Open Journal of Legal Science 法学, 2026, 14(11), 63-70
Published Online November 2026 in Hans. https://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2026.1411304

Hans汉斯

合理使用视角下生成式人工智能训练数据著作权风险与制度完善

冼天祺

湖北大学法学院，湖北 武汉

收稿日期：2026年9月30日；录用日期：2026年10月14日；发布日期：2026年11月11日

摘　要

生成式人工智能技术快速发展，训练数据著作权合规风险日益凸显，其数据驱动运行模式与现行著作权法律规范产生结构性冲突。针对人工智能数据输入、数据分析、内容生成全流程存在的著作权侵权争议，结合我国封闭式立法模式下合理使用规则难以适配大规模模型训练的现实困境，区分表达性行为与非表达性行为展开规范分析，尝试搭建“非享受性使用”与“转换性使用”联动的动态评估机制。研究发现，现有合理使用条款受制于适当引用、少量复制等适用条件，无法覆盖人工智能深度学习场景。依托现有立法修订路径，可在《著作权法实施条例》中增设人工智能深度学习对应的合理使用情形，并配套建立以使用性质与转换程度为核心的反向校验规则，为平衡技术发展与著作权人权益提供规范思路。

关键词

生成式人工智能，数据训练，合理使用，著作权

Copyright Risks and Institutional Perfection of Generative AI Training Data from the Perspective of Fair Use

Tianqi Xian

School of Law, Hubei University, Wuhan Hubei

Received: September 30, 2026; accepted: October 14, 2026; published: November 11, 2026

Abstract

With the rapid advancement of generative artificial intelligence (GenAI), copyright compliance risks

文章引用: 冼天祺. 合理使用视角下生成式人工智能训练数据著作权风险与制度完善[J]. 法学, 2026, 14(11): 63-70.
DOI: 10.12677/ojls.2026.1411304

concerning training data have become increasingly prominent. Its data-driven operation model creates structural conflicts with the existing copyright legal norms. In view of copyright infringement disputes arising throughout the whole process of artificial intelligence data input, data analysis and content generation, this paper conducts normative analysis by distinguishing expressive acts from non-expressive acts, considering the practical dilemma that fair use rules under China's closed legislative model can hardly adapt to large-scale model training. It attempts to construct a dynamic evaluation mechanism combining "non-consumptive use" and "transformative use". The study finds that the existing fair use provisions, constrained by application prerequisites such as appropriate quotation and partial reproduction, fail to cover the deep learning scenarios of artificial intelligence. Relying on the path of revising current legislation, a fair use circumstance corresponding to AI deep learning may be added to the Regulations for the Implementation of the Copyright Law of the People's Republic of China, with supporting reverse verification rules centered on the nature of use and the degree of transformation. This provides normative ideas for balancing technological development and the rights and interests of copyright holders.

Keywords

Generative Artificial Intelligence, Data Training, Fair Use, Copyright

Copyright © 2026 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

数字经济时代，数据已成为驱动社会发展的核心资源。这种新型生产要素具备四个关键特征：体量大、类型多、价值密度低及处理速度快的四大显著特征。以国产大模型 DeepSeek 为代表的开源项目通过代码、框架或工具的公开可获取，依托其突出的数据处理能力与知识生成能力，为社会大众开辟了全新的信息获取途径[1]。然而，这种数据驱动的技术跃进正面临严峻的著作权合规挑战。生成式人工智能需通过海量数据“投喂”完成训练，其语料库构建过程中大量使用网络公开文本、图像等作品，极易触发著作权侵权风险。2023 年出台的《生成式人工智能服务管理暂行办法》[1](后文简称《暂行办法》)在第 7 条中明确要求，生成式人工智能服务提供者在开展训练数据处理活动时，必须确保数据来源合法，并禁止侵犯他人依法享有的知识产权。但实践中“合法来源”认定标准模糊——当开发者爬取网络公开数据时，既难以逐项获取权利人授权，又无法有效区分作品著作权状态，导致著作权人与人工智能开发者利益严重失衡。

