深度伪造技术全链式刑事治理模式研究

刘期湘*

内容提要 深度伪造技术活动涉及技术研发、技术运用、生成物传播的整个链条，任一阶段技术活动的失范均会对个人法益与集体法益产生侵害。深度伪造技术的刑法规制存在三重挑战：宏观价值维度存在技术自由发展与刑事风险控制的动态张力，自由刑法观或安全刑法观的单一导向易使治理僵化；中观对象维度呈现“重末端、轻前中端”的治理盲区，传统事后惩罚机制难以应对深度伪造的多重刑事风险；微观规则维度存在罪名缺位、归责模糊及量刑机械等问题，加剧了刑法规范与技术特性间的结构性鸿沟。对此，应确立深度伪造技术全链式治理理念，在技术研发与信息获取阶段实现精细化治理，在技术运用阶段对使用行为进行独立评价与分类处置，在生成物传播阶段加强后果治理与罪名优化。

关键词 深度伪造 刑事治理模式 全链式治理 罪名优化
DOI:10.14111/j.cnki.zgfx.2026.03.006

目 次

* 湖南工商大学法学院教授。本文系2021年度国家社科基金重大项目“数字经济的刑事安全风险防范体系建构研究”（项目批准号：21&ZD209）的阶段性成果。

一、引　言

当今社会，以人工智能等为代表的信息技术对人们的生产生活方式带来了颠覆性变革，不断地推动着人类向数字化生存模式迈进。深度伪造作为依托于“生成对抗网络”大模型（Generative Adversarial Networks，简称 GAN）的高阶人工智能技术，[1] 在契合人类数字化生产生活需求的同时，也遭受“双刃效应”的挑战。《互联网信息服务深度合成管理规定》第 23 条将深度伪造规定为“利用深度学习、虚拟现实等生成合成类算法制作文本、图像、音频、视频、虚拟场景等网络信息的技术”。近年来，基于深度伪造制作的虚假信息以假乱真，技术滥用的案例也不胜枚举，包括换脸知名人士的成人影片、合成不存在的记者报道、伪造国家政要声音发布的仇恨言论等。这些行为已然对不同类型的法益产生了侵害风险，部分具备了刑事可罚性，故应结合深度伪造的特点作出回应，以满足人工智能刑事治理的时代要求。结合实践案例与技术利用的潜在风险来看，深度伪造可能造成三方面的危害。其一，侵害个人合法权益，其中可能涉及侵犯公民个人信息罪、制作、传播淫秽物品牟利罪等。其二，妨害网络空间的稳定秩序，可能构成非法利用信息网络罪、破坏计算机信息系统罪等。其三，严重、不当的深度伪造行为还将威胁国家安全与发展。如 2024 年印度大选中，内政部部长阿米特·沙阿声称要结束针对数百万贫困和低种姓印度人平权政策的虚假视频，引发了社会极大的恐慌。由此可见，深度伪造并非仅与技术相关，其风险已在犯罪领域显现，应及时、合理惩治不当行为，以预防可能发生的法益侵害。

深度伪造行为是一个围绕着深度伪造技术全链条中多种行为的概念集合体，且由于深度伪造与深度合成的技术原理相同，本文并不作实质区分。从技术内核来看，深度伪造以个人信息的获取、修改、组合、生成与处理为基础，其中包含了各类技术、算法的运用，最终利用真假难辨的内容侵犯权利人的合法权益。具体而言，深度伪造涉及三个不同阶段的行为：前端行为，主要涉及深度伪造技术所必要的信息获取与技术研发行为，如深度学习、虚拟现实等生成合成类算法的供给；中端行为，表现为利用深度伪造技术生成、合成并输出内容的行为；末端行为，包括但不限于利用、传播深度伪造内容等行为。由此，深度伪造的刑事治理不应单向度地聚焦某一阶段或某一部分行为，而应摆脱“头痛医头、脚痛医脚”式的治理思路，运用体系化思维把握深度伪造行为的“全链条”特征，从而科学、有效地治理涉及深度伪造技术的相关行为。

二、深度伪造技术刑事治理的困境与挑战

深度伪造行为所产生的负面效应已然超出民法和行政法的有效调整范围，在刑法层

(1)　参见刘宪权：《人工智能时代深度伪造行为的刑法规制》，载《政治与法律》2025 年第 11 期，第 49 页。

面具备了规制的必要性与紧迫性。(2) 目前,学界相关研究聚焦于“如何规制传播和利用深度伪造内容的行为”,如传播深度伪造的涉性信息、虚假信息等。质言之,如何治理深度伪造行为三阶段中的“末端行为”已成为学术研究的重点。作为一项前沿技术,其所引发的社会风险并不局限于单一群体或单一领域,已逐渐渗透至个人隐私、社会秩序、国家安全乃至国际关系等多个层面。是故,刑法理论与司法实践都应充分关注深度伪造行为的全阶段。有鉴于此,系统化、体系化、全面化地解构深度伪造行为,厘清其各阶段行为所面临的挑战,乃刑法全面、科学、有效治理深度伪造的题中之意。具言之,应先从宏观治理价值的选择入手,再厘定中观治理对象的范畴,最后关注微观治理规则的完善,实现深度伪造问题探查与挑战梳理的点线面结合。

(一)宏观价值维度:技术自由发展与风险控制间的张力

20 世纪 80 年代,贝克根据人类生活在后工业时代所面临的困境以及生存状况提出了“风险社会理论”。(3) 作为一项存在风险的新兴技术,深度伪造所带来的社会风险以及合理分配相关风险的需求,与风险社会理论高度契合。(4) 在经济与安全的双重要求之下,技术的自由发展与风险控制的价值张力日益凸显。在消解两者张力和克服现代技术风险的问题上,应在何种范围内运用刑法予以应对,成为一项重要课题。(5) 技术发展与技术风险之间的张力在一定程度上催生了自由刑法观与安全刑法观两种刑法理念。

为防止刑法过度介入导致技术发展受阻,在以深度伪造为代表的技术治理规范体系中,强调刑法谦抑性和保障功能的自由刑法观成为当下技术治理的宏观价值指引,其旨在强调刑法应在技术中立原则下审慎介入技术创新领域,具体表现在立法、司法和政策三个维度。从立法维度看,我国目前没有与深度伪造有关的刑事规则,而只有最高人民法院、最高人民检察院、公安部印发的《关于依法惩治网络暴力违法犯罪的指导意见》(以下简称《网络暴力意见》)中涉及刑事规制内容;从司法维度看,中国裁判文书网公布的裁判文书中涉及“深度伪造”的裁判文书主要为民事案件;从政策维度看,国家发展改革委等部门《关于促进数据产业高质量发展的指导意见》的附件中明确要求支持深度合成服务算法迭代创新,以促进深度合成技术发展为重点方向,并未涉及相关限制性条款。在学术界,学者们针对技术治理的问题也普遍支持自由刑法观,排斥安全刑法观。有学者认为,若将安全刑法观所强调的秩序作为首要价值,使自由刑法观退居其后,可能导致国家刑罚权的正当化根基受到冲击,规制深度伪造行为的价值选择亦不外如是。(6) 笔者认为,引发学者担忧的原因在于,安全刑法观强调为了控制社会发展过程中出现的新风险,必须将刑法定义为保证社会稳定的工具,并通过对危险的禁止来实现安

