

生成式人工智能数据训练的版权侵权风险应对研究

李婧瞳

南京理工大学知识产权学院, 江苏 南京

收稿日期: 2025年11月3日; 录用日期: 2025年12月3日; 发布日期: 2025年12月15日

摘 要

在生成式人工智能发展的过程中, 数据训练对最终生成结果的效果起到了关键作用。现阶段数据训练未经授权而使用海量受保护的作品, 版权侵权风险显而易见, 而学界及业界并未就数据训练的法律定位形成一致观点。通过对学界观点的分析比较和对域外立法经验的借鉴, 在我国合理使用制度框架下提出构建数据训练的合理使用规则和相应的配套措施, 能够在兼顾著作权人利益的同时保障人工智能产业的发展, 化解数据训练阶段的版权侵权风险。

关键词

生成式人工智能, 数据训练, 版权侵权风险

A Study on Addressing the Copyright Infringement Risks in the Data Training of Generative Artificial Intelligence

Jingtong Li

School of Intellectual Property, Nanjing University of Science and Technology, Nanjing Jiangsu

Received: November 3, 2025; accepted: December 3, 2025; published: December 15, 2025

Abstract

The development of generative artificial intelligence (AI) critically depends on large-scale data training, which fundamentally determines the efficacy of the outputs. However, the prevalent practice of utilizing massive amounts of copyrighted works without authorization for training purposes presents evident infringement risks. Currently, there is no consensus within either academia or

industry regarding the legal status of such data training activities. This study addresses this issue through a comparative analysis of scholarly perspectives and an examination of extraterritorial legislative approaches. It proposes a framework within China's existing fair use system to mitigate copyright risks at the data-training stage. The core argument is that establishing tailored fair use rules, supplemented by corresponding supporting measures, can effectively reconcile the interests of copyright holders with the developmental needs of the AI industry. This approach aims to foster innovation while providing a legal pathway to resolve the inherent copyright infringement risks associated with data training.

Keywords

Generative Artificial Intelligence, Data Training, Copyright Infringement Risks

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 问题的提出

近年来, 基于算法的发展和大数据的应用, 人工智能技术有了显著的进步, 而随着技术的更新和迭代, 人工智能的发展迎来了高峰。根据中国信息通信研究院发布的《人工智能发展报告(2024)》全球人工智能产业保持高速增长, 而根据 IDC 预测, 2024 年全球人工智能产业规模将达到 6233 亿美元, 同比增长 21.5%。¹ 人工智能已经渗透进人们日常生活的各个领域, 不仅包括文学艺术, 也包括气象、材料等自然科技。

随之而来的, 人工智能的迅猛发展给知识产权制度带来了许多新的挑战 and 难题。最初学界关注的焦点主要集中在人工智能生成内容的属性及生成内容的著作权归属问题, 而随着人工智能大模型的进一步发展, 其研发阶段数据训练的有关问题逐渐引发关注。人工智能大模型迅速更新迭代的背后, 高质量、大规模、多样化的数据集已经成为大模型发展的关键战略要素, 训练数据决定了模型能力的上限。在数据训练阶段存在着各种各样的风险, 例如知识产权侵权风险、数据偏见风险、虚假信息风险及个人信息泄露风险等, 而在其中最引人注目的当属知识产权侵权风险。海量用于训练的数据均是从互联网直接爬取, 其中大部分的内容实际上在著作权法的保护范围之内。但对这些数据的使用并未获得权利人的授权, 数据使用者也未向权利人支付合理的对价, 众多著作权纠纷也由此产生。例如在美国, 喜剧演员兼作家萨拉·西尔弗曼以及其他两位畅销书作家联合起诉 Meta 和 OpenAI, 指控这些公司通过使用他们的著作来训练人工智能语言模型侵犯了他们的版权;² 在德国, 一摄影师指控 LAION 公司提供的用于对生成式人工智能进行训练的数据集侵犯了其享有的涉案作品的著作权;³ 在我国“奥特曼案”中, 原告圆谷制作株式会社请求法院判令被告停止侵权的同时, 还请求将案涉奥特曼物料从被告的训练数据集中删除。⁴ 上述案件表明对数据训练的法律定位已是一个全球共同关注的问题, 而各方权利人在实践中也迫切需要明确的指引。