2026 年政府工作报告强调：“深化数据资源开发利用，健全数据要素基础制度，建设高质量数据集。完善人工智能治理。”[2]如今，历史的重大进程在一定程度上是由非人类智能所推动的，也正因如此，计算机网络的易错性才显得尤为危险。当计算机成为历史进程的重要推动者时，其出现的错误可能会引发灾难性后果[2]。综上，围绕生成式人工智能训练数据的著作权争议，本文以侵权风险的识别为切入点，对既有规制路径进行检视与反思，在论证合理使用制度对于规制生成式人工智能的数据训练过程的基础上，完善出针对生成式人工智能训练数据著作权侵权风险的规制路径。

[1]https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm
[2]http://lianghui.people.com.cn/2026/n1/2026/0313/c461827-40681711.html

2. 生成式人工智能训练阶段的著作权侵权风险及合理使用规制困境

2.1. 生成式人工智能运行机理与著作权风险审视

《暂行办法》第 22 条明确“生成式人工智能技术及服务”，是指“具有文本、图片、音频、视频等内容生成能力的模型及相关技术”，包括通过可编程接口等方式提供生成式人工智能服务。生成式人工智能的底层技术逻辑是依托现有数据的深度学习构建概率模型，进而实现创造性内容输出，其核心技术支撑包括机器学习、神经网络等前沿领域。数据训练的核心逻辑在于利用海量机器学习数据集(training data)，通过算法优化使 AI 系统具备预测数据规律与自主决策能力。训练用数据集的质量直接影响模型性能上限，优质训练数据成就优质人工智能，然而，训练过程中对受著作权保护作品的大规模采集与解析，已然引发法律合规性争议。

生成式人工智能训练运行过程分为三个阶段：数据输入、数据分析以及生成物输出[3]。在数据输入阶段，生成式人工智能通过数据爬取、清洗及标注等预处理流程，完成对原始数据集的格式标准化与持久化存储；进入数据分析阶段，系统会根据用户需求筛选相匹配的信息，经文本转换、词句识别等，实现系统对所搜集到信息的识别与归纳；最终，在生成物输出阶段，人工智能将基于解码器网络重构语义逻辑，输出符合语法规则与语用场景的文本。

1) 数据输入阶段的侵权风险

在数据输入阶段，人工智能的主要行为模式是数据抓取，对识别到的作品进行大量扫描和存储，并将其转换为机器可读的标准数据格式，以供后续深度学习使用。这一过程必然涉及对原始作品的复制行为，依据技术特性可分为两种形态：永久性复制和临时性复制。永久性复制指通过下载、存储等方式将作品固定于物理介质(如服务器硬盘、分布式存储系统)，形成可反复调用的持久化副本[4]。根据《保护文学和艺术作品伯尔尼公约》[3](后文简称《伯尔尼公约》)第 9 条及我国《中华人民共和国著作权法》[4](后文简称《著作权法》)第 10 条规定，此类复制行为明确纳入复制权规制范畴。临时复制则是指系统在处理数据时自动生成的瞬时性副本，例如网页缓存或媒体播放缓存，此类副本通常在设备重启或程序终止后即行消失。美国与欧盟将临时复制纳入复制权范畴，认为其存在侵犯复制权的潜在可能。但我国现行法律体系并未明确将临时复制列入侵权行为范畴，故在国内法框架下，侵权风险主要指向永久性复制行为。

