重庆大学学报(社会科学版) 2026年第32卷第2期
JOURNAL OF CHONGQING UNIVERSITY(Social Science Edition) Vol. 32 No. 2 2026

Doi:10. 11835/j. issn. 1008-5831. fx. 2026. 03. 004

欢迎按以下格式引用:李晓阳. 生成式人工智能服务提供者能否适用避风港规则——基于侵权主体适格性的探讨[J]. 重庆大学学报(社会科学版),2026(2):224-238. Doi:10. 11835/j. issn. 1008-5831. fx. 2026. 03. 004.

Citation Format:Li Xiaoyang. Can generative AI service providers apply the safe harbor rules:an exploration based on the suitability of infringing subjects[J]. Journal of Chongqing University (Social Science Edition),2026(2):224-238. Doi:10. 11835/j. issn. 1008-5831. fx. 2026. 03. 004.

# 生成式人工智能服务提供者能否适用避风港规则
## ——基于侵权主体适格性的探讨

李晓阳

(浙江工商大学 法学院,浙江 杭州 310018)

**摘要**:生成式人工智能服务在内容生产中的深度参与,对原有的网络版权架构造成强烈冲击,生成式人工智能服务提供者的侵权责任分配成为亟待解决的难题。避风港规则作为过去网络时代平衡技术创新与版权保护的基石,当前面临着是否能适用于这一全新主体的严峻考验。文章基于侵权主体适格性的视角,对此展开讨论。从制度演进历史来看,避风港规则的适用高度依赖于主体的适格性,其经历了网络服务提供者与内容提供者分离,到内容服务提供者兴起的变化。然而,生成式人工智能技术实现了从"发现信息"向"重组信息"甚至"创造信息"的跨越。生成式人工智能服务提供者直接参与内容生产环节,打破了传统网络服务提供者不参与内容的"中立性"前提;同时,由于其内容生成受制于用户指示,且跨模态的规模化生产导致内容来源难以回溯。它也无法被归类为版权法意义上的内容提供者。因此,传统的"网络服务提供者与内容提供者"二元主体框架在生成式人工智能时代陷入适用迷思。如果强行将生成式人工智能服务提供者拟制为传统网络服务提供者,并径直适用现有的避风港规则,将使其陷入合规的"两难"困境:一方面,受限于生成式人工智能的创作工具本质,其无法履行防范侵权作品传播的"注意义务";另一方面,由于AI生成内容是根据参数实时生成而非事先储存,服务提供者在技术上无法完成"通知—删除"规则下的断开链接或删除等必要措施。因此,破局点为摒弃修补式路径,在法律上承认生成式人工智能服务提供者是一类全新的独立责任主体。在其侵权责任的判断上,应当从传统避风港规则关注传播的"知情/不知情"推定,转向关注内容生成的"可知/不可知"标准。基于此,文章提出应当在当前内容生产者责任框架下,科学引入并重构专属于该类主体的"AI避风港"规则。具体包括:一是设立"AI训练免责"机制,明确将合法作品用于模型训练的技术处理行为视为非侵权使用;二是设立

**基金项目**:2024年度浙江省哲学社会科学规划课题"浙江省数据要素流通交易的法律规则研究"(24NDQN076YB)

**作者简介**:李晓阳,法学博士,浙江工商大学法学院(知识产权学院)特聘副教授。

“可知免责”规则，要求提供者通过添加水印标识和提供来源检索链接，确保生成内容及来源处于“可知”状态，并以此作为责任划分与抗辩的依据；三是设立“绕行免责”规则，鼓励服务提供者通过模型优化、重写与安全评估，主动偏离训练数据以规避侵权风险。通过这一套全新规则的构建，旨在保障AIGC技术持续创新发展的同时，有效维持现有版权秩序的制度平衡。

**关键词**：生成式人工智能；生成式人工智能服务提供者；网络服务提供者；避风港规则；版权责任

**中图分类号**：D923.4　　**文献标志码**：A　　**文章编号**：1008-5831(2026)02-0224-15

人工智能逐步成为推动新一轮科技和产业革命的重要驱动力。生成式人工智能(generative AI)是人工智能的一个重要发展方向，其特点是利用数据和算法生成人类可以理解的文本、图片、视频等内容(即人工智能生成内容，下称AIGC)，被应用于图像识别与合成、文本翻译、语音模拟等领域。这一特点使得人们开始忧虑，AIGC所涉及的知识产权风险是否在《著作权法》等法律体系框架下可控。避风港规则被誉为“网络时代的自由大宪章”，过去在著作权法领域，一旦面对新技术的冲击，各类主体都迫切向避风港规则“索要生产力”，划出利益均衡的缓冲带。这一境况会在AIGC时代复现吗？本次避风港规则恐难直接胜任。分析的视角是多元的，但从主体适格性切入，更易检视其本质。AIGC不再是一种“发现信息”技术，而是“重组信息”技术，令生成式人工智能服务提供者变得特别：它既不同于网络服务提供者，亦不同于内容提供者，这使得避风港规则无法以“旧瓶装新酒”的方式应对AIGC冲击。避风港规则的演进历程是理解这一变化的关键，将于本文第二章予以说明。要纾解生成式人工智能服务提供者的著作权风险，需以“新瓶装旧酒”的方式，将生成式人工智能服务提供者视为一种新的责任主体，而非某种避风港规则中特殊的责任主体，在内容生产者的框架下，重新构建“AI避风港”的新规则。

## 一、问题的提出

生成式人工智能服务尚未成熟，已遭诉累。2023年2月，Getty Images(图库摄影公司)作为原告对Stability AI在美国提起诉讼，指控其未经权利人许可，获取与利用其版权作品作为生成式人工智能服务提供者Stable Diffusion的“训练图像”。其中，原告直接质疑“AIGC生成新图像”的概念，认为Stable Diffusion只是提供了“一个复杂的拼贴工具”，并“将无数受版权保护的图像存储和合并为训练图像后……生成完全基于训练图像的‘新’图像”①。Getty Images的抨击充斥着作品交易市场受AIGC冲击的辛酸，也清晰地展示疑惑：类似Stability AI的生成式人工智能服务提供者，到底在版权法律关系中扮演什么角色？对于AIGC，它是“卖铲子”的网络服务提供者，还是“掘土”的内容提供者？

目前，法律没有作出清晰的回应。为了应对AIGC服务可能对目前网络监管体系带来的冲击，2023年7月，国家互联网信息办公室联合六部委发布了《生成式人工智能服务管理办法》(以下简称《管理办法》)，对生成式人工智能产品的开发和应用设定监管规则。其中第九条提出，“提供者应当依法承担网络信息内容生产者责任，履行网络信息安全义务”②。这一规定让“生成式人工智能服务提供者”的概念，结合了网络版权侵权中“网络服务提供者”(Internet Service Providers，ISP)的身份和与之对应的“内容提供者”(Internet Content Providers，ICP)的内涵。这两个概念来自美国1998年《千

① See Getty Images (US), Inc. *v.* Stability AI, Inc. and Stability AI, Ltd. Case: 1:2023cv00135 (2023).

