

生成式人工智能训练中直播数据利用的 著作权规制研究

徐沐阳, 李 想, 冯璐雯, 慕译家

上海理工大学出版学院, 上海

收稿日期: 2026年2月6日; 录用日期: 2026年2月23日; 发布日期: 2026年3月12日

摘 要

生成式人工智能的训练需要海量的数据, 而网络直播数据作为其中的重要组成部分, 便逐渐成为了训练重要的数据来源之一。此类数据通常同时具备视听作品等多重作品属性, 而人工智能训练过程中不可避免的数据复制与处理行为, 现实中极有可能落入著作权专有控制范围, 从而引发侵权风险。现行合理使用制度因依赖个案裁量、对市场替代效应敏感且难以适用于大规模商业化利用, 无法为AI训练提供稳定合规基础。在此背景下, 引入法定许可制度被认为是一种更具制度可行性的解决路径, 其允许AI开发者在支付合理报酬的前提下使用数据, 并通过集体管理组织整合权利主体、设计差异化版税分配机制, 以期在降低交易成本、避免“反公地悲剧”的同时, 保障权利人的获酬权, 实现技术创新与知识产权保护之间的平衡。

关键词

生成式人工智能, 著作权, 直播数据, 合理使用, 法定许可

Research on Copyright Regulation of the Use of Live Streaming Data in Generative Artificial Intelligence Training

Muyang Xu, Xiang Li, Luwen Feng, Yijia Mu

College of Publishing, Shanghai University of Science and Technology, Shanghai

Received: February 6, 2026; accepted: February 23, 2026; published: March 12, 2026

Abstract

The training of generative artificial intelligence requires massive amounts of data, among which

文章引用: 徐沐阳, 李想, 冯璐雯, 慕译家. 生成式人工智能训练中直播数据利用的著作权规制研究[J]. 电子商务评论, 2026, 15(3): 718-726. DOI: 10.12677/ecl.2026.153327

online live streaming data has gradually become an important data source. Such data often possesses multiple layers of copyright-protected attributes, including those of audiovisual works. During the process of AI training, the unavoidable acts of data copying and processing are highly likely to fall within the scope of exclusive rights under copyright law, thereby giving rise to significant risks of infringement. The existing fair use doctrine, due to its reliance on case-by-case judicial discretion, sensitivity to market substitution effects, and limited adaptability to large-scale commercial data utilization, fails to provide a stable and predictable legal basis for AI training activities. Against this backdrop, the introduction of a statutory licensing regime has been proposed as a more institutionally viable solution. Under this framework, AI developers are permitted to use copyrighted data upon payment of reasonable remuneration, while collective management organizations are employed to consolidate fragmented rights and design differentiated royalty distribution mechanisms. Such an approach aims to reduce transaction costs, avoid the “anticommons tragedy”, safeguard the remuneration rights of copyright holders, and ultimately achieve a balance between technological innovation and intellectual property protection.

Keywords

Generative Artificial Intelligence, Copyright, Live Streaming Data, Fair Use, Statutory License

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. 引言

随着生成式人工智能技术的爆发式增长，数据已成为数字经济时代的核心生产要素。网络直播数据包含视听、语音及弹幕等多模态信息，是训练及提升大模型拟人化能力的关键语料。然而，AI 开发者“先抓取、后训练”的普遍做法，与现行著作权法体系之间产生了显著冲突，使得海量主播面临作品被无偿使用的困境。

当前学界多试图通过扩张合理使用制度以化解危机。有观点认为，AI 训练旨在提取非表达性要素，应构成转换性使用[1]。也有观点强调，为了促进算法公平与技术创新，应当赋予 AI 训练广泛的文本与数据挖掘例外[2]。然而，这些逻辑在直播数据场景下并不能完全成立。与一般文本不同，直播内容承载了极高的人格与财产利益，且随着 AI 数字人技术的成熟，其生成物极可能在带货或表演领域替代真人主播。正如美国最高法院在 Andy Warhol 诉 Goldsmith 案中所指出的，一旦产生实质性市场替代，合理使用抗辩便难以成立。此外，当 AI 企业面对海量碎片化的数据时，若仍固守传统的一对一授权模式，高昂的交易成本会使其陷入“反公地悲剧”，阻碍技术创新。对此，有意见指出，应通过法定许可与集体管理机制，在保障权利人获酬权的同时限制其排他权[3]。

