

新华

不久前,某公众号运营者钟某为蹭“海啸预警”热点,利用AI生成“30万居民撤离上海”“数十万上海市民正排队登车”等耸人听闻的言论。这些假消息被大量阅读、转发,引发民众恐慌情绪。其后,钟某被警方拘留。

AI的广泛应用,为虚假信息的批量生产和病毒式传播提供了“温床”,给社会治理带来挑战。

9月15日,《人工智能安全治理框架》2.0版正式发布。中央网信办负责人表示,要强化规范引导,不断完善人工智能安全监管制度和标准规范体系,促进人工智能健康有序发展。

AI造谣高发

民生行业成谣言“重灾区”

人工智能时代,利用AI造谣日渐频繁。公安部曾发布多起典型案例。如,湖南公安机关查处编造“民警离职潮”谣言案,江西公安机关查处MCN机构使用人工智能工具造谣案,重庆公安机关查处使用人工智能工具编造“爆炸事故”谣言案,广东公安机关查处编造“广州限制外卖配送”谣言案。相关人员均被绳之以法。

清华大学新闻与传播学院新媒体研究中心2024年发布的《揭秘AI谣言:传播路径与治理策略全解析》研究报告指出,2023年以来,伴随AIGC技术的高速发展,一些造谣者不当利用AI工具,使得AI谣言量高速增长。

报告指出,在各类AI谣言信息量占比中,经济与企业类谣言和公共安全类谣言占比最多、增速最快,其中餐饮外卖、快递配送等民生行业成为谣言重灾区。

记者调查发现,人工智能技术让网络谣言内容更为逼真,常配有伪造的图片、视频甚至所谓“官方回应”,极具迷惑性。

中国(深圳)综合开发研究院院长助理、数字经济与全球战略研究所所长曹钟雄等业内人士认为,AI不仅能高效生成虚假信息,而且可以通过过滤、推荐机制,将虚假信息聚合形成“证据链”,给普通用户乃至专业人士造成误导。

谣言生成“提效” 批量生成虚假内容

不久前,某著名食品公司向安徽省合肥市公安局经开分局报案称,多个网络平台上集中出现大量虚假、带有诱导性的负面文章,恶意攻击公司品牌及产品质量。公安机关立案调查,民警在多个平台发现关于该品牌“二氧化硫残留”“霉变”等内容的文章200余篇,且内容高度雷同。

经公安机关核实,上述内容均为虚假信息。信息发布者陈某被行政拘留。陈某交代,今年8月,他在网上看到该公司相关信息后,想到知名企业涉及食品安全问题,容易引发公众关注。为博取流量,他使用AI工具编造内容,生成多篇虚假负面文章,在多个平台集中发布。

业内专家指出,传统谣言生产依赖造谣者人工操作。如今,造谣者使用AI工具,只要输入特定关键词或指令,便可快速生成内容逼真、欺骗性强的谣言,从而实现谣言的大规模、高效率生产。

去年3月,江西公安机关查处一起MCN机构使用人工智能工具造谣案。经查,王某某经营有5家MCN机构,共运营自媒体账号842个,长期组织公司人员选取热点文章,使用AI工具批量生成不实文章进行吸粉引流。

“他平均每天通过AI工具发布4000到7000条信息,最高的一条收入是700元。经初步估算,每天收入在1万元以上。”办案民警说。

AI批量造谣言背后藏着怎样「生意经」

造谣动机

流量变现与不正当竞争 催生黑色产业链

记者发现,多起AI造谣事件背后,不法者的造谣动机主要源于牟取互联网内容平台给予创作者的奖励,以及为电商平台经营引流等;一些造谣者还可以从广告收入、合同推广、直播带货或其他相关的商业活动中获益。

一名长期从事自媒体工作的博主对记者说:“以某大型内容生产平台为例,收益计算公式可以简化为‘有效播放量×互动系数×千次播放单价’。平台会对点击量高、互动性强的原创性内容给予更多收益分成。如果整体权重较高,一篇阅读或播放100万次的稿件,收益可超过1000元。”

此外,不正当商业竞争也进一步催生了AI谣言黑色产业链。一位头部消费电子企业法务相关负责人透露,“金主”企业放出任务信号后,公关公司会找来一批自媒体运营者、KOL(关键意见领袖),利用AI技术生成虚假信息,大批量、高频次在社交媒体、网络平台等释放抹黑竞争对手的内容;此后还有营销公司及流量商加入,进一步扩散传播谣言。

