DOI: 10.54254/3050-2160/2025.24043

综合治理背景下防范利用 AI 换脸技术实施诈骗行为研究

曹 钰[1*]，朱星瑶[1]，宋 萍[1]，马丽萍[1]

（1.南京理工大学知识产权学院，江苏省南京市，210094；*通讯作者，1812015865@qq.com）

摘 要：数字时代的浪潮奔涌向前，随着生成式人工智能技术的快速发展，AI 换脸技术在创造诸多便利的同时，也被不法分子用于诈骗活动，新型诈骗手段将不断涌现，这使得防范工作仍面临诸多挑战，因此，构建“制度-技术-主体-协同”的四位一体防范体系，强调治理主体的多元性和层级间的协同合作，提出系统性的解决方案是十分必要的。四位一体的体系，不只是冰冷的框架，更是万千守护者共同谱写的安全诗篇，只有通过持续不断的探索与实践，优化综合治理方案，才能在与 AI 换脸诈骗的博弈中占据主动，维护网络空间秩序与人民群众的合法权益。

关键词：AI 换脸；诈骗；综合治理；公共安全

引言

“水能载舟，亦能覆舟”——从张伯礼院士被 AI 换脸技术伪造宣传护肤品[1]，到专门针对年轻群体的“AI 培训”[2]，以及普通人因 AI 视频遭遇诈骗，AI 技术这一本应凝结科学技术的进步结晶，如今却被不法分子利用，沦为新型诈骗的工具。

华盛顿大学 Carl Bergstorm 教授曾指出：“当一项像 Deepfake 这种的新技术出现时，最危险的时刻是这项技术已经存世而不为公众所知，这是它最容易被加以利用的时候”[3]。2023 年 10 月，一诈骗团伙被提起公诉，因其假冒演员“靳东”获取受害人信任，再以投资、公益、恋爱等各种理由索要钱款，诈骗多名被害人共计人民币 30 余万元[4]；2024 年 2 月 4 日，香港警方披露首宗多人换脸 AI 诈骗案，一家总部在英国的跨国公司的香港分公司被 AI 换脸技术骗走 2 亿港元（折合约 1.8 亿人民币）[5]。由此可以看出，在 AI 换脸技术迅速发展的时代背景下，它的潜在风险与滥用问题日益显现，AI 换脸技术被用于诈骗、虚假信息传播等违法犯罪活动，已经严重威胁个人隐私与社会安全。

目前，学术界已经开始关注 AI 换脸诈骗的检测与防御。笔者查阅多篇文献，发现目前大部分文献多集中于技术研发、法律风险领域的研究，而关于在综合治理背景下，对利用 AI 换脸技术实施诈骗的行为如何进行规制，该领域仍然存在诸多空白。本研究在此基础上，分析 AI 换脸诈骗的技术原理、法律定性、现实问题及防御策略，结合综合治理模式，提出切实可行的解决方案，为防范不法分子利用 AI 换脸技术实施诈骗行为的政策制定和行业规范提供参考。

1 利用 AI 换脸技术实施诈骗行为的一般性分析

1.1 AI 换脸技术的技术原理

AI 换脸技术的核心是 DeepFake 技术。DeepFake 技术是基于深度学习的自编码器来实现换脸操作，它的换脸篡改方式是将源目标的人脸替换到目标人脸上去[6]。诈骗分子通过提取源视频或图像的面部特征，与训练好的模型进行匹配并替换，通过调整细节生成高度逼真的换脸视频或图像，最终达到以假乱真的效果，令人难以分辨。

当前，利用 AI 换脸技术实施的诈骗手法多样，包括冒充亲友视频通话骗取钱财、替换普通人脸进行敲诈勒索，以及伪装企业高管下达虚假指令，导致个人和企业遭受财产损失。

1.2 常见的诈骗人群

在通过线上网络平台“问卷星”开展的相关调研中，本次问卷共收集样本 174 份，受试人群年龄段丰富，包含儿童，青少年，青年，中年以及 65 岁以上老人，从调查中发现，94%的人都听说过 AI 换脸，82%的人都在社交 app 上刷到过疑似 AI 换脸的视频并且频率极高（见图 1）。

