Open Journal of Legal Science 法学, 2026, 14(8), 139-144
Published Online August 2026 in Hans. https://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2026.148241

Hans汉斯

“深度伪造”的刑法挑战与规制路径研究

朱周翔

江苏师范大学法学院，江苏 徐州

收稿日期：2026年7月13日；录用日期：2026年7月23日；发布日期：2026年8月24日

摘　要

人工智能“深度伪造”技术滥用引发的犯罪问题已对刑法体系构成严峻挑战。鉴于此类犯罪呈现法益侵害多样性的特征，我国现行刑法“事前打击 + 事后惩治”的规制模式已显现出明显的缺陷：事前打击模式对公开获取生物识别信息的行为存在规制空白；事后规制模式则面临罪名适用存疑的难题。对此，应坚持“技术中立性”理念，通过更新“公民个人信息”刑法内涵、强化伪造中间环节规制等路径，构建精准周延的刑法规制体系，在防范犯罪风险与保障技术发展之间实现平衡。

关键词

人工智能，“深度伪造”，刑法规制

Research on Criminal Law Challenge and Regulation Path of “Deepfake”

Zhouxiang Zhu

Law School, Jiangsu Normal University, Xuzhou Jiangsu

Received: July 13, 2026; accepted: July 23, 2026; published: August 24, 2026

Abstract

The criminal problem caused by the abuse of artificial intelligence “deepfake” technology has posed a severe challenge to the criminal law system. This kind of crime is characterized by the infringement of diversity by legal interests, and there are obvious defects in the regulation mode of “pre-strike + post-punishment” in China’s current criminal law: the pre-strike mode has no regulation on the public access to biometric information. However, the ex post regulation mode faces the difficult problem of the application of charges. In this regard, we should adhere to the concept of “technology neutrality”, update the criminal law connotation of “citizens’ personal information”, and strengthen the regulation of forged intermediate links to build an accurate criminal law regulation

文章引用: 朱周翔. “深度伪造”的刑法挑战与规制路径研究[J]. 法学, 2026, 14(8): 139-144.
DOI: 10.12677/ojls.2026.148241

system in GAI, so as to achieve a balance between preventing crime risks and ensuring technological development.

Keywords

Artificial Intelligence, “Deepfake”, Criminal Law Regulation

1. 引言

“深度伪造”(Deepfake)是人工智能深度学习(Deep Learning)和伪造(Fake)的融合产物，其以生成式对抗网络技术为基础，通过篡改或合成的方式生成非真实的音频和视频。随着近年来各方对生成式人工智能(AIGC)研发的投入不断加大和基础技术的迭代升级，深度伪造在短期内得到了长足的发展，在处理音视频的能力上已经达到了混淆视听的地步。

事实上，深度伪造技术已在教育、影视、医疗等多个领域展现出其独有的生产力价值。然而，技术永远是一体两面，深度伪造技术在给人们带来福利的同时，若被滥用也会给国家利益、社会利益和公民个人利益带来一系列的风险挑战。面对人工智能深度伪造所带来的犯罪风险与挑战，如何理解其背后的刑法逻辑，发掘与之对应的规制路径，已成为当前亟待解决的严峻课题。

2. “深度伪造”对我国刑法的挑战

2.1. 法益侵害多样化

人工智能深度伪造犯罪，是一种复合型的犯罪，其存在着鲜明的犯罪链条，可以分为伪造前端、伪造中间、伪造后端三个阶段。各环节均可能触碰不同类型的法益，最终形成多样化的法益侵害格局[1]。

在伪造前端数据取得时，犯罪人以实施深度伪造的目的搜集个人信息的行为，最先直接对准公民的个人信息权益，不管是经社交媒体爬取而来的还是其他途径获取的人脸、声音、身份等信息，统统都属于受法律保障的人格权范畴，此阶段的行为已经涉嫌侵犯公民的隐私权、个人信息受保护权，如果搜集的是特定身份信息或者敏感信息，极有可能触碰公民的人格尊严以及人身安全法益。

伪造中间内容合成阶段，尽管核心为技术加工行为，但此阶段对数据的篡改与合成本质上是对公民人格权的更进一步侵害，借助技术手段虚构公民未施行的言行、扭曲公民真实形象，已经侵犯公民的肖像权、名誉权以及荣誉权，特别是当合成内容关乎低俗、违法场景时，会对公民的精神利益造成直接损害，进而侵害公民的人格尊严法益。

