Open Journal of Legal Science 法学, 2025, 13(11), 2457-2463
Published Online November 2025 in Hans. https://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2025.1311336

Hans汉斯

生成式人工智能数据训练的著作权法因应

胡衍惠

湖北大学法学院，湖北 武汉

收稿日期：2025年9月26日；录用日期：2025年10月10日；发布日期：2025年11月6日

摘 要

生成式人工智能已经成为人工智能领域的研究重点，其生成作品的质量有时已经可以超越人类作品。这一现象离不开对生成式人工智能进行海量数据练，通过机器学习优化模型性能、提升创新能力。这些数据既有可能是公共领域的开源数据库也无法避免使用到在版权法保护范围内的作品数据，因而也就产生了数据输入端侵犯原作品著作权的问题。现有的授权使用规则与法定许可规则会极大增加AI运营商的开发成本，不利于人工智能产业的创新发展。为了保护新业态发展模式，在输入阶段，我国应对人工智能输入端数据设置合理使用规则，明确其适用的四要素标准和前提条件；在输出阶段，应当明确生成式人工智能输出内容侵权问题时AI运营商的过错责任承担，并且要呼吁AI运营商构建合理的预防机制与补救措施，如设置关键词过滤、举报投诉机制等。通过上述建议以期促进人工智能产业的发展，实现促进版权保护与维护公共利益的平衡。

关键词

数据训练，合理使用，四要素判断，利益平衡

Copyright Law Responses to Generative Artificial Intelligence Training Data

Yanhui Hu

School of Law, Hubei University, Wuhan Hubei

Received: September 26, 2025; accepted: October 10, 2025; published: November 6, 2025

Abstract

Generative artificial intelligence has become a research focus in the field of artificial intelligence, and the quality of its generated works sometimes surpasses that of human works. This phenomenon is inseparable from the massive data training of generative artificial intelligence, through which the performance of the model is optimized and the innovation ability is enhanced by machine learning.

文章引用: 胡衍惠. 生成式人工智能数据训练的著作权法因应[J]. 法学, 2025, 13(11): 2457-2463.
DOI: 10.12677/ojls.2025.1311336

These data may be from open-source databases in the public domain, but it is also inevitable to use works data within the scope of copyright protection, thus giving rise to the problem of copyright infringement of the original works at the data input end. The existing authorization and use rules and statutory licensing rules will greatly increase the development costs of AI operators, which is not conducive to the innovative development of the artificial intelligence industry. In order to protect the new business model, at the input stage, China should set reasonable use rules for the data input end of artificial intelligence, clarify the four-element standards and preconditions for their application; at the output stage, it should clarify the liability for fault of AI operators when the output content of generative artificial intelligence infringes on copyright, and also call on AI operators to build reasonable prevention mechanisms and remedial measures, such as setting up keyword filtering and reporting and complaint mechanisms. Through the above suggestions, it is expected to promote the development of the artificial intelligence industry and achieve a balance between promoting copyright protection and maintaining public interests.

Keywords

Data Training, Rational Use, Four-Element Judgment, Interest Balance

1. 问题的提出

2022 年 11 月，OpenAI 正式发布的世界第一款能够进行对话的人工智能产品——ChatGPT 成为 2025 年 3 月份全球下载量最高的应用 [1]；2025 年 1 月，我国杭州深度求索人工智能基础技术研究公司推出的 AI 助手——DeepSeek，上线 18 天内，累计下载量已突破 1600 万次 [2]。人工智能产业方兴未艾，在 ChatGPT、DeepSeek 等人工智能产品发布之后将人工智能产业推向了新的研究高潮，引领了新一轮的技术革命和产业升级。生成式人工智能是人工智能研究的重中之重。生成式人工智能生成的内容并非是对现有数据的简单复制、汇编、整合，而是通过数据喂养和训练算法模型，在迭代训练中不断完善结果，并独立生成、输出全新内容。例如，ChatGPT 具有强大的自然语言处理能力和多模态转换能力，不仅可以完成各种问答、写作等的文字处理工作，还可以进行编写和调试计算机程序、创作音视频作品，并在此基础上实现跨模态生成，形成创新蓝海。

