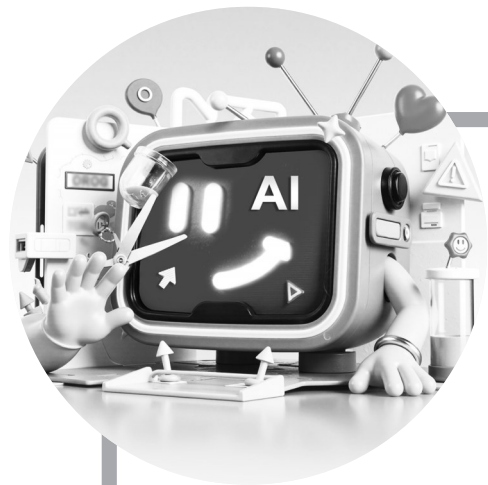




“AI生成画面”“疑似使用了AI生成技术,请谨慎甄别”“内容为AI技术制作”……不知你在浏览信息、刷短视频时是否注意到,一些图文、视频内容加上了这样的标识信息。

前不久施行的《人工智能生成合成内容标识办法》(以下简称《标识办法》)提出,对生成合成内容强制添加显式和隐式标识等规范要求,即用AI生成的每一段文字、每一张图片、每一条音频、视频,都必须强制标明“数字身份证”。

近日,公安部公布了10起打击整治网络谣言典型案例。其中便有一件案例为使用“高空坠楼、家庭悲剧”等关键词用AI生成视频片段,并经过二次剪辑后发布到网上。

截至10月31日,新规落地近两个月。尽管各大平台陆续上线标识功能并配套相应管理措施,但调查发现,短视频、图文笔记、直播间,仍有许多AI内容未亮明身份。

未标注的AI视频,是对社会信任基石的侵蚀

你刷的短视频,标注“AI生成”了吗?

你看的短视频是真是假?——AI已渗透生活各个角落

10月1日至10月18日,有媒体在六大短视频平台的推荐页面,分别随机浏览了100条内容。这六大平台上,约有一半的AI视频已有作者标注的创作声明,还有约三成内容,虽无作者声明,但被平台在显眼处标识为“疑似内容为AI生成”。

然而,未亮明任何标识的AI内容仍时常出现。10月12日,某社交平台博主“懒X”发布了一条短视频。画面中,一只卡通水獭头顶贝

壳,泡在加冰的柠檬茶里,憨态可掬。这条视频风格独特,高度疑似AI制作,但页面没有任何AI标识。

统计发现,自9月23日起,该博主已发布21条同类风格视频,均无任何AI“身份标识”。经联系核实,该博主承认视频确实为AI制作。

这并非个例。水獭属于国家二级保护动物,不少喜爱它的网友只能借助AI“云养水獭”。在该平台上,“#水獭”话题拥有超过8.1亿的浏览量。10月13日,在该话题下随

机浏览前20条内容发现,其中12条明显为AI制作。但在这12条中,9条没有任何标识,仅有3条为作者主动声明。

“目前平台对于AI生成内容的标识识别,主要依靠显性标识。”一位业内人士解释,受限于技术和算力成本,平台多依靠视频自带的AI水印或作者主动声明来识别。如果一段视频既无作者标注,画面又无水印,也非平台自家工具生成,识别难度将大大增加。

具体怎么进行“标注”?——显式标识和隐式标识

具体怎么“亮明身份”呢?《标识办法》明确,人工智能生成合成内容标识主要包括显式标识和隐式标识两种形式。

显式标识是指在生成合成内容或者交互场景界面中添加的,以文字、声音、图形等方式呈现并可以被用户明显感知到的标识,就像人脸一样可以快速区分,让普通用户“一眼可见”。如文本内容需在开头、结尾或关键位置标注“人工智能生成”或“AI生成”,字体需清晰可辨,不得刻意缩小或模糊等。

隐式标识更像指纹,是指采取技术措施在生成合成内容文件数据中添加的,不易被用户明显感知到的标识。它侧重于技术层面的追溯,要求在AI生成内容的文件内部嵌入元数据,且字段名称必须包含“AIGC”标识符号。元数据需涵盖AI生成确认状态、生成服务提供方信息、内容传播平台名称、唯一编号以及数字签名或哈希校验值等关键信息,为内容溯源与责任认定提供技术支撑。