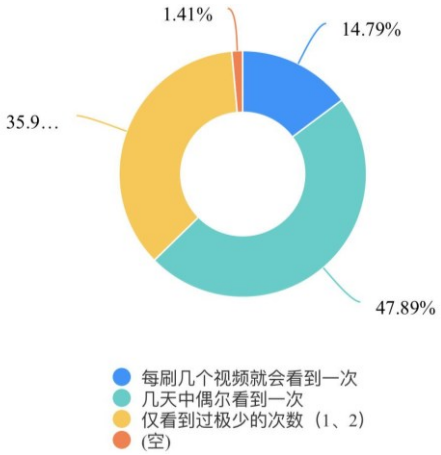

图 1

17%的人下载或使用过 AI 换脸照片/视频网站；不使用 AI 换脸的人中 65.28%是因为担心 AI 换脸采集其面部数据，导致隐私泄露（见图 2）。

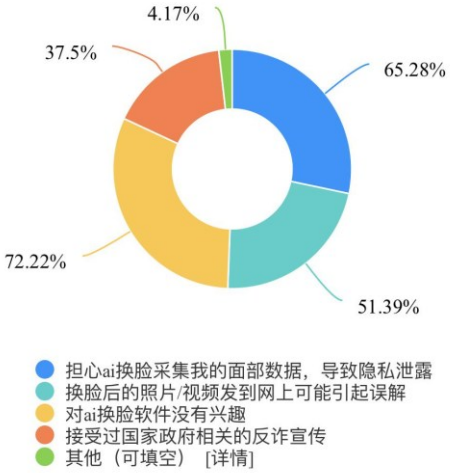

图 2

在 174 人中 85.6%听说过社会上利用 AI 换脸实施诈骗的相关新闻/案例；5.1%的人遭遇过不法分子利用 AI 换脸技术实施的诈骗行为；对于隐私的保护 83.33%的人选择不使用 AI 换脸软件；在一道四项选项均为 AI 换脸图片的多选辨识题中，仅有 16.7%的受试者正确选择了全部四个选项，其中 18-22 岁和 36-55 岁人群居多，18 岁以下的未成年群体和 55 岁以上的中老年人群体辨别能力较为薄弱，是易受 AI 换脸诈骗的重点防范人群[1]。

[1] 问卷调查，https://www.wjx.cn/vm/tc811g5.aspx# .

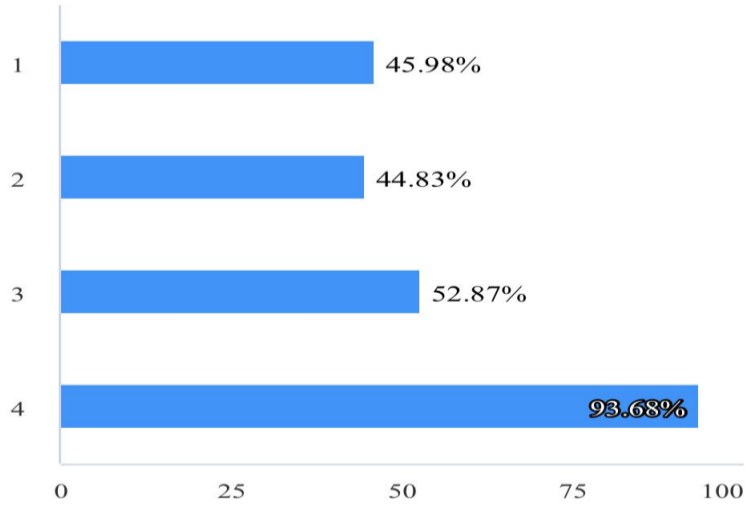

图 3

而从整体来看，当今社会人群遇到 AI 换脸和利用 AI 换脸实施诈骗的情况很多，但是防范意识和分辨能力还是较弱，对于自身的分辨能力认识不清，在实际中无法快速、正确的分辨，导致受骗几率大大增加。

2 不法分子利用 AI 换脸技术实施诈骗行为的法律定性

2.1 利用 AI 换脸技术诈骗行为是否构成一般的民事侵权

对 AI 换脸侵权行为的认定则需要进一步划分侵权行为与合法行为的界限，即讨论该使用 AI 换脸技术的行为是否满足构成侵权行为的四大要件：主观过错、加害行为、损害事实、加害行为与损害事实之间的因果关系。

首先，行为人应当具有主观过错。此处的过错并非强调行为人一定具备谋取非法财产的主观意图，而是强调行为人明知该行为会侵犯他人的合法权益而故意为之的主观心理状态。其次，行为人应当采取了相应的加害行为。由于 AI 换脸技术实质上是一种网络电子技术，其使用具有特殊性，在该情形下的加害行为应当指行为人在未经他人许可的情况下，利用他人的面部特征或相关受法律保护的影视媒体作品，合成照片、视频等文件的行为。根据法律关于肖像权和著作权内容的相关规定，此处并不要求行为人一定要将生成的相关媒体文件进行公开或展示。因为生成行为本身就已经违反了相关法律规定，若在此基础之上再增添公开的要求，其实是对法律原本保护权利的范围的缩限，是不恰当的。接着，行为人的行为应当产生了损害事实，并且行为人的加害行为与损害后果之间应当存在因果联系。若行为人经对方许可授权后，以合法的方式合成的图片遭到他人的不法利用，则造成的损害后果与行为人利用 AI 换脸技术合成图片的行为之间并不存在因果联系，这一行为就不可认定侵权，真正的侵权行为应当是他人的不法利用行为。