②《生成式人工智能服务管理暂行办法》第九条规定，“提供者应当依法承担网络信息内容生产者责任，履行网络信息安全义务。涉及个人信息的，依法承担个人信息处理者责任，履行个人信息保护义务”。

禧年数字版权法》("DMCA")确立的避风港规则(Safe Harbor Rules),其中将服务提供者划分为两大类:一类是参与内容制作的,被称为"内容提供者"(ICP);另一类则是不参与内容制作亦对内容不知情的"网络服务提供者"(ISP)[1]。随着生成式人工智能向人类知识对齐的能力达到了一个崭新的高度,引发了版权法上的迷思:生成式人工智能服务提供者如何承担内容生产者责任,应将其视作"网络服务提供者"或"内容提供者",还是一种新的责任主体类型,均无清晰的答案。

这带来了侵权责任认定的新难题:如果其生成内容,因他人使用构成版权侵权,生成式人工智能服务提供者应否承担共同侵权责任?这让人们不得不思考,生成式人工智能服务提供者与网络服务提供者是否具有同样的法律意涵?如果二者具有同样的法律内涵,生成式人工智能服务提供者在AIGC侵权纠纷中扮演什么角色?如果二者不同,避风港规则能否适用于生成式人工智能服务提供者?

## 二、避风港规则适用主体的演进检视

在版权法应对技术革新挑战的历程上,责任主体的演进与避风港规则的变迁,是紧密相连的[2]。新技术的出现,总是带来避风港规则与侵权主体适格性的新互动。两者互动的关键要素是什么,需从历史视角开始检视。

### (一)网络服务提供者与内容提供者的分离

目前,关于避风港规则的溯源,多从1984年美国Sony案③谈起,讨论"实质的非侵权用途"(Substantial noninfringing uses)标准,对避风港规则形成的影响[3]。而甚少被提及的是,Sony案后的十年内,大量美国的网络版权侵权纠纷,并没有出现避风港规则的适用雏形,法院甚至要求网络服务提供者,需确保自己经营控制内容的版权合法性,才能获得自身"生存的合法性"[4]。尽管被告曾试图从侵权行为的主观过错、对侵权行为的客观控制能力等方面,对网络服务提供者的角色地位进行辩解,但仍被法院判决承担版权侵权责任。其核心原因是,法院在审理涉及网络版权纠纷时,并不区分"网络服务提供者"与"内容提供者"。直到同类案件不断增多,关于"网络服务提供者"的模糊认识逐步厘清,才引发一系列议题讨论:侵权内容是否由网络服务提供者提供,网络服务提供者所提供的服务是否与侵权行为直接相关,网络服务提供者承担版权责任是否会威胁网络信息产业的发展等问题。新的思想开始被接纳。网络服务提供者显然不宜再承担版权侵权的直接责任或者共同侵权责任。首先是判例,其次是立法,才有了网络服务提供者"避风港"的例外[5]。

在我国,网络服务提供者、内容提供者和避风港规则,是一并出现的。2006年《信息网络传播权保护条例》正式引入避风港规则。从一开始,在网络侵权主体的界分上,就采取了"网络服务提供者"和"内容提供者"二分的基本框架。紧随其后,2009年《中华人民共和国侵权责任法》(以下简称《侵权责任法》)第36条规定"网络用户、网络服务提供者利用网络侵害他人民事权益的,应当承担侵权责任",将网络侵权责任主体限定为特定的网络服务提供者[6]。这一范式同样延续到2020年通过《中华人民共和国民法典》(以下简称《民法典》)"侵权责任编"中,在第1194条沿用了《侵权责任法》第36条的规定作为网络侵权的一般规则。同时,《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4条,通过对第一款中"与他人以分工合作等方式共同提供作品、表演、录音录像制品"和第二款中"仅提供自动接入、自动传输、信息存储空间、搜索、链接、文件分享技术等网络服务",再次确认了内容提供者与服务提供者的明确区分。

③ See Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417 (1984).

适格主体的分离，也标志着网络版权间接侵权制度正式形成。内容提供者和网络服务提供者④的划分，与著作权法中的侵权责任类型一致[7]：一种是直接侵权产生的自己责任，另一种是间接侵权产生的第三方责任[8]。避风港规则是成功的“舶来品”，据不完全统计，尼泊尔公约176个成员国，超90%引入避风港规则。适格主体的分离是关键：一方面，为网络服务提供者预留积极履行预防和制止侵权行为发生的机会，使其可免于承担网络版权间接侵权责任；另一方面为内容提供者和网络服务提供者之间建立起预防和制止网络侵权的合作沟通机制[9]，促使网络服务提供者与内容提供者合作，共同打击网络侵权[10]。

**（二）内容服务提供者成为一种特殊类型的网络服务提供者**

自避风港规则出现后，“新技术广泛应用→新责任主体出现→避风港规则调整→形成新的均衡状态”，就成为著作权应对新技术挑战的基本路径。这一过程中，既有不同利益的博弈，也有概念不明确产生的困惑[11]，正如内容服务提供者。随着数字内容在互联网的广泛传播，被认为“有能力控制内容呈现”的网络服务提供者遭受权利人的控诉。新的责任主体分类，又在全球范围热议：（1）把内容提供者所创作的内容直接上传、提供给用户（例如早期的门户网站直接设置内容栏目）[12]，或提供的服务让用户意识不到内容是来自第三方网站的（例如提供下载链接用户点击直接获得内容），是“内容服务提供者”，其因服务效果直接决定内容呈现而可能承担自己责任[8]；（2）不直接参与著作权内容的创作与分享，它被动、中立地存在于双方当事人之间[13]，仅提供接入、缓存、信息储存空间等技术服务，则为“技术服务提供者”，其被归于承担第三方主体责任[14]。相关的观点被司法实践吸收后，规则又有进一步的调整。《最高人民法院关于审理涉及计算机网络著作权纠纷案件适用法律若干问题的解释》第4、5条要求“提供内容服务的网络服务提供者”如在版权侵权案中被认定没有承担必要的注意义务和“通知—删除”义务，将承担共同侵权责任。澳大利亚的《广播服务法案》、印度的《信息技术法》也提出了相似的概念。

事实上，内容服务提供者的出现，改变了早前网络服务提供者的法律意涵。首先，内容服务提供者与内容提供者的边界，不如早前的网络服务提供者清晰。对于谁是“控制内容呈现”的主体，在内容提供者和内容服务提供者之间产生争议，由谁承担侵权责任，是个尚待讨论的问题。其次，网络服务提供者的角色功能，从一个“中介通道”，即为用户创作、发布、分享内容提供支持，成为网络空间中“看不见的手”[15]。这种高度的信息参与、内容参与以及经济利益参与，再次引发网络服务提供者是否要对自己控制的内容负直接侵权责任的疑虑[16]。

内容服务提供者的引入，却没有破坏“网络服务提供者/内容提供者”的二元主体框架。通过增设新的义务（如“注意义务”和“通知—删除”规则），内容服务提供者与内容提供者的争议，逐渐被搁置。承担新的义务后，内容服务提供者作为一种特殊的服务提供者，得以寄居于网络服务提供者的框架之下，继续适用避风港规则。