生成式人工智能的信息处理能力基于运用海量数据进行的模型训练，随着社会需求的提高，模型训练的所需要的数据的数量和质量也随之提高。在生成式人工智能的数据训练过程中不可避免地会对受版权保护的作品进行语料分析，如此 AI 运营商会面临输入数据和输出数据的版权侵权问题。在数据输入阶段，要将海量的文本、图像、音频和视频等原始数据输入至数据库，复制之后输送模型训练算法进行学习，这一阶段 AI 运营商受到著作权人复制权的限制。在数据输出阶段，人工智能通过预先的模型和算法进行结果输出，形成对他人作品的综合性使用并组成创作物表达，实现“洗稿”“重混”“拼凑”等生成活动[1]。如此，人工智能生成的内容就会在内容和结果上涉及他人的作品或者作品片段的表达，AI 运营商也存在侵犯已有作品的传播权的风险。

近年来，关于 AI 运营商侵犯著作权的案件频发。安德森等艺术家诉 Stability AI 等公司版权侵权案

[1]数据检索自 https://www.199it.com/archives/1750397.html，检索时间为 2025 年 10 月 8 日。
[2]数据检索自 https://wallstreetcn.com/articles/3740271，检索时间为 2025 年 10 月 8 日。

件[3]中，原告主张被告生成的作品实现对原有作品进行了复制，最终生成了与原作品存在竞争关系的演绎作品，侵犯了原作品的复制权、发行权、演绎权和传播权。在人工智能行业不断发展的当今社会，诸如此类的纠纷呈现出上升趋势，像环球音乐集团诉 Anthropic 案[4]等都指控 AI 运营商未经原作作品的同意存在对先前作品的非法使用问题。今年发生在广州互联网法院的奥特曼侵权案件，AI 运营商也被认定侵犯了原作的复制权、改编权和信息网络传播权。[5]人工智能的迭代升级离不开训练的大量数据，训练数据的质量会直接关系到最终的性能[2]。通过机器学习生成内容的人工智能高度依赖数据的自由获取和使用，但是《著作权法》也不应忽视在这一过程中可能存在侵权问题。为了实现版权保护和公共利益的平衡，所以我们应从人工智能数据训练的正当性出发，分析人工智能数据获取的合理使用规则适用标准，界定人工智能侵权的法律责任并为我国的人工智能治理司法实践提出合理的建议。

2. 生成式人工智能数据训练应得到社会支持的正当性分析

2.1. 生成式人工智能数据训练在《著作权法》中的合法性分析

我国《著作权法》的立法目的一是要保护作者的合法权益，二也鼓励作品的创作和传播，促进社会主义文化和科学事业的发展和繁荣。[6]这也就涉及到了个人利益和社会公共利益的平衡问题，《著作权法》并不是一味地鼓励、支持著作权人的利益，而是基于公共利益设置了合理使用、法定许可等限制著作权人权利的限制措施，这样的限制是为了更好地激励传播和次生创作，保障社会公众获取作品的权利，从而达到个人和公共利益的平衡点[3]。

生成式人工智能技术的迭代与应用需要成千上万的数据予以支撑，其数据训练的需求主要体现在数据数量、多样、质量、领域特定、多模态、实时、长期演进、平衡、合规以及多语言等方面[4]。根据我国颁布的《生成式人工智能服务管理暂行办法》和《生成式人工智能服务安全基本要求》对于训练数据合法性和语料内容的规定可知，生成式人工智能在进行数据训练时不得有侵犯知识产权的风险。我国《著作权法》也明确生成式人工智能服务提供者要消除著作权侵权风险，在未获得训练数据包含的权利人的许可下，只能使用公共领域数据进行数据训练。

文本与数据挖掘(Text and Data Mining，以下简称 TDM)是人工智能机器学习的底层技术，对数据的处理基本涵盖了信息搜寻、分析等处理活动。近年来，许多国家已经在修改法律，积极将使满足条件的 TDM 纳入合理使用范畴。日本在 2018 年对其著作权法进行修改时就确立了 TDM 例外的合理使用条款，极大地释放了文本与数据的潜力，达到了激励创新的效果。欧盟也在《数字化单一市场版权指令》也增加了两项 TDM 例外条款，缓和了法律对人工智能技术发展的阻碍，保障了大数据行业的发展。我国还没有关于 TDM 的例外条款，并且我国《著作权法》也在 2020 年进行了一次修改，频繁地修法不利于维护法律的稳定性，所以对于生成式人工智能数据训练中产生的问题需要进行个案分析。有学者指出可以将 AI 运营商开发的 AI 分为版权合规型 AI 和版权违规型 AI [5]。对于前者而言，其生成内容与先前作品不构成实质性相似，不存在侵权内容，这样的 AI 可以促进社会福祉的实现，也与《著作权法》的立法目的不谋而合。后者则是生成了侵权内容，不应对该种 AI 认定为合理使用。我国对于生成式人工智能的数据训练也颁布了《生成式人工智能服务安全基本要求》(以下简称《基本要求》)，其指出企业在采集和训练两个行为作出前，均需针对来源数据进行安全评估。对于如何进行安全评估这一问题《基本要求》也做出了明确回应，要求生成式人

