种草”了一篇苹果团队发布的视觉推理基准论文。这篇论文发表在公开的论文预印本平台上,同时也被苹果投向了国际表征学习大会(ICLR 2026)——这可是AI和深度学习领域的顶级学术会议之一。

都被公认“卷王”了,可见平时节奏有多快。可这项研究吸引杨磊,并让他心甘情愿“加班”的,恰是它的核心,和自己近期的工作相关。这篇论文本意是展示其最新的AI研究进展,“外行看热闹”的话,可以理解为出了套“基准测试”——用一套统一的数据集,类似于考题,来比较不同的AI模型,看谁的表现更好。

“可以这样理解,它让大模型去分析数据集里哪个地方错了,很像考试中老师给‘步骤分’。”杨磊告诉记者,“此前,我们也开发了专门的标注工具,产生了一批数据,并有了预研模型,很希望通过第三方平台来验证成果。”

可熬了一个周末的通宵来做适配,杨磊的心情只能用两个字来形容:沮丧!

“模型跑出来的点数极其之低,远低于预期。”他说,苹果团队也给出了其测试开源模型以及主流模型的榜单,杨磊所用的经过适配的模型,几乎要排名垫底了。

哪里出问题了?排除了自身模

型的问题后,杨磊仔细分析了论文里给出的代码,他被震惊了!照理说,这套“基准测试”的运行流程是,给个图、抛出问题、给出解答、解答里有步骤错误,模型需要准确找出这一步——也就是说,参加评测的必须是一个能输入图片的VLM(视觉—语言模型)。

可论文的开源代码里,输送的是图片的路径地址,而没有把路径解析成图片内容。这就会造成模型看不到图片,只能看到题目和错误之处。杨磊向论文作者反馈了这个“低级错误”,可得到的反馈却是:这是一段“伪代码”。

更令杨磊没有想到的是,在修复此处bug后,模型的点数更低了……

百思不得其解下,杨磊只能花费更多时间和精力,去分析到底哪儿出了问题。再看了前20道题中自家模型答错了的题后,结果大吃一惊——里面有6道题明确属于“参考答案”错误,从“参考答案”的错误风格来看,很可能是模型自动生成的“参考答案”加上质检严重不足,导致包含大量“幻觉”。杨磊初步估算了一下,“参考答案”错误率可能已经达到了30%。

“原本以为苹果是大厂,我们也比较信任。谁能想到自己花费了大量时间和精力,发现自己‘被坑了’,对方也很不负责。”杨磊如是说。

发布评论 对方撤稿

他在GitHub(全球最常用的代

码托管与协作平台)上,向论文作者反馈,指出其中的错误。6天后,作者“姗姗来迟”,简单回复后,便关闭了“讨论框”——这一举动意味着,问题被解决,可实际上并没有。

由于论文也被投向了国际表征学习大会,大会的审稿意见中,没有任何一位审稿人发现“参考答案”的质量问题,也没人发现论文中的例子存在幻觉和错误。

杨磊写了一份详细的公开评论,提醒大会审稿人和社区:这个数据集质量堪忧,极易误导研究方向。“我在这里评论是为了防止有兴趣的研究人员重复我经历的相同循环——看到第一个错误检测任务时的兴奋,运行它后的震惊和失望,以及追踪底层GT问题后的沮丧——从而节省每个人的时间和精力。”他这样写道。

再后来,论文撤稿了。作者“礼貌”地告诉杨磊,会积极修复这些问题,就没有然后了。

“众所周知,这几年人工智能领域特别火。每两年,投向顶级会议的论文数量几乎翻番。AI的进步,会使更多粗制滥造的工作容易‘批量生产’;加之审稿人都是‘义务劳动’,这些更不容易被发现。”杨磊解释。

杨磊的这一举动,在国内外各大机器学习社区,都引发了共鸣。他说,希望通过分享自己这次曲折的经历,能让AI社区更和谐,也让大家意识到,不要盲目信任表面包装,哪怕是来自大厂。

本报记者 郜阳

与城市细节“较真”

抓杆距地仅0.96米 呼叫装置安装过高

静安区检察院通过公益诉讼让多处问题公共无障碍设施更“规范”

一根L形安全抓杆,顶部距地仅0.96米——这个存在于静安区某公园公厕内的细节,与国际标准相差近半米。国际残疾人日前夕,静安区检察院通过一起公益诉讼案件,为修正这类“被忽视的高度”而行动,让城市温度体现在最细微处。

静安区是上海最早进入深度老龄化的城区之一,老年人口比例达42%,残疾人群体占比4.67%。今年8月起,区检察院联合区残联,对辖区内公园绿地等公共场所的无障碍环境建设开展专项调查。

