

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。

第三方咨询公司麦可思研究院近期发布的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0.01%的虚假文本时,模型输出的有害内容会增加11.2%;即使是0.001%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’。”暨南大学网络安全学院教授翁健说。

——AI本身“认知边界模糊”。翁健认为,人工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。

——人为调和和干预。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2.49亿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“清明·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,相关平台加强AI生成内容审核,提升检测鉴别能力。

“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。”

石家庄桥西检察院开展防范金融风险宣传活动

为进一步守护好群众的“钱袋子”,近日,石家庄市桥西区人民检察院联合多家单位,在盛世御城社区开展防范金融风险、打击非法集资宣传活动。

活动现场,检察干警通过张贴宣传海报、发放普法手册、开展现场咨询等方式,向

群众普及了非法集资和金融诈骗的常见手段及危害,帮助群众认清犯罪分子的惯用伎俩。本次活动共发放宣传资料1000余份,有效提升了群众识别金融风险和自我保护能力,取得良好的社会效果。

许明

中国人寿:筑牢反欺诈防线 守护保险市场健康发展

近年来,面对日益复杂多样的保险欺诈行为,中国人寿保险股份有限公司积极构建全方位、智能化、高效协同的反欺诈防御体系,持续夯实风险防控基础,切实维护消费者权益与行业诚信。

公司通过“三层防线”明确职责分工:治理层定政策、运作层强执行、监督层查漏洞,形成高效协同的内控机制。科技赋能方面,依托大数据与人工智能技术,实现从投保到理赔的全流程智能风控,做到事前预警、事中拦截、事

后回溯,有效识别和拦截欺诈风险。中国人寿还注重多方协作,与司法机关、行业组织建立常态化合作机制,积极参与案件打击与反欺诈宣传。近年来,该公司向公安机关报送线索200余条,助力破获多起诈骗案件,并通过各类宣传活动提升公众防范意识。

中国人寿表示,未来将继续以科技为支撑,深化内外协同,持续织密反欺诈防护网,为行业健康发展与社会信用体系建设贡献力量。

李丛 刘辉



近日,邯郸市公安交巡警支队邱县大队民辅警走进辖区小学,开展交通安全宣传活动。民辅警针对同学们日常出行特点和校园周边道路实际情况,详细讲解常见交通标志标识、交通法规等交通安全知识,引导同学们提高交通安全意识,安全文明出行。

苗文金 摄

女子大意丢失手提包 高速交警快速寻回

本报讯(王智勇 苑耀光)近日,一名女士一时大意,将手提包遗忘在京港澳高速沙河服务区洗手间,等返回找寻时包已不见踪影。该女士报警后,高速交警七里河大队迅速响应,快速找回了遗失手提包并物归原主,赢得了群众赞誉。

9月12日16时30分,高速交警七里河大队接到一名女士的

求助电话,其称在京港澳高速沙河服务区休息期间,不慎将手提包遗落在洗手间,包内装有重要证件及数千元现金,希望民警能够帮助寻找。接警后,高速交警七里河大队民警迅速响应,立即赶赴沙河服务区展开调查。

通过与该女士详细沟通,并结合对服务区保安、保洁人员的询问信息,民警判断手提包遗失

时间较短,有可能被其他旅客拾获。他随即协调服务区经理调取相关区域的公共视频进行仔细排查。经反复比对视频资料,民警发现一名旅客在经过洗手间时发现了这个手提包并带走,随即上车离开。民警根据车辆号牌信息通过公安系统进一步查到车主联系方式,并第一时间与其取得联系。对方接到民警

电话后,承认确实在沙河服务区捡到一个手提包。经核对,包内存放的证件和现金与求助者描述完全一致。

由于对方车辆已驶入高速,经民警协调,拾捡人员约定与失主在前方最近服务区进行物品交接。拿到失而复得的手提包后,求助女士立即给民警打来电话,再三表示感谢。

任丘市人民医院:包联共建“主动延伸”口腔健康科普进校园

近日,任丘市人民医院口腔科医生走进任丘市第一实验小学,为学生们开展健康口腔知识科普公益讲座。

讲座中,医生围绕口腔疾病患病特点以及如何保护牙齿、正确的刷牙方法等知识进行讲解,提高了学生们对口腔健康重要性的认识,也促进了学生们充分发挥“小手拉大手”的作用,助推全市居民健康素养水平不断提升。

今年以来,任丘市人民医院高度重视包联共建工作,构建“院级领导统筹—职能科室

推进—临床科室支撑”的包联架构,具体由每位分管副院长包联2至3个乡镇或社区,职能科室对口包联具体卫生院(卫生服务中心),精准对接需求,全力推动医疗服务从“被动响应”向“主动延伸”转变,持续开展义诊等惠民服务,定期举办健康讲座,为群众提供健康咨询和疾病诊断服务,助力群众提升健康意识与素养,树立科学健康观念。

图为医生正在为学生们讲解口腔健康知识。

文/图 席娟

工伤赔偿款遭公司拖欠 法官妥善执行全部到位

本报讯(于海伟)近日,承德县人民法院成功执结三起涉工伤赔偿系列仲裁裁决案,在保障劳动者合法权益的同时,兼顾涉案企业经营发展,达到了良好的执行效果。

申请执行人钱某、多某、姜某均系被执行人承德某建筑安装工程公司的木工。三人在工作中不幸受伤,经相关部门认定为工伤。后三人与公司就工伤待遇问题产生争议,

经仲裁委员会调解,双方达成和解协议:承德某建筑安装工程公司于今年3月31日前一次性支付三申请执行人就业补助金、停工留薪期工资合计10.9万元。然而,履行期限届满后,被执行人未能按期履行生效法律文书确定的义务。为维护自身合法权益,三申请执行人遂向法院申请强制执行。

执行立案后,承办法官第一时间向被执行人发出执行通

知书、报告财产令,但被执行人仍未按要求履行义务。为快速推进案件执行,承办法官依法对被执行人公司账户进行冻结并向被执行人送达相关法律文书。账户被冻结后,被执行人主动联系承办法官,表示该账户为企业经营账户,冻结后导致公司无法正常开展交易,严重影响经营运转,申请解除冻结并承诺先行支付部分款项,待公司项目尾款到账后结清剩余欠款。

办案法官秉持优化营商环境、善意文明执行的理念,在充分保障申请执行人合法权益的前提下,积极组织双方就还款事宜进行沟通。经协商一致后,法院依法解除了对被执行人账户的冻结。账户解冻后,被执行人信守承诺,先行履行了部分款项。公司项目尾款到账后,被执行人迅速将剩余欠款足额支付,三起案件顺利执结。