2) 数据分析阶段的侵权风险

数据分析阶段是生成式人工智能模型训练的核心技术环节，其核心在于对既有数据进行加工处理与特征抽取，所引发的著作权风险主要集中于改编权、汇编权及翻译权。改编权指向对原作品的改动与再创作，汇编权则关涉对作品或其片段的筛选与组合。具体而言，系统通常先对原始数据进行去重、清洗等基础筛选，再根据指令完成翻译、改编或汇编，以构建初始数据集，并在此基础上学习作品的思想、情感、视角及表达方式。若经上述筛选与编排形成的数据集在内容选择与结构编排上具备一定程度的独创性，那么此类对数据加以选择、整理和组合的使用行为，就可能引发对翻译权、改编权及汇编权的侵害风险[5]。

3) 生成物输出阶段的侵权风险

生成式人工智能在输出阶段面临多重著作权风险。如果系统生成内容与原作品构成实质性相似，可能侵犯复制权。当人工智能通过改编或翻译原作品形成新内容时，可能触发改编权或翻译权纠纷。若对多部作品进行选择编排形成汇编成果，可能损害原作者的汇编权。例如，人工智能可能侵犯原作者的信息网络传播权的常见表现之一，即为通过在线方式向公众提供生成内容。与此同时，实践中开发者常出

[3]https://ipr.mofcom.gov.cn/zhuanti/law/conventions/wipo/1/Berne_Convention.html
[4]http://www.npc.gov.cn/c2/c30834/202011/t20201119_308796.html

于获取算法反馈、共享训练数据或验证结果准确性等考量，将作品检索与分析过程公之于众，其中即时公开发布机器学习的数据分析结果可能侵犯作品的广播权，而延时发布相关分析结果则可能侵犯信息网络传播权[6]。

2.2. 我国现行合理使用规则尚无法涵摄生成式人工智能著作权侵权风险

合理使用(Fair Use)是著作权限制与例外制度中的一项核心规则，旨在平衡著作权人的专有权利与社会公众获取、利用作品的公共利益，是知识产权法中协调公平竞争与公共利益的重要机制[7]。现代国际著作权法上的通用评判标准，主要源于《伯尔尼公约》第 9 条第 2 款所确立的“三步检验法”(Three-step Test)。该条款针对复制权设定限制条件，要求各成员国仅在以下三种情形下允许对复制权设定例外：仅限于特定例外情形，不得妨碍作品的正常利用，亦不得无理损害权利人的合法权益。此后，这一评判框架被《与贸易有关的知识产权协定》[5]及《世界知识产权组织版权条约》[6]相继吸纳，成为国际通行的检验准则。

我国《著作权法》中可能涉及生成式人工智能数据训练合理使用的条款主要集中于第 24 条第 1 款的四项情形：第 1 项“个人学习研究例外”、第 2 项“适当引用例外”、第 6 项“教育科研目的下的使用”以及第十三项兜底条款。首先，适用该条款的第一项情形须满足双重要求：“个人”主体要件和行为属于“学习、研究”的性质要件。然而，上述前提在我国现行法规制下均难以成立。一方面，我国法律尚未赋予生成式人工智能以法律主体地位；另一方面，将带有明显具有商业目的的数据训练行为与“学习、研究”相等同，亦偏离了该条款的制度设计初衷，从这一角度思考，数据训练行为也无法被第六项“教育科研目的下的使用”所涵摄。

其次，针对第 2 项“适当引用例外”，从数据规模来看，生成式模型需要处理的训练数据量级远超传统合理使用条款的适用范围。我国《著作权法》第 24 条列举的“适当引用”、“少量复制”等标准，难以匹配人工智能训练所需的亿级数据规模。即便该条款第 13 项规定了兜底情形，但缺乏配套行政法规的细化解释，导致司法实践中无法直接适用。

