

DOI:10.16104/j.issn.1673-1883.2026.02.008

“AI 魔改”视频的版权问题审视 及治理路径

张艺博, 陈尚贤

摘要: 生成式人工智能技术的普及, 催生了以“AI 魔改”视频为代表的二次创作热潮, 在广泛传播的同时, 带来了显著的版权风险与治理挑战。文章沿“AI 魔改”视频创作的逻辑脉络展开规范分析, 探讨创作前“魔改”提示词的法律定性及训练数据的版权合规困境; 系统分析创作中该类视频的法律属性及对改编权、复制权等多重著作权的侵害风险, 并对其“合理使用”抗辩进行严格审视; 指出创作后平台应承担更高的注意义务, 实现从被动应对到主动防范的转变。为有效治理“魔改”等容易引发版权争议的二次创作乱象, 还需进一步推动法律规则调适、行业标准共建、授权机制创新与公众教育等多维度协同, 构建法律、技术与共治相结合的体系化治理路径, 在版权保护、技术创新和文化遗产之间寻求可持续的平衡。

关键词: “AI 魔改”视频; 二次创作; AI 提示词; 著作权侵权; 版权问题

中图分类号: D923.41 **文献标志码:** A **文章编号:** 1673-1883(2026)02-0075-12

收稿日期: 2026-01-10

作者简介: 张艺博 (2000—), 男, 安徽淮北人, 北京市律师法学研究会助理研究员, 中国人民大学法学院硕士研究生, 研究方向: 民商法, E-mail: wakaiz992@ruc.edu.cn; 陈尚贤 (2000—), 女, 安徽淮北人, 北京市海淀区人民法院民事审判二庭法官助理, 硕士, 研究方向: 民商法。

一、引言

随着生成式人工智能技术的快速发展, 尤其是 Sora、可灵 AI 等文生视频模型的应用普及, 视听内容的创作门槛大大降低, 视频二次创作 (以下简称“二创”) 行为变得异常活跃。其中, 对《红楼梦》《甄嬛传》等经典影视作品进行恶搞和娱乐化改编的“AI 魔改”视频, 在各大平台广泛传播, 成为当前备受关注的文化现象。所谓“AI 魔改”视频, 一般是指用户利用 AI 视频生成工具, 对影视作品中的角色形象、故事情节、台词对话进行颠覆性的改编重构, 制作出像“林黛玉倒拔垂杨柳”“甄嬛修仙大战”等十分荒唐的内容。这些视频凭借强烈的反差感和猎奇性, 在吸引了大量流量、为创作者带来可观收益的同时,

也助长了以恶搞、鬼畜为目的的创作风气,其背后更隐藏着复杂的法律风险和治理难题。对此,国家广播电视总局在2024年底发布《管理提示(AI魔改)》,并于2026年1月开展为期一个月的专项治理,体现出监管部门对该乱象损害经典文化、扰乱网络秩序的高度重视。表面上看,“AI魔改”视频是技术普及带来的创作自由,但实为AIGC时代的二创失序,暴露了从创作到传播过程中的一系列法律问题:生成视频的提示词性质如何认定?“AI魔改”视频是否构成侵权?平台应如何从被动的传播渠道,转变为更加主动承担责任的主体?这些问题相互关联,是一个亟须系统回应的法治课题。

二、创作前的法律定性:“魔改”提示词与训练数据

为使生成的视频呈现出某种荒诞或反差效果,用户输入的“魔改”提示词通常并非简单的操作指令,其文本需要包含对原作品角色、情节、场景乃至风格的具体化拆解与重组,实质上成为一种驱动AI进行特定内容生成的创作纲要。实践中,部分精心设计的提示词能稳定产出特定风格的生成物,由此促成了初具规模的分享社群与交易市场。这意味着,提示词可作为一种新型数字资产,在蕴含一定经济价值的同时,相关法律确权问题亦不容忽视^[1]。因此,准确界定“魔改”提示词的法律性质,是对后续生成内容进行侵权认定的前提与基础。此外,训练数据的合规问题也应重点关注,无论是开发方在模型的训练阶段使用数据,还是用户方自行上传数据对模型的输出进行训练或调试,均存在潜在的侵权风险。

(一)“魔改”提示词的可版权性分析

关于AI提示词能否成为受著作权法保护客体的讨论,理论界存在不同的声音,但基本围绕“思想与表达二分法”这一传统框架展开。支持派主张,判断的核心应聚焦于“人的贡献”,即关注用户通过关键词选择、结构编排所投入的智力劳动与审美判断,而非生成工具的角色或表达形式本身^[2],具备高度独创性的提示词撰写行为本身可被视为一种创作。反对派则认为,AI提示词只是表示了对生成结果的需求,没有运用“作品语言”,更符合思想或方法的提出,不是著作权意义上的表达行为。从政策层面看,生成式AI技术的发展与使用已具备内在动力,并无必要通过扩张版权保护范围给予创作激励^[3]。

