

别让AI“跑偏” 伦理重置不可缺

专家呼吁:人机互动需正向引导,避免人工智能产生“幻觉”

教育新观察

全国著名语文教育专家、复旦附中语文特级教师黄玉峰,多年来一直致力于对学生进行君子养成教育,但最近却遇到了一个新

问题——在人工智能时代,人形机器人将会越来越多地走入家庭、走进课堂、走向社会,在“它”面前我们是不是还要做个谦谦君子?是不是也要让“它”遵守起码的君子之道呢?这样的疑问在日前举行的第九届君子养成大会上引起了很多与会者的好奇。

AI不懂小我与大我

几天前,现任复旦五浦汇实验学校校长的黄玉峰在备课苏轼的《后赤壁赋》时,就文中的最后一句话“道士顾笑,予亦惊寤。开户视之,不见其处”,向大语言模型提了两个小问题:一是“惊”是什么意思,二是“其”又是指谁。“没想到人工智能的回答都是不准确的。比如对‘其’字的解释,显然人工智能是识别不了文章的本义,它既不会懂得这个‘其’代表的正是‘我’,更不能理解和表达‘我从哪里来,要到哪里去’这样的哲学命题。”黄玉峰说,君子的一个重要特征是善于学习,尤其是要具有“博学而笃志,切问而近思”的精神素养,但无论是AI搜索工具还是人形机器人,它们在“笃志”“切问”和“近思”上都不具备君子品质。一个不懂小我与大我,没有自信力的机器人,最终只是一个人类使用的工具而已。

“全知老师”只是“助教”

“人类是由各种行为实践上升到道德层面的,而机器人是依赖人类输入的指令来行事和‘思维’的,它需要依靠大量的例证才能学习人类的道

德规范,这就是人与机器的差别所在。”华东师范大学教育学部教授吴刚说,人形机器人也不会具备人类所特有的温情,所以,就算它将来能大量地走进教室,充其量也只是个“助教”的角色,类似一个拥有大数据库的“全知老师”,它会告诉学生现成的答案,会输出已有的知识,但它代替不了主讲教师传道授业的职责。因为,师生之间、同学之间的思想交锋、亲切交心和情感互动,是任何大语言模型数据库所不能做到的。

伦理操守需“双向奔赴”

上海市儿童发展研究中心主任杨雄提出了一个警醒:不要随意向DeepSeek、豆包之类的大语言模型输入无价值问题,如果喂给它的都是些乱七八糟的东西,AI不断接收后也会产生“幻觉”,认为“事情就该如此”。比如,当很多人不约而同地向它提问要不要去买某个股票时,它很可能会认为这个人都在追问的股票应该就是“绩优股”。因此,当人与AI互动时必须遵守一个基本的伦理规范,就是要维持正向的、高级的对话内容,防止它“一本正经胡说八道”。人形机器人只是个外挂学习工具,成不

了具有批判精神的君子。

有与会专家表示,开发者在模型训练和算法部署的每个环节,都应融入透明性和可验证性机制,从源头上防止虚假信息产生和传播。例如,在数据收集阶段,必须明确数据来源和用途,确保数据的真实性和可靠性;在模型训练过程中,必须采用可解释性的算法,使模型的决策过程能够被理解和监督,一旦出现虚假信息生成的情况,可以追溯到问题产生的环节。同时,应加快推出相关的法律法规,对利用AI制造并传播虚假信息的行为予以惩治。这就需要人与人工智能工具在基本的伦理操守上完成一次“双向奔赴”。

“当人形机器人开始进入我们的现实生活和真实工作场景时,就需要提前对整个社会的伦理规范进行一次重置,让AI工具也要遵守人类的公序良俗。”浙江之江实验室智能社会治理实验室副研究员吕明杰表示,就拿智能汽车来说,一旦在行驶过程中发生了事故,那么责任该怎么认定呢?所以说,人形机器人等智能工具正面临着伦理的重置与重塑,这是它们回避不了的,也是我们亟需重视和解决的问题。 本报记者 王蔚



新民随笔

AI是预测大师?

关尹

几天前的世界杯预选赛亚洲区18强赛,中国男足主场迎战澳大利亚队,好几个足球群都在讨论这场比赛,中国队能不能赢,有朋友煞有介事地告诉我:他“算”过了,国足不会输球。一打听,原来他问的是AI软件,后者表示:“两队实力有差距,但中国队急需拿分,有望借助强大的主场之利逼平对手,甚至1比0小胜。”

听后我不禁哑然失笑,在体育竞赛领域,所谓的AI预测胜负实在是太不靠谱,看过就算,哪里还能当真。果不其然,国足在杭州被澳大利亚队干脆利落地打了一个2比0,毫无胜机。

别以为是玩笑,如今市面上打着“AI预测”大旗的,还真不少。比如网络上充斥着大量“AI预测彩票号码”的视频,就吸引了不少渴望一夜暴富者的目光。一些不法分子借助AI之名,将AI预测彩票吹嘘成一本万利的“致富秘籍”。实际上,彩票中奖号码是随机产生,每次开奖结果相

互独立且毫无规律,AI根本无法精准预测,这是一个基本常识。所谓“AI预测彩票”,不过是某些人精心设计的骗局,是他们骗取钱财的手段罢了。一个很简单的道理:如果AI真的这么神,还要这么卖力地兜售软件吗?他们自己去买,不是早发财了?