¹ 人工智能发展报告(2024) [R]。中国信通院, 2024, <https://www.caict.ac.cn/kxyj/qwfb/bps/202412/P020241210548865982463.pdf>, 2025 年 1 月 11 日。

² Sarah Silverman is suing OpenAI and Meta for copyright infringement, <https://www.theverge.com/2023/7/9/23788741/sarah-silverman-openai-meta-chatgpt-llama-copyright-infringement-chatbots-artificial-intelligence-ai>.

³ LG Hamburg 310 O 227/23。

⁴ 广州互联网法院(2024)粤 0192 民初 113 号民事判决书。

目前,国外的立法对这些问题已有所涉及。例如,日本的著作权法中明确将并非为了使自己或他人享受作品所表达的思想或情感的、包含数据训练的“计算机信息处理行为”囊括在合理使用的范围内;欧盟通过《数字单一市场版权指令》,明确对文本和数据挖掘设定了例外规定,允许在特定情况下合法使用作品进行数据训练,还明确规定版权保护程度,并设置提供商的“数据透明”义务^[1];美国则在“四要素”框架下提出“转换性使用”,将人工智能数据训练纳入合理使用的情形。而在我国,目前立法中对于数据训练的侵权风险应对主要体现在《生成式人工智能服务管理暂行办法》第七条第一款和第二款,⁵规定了数据的来源应当合法,且不得侵害知识产权,但是该条的规定较为笼统,并未提供明确的应对路径。鉴于数据训练是现阶段人工智能发展的基础,而数据训练的法律定性将直接影响到人工智能产业的发展,因此近年来越来越多的学者对有关数据训练的法律问题展开探索和研究,并提出了不同的观点。鉴于此,持续进行对大规模数据训练的版权侵权风险的研究,厘清有关逻辑,提出应对策略,对于平衡各方利益、促进人工智能产业的发展具有重要意义。

2. 数据训练侵权风险观点之比较

由于数据训练行为的法律定位尚未明确,理论界对该问题展开了广泛的讨论,并提出了不同的版权风险应对策略。部分观点认为数据训练行为并非侵权行为,因此更无侵权风险;另一部分观点则认为侵权风险客观存在,并提出了不同的应对策略。接下来本文将对不同的观点进行介绍和分析,并讨论出较为可行的应对模式。

2.1. 否认侵权风险的存在

持有该观点的学者认为,对人工智能进行数据训练的本质是机器学习,而输入阶段和训练阶段是非表达型机器学习,即使使用受著作权保护的作品用于数据训练,该种行为也因“非作品性使用”而免于落入著作权保护范围,更不存在著作权侵权行为^[2]。但是不可否认的是,在数据训练的过程中势必会涉及对受保护的作品进行存储和复制的行为。即使提出“非表达性使用”的代表性学者 Matthew Sag 在其论述中也不止一次地承认数据训练涉及“复制行为”“复制”以及“制作完整的文字复制件”^[3]。尽管数据训练多数情形下可被解释为“非表达性使用”进而被主张不构成侵权,但这并不意味着这种使用完全不受著作权法的约束。著作权法对于作品的保护是多方面的,即使某种使用行为不属于对表达的直接使用,也可能在其他方面涉及著作权问题。例如,即使数据训练本身不构成侵权,但如果基于训练结果生成的内容与原作品存在实质性相似,不但可能会引发著作权侵权问题,而且此种情况下很难证明其没有对用于训练的作品进行表达性使用。如果复制行为没有经过版权人授权就进行使用,并且生成了模仿原作品的风格的生成内容,那么这种行为就符合《著作权法》上“复制”行为的基本特征,亦可能就直接侵犯了权利人的著作权^[4]。若著作权法对人类阅读和机器阅读在规则适用上采用双轨制,从长远看会引发一些根本性的问题,动摇著作权法的本质和目的,可能变相鼓励非法复制行为在机器人经济中的流通,同时也会对人类阅读和创作产生一定的负面影响。机器人读者应当同人类读者一样,其产生的复制等行为即使是非表达性的,也并不会当然因为其机器人使用身份而直接认定其不构成侵权。