调查发现,问题不止于抓杆。检察官在走访10余处公园绿地后,发现多个“失之毫厘”的细节:部分坡道缺乏清晰标识,有的厕所未设多功能台,救助呼叫装置安装过高……“这些对普通人或许是小事,但对坐轮椅的老人,残障人士而言,每次出行都可能充满忐忑。”承办检察官李昕学说。

根据《建筑与市政工程无障碍通用规范》,辖区内9块公园绿地的无障碍厕所区域,存在不符合标准的问题。依据相关法律法规,这些设施应当进行必要改造。

2025年9月,静安区检察院向职能部门制发检察建议,列出详细“改造清单”:救助呼叫装置需触手可及,轮椅坡道标识应鲜

明无误,厕所门应外开或水平滑动……建议将宏大的“无障碍”概念,分解为可执行、可检验的具体条款。

行政机关高度重视,积极沟通并制定整改计划。反馈照片显示,坡道旁已贴上醒目标识,呼叫按钮高度调至合适位置,外开式门为轮椅留出回转空间。整改推动无障碍设施从“有没有”向“好不好”转变。

前不久,检察官与区残联邀请残疾人志愿者,赴辅德里公园、中兴公园等地实地“验收”。“现在这个台盆高度正好,一伸手就能够到。”一位坐轮椅的志愿者体验后表示。改造后的多功能台、合规的L型抓杆、触手可及的呼叫标志,获得了最直接使用者的肯定。

“无障碍环境建设是持续过程,检察公益诉讼将延伸至为老中心、社区食堂等更多民生场所,让阳光照到每一处‘隐秘的角落’。”今年以来,静安区检察院已办理相关行政公益诉讼案件20件,督促改造无障碍标识、坡道、低位设施等120余处。

检察公益诉讼不仅关乎宏大的叙事,更聚焦市井生活。修正抓杆的高度,亦是搭建沟通的桥梁,推动社会从“有”到“优”转变,守护每个人平等参与生活的权利。

本报记者 郭剑烽



“沉浸式”学交规

今天是第十四个“全国交通安全日”,以“文明交通 礼行天下”为主题的上海市2025年“全国交通安全日”主题宣传活动在杨浦区YOUNG剧场举行。活动由市公安局联合多部门共同主办,通过沉浸式节目展演、公益表彰、全民倡议等形式,普及交通安全知识,传递遵礼守法、文明出行理念。

本报记者 张龙 摄影报道

“海鸥计划”助女大学生“破界”成长

“探索·破界·成长:智领她未来”上海市女大学生职业飞翔“海鸥计划”2025年度研讨会昨天在华东师范大学举行。近两百名高校师生、企业家与专家学者共同探讨智能时代下,青年女性如何突破边界、规划未来。

提供职前场景“预演”

许多大学生的职业起点,往往源于一个个具体的问题。上海师范大学信息与计算科学专业学生邹越分享了自己在职业探索过程中,如何面对迷茫、寻找方向,收获自我认知与成长。她感慨,职业规划并非一蹴而就,在“海鸥计划”职前研习坊这样的平台,通过与真实行业链接、在行动中试炼,才逐渐将“我想做什么”转化为“我能做什么”。

她的故事,也是“海鸥计划”试图回答的核心议题之一:在直面职

场之前,学生如何获得真实的“预演”?据悉,该计划通过与上海学生事务中心联合,通过社、企、校的紧密联动,为学生提供大量企业暑期实习的机会,在高校自身的职业规划大赛与个性化指导基础上,让学生在真实的职场场景中观察与实践,她们不仅能够更深入地认知自我优势与兴趣,更能在未来岗位选择上多些自信,明确各自的职业发展路径。

善于思维模式“破界”

圆桌对话中,“破界成长”的案例具体而动人。上海必是科技有限公司联合创始人兼COO李冉则讲述了思维模式的“破界”。她坦言,创业迫使自己完成了从大厂执行者到企业经营者的转变。从零开始,搭建团队、制度与文化的过程充满挑战,却也让她深刻体悟到,对团队

梦想与企业存亡负责的经营思维,这也是驱动个人能力边界突破的核心力量。

为积极开展巾帼就业创业促进行动,上海市妇联以“海鸥计划”为载体,聚焦女大学生群体,引导树立理性向上的就业观,拓展就业渠道、加强就业创业服务。2025年“海鸥计划”全年开展近100场活动。其中,“职飞起航”项目成功搭建起招聘平台、区妇联/人社(就促)与各大高校之间的信息桥梁,助力女大学生就业需求对接与资源联动,全年共联合896家企业提供岗位2932个,招聘需求27983人,收到简历超12.7万份。

据悉,2026年,“海鸥计划”将继续发挥高校“海鸥湾”工作站的枢纽作用,联动校内多部门协同参与,共同支持女大学生职业启航与成长发展。

本报记者 解敏