然而，内容服务提供者实际承担的法律义务，与避风港规则对网络服务提供者预设的法律义务，已经发生很大的改变。有学者认为，这是法律规则为了维持版权法在技术创新和权利保护间的协调，而不断让网络服务提供者承担更多的法律义务[17]。其结果是，避风港规则成为一个具有弹性的制度框架，为新技术的服务提供者，提供了一个可以躲避侵权风浪的港湾。而驶入这一港湾的条件是，作为一种特殊的网络服务提供者，内容服务提供者需积极履行注意义务，同时，这也为法院在个案中具体甄别网络服务提供者注意义务标准保有弹性空间，以动态调整内容服务提供者和内容

④ 内容提供者强调内容由自己创作、选择、编辑，再通过网络传播；网络服务提供者强调内容由网络用户产生和提供（user-generated），网络服务提供者仅仅为这些内容的存储、接入、搜索、链接、传播提供技术支持和便利条件。

提供者的责任边界[10]。

### (三)主体的适格性影响避风港规则的制度弹性

在历史的检视中,一个未被厘清的问题是:避风港规则可延展的范围到底有多宽?其关键的影响因素是技术中立原则的局限性、"通知—删除"有效性,抑或注意义务实现的成本?事实上,以上列举的三项因素,背后都指向数字技术。纵观避风港规则的演进历程,其与所涉数字技术的"第一链接点"为设立适格主体。例如,2012年卷积神经网络算法(CNN)兴起后,在接下来的四年内,利用深度学习的算法对内容进行推荐的网络服务,对内容分发的干预再次加深,令"内容服务提供者"走到一个新的十字路口。2018年,欧洲《数字化单一市场版权指令》通过第17条,提出了"在线内容共享服务提供商(OCSSP)",明确其责任限制的"四步法"⑤,其责任要求明显高于欧洲议会和理事会发布的《关于协调信息社会中版权及相关权若干方面的指令》(2001/29/EC号指令)对网络服务提供者的规定内容。OCSSP的出现并非偶然,在"新技术广泛应用→新责任主体出现→避风港规则调整→形成新的均衡状态"的路径演化之中,设立新的责任适格主体是"关键一跃":其定义是否能准确反映新技术的特征,与过往已有的责任主体形成区别;同时,对避风港规则进行有效的调整,令各参与方利益在动态博弈下走向新的均衡。

换言之,避风港规则的可延展范围,取决于责任适格主体的有效设立。这里容易产生误解,责任主体可以被"随意设立":把新业态的主体加上"XXX服务者"名称,再苛责以更高义务。检视历史的重要性再次凸显。从DMCA开始,在全球范围内,避风港规则的责任适格主体,均是"网络服务提供者"及其不断衍生的分支(新的特殊类型)。截至目前,适用避风港规则的各类责任主体在技术特征、技术标准上虽各有差异,但都具备开放性和互动性,并不断提升信息发现的效率[18]。但是,AIGC呈现出与一般网络服务截然不同的技术特征,甚至与推荐算法都存在鲜明的差别。这令生成式人工智能服务提供者,是否可视为"网络服务提供者"成为一个需要重新检视的命题。

## 三、生成式人工智能服务提供者的适用迷思

避风港规则的适用,以往通常涉及网络服务提供者、内容提供者以及网络用户三方主体。生成式人工智能服务提供者的特殊之处在于,它一边作为技术服务提供者,一边又深度参与内容生产,再次打破早前法律对网络服务提供者的能力想象,又让内容提供者的身份认定产生新的迷思,侵权责任分配原有的平衡同时被打破。因此,需进一步分析生成式人工智能的技术特征,将如何影响内容服务提供者参与版权责任的分配。

### (一)生成式人工智能服务提供者区别于传统网络服务提供者

传统网络服务提供者一般提供包括接入、储存、传输、搜索等在内的技术服务,这些服务围绕内容的传播和利用而衍生出各种形态,但都不直接参与内容生产。从理论上而言,网络服务提供者在侵权情形发生以前,对其服务的内容详情并不知情。这也是避风港规则适用的重要前提,因为一旦证明网络服务提供者对侵权内容知情或应当知情,即将触发其相关注意义务。即便算法推荐等技术加剧了网络服务提供者权力扩张,它仍是通过对用户的隐性规训、信息筛选等间接方式来强化其占有的技术资源优势[19],而非直接决定内容生产。避风港规则的重要目的之一,即维护网络服务提供者的中立地位,避免版权责任激励网络服务提供者过度参与网络内容的审查,让网络服务提供者

⑤ 参见:欧盟议会理事会《数字单一市场版权指令((EU)2019/790)》第17条:明确了OCSSP的责任限制机制,包括四步流程:尽最大努力获取授权、确保特定作品不被提供给公众(专业注意义务的较高行业标准)、立即执行通知和删除要求、防止被通知和删除的内容再次上传。

远离具体内容，以维护网络环境之自由[20]。

与之不同，生成式人工智能服务提供者，所提供的服务是直接参与内容生产。这令生成式人工智能服务提供者，与网络服务提供者，以及内容服务提供者，有本质上的不同。从技术看，生成式人工智能源于2017年谷歌几名技术人员提出Transformer训练框架，以ChatGPT为代表的生成式AI，证明了文字、图片、视频等类型的数据，通过不断迭代参数网络、训练次数和数据量，可以拓展语义网络，形成了“能说、会看、会画、会写”的生成式人工智能。从影响看，这一模式重塑了“人机交互”方式。在AIGC出现之前，传统的网络服务提供者，以“人找信息”的搜索方式为主；而内容服务提供者，以“信息找人”的信息流推送方式为主，但两者都没有改变信息自身的内容。AIGC的出现，改变了“信息”本身，这是一种新的“内容生产方式”[21]，用户使用互联网获取信息的方式，也会从“搜索信息”或“接受推送”，改变为“交互式对话”[22]。目前，AIGC产品通过对话、学习等方式，已经可以帮助用户实现从“灵感到内容”的转变。

直接参与内容生产环节，让生成式人工智能服务提供者，已无法适用到网络服务提供者的规范体系之中。以《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》为例，第七条规定，“网络服务提供者在提供网络服务时教唆或者帮助网络用户实施侵害信息网络传播权行为的，人民法院应当判令其承担侵权责任”。该条意在约束，网络服务提供者（尤其是内容服务提供者）与内容提供者的共谋，进行规模化侵权，但放在生成式人工智能服务提供者的语境下，“帮助”的边界变得模糊。用户任何刻意复现某个文字、图片、录音录像的意图，都会以“帮助”的形式进行，因为AIGC充当着“写手”“画手”“视频编辑手”的角色。又例如，该规定第四条，“网络服务提供者能够证明其仅提供自动接入、自动传输、信息存储空间、搜索、链接、文件分享技术等网络服务，主张其不构成共同侵权行为的，人民法院应予支持”。如果在不构成共同侵权行为的网络服务类型中，增设“自动生成”，以寻求生成式人工智能服务的豁免，其危害更甚，会径直破坏网络服务提供者与内容提供者之间的利益平衡。在过去，网络服务提供者不构成共同侵权，主要是基于“内容无涉”，即使是内容服务提供者，一旦采取“将热播影视作品等置于首页或者其他主要页面”等内容推广行为，也会推定为“应知”。但生成式人工智能服务提供者直接参与内容生产环节，这就产生了天然的不适，使之无法融入以网络服务提供者为适用主体的避风港规则体系之中。第四章将进一步论述该点。

