

当AI“一本正经胡说八道”……

新华社“新华视点”记者 颜之宏 胡林果

当前, AI正赋能千行百业,为人们的工作、学习、生活带来极大便利。与此同时,不少人发现,用AI搜索数据,给出的内容查无实据;用AI辅助诊疗,出现误判干扰正常治疗……AI频频上演“一本正经胡说八道”。社交平台上,AI幻觉引发热议。

AI好用但不时像是“中邪”了

用AI检索海量信息、让AI辅助查看三维病灶、打造AI互动课堂……如今, AI已深度融入现代生活,“人工智能+”产品赋能各行各业,从多个维度提供便利。

作为AI深度使用者,“95后”女生瑞希坦言,AI好用,但不时像“中邪”了一样胡说八道。“我让AI推荐10本高分小说,结果一多半都是它编的。反复确认后,它承认虚构了答案。”

现实生活中,不少人遇到相似情况。业内人士表示,这是由于AI幻觉导致。“AI可以快速给出答案,但生成内容可能与可验证事实不符,即凭空捏造;或生成内容与上下文缺乏关联,即‘答非所问’。”一名主流人工智能厂商技术人员说。

记者使用一款AI软件,让其给出某行业未来市场规模及信源,AI迅速回答称某投资机构预测2028年该行业的市场规模将达到5万亿美元,并提供相关链接,但链接页面找不到上述信息。记者看到,页面内容虽然包含该投资机构名称和5万亿美元表述,但预测数据并非该机构作出,且不存在2028年时间节点。

社交平台上,AI幻觉相关话题浏览量达数百万,网友吐槽涉及金融、法律、医疗、学术等多个领域。

第三方咨询公司麦肯锡研究院近期发布的2025年高校师生AI应用及素养研究显示,四千余名受访高校师生中,近八成遇到过AI幻觉。今年2月,清华大学新媒体沈阳团队发布的报告指出,市场上多个热门大模型在事实性幻觉评测中幻觉率超过19%。

AI幻觉已经影响了人们的生活与工作。

近期,一名国外男子被诊断出溴中毒。他此前询问AI,过量食用食盐不利于身体健康,有无食盐替代品,AI回答称可以用溴化钠代替。但溴化钠存在一定毒性,需要严格遵医嘱服用。该男子用溴化钠代替食盐三个月后出现精神错乱等症状。

这几年,美国多起案件中的律师因在法律文件中使用AI生成的虚假信息,被法院警告或处分。

AI幻觉为什么会发生?

受访专家认为,AI幻觉的背后存在多重因素。

——数据污染。AI“养成”过程中,数据“投喂”是关键环节。研究显示,当训练数据中仅有001%的虚假文本时,模型输出的有害内容会增加11.2%;即使是0001%的虚假文本,其有害输出也会相应上升7.2%。

奇安信集团行业安全研究中心主任裴智勇解释说,人工智能大模型需要海量数据,训练数据来自开源网络,难免会错误学习一些虚假、谬误数据,还有一些不法分子会恶意进行“数据投毒”。

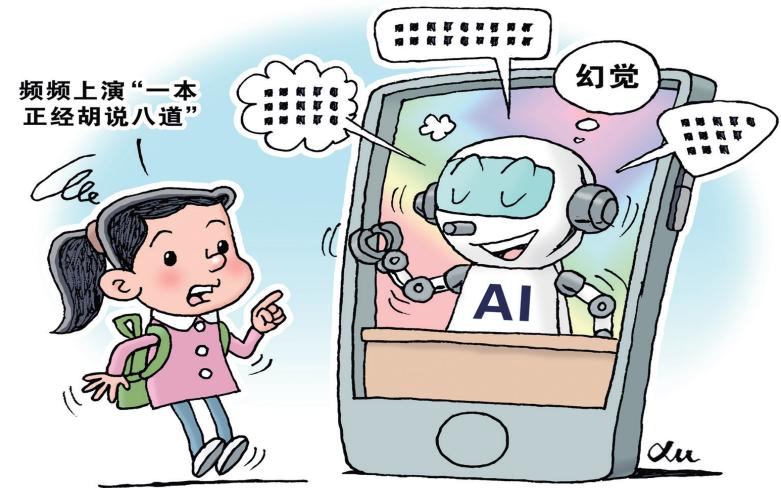
“如果把AI比作一个学生,数据污染就像是给学生看了错误的教科书,自然会导致‘胡说八道’。”暨南大学网络空间安全学院教授翁健说。

