

人工智能生成合成内容标识新规落地一个多月 未亮明身份的AI内容还在蒙人

9月1日,《人工智能生成合成内容标识办法》正式实施,AI生成内容进入“持证上岗”时代。按照规定,人工智能生成合成内容应有显式标识和隐式标识。然而,新规落地已有50天,记者调查发现,多家平台的短视频、图文笔记、直播间仍有许多AI内容未亮明身份。

短视频

疑似AI生成内容多 21条卡通水獭视频均无“身份”

新规落地后,短视频平台成为受影响最大的领域。10月1日至10月18日,记者在抖音、快手、小红书、哔哩哔哩、微博、视频号的推荐页面,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。另有约三成的内容虽无作者声明,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,小红书的一名博主发布了一条短视频。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。

记者统计发现,自9月23日起,这名博主已发布21条同类风格视频,均无AI“身份标识”。经记者联系核实,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,不少喜欢它的网友只能借助AI“云养水獭”。在这个平台

上,“#水獭”话题拥有超过8.1亿的浏览量。10月13日,记者在这个话题下随机浏览前20条内容,发现其中12条明显为AI制作。但是,在这12条中,9条没有任何标识,仅有3条为作者主动声明。

10月16日,为检验平台能否识别并标识AI生成内容,记者用“即梦AI”生成了一段视频,并将有AI水印和无AI水印的两个版本,分别发布在抖音、快手、小红书和视频号上,并均未主动标注AI创作。测试结果显示,抖音对两个版本均准确打上了“含AI生成内容”的标识,但快手、小红书和视频号,对两个版本的视频均未作出任何标识。

“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,识别难度将大幅增加。

图文笔记

图片文案全是AI生成 编纂育儿经验引导在线报班

“儿子三年级,全班第一,每晚的真实强度。”10月18日,博主“x x 爸爸带娃记”发布的一条育儿笔记出现在小红书的首页推荐上。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,之后话锋一转,推荐起某线上精读课程。

然而,翻看这名博主当天发布的其他8条笔记,令人困惑的情况出现了:他“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问题所在。记者仔细查看发现,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。

无标识的AI笔记可不少。在另一名博主“x x 妈妈有两娃”的主

页中,同样使用了大量相同的AI配图,画面中的母女在国贸餐厅就餐。这名博主在两天内连续发布的4条笔记中,女儿的年级在二、三、四年级之间不断变化,标题也与其他类似笔记高度雷同:“女儿三年级,全班第一,每晚的真实强度”“10年托举,女儿过上了普通人里最好的生活”等。同样,这些笔记也在文末推荐在线教育课程。

凭借精致的图片和诱人的话术,这名博主的笔记收获了不少互动,其中不乏“求精读课,非常感谢”“求推荐精读试听课”等积极回应。然而,这些笔记均未作任何内容标识,也有细心的读者质疑“怎么又是三年级呢”。

陈先生解释说,这是教育行业新出现的“AI全包广告”投放模式,即先用某款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。”

直播带货

美化图片货不对板 虚构商品效果顾客难辨真伪

如今,在直播间里,观众所见未必为实,AI技术正被频繁用于商品图片的美化甚至虚构,令普通消费者难辨真伪。

“树型品种可挑可选,一树一拍,今天拍今天发。”10月17日,抖音一个果树直播间里,主播正热情推销各品种龙眼树苗。商品链接中的图片显示,两株不足30厘米高的龙眼小树苗栽在约10厘米的花盆里,其中一株树苗树叶仅有十余片却结了近50颗果子。

“链接图是真的龙眼树苗吗?能种出这样的效果吗?”面对记者询问,主播信誓旦旦:“可以的!你看那树枝,每年都是高产、量产的,当年就能结果。”对此,北京农学院一名植物科学技术专家表示,龙眼树的小树苗不可能结出这么多果实,图片“高度疑似AI制作”。

这类夸大效果的AI图片,在毛绒玩具网上销售中也频频出现。“堪称诈骗!商品图竟然是AI做的!”网友“芝芝”近日花费60多元在淘宝购买了两只毛绒玩具猫挂件,收货

后发现实物眼歪嘴斜,与图片上毛发蓬松可爱的形象相差甚远。“完全用假图误导,货不对板。”

记者调查发现,鲜花种植类商品也是AI图片的“重灾区”。由于这类商品通常以种子或秧苗形式销售,消费者只能依靠商品图判断品质,而不少图片明显由AI制作,与现实存在显著差异。以某商家热销的垂丝海棠为例,这款商品获得近600人收藏,但其图片光影生硬,充满“塑料感”,花瓣细节与真实的垂丝海棠并不一致,素材明显取材于其他花卉品种。

此外,一些完全虚构的“AI植物”也在电商平台公开售卖。“冰糖葫芦枣树”就是典型例子。在拼多多上,一款收藏量达3774、周销量超400单的商品展示图中,一人高的果树上竟结满了巴掌大、形似葫芦的冬枣。对此,北京农学院的那位植物专家指出,真正的葫芦枣个头和普通枣差不多,只是中间有个“腰”,“图片上的‘枣树’其实是苹果树,苹果树上结枣儿,肯定不对。”



建议

提升验证手段 杜绝“漏网之鱼”

近日,北京市市场监管部门查处了首例利用AI技术进行虚假广告宣传的案件。这家公司通过AI技术剪辑知名主持人视频,加入自行设计的口播内容,在自有网络视频账号上以短视频等形式发布普通食品“深海多烯鱼油”广告,宣称其具有“解决头晕头痛、手麻脚麻”等医疗功效。这一行为违反了广告法相关规定,目前已受到行政处罚。

此类案件的出现,表明对AI生成内容的治理步伐正在紧紧跟上。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差距,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。

裴智勇建议,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。

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。

同时,平台自身也应建立AI内容识别效果评估体系,定期检验其拦截与核验能力的准确率,并将测评结果纳入平台安全评估。

据 北京晚报