AI“吸粉”乱象背后——

如何引导技术向善

外国人熟练演唱中文歌曲、银发奶奶传授养生秘诀、氛围感女孩分享穿搭美学……社交媒体上，一些AI生成的视频关注度颇高，“细节满满”让很多网友信以为真。有媒体调查发现，不少社交账号利用AI技术造假、博眼球快速“吸粉”起号，进而变现。

近期，多个平台对“AI起号”现象开展专项治理行动，清理违规内容并封禁部分账号。

现象 AI造假起号成新套路

广西市民王夏经常刷到这样的视频：“帅哥”“美女”高频率更新日常生活，并在评论区与网友频繁互动。由于熟悉AI软件，王夏很快发现这些视频的主角是AI数字人，“但视频没有任何AI生成提示字样，好几个账号显示来自同一家MCN公司”。

当前，生成式人工智能技术日益普及，不少内容创作者通过AI创作出更具想象力的作品。但也有一些人利用AI造假，无底线博眼球，作为起号变现捷径。

最近，跳水运动员全红婵因为“帮妈妈推销土鸡蛋”登上热搜。人气很高的乒乓球运动员孙颖莎、王楚钦也发声助阵，“孙颖莎”直言“我为婵宝家农家土鸡蛋代言”，“王楚钦”则是“收到婵妹私信”前来帮忙。

但媒体采访调查发现，“奥运冠军带货”子虚乌有，他们的声音是用AI工具合成的。在他们之前，知名演员靳东、刘晓庆，企业家雷军等公众人物，都曾遭遇相似问题。除了让被仿冒者口碑受损，造成网友、粉丝经济损失，技术造假如果得不到有效遏制，还会动摇社会的信任基础。有不少人担心：如今眼见、耳听都不一定为实，还能相信什么？

以此次“运动员代言土鸡蛋”事件为例，违法者只需一款AI软件，再找一段清晰的语音样本，就能快速合成出自己设定好的内容。几位经常出现在公众场合的知名运动员的语音几乎随手就能搜到。侵权账号也属于难以辨别主体、数量庞大的“账号矩阵”之一，被发现侵权后立即清空，将内容迁移至别的账号上。维权者要想在信息浩若烟海的互联网上锁定数字账户背后的侵权者、固定侵权证据，难度在不断提高。

措施 阻断AI造假起号利益链

当前，主流网络平台普遍升级AI内容识别系统，要求对AI生成作品添加标注，但仍有一些人采用各种手段绕过审核。

“平台甄别AI内容主要依靠特征性技术痕迹，一些起号者通过多种手段削弱特征痕迹，逃避平台标注。”重庆理工大学计算机科学与工程学院教授李彦说，如通过微信转压缩等方式改变视频文件代码结构，再上传至平台，就较难判定为AI生成内容。

新华社记者用一款AI软件生成一张人像图片，裁掉AI生图软件水印后发布在社交平台上。发布后，平台系统并未自动识别、提示添加标注。相当一部分网友辨别不出这是AI生成图，纷纷点赞、发送私信。

“零基础3天涨粉过万”“轻松月入过万”……近来，社交平台上一批打着“AI变现”旗号的教程正以极具诱惑力的宣传语吸引着不少人的目光。但据媒体报道，其实际内容多为网上随处可寻的基础AI工具使用方法，按教程操作不仅难达宣传效果，还可能因违规被平台限制流量，且商家常缺乏有效售后，甚至“售后即失联”。

在一些社交平台上，有不少网友分享了被这类课程“坑”过的经历——有人付了钱只拿到些零散资料，自己对着学半天也摸不着门道；有人尝试操作却屡屡碰壁，于是调侃“现在防AI如防狼”；更有人想找商家咨询，对方却早已没了踪影，只能自认倒霉。