(2) 参见熊波:《“深度伪造”的扩张化刑事治理风险及其限度》,载《安徽大学学报(哲学社会科学版)》2020 年第 6 期,第 108 页。

(3) 参见[德]乌尔里希·贝克:《风险社会》,何博闻译,译林出版社 2004 年版,第 7 页。

(4) 参见解志勇:《高风险人工智能的法律界定及规制》,载《中外法学》2025 年第 2 期,第 289 页。

(5) 参见吴宗宪、张进帅:《生成式人工智能:风险、挑战与刑事规制——以 ChatGPT 为例》,载《中国特色社会主义研究》2024 年第 3 期,第 71 页。

(6) 参见刘艳红:《中国刑法的发展方向:安全刑法抑或自由刑法》,载《政法论坛》2023 年第 2 期,第 66-67 页。

全,[7] 最终会促使刑法由“二次规范”演变成“一次规范”。

欧洲英、法、德等国在 20 世纪 90 年代出台《信息技术安全评价准则》后,[8] 长期重视安全刑法观,对深度伪造等技术的刑法扩张治理具有引导作用。无论是域外还是国内,对于深度伪造行为可能存在公民个人信息、网络空间秩序等法益侵害风险,已形成基本共识。因此,强调事前预防的安全刑法观应当更好地在深度伪造行为治理中发挥作用。特别是在面对人类社会数字化转型、技术革命阶段所产生的各种新风险时,安全刑法观作为总体国家安全观植入刑法后的内容与表征,在宏观价值层面对治理深度伪造行为发挥着极强的“指挥棒”作用。[9] 与科技发展所面临的双刃效应一致,自由刑法观与安全刑法观亦同样受制于双刃效应。若单纯选择其中一种刑法观作为深度伪造行为刑事治理的价值理念,不仅会导致治理理念的僵化,更会影响刑事治理的有效性。

具言之,自由刑法观之下的深度伪造治理更注重保障公民自由使用深度伪造技术,在此基础上,事后惩治运用深度伪造技术生成并传播伪造内容并造成实质危害的行为。但这较易忽视单纯的技术运用行为以及技术研发带来的风险。如,利用深度伪造技术生成淫秽视频,即使视频未传播,也可能构成制作淫秽物品牟利罪;深度伪造技术研发过程中不当的信息收集行为,涉嫌侵犯公民个人信息罪乃至非法获取国家秘密罪。因此,仅将深度伪造行为刑事治理的重心放在末端的利用、传播环节上,疏漏前端和中端的风险识别与阻断,会导致深度伪造的刑事风险控制显著失衡。在指引深度伪造行为刑事治理的过程中,自由刑法观过度注重个体自由可能会引发治理与实践脱节的现象,进而导致深度伪造冲破“制度的笼子”。然而,这并非直接表明应当秉持以风险预防为内核的安全刑法观,让刑法积极介入深度伪造的治理。应当肯定的是,刑法过早或者过多地介入社会生活,容易导致风险防控过度而限制公民权利或阻碍深度伪造产业的正常发展。因此,深度伪造行为刑事治理的价值选择,并非自由刑法观与安全刑法观的二元对立命题。

(二)中观对象维度:深度伪造技术刑事治理的阶段盲区

目前治理涉深度伪造案件的逻辑仍以“事后惩罚”为内核,强调“以个体法益保护为主线、以事后惩治为主张、以实害结果为主因”[10]。与此相印证的是,学界关于深度伪造行为所涉罪名的研究亦集中在侮辱诽谤罪[11]、诈骗罪[12]、传播虚假信息罪[13]等,聚焦于深度

(7) 参见[德]乌尔斯·金德霍伊泽尔:《安全刑法:风险社会的刑法危险》,刘国良编译,载《马克思主义与现实》2005 年第 3 期,第 38 页。

(8) 参见丁先存、王辉、段华洽:《论电子政务中信息安全及法律保护》,载《中国行政管理》2002 年第 10 期,第 11 页。

(9) 参见高铭暄、孙道萃:《总体国家安全观下的中国刑法之路》,载《东南大学学报(哲学社会科学版)》2021 年第 2 期,第 62 页。

(10) 莫晓宇:《人工智能的多元刑事治理:理论建构与路径选择》,载《厦门大学学报(哲学社会科学版)》2023 年第 4 期,第 54 页。

(11) 参见田宏杰:《人格权刑法保护:挑战与应对》,载《中国人民公安大学学报(社会科学版)》2023 年第 1 期,第 42-43 页。

(12) 参见敬力嘉:《作为行为不法类型的犯罪参与——兼论非法发布深度伪造信息的行为不法》,载《华东政法大学学报》2020 年第 6 期,第 87 页。

(13) 参见余永生、张昕、师胜男:《人工智能开发应用的伦理与法律协同治理研究》,载《科学管理研究》2022 年第 2 期,第 33 页。

伪造内容利用、传播等行为的治理。深度伪造行为链条中的末端行为成为实践中规制的重点和学术研究的焦点。然而,利用、传播深度伪造内容的相关行为仅是深度伪造行为链条中的最后一环。在治理对象上,缺少前端技术研发行为与中端技术运用行为,致使深度伪造行为的刑事治理存在阶段盲区。深度伪造囊括前端、中端、末端行为,刑法应充分、全面地治理各阶段的涉罪行为。

一方面,深度伪造技术前端研发和中端运用等涉罪行为产生的危害具有强隐蔽性,难以被立法者乃至社会公众识别,而只处罚造成实害结果的行为可能无法使刑法规范与技术迭代实现同频。例如,单纯开发可用于合成虚假视频的深度伪造软件,在尚未有人使用该软件合成并传播虚假视频的情况下,公众难以感知其社会危害性,且刑法尚未规制此类行为。但当不受限制的深度伪造行为泛滥时,会导致真假世界界限的崩塌,将社会推向失序的局面。(14) 即使当下深度伪造技术的运用还未完全渗透到各行业、各领域,其潜在的法益侵害风险还未彻底凸显,刑法也不应故步自封,而应当适时介入惩治那些在未来可能造成重大危害的行为。

另一方面,现有罪名适用门槛高、技术中立性抗辩掣肘等原因,也使刑法难以有效治理深度伪造技术研发与运用行为。传统刑法的结果本位主义与深度伪造技术的风险预防之间存在张力——若仅以发生实害结果作为处罚前提,将消解刑法风险预防的功能,进而导致深度伪造技术滥用风险向公共秩序、国家安全等集体法益领域渗透。因此,深度伪造行为的刑事可罚性应关注风险创设的盖然性与法益保护的周延性。例如,在深度伪造技术供给环节,专门为恐怖活动研发特定的深度伪造软件,尽管尚未造成实际损害结果,也可直接评价为准备实施恐怖活动罪。同理,在深度伪造技术运用阶段,若相关行为存在法益侵害风险,只要使用者明知技术运用行为可能导致法益侵害结果的发生,却未采取必要的风险防控措施,即使技术本身具有中立性,该行为仍应在刑法调整范围之内。然而,实践中的处理却恰恰相反。由于实害结果并未实际发生,刑法对深度伪造技术运用行为的治理难有作为。例如,运用深度伪造技术生成虚假身份信息的行为,可能导致这些信息被用于身份验证或身份溯源,若验证或溯源成功将导致“身份”所有者权益受到侵害。(15) 但当前只要未造成实质法益侵害结果,刑法就难以惩治此类行为。又如,虽然生成并传播深度伪造的淫秽视频构成传播淫秽物品罪,但单纯生成淫秽物品而没有传播的行为却不在刑法规制范围内,尽管单纯的生成行为同样创设了视频被传播的风险。