**（二）生成式人工智能服务提供者无法成为内容提供者**

这种内容生产的深度参与，让生成式人工智能服务提供者是否会演变为版权法意义的内容提供者，成为一个新的问题，因为两者的边界已无限模糊。从内容输入侧和输出侧，生成式人工智能服务提供者都无法替代成为内容提供者。

从输入侧而言，AIGC内容生产无法脱离用户指示。当前，GPT大模型实现AI从单专项能力超越人类，到通用能力逼近人类的跨越。无论AI的学习与输出能力获得如何跃升，它始终无法脱离用户指示，而凭空生成相关内容。尽管人工智能的研发者或者使用者对人工智能生成内容仅有间接影响，无法简单将生成内容这一过程形容为人类利用写作工具获得的结果[23]。至于如何理解用户指示和生成内容的关系，要回溯到中世纪时期对于作者和作品的讨论。18世纪以前的人们认为个人的写作能力和灵感源于神的启示，作者仅是传达真理的工具[24]。在这一论断下，人是无法脱离神的启示（即灵感）而进行独创性劳动的，作者亦非独立的主体。只有在经历宗教改革后，人们逐渐认识到人的理性可以取代神的信仰，作品的来源才不再是神的启示，而是人的创造[25]。当前，AIGC内容生产能力确实迎来了“量变”到“质变”的历史性拐点，令部分学者认为AI创作的兴起、人类作者的

淡出和作者的“祛魅”,作者是作品本质来源的“作者中心主义”受到根本挑战与动摇[26]。

但这一观点暂未得到足够多的事实支撑,更保守的观点占据上风。2023年3月16日,美国版权局发布了《版权登记指南:包含人工智能生成材料的作品》,引发了各国的关注。对于审查和注册包含人工智能技术生成材料的作品,美国版权局认为,“宪法和版权法中使用的‘作者’一词不包括非人类”,同时“科技工具可以是创作过程中的一部分,但作品表达的创造性必须是由人类控制的。如果只是AI技术根据人类的提示产生作品,则该作品缺乏人类作者身份,不受版权保护。2023年8月,美国联邦地区法院在“塞勒诉美国版权局案”⑥中否定了AIGC的作者身份,提出“人类在人工智能生成作品中扮演着不可或缺的角色”,并且强调了作者的人类身份,认定原告所使用的生成式人工智能既不能作为版权法意义上的作者,也不能因原告“雇佣”而视为作者。保守立场的坚固,源于科学界未能对生成式人工智能如何生成内容,给出有信服力的解释,对于这“有果无因”的技术产物,其性质认知总处在“不可知”的状态。这令AIGC产品的应用定位仍是“各行业知识工作者的新生产力工具”,这一距离足以将生成式人工服务提供者排除以内容提供者。

从输出侧而言,大模型使得AIGC内容生产实现“跨模态生成”和“规模化生产”转变,内容来源难以回溯,不具备承担直接责任的能力。这是一个技术事实。现实世界中,人类信息的传递总是依赖于某种载体作为媒介,包括文字、图片、音乐、信号等,在这一视角下,作品因作为知识产权对象的载体而受到保护[27]。这也是网络时代著作权权利体系转向“接触权”的重要原因[28]。面向人工智能,尽管早期的机器学习模型可对已有信息进行学习和内容生成,但传统模型只能基于单一模态进行训练学习和内容生成,如识别文字、识别图片。当前,生成式人工智能不仅可以对多种模态同时进行训练学习,从“一种模态”中学习到的能力,还可以有效应用到“另一种模态”,实现“文字到图片”、“图片到视频”等跨模态的内容生成能力[29]。这种“跨模态”的内容生成能力,源于生成式人工智能把不同媒介(如文字、图片、音乐、信号等)的理解转为参数,储存在模型之中,形成不同参数规模的模型,业界常说的“72B(billion)”“200B”就是在描述参数的数量。AIGC产生内容时,事实上是调用不同部分的参数预测下一个要生成的文字、像素点(预测机理暂不明晰),并非通过复现用以训练内容而生成。无法追溯内容来源,是由生成式人工智能的技术特征决定的。

如果内容来源难以追溯,那么,AIGC是否会因为事实上不可侵权,而令版权规则适用变得“杞人忧天”?有学者持肯定观点,“经过训练的机器模型,最终通常会产生与原始图像不同的新图像”[30],且AI是典型的技术理性,它的运作排除了情感、欲望的影响而仅遵从逻辑和算法[31]。换言之,AIGC输出内容侵犯版权是极小概率的事件。但现有的实证研究证明,侵权风险是不可避免。以Stable Diffusion模型为例,在一个子训练集(约1200万图片)的测试中,AIGC的生成内容与数据集作品相似度超过50%的可能性达到了1.88%[32]。鉴于庞大用户量和调用量,很难径直认为,生成式人工智能服务提供者无侵权之虞,尽管无法追溯内容来源,AIGC生产内容与现有作品也存在高度相似的概率。

### (三)“网络服务提供者/内容提供者”的二元框架或难适用于生成式人工智能服务提供者

在现阶段,生成式人工智能服务提供者的内容服务能力,已超越了过往案例中的内容服务提供者,前者不仅参与内容的传播环节,甚至参与到内容的生产环节,这“打破了著作权法的底层逻辑”[33]。

从功能看,生成式人工智能服务提供者彻底模糊了“网络服务提供者”和“内容提供者”的角色。

⑥ See Thaler *v.* Perlmutter, 22-cv-01564-BAH Document 24.

生成式人工智能服务提供者所提供的服务无法再从“内容”或“服务”中获得确定性归类，它不再无限围绕内容展开，而直接通过对已有作品和灵感的学习训练，实现最终内容的输出呈现。从影响看，这对于谁是“控制内容呈现”的主体产生巨大的理论冲击，如未能厘清，也很难继续沿用内容服务提供者对避风港规则的适用。但另一方面，AIGC并未真正达到人类的认知和推理水平，其对语料的学习认知过程存在一定的黑盒属性，产生知识跃迁和跨领域涌现的能力还未得到有效解释。这意味着如不对生成式人工智能服务提供者课以相应的责任，受预训练阶段语料质量、范围和模型结构限制而产生的版权侵权问题将因规范失序而被放大。生成式人工智能服务提供者并非传统意义的网络服务提供者，也不能认定为内容提供者，但却是AIGC内容生产和利用的重要主体，它如何参与版权责任的分配直接影响着AIGC的未来发展。

## 四、生成式人工智能服务提供者的版权责任承担路径

目前，对AIGC的版权保护，很难绕开人工智能是否构成适格的作者主体，以及作品是否满足独创性标准[34]。更难形成的是对这两个问题的共识。在科学界没有厘清AIGC的运作机理前，涉及生成式人工智能“作品”“作者”的讨论，只能停留在没有准确映射对象的“隐喻”之中。即使这一机理被发现，也需要相对较长的周期来完成法教义学沉淀。但现实纠纷已经迫在眉睫。如我国首例人工智能生成文章作品纠纷“腾讯诉盈讯科技侵害著作权纠纷案”⑦，以及最新针对OpenAI发起的“Doe诉Github Inc案”⑧等，已经涉及生成式人工智能服务提供者的版权责任问题。当前，生成式人工智能服务提供者，要在版权纠纷中径直适用避风港规则是困难的，现有的责任主体与之均不适配，应把“生成式人工智能服务者”视为一种全新的主体，并增设新的义务。

