2024年7月12日 星期五
责编 罗利 E—mail:aqwbzk@126.com
安庆晚報

场景一键生成、图文真假难辨

AI批量造谣防不胜防

近期，多地公安机关发布了多起利用AI工具实施造谣的相关案件。6月20日，上海市公安局城市轨道和公交总队官方微博@轨交幺幺零发布警情通报：两名营销人员为博取关注造谣地铁站发生持刀伤人的情况，被警方依法处以行政拘留。其中一人，使用AI软件生成视频技术，编造了地铁行凶的虚假视频等不实信息发布到网上，造成恶劣社会影响。

记者采访了解到，便利的AI工具大大降低了造谣的成本，提升了谣言的数量级和传播力。AI造谣呈现出门槛低、批量化、识别难等特点，亟待加强监管，斩断背后利益链条。

图片来源：网络

AI造谣频发

记者梳理发现，近年来，AI造谣在全国多地频发，传播快、门槛低、迷惑性强。

去年，在上海一女童走失事件中，一团伙以"标题党""震惊体"方式，恶意编造炒作"女孩父亲系继父""女孩被带往温州"等系列谣言。该团伙利用AI工具等生成谣言内容，通过114个账号矩阵，在6天内发布268篇文章，多篇文章点击量超过100万次。

AI造谣门槛很低。一名办案人员透露，此前在调查一起谣言案件时发现，谣言源头来自中部某省份的一个村落。经调查，该村有大量村民利用生成式AI撰写网络文章。

"部分村民通过口口相传获知，把搜索到的网络热点信息粘贴进AI模型后，可以生成新文章；在一些互联网平台上发布，即可通过流量赚钱。不少村民都自发从事该活动。"办案人员说。

迷惑性强，也是AI造谣的一大特点。

公安部网安局近期公布一起案例。2023年12月以来，一条"西安市鄠邑区地下涌出热水"的信息频繁在网络上传播，出现如"地下出热水是因为发生了地震""是因为地下热管道破裂"等谣言。经查，相关谣言是通过AI洗稿方式生成的。

记者发现，上述谣言采用类似新闻的行文，其中夹杂着"据报道""相关部门正对事故原因进行深入调查，并采取措施进行抢修""提醒广大市民在日常生活中要注意安全"等内容。普通人难辨真伪。

清华大学新闻与传播学院教授沈阳表示，AI的介入使得信息的生成和传播速度大大加快，同时也增加了判断信息真实性、准确性的难度。AI工具从技术上降低了不法团伙造谣、传谣的门槛，大幅提升了谣言的数量级和传播力，可能促使网络谣言"定制化"生成、"精准化"传播、"智能化"扩散。

清华大学新闻与传播学院新媒体研究中心今年4月发布的一份研究报告显示，近两年的AI谣言中，经济与企业类谣言占比最高，达43.71%；近一年来，经济与企业类AI谣言量增速高达99.91%，其中餐饮外卖、快递配送等行业更是AI谣言重灾区。

流量牟利

记者发现，多起AI造谣事件背后，造谣者的动机主要源于获取互联网内容平台给予创作者的点击量、阅读量奖励，以及为电商平台经营引流等。

浙江省绍兴市上虞区人民法院2023年宣判的一起案件中，不法分子为吸粉引流，通过非法渠道购买AI视频生成软件，将网上搜集的热门话题，通过AI一键生成视频，并发至多个热门视频平台上牟利。

经审查发现，截至案发，不法分子发布的《浙江工业园现场大火浓烟滚滚，目击者称有爆炸声！》等20余条虚假造谣视频，涉及浙江、湖南、上海、四川等多个省市，累计阅读观看量超过167万次。

记者调查发现，在自媒体平台上发布内容、博取流量已经成为一门生意。"图文创作，AI自动写文章，单号轻松日产500+，可多号操作，小白轻松上手""无需洗稿，AI一天100篇原创"之类的文章在互联网上广为流传。

以一家大型互联网内容平台为例，发布内容基础千次的阅读单价为0.12元，但通过一系列系数相乘，可以让千次阅读收益提高10倍。这意味着如果一篇文章阅读量达到100万次，创作者即能收益超过1000元。

一些人使用AI造谣的目的是为电商平台产品销售引流。今年2月，上海公安机关发现，在一家电商平台上出现了某艺人"命运多舛、含恨离世"等短视频，引发大量点赞和转发。

经查，该视频内容完全系伪造。视频发布者到案后交代，他在某电商平台上经营一家土特产网店。由于销量不佳，他便通过编造夺人眼球的虚假新闻给网店账号吸引流量。他不会视频剪辑，便利用AI技术生成文本和视频。

多管齐下强化治理

今年4月，中央网信办秘书局发布《关于开展"清朗·整治'自媒体'无底线博流量"专项行动的通知》，要求加强信息来源标注展示。使用AI等技术生成信息的，必须明确标注系技术生成。发布含有虚构、演绎等内容的，必须明确加注虚构标签。

专家表示，随着AI使用门槛降低，AI造谣问题可能加剧；建议压实互联网内容平台、AI内容生成平台等主体责任，阻断造谣背后的利益链条；加强法律研究和适用，提高造谣者、传谣者的违法成本。

北京师范大学新闻传播学院教授喻国明建议，AI内容生成平台要进一步完善对生成内容和内容来源的标识机制，例如附加不可删除的数字水印和文本说明。互联网内容平台可接入AI内容生成平台的大模型系统，实时进行数据识别比对。

中国社会科学院大学互联网法治研究中心主任刘晓春建议，互联网内容平台应加强对AI造谣行为的智能识别机制研究，进一步优化涉及AI内容的流量分发和收益分成机制，压缩别有用心者批量生产谣言的牟利空间。

喻国明认为，在辟谣方面，可利用AI辟谣应对AI造谣，"针对被确认为谣言的内容，互联网内容平台可利用AI技术向曾经浏览过相关内容的用户智能推送辟谣内容或进行风险提示，降低谣言破坏力。"

沈阳等业内专家表示，界定AI谣言产生是否存在主观恶意、认定被造谣方损失等存在一定困难，对AI造谣等行为在刑事处罚上容易出现法律适用难题，导致打击效果难以落地。"我们研究发现，在86起涉AI造谣相关案件中，惩处类型为刑事处罚的占比仅为15.66%。"沈阳说。

北京孟真律师事务所律师舒胜来认为，面对AI谣言等技术和犯罪形式的快速迭代，法律法规研究也要与时俱进。要进一步明确AI造谣犯罪的认定标准，加强行刑衔接，对AI造谣者"精准惩处"，做到罚当其罪、罚当其过，为广大网民营造风清气正的网络空间。

来源：经济参考报