(三)微观实践维度:深度伪造技术刑事治理的规则不足

前文剖析了深度伪造刑事治理在宏观与中观维度所面临的问题与挑战,旨在廓清“依何治”的观念性问题和“治什么”的对象性问题。基于此,最终还需梳理出“怎么治”这一方法论层面所面临的问题与挑战。从治理对象层面可以看出,深度伪造行为刑事治理在对象上过于关注深度伪造内容的传播行为,忽视了其他阶段行为的治理。是故,“治

(14) 参见郝雨、文希:《AIGC时代多重虚拟幻境的生成逻辑与制衡机制》,载《未来传播》2025年第1期,第61页。
(15) 参见曾磊:《电信网络诈骗案件证据规则的多维检视与分层优化》,载《江汉论坛》2025年第10期,第139页。

什么”层面上存在的周延性问题在一定程度上加剧了“怎么治”层面上刑法规则供给不足的难题。

从立法层面来看，未有专门规制深度伪造行为的罪名或直接涉及深度伪造的罪名。梳理相关司法解释及其他规范性文件可知，目前仅有《网络暴力意见》中将“深度合成”（即深度伪造）作为实施网络暴力违法犯罪从重处罚的情节。这表明，一方面，现行罪名体系以传统犯罪行为为规制对象，缺乏针对深度伪造各阶段行为的专门罪名。例如，AI换脸、人脸伪造等技术研发行为虽未直接服务于后续犯罪行为，但在涉及商业欺诈、身份盗窃等领域时，即便存在技术滥用的高度风险和法益受损的潜在威胁，但由于研发行为缺乏对后续犯罪的主观“明知”，现行刑法中并没有单独罪名可以对这种单纯的研发行为进行追责。这种立法空白导致深度伪造行为链条中的技术研发环节长期游离于刑法规制之外，形成了“前端风险敞口”。另一方面，现行刑法对深度伪造行为的规制以个人法益保护为主，对集体法益保护不足。在集体法益层面，虚构领导人国际表态、伪造外交官员外交声明等行为对国家安全存在极大威胁，捏造金融消息、政策消息等行为损害了公共秩序法益。(16) 针对此类杜撰虚假信息的行为，现行刑法中能够适用的“编造虚假信息罪”等罪名，多以“严重扰乱社会秩序”等实害结果为要件，难以周延应对技术运用的潜在风险。

从司法层面来看，与深度伪造相关联的罪名存在量刑机械与失衡的现象。一方面，危害结果量化标准单一。例如，在利用侮辱、诽谤罪规制深度伪造行为时，对于“情节严重”的判断以实际点击、浏览次数为依据。然而，网络空间“机器人水军”的转发、点击行为并未在真实世界中传播伪造信息、损害他人的名誉，但因达到入罪门槛而可能面临刑罚。相反，一条伪造证券市场导向的视频可能因转发量未达入罪门槛而逃脱刑罚，但其引发的金融市场动荡却远非传统量化标准所能衡量。另一方面，传统的区分制共犯体系无法准确衡量深度伪造各阶段行为人的责任，导致量刑失衡。例如，即使深度伪造行为链条中技术研发行为的危害性高于传播行为，技术研发者通常也只能按帮助犯轻判，而传播者却因直接实施构成要件行为而成立正犯，面临重罚。更深层次的问题在于，现行刑事制裁体系缺乏妥当应对深度伪造行为的手段。例如，目前对涉及深度伪造技术的犯罪行为，刑法仅能依赖自由刑与财产刑事后惩罚，并未设置从业禁止、技术禁用等非刑罚的预防性措施遏制源头风险。

从理论层面来看，传统理论与技术特性间的鸿沟加剧了深度伪造行为刑事治理的困难，而针对性的解释路径尚未明确。一方面，主观归责面临技术研发与技术应用相分离的挑战。深度伪造技术研发者若未与下游犯罪者共谋，仅因技术被滥用或深度伪造内容被滥用，能否推定其具有故意，未有统一的法律解释。由此，司法实践中，法官对行为人主观心态的判断需依赖技术的专门用途作出间接推定。这可能造成因间接推定的灵活性导致同案不同判问题。例如，某开发者仅设计了可绕开生物识别系统的算法，由于其

(16) 参见尚海涛：《“深度伪造”法律规制的新范式与新体系》，载《河北法学》2023年第1期，第30页。

未与诈骗团伙合谋,对于该技术是否将用于犯罪不知情,也很难满足帮助信息网络犯罪活动罪等犯罪的主观要件认定标准,故而可能因此而脱责。另一方面,因果关系认定因容易扩散性变得复杂。深度伪造内容经平台算法推荐、用户二次转发后,原始传播人对伪造内容广泛传播这一后果的责任边界如何划定?传统“条件说”因果关系理论难以应对传播链的不可控性,例如,行为人客观上在网络空间发布了少量伪造信息,但主观上未预见会引发大量转发进而扰乱社会秩序。此时,如何追究其刑事责任存疑。

三、深度伪造技术全链式刑事治理的理念确立

为有效应对深度伪造行为刑事治理的三重挑战,应避免以“重末端、轻前中端”或“末端为主、前中端为辅”的理念来指引深度伪造行为的刑事治理。笔者认为,因深度伪造行为涉及多个不同环节,故应运用体系化思维,以全链式理念指引深度伪造行为的刑事治理。全链式治理理念最先应用于食品安全治理,其强调治理应从“单一环节”转向“多个环节”,由“末端治理”延展至“全链治理”,[17] 旨在系统地把握事物之间的复杂联系,以及问题产生的非线性关系,构建与治理对象复杂性相适应的全面且充分的治理结构。[18] 对于深度伪造而言,其涉及多个犯罪节点、多个犯罪中心以及多数犯罪人员,应在复杂涉罪行为中注重对每个阶段犯罪行为的直接治理、间接治理、溯源治理等效能,实施穿透性归责。[19] 将全链式治理理念引入深度伪造行为刑事治理中,不仅能够有效应对深度伪造行为“高度真实性、泛在普适性和快速演化性”[20] 等特征所带来的涉技术犯罪风险升高的叠加效应,同时也有助于构建实现涉及深度伪造行为的穿透式刑事治理。

(一)深度伪造涉及三阶段行为应当开展全链式治理

深度伪造刑事治理目前仍以规制末端传播行为为重心,表现为注重危害行为与结果的“后链治理”,忽视了其他阶段行为所引发的刑法风险。“后链治理”成为深度伪造行为刑事治理的主导理念,其核心原因在于对深度伪造行为风险的阶段性、层次性认识不清,导致刑法对前端技术研发行为、中端技术运用行为治理乏力。全链式刑事治理理念能够有效因应上述问题,通过平衡自由与安全实现深度伪造各阶段行为的穿透式治理。

从涉及深度伪造技术的各阶段行为来看,不同阶段行为所蕴含的风险共同构成了系统性风险链条。前端技术研发行为可能为下游犯罪提供工具,如研发用于伪造人脸识别的算法;中端技术运用行为可能直接威胁法益,如生成虚假身份信息或不实金融资讯;末端内容传播行为可能通过社交平台算法推荐等方式放大危害范围,使深度伪造内容产生实质性危害。实践中,已出现前、中、末端行为“集中发力”危害法益的情况。例如,

(17) 参见易开刚、范琳琳:《食品安全治理的理念变革与机制创新》,载《学术月刊》2014年第12期,第45页。
(18) 参见赵吉:《总体国家安全观下城市社会安全的全链条治理》,载《天津行政学院学报》2023年第1期,第19页。
(19) 参见夏伟:《网络有组织犯罪的组织进化与治理转型》,载《法学论坛》2024年第1期,第69页。
(20) 王禄生:《论“深度伪造”智能技术的一体化规制》,载《东方法学》2019年第6期,第62页。