而到了伪造后端的内容发布与传播阶段，法益侵害便呈现出扩散与聚焦并存的特性，依据发布目的的不同，可能侵害的法益越发多样：如果用于网络诈骗，就会侵害公私财产所有权；如果用于虚假诉讼，便会破坏司法公正与诉讼秩序这一公法益；如果用于散布谣言、煽动对立，那就会侵害社会公共秩序与公共利益；如果用于敲诈勒索，既侵犯公民财产权，又侵犯公民的人身权利与安全。

2.2. 事前打击模式：存在空白与漏洞

《中华人民共和国刑法》(以下简称《刑法》)第二百五十三条之一规定了非法获取、非法出售或非法

提供公民个人信息将承担“侵犯公民个人信息罪”的罪责：前者是指以窃取或者其他方法非法获取公民个人信息的行为，后者即在违反国家规定的前提下，向他人非法出售或者非法提供公民个人信息，或将在履行职责或者提供服务的过程中获得的公民个人信息出售或者提供给他人的行为。而根据《最高人民法院、最高人民检察院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》[1](以下简称《侵犯公民个人信息解释》)第五条，如果行为人“知道或者应当知道他人利用公民个人信息实施犯罪”，仍向其出售或者提供该信息，则该行为应当被依法认定为刑法第二百五十三条之一规定的“情节严重”。由此可见，事前打击的规制模式是以非法获取的原始数据为跳板，通过伪造前端和伪造中间环节在原材料上的紧密供给关系，从源头阻断深度伪造犯罪的素材供给链条。从语义来看，只有“非法获取”行为才须要受到规制，刑法并不禁止“合法获取”公民个人信息的行为。但在司法实践中，犯罪人所收集的公民个人信息往往来自公开的互联网平台，这说明，犯罪人若想为中间环节收集素材，其实并没有触犯侵犯公民个人信息罪的风险，那么事前打击模式也就无法发挥规制中间行为的效果。

再看《侵犯公民个人信息解释》第一条，“‘公民个人信息’，是指以电子或者其他方式记录的能够单独或者与其他信息结合识别特定自然人身份或者反映特定自然人活动情况的各种信息，包括姓名、身份证件号码、通信通讯联系方式、住址、账号密码、财产状况、行踪轨迹等。”这里的“公民个人信息”，应当与“生物识别信息”作出明确的区分。“生物识别信息”是指“紧密依附于身体、直接体现公民个人身体特征和行为特征，对人身具有直接指向性的特殊公民个人信息[2]”，包括人脸、指纹、声纹、虹膜、DNA 等[2]。因此，生物识别信息仅可以被归类到《侵犯公民个人信息解释》第五条提及的“规定以外的个人信息”，需达到五千条以上才能构成犯罪。实际上，前者作为深度伪造的数据来源意义不大，后者才是伪造前端行为所主要侵害的对象，且其具有直接的人身指向性，往往与公民的人格权和依附在其上的财产权所直接挂钩，若坚持现行解释，则必然会留下规制漏洞。此外，绝大多数情况下，被害人的生物识别信息受到人工智能深度伪造犯罪侵害的只是关键的几条，达到五千条的入罪要求几乎不可能，这就导致被害人通过事前打击模式寻求救济这一途径落空。

由此见得，事前打击模式存在相当程度的空白和漏洞，仍然是不成熟的。

2.3. 事后规制模式：罪名适用存疑，罪与非罪有争

事后规制模式的实现方式，在于打击深度伪造内容生成以后的网络发布与传播所引起的关联下游犯罪。

首先，以“换脸视频”为典型，此类深度伪造行为侵害他人人格权，我国《刑法》第二百四十六条规定了侮辱罪与诽谤罪两种可能的追责路径。侮辱罪是指以暴力或其他方法公然贬低他人人格、破坏他人名誉，手段可包括语言、文字、图片、录像及网络发布等；诽谤罪则是故意捏造并散布虚构事实，贬损他人人格名誉。两罪均需达到“情节严重”标准，即涉案信息点击量、浏览量达 5000 次，或转发量达 500 次，或造成被害人及其近亲属精神失常、自残、自杀等严重后果[3]。该标准的立法初衷，是基于刑法谦抑性原则，避免过度追责，合理分配司法资源。但在深度伪造技术广泛应用的今天，这一高标准反而可能成为犯罪分子脱罪的借口，从而助长此类犯罪蔓延。

此外，当前网络犯罪多呈现协同运作模式，深度伪造的前、中、后三端往往分属不同主体。这对传统共同犯罪理论提出了挑战。传统共同犯罪中，行为人间的意思联络是必备要件。但在深度伪造催生的

