AI生成合成内容须添加标识

《人工智能生成合成内容标识办法》发布，强化全流程管理，引导技术向善

近日，国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局联合发布《人工智能生成合成内容标识办法》。

专家认为，办法是我国推进人工智能领域安全治理、促进产业规范健康发展、引导技术向善的重要举措，标志着我国在生成式人工智能领域迈出了构建安全可信生态的关键一步。

AI催生新型安全风险 四部门出台“标识办法”

《人工智能生成合成内容标识办法》聚焦人工智能“生成合成内容标识”关键点，通过标识提醒用户辨别虚假信息，明确相关服务主体的标识责任义务，规范内容制作、传播各环节标识行为，将于2025年9月1日起施行。

近年来，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、音频、视频等信息提供了便利工具，海量信息得以快速生成合成并在网络平台传播，在促进经济社会发展的同时，也产生了生成合成技术滥用、虚假信息传播扩散加剧等问题，引发社会关注。

“人工智能生成合成内容日益逼真，催生虚假消息传播、身份信息冒充、恶意内容生成等新型安全风险，并削弱着公众对网络传播内容的信任根基。”中国工程院院士、浙江大学教授陈纯认为，面对人工智能安全治理这一世界性难题，四部门适时出台《人工智能生成合成内容标识办法》，配套强制性国家标准《网络安全技术　人工智能生成合成内容标识方法》也同时发布，成为保障人工智能时代网络生态安全有序的关键手段。

国家网信办有关负责人表示，办法重点解决“哪些是生成的”“谁生成的”“从哪里生成的”等问题，推动由生成到传播各环节的全流程安全管理，力争打造可信赖的人工智能技术。

办法提出，人工智能生成合成内容是指利用人工智能技术生成、合成的文本、图片、音频、视频、虚拟场景等信息。人工智能生成合成内容标识包括显式标识和隐式标识。

针对服务提供者，办法明确应当对文本、音频、图片、视频、虚拟场景等生成合成内容添加显式标识，在提供生成合成内容下载、复制、导出等功能时，应当确保文件中含有满足要求的显式标识；应当在生成合成内容的文件元数据中添加隐式标识。

针对互联网应用程序分发平台，办法提出在应用程序上架或者上线审核时，应当要求互联网应用程序服务提供者说明是否提供人工智能生成合成服务，并核验其生成合成内容标识相关材料。

针对用户，办法要求使用网络信息内容传播服务发布生成合成内容的，应当主动声明并使用服务提供者提供的标识功能进行标识。

此外，任何组织和个人不得恶意删除、篡改、伪造、隐匿办法规定的生成合成内容标识，不得为他人实施上述恶意行为提供工具或者服务，不得通过不正当标识手段损害他人合法权益。

清晰的内容标识 有助于提升信息透明度

“此前出台的《互联网信息服务算法推荐管理规定》《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》提出了标识有关要求，办法进一步细化了标识的具体实施规范。”北京航空航天大学法学院副教授赵精武认为，办法明确了生成合成内容制作传播各主体的责任义务，用户能够清晰识别人工智能生成的内容，减轻人工智能生成合成技术滥用危害，防范利用生成合成内容传播虚假信息、实施电信诈骗等风险行为。

一方面，清晰的内容标识有助于提升信息透明度，增强用户的知情权和选择权，培养公众对人工智能技术的理性认知。另一方面，对于服务提供者及传播平台而言，内容标识制度既是责任更是机遇，在鼓励企业追求技术能力提升的同时，也将提升对产品社会影响的关注度，推动整个行业向更加规范、健康的方向发展。

良法善治，重在实施。专家认为，办法的落地需要各方协同配合，推动标识工作行稳致远。

公安部第三研究所副所长金波说，伴随标识管理与算法备案、安全评估等机制逐步实现有机衔接，生成合成内容标识合规或将成为相关部门开展人工智能监督检查、专项行动的重点关注领域。在此进程中，如何平衡好发展与安全、创新与责任，提升执法的专业化、精细化、智慧化水平，从而培育出安全、开放、公平的人工智能产业生态环境，还需要深入探究。

“此外，‘人的因素’应融入人工智能标识管理的全过程。”金波说，要着重提升公众对信息内容真实性、来源可追溯性的评估能力，积极培育公众的人工智能素养，确保人工智能技术成果普惠共享。

综合新华社等

相关新闻

刘德华打来视频？是AI换脸骗你呢

1980元就能换一张脸，有商家利用“名人脸”直播带货获取流量

刘志坤 岳致呈 实习生 王心铭 济南报道

名人“被代言” 公众遭欺骗

日前，全国人大代表、小米集团创始人雷军在全国两会期间呼吁立法严管“AI换脸拟声”滥用。“去年‘十一’，有网友说，放了7天假，被我整整骂了8天。”雷军说。雷军的声音、形象被不少网友利用AI技术伪造，用于制作恶搞视频甚至带货。雷军曾在个人社交媒体上公开表示：这种“雷军语音包”让自己非常困扰，带来了不好的影响和感受。

不止是名人，普通民众也深受AI换脸技术的侵害。去年12月，江西65岁老人通过抖音网恋AI“靳东”，欲贷款200万资助对方拍戏，被民警演示AI换脸技术后才意识到自己上当。