2024年英国ARUP公司因深度伪造视频会议被骗2亿港元的案例[21],其根源在于前端深度伪造工具泛滥与中端使用行为失控,危害直接来源于作为末端深度伪造内容的“数字人”之欺骗。由此应认识到,深度伪造涵盖了前、中、末三阶段行为,虽然实害犯主要出现在末端的深度伪造内容利用和传播阶段,但在规制末端行为的同时,亦应穿透至前端技术研发行为,从根源上防治犯罪风险。全链式治理的核心逻辑在于穿透性归责与风险分层防控。穿透性归责强调要通过后端传播行为的治理穿透至前端技术研发或中端技术运用行为的治理,要求追溯行为链条的每个节点,识别并阻断风险传递路径。风险分层防控旨在准确把握深度伪造前、中、末三端的技术研发、利用等环节的法益侵犯可能性,科学调控刑法干预的程度。具体从法益保护层面而言,前端可多以预备行为和帮助行为作为规制对象,侧重于国家安全、社会秩序、公共利益等集体法益;中端与末端主要规制实行行为,在个人法益与集体法益间取得平衡,并通过对危险犯的设立进一步防范人身损害和财产损失。这一分层模式既避免了“一刀切”式的压制技术自由,又通过阶段性风险控制实现法益保护的周延性。

整体而言,深度伪造行为全链式刑事治理要求治理过程中运用穿透式思维,对各阶段行为辩证地开展风险预防或风险控制,使全链式治理理念科学、有效、全面地贯彻到深度伪造行为刑事治理中。应当注意的是,全链式治理理念不是要求对深度伪造行为刑事治理“眉毛胡子一把抓”,更不是叠床架屋式的机械治理,其核心在于不能忽视深度伪造行为链条中的任何一环。

(二)自由与安全的价值协同需以深度伪造行为全链式治理为载体

对深度伪造行为的刑事治理,无论是采取自由刑法观所倡导的以个体法益保护为导向,还是选择安全刑法观所主张的以技术风险控制为核心,都属于较为固化的刑法观念,仅能覆盖深度伪造行为刑事治理的单一面向。这既难以有效应对深度伪造行为在网络空间中的异化,更无法实现对深度伪造三阶段行为在刑法领域的科学治理。因此,只有在秉持全链式治理理念的前提下,方能化解技术自由发展与刑法风险控制的张力,促进两种价值的科学、有效协同。

深度伪造行为的治理需以自由刑法观与安全刑法观的动态互补为核心,结合深度伪造各阶段行为的差异实现价值协同。具体来看,在技术研发阶段,行为主要表现为算法设计与工具开发,其法益侵害风险尚处于抽象、潜在状态。在技术运用阶段,核心行为是利用深度伪造技术生成虚假视频、图像等特定内容,其风险体现为法益侵害的具体化与紧迫化。在深度伪造内容利用与传播阶段,传播、扩散深度伪造内容直接造成了法益侵害风险,可能引发不可逆的财产损失、人格侮辱等个人法益损害以及金融市场动荡和国家安全危机等集体法益损害。

由于深度伪造行为不限于对深度伪造内容的传播行为,其行为链条中的任一行为

(21) 参见梁思琦:《“2亿诈骗案”一年后,AI换脸犯罪更疯狂了》,载凤凰网,https://i.ifeng.com/c/8hgnScpxF2s,2026年4月26日访问。

均可能对法益造成不同程度的侵害，故刑法不能遵循单一的价值指引来治理每一阶段行为，否则将导致治理的机械性与失衡性，不仅可能影响技术发展，也可能放任技术滥用造成社会危害。例如，开发用于影视特效的人脸合成算法属于技术自由范畴，但当算法被设计为专门绕过安全认证机制时，若仍从自由刑法观视角看待技术研发行为，机械地贯彻技术中立立场，则可能因放任一些具有法益侵害风险的行为，从而危害社会。从法益侵害的紧迫性来看，中端深度伪造技术运用阶段所实施的不当行为相较于前端技术研发阶段，其法益侵害的紧迫性更为明显，故在技术运用阶段更应强调自由价值与安全价值的均衡。同样是利用深度伪造技术生成他人肖像的换脸视频，若用于影视创作则可通过民法予以规制，确保法律在规制技术应用过程中尊重技术自由；但若用于伪造他人身份在银行开户，就应通过刑法阻遏这一危害行为，此时，“技术自由”这一价值应退居“秩序安全”之后。

综上所述，由于深度伪造不同阶段的行为具有不同性质，因此刑事治理的价值选择针对不同行为时应当有所差异。不能为了鼓励深度伪造技术的创新与发展，就以自由刑法观全面统摄；亦不能为了应对技术所带来的威胁就一味强调安全刑法观，不当扩大刑法处罚范围。全链式治理理念旨在关涉深度伪造行为的每一环节，而自由与安全任一价值均无法妥当因应涉及多重行为的深度伪造。是故，为实现自由与安全的价值协同、化解两者张力，应根据不同阶段的行为能动调整，把握好深度伪造全链式治理理念。

（三）自由与安全的辩证统一乃深度伪造行为全链式治理理念之内核

深度伪造行为的刑事治理以自由与安全价值的动态平衡为核心，贯穿技术研发、运用与传播的全链条，其本质是对技术自由所蕴含的创新价值与法益保护所要求的风险防控之间的张力进行规范调和。换言之，自由与安全的价值协同以深度伪造行为的阶段性特征为基础，以不同阶段的具体行为为分析框架，通过动态调整自由与安全的权重，实现深度伪造行为刑事治理在价值层面的辩证统一。刑罚的目的归根结底在于保护法益。因此，对涉及深度伪造的不法行为科以刑罚的前提在于“是否可能造成法益侵害”。依据行为链条中“损害的紧迫性”以及“损害的程度”，技术研发阶段需强调技术自由，技术运用阶段需平衡创新与风险，内容利用与传播阶段则需优先注重法益恢复。

在技术研发阶段，自由刑法观占据主导地位，但其边界受限于客观归责理论中的风险创设标准。根据客观归责理论，刑法应当对“制造法所不容许的风险”进行归责，而技术中立性原则要求对研发行为的干预保持克制。(22) 例如，开发通用型人脸合成算法虽存在潜在滥用可能，但因技术功能具备广泛的合法用途，其风险尚处于抽象、潜在状态，属于技术自由范畴，即如罗克辛所言的，缺乏危险创设时归责的排除。(23) 应当明确的是，当深度伪造行为创设法所不允许的风险，即便这种行为与结果不具有自然的因果性，但在既有的刑事评价中具有独立的归因性时，刑法就应当介入。具言之，若相关技术极易

(22) 参见魏超：《刑法中的注意义务规范保护目的：运用范围与认定方法》，载《法学》2023年第6期，第35页。
(23) 参见［德］克劳斯·罗克辛：《德国刑法学总论》（第1卷），王世洲译，法律出版社2005年版，第248页。

被用于实施违法行为，那么技术研发行为则因创设“不被容许的风险”丧失中立性保护。由此可以明确，技术研发自由与法益保护的临界点在于，当技术功能与违法目的高度关联时，自由价值需让位于安全价值。技术自由须以符合社会共同价值为限，若算法研发已带有除促进科技发展外的不法目的，此时则需基于安全刑法观提前干预。