造谣一张嘴,辟谣跑断腿。不少受访企业表示,辟谣工作给公司带来了繁重的负担,需要耗费大量人力、精力去做解释工作,以消除不利影响。

中国政法大学刑事司法学院教授于冲等法律专家指出,目前AI造谣已经形成商业化运作黑色产业链,一些造谣者以吸引流量变现牟取经济利益为目的。相关行业企业在应对AI谣言中面临维权难、舆论修复难等挑战。

专家建言

全链条治理 封住AI“黑嘴”

面对AI谣言治理中的取证难、定性难、追责难等问题,单一治理手段已力有不逮。多位受访业内人士建议,遏制“AI谣言”需要多方协同发力,构建全链条治理体系。

中国社会科学院大学互联网法治研究中心执行主任刘晓春说,新媒体平台应加强对AI造谣和“水军”行为的智能识别机制研究,进一步改革涉及AI内容的流量分发和收益分成机制,最大限度压缩通过AI造谣牟利的空间。

专家指出,平台可通过逆向传播,利用相同渠道传播准确信息,抵消谣言影响;鼓励用户积极举报可疑内容,通过设立奖励机制增加参与度;持续研究谣言生成和传播机制,开发有效监测和防治技术等。

于冲等建议,司法治理也需要“链条化”思维,针对AI造谣链条的前端发布主体、中端采取的技术手段和末端的后果危害,分别进行评价和判断,进行全链条打击和预防。

最高人民法院研究室副主任喻海松等表示,治理涉企AI谣言,仅仅依靠刑法是不够的,要进一步修复社会信任机制,营造自觉抵制谣言的良好环境,同时鼓励AI治理合作,进一步推广AI伦理标准、治理技术和创新国际合作模式。

上海市律师协会理事、上海博和汉商律师事务所律师执行主任王思维说,人工智能目前仍仅是一种技术手段,而非可承担法律责任的独立主体。对AI技术滥用行为的治理,应当坚持防惩并重原则,在确保科技创新积极性、AI技术合法合规使用的情况下,避免过度监管。

受访律师提示,无论是个人还是企业,都应该自觉遵守相关法律法规,正确使用AI工具,不造谣、不传谣。一旦利用AI造谣或从事其他违法犯罪活动,将受到法律的严惩。

筑牢反诈“防火墙” 守护市民“钱袋子” 兴业银行福州分行反诈在行动

“阿姨,您今天取这么多现金是要付购房款吗?现在购房一般都会通过转账,您需要现金是有什么特殊原因吗?”

“对方说必须用现金,签合同的时候直接给!”近日,年逾75岁的刘阿姨,神色紧张地走进兴业银行福州大名城支行网点,坚持要取款40万元用于“购房”,当柜员提出要联系其子女核实时,也被她果断拒绝。

经验丰富的柜员迅速意识到,这可能是一起诈骗案件,立即报告支行会计暨合规主管协助劝阻,同时启动警银联动机制,将线索同步至当地派出所及反诈中心。警方也迅速响应,与银行工作人员一同劝说老人:“对方是不是让您交保证金或者手续费?我们最近也处理过相关的案子,要求现金交易后人就消失了,您别急着给钱,我们帮您核实对方身份!”

在反复劝说下,老人最终放弃了取款。经警方调查,该案件确为电信诈骗,40万元资金损失得以挽回,险些被骗的刘阿姨送上锦旗:“是你们的坚持和负责,才使我避免了巨额损失!”

今年以来,兴业银行福州分行通过账户资金交易、外部线索等多维度综合研判,识别了福州地区疑似受骗客户超540户,为客户挽回损失超2000万元。

兴业银行福州分行相关负责人表示,近年来,电信网络诈骗花样不断翻新,在此温馨提示广大市民,务必提高警惕,筑牢反诈防线,凡自称客服以“取消会员”“退款”为由要求下载软件指导进一步操作的,切勿轻信;公检法机关不会通过电话、网络要求转账汇款到“安全账户”或缴交认证核查费用;转账汇款前务必多方核实对方身份,请勿向陌生账户转账,遇到可疑情况及时拨打110或全国反诈劝阻专线96110咨询求助。

唐昊制图