[1]https://flk.npc.gov.cn/detail?id=402881e45ffbbe41015ffbfc111a0350
[2]《人民检察院办理网络犯罪案件规定》第六十二条第六款。
[3]《最高人民法院、最高人民检察院关于办理利用信息网络实施诽谤等刑事案件适用法律若干问题的解释》第二条。https://www.cac.gov.cn/2013-09/07/c_133142246.htm

链式犯罪生态中，犯罪分子多素未谋面，借助互联网自发形成合作关系，各环节基于独立犯罪目的行事，呈现职责划分明确、流程化操作的特征[3]。一方面，若上游技术开发者与下游犯罪实施者缺乏共同犯罪故意与意思沟通，不应认定为共犯；另一方面，技术服务方对下游犯罪缺乏实际支配力，无法成立间接正犯。依据共犯从属性理论，若下游行为仅属行政违法而非犯罪，技术服务提供者的行为亦应认定为无罪，使其处于难以追责的“真空状态”[4]。

3. 人工智能“深度伪造”的规制路径刍议

3.1. 坚持“技术中立性”理念

技术中立性原则的核心要义在于，技术本身并没有善恶之分，其所产生的正面或负面价值取决于运用技术的主体与目的。深度伪造技术以生成式对抗网络为核心，本质上是信息生成与处理的工具，其算法逻辑与技术架构本身不具备刑法意义上的社会危害性。将犯罪归咎于技术本身，不仅违背技术发展规律，更会陷入“因噎废食”的误区。在科技浪潮的推动下，深度伪造技术如同一颗冉冉升起的新星，其发展前景与潜在风险并存，亟待社会各界以审慎态度进行探索与规范。刑法作为国家法律体系中的保障法与最后手段，其功能的发挥应当遵循谦抑性原则。如果仅仅因为该技术在实践中存在被滥用而引发犯罪的可能性，便进行全盘否定、一味限制乃至施以过重的刑法规制，可能会导致技术发展的“寒蝉效应”，使得研究者和企业不敢触碰相关领域，从而错失技术带来的发展机遇，更可能使刑法异化为追求短期治理效果的工具，从而背离刑法保障人权、维护社会公平正义的根本使命，削弱刑法理论的时代适应性与实践生命力。

3.2. 更新“公民个人信息”的刑法内涵

上文提到，因为《侵犯公民个人信息解释》中公民个人信息概念的不周延，对于生物识别信息的保护存在法律漏洞，导致了过高的入罪门槛。有学者曾主张，为了避免这种不周延，应当充分发挥法解释学方法论，将生物识别信息解释为“健康生理信息”，从而适用《侵犯公民个人信息解释》第五条第四项的规定[4]，适用500条的标准，以降低入罪门槛[5]，这显然不妥。首先，生物识别信息与健康生理信息的内涵存在显著区别。在《信息安全技术 个人信息安全规范》[5]中，对生物识别信息与健康生理信息做出了明确分类。单从法律设置的逻辑上来看，这两者之间就不存在相互解释的空间，强行将前者解释为后者，忽视了生物识别信息的独特敏感性以及它背后所蕴藏的特定风险[6]。此外，由原先的5000条降低到500条，看似降低了门槛，但上文已经指出，生物识别信息与人格权、财产权等利益深度绑定，仅仅一条的泄露，便能够引发雪崩效应，造成巨大的损失，而如今人工智能深度伪造犯罪的成本、门槛又如此之低，犯罪人不需要网罗大量信息，凭借几条生物识别信息便可以实施较为完善的犯罪。在此背景之下，如此强调数量标准，本身就失去了意义。

《中华人民共和国个人信息保护法》(后文简称《个人信息保护法》)将公民个人信息分类为“一般个人信息”和“敏感个人信息”。其又在第二十八条通过列举的方式释明了何为“敏感个人信息”，其中，生物识别信息更是被摆在了首位[6]。由于我国《刑法》中并不存在“生物识别信息”的概念，更没有“敏

[4]《最高人民法院、最高人民检察院关于办理侵犯公民个人信息刑事案件适用法律若干问题的解释》第五条第四项规定，“对于非法获取、出售或者提供住宿信息、通信记录、健康生理信息、交易信息等其他可能影响人身、财产安全的公民个人信息五百条以上的”，即可以被认定为《刑法》第二百五十三条之一规定的“情节严重”。

[5]https://std.samr.gov.cn/gb/search/gbDetailed?id=A0280129495AEBB4E05397BE0A0AB6FE