综上所述，若行为人采取的是合法的方式，经过他人授权，仅仅是为了娱乐或者其他合理需求而使用 AI 换脸技术进行的深度合成、公开传播行为不能认定为侵权行为。

2.2 利用 AI 换脸技术诈骗行为是否构成诈骗罪

在目前我国的现有的刑法中，对于利用 AI 换脸技术进行诈骗虽然没有专门的罪名，但利用 AI 换脸技术进行诈骗的行为，性质上可归类为计算机信息系统犯罪。根据具体情节、诈骗金额大小、造成的社会影响和受害者损失程度不同，可能构成现有的一些犯罪。

不法分子利用 AI 换脸技术实施诈骗，符合刑法上诈骗罪实施行为中的虚构事实、隐瞒真相，使得受害人对不法分子的身份产生误解，从而骗取钱财。若客观上侵犯了公私财产所有权，骗取财产数额较大的（根据有关的司法解释精神和司法实践经验，是指个人诈骗公私财物价值人民币 2000 元至 4000 元以上），即可构成诈骗罪。

当判断利用 AI 换脸技术实施诈骗是否构成诈骗罪时，可以将“利用 AI 换脸技术实施诈骗过程”看作是一种特殊的犯罪实施手段，其符合诈骗罪的核心，即“虚构事实、隐瞒真相”，从而以刑法中的诈骗罪对其进行规制。

2.3 利用 AI 换脸技术实施诈骗行为是否构成其他犯罪类型

根据 AI 换脸后具体实施的行为不同，如搜集他人个人信息，通过 AI 技术制作成“不雅视频”，精准发送给受害者，并以举报、散播、告知家属等方式要挟受害者“花钱消灾”的，可构成敲诈勒索罪；在网上收集他人提供的女明星或普通女性的生活照，通过 AI 技术“定制”淫秽视频和图片，甚至同时出售相关制作软件及使用教程的，又可构成传播淫秽物品牟利罪。

根据危害后果和情节严重程度不同，如利用 AI 换脸技术实施诈骗，即使诈骗金额较小，不构成刑法意义上的诈骗罪的，若情节特殊，造成社会影响巨大，严重扰乱社会秩序的，诸如编造虚假的险情、疫情、灾情、警情，在信息网络或者其他媒体上传播的，根据《中华人民共和国刑法》第 291 条之一第二款的规定，可构成编造、故意传播虚假信息罪。而不法分子利用 AI 换脸，在并没有得到授权的情况下，也相当于泄露了他人的面容信息，情节严重，给被害人造成物质损失或精神损害的，根据《中华人民共和国刑法》第 253 条之一规定也可构成侵犯公民个人信息罪。

若涉及到特殊主体，在司法实践过程中，有不法分子利用 AI 换脸技术，冒充国家公职人员发布新闻或参加有关线上会议，虽然没有造成有关的财产损失，却对政府的公众威信力造成损害，同时还有可能危害到政府正常的管理活动，潜在社会危害性极大，目前《刑法典》中相关犯罪仅有招摇撞骗罪，在有关法律条文上还有所缺失。

即使因情节不同构成罪名有所不同，利用 AI 换脸实施诈骗的核心即是利用虚假的面部特征带来的身份实施犯罪行为，所谓“虚构事实、隐瞒真相”。

3 目前防范利用 AI 换脸技术实施诈骗行为存在的问题

3.1 承担法律责任层面

2023 年公安部发布 AI 换脸诈骗案件同比增长 37%的数据，揭露了现行法律体系对此留存的三重制度性缺陷。法律定性模糊位居首位，《民法典》第 1019 条未明确深度伪造技术之定性，致使司法实践中“换脸行为”常被简单归入一般侵权[7]；随之，《刑法》第 266 条也难以体现虚拟空间的身份冒用行为的时代特殊性，24 年杭州互联网法院“AI 换脸冒充公司高管诈骗案”被告就仅以普通诈骗罪判处；除此之外，现行司法解释对“严重情节”的量化标准明显滞后技术发展，24 年香港最大 AI 诈骗案[8]中，犯罪集团伪造领导人形象，涉案金额达 2.3 亿港元，但因缺乏明确量刑指引，主犯仅获刑 13 年，与其社会危害性明显不符。

3.2 技术层面

现有的 AI 换脸检测技术容易出现误判情况，容易将正常的面部图像或视频错误地认定为经过 AI 换脸的内容。

由于现有检测技术缺乏实时检测能力，对于一些场景，目前的检测技术难以做到实时监测和预警。2023 年 4 月，福建某科技公司法人代表郭先生的微信好友通过联系他，称竞标需 430 万元保证金，郭先生通过视频“核实”身份，10 分钟内分两笔转账 430 万元，事后致电好友才知被骗[9]。在此案中，诈骗分子就是利用检测技术的这一漏洞，在与受害者实时交流过程中通过 AI 换脸实施诈骗。