比较法视野下的制度设计呈现出差异化特征。欧盟通过《人工智能法案》[7]衔接《数字单一市场版权指令》[8]，建立“原则允许 + 例外保留”的双轨机制：文本与数据挖掘行为原则上构成合理使用，但允许权利人通过明示退出机制保留作品使用权。日本《著作权法》[9]创新性地引入“非享受性使用”概念，将不涉及作品表达内容消费的机器学习行为纳入例外范畴，形成更包容的技术发展空间。美国司法体系通过判例法发展出四要素合理使用制度适用性的标准，包括“使用的目的及性质”、“享有版权保护作品的性质”、“同整个被使用作品相比，所使用部分的数量和内容的实质性”及“使用行为对原作品潜在市场或其价值的影响”，其中“使用的目的及性质”是判断使用行为正当性的基础，通常作为首要因素予以考虑[8]。值得注意的是，美国法院在后续判例中提出了“转换性使用”标准，提出在后作品通过对原作的利用赋予其新的价值或功能，从而区别于原作的初始功能或目的。美国最高院在 Authors Guild v. Google 案中强调，该标准的认定取决于使用行为“是否且于何种程度上”在目的与性质层面区别于原作，属于程度而非类型问题。该案裁判观点认为，谷歌对图书开展数字化处理用于模型训练，并在此基础上实现图书片段的检索与展示，该系列行为应当归入合理使用范畴，为商业人工智能开发开辟数据训练合规通道[9]。反观我国生成式人工智能在数据训练环节对作品的使用，目前尚难以借助法律解释途径归入

[5]https://ipr.mofcom.gov.cn/zhuanti/law/conventions/wto/trips.html
[6]https://ipr.mofcom.gov.cn/zhuanti/law/conventions/wipo/1/WCT.html
[7]https://chinawto.mofcom.gov.cn/article/jsbl/zcfg/202608/20260803640094.shtml
[8]https://www.wipo.int/wipolex/zh/legislation/details/18927
[9]https://www.wipo.int/wipolex/zh/legislation/details/24012

《著作权法》的合理使用范畴，我国著作权法在此领域的制度供给仍显匮乏。

3. 合理使用制度规制的正当性证成

如若人工智能数据训练完全纳入合理使用范畴，则互联网平台依托技术优势将大规模无偿使用著作权保护的作品，著作权人必然会产生自身的正当权益和作品的正当使用将无法得到保障的质疑[10]。当前争议的焦点并非是否保护著作权人权益，而是如何在“技术跃进”与“权利保护”间寻求动态平衡。为此，需从制度构建、战略协同以及实践选择三个维度，系统论证合理使用规则的正当性基础。

3.1. 制度构建

生成式人工智能高度依赖数据供给，该技术特征决定了其在数据获取与输入环节一旦遇阻，将直接影响人工智能产业的发展。同时，生成式人工智能在数据训练环节需要海量作品，逐一授权模式在操作上几近不可行，由此产生的高昂交易成本不仅加剧了市场供需矛盾，亦潜藏司法与社会资源浪费的风险[11]。当许可成本远高于技术研发收益时，合理使用制度能有效破除“反公地悲剧”，其中“免费使用”模式能解决数据训练的实践成本和资金成本问题，避免因授权障碍导致算法偏见加剧或技术发展停滞优化[12]。兼顾技术发展诉求与法治体系建设需求，以合理使用制度规制数据训练行为具备现实与法理层面的合理性。

3.2. 战略协同

实现人工智能技术创新与数字产业良性发展，离不开系统完备的生成式人工智能制度保障。欧盟在《数字化单一市场版权指令》中指出，“文本与数据挖掘”技术的法律确定性与其区域竞争力密切相关。我国“十五五”规划将人工智能、新质生产力、数字治理作为核心板块，明确要求构建适配人工智能产业的现代化治理框架。在《中华人民共和国人工智能法(学者建议稿)》[10]中也体现了人工智能数据训练构成合理使用的可能性[13]，认定某一使用行为属于数据合理使用的条件为“该使用行为与数据原本的使用目的或者功能不同，且不影响该数据的正常使用，也没有不合理地损害数据权利人的合法权益”。我国政策导向已释放出对生成式人工智能数据训练适用合理使用制度的积极信号。该制度能够有效回应模型训练的数据需求，为人工智能产业提供稳定的制度支撑，对于未实质损害著作权人权益的数据使用行为，法律评价上更倾向于确认其正当性。