[6]《中华人民共和国个人信息保护法》，第二章个人信息处理规则中，第一节为一般规定，主要规定了一般个人信息的处理规则，第二节为敏感个人信息的处理规则，突出对敏感个人信息的着重保护，彰显了《个人信息保护法》对于一般个人信息与敏感个人信息区分保护的立法态度。《个人信息保护法》第二十八条第一款：“敏感个人信息是一旦泄露或者非法使用，容易导致自然人的人格尊严受到侵害或者人身、财产安全受到危害的个人信息，包括生物识别、宗教信仰、特定身份、医疗健康、金融账户、行踪轨迹等信息，以及不满十四周岁未成年人的个人信息。”

感个人信息”的分类，即便违反了《个人信息保护法》，也无法寻求刑事救济。基于法秩序统一原理以及此种分类的合理性，在刑法中引入此等分类制度来更新公民个人信息的内涵，具有一定的规范性和科学性[7]。

3.3. 伪造中间阶段的规制

在人工智能深度伪造的犯罪链条中，非法运用该技术的中间环节扮演着承上启下的核心角色。尽管从形式上看，这类行为带有明显的帮助犯特征，但其在犯罪中扮演的关键角色和独特社会危害性，使得它与正犯行为的传统关系发生了本质变化：中间行为实际上已对上游侵害行为的滋生和下游危害结果的蔓延起到主导作用。这种主导性体现在犯罪链条的双向传导中：一方面，深度伪造技术对生物识别信息的刚需，成为了非法获取、售卖及提供公民生物识别信息等犯罪高发的根本诱因；另一方面，利用非法获取的信息深度伪造图像、视频，是行为人实施下游各类犯罪的必要前置环节。

当前，上游的侵犯公民个人信息罪与下游多种传统罪名，虽已覆盖了人工智能深度伪造犯罪的部分行为，但这种“事前打击 + 事后惩治”的规制模式，难以全面覆盖深度伪造技术滥用的完整犯罪链条。尤其是对于滥用技术合成深度伪造内容的中间行为，现有罪名体系已经显得力不从心。因此，要实现对这类犯罪更有效的预防与惩戒，就必须转变以往的思路，适时将深度伪造技术滥用行为本身纳入刑法规制体系，构建更精准、周延的法律规制网络，从源头上遏制人工智能深度伪造带来的多重犯罪风险。

对此，可以着眼《刑法》体系本身，在与人工智能深度伪造犯罪相贴近的几个罪名中寻求其生存空间。

以帮助信息网络犯罪活动罪(以下简称帮信罪)为例，根据《最高人民法院、最高人民检察院关于办理非法利用信息网络、帮助信息网络犯罪活动等刑事案件适用法律若干问题的解释》[7](以下简称《信息网络犯罪解释》)的规定，“提供专门用于违法犯罪的程序、工具或者其他技术支持、帮助的”可以认为行为人“明知他人利用信息网络实施犯罪”。这里的“专门用于”的表述限制了帮信罪在人工智能深度伪造犯罪中的成立空间。根据技术中立性视角，人工智能深度伪造技术本身不具有合法或者非法的性质，所以技术的提供者、服务商难以成立帮信罪；至于人工智能深度伪造软件的开发者和开源者，因其不具备犯罪目的，不能成立“明知”，往往也没有可归责性[8]。虽说按照严格的罪刑法定原则本应如此，但司法实践中，为了提高司法效率，也限于司法人员的认识水平和高新技术飞速发展之间的不同步性，对于某项技术是否是“专门用于”违法犯罪的判断则要宽松得多。往往只根据《网络音视频信息服务管理规定》第十一条的规定[8]，进行形式判断，无需对技术的性质进行认定。只需要推定出技术服务人员对于其提供的技术帮助可能用于犯罪存在一定的概括故意或间接故意，就可以按照帮信罪对其定罪处罚，从而使得深度伪造中间环节的主体得以“帮助犯正犯化”，此种司法认定路径虽缺乏规范层面的直接依据，但客观上填补了伪造中间环节的追责空白。此外，根据《信息网络犯罪解释》第十二条[9]，帮信罪“情节严重”的认定还比较苛刻，因此对于利用人工智能深度伪造技术实施犯罪帮助行为的规制仍然有待提高，

[7]http://gongbao.court.gov.cn/Details/2aa922f042f1cfc097b56d8b61f686.html

[8]《网络音视频信息服务管理规定》第十一条：“网络音视频信息服务提供者和网络音视频信息服务使用者利用基于深度学习、虚拟现实等的新技术新应用制作、发布、传播非真实音视频信息的，应当以显著方式予以标识。网络音视频信息服务提供者和网络音视频信息服务使用者不得利用基于深度学习、虚拟现实等的新技术新应用制作、发布、传播虚假新闻信息。转载音视频新闻信息的，应当依法转载国家规定范围内的单位发布的音视频新闻信息。”