2.2. 认为侵权风险存在

目前较为普遍的观点是认定生成式人工智能数据训练的过程中存在著作权侵权问题,但是考虑到生

⁵《生成式人工智能服务管理暂行办法》第七条:生成式人工智能服务提供者(以下称提供者)应当依法开展预训练、优化训练等训练数据处理活动,遵守以下规定:(一)使用具有合法来源的数据和基础模型;(二)涉及知识产权的,不得侵害他人依法享有的知识产权;(三)涉及个人信息的,应当取得个人同意或者符合法律、行政法规规定的其他情形;(四)采取有效措施提高训练数据质量,增强训练数据的真实性、准确性、客观性、多样性;(五)《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律、行政法规的其他有关规定和有关主管部门的相关监管要求。

成式人工智能的发展和更新的需要，直接认定侵权的模式基本已被排除，目前学界与实务界更加关注的问题是对于侵权行为如何认定的问题，即如何为数据训练提供合理的抗辩。

2.2.1. 构建法定许可制度

该观点认为，数据训练过程中的侵权风险是存在的，但是可以通过构建法定许可制度进行应对。也就是说，法律允许在未经著作权人授权的情况下将其作品用于生成式人工智能的数据训练，但是使用者仍需支付报酬。这种方式在一定程度上对著作权人的权利进行限制，给予使用者在法律规定的框架内利用受保护的作品开展数据训练的权利，旨在平衡著作权人与使用者的利益，缓和双方矛盾。但是，此种路径并未解决实际运用过程中可能会产生的诸多现实困境。

其一，数据训练所涉及的作品来源复杂，作品的种类繁多，包括但不限于文本、图像、音乐、视频等多种形式。这些作品可能来自不同的创作者、不同的平台，且其著作权归属和授权情况各不相同。例如，一些作品可能已经进入公有领域，而另一些作品则受到严格的著作权保护。这种复杂性使得在数据训练过程中，追溯不同作品的相应著作权人变得十分困难。其二，同一受保护作品可能被不同模型开发者用于数据训练，而基于各自经济实力和市场预期，这些开发者所愿意负担的用于支付报酬的成本往往会有较大差别。对于中小企业而言，与头部企业相比其自身综合实力不足，因而在竞争中常常处于劣势，甚至可能因无力承担高昂报酬费用而放弃使用关键作品进行数据训练，这种情况不利于行业的创新和多元化发展。其三，由于作品价值在不同应用场景、市场环境和技术背景下具有高度不确定性，规定一个既能合理体现著作权人创作价值，又能为使用者所接受的报酬数额极为困难。此外，要求使用者与著作权人协商并达成合意在现实中缺乏可行性，这种方式会耗费大量时间与精力，极大地降低数据获取效率，不利于生成式人工智能的发展。

综上所述，法定许可仅解决了授权这一步的问题，而对后续实际应用过程中可能引发的问题考虑并不周延，因此如果贸然设置法定许可制度，极有可能引发相反效果，进一步加剧权利人与使用人之间的矛盾冲突，滋生更多法律纠纷和社会问题。

2.2.2. 构建合理使用制度

该观点认为，数据训练中的侵权风险是存在的，但是可以通过构建合理使用制度进行应对，即将受保护的作品用于数据训练的行为属于合理使用，既不需要获得著作权人的许可，也无需向权利人支付报酬。该种观点一方面解决了作品来源合法性的问题，另一方面解决了权利人报酬支付困难的问题。合理使用原则避开法定许可，否认数据训练特定阶段的直接侵权责任，能够更深层次地减少数据训练各利益方的摩擦，最大化社会公益。但是，此种方法也招致了另一种担忧，即人类作者无偿地为生成式人工智能模型训练提供了海量、优质的版权作品，模型的成功不但没有“反哺”人类，反而可能压缩人类作者的生存空间[5]。