国内外司法实践对此亦持审慎态度。美国版权局在官方文件中指出,像Midjourney这类模型并不将提示词视为创建特定结果的具体指令,而是将其解构为统计意义上的参数标记^[4]。在Théâtre D'opéra Spatial案中,尽管用户输入了数百次文本提示并进行了后期处理,美国版权局仍认为其不构成作品,理由是生成结果最终取决于系统对提示词的处理方式,而非用户的具体表达意图^[5]。简言之,美国倾向于将提示词视为技术性的要素罗列或触发参数,因为其缺乏人类作者对生成内容的充分控制与创造性贡献,故不予版权保护。在我国的司法实践中,有法院明确认可了原告在设计角色、组织提示词、调整参数及筛选图像等流程中所付出的多种智力投入,并将提示词本身作为判断最终生成内容是否具有独创性的重要考量因素^①。尤其是在“上海首例提示词案”中,法院说理部分聚焦于提示词文本,认为其如果仅对画面元素、风格进行基础的常规性描述,则本质上属于不受保护的“思想”;若提示词经过有目的的构思和语义设计,能体现一定的形态细节、风格特点、情感氛围,呈现出独特的审美取向,则可能构成受保护的“表达”^②。我国司法机关主要立足于著作权法的文本与构成要件,严格审视提示词是否符合独创性表达的判断标准。就“AI魔改”视频的提示词来看,大多数是为达成特定荒诞效果而对原作品要素进行的直接搬运与组合。例如,“让林黛玉倒拔垂杨柳”的指令本质上是将原作角色与另一个文化符号进行拼接,

并未在语言表达层面形成具有审美意义的独创性文本,难以满足著作权法对“表达”的要求。退一步来说,即便存在少数经过精心设计、在语言层面具备独创性表达的提示词,也仍需谨慎考量其受保护的边界。“魔改”提示词的核心功能指向对原作品的改编,它的价值和意义,系于其所生成的视频内容,而非作为独立的文本存在。换言之,“魔改”提示词的法律地位在事实上从属于其所驱动的生成物,当生成内容本身容易引发侵权争议时,单独保护驱动它的工具,在逻辑上是矛盾的。更何况,“魔改”的生成结果大多与文化遗产的严肃性相悖,本身便缺乏著作权法所鼓励的“有益于社会主义文化创新与传播”的正向价值。在此前提下,就算个别提示词在形式上具有一定的创意,法律也缺乏对其提供独立版权保护的必要。否定“魔改”提示词的可版权性,并非否认其可能存在的智力投入,而是基于著作权法的目的考量,将有限的司法资源与法律保护聚焦于那些真正贡献于文化创新的创作,将治理焦点集中于最终侵权内容的认定与平台责任的落实,这或许是更有效的路径。

(二)“魔改”训练数据的合规问题分析

“AI魔改”视频的生成不仅依赖于精心编排的提示词,更离不开对其背后模型的数据训练。而这一训练过程通常需要使用影视作品片段作为学习素材,进而引发了数据版权的合规困境。首先,生成式AI的训练需要数以百万计的作品,逐一获取授权在实践中并不具备可操作性,“先授权后使用”的传统许可机制失灵。这也触及了著作权法在数字时代的功能定位问题:当技术使用规模远超立法预设时,完全的事前许可机制是否仍具有合理性?是否有必要引入关于AI训练的“合理使用条款”或扩展性集体管理等替代性机制?关于这些问题的讨论已成为学界与实务界争论的焦点,目前尚无准确定论。其次,训练过程具有非表达性、算法黑箱和分散化的特点,这对传统侵权判定框架构成了三重挑战。第一,著作权人难以证明模型接触了自己的特定作品,尤其当训练数据来源混杂且未经记录时;第二,生成内容与训练数据之间往往不存在直接的表达性相似,而是更多体现为风格、情节架构等不受保护元素的相似,使得“实质性相似”标准难以适用^[6];第三,训练行为本身是否为著作权法意义上的“复制”,在法理上仍存争议。特别是当数据仅被用于非表达性的特征提取与参数优化时,其法律定性悬而未决。

在我国,关于训练数据侵权的司法实践正处于审慎探索期,虽暂无统一认定逻辑,但整体上呈现出“侧重输出端认定、包容训练行为但强化平台责任”的倾向。在“上海美杜莎案”中,法院认为用户利用平台训练的LoRA模型所生成的内容不构成作品,故未支持改编权主张;同时,认定平台提供的是“中立技术”,在履行“通知—删除”义务后不承担责任^[7];在“广互奥特曼案”中,法院认定侵权的核心在于生成内容与原告作品存在实质性相似,且平台未预设合理的技术防范措施。对于原告要求将作品从训练数据中删除的诉请,因平台未实际实施训练行为而未予支持,这体现了法院对训练行为认定的谨慎态度^③。

具体到“AI魔改”视频现象,平台责任因用户的直接参与而被进一步放大。用户在“魔改”视频时,除了输入指令外,一般还会自行上传影视片段进行内容的多次训练、调试及最终输出,而此行为本身已涉嫌对作品复制权与改编权的直接侵害。正如在“杭州奥特曼案”中,平台仅将技术接口作为工具开放给用户自行训练,对训练数据的控制力较弱,不构成直接侵权。但因平台的商业模式使其对侵权内容具备预见的可能性,且未在前端采取必要的技术过滤、风险提示等防控措施,最终被认定为未尽到合理注意义务,构成帮助侵权^④。虽然本案的直接侵权主体是用户,但考虑到保护权利人的利益,将责任归于平

台更加高效、务实且节约成本,符合《生成式人工智能服务管理暂行办法》规定的以“服务提供者”(包括开发者、提供商)为中心的归责体系。这实质上明确了平台在“AI魔改”视频这类用户深度参与场景下的前端合规义务底线。

三、创作中的侵权风险:“魔改”视频的版权问题审视

当指令输入完成、AI工具执行并输出“魔改”视频后,我们所面对的便是一个既成的、可供传播的视听文件。问题的核心从“行为如何发生”转向了“结果具有何种法律属性”。这一由技术与用户意图共同塑造的产物,能否被纳入现有著作权法的评价框架?它究竟是对原作品的合法演绎,还是构成侵权?