——AI本身“认知边界模糊”。翁健认为,人类智能的一个重要特征是“元认知”能力——知道自己懂什么、不懂什么,而当前AI技术架构缺乏这种自我认知机制。

翁健解释称,AI可以博览群书,但并不一定理解书里的内容,只是根据统计规律把最有可能的词语组合在一起,在准确评估自身输出的可信度方面尚存盲点。

——人为调校和干预。在中国通信学会数据安全专业委员会副主任委员左晓栋看来,相较于事实真相,AI更在意自己的回答是否契合用户需求,从而导致AI有时为了“讨好”用户而编造答案。

“针对不同需求,AI的训练、打分方式也不同。”一位从事大模型训练的技术人员说,当面对写作等创意性需求时,偏理性的事实严谨在打分系统中占比相对较低,偏感性的词语优美、富有感情色彩等占比更高。“所以可能会出现一篇辞藻华丽但词不达意的文章,里面内容甚至与事



实相悖。”

多方合力减少AI幻觉

第55次《中国互联网络发展状况统计报告》显示,截至去年12月,有249亿人使用过生成式人工智能产品,占整体人口的17.7%。受访专家表示,应通过多方合力应对AI幻觉带来的风险挑战。

今年4月,中央网信办印发通知,在全国范围内部署开展“清朗·整治AI技术滥用”专项行动,训练语料管理不严、未落实内容标识要求、利用AI制作发布谣言等均为整治重点。

“可靠、可信、高质量的数据对降低AI幻觉非常重要,应优化人工智能的训练语料,用‘好数据’生成‘优质内容’。”左晓栋认为,可以加快推动线下数据电子化,增加“投喂”的数据量;同时探索建立具有权威性的公共数据共享平台,“各大厂商也应加强优质数据筛选,提升训练准确性”。

多家主流人工智能厂商已经采取措施,从技术层面减少AI幻觉发生。

豆包升级深度思考功能,由先搜后想变为边想边搜,思考过程中可以基于推理多次调用工具、搜索信息,回复质量明显提升;通义千问在20多个通用任务上应用强化学习,增强通用能力的同时纠正不良行为;元宝持续扩充引入各领域的权威信源,在回答时交叉校验相关信息,提高生成内容的可靠性。

翁健建议,建立国家级人工智能安全评测平台,就像生物医药新药上市前要做临床试验一样,大模型也应该经过严格测试;同时,相关平台要加强AI生成内容审核,提升检测鉴伪能力。

“AI可能‘欺骗’用户,公众应客观认识人工智能的局限性。”左晓栋等专家提示,可以通过改进使用方式,如给出更加明确的提示词、限定范围等避免AI幻觉。“无论是工作、学习还是生活,现阶段的人工智能还不能全面替代人类的认知和创造能力,大家在使用AI时要保持怀疑态度和批判思维,不过度依赖AI给出的回答,多渠道验证核查。”

透视第三届全国技能大赛:技能人才如何向“新”?

新华社“新华视点”记者 张晓洁 孙清清 刘振坤

金秋时节,能工巧匠收获劳动果实。

第三届全国技能大赛23日落下帷幕,348名选手获得106个项目的金银铜牌。从占比过半的新领域赛项,到应变求新的参赛选手,本届大赛新意迭出。

大赛选拔了哪些紧缺技能人才?技能提升如何与产业换新同频共振?又该如何把握新机遇、开辟职业发展新天地?