[9]《最高人民法院、最高人民检察院关于办理非法利用信息网络、帮助信息网络犯罪活动等刑事案件适用法律若干问题的解释》第十二条：明知他人利用信息网络实施犯罪，为其犯罪提供帮助，具有下列情形之一的，应当认定为刑法第二百八十七条之二第一款规定的“情节严重”：(一) 为三个以上对象提供帮助的；(二) 支付结算金额二十万元以上的；(三) 以投放广告等方式提供资金五万元以上的；(四) 违法所得一万元以上的；(五) 二年内曾因非法利用信息网络、帮助信息网络犯罪活动、危害计算机信息系统安全受过行政处罚，又帮助信息网络犯罪活动的；(六) 被帮助对象实施的犯罪造成严重后果的；(七) 其他情节严重的情形。

可以通过降低帮信罪的入罪门槛，提高公众警惕性与刑事制裁的威慑力[9]。

又以侵犯公民个人信息罪为例，虽然正如上文所说，我国对于公民个人信息的保护还不周延，但本罪在适用上仍有探讨的价值。我国《刑法》将本罪中的“侵犯”解释为“非法获取、出售或提供”，也正因此本罪的射程只能覆盖深度伪造的前端阶段。倘若能够将“非法使用”也纳入“侵犯”的语义当中，便有了将本罪的触手伸进伪造中间环节的正当性。此举不仅突破了证明伪造中间环节主体实施下游犯罪意图的困难，并且让“非法适用”和“非法获取、出售或提供”适用相同的法定刑，也恰恰体现了刑法的平等性和谦抑立场。

4. 结语

深度伪造刑法规制困境的根源在于，现行刑法规范未能适配生物识别信息的特殊敏感性与流水线网络犯罪分工模式，公开渠道获取生物素材、单纯伪造合成行为缺少规范约束，传统共犯、量化入罪标准难以适配此类新型犯罪形态。因此，要跳出单一完善后端罪名的思路，搭建分层全链条规制方案。首先要统一《个人信息保护法》与刑法规范逻辑，将生物识别信息划定为独立敏感信息类型，摒弃以信息条数为核心的入罪标准，补齐前端素材获取的规制空白。其次通过扩张解释，将非法使用生物信息实施深度伪造的行为纳入侵犯公民个人信息罪调整范围，填补中端合成环节追责漏洞。最后要细化帮信罪涉案工具司法推定规则，适度下调入罪门槛，厘清无直接意思联络技术服务主体的归责路径。

区别于笼统平衡技术发展与风险防控的既有研究，以技术中立划定刑法介入边界，不片面限制深度伪造技术研发，依托现有罪名解释优化实现全链条治理，推动刑法规制由事后追责转向全流程事前预防，化解人工智能生成技术创新与多重法益保护之间的规范矛盾。

人工智能技术的演进具有无限性，可以预见，深度伪造带来的刑法挑战也将呈现动态升级特征。这要求法律研究需突破传统教义学框架，强化技术认知与规范研究的交叉融合，推动刑法规制从“事后回应”向“事前预防–事中管控–事后惩治”的全流程转型。通过构建兼具适应性与前瞻性的刑法规制体系，实现法律对技术发展的规范引导与保障功能，最终达成技术创新与社会治理的法治协同。

参考文献

[1] 蔡士林. “深度伪造”的技术逻辑与法律变革[J]. 政法论丛, 2020(3): 131-140.

[2] 王德政. 针对生物识别信息的刑法保护: 现实境遇与完善路径——以四川“人脸识别案”为切入点[J]. 重庆大学学报(社会科学版), 2021, 27(2): 133-143.

[3] 刘双阳. 犯罪协作视域下破解人脸识别认证行为的刑事规制[J]. 河南财经政法大学学报, 2021, 36(4): 41-51.

[4] 刘涛. 网络帮助行为刑法不法归责模式——以功能主义为视角[J]. 政治与法律, 2020(3): 113-124.

[5] 冯明昱, 姜涛. “深度伪造”滥用行为的刑法规制[J]. 湖北社会科学, 2023(4): 127-135.

[6] 刘宪权. 人工智能时代深度伪造行为的刑法规制[J]. 政治与法律, 2025(11): 48-61.

[7] 李怀胜. 滥用个人生物识别信息的刑事制裁思路——以人工智能“深度伪造”为例[J]. 政法论坛, 2020, 38(4): 144-154.

[8] 孟醒. 深度伪造视听类电子数据的民事审查困境与应对[J]. 大连理工大学学报(社会科学版), 2026, 47(1): 86-95.

[9] 陈冉. 深度伪造涉性信息的刑法规制[J]. 法学, 2024(3): 76-90.