[3]参见 Andersen v. Stability AI Ltd., 2023 U.S. Dist. LEXIS 194324, 2023 WL 7132064.
[4]数据检索自 https://www.bjnews.com.cn/detail/1703757216129948.html，检索时间为 2025 年 10 月 8 日。
[5]参见广州互联网法院(2024)粤 0192 民初 113 号民事判决书。
[6]《中华人民共和国著作权法》第一条 为保护文学、艺术和科学作品作者的著作权，以及与著作权有关的权益，鼓励有益于社会主义精神文明、物质文明建设的作品的创作和传播，促进社会主义文化和科学事业的发展与繁荣，根据宪法制定本法。

工智能服务提供者采用关键词、分类模型、人工抽检等方式，充分过滤训练数据中违法不良信息。这也说明了我国对 AI 数据训练采取了较为宽松的标准，并未否定在数据训练在法律上的合法性。

2.2. 生成式人工智能数据训练应当受到《著作权法》支持的必要性分析

数据训练是生成式 AI 的核心环节之一。通过大量的数据输入，实际上，用于数据训练的语料中，收到《著作权法》保护的作品比公共领域数据更多且质量更高，更符合生成式人工智能的训练需求。但是需要训练的数据之多，让生成式人工智能服务提供者获得许可的成本过高，可能会导致学习因著作权保护而终止。但是模型通过学习特定领域内的模式和规律，从而提高其预测能力和生成质量。例如，在自然语言处理(NLP)中，经过充分训练的模型能够更准确地理解语义、语法结构以及上下文关系，进而生成更加流畅、自然的语言表达。

人工智能技术的发展已经可以运用到各种不同的场景，也对生成式 AI 提出了不同的要求。但每个领域都有自己独特的规则和标准，针对具体行业的定制化数据集训练显得尤为重要。通过收集和标注行业特定的数据集，AI 运营商可以让模型更好地适应特定任务的需求，提供更为专业和精准的服务。这不仅提升了用户体验，也为各行业带来了更高的效率和价值。一个经过良好训练的生成式 AI 模型可以在未见过的数据上进行合理的推断和创造，而不会局限于训练集中已有的样本。数据训练为研究人员提供了探索新技术、新算法的平台。通过实验不同类型的训练数据和参数设置，才能发现潜在的改进方向，推动整个领域的进步。此外，数据训练不仅是技术层面的工作，还涉及价值观的传递。通过精心挑选和构建训练数据集，我们可以引导模型遵循正面的社会准则，避免产生有害内容或偏见。例如，在社交媒体平台上，使用经过审核的正面评论作为训练数据，可以帮助聊天机器人学会积极友善地与用户交流，营造和谐健康的网络环境。

生成式人工智能的数据训练不仅是实现高效、准确模型的基础，更是推动技术创新、满足多样化需求和履行社会责任的重要手段。随着技术的不断进步，我们将看到更多基于高质量数据训练的生成式 AI 应用于各个领域，为人类带来前所未有的便利和发展机遇，所以对于生成式人工智能的数据训练的正当性是无法忽视的，所以我们需要增加保障数据训练的法律规定，我国《著作权法》在近几年已完成了修改，不适宜再进行大幅修改，所以我们可以已有的合理使用规则为其提供保障。

3. 合理使用规则在生成式人工智能数据训练中的适用

尽管目前世界上还未对生成式人工智能数据训练的著作权法争议形成共识，对其侵权认定也因司法管辖而存在差异。美、韩两国多采用四要素法将利用作品数据进行训练的生成式人工智能归入合理使用的范畴；欧盟有条件地将 TDM 例外适用于生成式人工智能训练场景为 AI 运营商提供了合理预期。我国《著作权法》第 22 条虽然对合理使用采取了半封闭式的立法，但是在我国司法实践中已经有可以对合理使用进行扩展的可能[6]。同时最高法的司法政策也明确了法院在认定合理使用时可以参照四要素标准。[7]所以在发生数据输入侵权纠纷时，我国法院应该重视四要素标准对合理使用抗辩的进行有效性评估，并以此做出合理的判决，以保护作者著作权和人工智能产业的发展，实现利益平衡。