3.3. 实践选择

由上文所述可知，全球主要司法辖区在生成式人工智能数据训练的著作权规制方面呈现差异化路径。各国立法实践虽存在进程差异，但制度设计逐渐显现趋同特征。2022 年 12 月，美国和欧盟等多家权威机构于核心期刊发表联名文章指出，当前著作权法对文本和数据挖掘活动构成了一定制约。文章还提出，经济发展水平越高的国家，在著作权法立法条款或司法惯例上对文本和数据挖掘行为越持开放态度；相反，经济相对落后的国家在此方面的著作权法规则则更为僵化[14]。由此可合理推测，在未来的国际视野下，兼具实操可行性与实践适配性的合理使用规则或将成为各国处理该类治理问题的首选方案。

4. 生成式人工智能场景下著作权合理使用制度的完善路径

生成式人工智能训练数据的著作权规制困境，并非单一法律要件无法适用所致，而是传统合理使用制度的主体、行为性质、使用规模与利益衡量标准，与人工智能数据训练的技术逻辑之间存在系统性错

[10]https://sdbdra.cn/newsinfo/6966473.html

位。具体而言，人工智能训练行为既难以满足“个人学习、研究”条款中的主体要件和行为性质要件，也难以适用“适当引用”“少量复制”等以小规模、可识别使用为预设的传统判断标准；与此同时，由于训练过程所涉及的海量数据处理具有明显的技术性、批量性和非消费性，若仅依赖既有列举式合理使用情形，难以对其进行准确、有效的法律评价。

通过对域内外相关制度与实践的比较考察，以调整合理使用规则来适配生成式人工智能数据训练过程规制问题，在制度设计上具备现实性。然而，审视我国《著作权法》第 24 条第 1 款的前 12 项条款，尚未针对新技术发展所催生的需求增设合理使用情形，第 13 项兜底条款虽然提供了制度衔接空间，但由于缺乏具体适用标准，仍难以直接承担人工智能训练行为的规范功能。基于此，有必要以“表达性/非表达性”二元判定为逻辑起点，以“非享受性使用”与“转换性使用”联动的动态评估为核心，并辅之以市场替代效应等因素构成的反向校验机制，逐步建立适应生成式人工智能技术特征的合理使用认定框架。

4.1. 区分“表达性/非表达性”二元判定标准

在生成式人工智能训练数据合法性审查框架中，使用行为的法律定性是侵权判定的逻辑起点。因此，根据使用行为与作品独创性表达的关联程度，需要对训练数据的表达性使用与非表达性使用进行细致区分[15]。表达性使用是指以接触、再现作品独创性内容为核心目的，表现为直接体验性使用行为(如阅读文学作品的完整章节)和表达再现性使用行为(如传播影视作品关键帧)。非表达性使用是指以获取作品功能性要素或数据特征为技术目标，不以获取作品内容或向公众传播作品内容为目的的使用行为(例如算法压力测试中的数据处理，或将作品纳入研究数据集进行统计分析)，并不直接针对作品的独创性表达部分加以利用，此类使用因未触及作品“独创性表达”内核，通常不产生侵权责任。著作权法的规制以人工智能大模型对作品实施表达性使用为前提，合理使用制度亦仅在此场景下才有讨论空间。通常情况下，生成式人工智能依循用户指令完成数据处理与内容生成，并未涉及对作品的独创性表达利用，故归入非表达性使用范畴。但特殊情形下，如根据用户需求提炼小说核心情节或基于原作生成虚拟角色，则构成表达性使用，侵权风险显著增大。