综上所述，本文主张引入 Calabresi 的责任规则理论[4]，允许 AI 开发者在支付合理费用的前提下利用相关数据，而非赋予权利人绝对的禁止权。换言之，本文尝试论证并构建一种适应 AI 时代的法定许可与版税分配机制，以期在保障产业发展的同时，实现知识产权的价值回归。

2. 直播数据的作品属性与 AI 侵权风险

2.1. 直播数据的构成与作品属性界定

在探讨 AI 训练的合法性之前，需要厘清网络直播数据在著作权法上的地位。网络直播数据并非单一

的数据，而是一个包含了画面、声音、文字及底层代码的多模态复合数据集。其法律属性的界定直接关系到权利的归属与侵权的判定。本文认为，直播数据主要由三个部分构成：直播的内容与素材画面、主播的表演与解说、用户的弹幕与评论。

2.1.1. 直播画面及其视听作品属性

依据《中华人民共和国著作权法》第三条，具备独创性的“固定于介质上的动态影像集合，由一系列有伴音或者无伴音的画面组成，并且能够借助技术设备进行传播”属于受保护的视听作品。但并非所有直播画面都满足此要求，司法实践中通常需要根据其是否含有独创性进行区分[5]。例如带货直播、赛事直播、知识分享或才艺表演等场景通常包含脚本设计、镜头调度、音乐编配以及主播与观众的互动，这些创造性的安排体现了作者的智力投入，符合视听作品的独创性要求[6]。而若直播画面仅呈现单一机位的监控画面或未经人为选择的风景，则因缺乏创作表达，通常仅被视为录像制品。

2.1.2. 主播的演绎与虚拟主播形象的作品属性

在直播画面之上，是主播的表演与解说。以游戏直播为例，在非剧情类、强竞技类游戏中，游戏主播的战术布局、即兴操作以及特别的解说风格，体现了其独特的思考和个性化表达，其符合独创性的要求[7]。再例如，虚拟主播的视觉形象通常构成美术作品，其在直播中即兴表演舞蹈、歌曲等连续画面可构成视听作品，若虚拟主播对产品进行评述，其底稿或口述内容可构成文字作品或口述作品[8]。

2.1.3. 用户交互数据与平台的汇编权益

是网络直播中的弹幕与评论这类独特的交互数据。绝大多数弹幕和实时评论因过于短小或仅是通用情感表达，缺乏独创性，属于公有领域的日常用语，不构成文字作品。而部分用户发布的具有一定长度、逻辑完整、体现个性化观点的长评，其独创性可能达到文字作品的高度，受著作权法保护[9]。再者，虽然单条弹幕与评论可能不构成作品，但海量弹幕或评论经过特定时间轴的排列组合，形成了反映特定群体情绪与文化氛围的整体，其可能会被视为汇编作品。

2.2. AI 训练使用直播数据的主要矛盾

AI 训练利用直播数据的主要矛盾，在于数据挖掘技术与复制权之间的冲突。从技术本质上看，AI 模型对直播数据的利用并非简单浏览，而是一个涉及提取、转换和存储的数据处理过程[10]。但这一必要步骤无法避免地会触及复制权的规制范围，从而在技术中立原则与权利排他性之间形成了难以弥合的冲突[11]。

从具体技术流程来看，AI 获取直播数据通常遵循 ETL 即提取(extract)、转换(transform)、加载(load)；有时也采用 ELT 技术，转换与加载的顺序发生了变化，但其核心都离不开转换这一环节。直播中的视频、弹幕等内容原是为人类观众设计的，计算机无法直接理解，因此必须将其转换为图像张量、文本向量等机器可读形式[12]。这就从客观上导致了两种复制结果：一是永久性复制，即转换后的数据集被永久保存在硬盘或云端；二是临时复制，即数据处理过程中在内存内生成的临时副本。其中，临时复制是否构成侵权，在国际和国内层面都存在显著争议。