3.1. 四要素判断标准

四要素标准包括作品使用行为的目的和性质、被使用作品的性质、被使用部分的数量和质量、使用

[7]最高人民法院 2011 年颁布的《最高人民法院关于充分发挥知识产权审判职能作用推动社会主义文化发展大繁荣和促进经济自主协调发展若干问题的意见》第 8 条规定：在促进技术创新和商业发展确有必要的特殊情形下，考虑作品使用行为的性质和目的、被使用作品的性质、被使用部分的数量和质量、使用对作品潜在市场或价值的影响等因素，如果该使用行为既不与作品的正常使用相冲突，也不至于不合理地损害作者的正当利益，可以认定为合理使用。

行为对作品潜在市场或价值的影响。

3.1.1. 使用行为的目的和性质

美国著名的谷歌图书案[8]，联邦最高法院认为，使用行为的目的和性质不具有决定性意义，转化性使用是合理使用更重要的判断标准。美国版权联盟发布的《著作权与人工智能基本原则》中，使用受著作权保护的作品训练大模型或创建数据集构成合理使用[7]。转化性使用是判断使用行为与目的的重要方法，它是指新作品的目的并非是为了取代原作品，而是向原作品中加入新表述、新含义、新信息，使其目的或性质得以转变，以达到著作权法扩充公众知识的总体目的。转化性使用产生于美国自坎贝尔诉艾克夫柔丝音乐公司案确立其为法院判断合理使用的核心。转化性使用的本质就是生成式人工智能的数据从输入到输出结果的过程是对原作品添加了新的价值的过程，生成式人工智能的数据训练不是为了实现原作品本身的文学艺术目的，而是为了对输入数据进行数据挖掘，通过机器学习和模仿人类语言输出不同于原先作品的新内容，如此也可以起到激励社会创作的效果，从而实现《著作权法》的目的之二，也是生成式人工智能公共利益属性的体现。在生成式人工智能的数据训练、挖掘过程中，作品只是用来教导模型图像元素之间统计关系的数据，构成转化性使用；并且在数据挖掘中生成式人工智能仅学习作品代表什么，并据此创造出新文本和图像；生成式人工智能输出的作品存在于数据市场中，原作品则属于创意输出市场，数据训练并不会取代原作品，不会与原作品在相关市场上形成竞争关系。

3.1.2. 被使用作品的性质

被使用作品的性质是判断是否构成合理使用的第二要素，但并非是决定性要素。被使用作品既包括已出版的作品也包括未出版的作品，这一要素的判断依附于第一要素即转化性程度，因为当转化性程度非常高时，被使用作品的性质对于是否构成合理使用的判定就微乎其微。在谷歌图书馆案件中，法院认为谷歌公司对图书扫描的行为具备高度的转换性，所以无论谷歌公司扫描的书籍类型是什么，都不会影响合理使用的认定。并且生成式人工智能获取数据的途径主要是开放型数据库、公共领域数据资源、网络爬虫等途径，所以涵盖了几乎我们可见的全部作品类型。无论数据输入是何性质，只要被认定具有了高度的转换性使用，被输入作品的性质也不应当影响合理使用的认定。

3.1.3. 被使用部分的数量和质量

被使用部分的数量和质量也是基于转换性使用进行判断，当然这一因素还要求对作品的使用要以“没有超过必要的限度”为准。合理使用并不等同于少量使用，如果已经可以认定转换性使用，那么即使大量使用也可以构成合理使用。还是基于谷歌图书馆扫描案，谷歌公司对书籍进行全部扫描整体复制是有必要的，谷歌公司推出的关键词搜索和片段浏览都是以此为基础的，所以即使是进行了全篇复制也是可以认定为合理使用的，并且生成式人工智能生成的内容与训练数据之间也不会存在实质性相似问题，因为公众本来就无法接触原先作品的表达，那么输出内容也就难以成为对原作的竞争性相似问题。