此种担忧并不是毫无道理，但是生成式人工智能的发展已成为不可抗拒的趋势，模型没有“反哺”人类的观点也值得推敲。事实上随着大模型的一次又一次迭代更新，生成式人工智能的应用已不局限于创作一幅图画或生成一段文字，其在自然科学、医学领域的发展也颇为瞩目，为人类日常生活的许多方面带来了便利，这何尝不是大模型服务于人的体现。但生成式人工智能在机器学习中稍有不慎就会触碰到“个人利益”的边界，若只是机械地保护原作者的权益，显然不利于该技术在“公共利益”方面带来的效益，有悖于“利益平衡原则”[6]。并且著作权不能被理解为一种对作品的全部价值进行绝对保护的权力，应当明确坚持“利益适度原则”，寻找能够促进文化繁荣发展、增进社会福利的最佳利益平衡点，塑造合理的著作权边界[7]。生成式人工智能的数据训练阶段是一个隐秘的过程，也不会直接地产生经济收益，因此一般不会对权利人的财产权利造成损害。而要求训练数据进行版权披露则能够确保作者的精神

权利受到尊敬和认可[8]。在实际应用过程中,可以采取多种手段降低对著作权人的权利的侵害。

经过上述分析与比较,合理使用制度的构建不仅能够平衡各方利益,而且具备现实可行性,是对生成式人工智能数据训练版权困境的最优回应。

3. 数据训练合理使用的正当性证成

我国《著作权法》中对合理使用的规定主要集中在第二十四条,⁶在“三步检验法”的框架下对合理使用的情形进行了明确的列举,并设置第一款第十三项“法律、行政法规规定的其他情形”的弹性条款,为将数据训练列为合理使用的新情形提供了制度接口。接下来,本文将从多个方面对数据训练合理使用的正当性进行论证。

3.1. 生成式人工智能发展的需要

前文提到,数据训练影响着生成式人工智能的能力和应用潜力,因此对于用于数据训练的作品,在保证一定数量的同时还需要注重质量。在合理使用的框架之下,可以最大限度地为生成式人工智能提供足够多且足够好的作品,从而保障其健康发展[9]。优质的训练数据不足将导致算法偏见,甚至在输出端生成带有歧视性、侮辱性的观点和偏见,这不利于人工智能的长期发展。在现实中,因输入不良数据而导致输出歧视性内容的案例早在2016便有发生,微软的聊天机器人“Tay”在学习了一些煽动性和种族主义的数据资料后因发表不恰当的帖子而被迫停止服务。由此可见,训练数据的质量和丰富性对于生成式人工智能具有重大意义,我国《生成式人工智能服务管理暂行办法》第七条也规定,生成式人工智能服务提供者应当采取有效措施提高训练数据质量,增强训练数据的多样性。⁷

3.2. 作品使用许可市场失灵问题的存在

在生成式人工智能数据训练中,市场失灵的客观存在是认定合理使用的关键理由,而市场失灵主要体现在以下几个方面。

首先,合理定价困难且交易成本过高。例如,相关数据显示文本到图像生成模型 Stable Diffusion 使用非营利组织 LAION 收集的三个大型数据集进行训练,包括 58.5 亿个图像-文本对。自然语言处理模型 GPT-3 从 45TB 原始数据中过滤出 570GB 数据进行训练,这些数据包括网站抓取数据集、网页文本数据集、图书语料库和英语维基百科,共设置了 1750 亿个参数数据训练阶段需要海量数据[10]。与这些庞大的数据相对应,权利人的数量也相当庞大且分散,要求使用者寻找相应的权利人,再分别商讨并支付许可费用,将会产生高昂的交易成本。哪怕诉诸集体管理组织,也无济于事,因为并非每一作品的种类均有对应的集体管理组织。且即使建立起了这样的机制,执行成本也会非常高,从人工智能企业收到的费用中的一大部分都会被用于支付给集体管理组织,著作权人收到的费用将非常有限,无法向他们提供有意义的经济支持。