AI培训课程虚假宣传乱象丛生，已成为侵害消费者权益的重灾区，众多学员陷入“交钱易、维权难”的困境。在社交平台，大量网友吐槽被AI培训课程“画饼”欺骗，课程宣传时承诺“AI起号轻松变现”“月入过万不是梦”，甚至晒出虚假高收益案例，同时，这些课程常以“低价体验课”引流，后续层层诱导付费，用“限时优惠”制造紧迫感迫使学员冲动购买高价课，不少学员被诱导二次消费，陷入“不续费没服务，续费更人坑”的连环套。而当学员发现课程与宣传不符要求退费时，商家会百般推脱，甚至直接失联跑路，还会解散学习群、删除联系方式，切断学员维权沟通渠道。

重庆公孝律师事务所执行主任徐斌表示，AI造假起号灰色产业链是技术异化的结果，可能造成低俗猎奇信息泛滥，加速虚假新闻、谣言传播，扰乱网络空间秩序，亟待加强清理整治。

今年4月以来，中央网信办部署开展“清朗·整治AI技术滥用”专项行动，聚焦AI换脸拟声侵犯公众权益、AI内容标识缺失误导公众等AI技术滥用乱象开展重点整治。第一阶段累计处置违规小程序、应用程序、智能体等AI产品3500余款，清理违法违规信息96万余条，处置账号3700余个。

前不久，多家互联网平台发布专项治理公告，重点整治AI批量造假、AI起号引流带货、转让销售AI虚拟账号等违规行为。

北京邮电大学互联网治理与法律研究中心执行主任谢永江建议，技术提供者强化技术的安全性和合规性研发，提供技术支持和解决方案；平台提高对深度伪造内容的检测精度，进一步完善相关规则，明确对AI造假起号等违规行为的界定和处罚标准等。

反思 技术向善，任重道远

既然是违法乃至犯罪行为，不该也不能成为无解难题。随着技术发展，此类违法行为还会增加并不断变化，必须从技术和制度层面上筑牢防火墙，才能有效惩治违法行为，维护互联网安全。

以技术治理技术是一条可取之道。如今一些社交媒体平台已经开始试点在AI生成内容上标注提醒，今年9月1日起将实施的《人工智能生成合成内容标识办法》则明确要求“生成合成内容添加显式标识”“提醒用户主动声明发布内容中是否包含生成合成内容”。同时，平台也应利用AI技术和大数据信息，及时分析比对违法内容，主动拦截、及时下架并保存数据作为证据。

社交平台对这类“AI起号教程”的发布、售卖应承担怎样的审核和监管责任？中国人民大学法学院副教授黄尹旭表示，社交平台应承担的审核和监管责任包括：

第一，内容合规审查责任。平台需依据《互联网信息服务管理办法》《网络信息内容生态治理规定》《人工智能生成合成内容标识办法》等规范，对AI起号教程进行前置审核，确保其不包含虚假宣传、账号买卖，以及伪造身份、批量起号等违规操作指导。

第二，标注与真实性核查责任。若教程涉及AI生成内容，平台应要求发布者明确标注“AI生成”，并核查其宣传真实性，防止夸大或欺诈性营销。若平台未核验AI起号教程的AI生成标注，法院可推定其未尽合理注意义务；若平台放任未标注AI起号教程传播的，可认定其默许风险发生。

第三，动态监测与处置责任。建立技术监测机制，提升深度伪造内容检测精度，完善AI造假行为界定标准和处罚规则，提高虚假内容识别效率，从源头上斩断传播链条。对已发布的教程进行持续筛查，发现违规后采取下架、限流、封号等措施，并向监管部门报告。

加强AI技术治理，绝不是限制技术发展。技术本无错，为牟利踩踏法律红线、践踏他人权益的是藏在技术背后的人。加强治理利用AI技术违法犯罪，不仅是保护个体权益，也是为AI健康发展铺平道路。将技术创新纳入法律规范之中，是未来发展的必由之路，也考验我们的治理智慧。

（综合新华社、《解放日报》、法治网等）

AI造假“起号”博流量，如何封堵技术漏洞？

◎ 刘霁月

近日，社交媒体上AI生成的虚拟网红、明星仿声视频引发广泛关注。从“奥运冠军带货”到“银发博主授课”，一批借助人工智能技术造假吸粉、违规变现的账号浮出水面，平台专项治理随之启动。这既折射出生成式人工智能技术的广泛应用，也暴露出技术滥用对社会信任和网络秩序的严峻挑战。