3.1.4. 使用行为对作品潜在市场或价值的影响

使用行为对作品潜在市场或价值的影响，这一要素不仅要综合考量数据训练行为对传统市场的损害，还要考察其对原作品潜在市场造成的不利影响。换言之，数据训练行为若给原作品带来了竞争性替代的风险[8]，损害了著作权人的合法权益，造成了实质性收入的减少，那么再将该训练行为认定为属于合理使用则存在一定的障碍。当然这一要素仍然还要以转化性使用为前提，转换性程度越高，数据训练复制行为构成实质性替代的可能性就越小。实际上目前生成式人工智能与原作品实际上很难构成有竞争性的原作替代品，因为生成式人工智能基本上是是对原作品的简单概括，并不涉及段落或章节摘取。虽然现

[8]联邦最高法院承认谷歌将受版权保护的书籍数字化并收录到其搜索引擎中构成合理使用，优先考虑使用的变革。

在有生成式人工智能可以形成与艺术家风格一致的作品，但是我们要严格按照思想表达二分法，风格只是思想的领域，生成式人工智能实际创作的不与原作相同的独创性表达，所以可以认定生成式人工智能的数据训练行为属于合理使用。

3.2. 明确合理使用规则的适用前提

除了这种客观的判断标准，也有学者指出判断生成式人工智能数据训练是否应该适用合理使用规则还需要具备三个前提条件。第一、未经许可利用作品训练生成式人工智能，只有当这种侵权行为落入著作权专有权控制的范围时，才可能构成侵权。对于“非表达型机器学习”，尽管在输入和训练阶段涉及对作品的复制和转换，但这些行为并未直接展示作品的表达性内容，因此不应被视为作品性使用，也不构成侵权。相反，“表达型机器学习”中的输出阶段，如果生成的内容与原作品相同或相似，则可能侵犯著作权人的权利，属于“作品性使用”，需考虑是否适用合理使用规则。第二、遵循目前的授权使用规则是否会限制生成式人工智能产业的发展，合理使用规则因对作品保护明显限制了公众接触作品的可能，就必须使用其他机制打破这种限制，我国只有针对于个人利益限制著作权保护的合理使用规则，并未机器学习的合理使用规则，所以机器学习仍然应当遵循“先许可 + 付费，后使用”的授权使用规则，这时就需要考虑著作权法对作品的保护是否限制了机器人工智技术的发展，若该保护限制了人工智能技术的发展就需要平衡公共利益与个人利益之间的平衡，引入合理使用规则，反之，则不需要动用合理使用规则。最终实现促进社会公共利益与著作权人个人利益的相互平衡。第三、衡量其他简化授权机制是否难以平衡各方利益，《著作权法》除了有相关的“先许可 + 付费，后使用”的授权使用规则，也存在法定许可和合理使用的规则。但是合理使用规则对于著作权人权利限制较大，所以应当在其他授权机制不能满足公众接触作品的可能与著作权保护相适应的目标时，才可以引入合理使用规则。但如果法定许可等机制能够较好地解决训练数据的版权问题，则可以直接适用其他规则，而不必要引入合理使用规则[9]。

4. 生成式人工智能在输出数据侵权时责任承担

生成式人工智能在文本、图像、音频等多个领域的应用已经日益广泛，但是生成式人工智能输出的内容还存在偶发性的侵权问题，前文已经论述了生成式人工智能在数据输入阶段的数据训练可以基于合理使用规则被认定为合法，那么在生成结果也即数据输出上还需要进一步考虑其合法性。

目前判断是否构成著作权侵权的关键在于两个要素：接触与实质性相似。对于人工智能生成的内容，这一原则同样适用。首先，必须证明 AI 在生成过程中“接触”了受版权保护的作品，即训练数据中包含了该作品或其部分片段。其次，生成的内容必须与原作品之间存在“实质性相似”，即两者在表达上有显著的相似之处。然而，即使满足这两个条件，AI 运营商也不必然构成侵权，具体情况还需进一步分析。并非所有与原作品构成实质性的内容都构成侵权。在 AI 生成的内容基于的是已经进入公有领域的作品，那么即便存在相似性，也不涉及版权问题。此外，许多创作者通过知识共享许可协议(如 Creative Commons)授权他人在一定范围内使用其作品。在这种情况下，AI 运营商只要遵守许可协议中的条款，使用行为就是合法的。