(一)“魔改”视频的可版权性分析

在深入分析“AI魔改”视频可能侵害的各项具体权能之前,有必要先厘清其自身的法律属性。此类视频的制作大多并未获得原作品权利人的授权,而这一“来源非法性”从根本上否定了其作为合法演绎作品受到保护的前提。那么,倘若不考虑授权因素,视频本身是否能够满足独创性要求,从而可以作为独立的作品获得保护?司法实践表明,AI生成内容构成作品的核心判断标准在于人类是否进行了“主导性创作干预”,即通过精细化提示词、多轮参数调整、反复更新以及后期优化,对最终画面的具体表达要素形成实质性控制,体现出作者的个性化判断^⑤。反之,如果用户参与仅停留于机械输出的层面,且无法提供创作过程记录以证明智力投入,则生成物不具独创性,不受著作权法保护^⑥。

若对照此标准审视“AI魔改”视频,其生成过程虽可能包含多次尝试,却通常难以达到“主导性创作干预”的高度。正如前文对“魔改”提示词的论述,用户对此类提示词的反复调整,本质上仍是在主题设定与元素组合层面进行操作,人物姿态、光影分布、镜头衔接等具体的表达细节,很大程度上取决于模型内部的计算逻辑,而非用户的精确控制。即使用户输入完全相同的提示词,也无法稳定再现一致的内容,这也恰恰说明其对最终表达要素缺乏“可预见的控制力”^⑧。此外,“魔改”视频的创作目的使其自身的独创性受到限制。该类视频的卖点正是其改编后所带给人的视觉冲击,这也决定了它必须高度依赖并忠实于原作品的可识别性。观众之所以觉得好笑,正是因为认出了“魔改”后的角色或者影视剧本身。换言之,生成视频的表达内核始终被限制在原作品所设定的框架之内,其新意更多体现为对既有情节与角色设定的拆解重组,而非从无到有地创造出全新的艺术形象或叙事世界。

不过,这并不意味着“AI魔改”视频绝对不具备独创性。理论上来说,若用户能够证明其通过巧妙的提示词设计、数百次的参数调试以及对生成结果的实质性后期修改,对最终画面的光影、构图、节奏等具体表达要素形成了稳定且可预见的控制,则该生成视频确有满足独创性要求的可能。但从现实情况来看,如今在网络上广泛传播的“AI魔改”视频,大多受流量驱动,内容同质化严重,普遍难以满足独创性的基本要求。因此,将这类视频认定为难以独立产生新著作权的衍生生成物,是更为审慎的判断。而这一确定性,为下文系统审视其对原作品权利人具体专有权利的侵害风险奠定了逻辑前提。

(二)“魔改”视频对多重著作权的侵害风险

第一,对复制权的侵害风险。用户将影视片段上传至AI工具进行处理,或AI模型根据提示词从其训练数据中调用相关作品的视觉元素时,均可能构成对原作品数字内容的复制行为。而该复制往往并非为个人学习、研究等法定合理用途,而是服务于后续的颠覆性改编与网络公开传播,即便仅截取原作少量

片段,也难以纳入私人复制的豁免范畴。若影视片段本身通过盗版、规避技术保护措施等方式获取,该复制行为的违法性则更为显著⁹⁹。

第二,对改编权的侵害风险。改编权即改编作品、创作出具有独创性的新作品的权利。“AI魔改”对原作品表达性内容的定向变更,已触及著作权人的改编权边界。通过提示词指令让AI修改原作人物行为、叙事基调、场景逻辑的行为,显然超出作品引用、参考的合理范畴,构成对原作表达的擅自改编,而改编权专属于著作权人,未经许可的该类行为显然存在侵权风险。

第三,对保护作品完整权的侵害风险。该权利作为著作权人身权的核心,旨在禁止任何歪曲、篡改原作并破坏其艺术内涵与创作初衷的行为。在司法实践中,已有将经典影视角色的歪曲性改编认定为侵权的裁判,即通过修改角色形象、台词风格导致作品核心表达与精神内涵被贬损的,均构成对该权利的侵害⁷。而多数“AI魔改”视频对影视作品的戏谑性、颠覆性重构,恰恰割裂了原作角色与故事主题、艺术风格的固有联系,构成对作品原意的明显歪曲。

第四,对信息网络传播权的侵害风险。当“AI魔改”视频制作完成后,创作者将其上传至网络平台的传播行为,进一步触及信息网络传播权的规制范围。司法实践已经明确,侵害信息网络传播权的“提供行为”以公众能否获得作品为判断标准,并不关注实际获得内容的完整性。视频片段作为作品的组成部分亦能体现独创性表达,未经许可的片段化使用即构成对信息网络传播权的侵害⁸。因此,或许“AI魔改”视频仅使用原作的小部分画面,但其作为包含原作独创性表达的传播物,经上传后使公众可在选定的时间、地点浏览获取,存在一定的侵权风险。