在技术运用阶段，核心矛盾在于技术合法使用与潜在滥用的交织。“根据法益侵害说，受侵犯的法益的重要程度不同以及受侵犯的程度不同，决定了不法程度的差异。”(24)由此，也可根据行为显著增加法益侵害的盖然性高低来判断刑法是否应当介入。应当肯定的是，若行为显著增加法益侵害的盖然性，且已达到“法不容许的风险”的程度，即使实害结果未发生，亦可归责。质言之，对于深度伪造技术的使用应在尊重个人技术使用自由和防范技术使用风险之间取得平衡。此时，自由与安全的价值协同体现为：刑法干预的限度应以行为本身与风险升高之间因果关系的判断为基准。

在内容传播阶段，深度伪造内容的扩散可能引发不可逆的集体法益或个人法益损害。此时，应以“安全刑法观绝对主导，自由刑法观退居其次”来实现刑法对深度伪造治理价值协同的辩证统一。以往观点认为，对于集体法益而言，只有当危险行为被持续且大量实施时，才能使集体法益的功能受到损害，即只有累积危险行为被普遍实施，才会真正地动摇集体法益的根基。(25) 基于此，有观点认为，是否侵犯集体法益（如公共秩序）可借鉴《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》（以下简称《网络诽谤解释》）中对于诽谤罪“情节严重”的量化定罪标准，即将“网络虚假信息被浏览超过 5000 次或转发超过 500 次”认定为侵犯集体法益。(26) 然而，此种以“点击量”“转发量”为核心的量化标准难以反映实质风险。例如，传播伪造央行降息政策的视频，虽因转发量未达入罪门槛而出罪，但可能严重扰乱金融秩序。此时，安全刑法观应当作为刑事治理深度伪造的首选价值，突破形式化标准，将此类行为直接以符合“严重扰乱社会秩序”要件定罪。互联网时代下的深度伪造内容传播独有的危害特点表现为风险的累积，刑法对其予以规制的目的在于防堵可能发生的“溃堤效应”，以避免发生难以修复的社会失序。因此，刑法在治理深度伪造内容传播行为时应防微杜渐，除法律另有规定外，应秉持安全刑法观绝对优先立场。

四、深度伪造技术全链式刑事治理的模式展开

深度伪造行为涉及技术研发、技术运用、内容利用与传播三阶段行为，为避免治理疏漏引发的阶段规制盲区问题，全链式治理模式可作为刑事治理深度伪造行为的选择。

(24) 张明楷：《具体犯罪保护法益的确定标准》，载《法学》2023 年第 12 期，第 82 页。

(25) 参见［日］关哲夫：《现代社会中法益论的课题》，王充译，载赵秉志主编：《刑法论丛》（第 12 卷），法律出版社 2007 年版，第 338 页。

(26) 参见章书海、雍自元：《编造、散布网络虚假信息型寻衅滋事罪构成要件研究》，载《法治文化》（第 2 卷），第 54 页。

（一）全链式治理应统筹法益侵害大小与风险升高程度

全链式治理理念所强调的全面治理，要求能够辩证地协同自由与安全在深度伪造治理中的价值选择。关于自由刑法与安全刑法，其二律背反关系本身并非不可调和，尤其是在贯彻全链式治理理念下的深度伪造行为刑事治理中。在总体国家安全观的语境下，刑法中安全与自由的平衡并非表现为简单的先后、强弱问题，其辩证统一的根本在于，国家为实现某一领域的安全可以在多大的程度上通过刑法对个体自由予以限制。(27) 以何种刑法观作为深度伪造治理的价值指引，核心在于确定刑法对涉及深度伪造行为的干预程度。深度伪造行为的刑事治理必须以刑法所认可的法益侵害作为判断标准，同时还需根据法益侵害程度决定是否应当由刑法进行治理，以及采用何种刑罚进行治理。这一问题的解决，应重点考虑深度伪造不同行为阶段自由与安全价值指引的动态平衡。此种价值动态平衡的判断标准在于法益位阶高低和风险紧迫程度。进一步而言，刑法在实质层面实现自由与安全的辩证统一，应在贯彻比例原则的前提下，确保手段必要性和利益均衡性，以免社会治理过度依赖或轻视刑法。

一方面，深度伪造行为中任一行为的实施一旦导致了法益侵害结果发生，或创设了刑法所不容许的法益侵害风险，刑法介入便具有了正当性，这是由刑法“惩罚犯罪，保障人权”的根本任务所决定的。“犯罪的本质是法益侵害，惩罚犯罪不仅要求不法，而且要求责任。”(28) 深度伪造行为自出现以来，已经在不同领域、不同阶段以各种形式产生了实害结果，其中涉及公民的名誉权与财产权、社会秩序以及国家安全等诸多法益，直接或者间接影响人的自由发展。刑法所保护的法益存在位阶区别，体现为：在责任形式相同的情形下，法定刑较重犯罪的保护法益，其位阶不应低于法定刑较轻犯罪的保护法益。因此，刑法在治理深度伪造时，应从涉及技术的三个不同阶段入手，根据情况的不同，秉持不同的治理理念和遵循不同的价值指引。具言之，以现行刑法为基础，严格区分法益位阶，对涉及深度伪造的不法行为以其所侵犯的法益位阶高低作为刑法干预程度的判断标准。应当注意的是，在深度伪造涉及的三阶段行为中，技术研发行为法益损害的位阶通常较低，深度伪造内容利用和传播行为法益损害的位阶最高，但在实际刑法评价时，应针对具体行为对法益损害予以客观分析。

另一方面，在考量法益位阶时，不应忽视尚未现实化的法益侵害风险。深度伪造行为中的任一行为引起了法不容许的紧迫风险，即使某些行为没有造成实害结果，也不能必然推导出刑法应对该行为保持绝对的谦抑。从对深度伪造行为的治理而言，刑法应当在危害结果发生之前提前介入，预防风险的现实化。预防刑法摒弃以实害结果为导向的追责模式，其核心在于对未来犯罪风险的防范和社会安全的关注，避免潜在的法益侵害危险，最终实现有效的社会控制。(29) 这表明，从安全预防的角度出发，实害结果的发生也并非刑法介入深度伪造行为治理的唯一评判标准，如果深度伪造的某些行为可能会导致

(27) 参见李翔：《总体国家安全观的刑法表达》，载《法学》2025年第3期，第81页。
(28) 郭旨龙：《论预备行为的风险评价》，载《中外法学》2024年第5期，第1304页。
(29) 参见何荣功：《预防刑法的扩张及其限度》，载《法学研究》2017年第4期，第138-139页。

重大个人法益遭受损害，即使没有造成具体的危害结果，只要创设了刑法不容许的风险，就需要刑事治理。(30) 当然，预防刑法的扩张须以“法益关联性”为限，避免脱离具体法益盲目入罪与抽象归责。

（二）前端应采取以风险预防为核心的审查式治理

在风险社会理论的基础上，风险刑法之罪刑结构的改变之一是由结果本位转向行为本位，即无需评判行为造成的实害结果，强化抽象危险犯的治理。(31) 在此框架下，技术研发的合法性成为刑法对深度伪造前端行为审查的重点，不具有合法性或未通过合法性审查的技术研发将面临危险的暴露。“现代社会的社会成员对于安全的欲求极为强烈，对于暴露的危险非常敏感，社会成员热切希望消除、减少这种高度、广泛的危险，热切希望在这种危险现实化之前，国家介入社会成员的生活来消除、减少这种危险。”(32) 这种期待在深度伪造刑法领域具象为“预防”治理思路。