当前，针对 AI 换脸诈骗的防范技术发展仍然存在显著不足，这一技术缺口使得诈骗分子能够轻易采用对抗性攻击手段，对检测系统实施干扰和破坏，导致技术部门的应对进入“被动防御”的不利局面。

3.3 社会层面

由于宣传缺乏持续性，公众在宣传活动结束后，随着时间推移，对相关防范知识的记忆和重视程度会逐渐下降，难以形成持久的防范意识；并且宣传针对性不强，不同年龄、职业、文化程度的人群对 AI 换脸诈骗的认知和接受能力存在差异，但目前的宣传工作往往缺乏针对性，没有根据不同群体的特点制定个性化的宣传方

案，导致宣传效果大打折扣；各类宣传渠道之间缺乏有效的整合和协同，传统媒体、新媒体、社区宣传等各自为政，没有形成宣传合力，这使得宣传内容难以全面、及时地覆盖到所有受众，影响了宣传的广度和深度。

在防范 AI 换脸诈骗过程中，需要不同机构和部门之间进行数据共享，以便更全面地掌握诈骗线索和信息。然而，数据共享可能会引发隐私保护问题，如何在两者之间找到平衡是一个难题。如果过于强调数据共享，可能会侵犯用户的隐私权；而过于注重隐私保护，又可能导致数据流通不畅，影响防范工作的效率[10]。

3.4 防范主体层面

政府职能部门之间存在治理壁垒。AI 换脸诈骗的防范工作涉及公安、工信、网信、市场监管、司法机关等多个职能部门，但各部门的职能定位和治理逻辑并不完全相同。由于各职能部门之间存在职能定位的差异，导致三方面问题。其一，信息共享程度低。公安机关的诈骗案件信息、工信部门的技术风险预警信息、网信部门的舆情监测信息没有实现动态共享。其二，权责界划分“真空”。对 AI 换脸技术的“事前防范—事中防范—事后惩戒”的三个阶段，尚未厘清各环节的主管部门。其三，步调难以协调。司法机关对于 AI 技术犯罪的法律定性还在探索过程中，公安机关对于案件快速处置有着现实压力，容易出现快慢、虚实、轻重不分的问题。

社会组织参与防范的可行性以及有效性存在问题。社会组织尚未充分发挥其应有作用，主要原因有三，一是尚未健全政策激励机制。尽管有个别试点，但总体来看，除政府购买服务项目外，大多数社会组织不具备政府购买服务资质，资金来源主要依靠企业赞助，缺乏持续性。二是尚未打通参与防范“公安-工信-社会”通道。三是参与防范能力建设不足。中小社会组织普遍缺乏 AI、法律等方面专业人才，开展技术监测、公众教育的能力不足，亟需建立健全“政府-高校-社会”协作机制。

4 综合治理背景下防范利用 AI 换脸技术实施诈骗行为的思路

综合治理是党和国家统筹社会问题的重要方式，是指在各级党委和政府的统一领导下，多主体协同参与，综合运用政治的、经济的、行政的、法律的、文化的、教育的等多种手段[11]，强调治理主体的多元性和层级间的协同合作，要求社会公众共同参与，并建立科学系统的治理体系，实现资源共享和手段配合。

下面是贯彻落实综合治理的具体措施。

4.1 制度层面

从民法对举证和担责主体需进一步完善补充，法院在对相关案件进行裁判时，应以过错推定原则为指引。举证责任分配上则应适用《民法典》第 1165 条的特殊规则，即被侵权人能初步证明存在深度伪造行为的，推定技术使用者及传播平台对此存在过错。实施可参考 24 年北京互联网法院“AI 换脸侵权案”的裁判经验，即原告需证明基础事实（如视频存在换脸痕迹）、被告需证明技术正当性（如获得授权）、法院综合评估损害后果的“三步法”来展开证明[12]。在《民法典》人格权编中第 1035 条之一增加规定“任何组织或个人不得利用深度伪造技术侵害他人肖像权、名誉权及其他人格权益[13]。”进一步参照欧盟《数字服务法》相关规定，建立技术提供者与内容传播者双轨责任机制，明确责任主体，进而要求 AI 换脸工具开发者进行必要的技术限制。