第五,对表演者权益的侵害风险。根据《中华人民共和国著作权法》(简称《著作权法》)第三十九条,表演者有权保护其表演形象不受歪曲。如果“AI魔改”视频不仅篡改了剧情,还通过深度合成技术对演员的面部表情、口型、声音进行了违背原表演意图的修改,导致其表演形象被丑化或用于不当情境,则可能构成对表演者权益的侵害。特别是在修改内容具有贬损性质时,即使未造成广泛的社会负面评价,也可能因损害了表演者与其表演形象之间的精神联系而构成侵权。原作品的表演者若未就此提起诉讼,多是出于诉讼成本、舆论环境等现实考量,并不代表相关侵权风险得以排除。

(三)“合理使用”抗辩的严格审视

我国《著作权法》第二十四条规定了十余种合理使用情形,包括为个人学习、研究而使用,为介绍、评论作品或者说明问题而适当引用,为报道新闻而不可避免再现等。理论上,部分“AI魔改”行为可归入上述范畴,但这类视频实际上多以娱乐恶搞和流量获取为目的。因此,视频的制作人或传播者在面对侵权指控时,最可能援引的只能是前述的“适当引用”条款,主张其行为属于法律允许的评论或说明性使用。但根据“三步检验法”为核心的合理性判断标准,绝大多数“AI魔改”视频难以通过该审查。

第一步,使用行为属于法定情形。“AI魔改”视频的核心驱动多为博取关注和娱乐消遣,尚且缺乏对原作品进行深度解读、评论或说明的实质意图,难以被纳入“介绍、评论或说明”的法定情形。尽管管理论界存在对“适当引用”目的进行扩张性解释的呼声,但在现行立法模式下,突破空间仍较为有限¹⁰⁰。

第二步,不得影响该作品的正常使用。一小段的“AI魔改”视频虽难以直接替代原作,但其规模化、碎片化的传播以及对角色形象与情节的扭曲性呈现,可能会干扰公众对于原作的完整认知与审美体验,进而干扰作品的正常文化传播与市场运行,构成对“正常使用”的实质妨碍。尤其对于青少年群体而言,

他们正处于价值观形成与审美能力发展的关键阶段,若长期接触被恶意篡改、低俗化或片面解构的魔改内容,不仅容易对原作情节、人物与主题产生认知偏差,也可能导致其对原作失去兴趣或产生误解,从而影响作品在年轻群体中的正常传播、接受与价值实现,这实质上削弱了作品在部分受众层面的正常使用效果。

第三步,不得不合理地损害著作权人的合法权益。就财产权益而言,若“AI魔改”视频借助原作的知名度与影响力吸引流量、实现变现,或多或少会挤占原作及合法演绎内容的市场空间,造成财产权益的损害。在人身权益方面,正如前文所述,此类行为直接接触及保护作品完整权、改编权等著作人身权。此外,如果被“魔改”的作品属于承载集体记忆与文化认同的经典文艺作品,此类歪曲性改编还会损害作品的公共文化价值,妨碍社会公众接触和传承真实、完整的文化遗产,这将进一步削弱“AI魔改”的合理性。

不过,有观点试图援引美国判例法的“转换性使用”理论,认为只要为原作注入新的表达、目的或信息,就能赋予二作品新的价值^[11]。然而,真正的“转换性”应体现为对原作品进行批判甚至全新阐释的智力创造,从而产生新的美学价值或社会评论意义。相比之下,“AI魔改”视频的新意往往仅停留在表面,未能基于原作贡献新的实质内容或提供深刻的批判视角,更多的是技术驱动下的感官刺激。这种“转换”并未形成有意义的新的艺术表达或思想观点,难以被认定为“转换性使用”。

四、创作后的责任重构:算法时代平台义务的强化

在传统著作权制度下,“避风港”原则为网络服务提供者设置了责任豁免条件,即平台对用户的侵权行为“不知情且无合理理由应知”。然而在算法时代,平台的角色发生了实质性变化,它不再仅仅是信息的存储者或传递者,而是成为通过算法主动干预内容筛选、排序与推送的控制主体。这一变化不仅挑战了“技术中立”与“技术不能”的传统前提,也使单纯依靠“通知—删除”的事后处理模式,难以应对当前侵权内容数量庞大、传播隐蔽且迅速的现状^⑨。因此,面对“AI魔改”视频这类侵权特征显著、传播范围广泛的典型现象,平台必须承担与其技术控制力相适应的更高注意义务,实现从被动应对向主动防范的转变。这并非单纯加重平台负担,而是要求其法律责任与自身实际具备的内容管控能力、所引发的传播风险相匹配。

(一) 平台“应知”状态的认定强化

认定平台是否承担侵权责任的关键在于其主观过错,尤其是对“应知”状态的判断,而对于“AI魔改”视频这种行为,平台的“应知”状态更容易被推定成立。其一,基于被利用作品的知名度与侵权行为的明显性。被频繁“魔改”的经典影视作品,往往具有较高的公众认知度与文化影响力,平台作为专业的网络视听服务提供者,理应对此类核心文化产品的版权状况具备基本认知^[12]。当平台上持续、大量出现对这类经典作品角色和情节进行明显歪曲、篡改的视频时,其侵权性质如“红旗”般醒目,平台难以主张“不知情”。其二,算法推荐机制构成了对侵权内容的主动干预。平台并非中立地向用户呈现内容,而是利用预设的算法机制,对用户画像进行分析,主动将“AI魔改”视频推荐给潜在的兴趣用户,甚至为其设置话题、标签^[13]。这种主动的、目标明确的推荐行为,使得平台对侵权内容的传播范围和效果具有实质性控制力与助推作用,其过错程度也因此加深。其三,平台与侵权内容之间可能存在直接的经济利