其次,权利人的许可意愿有限且存在差异。权利人普遍会对自己的作品被滥用而担忧,并且为了保持一定的竞争优势,无论是企业还是权利人等主体都希望锁住自己生产或持有的信息,也就导致权利人的许可意愿有限。此外权利人之间的许可意愿也存在差异,许可意愿较低的权利人往往认为自己作品的价值较高。许可意愿的有限性一方面限制了大模型的丰富性和全面性,另一方面也将导致大模型缺乏高质量数据而生成偏颇或歧视性内容。

⁶《中华人民共和国著作权法》第二十四条:在下列情况下使用作品,可以不经著作权人许可,不向其支付报酬,但应当指明作者姓名或者名称、作品名称,并且不得影响该作品的正常使用,也不得不合理地损害著作权人的合法权益……(十三)法律、行政法规规定的其他情形。前款规定适用于对与著作权有关的权利的限制。

⁷同上注 4。

最后,生成式人工智能具有公益性。生成式人工智能的发展涉及公共利益,其影响远超著作权领域,因此不宜仅由市场调节。在医疗、教育等关键领域,生成式人工智能也发挥着重要作用,可以增加社会福利。此外,2023年共有149个基础模型发布,其中65.7%是开源的。也就是说,人工智能开发者将源代码、训练数据、模型权重等关键资源对公众开放,这一举措促进了技术的共享和创新。并且开源软件通常免费或低成本提供,这对于初创公司、教育机构和个人开发者而言具有重要意义。即使是出于公共利益的需要,也应当尽量减少数据训练环节的阻碍,使得生成式人工智能大模型能够更好地服务社会。

3.3. 合理使用不会影响作品的正常使用

生成式人工智能的创作呈现“学习-创建规则-创作”的三层结构,第一层“学习”和第二层“创建规则”是内在的、隐含的,原则上不与公众接触^[11]。众所周知在人类创作的过程中,参考借鉴他人作品的行为十分普遍,但这并不意味着最终产出的成果会与参考借鉴的作品构成实质性相似。数据训练阶段也可以视作为人工智能参考借鉴的准备阶段,用于训练的作品旨在为人工智能提供学习资料,但对其日后输出内容的情况则无法预测。而人工智能将数据训练的过程可以视为一个“黑箱”,著作权法只需要关心黑箱输出了什么,而在黑箱内部发生了什么,作品是否被使用和如何被使用,则没有进行探究的必要性和可能性^[12]。此外,“转换性使用”也能为合理使用规则提供一种新的可能。“转换性使用”即指行为人对他人享有合法著作权的作品予以利用时,其行为要么具备作品再生产的功能属性,要么在作品内容的使用模式、目的指向或功能效用层面,与著作权人对原作品的利用形态形成本质性差异的使用行为范式。在适用合理使用一般规则在内的侵权认定规则时,依据使用行为的“目的性/功能性转换”考察,裁判者可以将上述符合公共政策目标的合目的性价值引入至规则体系以内,更有必要做到既充分重视著作权人的相关市场利益,又兼顾“公民基本自由以及科学进步和文化、社会或经济发展等方面的其他公共利益”^[13]。

正是由于数据训练阶段的隐秘性,确保了作品的使用仅限于技术开发和模型训练的目的,并不会被传递给公众,因此作品本身的完整性不会遭到破坏,其艺术价值也不会因此贬损。而合理使用规则,更加符合当下人工智能快速发展的社会现状和公共利益要求。数据训练的目的也不是产生原作品的替代品,因此不会影响原作品的市场销售或者其他使用行为。