应构建刑事规制体系完善方案[14]，立法机关在明确 AI 换脸诈骗犯罪构成要件的同时，增设加重量刑的“特别严重情节”。在《刑法》第 266 条中明确客观要件：以“利用深度伪造技术虚构身份实施诈骗”为犯罪行为，“诈骗数额特别巨大或有其他特别严重情节”为犯罪结果。量刑则可将“诈骗数额 50 万元以上”、“造成被害人自杀等严重后果”、“伪造国家机关工作人员形象”、“针对老年人、未成年人实施诈骗”等情形认定为特别严重情节[2]，作为量刑标准，加重刑罚[15]。

同时，国家网信办相关政策，也应当进行完善。相关政策明确提出了服务提供者定期审核算法的要求，但未指明具体的审核周期以及未来的标准更新机制，对此应当在后续文件中进一步补充相关内容。此外，规定、政策中强调了平台责任，但却未涉及实际诈骗行为中的技术提供商、数据供应商等其他多方主体，即存在责任

[2] 《2023 年移动支付安全大调查报告（中国银联）》显示，75.7%的大学生和 77.7%的大龄群体收到过诈骗信息。2023 年至 2025 年，在 AI 诈骗案件中，老年人和年轻人各占 27%，学生和职员群体各占 10%。

链条模糊的不足，该漏洞可以通过引入“过错推定原则”进一步扩大责任主体的范围，防止“甩锅”、“背锅”现象。在用户层面，鉴于尚未明确商务数据加密储存标准，用户个人信息在“合法处理”过程中仍然存在数据泄露风险，对此可结合国际经验和专家建议制定合理且实际可行的储存标准，保障个人信息安全。

综合治理协同体系的构建，也是十分必要的。首先，借鉴外国经验建立技术溯源机制。由工信部牵头制定《深度伪造技术应用管理办法》，要求 AI 换脸工具必须嵌入数字水印标识。借鉴美国 NIST 的深度伪造检测标准，建立国家级检测认证中心，对市场流通的换脸软件进行定期安全评估。其次，完善跨境司法协作。结合 2024 年中柬联合破获的“东南亚 AI 诈骗集团[3]” 案件，参照《联合国打击跨国有组织犯罪公约》第 18 条，推动建立“一带一路”数字取证协作平台。

4.2 技术层面

相关技术部门可以研发 AI 换脸内容的检测工具。蚂蚁数科的 ZOLOZ Deeper[16]、英特尔与纽约州立大学宾厄姆顿分校合作开发的实时 Deepfake 检测器（称为 FakeCatcher）[17]等工具，为相关技术部门的研发提供了技术经验。技术部门在研发工具的同时，也可以利用深度学习、数字水印等技术识别虚假内容。相关科研机构、企业能够为研发软件提供技术支持，技术部门可以与这些企业进行合作，共享反欺诈技术，持续优化 AI 模型，提高检测工具对换脸内容的识别能力，减少 AI 换脸误判和漏判。技术部门可以联合相关部门建立诈骗行为黑名单，共享可疑 IP、账号和设备信息，与社交媒体、视频平台进行合作，快速识别和拦截虚假内容。

一些受害人轻易点击陌生人发送的链接，在这一过程中，木马病毒便入侵到受害人的通信工具中，对其个人信息和生活动态进行监控和提取[18]。犯罪分子便是利用这种手段获得受害者的信息并实施犯罪。针对这些情况，技术部门应当利用 AI 技术加强内容审核，快速识别并下架虚假换脸内容，建立便捷的举报渠道，鼓励用户举报可疑内容，并快速响应处理。

技术部门在 AI 的研发与应用过程中也要严格遵守伦理及法律规定，保证 AI 技术的创新与运用既符合道德规范，又满足法律法规要求[19]。着力构建人 工智能伦理规范与法律法规的衔接融合机制，推进人工智能伦理法治化进程[20]。AI 换脸技术的设计者和使用者都必须遵守相关的法律规定，遵守以《中华人民共和国电子商务法》、《中华人民共和国数据安全法》等法律法规为基础的初步网络法律体系[21]。

随着人工智能系统在日常生活中的全面应用，如何切实保护个人隐私，避免数据被窃取和非法利用将愈发成为一个极具挑战性的课题[22]。技术部门应当通过技术手段向公众普及 AI 换脸诈骗的识别方法，提高防范意识，同时在平台中嵌入警示系统，当用户接触可疑内容时自动弹出提醒，并在程序醒目位置以醒目字体标注“该内容可能涉及 AI 诈骗”，以此提醒用户注意防范。