### （一）生成式人工智能服务提供者无法直接适用现有的避风港规则

生成式人工智能服务提供者，需要一个“避风港”。2023年12月，美国作协与17名作家诉OpenAI侵害其版权⑨，认为OpenAI在未经授权的情况下，使用版权作品训练大模型，提出“训练这一词汇是对复制和再现在技术上的委婉表述”，使其可能总结、复述及模仿这些作品的衍生作品。尽管暂未作出最终裁决，但新技术对版权秩序的冲击已昭然若揭。纽约时报诉OpenAI⑩与之相似，两案原告均认为未经授权使用作品训练大模型，是对其版权的侵害。一个被忽略的事实是，自大模型出现之前，作品就一直以人工智能的训练作为数据集储存在服务器，作为训练素材在服务器进行预处理，作为模型调优样本在处理器运行。这几种作品的利用行为，自2014年循环神经网络（RNN）兴起时就存在，成为模型训练的基本范式。2017年Transformer框架提出后，训练规模和效率有了指数级增长。但在训练过程中，对作品使用的方式从未改变。为何2023年OpenAI发布GPT3. 5之前，这种“对复制和再现在技术上的委婉表达”，从未遭诉累，亦没有侵权讨论？问题的背后，揭露着版权人权利主张的“实质”：一方面，试图悄然扩张权利范围，把“将作品用以训练行为”纳入侵犯版权之列；另一方面，表达了对大模型创作能力的不安，对作者群体已造成直接或间接的利益损害。当前，如何界定“将作品用以训练行为”是关键。事实上，骤然扩张版权的权利范围是不可取的。大模型得以出现，主要是基于训练规模的不断扩张，一旦作品用以训练需经版权人同意，大模型发展将因边际成本不断叠加而戛然而止，哪国采取这一主张，其大模型产业无疑将直接“出局”。因此，要防

⑦ 参见：（2019）粤0305民初14010号民事判决书。

⑧ See Doe *v.* Github, Inc., No. 22-cv-06823-JST, 2023 U.S. Dist. LEXIS 86983 (N.D. Cal. May 11, 2023).

⑨ The Authors Guild v. OpenAI, No. 1:23-cv-8282-SHS, 2023 U.S. Dist. Court of South Dist. NEW YORK (Dec. 4 2023).

⑩ The New York Times v. OpenAI, No. 1:23-cv-11195, 2023 U.S. Dist. Court of South Dist. NEW YORK (Dec. 27 2023).

止诉累频现,生成式人工智能服务提供者需要新的“避风港”,在版权领域将训练行为合法驶入“安全港湾”,并针对作者权利受损的行为进行制度的调整安排。

当前面临的首个挑战是,生成式人工智能服务提供者无法直接适用现有的避风港规则。现有关于生成式人工智能服务提供者的侵权责任讨论,多从其过错判断标准以及注意义务要求的高低展开[35-36],更多将其视为“内容服务提供者”,继续适用避风港的旧框架下进行。事实上这行不通。区分“网络服务提供者”和“直接实施版权侵权的行为”两个角色,是避风港规则适用的前提。即使内容服务提供者出现,也是作为网络服务提供者的一种特殊类型。但如前所述,生成式人工智能服务提供者本身既非网络服务提供者,也非侵权行为人。如果生成式人工智能服务提供者径直适用《民法典》《中华人民共和国刑法》《中华人民共和国著作权法》《中华人民共和国网络安全法》以及《信息网络传播权保护条例》等现行法律法规关于“网络服务提供者”的相关规定,会造成合规的“两难”。

首先,生成式人工智能服务提供者无法履行网络服务提供者的注意义务。避风港规则对网络服务提供者设定注意义务要求,以解决其对侵权内容所负担责任的范围与程度,同时也激励其在预防成本与侵权损失的合理衡量上做出理性选择[37]。从现有法律规定来看,网络服务提供者在著作权领域的注意义务主要集中于内容的传播环节,一是防止用户通过网络服务扩大侵权作品传播,二是预防网络服务提供者对侵权作品的传播形成激励。对生成式人工智能服务提供者而言,二者均处于技术不能状态。《信息网络传播权保护条例》第23条,的确对“提供内容服务的网络服务提供者”提出注意义务要求⑪。但是,AIGC经过充分训练后,是依据用户给出的指令输出最终内容,它仅包含作为训练数据的元素及特征,与训练数据始终保持差异。用户无法在使用AIGC网络服务过程中,故意或放任其对他人著作权的侵犯,生成式人工智能服务提供者的注意义务难以履行。其次,注意义务关注内容的传播,并非因其忽略了内容的生成,而是根据过去的技术条件。这有一个假设的前提:“笔”在用户手上、“纸”在网络传播手中。著作权法对网络服务提供者课以注意义务,是一种管住“纸”的规制思路,缓解侵权作品大规模传播导致的作品创作抑制。而作为一种工具,生成式人工智能服务提供者的角色互换:它是一杆“笔”,使用该服务的用户成了“纸”。生成式人工智能服务提供者履行注意义务的“技术不能”只是表象,本质是用管住“纸”的办法,去管住“笔”,只会带来“因噎废食”的效果。另一方面,生成式人工智能服务提供者,无法完成“通知—删除”规则项下的多类必要措施。我国《民法典》第1195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,《信息网络传播权保护条例》第15条亦有相关规定。但AIGC技术是通过预训练捕捉到不同场景中不同单词间的关系,再通过不断重复词组的预测和输出,保持文本、图像等形式输出的连贯性和一致性。换句话说,AI生成内容并不会完整储存在某个数据库中,而是根据模型参数实时生成,即使生成内容构成侵权,由于这段内容未事先储存,生成式人工智能服务提供者也无法删除、屏蔽、断开链接。因此,生成式人工智能服务提供者如因拟制为网络服务提供者,径直适用《民法典》《信息网络传播权保护条例》之规定,很可能因未及时采取必要措施,而承担连带责任;如需通过法律解释,扩大“提供作品”“必要措施”之含义,或因文义差异过大,滋生非议。

⑪ 参见:《信息网络传播权保护条例》第23条,“网络服务提供者为服务对象提供搜索或者链接服务,在接到权利人的通知书后,根据本条例规定断开与侵权的作品、表演、录音录像制品的链接的,不承担赔偿责任;但是,明知或者应知所链接的作品、表演、录音录像制品侵权的,应当承担共同侵权责任”。

### （二）承认生成式人工智能服务提供者是一种全新的责任主体

生成式人工智能服务提供者直接适用避风港规则的种种困难，系因其既非网络服务提供者，亦非内容提供者，它是一种全新的主体类型。在部分学者看来，将生成式人工智能服务提供者视为一种全新主体，多少有些“削足适履”的意味[38]。自互联网技术诞生以来，搜索引擎、云计算等新技术层出不穷，但在避风港规则的适用问题上，均未完全脱离网络服务提供者的规范框架。这是因为上一代的互联网技术是以“信息有效发现”为核心，而新一代人工智能技术是以“信息有效重组”为核心，这一特征将AIGC与2014年、2017年以算法推荐、算法识别为核心的“小模型”区别开来，算法推荐等技术仍能在网络服务提供者的规范框架内解决。AIGC的不同之处在于，它是通用人工智能的第一个里程碑，是从“信息有效重组”迈向“信息有效创造”的早期阶段，为此，将生成式人工智能服务提供者视为全新的主体，其深层次原因是，法律需要为通用人工智能的到来，探索一套新的权利义务规范。