3.4. 合理使用不会不合理地损害著作权人的合法权益

目前人工智能的发展尚处于弱人工智能的阶段,生成式人工智能并不以复制、再现现有作品为目的。即使最终生成的内容与现有作品构成实质性相似,也应当从输出端对其进行规制,如明确使用者责任和侵权损害赔偿责任。此外,著作权法的重要目的是保障作者从作品利用中获得经济报酬,著作权人获取作品利益的方式是控制作品复制和传播,并从作品的每一次复制和传播中收取相应许可费用^[14]。但正如前文所述,因为数据训练阶段的内在性和隐含性并不会导致作品的复制件在市场上广泛地传播,所以作者的财产权益并不会受到损害。并且只有在输出端才有可能出现相对于著作权人产生市场替代效果的内容,才符合著作权原本的激励范围^[15],数据训练阶段不会与著作权人形成市场竞争关系。生成式人工智能训练作品旨在对作品进行分析并学习,其最终生成的内容往往是经过算法和模型加工后的全新表达,这种表达可能反而会激发人们的一些创新灵感,拓展新的创作思路。

除以上分析的四点之外,生成式人工智能作为新质生产力的一种,对于推动我国的高质量发展具有重要意义。而立足国际视野,人工智能发展浪潮席卷全球,欧盟、美国、日本等均直接或间接地通过设置合理使用条款为人工智能数据训练提供制度保障,为其发展提供较为宽松的法律环境,以提升自身的国际竞争力。因此,数据训练的合理使用制度无疑为当下最为契合现实的选择。

4. 数据训练合理使用规则的构建

我国《著作权法》的合理使用条款已规定了较为开放的兜底性条款，因此可以在现有的框架下，借鉴国际经验，针对生成式人工智能构建新的合理使用规则。此外也应当认识到，合理使用条款固然能为生成式人工智能的发展提供相应的保障，解决数据训练中的版权问题，但是“宽泛的版权例外”会使权利人陷入弱势地位，并不能获得“相应保护”。技术的发展不能竭泽而渔，因此仍要坚持以人为本的价值导向和利益平衡原则构建数据训练合理使用规则，以保障数据训练的顺利进行，促进技术创新的同时切实维护著作权人的利益。

4.1. 数据训练主体和范围的规定

如前文所述，目前世界上不少国家和地区已将人工智能训练数据纳入合理使用制度，但是不同国家和地区对人工智能数据训练合理使用的主体的规定并不相同。欧盟《单一数字版权法》第三条规定了科研机构和文化遗产机构数据挖掘行为的例外，但是并没有对其他主体进行排除。⁸英国则对受版权保护的材料进行合法的数据挖掘限定于非商业用途和非商业机构，但英国知识产权局已对扩大数据挖掘版权例外进行磋商，允许所有用途数据挖掘的许可或例外情形将适用于更多的人工智能程序的训练。⁹随着生成式人工智能应用场景的增加，数据训练的主体也在拓展。合理使用制度关注的重点在于权利人的利益与社会公共利益的平衡，因此在生成式人工智能数据训练并不会影响作品使用、不会不合理地损害权利人的合法权益的情况下，应当基于社会公共利益的考量和人工智能创新发展的需要，避免对数据训练的主体过度限制。

与数据训练主体宽松的规定规则不同，对于数据训练合理使用的范围需要采取较为严格的限制。基于前文的分析，数据训练合理使用规则的重要正当性基础之一是数据训练阶段并不会与公众接触，在对范围进行限时时依旧需要贯彻这一基础。首先，合理使用规则仅适用于训练阶段，而不包括“数据采集”和“内容生成”阶段，因为数据采集阶段并不是对数字化作品的直接利用，不符合作品合理使用的前提条件。而在内容生成阶段，如果生成的内容与受保护的作品构成实质性相似，同样不能利用合理使用制度给予豁免。其次，数据训练的行为应当限定为复制、翻译等发生在人工智能内部神经网络供其“学习”的行为，而一旦行为涉及将用于训练的作品向公众公开，则是对“不与公众接触”基础的打破，自然无法适用合理使用规则。就使用目的而言，仅当使用作品直接服务于模型训练构建数据集时，方可纳入非欣赏性使用范畴；而以对外的商业交易为目的制作、销售数据集的行为因涉及数据库作品著作权人的排他性权利，应当被排除在该规则之外^[16]。再次，对合理使用规则予以精细化设计，需区分商业性与非商业性主体。但是在当前生成式人工智能的研发迭代以商业主体为主的情况下，仍可依托非欣赏性使用的判断依据，明确商业主体若未使用作品表达的思想情感，或仅在特定情形下少量使用作品表达，可认定为合理使用。最后，不宜为著作权人设置退出机制。若允许版权人声明权利保留，从而排除数据训练的合理使用，该合理使用条款将形同虚设。综上所述，数据训练合理使用条款可以设置为“以数据训练为目的使用他人已发表的作品，为合理使用。”