对于前端深度伪造技术研发行为而言，以风险预防为核心的审查式治理模式强调通过技术研发的合法性审查，从源头阻断技术滥用的风险传递。质言之，应将技术研发合法性审查作为刑法介入的“前置过滤器”，将行业自律与法律强制相结合，实现技术促进创新、改善生活的目标。(33) 通用性技术，如开源人脸合成算法 StyleGAN，因具备影视特效、医学仿真等广泛合法用途，受技术中立性原则保护，可以直接通过技术研发合法性审查。对于专用性技术，如专用于伪造生物识别的 GAN 模型，则可能因其属于高风险 AI 技术，难以通过技术研发合法性审查。若有绕过技术研发合法性审查的行为，即使不属于专用于伪造生物识别的 GAN 模型，也可推定该项技术的研发行为已创设“法所不容许的风险”，具有被刑法规制的合理性。对于特定技术的研发，如行为人意图开发深度伪造技术进行合法的国际商业行为，但该技术研发需进入相应的网站后台采集特定的国家事务信息，此时，该行为就已具有高度具体的危险。为预防这一风险现实化，应对非法进入国家事务信息网站后台采集信息的研发行为进行规制，评价为非法侵入计算机信息系统罪和非法获取计算机信息系统数据罪。前端技术研发行为产生的法所不容许的风险是对社会秩序等集体法益的侵犯。无论是通过技术研发合法性审查防范风险，还是刑法直面技术研发行为，之所以注重保护此类集体法益，是因为集体法益作为阻挡层法益，是个人法益的屏障，(34) 因而保护集体法益终归是在保护重要个人法益。此种底层逻辑的本质体现为自由刑法观与安全刑法观的张力。当安全刑法观成为指引深度伪造全链式治理的统摄价值时，风险自然较低。换言之，“在一个社会系统中，越是强调内在的秩序和安全（确定性），风险发生的可能性就越小”(35)。应当注意的是，由于我国刑法并未完全实

(30) 参见王良顺：《预防刑法的合理性及限度》，载《法商研究》2019 年第 6 期，第 52 页。
(31) 参见姜涛：《为风险刑法辩护》，载《当代法学》2021 年第 2 期，第 93 页。
(32) 前注 ，关哲夫文，第 338 页。
(33) 参见黎四奇：《数据科技伦理法律化问题探究》，载《中国法学》2022 年第 4 期，第 126 页。
(34) 参见蓝学友：《规制抽象危险犯的新路径：双层法益与比例原则的融合》，载《法学研究》2019 年第 6 期，第 134 页。
(35) 姜涛：《社会风险的刑法调控及其模式改造》，载《中国社会科学》2019 年第 7 期，第 117 页。

现自由刑法观的主导与统摄，要避免在安全刑法的挤压下造成对集体法益概念的误读以致适用上的泛化，使得其恰恰成为刑罚不当扩张的推手。(36) 加之深度伪造作为前沿的人工智能技术，从保障技术发展角度出发，自由价值在技术研发层面应优位于安全价值，前端的风险预防全链式治理应重点关注技术研发合法性审查。

（三）中端与末端应实施以风险控制为核心的链条式治理

涉及深度伪造中端与末端的行为，应实现“风险预防—风险控制—溯源追责”的链条式治理，既防范新风险的生成，又遏制传统风险的扩散，并通过穿透性机制追溯至前端技术研发或中端技术运用的责任源头。在中端技术运用阶段，追求自由与安全的动态平衡，其核心在于结合具体场景与风险等级灵活应对；在末端通过技术运用生成深度伪造内容并对其传播或利用阶段，强调安全刑法为主导的刑事治理价值指引，本质上是追求严格的风险控制与实害惩治。由此，在对中端和末端行为实施穿透性治理时，也应在刑法谦抑性的基础上考量规制和穿透的必要性。

具体来看，对于中端的技术运用行为，既要防范风险向实害转化，又要通过溯源机制追溯至前端技术研发的责任源头，形成“风险控制—责任穿透”的双重治理效能。根据理查德·塞勒和卡斯·桑斯坦提出的“助推理论”，无需采用禁止或明显的刺激方式，而在人们无意识的情况下通过对规则的设计来引导人们改变行为。(37) 在刑法规制深度伪造中端的技术运用行为时，应通过精细化规范设计引导技术运用行为合法合规，而非简单禁止。质言之，在中端技术运用阶段，需结合场景差异动态平衡自由与安全。在风险控制层面，若行为人完全出于正当理由，即合法地使用深度伪造技术，如生成虚拟教学视频与人物追思视频，自由刑法观应占据主要地位。相反，若行为人使用深度伪造技术大量生成虚假身份，此时抽象危险犯应成为刑法规制的“利器”。对于末端深度伪造内容利用和传播行为，除通过传统的诈骗类（如诈骗罪、招摇撞骗罪等）和传播类（如故意传播虚假恐怖信息罪、故意传播虚假信息罪、传播淫秽物品牟利罪等）犯罪予以规制外，其需以风险控制为治理核心，通过实质化危害评价遏制虚假内容的扩散。同样以网络诽谤犯罪为例，两高在《网络诽谤解释》中对传播行为的规制主要依赖“点击量”“转发量”等量化指标，但深度伪造内容因算法推荐、自动化传播等特性，其危害往往远超形式化标准的覆盖范围。为此，风险控制不应陷入“机械量化逻辑”之泥淖，应回归行为本身，以法益损害为判断起点进行实质评价，即以行为对法益的实质威胁程度作为入罪标准。质言之，名义上的“点击量”“转发量”不一定能反映诽谤内容扩散的实际情况，还需考虑深度伪造内容是否成为算法推荐、自动化播放的内容等其他情形。

对于中端和末端行为的治理，风险控制的重点是穿透性归责。穿透性治理的核心在于通过技术手段与法律机制追溯责任链条。例如，行为人为了利用深度伪造技术制作并传播淫秽物品，在公开网站上获取他人的人脸、身体等信息，但未向他人出售或者提供，

(36) 参见马春晓：《现代刑法的法益观：法益二元论的提倡》，载《环球法律评论》2019年第6期，第145页。

(37) 参见［美］理查德·塞勒、卡斯·桑斯坦：《助推（终极版）》，姜智勇译，中信出版社2023年版，第289页。

这种情况并不满足侵犯公民个人信息罪的“出售”或“提供”要件,且获取信息的方式也是合法的。从犯罪构成层面来看,若行为已经具备了利用深度伪造技术制作淫秽物品的主观故意与能力,便应当延伸性认定该行为构成传播淫秽物品罪的预备。即使没有淫秽物品存在,相应的技术能力已经使其具备了传播淫秽物品的可能性。应当注意的是,刑法处罚预备行为应采取审慎克制态度,必须是预备行为确实创设法益侵害的危险时,才能进行处罚。

整体而言,深度伪造行为全链式治理需构建“前端预防—中端平衡—末端控制”的溯源穿透的链条式治理。前端应围绕技术研发合法性审查做好风险预防,中端通过动态归责逻辑区分合法与非法使用场景,末端则应重点强调实质性风险控制。穿透性治理应把握限度,遵循妥当性、必要性、均衡性等限制性原则,防止刑法介入的过度化、超前化与泛化。