首先，避风港规则对网络服务提供者，主要关注“作品传播”的著作权侵权风险，而生成式人工智能服务提供者，应该关注“作品生成”著作权侵权风险。在避风港规则之下，网络服务提供者的责任分配，首先被推定为一种“不知情”的法律状态，而注意义务、“通知—删除”规则设立，核心就是用以推定网络服务提供者是否实质处于“知情而放纵”的状态，进行对其责任进行认定[39]。而生成式人工智能服务提供者则无法依“知情/不知情”推定其责任状态，因为对于是否存在侵犯著作权的情形，使用AIGC的用户是第一责任人。这里容易产生的误解是：一旦侵权责任转移至用户，生成式人工智能服务者，会因推定其“不知情”而放纵著作权侵权。误解的产生，源于容易把用户获得的网络服务，推定为单一责任主体提供。例如，在抖音，利用平台提供的AIGC生成视频，并在平台内进行分享，如果视频内容侵犯第三人著作权，一旦使用AIGC的用户是主要责任人，那么，抖音将不承担著作权侵权风险。这一认识是错误的。原因是用户在抖音生成AI视频并发布的行为，涉及两类主体的服务：一类涉AIGC，是由作为生成式人工智能服务提供者的抖音所提供，无需推定其对生成内容“知情”；另一类涉及内容分发，是由作为内容服务平台的抖音所提供的，其将可能因未履行AIGC生成视频的注意义务，而承担著作权的间接侵权责任。一旦在同一平台，两种服务的提供主体未能区分（如服务背后不存在独立法人或混同），则推定其为内容服务提供者，而非生成式人工智能服务提供者。换言之，生成式人工智能服务提供者视为全新的责任主体，反而跳脱出“避风港规则”的框架，令“生产”和“传播”环节进行有效的互动。

其次，作为一种全新的责任主体，对生成式人工智能服务提供者，应对推定的法律状态是“可知/不可知”。如果认为AIGC生成物侵权，用户是主要侵权行为人，这也意味着，用户首先要对自己的行为知情，才能承担相应责任。此时，生成式人工智能服务提供者，应当确保用户、第三人对AIGC生成物的相关信息处于“可知”的状态，至少要确保两点“可知”。

一是确为AIGC生成物。即使用AIGC生成的内容，根据不同使用场景，应当进行“明/暗”水印标识或有公开标识，以便用户主张特定权利或第三人知晓。这一点在2024年9月国家网信办发布的《人工智能生成合成内容标识办法（征求意见稿）》中已有体现。如未提供相关标识，使第三人对生成内容是否属于AIGC，处于“不可知”的状态，生成式人工智能服务提供者应承担间接侵权责任；如其已提供标识服务，被用户以某种技术手段加以破坏，应豁免生成式人工智能服务提供者相关责任，由用户承担。事实上，2024年初，美国版权局收集AI版权问题时（约10 000条），AI生成内容是否需标识或公开身份已是热议问题，该项措施是否应当成为法律义务，则是核心争议[40]，主要担忧是各类水印的添加会导致AIGC在不同场景变得“不可用”。例如，绝大部分商家都不会使用有水印

的AIGC图片销售商品。这主要涉及识别的阈值,即特定场景的标识,应当辅以一定查验手段显示。

二是AIGC生成内容的引用来源可知。如前所述,通知生成式人工智能服务提供者,在大模型内删除某项作品,在技术上是无法实现的。但如果要求,AIGC生成内容时,展示网络检索、作品检索的相关内容,却是可行的。目前在中国,通义千问、秘塔搜索等AI产品在提供服务时,已对生成内容提供来源引用,这帮助用户进行侵权风险的判断。目前,这一"来源可知"要求,用以文字生成较易满足,但用以图片、视频却更难。值得说明的是,对引用来源的"可知",不应要求引用来源均为版权作品或获得授权,仅提供有效检索链接即可。

**(三)"AI避风港"的规范构建及可行路径**

承认生成式人工智能服务者的独立主体地位,意味着两点:一是作为新的责任主体,生成式人工智能服务提供者,需要新的责任规则,不再将网络服务提供者适用避风港规则的各项义务悉数套用;二是生成式人工智能服务提供者,需要与避风港进行有效衔接。

首先,从落地层面看,在我国,构建"AI避风港"规则的最佳试验场,在《生成式人工智能服务管理办法》,而非《信息网络传播权保护条例》。一方面,生成式人工智能服务提供者,其所承担的责任不是单一的责任,而是涵盖版权责任、内容安全、数据安全、个人隐私等在内的复合型责任。"管理办法"是一个交汇点,将帮助生成式人工智能服务提供者,在AIGC内容的侵权问题上完成版权侵权与内容风险、数据安全以及个人隐私等相关制度要求的衔接。另一方面,《管理办法》规定了"内容生产者责任",却没有找到责任界定的切入口。自2023年《管理办法》颁布以来,生成式人工智能服务已完成备案4批,超200多个大模型⑫,却从未在现有司法实践、行政监管看到"内容生产者责任"的身影。原因在于,这一责任范围过于庞大,远远超过了当前生成式人工智能服务提供者的责任承担能力范围,严格施行只会带来"普遍性违法"的窘境,需要小切口。

其次,要构建一套"AI避风港"规则,至少需要三方面为其增设权利义务:第一,设立"AI训练免责"机制。把含有版权作品的数据集进行清洗、标注及预处理,并选择特定模型框架进行训练的行为,视为不构成著作权意义上对作品的使用。前文案例中,美国作协、纽约时报均以大模型是否能高度复现作品内容和作品风格,作为是否未经许可把版权作品,用以大模型训练侵权的直接证据。如果明确模型训练的技术处理,为了让机器阅读、学习,而非对不特定公众提供,其表现形式对人类既然无可读性,也无法正常获取。那么,大模型训练侵权行为证据链不再成立。事实上,这并非对版权保护的突破,而是纠偏,只是"把作品用以模型训练"行为将回到AIGC出现之前,回到"默许"之中。此外,对含大量作品数据集的使用,是否构成特定数据权益的侵害,应另行讨论。第二,设立"可知免责"规则。按现行法律,用户因特定内容构成著作权侵权,承担侵权责任,并无异议。真正的问题是,权利人以帮助侵权为由,起诉生成式人工智能服务提供者,责任应如何划分。以两个"可知"作为责任划分的依据,有两重效果:一是与用户主观意志的判定,如生成式人工智能服务提供者,已提供引用来源,则可推定用户对内容来源"知情",作为判定其主观意图的依据。二是与避风港规则的衔接,如生成式人工智能服务提供者,已提供水印或标识,则网络服务提供者/内容提供者均对该AI内容"知情"。至于是否针对AI作品的传播,设立新的注意义务,可在避风港规则框架下进一步讨论。事实上,目前美国部分企业已通过保障"知情权",作为侵权风险的抗辩。例如,Stability AI公司近期表示将修改《用户协议》中"数据库不得加入或退出"的规定,用户或者版权人可在"Have I Been Trained"网站上找到自己的作品。

⑫ 国家互联网信息办公室.关于发布生成式人工智能服务已备案信息的公告(2025年9月至10月)[R/OL].(2025-11-11)[2025-12-12]. https://www.cac.gov.cn/2025-11/11/c_1213284756.htm.