中国电子技术标准化研究院

副院长范科峰认为,《标识办法》和《网络安全技术 人工智能生成合成内容标识方法》协同发布,开创了技术标准与管理制度配合实施的新型治理模式。作为我国人工智能领域首部“办法+强标”组合拳,其在管理维度明确生成、传播、分发全链条主体责任,在监管维度建立审核、监测、追责的闭环机制,在技术维度规范显隐标识体系并预留创新空间,这种三位一体的治理架构为人工智能安全治理提供了清晰的实施路径。

我们面临哪些挑战?——当“虚构”穿上“真实”的外衣

“儿子三年级,全班第一,每晚的真实强度。”10月18日,博主“复旦爸爸带娃记”发布的一条育儿笔记出现在某社交平台的首页推荐上。笔记先分享了孩子“不报补习班、早睡早起却能保持优异成绩”的时间管理经验,后话锋一转,推荐起某线上精读课程。

然而,翻看该博主当天发布的其他8条笔记,令人困惑的情况出现了:其“儿子”的年级在一至六年级之间来回变换,先后出现“儿子二年级,全班第一”“四年级全班第一的周末,学习和玩两不误”等相互矛盾的说法。而且,9条笔记使用的是同一张配图。

“这是典型的AI图配AI文。”某新媒体广告公司招商负责人陈先生点出。仔细查看发现,配图中孩子需要默写的“词语表”实际上全是乱码,显示出明显的AI生成痕迹。然而,这条笔记既没有作者主动标注的AI创作声明,也没有平台添加的疑似AI标识。

陈先生表示,这是教育行业新出现的“AI全包广告”投放模式,即先用AIGC软件批量合成图片和生成文案,再将其批量投放至粉丝数量低于1000人的普通用户账号发布,“看似经验分享,其实是产品引流,一条笔记的全包收费在150元至200元之间,比起

之前的人工种草笔记,内容端能省不少钱。”

生成式人工智能、深度合成等新技术快速发展,为生成合成文本、图片、音视频等提供了便利工具,在促进经济发展、丰富网上内容、便利公众生活的同时,也造成虚假信息传播、破坏网络生态等问题。

例如,一些未明确标识的AI生成合成的视频正扮演着“数字孙子”角色,通过提供情绪价值,将许多不明真相的老人哄得心花怒放,部分老人不顾家人劝阻,执意购买这些“孙子”在视频里推荐的产品,最终陷入骗局。

我们该如何应对?——在狂飙中建立“护栏”

技术的精进很快催生了新的商业形态。今年清明节前后,淘宝、拼多多等电商平台涌现出大批“AI复活逝者”服务,价格从99元到599元不等。商家宣称,只需提供一张清晰照片和100字以上的人物生平,就能生成“开口说话”的30秒视频。某店铺显示,其月销量已突破2000单。评论区里满是“谢谢让我再听外婆唠叨”“妈妈笑得和以前一样”的感慨。

事实上,监管已在加速跟进。2023年8月施行的《生成式人工智能服务管理暂行办法》明确规定,AI服务不得侵害他人肖像权、名誉权,生成内容需按规定标识。今年,浙江省出台《网络直播营销行为规范指引》,首次对AI数字人直播作出规范,要求“不得通过AI伪造他人声音肖像制作引流广告”。北京查处的李梓萌AI代言案,更是首次将AI虚假宣传纳入《中华人民共和国广告法》监管框架。

司法实践也在探索边界。杭州

互联网法院曾审结我国首例虚拟数字人侵权案,认定未经授权使用AI数字人构成著作权侵权,判决赔偿12万元。在“奥特曼虚拟形象侵权案”中,法院明确AI生成的虚拟形象受著作权法保护,侵权者需赔偿6万元。这些判例为AI侵权纠纷提供了司法参照。

奇安信集团行业安全研究中心主任裴智勇表示:“当前AI内容管理规定与实际执行之间仍存在落差,主要原因是缺乏有效的技术验证手段,导致仍有‘漏网之鱼’。”他表示,目前平台主要依赖末端检测,如人工观察或AI特征识别,这种方式算力消耗大、效率也不高。