在判断某一使用行为是否构成非表达性使用时，可综合考量以下要素：第一，使用行为是否触及作品的独创性表达部分。若使用行为仅涉及作品中的事实性信息、数据特征或语言统计规律，而未对作品的具体表达形式进行再现或传播，则更倾向于构成非表达性使用。第二，使用过程中作品内容是否被“感知”或“再现”。若系统在训练过程中未将作品内容向任何主体展示或传播，仅将其作为数据样本进行特征提取，则非表达性使用的特征更为明显。第三，使用的技术目的是否仅为获取数据特征或功能性要素。若使用行为的目的是训练模型识别语言模式、图像特征等技术性目标，而非获取作品所承载的思想情感内容，则属于非表达性使用。第四，使用结果是否导致作品内容向公众传播。若训练完成后的模型输出内容与特定原作不构成实质性相似，则训练阶段的使用行为更可能被认定为非表达性使用。

4.2. 构建“非享受性使用”与“转换性使用”联动的动态评估机制

日本《著作权法》第 30 条确立的“非享受性使用”规则为在立法中明确规定人工智能训练数据的著作权例外提供了重要参考。该条款创新性引入了美国的转换性使用原则，以使用行为的功能性目的为核心标准，认定“非享受性使用”不以获取作品承载的情感、思想内容作为目标，在限定合理范围之内，行为人可采取各类途径对受保护作品加以利用。但该条款也设置限制性例外规则：倘若该使用行为会不合理侵蚀著作权人的权益，即便行为人不存在享受的主观目的，依旧无法通过伯尔尼公约规定的“三步检验法”，合理使用抗辩仍不成立。日本著作权法相关条款的立法逻辑与生成式人工智能技术特性高度契合：训练过程通过数据清洗、特征提取、参数调整等技术手段，将作品转化为机器学习模型中的算法

参数，本质上属于对作品功能性价值的利用而非对作品独创性表达的直接消费。日本文化厅将其归类于“信息分析”范畴，正是基于技术应用中作品价值的转换特性。

然而，单纯的功能性使用标准难以应对复杂的技术场景。为此，需引入美国司法实践中成熟的转换性使用理论，构建“使用目的 + 内容标准”双重维度的动态评估体系。“非享受性使用”是“转换性使用”分析的前提条件。前者解决的是“使用行为是否进入合理使用讨论范围”的问题——若使用行为本身即构成对作品思想情感的消费，则无需进入转换性程度的评估，该行为原则上不构成合理使用。在具体制度设计层面，建议通过修订《著作权法实施条例》增设人工智能技术开发的合理使用条款。现行《著作权法》第 24 条第 13 项规定“法律、行政法规规定的其他情形”为制度衔接提供了法律接口，可将人工智能深度学习技术纳入该兜底条款的适用范围。这一调整既赋予法官灵活运用法律原则的空间，又为适应技术发展预留规则拓展通道。在制度衔接路径上，可选择《著作权法实施条例》《信息网络传播权保护条例》[11]等行政法规，明确列举“基于数据分析与算法建模的人工智能深度学习行为”作为合理使用情形。具体条款可参照日本《著作权法》第 30 条的“非享受性使用”规则，将技术开发中的非表达性数据分析纳入法定例外范畴。需要特别注意的是，合理使用认定范围不应局限于非商业领域。我国政策明确支持产学研协同创新，公立科研机构与企业的商业化技术研发已成为主流模式，唯有承认商业性技术开发的合理使用空间，才能有效落实国家人工智能发展战略，为技术创新提供制度保障。

设计生成式人工智能训练数据对应的合理使用规范，应当兼顾《伯尔尼公约》国际义务与国内著作权立法的规则框架。基于上述制度设计需求，《著作权法实施条例》修订时可选增设专门条款，明确人工智能技术开发中的合理使用范围，规定：“人工智能系统基于数据分析进行创作时，可依法存储与复制他人作品，但需确保生成内容最终向公众开放。”这一条款设计既确认技术研发的合法性基础，又通过结果公开要求实现权利平衡。此外，需建立“非享受性使用”与“转化性程度”的反向校验机制。当技术应用可能实质性替代原作市场(如 AI 生成的网络小说挤压原作电子书销售),即便符合“转换性使用”标准，也需通过三步检验法中的“市场影响测试”进行限缩。此时应结合个案中作品类型、使用比例、市场替代可能性等因素，对合理使用范围进行动态调适。