最后，设立“绕行免责”规则。以对AIGC模型的优化与重写，使其主动偏离训练数据⑬，可作为生成式人工智能服务提供者的抗辩理由。因为AIGC内容服务的技术方案，将极大决定内容风险责任所指向的是生成内容的结果性要求还是过程性要求。目前，对AIGC内容服务技术方案指导效果最好的，主要是依《管理办法》第17条，对具有舆论属性或者社会动员能力的生成式人工智能服务进行安全评估备案，并完成的“双新评估”⑭及“算法备案”⑮。这是因为生成式人工智能服务提供者迫切需要让应用上线。把“绕行免责”规则的要求，内化于“内容生产者责任”后，作为安全评估的一部分，将减少版权侵权风险，转化为AIGC内容服务技术方案之中。对内容责任风险的预防，一般认为，传统媒体采取的是“先审后发”的封闭式把关模式，而网络服务提供者是采取实时交互开放式传播方式，实行筛查式过滤模式[41]。

现有技术框架下，生成式人工智能服务提供者无法精细管控生成内容与其用户之间的链路，引发人们对其应用感到忧虑的实际上是生成式人工智能技术对网络服务市场已有秩序的潜在挑战。在内容生产者责任中引入间接侵权责任，将完成生成式人工智能服务提供者正在承担的复合型内容合规义务与侵权责任分配之间的制度衔接。这一方面能够激励生成式人工智能服务提供者主动积极地在AIGC内容生成和传播中履行注意义务，同时也能平衡过高的义务设置所产生的创新抑制效应。

## 结语

生成式人工智能服务提供者的义务完善是一个动态演进的过程。作为网络版权世界中全新的主体，生成式人工智能服务提供者的各项义务是否切实可行，需等待现实诉讼纠纷进行校验与调整。目前，生成式人工智能服务提供者的类型特征尚不清晰。AIGC产业发展依赖于大算力、大数据、大模型的良性循环和“飞轮效应”，这一新型循环在全球范围内仍不成熟，可预见的变化是它与过往的网络服务有显著区别，但变化的特征尚未浮现。法律规则无法在这样的条件下将其运转的关键环节，转化为法定义务的规制对象。从避风港规则的演进历史可以看到，网络服务提供者因重大技术进步改变了侵权法所预设的前提，导致权利人和相关公众在预防侵权方面的相对成本发生变化，进而推动了规则变革[42]。生成式人工服务提供者与避风港规则的互动，仍可以留给尚待发生的法律纠纷，这将会为日后我国人工智能专门立法研究，提供有益的制度经验和贡献。

生成式人工智能技术在内容生成上的突破将深刻改变数字信息等众多产业的未来发展与竞争格局，大幅提升知识获取和利用的效率，从而在千行百业中引发新的“效率革命”。知识产权很容易成为市场应对新技术冲击的工具，从而造成对创新和竞争的抑制。如果急于对AIGC可能引发的版权问题进行全面围堵，可能导致监管不得不选择性执法的局面，同时，违法行为的发现成本过低，会催生大量以投诉、举报牟利的黑灰产群体，限制AIGC产业的发展。在这样的局面下，版权责任的分配甚至是知识产权应当坚守鼓励创作、支持创新的根本目标。为此，应当坚持我国著作权法、民法典关于网络侵权责任的相关规定，创造性地适用避风港规则，作为内容使用合法性判定标准，平衡

⑬ See Berns S, Colton S. Bridging generative deep learning and computational creativity [R/OL]. (2020) [2024-05-20]. https://computationalcreativity.net/iccc20/papers/164-iccc20.pdf.

⑭ 目前，企业在申报双新评估时，需要提交的申报材料动辄几百页，例如《互联网新技术新应用安全评估申请表》第61页，要求填写的附件达200-300页。

⑮ 根据《生成式人工智能服务管理办法》第六条规定，利用生成式人工智能产品向公众提供服务前，应当按照《具有舆论属性或社会动员能力的互联网信息服务安全评估规定》向国家网信部门申报安全评估，并按照《互联网信息服务算法推荐管理规定》履行算法备案和变更、注销备案手续。

著作权人、生成式人工智能服务提供者与用户的权利义务关系,通过推动大模型开发应用推动知识传播。要洞察避风港规则在新技术的适用效果,需要理解法律所面对的问题的交互性,这有待多方利益主体在诉讼进程中进行理性对话。在技术与法律的互动中,技术不仅对法律产生影响,也可以凸显法律的已有特征,帮助人们看到法律中稳定的方面。从历史的眼光来看,正是避风港规则对平台灵活的责任设置,才造就了今日美国极具竞争力信息服务行业。尽管避风港规则要良好地适用于AIGC内容服务提供者,目前缺乏经验理性的支持,但从制度弹性来看,避风港规则将是解决AIGC版权冲突的重要制度框架。

**参考文献:**


[1] 刘文杰. 网络服务提供者的安全保障义务[J]. 中外法学,2012(2):395-410.

[2] 沈伟伟. 技术避风港的实践及法理反思[J]. 中外法学,2023(4):906-922.

[3] Decherney. Hollywood's copyright wars: from Edison to the internet[M]. Columbia University Press, 2013: 169.

[4] 王迁.《信息网络传播权保护条例》中"避风港"规则的效力[J]. 法学,2010(6):128-140.

[5] 李明德. 美国知识产权法[M]. 法律出版社,2003:441.

[6] 吴汉东. 侵权责任法视野下的网络侵权责任解析[J]. 法商研究,2010(6):28-31.

[7] 熊文聪. 避风港中的通知与反通知规则:中美比较研究[J]. 比较法研究,2014(4):122-134.

[8] 梁志文. 网络服务提供者的著作权责任:文本解释与比较分析[J]. 法治研究,2011(2):74-82.

[9] 王利明. 论网络侵权中的通知规则[J]. 北方法学,2014(2):34-44.

[10] 崔国斌. 网络服务商共同侵权制度之重塑[J]. 法学研究,2013(4):138-159.

[11] 李晓阳. 重塑技术措施的保护:从技术措施保护的分类谈起[J]. 知识产权,2019(2):69-80.

[12] 闫宇晨. 公共图书馆电子借阅服务著作权侵权风险与对策研究[J]. 国家图书馆学刊,2023(2):3-14.

[13] 刘家瑞. 论我国网络服务商的避风港规则:兼评"十一大唱片公司诉雅虎案"[J]. 知识产权,2009(2):13-22.

[14] 吴汉东. 论网络服务提供者的著作权侵权责任[J]. 中国法学,2011(2):38-47.

[15] Cheung A, Weber R H. Internet Governance and The Responsibility of Internet Service Providers [J]. Wisconsin International Law Journal, 2008, 22: 403-478.

