用AI造谣“女儿被人抱走”，有人被行拘

AI整蛊别让“玩梗”越过红线

流浪汉闯入家中、女儿被人抱走……这样令人触目惊心的事情，竟然只是个玩笑。

近日，安徽铜陵的张先生在外地收到妻子李女士发来的图片，画面中一名流浪汉闯入家中，吓得他报了警。民警迅速出警，却发现是李女士用AI合成了流浪汉坐在自家餐厅的图片，试图“教训”爱喝酒的丈夫。经民警批评教育，李女士认错并公开道歉。

但闹剧并未收场，反而掀起模仿狂潮。有网友用AI制作“流浪汉闯入餐厅后厨”，老板吓到喊来保安；还有人为博关注用AI生成“女儿被人抱走”的“寻人启事”并发布到社交平台，最终被行政拘留……

如今，AI换脸、语音合成等技术触手可及，也让制造虚假信息变得轻而易举。若下一次真有危险发生，人们是否会因“狼来了”而迟疑？

李女士发送给丈夫的“流浪汉”闯入家中的照片（系AI生成）

1 10秒就能无中生有

图片视频“真假难辨”

在各大社交平台，依然还能搜出“流浪汉”相关AI指令和教程。这些帖子讲述的，都是如何用AI跟家人、朋友开玩笑的过程，并得意地附上聊天记录。看到对方焦急而紧张地回复，还会沾沾自喜。

记者尝试用某知名AI软件进行实测。先拍摄一张自家客厅沙发的照片上传，再输提示词：把一名流浪汉P到沙发上坐着。随后AI自动补充了画幅比例、高品质等常见指令。不到10秒，流浪汉坐在记者家沙发上的照片便出现在眼前。AI还给出了两种不同选择：一张是一名流浪汉单脚踩在沙发上，目光盯着镜头；另一张，有两名流浪汉坐在沙发上看电视。

记者对两名流浪汉的那张照片再次输入指令：把这张照片做成视频，让流浪汉动起来。AI回复：本次使用高品质模型生成，预计等待1~3分钟。大约1分钟，视频生成好了，两名流浪汉动起来了，一人左顾右盼，另一人抠手挠头，背景还营造出了手持相机的抖动感，甚至玻璃窗上也有倒影。视频时长5秒，不仅画面高清，还十分逼真。

在提前告知家人图片和视频是AI生成的情况下，家庭群里个个看了都觉得不可思议，也有急得立马打来电话，再三确认没事后才放心。若不是右下角有“AI生成”的字样，别说老人，许多年轻人也“真假难辨”。

简单几个指令便能“以假乱真”，不禁令人担忧，一些别有用心的人，会不会借此骗取他人信任并牟利。

2 AI标识并非万能

普通人该如何分辨

去年9月1日起，由国家互联网信息办公室等发布的《人工智能生成合成内容标识办法》开始施行，要求所有AI生成的文字、图片、视频等内容，都要添加标识以明确其来源和真实性。

“疑似包含AI创作信息，请谨慎识别。”记者在浏览某短视频平台时，一些AI生成的短视频文案下方，被平台标注了这样的提示，确保用户与监管方能够清晰识别AI属性。

“除了这种显式标识，隐式标识更重要。”杭州电子科技大学网络空间安全学院、浙江-法国数字媒体取证联合实验室主任乔通表示，以图像为例，像素之间存在大量肉眼无法察觉的信息，隐式标识可以通过微调部分像素值，将标识信息偷偷嵌入其中，“一旦有人利用AI生成伪造视频或者图像等内容，并在互联网进行不良传播，隐式标识就像一把‘定位钥匙’，帮助快速追溯内容来源。”乔通说。

此外，许多文档、图片、视频等内容本身都自带“元数据”，其中记录了创建时间、存储位置等基础信息，这些“元数据”也是“隐式标识”，等于“AI生成”的标识。但AI生成标识并非“万能”，用户可以通过“P图”“转码”等办法进行一定程度的规避。

“即使没有标识，仍然可以通过技术手段完成鉴别。AI生成的图片与真实图片的高频分量周期性存在差异，通过技术手段可挑出关键特征。”乔通说。

不具备技术手段的普通人该如何分辨？

乔通表示，AI生成的图片中，色彩、光线、人物、背景等会看起来非常完美，给人一种“P过头”的感觉；AI生成的高清视频，由于技术原因还不能达到非常高品质的程度，牙齿、脸部边缘等细节会有缺陷和不流畅。对于AI生成的音频，他认为，现在光靠人耳是很难分辨是AI还是真人，尤其是简短的语音，“总之，别转钱。”他提醒道。

如今，有许多AI产品通过真假参半来打“擦边球”。如一些宠物博主通过实拍和AI结合，让小猫一脚将球踢进球框，点赞数上百万。虽然视频只为给大众图个乐子，但是此类视频往往没有AI标识，让人难以鉴别、浮想联翩。“若发现内容与日常认知存在偏差，逻辑或感官上‘别扭’，就需带着怀疑态度看待与核实。”乔通说。

去年上半年，杭州电子科技大学网络空间安全学院与杭州公安联合建立深度伪造检测服务机构，只要有质疑，需要科普的都可免费监测。

3 开玩笑也要有底线

市场需要针对AI的“反诈APP”

玩笑的本质是善意互动，如果用AI玩笑开大了，则暗藏巨大风险。

金道律师事务所律师王杰认为，从法律层面看，行为人在被“整蛊”后陷入错误认知而报警，其本人不构成报假警，但整蛊者可能涉嫌扰乱公共秩序，根据《治安管理处罚法》，将面临5~10天拘留的处罚，最高可处1000元罚款。如果视频被转发从而引发社会恐慌，可能触犯刑法的“编造虚假信息罪”，最高能判5年以上有期徒刑。

“AI属于新事物，通过AI生成的图片或视频，真假越来越难以辨别。”王杰指出，针对人工智能的相关法律法规在逐步出台和完善，目前已有《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》《人工智能生成合成内容标识办法》等规范，对于AI生成物的用途、审核管理机制、数据和基础模型来源的合法性、知识产权及个人信息保护、显著标识添加等提出了具体化的要求。

随着人工智能引发的致损事件不断增加，王杰认为，首先要让更多民众了解AI能做到什么样，就像人人都知道电信诈骗，便会时刻有警惕。他呼吁，市场需要一个像“反诈APP”的官方平台，用来识别AI图像、音频等真伪；人工智能服务提供者，也需要加强内容审查，确保技术合法合规使用。

最高法院刑事审判第三庭庭长陈鸿翔表示，针对AI深度伪造带来的新情况新问题，最高法将加强研究，会同相关部门及时出台规范性法律文件或实施细则，完善法律适用。针对电信网络诈骗犯罪这一AI侵权的重要表现形式，他指出，诈骗分子利用AI深度伪造新技术不断翻新诈骗手法，防范打击难度越来越大。对此，人民法院将积极应对挑战，强化法律支撑，完善相关法律规范。

娱乐有尺度，玩笑有底线。AI整蛊的边界，从来不是技术能力的上限，而是法律与伦理的底线。在享受AI带来的便利与乐趣时，别让“玩梗”越过法律红线，别让技术凉了人心。

据《钱江晚报》