4.3 社会层面

防范利用 AI 换脸实施的新时代诈骗，要通过宏观和微观共同发挥作用，形成“线上+线下”多维度、全覆盖的宣传与防范 AI 换脸诈骗体系。

首先是线下治理措施。营造多主体的协作氛围，着手打造“家庭+学校+社区+企业”共同发展的反诈宣传系统，2025 年 3 月，重庆理工大学学生党员志愿者在龙骏社区开展“AI 反诈课堂”，通过动态 PPT 演示向群众展示 AI 换脸模拟诈骗案例，揭露“假冒亲友求助”等骗局[23]。通过在线下举办讲座、研讨会、在线课程等形式，普及人工智能安全知识，增强社会各界的安全意识，形成全社会共同关注、共同参与人工智能安全治理的良好氛围[24]。同时要对特定群体进行特殊保护，重点关注 18 岁以下的未成年群体，使其在义务教育阶段就从家庭教育和学校讲座中产生对 AI 换脸诈骗的警惕意识，在宣传时可结合青少年的偏好，将相关案例制作成“闯关游戏”、“答题比拼”等更加富有乐趣的形式；对于 55 岁以上的中老年人群体，其获取社会热点的媒介较窄，不像年轻人般迅速获取“AI 换脸”诈骗案例的热点资讯，社区的老年活动室也应定期与当地基层警察局有效联动，合作组织线下面对面宣讲活动，例如 2024 年 11 月，北京市丰台区芳古园社区社工就与民警联动，成功拦截一起利用 AI 换脸冒充马云对老年人实施的诈骗。诈骗分子通过 AI 换脸技术伪造视频通话，以“扶贫项目分红”为诱饵要求老人下载 QQ 软件并转账。民警现场与诈骗分子视频连线拆穿了骗局，避免了经济损失[25]。在进行反诈宣传时，要侧重于发放宣传手册、宣传物品等老年人更易接触到的实物，在社区公

[3] 中柬联合破获东南亚 AI 诈骗集团案例。

告栏等醒目处张贴反诈宣传海报，加强中老年人的防范意识和能力。使得 AI 换脸反诈宣传能基本覆盖到全年龄段，在日常生活中潜移默化地提升全社会对于 AI 换脸诈骗的识别能力。

由于 AI 换脸诈骗依托于互联网，具有线上网络媒介的强依附性，在线下全社会普及宣传教育时，线上的宣传和环境治理更是不可或缺。

除了线下讲座、社区宣传等传统渠道，还可以借助新媒体，在用户众多的网络论坛、短视频平台上投放反诈宣传广告，鼓励官方平台录制基于真实案例改编的宣传教育歌曲或视频。并在这些平台上加强以国家反诈中心为代表的反诈 APP 的宣传，反诈软件中有着丰富的防诈知识与案例展示，能增强群众识骗防骗意识，提升自我保护能力。同时，反诈软件能精准识别 AI 换脸平台的网址，及时预警，极大降低民众受骗风险。同时对接网络监管部门，打造数据安全和隐私保护的网络保障机制。数据是人工智能发展的基础和核心，确保数据的安全性和隐私性是人工智能安全治理的重要内容。应进一步完善数据安全和隐私保护相关法律法规，健全数据治理制度，严格规范数据的采集、存储、使用和共享，提高数据质量并且增强数据的真实性、准确性、客观性、多样性[26]，防止数据滥用和隐私泄露，还可建立用户信用评级体系，对于信用评级较低的用户，限制其部分功能，同时加强对其发布内容的审核，为人工智能发展营造安全可信的环境。

在当今社会，由于利用 AI 换脸与诈骗相结合尚为一种较为新颖的手段，初步建立的数据安全保障机制难免有不足之处，在总体完善网络保障机制的同时可引入民众的主观能动作用弥补体系无法注意到的细微之处，发挥群众的监督作用，采用举报奖励机制，鼓励用户手动对可疑内容进行举报，利用反诈软件的线索收集、反馈渠道，群众发现涉诈线索能一键举报，或拨打反电信网络诈骗专用号码 96110，为公安机关打击犯罪提供有力支撑，形成全民参与反诈的良好局面，推进共筑人工智能良好应用环境。

4.4 主体协同层面

AI 换脸技术的事前研发企业、网络平台应发挥技术链主要责任主体作用，强化主体责任意识。一方面，在技术层面，采取“端到端”的安全措施，从算法源头上植入多重安全机制，运用数字水印、生物识别等技术，第一时间发现“异常操作”——频繁、异常的换脸请求，及时熔断并推送预警[27]。另一方面，保持与监管部门的沟通畅通，共享底层技术参数和场景落地措施，配合相关部门完成安全审计和临时检查。及时处理违法使用可能，如发现黑产侵入，第一时间冻结接口，并及时进行电子取证，与网安部门联合开展溯源打击，斩断“滥觞之祸”。

国家实验室应当发挥技术优势，打造动态防御引擎，研发多模态融合特征，提高模型判断能力，将响应时间控制在毫秒以内。同时建立实时更新的深度伪造样本池，通过对抗训练不断完善检测引擎，伪造内容全网传播必留痕迹。最重要的是，建立产学研用合力，与互联网头部平台合作设立检测节点，与国家实验室合作建立开放数据集，多方共建 AI 安全矩阵，形成内容全网追踪、技术全网阻断的打不断、骗不来的反诈生态。