1980元一张AI脸 识别越来越难

近日，记者调查发现，网络上存在一种“AI换脸”的定制服务。这种AI脸可用于各大平台，视频通话和直播都能胜任。先在平台发帖宣传，再通过平台群聊发布商家个人联络方式，这项服务的交易最终在微信进行沟通，以银行卡转账的方式进行。

“想要的话就发一张正脸照片，五小时之内就能做好。”一家**刘德华给你打来了视频电话？别信，这或许是AI换脸在骗你。AI换脸技术诞生之初，人们将自己的脸“移植”到影视角色、明星身上进行娱乐。如今这一技术却沦为诈骗、侵权，尤其是直播带货骗局的“新型作案工具”。**


商家展示AI换脸效果

AI换脸公司的工作人员李立（化名）介绍，“AI换脸”定制服务以“先定金后尾款”的方式进行。一张AI脸的价格全款1980元，定金500元，已经有成品的名人AI脸则只需980元。支付尾款后卖家还会免费提供安装教学和变声器。据李立介绍，近期来“画皮”的人很多，公司已超负荷运行。“我们这里有七八台设备，每天都要做到凌晨。”

不同于以往的AI换脸，新技术已经实现声音与图像的实时同步克隆，使得伪造视频在视觉和听觉上都达到了近乎完美的程度，识别难度大大提升。

“已经跟直接录制人像视频没什么区别了，就是专业的技术人员都认不出来。”齐鲁壹点大数据中心专家王智直言，目前AI换脸技术已经相当成熟。“尤其是中老年人，他们的分辨能力比较差，而且跟不上AI技术的发展，这方面遇到的问题会更多。”

AI乱象还需 加强法律制约

过去人们往往通过挥手来大幅遮挡面部，观察面部图像闪现、抖动等异常来识别对方是否为AI合成的假脸。然而随着技术进步，这些传统手段渐渐不再有效。

“分辨的方式主要是观察嘴型跟声音。”王智指出，AI合成的人像嘴型和正常人仍然存在区别，声音往往似像非像。如果视频中的人物一直坐着不动，不进行转身，或者手部动作很少，那么该视频有很大概率是使用了AI技术。

“我们对AI的担忧，其实是技术发展进步过程中常有的现象。”山东大学哲学与社会发展学院硕士生导师王元超解释，AI技术发展本身就会带来一定的问题，但我们需要讨论的不是技术本身，而是社会该如何去引导它的发展。

“这个工具确实很好。”王智说，我们期待伦理和法律追上甚至引领技术的发展，打击技术滥用，更好地保护消费者权益，重塑数字时代的信任基石。

这些穿搭咋这么完美？原是“AI模展”

商家采用AI模特试穿服装展示，消费者小心踩雷

李苗 路董萌 济南报道

发光的头发、完美的身材、恰好合身的衣服……很多购物平台的模特都美得像一幅画。记者发现，不少商家在电商平台采用AI模特试穿服装展示商品。

“都是‘大波浪反光发型、程式化微笑’的AI形象。”消费者严女士表示，自己在购买毛衣时，发现好几家店铺使用的模特风格都特别相似，但仔细辨别就会发现这些模特似乎并非真人。

严女士透露，其中一家拥有110万粉丝的店铺，上千成交量的商品，但商品详情页使用的都是AI图，其他图片仅靠镜面或放大进行充数，缺少商品详情的平铺图、细节图等重要信息。“根本看不到衣服细节，并且商品所有评论都是好评，其中带图好评展示的也是AI生成图片。这明显不符合真实交易逻辑。”

事实上，严女士的遭遇并非个例。在社交平台上，还有不少网友表示，购买此类衣服后，出现付款后长期不发货、空包裹或者货不对版等问题。大家普遍认为，AI模特展示衣服增加了网购踩雷概率。

对此，齐鲁晚报·齐鲁壹点记者联系了该平台工作人员。对方称目前平台尚未针对AI模特图的使用制定明确规定，只能先行收集消费者反馈，等待后续整合改进。不过，若出现货不对版、空包裹等情况，消费者可在平台申请售后，权益能得到保障。

记者以“早秋女装”等为关键词在各购物平台进行检索，发现大量商家宣传图为AI模特图，部分模特手部存在模糊、畸形等问题。而且，多家店铺图片高度相似，可店家均未对AI模特图进行明显提示与标注，全靠消费者自行分辨。

“商家通过AI模特图不能真实展现商品材质、款式、尺寸等关键信息，仅用单一的虚拟图片来代替实物细节的展示，必然会涉及损害消费者的知情权，违背了经营者对消费者应当承担的商品或服务信息全面、真实、准确的义务，甚至有可能涉嫌虚假。”上海段和段（济南）律师事务所合伙人、律师魏方丹说。

济南大学商学院教授朱瑾分析，AI模特试穿展示属于技术驱动下的商业模式变革。然而，当下AI技术尚不成熟，在逼真度与现实场景还原上存在差距。部分商家急于应用该技术，平台又未明确管理边界，加之许多消费者对AI图不够敏感，极易产生困惑与误导。

专家指出，近日，国家网办等四部门联合发布了《人工智能生成合成内容标识办法》，店家应明确告知消费者模特图由人工智能生成，履行标识义务，让消费者自主判断，减少误解。“在双方都努力的情况之下，才能够使得互联网交易的环境变得更加健康可持续，这种信任的氛围会越来越好。”