5. 结语

生成式人工智能的迅猛发展正深刻重构数字社会的生产范式与创新生态，其作为新一代基础设施的战略价值已在全球范围形成共识。然而，技术跃进背后潜藏着严峻的著作权合规困境：海量训练数据的获取与使用，在激发算法潜能的同时，亦使著作权保护面临“大规模复制”与“隐蔽性利用”的双重挑战。本文在论证合理使用制度对于规制生成式人工智能的数据训练过程的基础上，对合理使用制度的完善提供具有可操作性的建议。完善现行合理使用规则的第一步，是对数据训练过程中两类作品利用行为——表达性使用、非表达性使用作出清晰界分；其次在制度完善路径上，主张引入“非享受性使用”与“转换性使用”联动的动态评估机制。随着科学技术的变革与迭代，以“非享受性使用”与“转换性使用”联动的动态评估机制为核心，可逐步发展为适应多元技术场景的治理范式。未来，生成式人工智能技术的持续迭代与应用场景的不断拓展，仍将对著作权制度提出新的挑战，需进一步完善和细化合理使用的适用规则，最终实现生成式人工智能产业创新与著作权保护的长效平衡。

参考文献

[1] 王新雷, 饶宇锋. 数智时代的关键基础设施: 开源大模型普及的潜在风险与防范进路——以 DeepSeek 为例[J]. 决策咨询, 2025(2): 64-69+77.

[11]https://www.gov.cn/zwgk/2006-05/29/content_294000.htm

[2] [以]赫拉利. 智人之上: 从石器时代到 AI 时代的信息网络简史[M]. 北京: 中信出版集团股份有限公司, 2024.
[3] 吴汉东. 人工智能生成作品的著作权法之问[J]. 中外法学, 2020, 32(3): 653-673.
[4] 金宇菲. 生成式人工智能训练数据著作权侵权风险及应对[J]. 出版与印刷, 2025(3): 38-49.
[5] 张妍. 数据纠纷知识产权规制路径的构建[J]. 武汉理工大学学报(社会科学版), 2024, 37(5): 64-73.
[6] 焦和平. 人工智能创作中数据获取与利用的著作权风险及化解路径[J]. 当代法学, 2022, 36(4): 128-140.
[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] 李燕. 《著作权法》修改背景下合理使用制度的完善路径[J]. 出版发行研究, 2019(4): 56-58+67.
[9] 金海军. 合理使用认定中“转换性使用”的重新界定——基于“戈德史密斯案”的思考[J]. 中国版权, 2024(2): 36-50.
[10] 王文敏. 人工智能对著作权限制与例外规则的挑战与应对[J]. 法律适用, 2022(11): 152-162.
[11] 徐聪颖. 著作权法规制生成式人工智能数据训练行为的理论争议与路径选择[J]. 河北法学, 2026, 44(3): 97-113.
[12] 申卫星. 数据产权的力度和限度[J]. 东方法学, 2025(1): 32-45.
[13] 黄汇, 翟鹏威. 公共领域视野下大模型数据使用行为的著作权定性[J]. 知识产权, 2025(1): 94-110.
[14] Fiil-Flynn, S.M., Butler, B., Carroll, M., Cohen-Sasson, O., Craig, C., Guibault, L., *et al*. (2022) Legal Reform to Enhance Global Text and Data Mining Research: Outdated Copyright Laws around the World Hinder Research. *Science*, **378**, 951-953. https://doi.org/10.1126/science.add6124
[15] 袁帅. 数字化背景下作品非表达性使用的著作权法应对[J]. 知识产权, 2024(9): 110-126.