五、深度伪造技术全链式刑事治理的规则构建

由于我国目前刑事治理深度伪造过于注重末端环节的制作与传播相关深度伪造内容的行为,导致在治理效能上出现了“重结果、轻过程”的问题。但这并非表明我国对于深度伪造内容的传播、利用等行为的治理已臻完善。在优化具体治理路径层面,仍要贯彻全链式治理模式,针对各个可能涉罪领域深入分析,弥补既有治理模式中的阶段性缺憾和治理效能不足。诚然,从规则层面完善深度伪造的刑事治理,无外乎是从解释学或立法论两维度展开。由于深度伪造的特殊性,解释论视角的确可为深度伪造刑事治理贡献力量,但多维度直面深度伪造的现实和未来问题,则更能确保治理的周延性与科学性。应当注意的是,若要对深度伪造涉及的三阶段行为予以入罪考虑,应检视该入罪思路是否遵循刑法规制的妥当性、必要性、均衡性原则,以确保刑事治理深度伪造的科学性无虞。

(一)推动前端技术研发中信息获取与算法研发的精细化规制

信息获取与算法研发是深度伪造行为的源头。有学者指出,深度伪造具有技术中立性,使用该技术生成图片、音像、视频已经成为一种新兴产业,法律不应禁止深度伪造技术的使用。[38] 尽管纯粹的自由刑法观秉持技术中立性原则,但从学理分析和现实情况来看,深度伪造的技术开发也有超过容许限度的可能。如果仅将自由主义刑法观作为信息获取与技术研发的治理的主导价值,可能会出现该技术刑法规制的“真空地带”。刑法作为最严厉的法律,需要对深度伪造这种新技术出现的新问题及时做出适度应对。深度伪造技术研发行为不仅可能在信息收集环节单独构罪,也可能构成其他犯罪的预备行为或帮助行为。此外,我国在网络信息犯罪治理中已存在可适用于深度伪造技术的预备行为实行化和帮助行为正犯化的立法例,表明可以通过增设相关罪名解决深度伪造前端部分行为的规制问题,只是仍有精细化的完善空间。

(38) 参见王利明:《〈民法典〉人格权编的实施与未来展望》,载《中国法学》2026年第1期,第130页。

深度伪造技术得以切实运用的前提,在于相关人员的脸部特征、姓名、性别等具有特定人格识别性信息的获取,故而相关身份信息获取行为是否合法成为治理的重点。从类型上看,身份信息获取有合法获取和非法获取两种,理论上将不合法获取的情况称为身份盗窃,(39) 现实中还存在获取方式合法但非法使用这一情况。上述分析虽然未超出既有刑法体系,但并非表明现行刑法体系已经能够涵盖全部身份盗窃与身份冒用的行为。以身份冒用为例,行为人获取了他人的身份信息后意图实施诈骗行为,若仅停留在身份信息获取或尚未明确实施诈骗阶段,此时,控方需证明被告人持有他人身份信息且有诈骗故意,才会构成诈骗罪预备或者未遂。从各国立法来看,在诈骗行为未实施前证明行为人存在诈骗之故意的难度系数非常高,这容易导致因证明困难而造成定罪上的障碍。对于身份冒用和身份盗窃是否需要单独入罪的问题,首要考虑的是行为所导致的风险升高是否达到法所不容许的状态且有新的法益需要保护。诚然,若仅从上述案例来看,可能会认为既然无法证明诈骗的故意本就不应通过诈骗罪予以规制,以侵犯公民个人信息罪定罪更为合适。但是,此种观点没有考虑到数字时代深度伪造行为涉罪的特殊性。无论是身份盗窃抑或身份冒用,均存在“非法获取并冒充他人”这一环节,也均造成了法益侵害危险。需要注意的是,对于深度伪造的前端行为规制虽应秉持安全刑法观的治理理念,但不能为了保护被害人的权益就加重国民面对国家刑法约束的负担,避免刑罚范围扩大而导致人民的权利遭到侵害。由此,对于身份冒用和身份盗窃入罪应设置相应的严格条件,例如身份信息获取的数量、不法使用的情形,不能将一次获取以及轻微不法使用等情形入罪。

而对于深度伪造的前端算法研发行为,应兼顾算法类型与研发意图。对于专用于生成违法内容的算法研发,例如涉性换脸AI,若只将其作为传播淫秽物品罪的预备行为,由于我国对预备犯罪的打击力度较轻,难以对纯粹的研发行为予以有效规制,结果可能导致大量的涉性换脸AI技术在网络中传播。由此引发的连锁反应则是大量的涉性深度伪造视频流传于网络。因此,应以安全刑法观为理念引导,遵循“预备行为实行化”思路,在不影响技术创新发展的前提下,将开发用于制作涉性换脸视频的深度伪造算法之行为入罪。同时,还可增加相关出罪事由,如用于教学或自行观看等。此种“预备行为实行化”思路是以算法类型与设计目的之直接违法性为核心入罪要素,无需等待实际危害结果发生。此外,还应当注意,为了避免深度伪造算法研发规制的泛罪化,须将算法类型和研发意图限定在国家安全、涉性信息、公共秩序等领域,谨防以技术安全为由损害技术和产业的发展与创新。相反,对于具备合法用途的通用算法,只要其履行了技术研发安全审查义务,则可成为出罪事由。但若深度伪造算法设计者故意不履行技术研发安全审查义务且造成严重后果,则可考虑依据帮助信息网络犯罪活动罪等罪名予以追责。

(39) 参见陈玲:《身份识别信息“使用”的界定纷争——基于美国加重身份盗窃犯罪的判例考察》,载《政治与法律》2024年第1期,第160页。

（二）明确中端技术使用行为的独立评价与分类处置

基于自由刑法观与安全刑法观在中端层面的能动协同，深度伪造技术使用行为需突破传统“预备—实行”的线性思维，根据生成内容性质与使用场景进行独立评价。对于一般意义上的技术使用行为，若仅生成无害娱乐内容或教学视频，则刑法应保持谦抑，当出现侵权等其他违法行为时，可通过民事或行政手段解决纠纷。例如，用户未经授权使用明星肖像制作搞笑视频，虽然广义上可能存在身份冒用之嫌疑，但个别冒用或者轻微违法使用的行为可通过《民法典》肖像权条款追责，无需刑法介入。进一步来看，若技术使用涉及特殊敏感内容，如生成虚假影像或与国家安全相关虚假信息，即便尚未传播，其行为本身已具备实质危险。

考察世界各国深度伪造相关立法可以发现，对深度伪造技术使用行为的独立评价具有一定的可行性。目前，德国秉持无特别立法必要之立场。美国在《深度伪造责任法案》《禁止恶意深度伪造法案》等法律中明确对深度伪造使用行为予以规制。(40) 韩国等国家因涉性深度伪造行为造成了巨大危害，针对此类行为在刑法层面予以特别规制。(41) 相关立法实践中，为了将调整范围限缩至深度伪造领域，则须限定为以“计算机合成或其他科技方法”实施。对于“制作”行为独立评价的必要性主要是基于人格尊严与隐私保护，避免制作行为给被害人造成难堪、恐惧等身心创伤。并且，通过深度伪造技术将被害人的脸部移接于他人生成涉性视频，并不以造成与性隐私有关的人格权侵害为要件，只要具有相关不法“意图”且“足以损害他人”即可，属于典型的“抽象危险犯”。(42)