人工智能技术近年来发展势头迅猛,被应用到越来越多的领域。不可否认,它的出现推动了社会进步,也确实给人们带来了便利。但同时,过度依赖AI也正在成为一种“社会病”。比如此前大众热议的“AI将代替医生”,有病不用去医院,只需在DeepSeek上搜索一下,就能知道诊断结果、如何用药,因此而吃亏的不在少数。只能说,AI可以成为医生的得力助手,却无法完全取代医生。

所以,完全依赖互联网、人工智能的行为都是不可取的。尤其在探索未知的“预测”领域,AI更不可能是大师,千万不要迷信它是万能的,否则将追悔莫及。

“复旦源”落成在即 一源五馆启动公募

本报讯(记者 王蔚)在纪念建校120周年之际,复旦大学标志性教育历史文化功能区“复旦源”揭幕在即,将形成包括校史馆、博物馆、艺术馆、科技成果馆、校友馆在内的“一源五馆”新格局。昨天下午,“复旦源”教育历史文化功能区内涵建设项目捐赠签约仪式暨公募启动仪式举行。上海市教育发展基金会资助“复旦源”内涵建设项目1200万元,同时支持依托基金会开展“复旦源”公募活动,面向社会公众筹集内涵建设资金。

登辉先生亲赴南洋,募集15万银元,为学校在上海江湾先后购地70余亩作为校地(今复旦大学邯郸校区)。1922年,复旦大学江湾新校舍建成,这片校舍就是相辉堂大草坪四周区域。2023年底,学校启动“复旦源”历史文化功能区建设。这片以相辉堂草坪为中心的区域,是现在复旦邯郸校区的发源地、复旦精神的溯源地,综合地理、历史和文化因素,命名为“复旦源”。复旦大学要将“复旦源”历史风貌区打造为上海教育博物馆的重要承载地、先行区。

复旦可信具身智能研究院揭牌

现场发布3个代表性科研成果

本报讯(记者 张炯强)昨天,复旦大学可信具身智能研究院揭牌。今年1月,复旦大学可信具身智能研究院(筹)实体运行科研机构成立。研究院专注具身智能的前沿研究与应用落地,是复旦大学面向世界科技前沿的重要战略布局。通过整合计算机视觉、自然语言处理、机器人学、控制系统及科技伦理等多学科力量,研发具有自主探索能力、持续进化特性且符合人类价值观的智能体,为未来人机协同与智能社会建设提供核心驱动力。面向具身智能的未来发展,研究院重点关注基础模

型、数据引擎、具身交互、本体研制、可信机制五大方向,并将整合校内外科研力量,面向全球招聘高端人才。揭牌现场发布了3个代表性成果:邱锡鹏教授发布了研究院视觉与语言团队联合构建的多模态具身模型,该成果有望突破具身模型的交互瓶颈;吴祖煌副教授发布了细粒度时空描述数据集和大规模复杂任务操作数据集;马兴军研究员发布了可信人工智能开放社区平台OpenTAI。复旦大学携手北京字跳网络技术有限公司等4家企业,共同建设的4个校企联合实验室也集体亮相。



昨天下午,第五届“爱在静安·点亮星梦”世界孤独症关注日主题活动,在静安区南京西路中信泰富广场举办。一首表演演唱《青春舞曲》拉开了活动帷幕,来自小马过河启音天使合唱团的孩子们用歌声传递出孤独症群体对生活的热爱与渴望,赢得了现场观众的阵阵掌声。 本报记者 张龙 摄影报道

「星梦」童声唱出热爱

首届国际智慧 物业博览会启幕

本报讯(记者 杨玉红)昨日,2025上海国际智慧物业博览会(简称“上海物博会”)在新国际博览中心开幕,这也是物业行业首个“物业 Plus”系列展。展览打造了物业与城市服务、智慧社区、智慧停车等10大专业展区,推出超1000款科技新品,展示物业上下游产业链各领域最前沿的产品、技术以及解决方案。展览将持续至4月3日。在智慧科技展区,一款人形智能安保机器人吸引了众多参观者。它能够精准识别业主身份,

人形机器人当保安 千款科技新品亮相

主动上前开门热情迎接;自动扫描访客二维码,高效完成身份核验流程。一旦遭遇灾害性天气等突发状况,它便自动播报温馨提示。“一台人形智能安保机器人售价低于10万元,设计使用寿命为5年,某些方面的工作效能可替代2至3名保安。目前,这款机器人已在多个社区‘正式上岗’,表现十分出色。”该展位负责人朱思宁介绍道,他们是一家来自北京的科技公司,首次来上海参加大型物业行业展会。开展首日,他们已接待了多家物业企业的负

责人,围绕产品应用、合作模式等事宜进行洽谈,探索智能安保机器人在物业领域的应用前景。同时,京东物流、宇泛智能、新视窗等众多优秀科技企业纷纷亮相,带来了24小时在线响应的AI管家系统、动态监测社区能耗的智慧能源平台、AI数字人智能体、无人巡检设备等前沿科技产品。展会还特别设立了“以旧换新”展区。一大批电梯企业和智慧照明节能企业展示了电梯更新换代技术、照明节能解决方案,以及相关优惠政策与服务。