

AI频频上演“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。

1 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。

2 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。”

3 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作临床试验一样, 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

外卖平台服务“新国标”向社会征求意见

新华社北京9月24日电 记者24日从市场监管总局了解到, 由市场监管总局组织起草的《外卖平台服务管理基本要求(征求意见稿)》, 24日向社会公开征求意见。

征求意见稿聚焦平台收费、促销行为等重点问题, 帮助外卖平台企业规范服务管理、提升服务质量,

减轻商户经营负担, 引导平台企业公开有序竞争。据介绍, 国家标准计划《外卖平台服务管理基本要求》, 由全国平台经济治理标准化技术委员会归口, 主管部门为市场监管总局, 主要起草单位有中国标准化研究院、中国网络安全审查认证和市场监管大数据中心等。

全球首个2000伏特光伏实证认证基地落地海南

据新华社海口9月23日电 记者23日从由中国质量认证中心等单位联合主办的“光伏实证认证暨2000伏特光伏技术创新研讨会”获悉, 此次会议期间, 全球首个2000伏特光伏实证认证基地在文昌市揭牌, 由大唐海南能源开发有限公司建设并运营。该基地具备全周期数据采集与系统化性能评估能力, 标志着我国在高电压等级光伏实证领域实现重大突破。

据了解, 当前光伏产业在高质量转型中, 面临2000伏特等高电压技术推广缺乏统一标准体系问题。2000伏特光伏技术作为未来大型地

面电站的发展方向, 虽较传统1500伏特系统减少电缆用量、降低度电成本, 但其长期运行稳定性、极端环境适应性及设备兼容性仍需权威实证支撑。

大唐文昌光伏实证基地将系统开展高温高湿、高盐雾典型气候环境下光伏产品的长期可靠性验证。通过统一测试边界条件, 实时监测发电性能、辐照度等20余项参数, 为光伏组件、逆变器、支架等关键设备提供科学的户外运行数据, 有力支撑产品选型、技术迭代与投资决策, 推动光伏产业从“量”的扩张转向“质”的升级。

新的世界第一高桥9月28日正式建成通车

新华社贵阳9月24日电 经过三年多建设, 新的世界第一高桥——贵州花江峡谷大桥将于9月28日正式建成通车。

记者24日从贵州省人民政府新闻办召开的新闻发布会上获悉, 横跨贵州黔东南州与安顺市交界处的花江峡谷大桥是六枝至安龙高速公路的控制性工程, 大桥全长2890米, 主跨径1420米, 是目前世界山区峡谷第一大跨度钢桁梁悬索桥; 其桥面距水面垂直高度625米, 超越北盘江第一桥成为新的世界第一高桥。

贵州省交通运输厅党委书记、

厅长张胤说, 一线建设者攻克峡谷强风、复杂地质等难题, 让花江峡谷大桥屹立于被誉为“地球裂缝”的花江大峡谷之上。大桥在抗风设计、高空施工等领域实现了多项技术突破, 获得授权专利21项, 多项技术成果纳入国家桥梁建设标准。

据了解, 截至目前, 贵州已建和在建桥梁超过3.2万座, 在世界高桥前100名中, 贵州占据近半数席位。从崇山峻岭到万桥飞架, 从千沟万壑到“高速平原”, 不仅展现了贵州自信、开放的形象, 也是中国基建实力的缩影。

金饰克价站上1100元

新华社北京9月23日电 随着美联储降息“靴子”落地, 国际金价延续连日来的涨势。受此带动, 以人民币计价的黄金价格再创新高。23日, 有品牌足金饰品报价首次达每克1100元。

9月23日, 伦敦现货黄金盘中一度升至每盎司3750美元上方。受此影响, 上海黄金交易所现货黄金价格23日盘中来到每克850元上方, 上海期货交易所上市的黄金期货主力合约一度涨至每克855.6元, 均创新高。

今年以来, 国际现货黄金涨幅已近43%, 以人民币计价的现货黄金价格累计上涨超36%。

“美联储如期降息后, 市场关注点转向年内后续降息路径。”东方金诚分析师瞿瑞认为, 美联储降息将直接降低持有无息黄金的机会成本, 同时打压美元指数, 此外地缘冲突仍然频发, 这些因素均对金价构成支撑。

贺利氏贵金属中国区交易总监陆伟佳表示, 8月份, 上海黄金交易所黄金实物出库量出现逆季节性下滑, 高金价对亚洲地区黄金实物消费的抑制逐渐显现。从当月黄金ETF来看, 尽管全球黄金ETF保持净流入, 但亚洲地区的投资者反而减持4.8吨, 这导致亚洲交易时段黄金涨势相对乏力。