## “去水印”绕过监管 “反标识”生意红火

# AI生成内容“持证上岗”为何难落地

针对AI生成内容使用乱象,2025年9月1日《人工智能生成合成内容标识办法》施行,明确AI生成内容必须添加标识,标志着我国AI生成内容迈入“持证上岗”的规范化时代。

如今新规落地已过百日,尽管各大平台陆续上线标识功能并配套相应管理措施,但记者调查发现,短视频、图文笔记、直播间,仍有许多AI内容未亮明身份,甚至一些原本显性的违规行为逐渐转向暗箱操作,深度伪造技术与黑灰产结合愈发紧密……标识制度在执行层面仍面临重重挑战。


记者从社交平台及电商平台搜索发现的AI去水印服务及工具。 制图 马莎

### 标识落地显成效 行业治理迈新阶

家住甘肃省兰州市的王浩(化名)去年8月在社交平台发布了一条“喜提迈巴赫”的AI合成视频后,多年未联系的“老同学”打来电话祝贺并向其张口借钱。王浩坦言,这条视频如果晚发一个月,就不会闹出这样的笑话。“没有AI标识,很多人当真了。”

“作品由AI生成”“内容存在AI成分注意甄别”……自去年9月《人工智能生成合成内容标识办法》正式施行以来,AI生成内容全面进入“持证上岗”的时代。一方面,针对可能导致公众混淆或者误认的内容,要求添加显式标识;另一方面,要求服务提供者在生成合成内容的文件元数据中添加隐式标识,为内容溯源与责任认定提供技术保障。

中国互联网络信息中心发布的《生成式人工智能应用发展报告(2025)》显示,截至2025年6月,我国生成式人工智能用户规模达5.15亿人,较2024年12月增长2.66亿人,用户规模半年翻番。

记者梳理发现,各大网络平台均已构建起各具特色的AI标识实施机制。抖音、今日头条、快手等内容平台在发布界面增设“AI生成内容声明”选项,用户勾选后将在作品标题下方显示统一标识。喜马拉雅等音频平台也采用了“片头提示+文字标注”等形式,对AI合成语音内容进行明确标识。

西部一所高校的AI治理团队2025年四季度做的一项抽样调研显示,AI标识政策落地后,用户对未知来源内容的“质疑意识”提升了近40%。此外,由于隐式标识能快速锁定内容的生成工具和传播节点,该团队参与的一起跨境AI虚假新闻溯源案例中,追责周期从过去的平均72小时缩短至12小时。

甘肃慧联信息科技发展有限责任公司总经理王雪莲表示,AI标识政策的实施,破解了AI内容“难识别、难追溯”的行业痛点,在政策引导下,从平台合规到用户认知,从技术研发到行业自律都呈现出向好的积极态势。

### 黑灰产牟利营 “反标识”乱象升级

“但AI伪造仍未根除,技术更高明,手段更隐蔽,标识治理面临‘道高一尺魔高一丈’的现实挑战。”王雪莲说。

“AI伪造烂水果骗取‘仅退款’薅羊毛”“一位明星穿梭8个直播间来回带货”……业内人士认为,AI伪造已从“一眼假”逐步实现高拟真段位升级,且背后已滋生分工明确、收益可观的完整黑灰产链条。

记者在各大电商平台和社交媒体上搜索“AI去水印”“无痕除标识”等关键词发现,从几十元的基础工具到上千元的定制服务,“明码标价”的规避标识“生意”形成黑灰产,各类绕过平台检测的“反标识”技术和服务“隐身术”持续迭代升级。

在某社交平台上,一则“AI生图变现指南”的评论区里,一条分享“AI去水印神器”的评论获得高赞,点进该用户主页,就有明码标价9.9元、19.9元等不同价位的去水印工具。记者私信该用户询问商品购买方法时,对方表示要换平台加好友下单,且提醒称“产品属性特殊,发货后不支持退款,若不想使用需自行处理或赠予朋友。”

记者在某电商平台搜索相关产品,一款“万物皆可去水印”的商品卖家介绍,这款工具能无痕去除98%的AI生成痕迹,尤其是AI视频的处理。

除不法商家低成本牟利,部分违规者还利用不同平台标识规则差异,进行跨平台“零成本”规避监管。

“由于各平台的AI识别技术水平和标识要求不同,在A平台被要求标注的内容,通过格式转换后在B平台发布就很有可能绕过检测。”甘肃省计算中心副主任沈玉琳说。

记者用一款AI软件生成一张矿泉水瓶损坏的图片,保存至手机相册后裁掉AI标识,再将其发布到社交平台时,并未被检测出AI元素。

沈玉琳认为,“反标识”技术正从单一手段向“多元技术路线与商业模式并行”的完整产业链形态升级,催生元数据“深度清零”、格式多轮转码、标识“露而不显”等“升级版”AI仿造乱象。

甘肃政法大学民商经济法学院副教授盛玉华介绍,AI标识政策落地一百多天以来,主流AI服务已普遍添加显式标识,溯源能力增强,但处罚标准模糊、技术检测不足仍制约其成效进一步释放。

“一方面,《人工智能生成合成内容标识办法》第十三条仅规定‘依据现有法规处理’,缺乏针对未标识、恶意篡改标识等行为的具体罚则;另一方面,部分平台缺乏隐式标识核验工具,难以有效识别违规内容。”盛玉华说。

### 多维协同筑防线 破解AI治理难题

受访专家及业内人士呼吁,AI治理需在“标识可见”的基础上,进一步筑牢“可识别、可追溯、可问责”的纵深防线,从技术创新、责任划分、协同治理等多维度发力,应对技术演进带来的监管挑战。

——完善“不可篡改、全程可溯”技术硬屏障。北京师范大学中国教育与社会发展研究院助理研究员蒋艳双认为,目前大多数平台监管技术仍存薄弱环节,未来应加快推进AI标识技术的标准化,细化不同平台、不同类型内容的标识技术规范,避免因技术差异导致监管漏洞。

此外,还需进一步强化隐式标识的抗篡改技术升级,确保标识信息在文件格式转换、二次编辑等操作后仍能有效识别,整合不同AI工具的生成痕迹,提升对各类规避手段的识别精准度。

——厘清“生成、传播、使用”权责边界。盛玉华等人建议,应进一步细化《人工智能生成合成内容标识办法》中的主体责任,明确生成合成内容服务提供者、传播平台、分发平台等不同主体的具体义务。

同时,强化对黑灰产的处罚力度,对恶意删除标识、伪造标识、提供规避标识工具、批量篡改标识等行为从重处罚,可探索建立违规主体“黑名单”制度,将多次实施规避标识行为的个人和企业纳入黑名单,从利益根源上遏制黑灰产。

——构建“政府、公众、平台”治理闭环。兰州大学信息科学与工程学院教授杨裔建议,需进一步完善举报与监督机制,鼓励公众参与AI内容治理,建立健全跨平台投诉举报联动机制,实现违规账号信息共享,避免违规主体“换平台重来”。

此外,还需持续强化公众AI素养科普,普及AI内容标识识别方法、常见造假手段及维权路径,引导公众养成“看标识、辨真伪”的AI使用习惯。

“多管齐下、层层设防,各方形成合力织就一张人人参与的治理大网,引导AI技术告别野蛮生长的浮躁,真正成为赋能千行百业的好工具、好帮手。”蒋艳双说。

(据《经济参考报》)