益关联。若“AI魔改”视频带来的巨大流量能够为平台创造直接的广告收入、用户激励分成,或显著提升平台活跃度与用户黏性,则平台与该类侵权内容存在更紧密的利益捆绑。此种直接获利关系,也是认定平台应负有更高注意义务并倾向于判定其构成“应知”的重要考量因素^⑩。

(二) 平台“必要措施”义务的拓展

传统的“通知—删除”规则是一种事后、被动的救济机制,高度依赖权利人发现并通知。在大量且快速生成的“AI魔改”视频面前,这种机制存在一定的滞后性。权利人逐一发现并通知平台的难度极大,而且侵权视频可能在通知前就已经大规模传播。因此,平台所应采取的“必要措施”不应再局限于侵权发生后的被动响应,而应当延伸至传播链条的前端与中端,形成一种体系化的应对路径。

首先,建立前端的规则明晰与合规引导。具体而言,平台需在用户协议及社区规范中划清底线,对恶意篡改、歪曲经典影视作品等行为予以明确禁止。同时,在用户使用工具或上传相关内容的环节,设置清晰的法律风险提示,履行告知义务。

其次,构建中端的技术过滤与推荐约束机制。平台需根据其技术实力,建立分级分类的管理机制:对于直接使用知名作品片段进行明显荒诞改编的视频,平台可利用视频指纹、图像识别等技术,建立正版的作品特征库进行比对,对高度疑似侵权的内容进行主动拦截或转入人工通道重点审核;针对平台自身提供或内嵌的AI视频生成服务,应设置提示词过滤机制,对明显指向“魔改”的指令进行风险警示或生成限制。此外,还需要约束算法的推荐行为,避免对已被标识的“魔改”关键词或话题进行流量加持与热门推荐,在传播路径上抑制侵权内容的扩散^[4]。

最后,完善后端的处置追责与生态协同机制。平台应建立高效的侵权投诉处理通道,对确认侵权的“AI魔改”视频及时采取删除、屏蔽、断开链接等措施。如有个别账号多次发布侵权内容或违规情节严重,应依据平台规则对其采取限制功能、暂停服务甚至封禁账号等处置。同时,平台应定期公开相关治理数据,并探索与版权方建立常态化的版权保护合作机制。

需要注意的是,平台责任的强化需把握合理限度。要求平台对全部内容进行毫无差别的实质性审查,既不具备技术可行性,也会带来过高的运营成本,甚至可能抑制技术创新和合理的表达空间。因此,责任标准的设定应遵循“比例原则”,考量平台类型、规模、技术能力、侵权风险的可预见性以及采取预防措施的成本与收益^[5]。对于用户基数庞大、算法影响力强的头部平台,理应承担与其能力相匹配的更高的注意义务和更积极的防护责任;对于中小型平台,则可设定以有效响应权利人通知为主的义务标准,避免其因过高的合规成本而难以生存发展。简而言之,治理资源应优先聚焦于对经典、高知名度作品进行恶意篡改这类侵权性质明显、社会危害性大的行为。

总之,平台责任的重心应从被动免责转向主动治理。通过强化在算法环境下的“应知”认定标准,并构建一个与其技术能力相称的、全链条的“必要措施”义务体系,推动平台切实履行起内容生态守门人的职责。这不仅是保护著作权人合法权益、维护文化经典严肃性的需要,也是在技术变革中重建利益平衡、促进网络视听行业健康有序发展的关键所在。

五、构建法律、技术与共治相结合的体系化治理路径

通过对“AI魔改”视频创作的前、中、后全过程分析可见,这一现象背后,是技术发展与法律规制

之间尚未弥合的深层鸿沟。其带来的问题，并不局限于对著作权的侵权风险，甚至进一步影响了文化传承的严肃性、创作生态的健康发展以及网络空间的整体秩序。因此，对“AI魔改”现象的治理不能仅停留在个案侵权认定或事后平台追责上，而是需要用体系化的应对思路，做好前瞻性的制度构建。这意味着，除了前文重点分析的平台责任规制外，还需从更宏观的视角，构建一个兼具完备法律规则、多方协同共治与技术伦理引导的综合治理体系，以期在保障技术创新活力的同时，有效遏制乱象，推动AIGC二创走向规范与繁荣。

（一）法律规则的调适与应对

生成式人工智能技术在独创性认定、权利归属与合理使用边界等方面对我国现行著作权法律体系构成了挑战，但这并不意味着法律对此束手无策。治理的关键，并非简单地创设大量新规或施加严苛的绝对责任，而是在现有法律框架的基础上，针对新技术特性进行规则的调整与解释。目标是使法律能够更精准地识别和评价相关行为，为司法裁判提供更清晰的逻辑，同时为产业发展划定更明确的合规红线与安全区。