[16] 邹晓玫. 网络服务提供者之角色构造研究[J]. 中南大学学报(社会科学版),2017(3):63-69.

[17] 梁志文. 网络服务提供者的版权法规制模式[J]. 法律科学(西北政法大学学报),2017(2):100-108.

[18] Weller M. 25 Years of Ed Tech[M]. AU Press, 2020: 16.

[19] 钟晓雯. 算法推荐网络服务提供者的权力异化及法律规制[J]. 中国海商法研究,2022(4):63-72.

[20] 李安. 智能时代版权"避风港"规则的危机与变革[J]. 华中科技大学学报(社会科学版),2021(3):107-118.

[21] 喻国明,李钒. 内容范式的革命:生成式AI浪潮下内容生产的生态级演进[J]. 新闻界,2023(7):23-30.

[22] Nagmias Y, Perel M. The oversight of content moderation by AI: impact assessments and their limitations[J]. Harv. J. on Legis., 2021, 58: 145.

[23] 王迁. 再论人工智能生成的内容在著作权法中的定性[J]. 政法论坛,2023(4):16-33.

[24] 李雨峰. 从写者到作者:对著作权制度的一种功能主义解释[J]. 政法论坛,2006(6):88-98.

[25] 林秀芹,刘文献. 作者中心主义及其合法性危机:基于作者权体系的哲学考察[J]. 云南师范大学学报(哲学社会科学版),2015(2):83-92.

[26] 林秀芹. 人工智能时代著作权合理使用制度的重塑[J]. 法学研究,2021(6):170-185.

[27] 张玉敏,易健雄. 主观与客观之间:知识产权"信息说"的重新审视[J]. 现代法学,2009(1):171-181.

[28] 熊琦. 论"接触权":著作财产权类型化的不足与克服[J]. 法律科学(西北政法大学学报),2008(5):88-94.

[29] Radford A, Kim J W, Hallacy, at al. Learning transferable visual models from natural language supervision [C]//

International Conference on Machine Learning. PMLR, 2021: 8748-8763.
[30] Guadamuz A. Do androids dream of electric copyright Comparative analysis of originality in artificial intelligence generated works[J]. Intellectual Property Quarterly, 2017.
[31] 龙文懋. 人工智能法律主体地位的法哲学思考[J]. 法律科学(西北政法大学学报),2018(5):24-31.
[32] Somepalli G, Singla V, GoldblumO M, et al. Diffusion art or digital forgery Investigating data replication in diffusion models[EB/OL]. (2022-12-12)[2026-03-14]. https://doi. org/10. 48550/arXiv. 2212. 03860.
[33] 司晓. 奇点来临:ChatGPT时代的著作权法走向何处:兼回应相关论点[J]. 探索与争鸣,2023(5):79-86,178-179.
[34] Kaminski M E. Authorship, disrupted: AI authors in copyright and first amendment law[J]. UCDL Rev. 2017, 51: 589.
[35] 刘晋名,艾围利."避风港规则"的法律适用困境及消解路径[J]. 南京社会科学,2020(8):95-99,116.
[36] 徐伟. 论生成式人工智能服务提供者的法律地位及其责任:以ChatGPT为例[J]. 法律科学(西北政法大学学报),2023(4):69-80.
[37] 刘艳红. 人工智能法学研究的反智化批判[J]. 东方法学,2019(5):119-126.
[38] 司晓. 网络服务提供者知识产权注意义务的设定[J]. 法律科学(西北政法大学学报),2018(1):78-88.
[39] 刘艳红. 人工智能的可解释性与AI的法律责任问题研究[J]. 法制与社会发展,2022(1):78-91.
[40] 李雨峰,邓思迪. 互联网平台侵害知识产权的新治理模式:迈向一种多元治理[J]. 重庆大学学报(社会科学版),2021(2):155-165.
[41] Samuelson P. U. S. Copyright Office's questions about Generative AI[J]. Communications of the ACM, 2024(3): 25-28. DOI: 10. 1145/3637627.
[42] 崔国斌. 论网络服务商版权内容过滤义务[J]. 中国法学,2017(2):215-237.

# Can generative AI service providers apply the safe harbor rules: an exploration based on the suitability of infringing subjects

Li Xiaoyang

(*Law School*, *Zhejiang Gongshang University*, *Hangzhou 310018*, *P. R. China*)

**Abstract:** The deep involvement of generative AI (GenAI) in content production has severely impacted the existing internet copyright legal framework, making the allocation of infringement liability for GenAI service providers a core challenge that urgently needs to be addressed. As the cornerstone of balancing technological innovation and copyright protection in the internet era, the safe harbor rules are currently facing a severe test regarding their applicability to this new subject. This paper conducts an in-depth exploration of this issue from the perspective of the suitability of infringing subjects. From the history of institutional evolution, the application of the safe harbor rules highly depends on the suitability of subjects, having undergone the separation of internet service providers (ISPs) and internet content providers (ICPs), as well as the rise of content service providers. However, GenAI technology has achieved a leap from discovering information to reorganizing information and even creating information. GenAI service providers directly participate in the content production process, breaking the neutrality premise of traditional ISPs that do not participate in content creation. Meanwhile, because its content generation is subject to user instructions, and the cross-modal, large-scale production makes it difficult to trace the source of the content, it cannot be classified as an ICP within the meaning of copyright law. Therefore, the traditional dual-subject framework of internet service providers and content providers has fallen into a dilemma of applicability in the GenAI era. If GenAI service providers are forcibly regarded as traditional ISPs and the existing safe harbor rules are directly applied, it will plunge them

into a dilemma of compliance. On the one hand, constrained by the technological essence of GenAI as a pen rather than paper, it cannot fulfill the duty of care to prevent the dissemination of infringing works. On the other hand, since AI-generated content is generated in real-time based on parameters rather than stored in a database in advance, service providers are technically unable to take necessary measures such as disconnecting links or deleting content under the notice and takedown rules. Therefore, the key to breaking the deadlock lies in abandoning the patching approach and legally recognizing GenAI service providers as a brand-new, independent liable subject. In determining their infringement liability, the focus should shift from the traditional safe harbor rules' presumption of actual knowledge (informed or not) regarding dissemination to the standard of knowability (knowable or not) regarding content generation. Based on this, this paper proposes that a new AI safe harbor rule exclusively tailored for such subjects should be scientifically introduced and reconstructed within the framework of content producer liability under the Interim Measures for the Management of Generative Artificial Intelligence Services. Specifically, it includes three core mechanisms: First, establishing an AI training exemption mechanism, explicitly treating the technical processing behavior of using lawful works for model training as non-infringing use. Second, establishing a knowable exemption rule, requiring providers to ensure that the generated content and its sources are in a knowable state by adding visible and invisible watermarks and providing source retrieval links, and using this as a basis for liability apportionment and defense. Third, establishing a circumvention exemption rule, encouraging service providers to proactively deviate from training data through model optimization, rewriting, and security assessments to evade infringement risks. The construction of this set of brand-new rules aims to effectively maintain the institutional balance of the existing copyright order while safeguarding the continuous innovative development of GenAI technology.

**Key words:** generative AI; generative AI service providers; internet service providers; safe harbor rules; copyright liability

(责任编辑 刘 琦)