2.2.1. “临时复制”的定性争议

回顾国际立法历程，在 1996 年制定《世界知识产权组织版权条约》时，欧美国家曾试图将“临时复制”明确纳入复制权范围，但遭到包括中国在内的发展中国家反对，最终条约文本未作明确规定，留下了解释空间[13]。此外，这种分歧也体现在各法域的不同选择上，欧盟法律明确复制权覆盖临时复制，原则上认定其构成侵权；而中国法律在此问题上则保持了一定的留白，尚未形成定论。这使得 AI 训练在全

球面临合规不一致的困境，在某些地区，临时复制可能无需授权；在另一些地区，则必须有法律豁免才能合法进行。

2.2.2. “非表达性使用”的理论尝试与现实局限

为调和关于“临时复制”的矛盾，学界提出了“非表达性使用”理论作为补充路径。该理论主张，AI复制数据的目的并非为了再现作品的独创性表达，而是为了分析其中的统计规律，因此不应受复制权的严格限制[11]。然而，这一理论在实践中充满了局限。因为著作权法通常从技术行为本身界定复制，并不深究使用目的。尤其当AI企业为提升效率而建立包含永久性复制的大型数据集时，“非表达性使用”理论难以直接否定其侵权风险，它更多只能作为判断是否构成“合理使用”的考量因素之一[14]。这进一步凸显了在现有版权框架内解决AI训练数据问题的局限性，也指向了探索新机制的必要性。

3. 直播数据训练的合理使用困境与法定许可适用

3.1. 合理使用制度适用的困境分析

面对AI训练引发的版权侵权指控，业界通常将视角置于合理使用制度之上。但是本文认为，现阶段的合理使用制度在直播数据挖掘场景下并不完全适用，不足以成为目前AI企业完全的免责理由。

3.1.1. 合理使用制度的特征及局限

从制度定位看，合理使用并非为产业运行提供稳定授权的常态性机制，而是一种具有高度弹性与补救属性的版权例外规则。其功能在于通过个案判断，对著作权专有权作出必要限制，以防止过度行使权利从而妨碍公共利益。正因如此，合理使用在制度结构上存在若干普遍性特征，而这些特征也在客观上限制了其作为稳定制度基础的可能性。

合理使用的运用高度依赖事后判断与司法裁量。无论采取多因素衡量还是原则性标准，其适用结果通常需在具体诉讼中由法院最终认定。这种结果导向的机制，使利用者往往难以在行为前获得明确的合法性预期，因而使该制度更接近于一种事后的免责抗辩，而非可预先遵循的行为规范[15]。

同时，合理使用本身具有鲜明的例外属性。该制度不是为持续性、系统性的大规模利用设计的，而是对特定目的、特定方式的有限使用作出的例外安排。相关判断标准中对使用范围、使用程度以及对权利人利益影响的考量，均建立在适度使用的前提之上，使其难以适配高频、可复制、大规模的数据利用模式[16]。另外，合理使用对潜在市场影响极为敏感。在各类法律体系中，是否损害作品的正常使用或潜在市场，始终是判断合理使用成立的关键因素之一。一旦相关使用可能对既有或可预期市场形成替代，其正当性基础便会显著削弱[17]。而在新兴技术应用中，市场影响常具有滞后性与不确定性，这使得合理使用的边界更加模糊。

此外，合理使用不以获得报酬或利益再分配为目标，其重心在于免责而非补偿[18]。因此，在权利主体众多、利益结构复杂的利用场景中，该制度难以回应收益分享与利益协调方面的现实需求。

3.1.2. 合理使用抗辩在AI侵权诉讼中的不足

当合理使用制度被直接移用于以大规模数据利用为特征的AI训练场景中时，其内在缺陷将在侵权诉讼中被进一步放大，从而进一步削弱该制度作为主要合规路径的可行性。

合理使用并非事前授权规则，而是需要在纠纷发生后，由法院结合具体案情作出判断的事后抗辩理由。从司法实践来看，对于转换性使用等关键标准，不同法院的判断并不一致。例如，在 *Authors Guild v. Google* 案中，谷歌图书扫描被认为具有转换性，且对原作品市场影响有限，从而构成合理使用。但在近年来的 *New York Times v. OpenAI* 等案件中，权利人强调生成式AI输出内容与原作品在功能和市场上