1. 规避源头侵权风险，强化过程性合规要求

关于AI提示词的法律定性，虽然实践中个案判断存在差异，但“AI魔改”所用提示词的功能，最终指向的是可能构成侵权的生成内容。未来更务实的路径，或许不必纠结于其是否构成“作品”，而是可以从工具属性的角度对其法律地位加以界定。在侵权认定中，提示词可以被视为实施侵权行为的“工具”，而非独立的“表达”。这并不意味着要在立法层面直接排除某类提示词的版权资格，而是为司法裁判提供一种分析思路，即当提示词的价值与功能高度附着于其所驱动的侵权争议内容时，其法律地位应当从属于其所服务的侵权事实。循此逻辑，对“工具”的规制不能仅停留在事后追责，更应在前端使用环节进行风险防范。承接前文所述，平台作为AI生成服务的提供者，对提示词的输入具有技术上的可控性，理应承担相应的管理责任。为此，可将平台是否建立有效的提示词过滤机制纳入“避风港”规则的考量因素。平台若对明显指向“魔改”经典作品的高风险指令采取了合理的技术防范措施，可在侵权认定中作为尽到注意义务的证明；反之，若明知此类提示词泛滥而不作为，则应承担相应的法律责任。而针对创作前端存在的算法“黑箱”特性，法律规制的重心应从追究难以证明的“结果侵权”，转向构建可验证的“过程合规”体系。目前，我国虽然尚无关于AI开发训练的合理使用规定，但在《生成式人工智能服务管理暂行办法》《人工智能拟人化互动服务管理暂行办法（征求意见稿）》中均已确立了数据来源合法的合规义务，体现了积极的“责任前置”导向，司法裁判也强化了这一趋势。未来的法律细化方向，应聚焦于这一原则性义务，继续将其转化为可验证、可审计的过程性要求^[16]。

2. 补齐合理使用制度短板，平衡保护与创新

鉴于我国在AI训练阶段合理使用制度的缺位，若只强化过程合规，长远来看可能过度加重企业合规责任，抑制产业创新活力。因此，未来的制度调适可借鉴域外的有益经验，修订专门的合理使用条款。例如，日本的《著作权法》第三十条之四允许“非享受目的”的数据训练，但以“不得当损害著作权人利益”为限；行政指引则进一步指出，判断是否构成“不当损害”需结合作品类型、利用方式等综合考量^[17]。据此，日本政府于2025年10月表态，要求AI企业切实重视著作权及文化资产保护^[18]。类比“AI魔改”视频，这种监管思路具有一定的启发性：若“魔改”的对象是承载重大文化价值的经典作品，其

扭曲性传播很可能会对原作的声誉与市场造成显著损害,那么在训练阶段使用相关数据的行为,难以依据合理使用条款进行抗辩;反之,若“魔改”的对象仅为普通作品且影响程度有限,则不宜在训练阶段过度追责。至于最终的“魔改”视频是否构成侵权,可于生成阶段单独判断。如此区分处理,既能从源头规避部分侵权争议,也可减轻企业在训练环节的合规负担。

3. 建立类型化裁判指引,引入文化价值损害考量

在现行著作权法下,大多数未经授权生成的“AI魔改”视频构成侵权的概率较高,无法通过合理性检验。不过,司法实践仍需要更加具体的指引来应对AIGC生成物在侵权判断上的模糊地带,以防止类案不同判。对此,未来的法律调适可以侧重以下两点。一方面,通过司法解释或典型案例发布,对利用AIGC技术侵害经典作品相关权利的行为进行“类型化”描述。例如,明确将“利用深度合成等技术,对已公开传播的经典影视作品中,具有高知名度的角色形象、标志性情节进行实质性篡改,导致其原有艺术定位、人格特征或精神内涵被严重贬损的行为”,作为认定侵犯保护作品完整权或构成不正当竞争的典型情形予以列举。这种列举并非创设新权利,而是对现有法律原则在新技术场景下的具体应用进行阐明,有助于统一裁判尺度,并对潜在侵权者形成有效震慑。另一方面,在确定侵权赔偿数额时,引入对作品“文化价值损害”的考量因素。对于“魔改”经典作品的行为,其损害远不止于市场份额的替代,更在于对作品文化声誉和传承价值的贬损。在确定赔偿数额时,法官可在法定幅度内,将被“魔改”作品的经典程度、文化影响力、侵权行为对作品公共形象的破坏性等因素,作为衡量其主观恶意和损害后果的重要情节予以考量。这能使法律责任更好地反映侵权行为的危害性,也符合《中华人民共和国著作权法》维护社会主义文化繁荣发展的立法目的。

(二) 治理机制与标准的协同建设

除了完善具体的法律规则,构建体系化治理路径还需要建立一系列配套的机制与标准,推动治理从被动应对、个案处置转向主动预防、系统治理。这要求政府、行业组织、技术社群等多方主体协同参与。

第一,推动建立跨平台协同治理标准与行为规范。目前,我国已出台《人工智能生成合成内容标识办法》及配套强制性规定,为AIGC的溯源提供了制度支撑。但实践中,侵权视频在一个平台下架后,稍作修改便可能在其他平台重新上传,存在打地鼠式的治理难题。对此,广电、网信等相关部门可牵头制定跨平台协同封堵的统一技术标准,明确侵权内容特征码的生成规则、数据格式与共享接口。当某个平台依据权利人通知或主动审核认定某“魔改”视频构成侵权时,将该视频的特征码同步至共享数据库,其他接入平台据此在前端进行拦截过滤,实现一处认定、全网封堵。同时,可联合发布《网络视听平台版权协同治理行为规范》,明确平台在跨平台治理中的权利与义务,形成覆盖发现、认定、同步、封堵各环节的协同标准体系。

第二,构建弹性监管模式并赋能基层治理。针对AIGC技术的快速迭代所带来的监管滞后与基层执法标准不一问题,监管部门可探索设立“监管沙盒”,在可控范围内允许平台企业、开发商与版权机构就新型过滤算法、权利标识系统或授权交易模式等进行创新测试,通过动态评估将验证有效的做法转化为可推广的行业指引^[9]。与此同时,为提升基层执法部门与平台企业在一线治理中的操作一致性与合规效率,有必要由行业主管部门联合高校、司法机构及头部平台,共同编制分层分类的治理指南。该指南应以可视化流程图为辅助,区分不同二次创作形式的侵权风险等级,明确认定要点,并根据平台规模与技术能

