

# 当AI“一本正经胡说八道”……

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。”

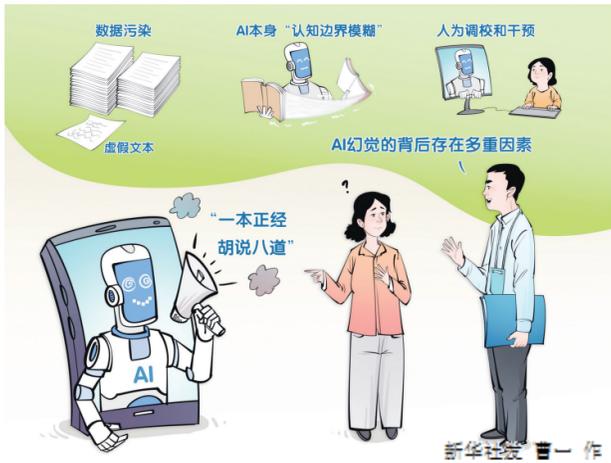
现实生活中,不少人遇到相似情况。业内人士表示,这是由于AI幻觉导致。“AI可以快速给出答案,但生成内容可能与可验证事实不符,即凭空捏造;或生成内容与上文缺乏关联,即‘答非所问’。”一名主流人工智能厂商技术人员说。

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,四万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制作发布谣言等均列为整治重点。

“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。”

(新华社广州9月24日电)

# 两部门紧急预拨1.5亿元支持粤琼闽做好防汛防风应急抢险救灾工作

新华社北京9月24日电 受超强台风“桦加沙”影响,华南多地发生严重暴雨洪涝、地质灾害。记者24日从财政部了解到,财政部、应急管理部当日紧急预拨1.5亿元中央自然灾害救灾资金,支持广东、海

南、福建做好防汛防风应急抢险救灾工作,用于转移安置受灾人员、排危除险等应急处置、开展次生灾害隐患排查和应急整治等,最大程度减轻灾害影响,切实保障人民群众生命财产安全。

# 农业农村部部署台风“桦加沙”防范应对工作

新华社北京9月24日电 针对今年第18号台风“桦加沙”将于9月24日在广东沿海登陆,给多地带来明显风雨影响,农业农村部近日部署了防范应对工作。

农业农村部近日印发通知,要求各级农业农村部门牢固树立底线思维、极限思维,以时时放心不下的责任感,切实抓好防御工作。同时,联合中

国气象局发布农田渍涝和风灾风险预警,并对福建、广东、广西、海南4省区启动农业防汛防风四级应急响应,指导各地在强降雨来临前抓紧抢收成熟果蔬、疏通田间沟渠、加固农牧渔业设施,雨后尽快排水清淤、加强肥水管理,并做好病虫害监测防治。组织出海作业渔船及时返港避风,以及海上养殖人员转移上岸。

# 神舟二十号航天员乘组将于近日择机实施第四次出舱活动

新华社北京9月24日电 记者从中国载人航天工程办公室获悉,根据计划安排,神舟二十号航天员乘组将于近日择机实施第四次出舱活动。

自8月15日圆满完成第三次出舱活动以来,神舟二十号航天员乘组稳步推进空间生命科学与人体研

究、微重力物理和空间新技术与应用等领域实(试)验任务,完成医疗救护在轨训练及全系统压力应急演练,持续开展站内环境监测、设备检查维护、物资整理等工作。

目前,空间站组合体运行稳定,神舟二十号航天员乘组状态良好,已做好出舱活动各项准备工作。

# 今年前8月全国新开工改造城镇老旧小区2.17万个

新华社北京9月24日电 记者从住房和城乡建设部获悉,今年前8个月,全国新开工改造城镇老旧小区2.17万个。

2025年,全国计划新开工改造城镇老旧小区2.5万个。据介绍,2019年至2024年,全国累计开工改造老旧小区28万个,惠及居民4800万户、超过1.2亿人,共改造提升各类老化管线36万公里,增设停车位387万个,建设养老、托育等各类社区服务设施7.8

万个。

住房和城乡建设部相关负责人表示,下一步,住房和城乡建设部将在完成“十四五”城镇老旧小区改造任务的基础上,继续深入推进这项工作,重点做好小区内老化管线改造,消除安全隐患。开展建筑物屋面、外墙、楼梯等公共部位维修,小区环境及配套基础设施改造,小区内建筑节能改造,支持有条件的楼栋加装电梯,改善群众居住条件和居住环境。

# 我国成功发射吉利星座06组卫星

9月24日15时56分,我国太原卫星发射中心在山东日照附近海域使用捷龙三号运载火箭,成功将吉利星座06组卫星发射升空,卫星顺利进入预定轨道,发射任务获得圆满成功。

新华社发 安迪 摄



# 潍宿高铁山东段建设稳步推进

9月24日,在潍宿高铁山东段,中铁十四局施工人员进行安丘汶河特大桥40米简支梁浇筑施工。

9月24日,国家重点项目建设(坊)宿(迁)高速铁路山东段安丘汶河特大桥十榀40米简支梁全部浇筑完成,项目承建方抢抓工期与节点,稳步推进标段施工。

潍宿高铁是京沪高铁辅助通道的重要组成部分,正线全长约399公里,设计时速350公里。新华社记者 朱峥 摄

中国移动 China Mobile 移动爱家 让家更有AI

移起AI 移动爱家计划 爱家让家更有AI

欢迎礼 爱家礼 聚会礼

AI家有心意 爱在亲情圈

1站式管全家通信消费 自己/家人/业务全搞定

5类资源全家一起用 设备/业务/流量/应用/权益

亲朋好友同享3重礼 欢迎礼、爱家礼、聚会礼,加入/体验/分享都有礼

让我们一起加入 移动爱家计划吧

郭晶晶 移动爱家大使

广告