的高度重合，法院态度明显趋于谨慎。这种裁判标准的不稳定，使企业难以在训练初期准确判断自身行为的合法性[19]。

进一步来看，在网络直播数据的 AI 训练与使用场景中，转换性使用的成立空间十分有限。直播内容承载着主播独特的语言风格、表达节奏与个性化表达，完全可能构成具有独创性的表达。AI 对直播数据的学习，并非简单提取事实信息或技术特征，而是大量吸收主播具有独创性的作品要素。受制于直播带货、服务等特定场景的功能约束，AI 据此生成的内容，很难形成新的思想视角、美学表达或使用价值，也难以真正实现转换性创作。相反，其生成结果更容易在表达风格、功能用途上与原直播内容形成高度近似，甚至产生直接的市场替代效果。因此，在网络直播数据使用中，以转换性使用作为侵权抗辩，缺乏足够的法理基础与现实合理性。

此外，AI 模型训练具有明显的不可逆特征。一旦模型完成训练，相关成本即已无法回收，若企业在事后被认定不构成合理使用，其可能面临模型下架、删除数据甚至重新训练的风险，这将造成巨大的损失。在这种情况下，合理使用并未真正降低风险，反而将法律风险推迟至事后集中爆发。因此，在直播数据挖掘等场景下，仅凭合理使用原则难以为 AI 训练提供稳定可靠的法律基础。

3.2. 法定许可制度作为解决方案的优越性

在版权法体系中，法定许可制度是指法律预先规定在特定情形下，作品可以在无需事先取得权利人同意的情况下被使用，但使用者须依法支付合理报酬的一种制度安排。与完全依赖个别授权的自愿许可不同，法定许可通过法律强制力统一设定使用条件和报酬规则，兼具效率与公平双重目标[3]。在 AI 训练这一高度依赖大规模数据输入的场景中，合理使用制度由于判断标准不统一结果高度依赖个案裁量，难以化解直播数据利用所引发的侵权风险，因此，本文认为，有必要引入一种兼顾技术创新激励与权利人利益保护的替代机制。从法律经济学视角出发，并参考域外的立法实践，法定许可制度在降低交易成本与实现利益平衡方面具有明显优势。

3.2.1. 降低交易成本

网络直播数据通常同时包含画面、音乐、解说等多种内容，对应多个权利主体。对于 AI 训练而言，这种高度分散的权利结构会直接推高授权成本。如果仍采用逐一协商的授权方式，开发者需要与大量权利人分别谈判，不仅成本高昂，而且往往难以在合理时间内完成授权，最终使数据无法被有效利用。这正是“反公地悲剧”在数据利用场景中的典型表现[20]。

针对这一问题，不同国家采取了不同制度路径。日本通过立法直接允许以信息分析为目的的作品使用，将 AI 训练排除在许可体系之外，从而彻底消除了授权成本[21]。但这一做法同时剥夺了权利人的报酬请求权，与我国强调创作者利益保护的立法方向并不一致。相比之下，北欧的延伸性集体管理制度在降低成本的同时保留了权利人的经济利益。在该制度下，集体管理组织与使用者达成的许可协议，可以依法适用于未加入组织的权利人，从而避免逐一谈判的问题[22]。这一思路可为直播数据的 AI 训练提供参考：通过设定法定许可，将分散的主播及相关权利人统一纳入集体管理框架，由组织集中授权并分配报酬。这样，AI 企业只需面对单一主体即可合法使用大量数据，交易成本随之大幅降低。从制度效果看，法定许可并非否认权利人的存在，而是通过法律手段简化授权流程，将原本难以完成的个别交易转化为可执行的制度安排，从而为大规模数据利用提供现实可行的路径。

3.2.2. 实现利益平衡

在传统著作权制度中，权利人原则上享有是否许可他人使用作品的决定权。但在 AI 训练场景中，如果每一位权利人都可以单独拒绝授权，任何个别拒绝行为都可能影响整体训练进程，甚至使项目无法继

续。这种结果并不符合数据利用所带来的整体社会利益。

法定许可的作用，正在于对这种情况进行制度调节。通过法定许可，AI 开发者在符合法律规定条件的情况下，可以无需事先取得许可而使用相关数据，但必须依法向权利人支付报酬[1]。这意味着权利人不再拥有完全的否决权，但其获得合理报酬的权利仍受到保障。由此，制度在防止数据被个别权利人阻断的同时，也避免了技术企业无偿使用创作成果。与合理使用主要依赖司法在个案中进行判断不同，法定许可通过立法事先明确适用范围、使用条件和报酬机制，使数据利用的法律后果更为清晰。这种明确性可以显著降低 AI 企业面临的合规风险，也减少因法律不确定性而产生的诉讼成本。因此，法定许可不仅是一种提高效率的制度设计，更是在 AI 训练这一特殊场景下，对权利保护与数据利用之间关系所作出的较为稳妥的平衡安排。