力分别列出必要措施清单与推荐性做法,梳理出成熟可落地的技术防范工具供基层单位与中小平台选择使用。在此基础上,还应建立常态化的基层赋能与反馈机制。通过培训指导、问题直报通道等方式,使基层实践中遇到的制度性障碍与新型侵权手法能够及时反馈至规则制定层面,亦可定期发布平台与监管部门协同治理的成功案例,以经验共享推动治理能力的整体提升,从而形成实践与规则动态调适的治理闭环。

第三,探索建立新型授权与补偿机制。由于“先授权后使用”的传统模式在AIGC海量、碎片化使用的现实面前已显失灵,除了修订专门的AI训练合理使用条款外,可由著作权集体管理组织牵头,与主要的AI模型开发者或平台服务商,谈判达成“一揽子许可+补偿金”的框架协议。该机制允许开发者在支付一笔总的补偿金后,在约定范围内使用作品库进行模型训练,或为用户生成内容提供素材库支持^[20]。但是,这一机制的核心前提和设计关键,必须包含对生成内容性质的严格约束。就AIGC二创领域来说,协议中应明确规定,此许可绝不覆盖,也不得用于生成对许可作品进行“魔改”的内容。平台或开发者必须在技术层面和用户协议中落实这一限制,对指向此类侵权用途的提示词或生成结果进行过滤与拦截。而在补偿金的计算与分配方面,若经监测发现某平台利用该许可产生的“魔改”类侵权内容比例异常偏高,则该平台需支付更高的补偿金,甚至可能面临许可中止的处理。如此一来,新型授权机制便不再仅是简单的“付费使用”,而是将合规生成、尊重原作作为获得低成本、高效率数据利用许可的对价。这有助于引导技术向善,并从经济逻辑上抑制“AI魔改”乱象的发生。

第四,加大对公众的媒介素养教育。首先是面向广大AIGC用户与潜在创作者的普法教育与技能培训。平台、行业协会应与司法行政机关、高校等进行合作,聚焦于“AI创作可能踩中的法律雷区”“尊重版权与创意自由的边界”“经典作品的文化价值”等核心议题,将法律知识转化为创作者易于理解和遵守的操作性建议。其次是提升全社会,特别是青少年的数字素养与媒介批判能力,将“识别深度合成内容”“理解AIGC技术原理与伦理风险”“培养对优秀传统文化与经典艺术的尊重”等内容纳入相关教育课程,帮助年轻一代建立起对AI技术更全面、理性的认识,使他们成为健康网络文化的守护者和传承者。对AIGC时代的二创行为,我们不仅要明确其法律边界,还要创新治理机制以提升系统应对能力,更要借助行业自律与公众教育,从根本上培育一种尊重原创、敬畏经典、善用技术的社会文化氛围。

六、结语

“AI魔改”现象本质上是技术快速发展、法律边界模糊与商业流量驱动共同作用的结果。对其治理不能仅依靠事后追责或单一法律条款的修订,而须采取体系化的应对思路。这意味着,需要在现有法律框架内进行规则调适与解释,同时推动行业技术标准、新型授权机制与公众媒介素养教育的协同建设,构建法律规范、技术赋能与多元共治相结合的综合治理路径。面对技术进步带来的创作自由,我们既要依法规范那些损害版权、扭曲经典的“魔改”乱象,维护好文化遗产的严肃性与创作者合法权益,也应为那些具有创新价值和公共意义的良性二创留出发展空间。最终目标是在版权保护、技术创新与文化繁荣之间找到可持续的平衡点,引导人工智能时代的创作走向更加规范、健康与富有活力的未来。

注释:

- ① 参见北京市互联网法院(2023)京0491民初11279号民事判决书、江苏省常熟市人民法院(2024)苏0581民初6697号民事判决书。
- ② 参见上海市黄浦区人民法院(2025)沪0101民初14775号民事判决书。
- ③ 参见广州互联网法院(2024)粤0192民初113号民事判决书。
- ④ 参见杭州互联网法院(2024)浙0192民初1587号民事判决书、浙江省杭州市中级人民法院(2024)浙01民终10332号民事判决书。
- ⑤ 参见北京互联网法院(2023)京0491民初11279号一审民事判决书、江苏省常熟市人民法院(2024)苏0581民初6697号民事判决书。
- ⑥ 参见张家港市人民法院(2024)苏0582民初9015号民事判决书。
- ⑦ 参见北京知识产权法院(2016)京73民终587号民事判决书、重庆市第五中级人民法院(2019)渝05民初3828号民事判决书。
- ⑧ 参见北京市高级人民法院(2022)京民再62号民事判决书。
- ⑨ “技术中立”的核心内涵可归纳为两点:其一,技术本身并无价值倾向,提供者通常难以实际控制其具体应用,故不应为他人不当使用的行为承担责任;其二,网络服务提供者应采取中立立场,公平处理各类信息,以维护网络信息的自由传播。“技术不能”的法律依据:《最高人民法院关于审理侵害信息网络传播权民事纠纷案件适用法律若干问题的规定》第九条规定,人民法院应当根据网络用户侵害信息网络传播权的具体事实是否明显,综合考虑以下因素,认定网络服务提供者是否构成应知:(一)基于网络服务提供者提供服务的性质、方式及其引发侵权的可能性大小,应当具备的管理信息的能力;“通知—删除”规则的法律依据:《中华人民共和国民法典》第一千一百九十五条规定,网络用户利用网络服务实施侵权行为的,权利人有权通知网络服务提供者采取删除、屏蔽、断开链接等必要措施。通知应当包括构成侵权的初步证据及权利人的真实身份信息。
- ⑩ 参见北京互联网法院(2021)京0491民初10052号民事判决书。