裴智勇认为,可借鉴网络安全领域的经验,建立视频内容溯源与可信认证体系,以低成本验证AI生成内容。“例如采用数字签名技术,要求视频生产软件自动在文件中嵌入标识来源和作者的数字签名与标签。”他解释,这类标签对用

户无感,但平台可借此区分人工与AI内容。同时,平台应对未带标签或标签不合规的视频实施限流,通过技术实现客观标识。

据了解,许多大模型厂商已经开始对AI生成合成内容添加标识。如DeepSeek发布公告称,已在平台内对AI生成合成内容添加标识,并提醒用户相关内容由AI生成。用户不得恶意删除、篡改、伪造、隐匿此类生成内容标识,不得利用AI制作、传播虚假信息、侵权信息以及从事任何违法违规活动。

在中国社会科学院大学互联网法治研究中心主任刘晓春看来,除技术手段外,还应建立常态化的动态评估与风险分级机制。她建议,监管部门可按应用场景风险等级实施差异化监管,如将新闻资讯、商业推广列为较高风险领域,执行更严标准;创意娱乐等列为一般风险。

(综合《人民日报海外版》《北京晚报》《济南晚报》等)



为AI生成内容亮明身份是底线更是责任

◎ 张峻榛

AI生成内容隐匿身份的背后,是技术便利性与监管滞后性之间的博弈。当一部手机、一个软件就能快速生成以假乱真的图文视频时,部分内容创作者选择性地“遗忘”了标识义务。有的或许是出于省事,有的则可能怀有更复杂的动机——无论是商业推广中的“软性植入”,还是刻意营造的“真实分享”,隐匿AI身份都能让内容更具“亲和力”与“说服力”。这种刻意的模糊,正在悄然侵蚀着信息环境的真实性基础。

必须认识到,为AI生成内容标识身份,绝非多此一举的形式主义,而是维护网络信息生态健康的底线要求。在传统媒体时代,“本报讯”“本台记者”等信源标识是内容可信度的基础;在AI时代,明确标识生成方式同样应当成为数字内容传播的基本规范。这不仅关乎受众的知情权与选择权,更关系到整个社会的信息安全与信任体系。当AI内容可以与真人创作毫无二致时,缺乏明确标识就如同允许虚拟商品以真品名义流通,长此以往,必将损害公众对数字内容的整体信任。

平台作为内容传播的重要渠道,在AI内容标识方面承担着不可推卸的主体责任。当前部分平台主要依赖创作者主动声明的做法,显然不足以应对复杂的内容生态。面对海量上传内容,平台需要投入更多技术力量,开发更高效的AI内容识别系统,同时建立严格的审核与惩戒机制。技术瓶颈固然存在,但不能成为放任的借口。正如网络安全建设走过的路程,从无序到有序需要持续的技术投入与制度创新。

从更广阔的视角看,AI内容标识只是构建健康数字生态的第一步。随着技术发展,AI生成内容将更加深入地融入日常生活。建立健全从生成、传播到识别的全链条管理机制,需要监管部门、技术企业、内容平台和广大用户的共同参与。监管部门需不断完善细则,技术企业应积极研发更可靠的标识与识别技术,内容平台要切实履行审核责任,而用户也需提升数字素养,对未标识的可疑内容保持警惕。

人工智能技术开辟了创意表达的新天地,但真实永远是信任的基石。让每一幅AI生成的画面、每一段AI创作的文字都拥有坦荡的“数字身份”,这不仅是对法律的遵守,更是对每一位信息接收者的尊重,对我们共同构建的数字未来的负责。在技术狂飙突进的时代,为AI内容建立清晰的标识体系,就是在为数字文明铺设坚实的信任轨道。

声音

现在刷视频得自带“AI检测仪”:明星突然卖杂粮?专家话里插广告?先打个问号准没错。

——网友@一块抹布

鉴于AI生成合成技术可大规模产出与人类创作高度相似的内容,多数国家通过立法强制性要求企业履行内容标识义务,通过技术溯源以及显式和隐式标识,确保公众能快速识别AI生成合成内容的来源,抑制虚假信息的泛滥。

——浙江大学教授 王春晖

AI可以模拟现实,但无法替代真实的思考与判断;技术可以模糊边界,但不能瓦解信任的基石。唯有在创新与规范间找到平衡,才能避免被AI的莫比乌斯环困住,让技术真正服务于人类。

——南方网、粤学习评论员 曹晓静