4. 法定许可的制度构建与机制设计

在明确了引入法定许可必要性后，亟需解决的是该制度在直播数据 AI 训练场景下的具体构建问题。因此，本章将视角由“理论应然”转向“制度实然”，具体构建适用于网络直播数据训练的法定许可机制。

4.1. 法定许可适用行为的限定

并非所有与直播数据相关的使用行为，都有必要纳入法定许可的调整范围。作为对著作权专有权利的例外性制度安排，法定许可在适用对象的界定上应保持克制。基于使用目的与技术功能的区分，法定许可的适用对象应限定于以模型训练与算法优化为目的的数据利用行为[23]。此类行为的共同特征在于，其利用结果并不以向公众传播或再现作品内容为目标，而是通过统计分析、特征提取、参数优化等方式，将作品转化为抽象化的数据特征，用以提升模型的能力。

相应地，若 AI 企业将直播数据直接用于生成可替代原直播内容的商业化产品，或将训练数据本身对外提供、交易，则已实质性进入了对作品正常利用与潜在市场产生直接影响的领域。在此情形下，继续适用法定许可将削弱著作权专有权利的排他效力，因此应当回归一般授权规则或侵权责任体系加以规制。通过对使用目的与利用方式的严格限定，可防止法定许可被不当扩张而异化为一般性免责机制，并维持其作为版权例外制度的定位。

4.2. 权利主体整合与集体管理机制

4.2.1. 多元权利主体整合的必要性

与传统的由单一作者完成的作品不同，网络直播数据通常涉及多个权利主体。一场直播中，既包含主播对其表演内容所享有的相关权利，也可能涉及平台基于合同安排所享有的信息网络传播权益，同时还常常使用到背景音乐、图片或特效素材等第三方作品。这种多主体共同参与的创作与传播模式，使直播数据在权利归属上呈现出较为复杂和分散的特点。

在这种情况下，若缺乏相应的权利整合机制，即便引入法定许可制度，其在实际运行的过程中仍可能受到较大限制。一方面，AI 企业在进行数据合规时，往往难以准确识别并逐一联系所有相关权利主体；另一方面，分散的个体权利人普遍缺乏与大型技术企业进行谈判的能力，信息不对称问题较为突出，客观上提高了制度运行成本。由此可见，单纯依赖法定许可而不解决权利分散问题，难以有效回应直播数据大规模利用的现实需求。

因此，引入著作权集体管理组织具有现实必要性。通过由集体管理组织集中行使相关权利，可以在权利人与 AI 企业之间建立相对稳定的对接渠道，在一定程度上降低授权和合规成本，同时也有助于保障个体权利人的基本利益。该机制为法定许可制度在直播数据 AI 训练场景中的实际适用提供了重要的制度

支持[24]。

4.2.2. 延伸性效力的引入

若法定许可仅对加入集体管理组织的权利人发生效力，则仍可能因部分权利人消极观望或策略性退出，导致许可链条被阻断，从而重现“反公地悲剧”。

延伸性效力的基本做法，是在满足一定条件的情况下，使集体管理组织与使用者达成的许可安排，对未明确表示反对的同类权利人同样发生效力。该机制并非否定个体权利，而是通过制度化方式，弥补在大规模数据使用情形下逐一授权难以实现的问题[22]。在直播数据 AI 训练中，引入延伸性效力具有现实的合理性。一方面，模型训练过程具有整体性，AI 企业难以在模型层面区分已授权数据与未授权数据；另一方面，直播内容涉及的权利主体数量庞大，逐一征得许可在实践中几乎不可能。因此，延伸性效力是对法定许可制度得以有效发挥其功能的重要补充。