参考文献:

- [1] FLEMINGC. 提示词(Prompt)交易是伪命题还是新蓝海? 值得深思[EB/OL].(2025-03-17)[2026-01-10].https://mp.weixin.qq.com/s/Wb3gg_PaZysO3ISdI8sQ-A.
- [2] 蒋舸.论人工智能生成内容的可版权性:以用户的独创性表达为视角[J].知识产权,2024(1):36-67.
- [3] 李琛.重思“表达”概念的规范意义——兼论人工智能用户指令行为的法律性质[J].知识产权,2025,(5):3-20.
- [4] U.S.Copyright Office.Decision on the Registration for "Zarya of the Dawn"[EB/OL].(2023-02-21)[2026-01-10].<https://www.copyright.gov/docs/zarya-of-the-dawn.pdf>.
- [5] 陈亮,朱晟男.激励前置与梯度赋权:生成式人工智能提示词保护的初步设想[J].北京工商大学学报(社会科学版),2025(4):131-142.
- [6] 许春明,刘桐瑞.生成式人工智能数据训练的著作权合理使用适用[J].上海广播电视研究,2025(4):95-103.
- [7] 上海市高院.《斗破苍穹》美杜莎形象被抄袭人工智能大模型著作权侵权案一审落槌[EB/OL].(2025-11-03)[2026-01-10].<https://mp.weixin.qq.com/s/Plae0snaOEsqmodLU9j4g>.
- [8] 刘维.著作权法中的创作控制论——兼论人工智能生成物的可版权性[J].财经法学,2026(1):3-18.
- [9] 刘丁勤.AI“魔改”经典影视作品的版权风险及治理路径[J].出版发行研究,2025(11):71-80.
- [10] 刘小砚.智能时代数字版权合理使用的司法扩张适用[J].出版发行研究,2021(7):64-69.
- [11] 朱文玉,姜彬彬.转换性使用视角下二次创作短视频合法性探析[J].学术交流,2022(3):55-66.

- [12] 姜丽娜.平台二次利用侵权内容的行为定性和责任认定[EB/OL].(2025-07-21)[2026-01-10].<https://mp.weixin.qq.com/s/CI-JMwvyX5X3p8DasyzeEmg>.
- [13] 孙莹,周宏宇.算法推荐背景下网络服务提供者的侵权认定——全国首例算法推荐案评析[J].时代法学,2023(5):18-29.
- [14] 李晓宇.短视频平台算法推荐侵权的体系化规制[J].河北法学,2025(6):147-163.
- [15] 杨帆.人工智能治理中平台社会责任的法律定位[J].行政法学研究,2025(3):146-159.
- [16] 朱尉贤.中国数据合规2025年观察:深化、协同与激活[EB/OL].(2026-01-09)[2026-01-10].<https://mp.weixin.qq.com/s/dCd3DXeMgEjn8GzxHTC2Dw>.
- [17] AIと著作権Ⅱ-解説「AIと著作権に関する考え方について」[EB/OL].(2024-08-09)[2026-01-10].https://www.bunka.go.jp/seisaku/chosakuken/pdf/94097701_02.pdf.
- [18] 政府、OpenAIに著作権侵害防止を要請——「Sora2」問題で平デジタル相は“オプトイン方式”を提言[EB/OL].(2025-10-14)[2026-01-15].https://ledge.ai/articles/openai_sora2_government_copyright_request_oct2025.
- [19] 杨丰一.人工智能“监管沙盒”:理论溯源、实践考察与制度构建[J].中国行政管理,2024(12):25-35+56.
- [20] 袁真富,夏子轩.机器学习作品中作品利用的著作权补偿金制度研究[J].科技与出版,2024(7):28-36.

Examination of Copyright Issues and Governance Approaches for "AI-altered" Videos

ZHANG Yibo, CHEN Shangxian

Abstract: The widespread adoption of generative artificial intelligence technologies has spurred a wave of secondary creation, typified by "AI-altered" videos. While these creations circulate extensively, they also pose significant copyright risks and governance challenges. Following the creative logic of "AI-altered" videos, this paper conducts a normative analysis to explore the legal characterization of prompts prior to creation and the compliance challenges associated with copyrighted training data. It systematically examines the legal attributes of such videos during the creation process and the risks of infringing on multiple copyrights, including the rights of adaptation and reproduction, while rigorously scrutinizing the "fair use" defense. The paper further argues that platforms should bear enhanced duties of care in the post-creation phase, shifting from passive responses to proactive prevention. To effectively address secondary creation controversies such as "AI-altered" content, which are prone to copyright disputes, it is essential to promote multidimensional coordination, including legal refinement, the establishment of industry standards, innovation in licensing mechanisms, and public education. A systematic governance framework integrating law, technology, and collaborative governance should be developed to achieve a sustainable balance among copyright protection, technological innovation, and cultural preservation.

Keywords: AI-altered videos; derivative works; AI prompts; copyright infringement; copyright issues

责任编辑:周兴涛