4.2. 生成式人工智能提供者的义务

著作权法中的“合理使用”制度本旨在调和作品保护与公共利用的价值冲突，但该制度的适用若忽视对创作者权益的审慎评估，易造成权利保护的失衡。因此，在照顾社会公众对作品传播、二次创新与

⁸曹建峰，史岱汶。欧盟《单一数字市场版权指令》全文中文翻译[EB/OL]。腾讯研究院，2019-03-26 [2025-1-12]。

<https://www.secrss.com/articles/9879>

⁹中国保护知识产权网。英国知识产权局发布关于人工智能发明和版权的最新指南[EB/OL]。2022-07-08[2025-1-12]。

<https://ipr.mofcom.gov.cn/article/gjxw/gbhj/ozqt/yg/202207/1971658.html>

知识共享的需求的同时，也要保障创作者基于智力劳动享有的合法收益与权利边界，避免其创作积极性因不合理使用受损。鉴于生成式人工智能的发展和利用过程中可能产生侵权行为，为更好地保护著作权人的合法权益，还应对生成式人工智能提供者的义务进行明确规定。

首先，生成式人工智能提供者应尽到相应的版权过滤措施。人工智能生成的内容具有不确定性，即使是提供者也无法预测最终的输出结果，因此提供者应当做到在技术层面人为强化训练数据和生成内容的差异^[17]，以避免侵权结果的发生。但考虑到提供者无法对模型训练涉及的数据和文本进行细致的审核，因此不宜过度加重提供者的义务，提供者以现有技术为标准采取相应的版权过滤措施即可认定为尽到了必要的义务。

其次，生成式人工智能提供者应在用户使用有关产品时提示尊重知识产权。相比较提供者，用户在使用大模型时输入的指令对最终生成的结果有更强的干预效果，且生成结果往往是为满足用户个人的期待和需要，因此生成式人工智能提供者应提示用户尊重知识产权，并采取一定措施来减少诱导式提问可能引发的侵权问题。

最后，生成式人工智能提供者应当建立投诉举报渠道。在“全球 AIGC 平台侵权第一案：奥特曼案”中，法院就认定被告未建立相关投诉举报机制，使得权利人难以通过投诉举报机制来保护其著作权，并据此认定被告作为服务提供者未尽到合理的注意义务。¹⁰因此，生成式人工智能的提供者应当建立起有效的投诉举报渠道，并在接到相关投诉后及时采取技术手段制止侵权，以保护著作权人的合法权益。此外，生成式人工智能提供者还应在合理期限内对导致侵权发生的算法进行纠正，以避免针对同一受保护作品的侵权行为再次发生。

4.3. 配套制度的设计——数据训练版权信息披露

如果想要对数据训练过程进行监管，则必须对数据训练的透明度提出要求。欧盟最近通过的《人工智能法案》专门规定了人工智能提供者的透明度义务——公开提供受版权法保护的训练数据或其他材料使用情况的详细记录^[18]。