此外，延伸性效力的适用还应受到严格限制，并配套必要的保障程序。第一，集体管理组织应具备充分的代表性与公信力，其成员结构、治理规则及收费机制需接受公共监督；第二，应建立明确的事前公告与异议程序，允许权利人在合理期限内明确表示反对；第三，应赋予权利人合理的退出权与充分的谈判空间，以避免延伸性效力演变为事实上的强制许可。通过这些安排，可以在提高制度可执行性的同时，避免对权利人利益造成不当影响。

4.3. 报酬标准与分配机制

4.3.1. 报酬标准的确定原则与差异化定价

法定许可的最重要的环节之一，即在于报酬机制的合理性。在直播数据 AI 训练场景中，数据价值具有明显的差异，很难以单一标准进行衡量。因此，报酬标准不宜采取“一刀切”的定价方式，而应进行差异化定价。具体而言，可在国家有关部门的指导下，建立统一但具有弹性的指导费率框架，并通过权重因子进行动态调整。相关权重因子可包括数据类型、数据质量、数据贡献度。

首先，在数据类型层面，不同媒介形态所承载的信息量与训练价值存在客观差异，这种差异应当在报酬标准中得到体现。一般来说，视频数据同时包含图像、音频等多种信息，其价值理应高于单一音频或文本数据。

其次，在数据质量方面，直播内容的清晰度与稳定性会直接影响其被模型有效利用的程度。高清、稳定且制作相对规范的直播数据，更有利于模型的学习，其利用价值理应高于低清或高度碎片化的内容。

最后，在数据贡献度层面，区分头部主播与普通用户，有助于在报酬设计中反映直播内容在现实中的影响力差异。头部主播的直播相比普通用户通常具有更高的观看量和互动量，其数据在训练中所体现的价值也相对更高。

在此基础上，还可采用“基础费率 + 流量乘数”的计算模式。基础费率用于保障最低补偿水平，确保数据不被无偿使用；流量乘数则可根据粉丝规模或观看量等指标进行调整。

4.3.2. 报酬分配的透明化与预存机制

在报酬分配层面，考虑到权利人高度分散、训练行为难以精确对应具体作品，集体管理组织应建立相对透明且可解释的分配规则。对于无法逐一核算的使用情形，可通过抽样、权重估算等方式进行分配，但应定期向权利人披露收支情况，以确保整体公平性。同时，可引入智能合约技术作为辅助工具，依托区块链平台，将费率标准与分配规则以程序代码形式固化，实现自动结算与留痕记录。这样既可以减少人工操作带来的不透明问题，也便于权利人事后查询和监督。当然，考虑到商业秘密与数据安全问题，可采取“链上记录规则与结果、链下保存具体数据”的方式，在提高透明度的同时避免过度公开。

此外,为保障制度运行的稳定性,还可引入版权费预存机制。即由 AI 企业在使用直播数据前,向集体管理组织预存一定数额的版权费,并根据训练数据规模变化进行动态调整,以降低事后追偿与纠纷解决成本[25]。

5. 结论

在生成式人工智能高度依赖大规模数据输入的背景下,网络直播数据已经成为 AI 模型训练和算法优化的重要资源,但随之而来的著作权问题也愈加凸显。本文认为,网络直播数据兼具视听作品属性与生产要素价值,在当前的 AI 训练过程中,其对直播数据的规模化复制与利用在现行著作权法框架下存在较高的合规风险。尽管合理使用制度在价值层面强调利益平衡,但其高度依赖事后裁量,且对商业化利用高度敏感,难以适应生成式 AI 所需的持续性、大规模数据利用需求。

相比之下,法定许可制度展现出了更强的制度适配性。通过在特定条件下允许使用并要求支付报酬,法定许可既能避免个别权利人阻断数据利用,又可保障创作者的合理收益,有助于降低授权成本,缓解数据利用中的结构性障碍。因此,本文主张构建面向 AI 训练场景的专门法定许可机制,并在适用范围、主体整合、费率设计等方面作出具体安排。通过引入集体管理与差异化报酬机制,可以在保障权利人合理收益的同时,降低技术创新的制度阻力,实现数据利用效率与版权保护之间的动态平衡。这一制度路径,有助于在促进技术发展的同时,维护内容产业的长期秩序与活力。