对 AI 换脸技术的宣传很大程度上源于人们对于 AI 换脸技术的知之甚少，所以宣传部门可以通过电视、广播、网络等媒体进行全方位、多层次的宣传。宣传部门也可以通过舆论引导营造良好的网络环境，利用主流媒体和社交平台宣传 AI 换脸技术在影视制作、虚拟现实等场景中进行的合法合规的应用，引导人们正确认识和运用 AI 换脸技术。同时，对那些利用 AI 换脸技术实施诈骗的行为进行及时曝光，形成舆论压力，以震慑不法分子。社区作为基层治理单元，在预防 AI 换脸诈骗中发挥着重要作用。社区工作人员可组织志愿者深入小区、商场等地发放宣传手册，面对面为居民宣讲防范知识。针对老年人、学生等易受骗人群开展专项讲座，以通俗易懂的语言、真实案例，提升其识别能力和应对技巧。此外，建立社区预警机制，及时收集、发布周边诈骗信息，使居民时刻保持警惕。同时，高校作为知识创新和人才培养的重要场所，应积极参与防范宣传活动。在课程教育方面，在相关网络安全课程中加入 AI 换脸诈骗防范的内容，培养学生安全意识及法律素养。引导学生参与科研项目，开发出可以防范 AI 换脸诈骗的技术和产品。组织学生志愿者进社区、进乡村开展宣传活动等。

目前，我国《网络安全法》《个人信息保护法》等法律法规已对网络空间行为进行初步规制，但尚未针对 AI 换脸技术进行专门规制。因此，行政机关应当加强对 AI 换脸技术供给与使用端的治理。各部门应当建立 AI 换脸技术登记造册制度，要求 AI 换脸技术提供者与使用者在产品发布前进行登记造册，并确保其产品用途的正当性。对于未经登记造册或违法使用 AI 换脸技术的主体，则应依法予以惩戒。司法在规制 AI 换脸技术诈

骗中的作用包括：加大对利用 AI 换脸技术实施诈骗行为的打击力度。司法机关应当设立网络犯罪侦查部门,通过大数据、人工智能等技术，追踪利用 AI 换脸技术实施的诈骗犯罪，及时惩处犯罪分子。明确利用 AI 换脸技术实施诈骗行为的法律责任。应当依据利用 AI 换脸技术实施诈骗行为的情节轻重，依法惩处犯罪行为。对于提供技术支持，明知客户所利用 AI 换脸技术应用在违法犯罪活动中的平台或个人，应承担相应责任。

5 结语

技术的迭代更新在拓展人们权利边界的同时也伴随着具有时代特色的不法侵害隐患，人工智能技术的发展导致网络环境下的诈骗犯罪呈现出全新的复杂发展态势，其 AI 换脸的手段在侵犯传统的财产权益的同时，更涉及人格权等多重权益。为保障公民的合法权益，回应保障安全、打击不法侵害的社会需求，针对当前社会面临的挑战，本研究基于社会调查和文献分析，结合《民法典》及《刑法典》等相关法律法规，分析不法分子利用 AI 换脸技术实施诈骗行为的法律定性，阐述现有防范 AI 换脸技术实施诈骗行为在法律追责、技术监测、社会宣传及防范主体责任等方面存在的不足，旨在构建多元化的系统治理体系，通过强化制度衔接，构建跨领域融合的多方监管机制，把握技术创新和风险防控的动态平衡，明确和强调政府部门、技术平台、社会组织和公众的责任分工，探索一条治理人工智能换脸诈骗的多主体协调路径，持续完善“制度-技术-主体-协同”的四位一体防范体系，为保障数字时代的公共安全和社会稳定提供理论支持。

基金项目

南京理工大学大学生创新创业训练计划省级立项资助（202410288260Y）

参考文献

[1] 张伯礼. 央视曝光 AI 假冒名人明星，被 AI 伪造宣传护肤品[EB/OL]. (2025-03-16) [2025-04-1]. https://news.qq.com/rain/a/20250316A04KG000.

[2] 新京报传媒. 央视 3·15 晚会曝光换脸骗局, AI 套路远远不止这些[EB/OL]. (2024-03-16) [2025-04-01]. https://news.qq.com/rain/a/20240316A06GRY00.

[3] O'CONNOR C, WEATHERALL J O. How misinformation spreads—and why we trust it[J]. Scientific American, 2019, 321(3): 54-61. DOI:10.1073/pnas.2104239118.

[4] 成都商报红星新闻. 每天问候、深夜跟“姐姐”谈心！假冒演员靳东实施诈骗, 8 人获刑[EB/OL]. (2024-02-03) [2025-01-31]. https://news.qq.com/rain/a/20240203A0374U00.