一方面通过信息披露能够实现人工智能法律问责，另一方面借由信息披露能够促进人工智能技术改良。具体而言，相比于通过对输出的内容进行分析进而反推生成式人工智能提供者在数据训练阶段具体使用了哪些作品，信息披露能够提供作品使用的直接证据。一旦输出内容被认定构成侵权，著作权人可以通过侵权方所披露的信息直接证明其作品被用于数据训练，从而减轻著作权人的举证责任，有利于案件事实的调查和纠纷的解决。而通过信息披露，能够帮助识别出虚假、劣质、偏见信息，及时将该部分信息剔除有利于提高数据资源的质量，进一步提升输出内容的质量。此外，根据版权法“接触+实质性相似”的传统侵权判断标准，版权人应当具有提供被告构成“接触”的举证可能以维护自身合法权益。因此，无论是从版权人知情权的角度出发，抑或是从版权人举证的客观可能性出发，均应当对人工智能开发者提出合理的信息披露义务要求，要求对人工智能开发者在合理限度内公开技术原理，而合理限度可进一步界定为使版权人具备知晓其作品存在侵权风险的可能性^[19]。

信息披露是对作者精神权利的认可和尊重。我国《著作权法》合理使用制度的规定本就明确了“指明作者姓名或者名称、作品名称”的义务，¹¹鉴于数据训练合理使用规则下作者的奉献与牺牲，此种义务在人工智能场景下更应得到贯彻。而尊重并认可署名等精神权利是对作者心情的体察，是对作者感情的关怀，事关作者尊严的承认，具有更为根本性的内涵，人工智能训练数据的版权治理应对此给予充分重视。

¹⁰同上注 3。

¹¹同上注 5。

参考文献

- [1] 陈亮, 张翔. 欧盟生成式人工智能立法实践及镜鉴[J]. 法治研究, 2024(6): 105-118.
- [2] 魏远山. 生成式人工智能训练数据的著作权法因应: 确需设置合理使用规则吗? [J]. 图书情报知识, 2025, 42(1): 78-88.
- [3] 刘水美. 人工智能数据训练著作权合理使用法律规则路径探究[J]. 暨南学报(哲学社会科学版), 2024, 46(11): 60-73.
- [4] Grimmelmann, J. (2015) Copyright for Literate Robots. *Iowa Law Review*, **101**, Article 657.
- [5] 高阳. 人工智能训练数据侵犯著作权行为规制[J]. 中国出版, 2024(15): 12-18.
- [6] 詹爱岚, 田一农. 生成式人工智能机器学习中的著作权风险及其化解路径[J]. 电子知识产权, 2023(11): 4-14.
- [7] 张吉豫, 汪赛飞. 大模型数据训练中的著作权合理使用研究[J]. 华东政法大学学报, 2024, 27(4): 20-33.
- [8] 李安. 人工智能训练数据的版权信息披露: 理论基础与制度安排[J]. 比较法研究, 2024(5): 136-152.
- [9] 汤贞友, 谢艺婕. 生成式人工智能数据使用的版权侵权风险及其治理[J]. 中国出版, 2024(21): 56-61.
- [10] 张涛. 生成式人工智能训练数据集的法律风险与包容审慎规制[J]. 比较法研究, 2024(4): 86-103.
- [11] 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, 43(6): 170-185.
- [12] 徐小奔, 薛少雄. 生成式人工智能服务提供者版权注意义务的法律构造[J]. 科技与出版, 2024(7): 48-58.
- [13] 李杨. 著作权侵权认定中的转换性使用理论适用阐释[J]. 北方法学, 2023, 17(3): 42-56.
- [14] 刘云开. 人工智能训练作品的著作权合理使用进路[J]. 东北大学学报(社会科学版), 2025, 27(1): 117-126.
- [15] 刘晓春. 生成式人工智能数据训练中的“非作品性使用”及其合法性证成[J]. 法学论坛, 2024, 39(3): 67-78.
- [16] 李可心, 肖冬梅. 日本生成式人工智能训练数据合理使用规则及其启示[J]. 图书馆论坛, 2025, 45(9): 93-101.
- [17] 阮开欣, 黄歆瑜. 生成式人工智能数据训练中的版权问题研究[J]. 中国版权, 2024(5): 61-72.
- [18] 韩荣. 生成式人工智能作品利用的“合理使用+单纯获酬权”模式探析[J]. 出版广角, 2024(19): 75-80.
- [19] 储翔, 周怿霖. 适应生成式人工智能数据训练的版权合理使用规则完善[J]. 中国出版, 2025(6): 3-7.