研究中心主任张成刚说。

他认为,一方面,人工智能训练师、碳排放管理员等专业性较强的新职业为高技能、高学历人才提供了施展才华的新舞台;另一方面,数字经济、平台经济蓬勃发展,直播带货、数字营销、短视频剪辑等岗位对应的新职业打破传统职业边界,为更多群体创造了进入劳动市场的机会。

新技术:折射人才需求新趋势

本届大赛比拼了哪些新技能?选手们的“真功夫”中,折射出怎样的人才需求新趋势?

“机电一体化项目融合机械、液压气动、传感器、计算机等多学科技术,比赛难度一届比一届加大,要求选手技能不断提升,具备设备组装、电气连接、程序编写与调试等多方面能力,与企业自动化生产线更加融合。”机电一体化项目裁判宗玉说,不少裁判都感到选手技能水平大幅提高。

硕士毕业入职首都航天机械有限公司的选手马辰昊介绍,智能制造工程技术项目考察选手在智能产线开发、生产管控、智能运维等领域的综合应用能力,既要在技术上懂工艺流程和工具设备,也要在技能上稳定高效落地实施。

“这就意味着我们要有更广阔的知识体系、前沿技术的创新思维、扎实稳定的实践能力,努力成为学科交叉、能力全面的复合型人才;保持学习力与创新力,从理论知识出发,在项目实践中掌握、到应用场景中创新。”马

辰昊说。

本届大赛参赛选手中,博士、硕士、本科人数占比33%,高学历参赛选手主要集中在智能制造、高端装备等前沿技术技能领域。“这表明,技术技能融合发展的大趋势,对不同学历层次人才技术技能水平提升的需求加大。”人力资源社会保障部职业能力建设司司长吴礼舵说。

参与污水处理项目比拼后,郑州市南曹污水处理厂一期工程项目部副经理陈超越感到,技能提升需久久为功。“污水处理正向数字化、智能化不断发展。我们在工作中要对AI调控等持续加强学习,更好适应行业发展趋势。”

大赛明确规定各代表团国赛精选项目选手中,企业职工占比不低于50%。“目的就是通过大赛更好助力职工创新创造,让企业职工认识到自身价值,带动更多职工钻研技术、提高技能。”吴礼舵说。

新机遇:开辟职业发展新天地

大赛期间,浙江海天教育科技有限公司总经理陈兴对与企业所在领域相关的项目颇为关注。“作为深耕行业50多年的注塑机制造商,海天塑机集团需要大量精通数字化运维、自动化调试的现场工程师,对人才之‘渴’、对技能之‘盼’,感受尤为深切。”

近日发布的《北京市新质生产力人力资源开发目录(2025年版)》显示,新一代信息技术产业的5G-A技术研发与应用、人工智能产

业的具身智能技术研发等18个核心领域人才紧缺程度最高。根据《成渝地区双城经济圈人才需求目录(2025)》,四川六大优势产业人才需求占比最高前两位为医药健康、装备制造。

在大赛期间,中华职业教育社副理事长谢经荣建议,加强产教融合,让企业深度参与人才培养全过程,通过开设开拓思维、发明创造类课程,搭建实践创新平台,引导学生将理论知识与生产实际相结合,助力企业攻克技术难关、实现技术革新。

黑龙江林业职业技术学院教师段嘉琪此次参与了工业互联网工程技术项目比拼。“项目紧密贴合企业实际需求和行业发展趋势,让我对工业数据采集、工业互联网平台搭建等有了进一步理解,赛事中的典型任务可以转化成课堂实例和实训项目,让学生在解决问题中掌握核心技能。”

让技能成才、技能报国之路更顺畅,离不开薪酬待遇的改善、职业发展通道的畅通。本届大赛在奖励政策上不仅为获奖选手颁奖,还对国赛精选项目的金牌职工选手授予“全国技术能手”称号,让技能人才有荣誉、有获得感。

记者从大赛中了解到,人力资源社会保障部将全面推行“新八级工”职业技能等级制度,高质量开展技能等级评价,完善“新八级工”配套保障制度,推动职业资格、职业技能等级与相应职称、学历比照认定,积极搭建技能人才成长“立交桥”,不断提升技能劳动者的待遇和声望。