我国关于深度伪造的刑事治理亦可借鉴上述经验，适度拓宽立法领域，将此种规制深度伪造技术使用行为的立法模式扩展至国家安全、公共安全等领域。在国家安全和公共安全领域的技术使用行为的危险性尤为凸显。例如，利用深度伪造生成虚假的军事调动指令或领导人外交声明，即便尚未扩散、传播或用于其他使用行为，也可能破坏国家决策系统的可信度，甚至具有对国家安全的抽象危险。是以，在单独设立罪名时，可以适当考虑不以“造成公共秩序混乱”“危害国家安全”为要件，径行入罪。但应当注意的是，这种扩张性治理需坚守刑法谦抑性原则，如果使用深度伪造形成的影音或视频未危及国家或社会公众安全，刑事治理不宜过度介入。如涉性类视频，其目的仅是自行观看，而非实施传播等其他扩散类利用行为，则不该当构成要件，亦不应遭受刑罚。

（三）末端深度伪造内容利用行为的后果强化与罪名优化

利用深度伪造技术产生虚假内容，是深度伪造技术对国家、社会以及个人造成法益侵害的最终环节。深度伪造作为依托于计算机、互联网、大数据而问世的生成式人工智能技术，其造成的危害与未涉及科技手段的犯罪相比，结果形成更迅速、影响更广泛、程

(40) 参见谢波、曹亚男：《深度伪造技术犯罪风险的典型样态、生成逻辑及治理路径》，载《中国人民警察大学学报》2026年第1期，第42页。

(41) 参见王刚：《韩国修法严惩“深度伪造”犯罪》，载《法治日报》2024年10月21日，第6版。

(42) 参见陈俊伟：《重思深度伪造影音的入罪化——以人工智能的科技风险规制为思考方向》，载《月旦法学杂志》2023年第2期，第24页。

度更深刻。在德国,有学者主张应当考虑基本权利保障,对于具有言论自由、艺术自由等正当理由的内容不应禁止。(43) 易言之,单纯运用深度伪造等生成式人工智能并不会产生不法,即使存在“不法”,其来源只能是深度伪造技术被滥用后所生成的内容,故应针对生成内容后的相关行为予以治理。虽然德国此种立场对于我国目前刑法规制深度伪造而言较为保守,但不可否认的是,其也指出刑法对于深度伪造治理的重点在于生成内容的利用行为,故在针对某些利用型犯罪时应考虑调整其相应的加重情节与入罪门槛,或者构建专门的罪名予以规制。

一方面,与传统刑事治理中的伪造、编造物不同,深度伪造生成物最大的特点为高度真实性,几乎到了难辨真假的程度,故社会公众更易遭受其欺骗。随着科技日新月异,诈骗者利用深度伪造技术制作虚假的音视频等内容,用以诈骗社会公众,其手段恶劣且难识别,加之虚假音视频依托于互联网传播,其有导致不特定人或多数人受诈骗之可能,进而对社会造成严重影响,确有加重处罚之必要。为此,对于利用深度伪造实施诈骗犯罪,也可考虑增设相关量刑条款,将“利用深度伪造实施诈骗的”作为相关犯罪的加重情节,防止深度伪造生成物过度泛滥导致诈骗问题频发。并且,将使用深度伪造技术实施犯罪作为加重情节的不应局限于诈骗类犯罪中,对于影响民主、选举、颠覆国家政权等涉及国家安全和社会安全的犯罪中,亦应增设相应的加重情节。与此对应,我国也应修正“积量构罪”(44)标准,排除机器人水军刷量等未实际降低被伪造者社会评价的虚假传播数据的影响,仅以真实点击、转发量作为入罪依据,防止归责范围的不当扩大。

另一方面,对于传播深度伪造内容危害国家安全、经济安全等集体法益或超个人法益的行为,应突破现有罪名框架单独设罪。相较于传统伪造或者变造行为,深度伪造技术风险高、传播速率快,对于各种法益的侵害范围与程度更大也更容易,域外许多国家及地区已经将以营利为目的,传播、散布、公然陈列深度伪造内容的行为进行单独入罪。尤其是涉性犯罪问题上,韩国与我国台湾地区均将意图非法使用深度伪造技术生成涉性视频的行为通过刑法独立进行评价。(45) 基于举轻以明重的逻辑,我国可以考虑对于损害重大集体法益的行为增设情节,更可单独设罪。如参考《刑法》第105条第2款“煽动颠覆国家政权罪”,以及第291条之一规定的“编造、故意传播虚假恐怖信息罪”和“编造、故意传播虚假信息罪”的立法技术,增设“传播虚假国家安全信息罪”,将“利用深度伪造技术制作、传播伪造的国家安全决策信息”作为抽象危险犯处理,无需证明实际危害结果。

整体而言,目前我国刑法体系对于深度伪造内容利用行为的规制可在加重情节和罪名单立两个层面予以完善,实现对深度伪造行为的精准打击。同时,也应避免象征性立法或象征性情节增设,确保新增罪名和情节的明确性与必要性,并通过动态解释适应技术迭代的挑战,最终实现数字时代法治的精准化与人性化。

---

(43) 参见维克托莉亚·克雷齐格:《针对人工智能生成影像侵权的保护》,邓环宇译,载《南大法学》2025年第2期,第16页。

(44) 参见皮勇:《论新型网络犯罪立法及其适用》,载《中国社会科学》2018年第10期,第135页。

(45) 参见陈冉:《深度伪造涉性信息的刑法规制》,载《法学》2024年第3期,第77、79页。

六、结　语

在数字化浪潮汹涌澎湃的今天，深度伪造技术作为一项前沿的人工智能技术，正如一枚硬币的两面，在人们享受技术带来“数字红利”的同时，深度伪造技术的不当应用也带来了诸多负面影响，从个人隐私的侵犯到社会秩序的扰乱，乃至国家安全的威胁，都迫使刑法必须正视其潜在的风险。未来，深度伪造技术的不断发展将带来更加复杂和多元的刑法风险。因此，需秉持深度伪造行为治理的“全链式理念”，不断完善刑事治理体系和治理规则。深度伪造技术应用范围的宽度、向度、广度难以预料，应在着眼未来的基础上，重点关注深度伪造技术在其他领域应用所带来的风险。例如，在司法领域，深度伪造技术可能被用于伪造证据、干扰司法公正等，这些行为在未来的研究中应给予足够的重视和关注。一言以蔽之，深度伪造行为的刑事治理是一项长期而艰巨的任务，应以开放、包容、审慎的态度面对这一挑战，为数字时代的法治建设贡献智慧和力量。

Abstract: The utilization of deepfake technology encompasses the entire continuum of research and development (R&D), implementation, and propagation of generated content. Any misconduct throughout this continuum has the potential to infringe upon legally protected interests of individuals and collectives alike. The regulation of deepfake technology through criminal law is confronted with three principal challenges. On a macroscopic scale, there exists a dynamic tension between the freedom of technology and the control of criminal risks. Adherence to either a liberal or a security-centric criminal law paradigm may result in governance becoming rigid. On a mesoscopic level, a governance deficiency emerges due to an overemphasis on downstream activities, which is to the detriment of upstream and midstream phases. Traditional punitive mechanisms, which are reactive in nature, find it challenging to effectively address the multifaceted criminal risks presented by deepfake technology. On a microscopic level, the lack of specific offenses, the indistinct attribution of responsibility, and the rigid sentencing practices further exacerbate the structural disparity between criminal law principles and the nuances of technology. In response, a holistic legal regulatory framework should be implemented, which includes: meticulous governance during the R&D and information acquisition phases; independent assessment and differentiated management of usage behaviors during the implementation phase; and improved consequence management coupled with offense-specific optimization during the application and propagation phases.

（责任编辑：王　楷）