参考文献

- [1] 陶乾. 基础模型训练的著作权问题: 理论澄清与规则适用[J]. 政法论坛, 2024, 42(5): 152-164.
- [2] Geiger, C., Frosio, G. and Bulayenko, O. (2018) Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data? Legal Analysis and Policy Recommendations. *IIC-International Review of Intellectual Property and Competition Law*, 49, 814-844. <https://doi.org/10.1007/s40319-018-0722-2>
- [3] 魏丽丽. 生成式人工智能数据训练中著作权限制制度的适用选择[J]. 政法论丛, 2025(5): 159-172.
- [4] Calabresi, G. and Melamed, A.D. (2013) Property Rules, Liability Rules, and Inalienability: One View of the Cathedral. In: *Modern Understandings of Liberty and Property*, Routledge, 139-178.
- [5] 孙昊亮. 论视频的分类与独创性认定[J]. 法律科学(西北政法大学学报), 2024, 42(6): 36-48.
- [6] 何渊, 周琴生, 周长福, 等. 网络游戏直播画面著作权归属问题研究[J]. 河南科技, 2022, 41(12): 155-158.
- [7] 林汶欣. 类型化视角下网络游戏直播画面的著作权归属问题[J]. 法学, 2023, 11(6): 5425-5430. <https://doi.org/10.12677/OJLS.2023.116776>
- [8] 韩思宇. 虚拟数字人表演的著作权保护问题研究[J]. 争议解决, 2024, 10(6): 163-169. <https://doi.org/10.12677/ds.2024.106319>
- [9] 薛佳. 新媒体场域用户弹幕可版权性研究——以 Bilibili 网站为例[J]. 河南科技, 2023, 42(5): 120-124.
- [10] 周啸. 生成式 AI 数据使用中的著作权侵权风险研究[J]. 法学, 2024, 12(5): 3261-3267. <https://doi.org/10.12677/ojls.2024.125463>
- [11] 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.
- [12] Triaille, J.-P. (2014) Study on the Legal Framework of Text and Data Mining (TDM).
- [13] [匈]米哈依·菲彻尔. 版权法与因特网[M]. 郭寿康, 万勇, 相靖, 译. 北京: 中国大百科全书出版社, 2009: 650-651.
- [14] 焦和平, 李泽儒. 合理界定数据训练中使用作品的行为[N]. 中国知识产权报, 2025-02-20(03).
- [15] 王果, 曹梓怡. 我国著作权合理使用的认定困境与规则完善——基于 142 份判决的分析[J]. 重庆邮电大学学报(社会科学版), 2025, 37(2): 76-84.
- [16] 黄荣楠, 刘佳迪, 赵怡青. 再议 AIGC “版权合理使用例外”: 美国 Anthropic 案解读[EB/OL]. 2025-07-03. <https://www.junhe.com/legal-updates/2727>, 2026-01-06.
- [17] 王玉琪. 合理使用要素分析在司法实践中的适用与反思[J]. 社会科学前沿, 2024, 13(5): 367-373.

<https://doi.org/10.12677/ass.2024.135403>

- [18] 黄炜杰. 论著作权合理使用规则的优化——基于界权成本的有效配置[J]. 学术研究, 2024(10): 67-72.
- [19] Didsbury, H. and Zhu, X.A. (2025) Transformative Training: An Analysis of AI Training Data and Fair Use in Authors Guild v. OpenAI Inc. *Publishing Research*, 4, e002. <https://doi.org/10.48130/pr-2025-0001>
- [20] 陈艺. 用户生成内容中的反公地悲剧及其应对[J]. 暨南学报(哲学社会科学版), 2024, 46(11): 74-88.
- [21] 万勇. 人工智能时代著作权法合理使用制度的困境与出路[J]. 社会科学辑刊, 2021(5): 93-102.
- [22] 孙新强, 姜荣. 著作权延伸性集体管理制度的中国化构建——以比较法为视角[J]. 法学杂志, 2018, 39(2): 34-42.
- [23] 梁远高. 论人工智能大模型训练数据风险的分层规制[J]. 郑州大学学报(哲学社会科学版), 2025, 58(3): 61-67+144.
- [24] 孙山, 曾翊展. 从“分散许可”到“集中治理”: 法定许可制度因应人工智能训练行为的优化方案[J]. 科技与出版, 2025(10): 105-118.
- [25] 汤广花. 人工智能时代版权保护三大难题如何破解? [N]. 中国新闻出版广电报, 2025-07-17(005).