AI造假之所以能形成灰色产业链，源于其低门槛与高流量的诱惑。通过合成人脸、模拟人声，侵权者可以快速打造“爆款人设”，短期内聚集大量粉丝进而牟利。而深度伪造内容难以辨识、侵权主体隐蔽性强、证据固定难等特点，更使得此类行为屡禁不止。尤其值得警惕的是，部分所谓“AI变现课程”以夸大宣传诱骗用户，实质是以“技术培训”为名行“收割韭菜”之实，严重损害消费者权益。

技术本身并无善恶，但如何使用技术却关乎伦理与法律底线。AI造假内容不仅侵犯公民肖像权、名誉权，还可能衍生诈骗、虚假广告等违法犯罪行为，侵蚀社会信任基石。如果“AI造假”成为公众共识，甚至催生“怀疑一切”的认知危机，将严重破坏数字时代的信任生态。

治理AI滥用乱象，需打出“技术+制度+平台”的组合拳。技术上应强化深度伪造检测与溯源能力，通过数字水印、内容标识等技术手段提高AI生成内容的可辨识性；制度上须加快完善人工智能应用法律法规，明确技术开发者、内容生产者、平台运营者及用户各方责任；平台上则应建立从内容审核到违规处置的全周期治理机制，切断违规AI账号的利益链。自今年4月“清朗·整治AI技术滥用”专项行动开展以来，已有数千个违规账号及应用程序被处置，显示出治理行动的初步成效。

我们既要看到人工智能赋能创意生产、提升内容质量的积极作用，也要警惕技术滥用带来的风险。推动人工智能技术向善发展，不仅需要有效的监管和法治保障，更需要行业自律、公众监督和教育引导，形成多方协同的治理格局。

归根结底，技术发展的终极目标是为了人的福祉。唯有将人工智能纳入规范发展轨道，使技术创新行驶在伦理与法律的轨道上，才能真正实现科技造福人类的美好愿景。人工智能的未来，不仅取决于技术突破的高度，更取决于我们治理智慧的深度。

新闻多一点

《人工智能生成合成内容标识办法》的实施，能改变当下AI出现滥用的情况吗？

几天后的9月1日，网信办等四部门印发的《人工智能生成合成内容标识办法》（以下简称《标识办法》）将开始施行，它明确规定服务提供者，应当对文本、音频、图片、视频等生成合成内容，在适当位置添加显式标识，也就是说今后所有AI作品必须亮明身份，标注是AI生成。它的实施，能改变当下AI出现滥用的情况吗？

在某短视频平台上，一条生成式视频的左下角“标注着”作者声明：内容由AI生成，而这样的标注义务，在即将于9月1日实施的《人工智能生成合成内容标识办法》中，得以进一步明确，在人工智能生成合成内容的全生命周期，都要履行标识义务。

北京抖音信息服务有限公司副总编辑全森表示，这个标识办法要求AI技术服务提供者，对文件（元数据）内容进行隐式标识。这样一个核心的链路解决了最大的卡点，这个卡点就是技术服务平台与内容传播平台之间，数据难以互通的一个卡点。解决了这个卡点之后，我们预计很大程度上解决了这个AI难以识别的问题。

马上就要实施的这个《标识办法》，从人工智能生成合成内容的生产、传播和使用等环节入手，对标识制度进行了系统性的细化和规范。随着深度合成技术日益逼真，逃避审查的手段不断升级，人工识别的难度越来越高，相应的法规和机制，也需要不断应对挑战。

推进生成式人工智能标识制度的建立，已经成为世界各国的共识。在专家看来，中国在该领域的立法起步相对较早，2023年施行的《互联网信息服务深度合成管理规定》，就在世界范围内，在相关法规中首次明确提出生成式人工智能标识义务。

在专家看来，即将实施的《标识办法》，还只是一个规范性文件，在整个立法体系中位阶较低。如果站在更高的维度观察人工智能立法，安全、隐私、公平等都是必须直面的核心问题。

（据央视新闻客户端）

本版制图 刘英卉