[5] 央视网. 防范“AI 换脸”诈骗，你需要的知识都在这儿了[EB/OL]. (2024-02-26) [2025-02-15]. https://news.cctv.com/2024/02/26/ARTIfJJNnT6fAdR8jRKPOgBe240226.shtml.

[6] 王桔波. AI 换脸检测关键技术研究[D]. 重庆:重庆邮电大学, 2021.

[7] 重庆市渝北区人民法院. (2023) 渝 0105 民初 36081 号[Z]. 2023.

[8] 陈曦. 震惊！“变脸”冒充 CFO，骗走两个亿！香港最大 AI 诈骗案细节曝光[EB\OL]. (2024-02-06) [2025-03-19]. https://www.thepaper.cn/newsDetail_forward_26284950.

[9] 寿光市人民政府. 先声音再视频 所谓“熟人”来敲门？——“AI 换脸”诈骗乱象风险预警[EB/OL]. (2023-12-04) [2025-15]. https://www.shouguang.gov.cn/credit/f xts/zdlyfxts/202312/t2 0231204_47585.html.

[10] 王威力, 王玥珺. 全球 AI 治理中的社会技术想象——人工智能发展规划的国际比较分析[J/OL]. 情报资料工作, 1-32[2025-02-27]. http://kns.cnki.net/kcms/detail/11.1448.G3.20250227.1253.004.html.

[11] 张凌寒. 深度合成治理的逻辑更新与体系迭代——ChatGPT 等生成型人工智能治理的中国路径[J]. 法律科学(西北政法大学学报), 2023(3):38-51.

[12] 戴敏敏. “换脸视频” 智能技术的法律规制研究[J]. 科技与法律, 2021(3): 89-96.

[13] 樊芸. AI 换脸精准诈骗难以防范——全国人大代表建议推进立法强化人工智能技术安全应用[N]. 光明网, 2024-04-10.

[14] 中国犯罪学学会. 追踪丨 AI 换脸猖獗、95% 的诈骗导流第三方——中国犯罪学学会副会长皮勇：犯罪在变, 教材没有及时更新[N]. 每日经济新闻, 2025-03-15.

[15] 最高人民法院. 《关于审理涉人工智能犯罪案件适用法律若干问题的解释（征求意见稿）》[Z]. 2024.

[16] 薛世轩. 蚂蚁数科发布反深伪产品 ZOLOZ Deeper,设置超百万奖金池支持挖掘漏洞[EB/OL]. (2024-04-16) [2025-4-26]. https://www.163.com/tech/article/IVTKCH7L00098IEO.html.

[17] IT 之家. 英特尔推出 FakeCatcher，通过检测血流判断是否是 AI 换脸[EB/OL]. (2022-11-15) [2025-3-16]. https://news.qq.com/rain/a/20221115A01W9B00.

[18] 于顺, 滕宝毅. “AI 换脸”冒充身份诈骗犯罪防治对策研究[J]. 法制博览, 2024(9): 130-132.

[19] 吴仲琦, 曹亚领, 代涛. 负责任人工智能治理体系研究[J]. 世界科技研究与发展, 2024(4): 536-547.

[20] 张欣. 人工智能治理的全球变革与中国路径[J]. 华东政法大学学报, 2025, 28(1): 18-32.

[21] 沈芳君. 生成式人工智能的风险与治理——兼论如何打破“科林格里奇困境”[J]. 浙江大学学报(人文社会科学版), 2024, 54(6):73-91.

[22] 董双. 全球人工智能治理建设性方案研究[C]//2024 世界智能产业博览会人工智能安全治理主题论坛论文集. 2024: 59-61.

[23] 晏红霞. AI 赋能志愿服务——校社联动追“锋”前行 [EB/OL]. (2025-3-14) [2025-4-15]. https://ne ws.qq.com/r ain/a/20250314A087XS00.

[24] 李超, 郑毅平. 提高人工智能安全治理水平[N]. 重庆科技报, 2025-01-21(002).

[25] 北京日报. 利用 AI 换脸对老人实施诈骗, 丰台一社区社工与民警联动成功拦截！[EB/OL]. (2024-11-26) [2025-2-13]. https://www.163.com/dy/article/JHUP2D1M0552UKT5.html.

[26] 寿晓明. 生成式人工智能风险、规范和治理路径——以融贯流程为视角[J/OL].南京邮电大学学报(社会科学版), 1-11[2025-02-27]. https://doi.org/10.14132/j.cnki.nysk.20250225.001.

[27] 国家市场监督管理总局，国家标准化管理委员会. 信息安全技术人脸识别数据安全要求: GB/T 41819-2022[S]. 北京: 中国标准出版社, 2022.