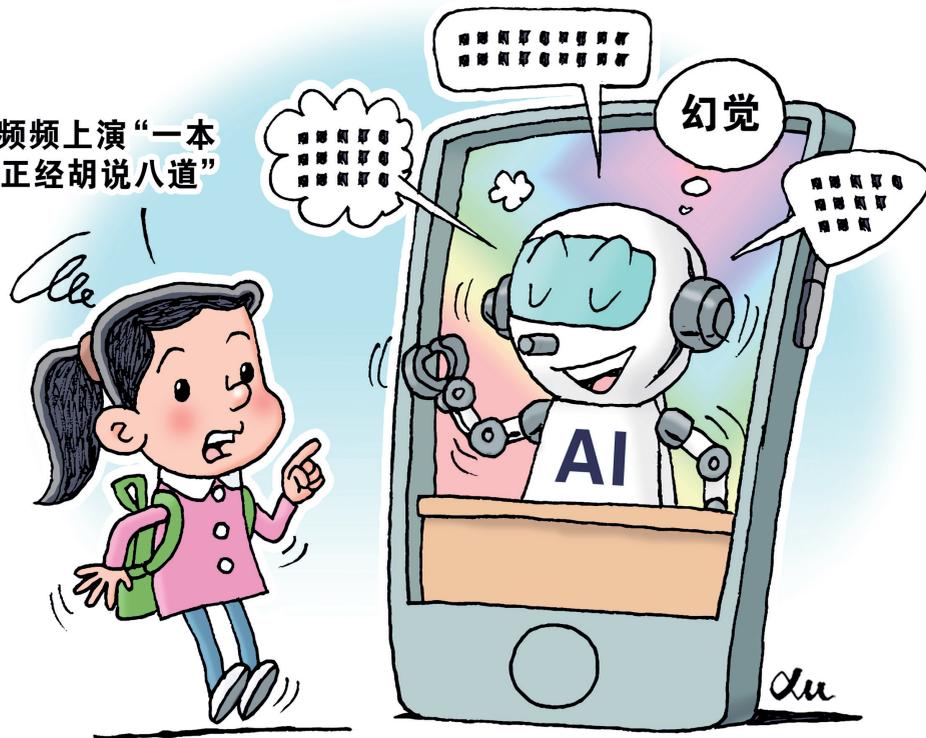


当AI“一本正经胡说八道”……

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。

频频上演“一本正经胡说八道”



频频出现的AI幻觉。新华社发 徐俊作

2024年我国哪吒企业数量达390家

据新华社天津9月24日电(记者郭方达、梁姊)2024年我国共有390家哪吒企业, 其中新晋367家, 有19家同时为独角兽企业、189家同时为潜在独角兽企业。这是记者24日在2025中国民营企业投融资洽谈会上了解到的消息。

哪吒企业指成立时间不超过3年(含)、A轮(含)前累计获得私募融资金额1亿元人民币及以上, 具备技术颠覆性强、赛道新颖、场景裂变快等硬核特征的科技创新型企业。从产业方向来看, 哪吒企业已经成为发展新质生产力的重要力量, 共分布于35个赛道, 超六成聚集在集成电路、创新药、清洁能源、机器人、人工智能领域。

报告显示, 哪吒企业群体累计融资额超2100亿元人民币, 平均融资额超5亿元, 大额融资主要聚集在集成电路和人工智能赛道。国资成为投资主要力量, 有66家国有产业型大企业参投68家哪吒企业, 765家国有基金参投296家哪吒企业。超六成哪吒企业聚集在北上苏深。

北斗卫星导航定位技术在大众消费市场占比超过七成

据新华社长沙9月24日电(记者刘芳洲、谢奔)第四届北斗规模应用国际峰会24日在湖南株洲开幕, 会议期间, 《北斗产业发展蓝皮书(2025年)》正式发布。蓝皮书显示, 北斗卫星导航定位技术加速融入普通民众生活, 在智能手机、可穿戴设备、共享出行等领域落地产品日益丰富, 在各类终端中占比超七成。

当前, 北斗系统在大众消费领域的应用布局持续深化, 在智能手机、可穿戴设备、共享出行、电动自行车等核心场景展现出蓬勃发展活力。

蓝皮书显示, 2024年全年国内智能手机出货量2.94亿部, 同比增长6.5%, 其中约2.88亿部智能手机支持北斗定位功能, 北斗安装率接近98%, 智能手机出货量和北斗安装率均有所提升, 成为北斗在大众消费领域最大的应用载体。

与此同时, 可穿戴设备与北斗系统深度融合。可穿戴设备已成功跃升为北斗应用的第二大市场。根据中国信息通信研究院预测, 2028年, 全国支持北斗的可穿戴设备年出货量将增长到4795万台, 可穿戴设备北斗安装率将逐步增长至74%以上。

AI好用但有时像是“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, 但有时像“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。

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。”

(据新华社广州9月24日电“新华视点”记者颜之宏、胡林果)