当生成式人工智能的输出结果确实构成侵权的情形，AI 运营商就应当根据现行法律承担责任。笔者认为 AI 运营商的责任判定应遵循过错责任原则，即，只有在运营商存在主观过错的情况下，才需要承担侵权责任。具体而言，法院应考察运营商是否尽到了合理的注意义务，包括事前预防措施和事后纠正措施。AI 运营商应当采取必要的版权过滤措施，以避免生成内容侵犯他人版权。这包括但不限于对训练数据进行筛选，确保不包含未经授权的作品；以及在技术层面上优化算法，减少生成内容与已有作品的相似度。虽然现有的 AI 技术无法完全杜绝侵权内容的生成，但运营商应尽可能采取措施降低风险。例如，

通过设置关键词过滤、内容审查机制等方式，防止明显的侵权行为发生。除了事前的预防措施，AI 运营商还可以通过及时采取纠正措施来减轻或者免除责任。首先，运营商应建立有效的举报投诉机制，方便版权人或其他用户发现并报告侵权行为。一旦接到侵权通知，运营商应在合理期限内进行调查，并采取相应的处理措施，如删除侵权内容、修改算法等。其次，运营商还应对算法进行调整，防止类似侵权内容再次生成。例如，在“奥特曼案”中，被告因未建立有效的投诉举报机制而被法院认定为存在过错，最终被判承担赔偿责任。

值得注意的是，AI 技术本身具有一定的局限性，尤其是在算法黑箱的情况下，运营商难以对生成内容进行全面审查。由于 AI 模型的复杂性和输入数据的庞大性，AI 运营商无法保证每一项输出都完全符合《著作权法》的要求。因此，在评估 AI 运营商的责任时，法院应当考虑到技术水平的限制，不宜过度加重其义务。若 AI 运营商已经采取了合理的预防和纠正措施，但由于技术原因仍未能完全避免侵权内容的生成，法院可以酌情减免其责任，甚至免除赔偿责任。这种做法不仅有助于促进 AI 技术的发展，也为行业创新提供了更为宽松的法律环境。

5. 结论

生成式人工智能的数据训练不仅是实现高效、准确模型的基础，更是推动技术创新、满足多样化需求和履行社会责任的重要手段。随着技术的不断进步，会有更多基于高质量数据训练的生成式人工智能将要应用于各个领域，为人类带来前所未有的便利和发展机遇。正确认识生成式人工智能的数据训练的正当性是无法忽视的，必须在法律框架内为其提供必要的保障。应当为其在《著作权法》内寻找法律支撑——合理使用规则。在人工智能进行机器学习的非表达型机器学习不应当被认定为是对已有作品的侵犯，当然这一认定规则也不能无限度地加以适用，还需要综合考量合理使用规则的适用前提与认定标准，确保在数据输入阶段既不阻碍技术发展，又不损害著作权人的合法权益。并且还要明确在数据输入与输出发生侵权行为时，AI 运营商应当按照过错责任原则和“安全港”原则承担相应的责任，采取合理的预防和纠正措施，包括建立版权过滤措施、内容审查、举报投诉机制等，以减少侵权行为的发生。法院在评估运营商的责任时，应考虑到技术水平的限制，不宜过度加重其义务，从而为行业创新提供更为宽松的法律环境。生成式人工智能的发展离不开高数量和高质量的数据训练挖掘，大数据技术的发展不止是企业之间的竞争，更是国家之间的竞争，合理的法律规制是这一进程的重要竞争，法律法规不断完善可以为生成式人工智能的发展提供坚实的支撑，同时也可以促进整个人工智能行业健康、有序发展。

参考文献

[1] 刘强, 孙青山. 人工智能创作物著作权侵权问题研究[J]. 湖南大学学报(社会科学版), 2020, 34(3): 140-146.

[2] 陈锐, 江奕辉. 生成式 AI 的治理研究: 以 ChatGPT 为例[J]. 科学学研究, 2024, 42(1): 21-30.

[3] 冯晓青. 论著作权限制的合理性及其在著作权制度价值构造中的意义[J]. 湖南社会科学, 2011(5): 49-52.

[4] 张平. 人工智能生成内容著作权合法性的制度难题及其解决路径[J]. 法律科学(西北政法大学学报), 2024, 42(3): 18-31.

[5] 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.

[6] 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.

[7] 廖小莉, 潘凤湘. 生成式人工智能数据挖掘合理使用适用性及规范路径[J]. 产业创新研究, 2025(6): 14-18.

[8] 张镇涛. 人工智能生成作品的著作权之问[J]. 法制与社会, 2020(2): 214-215.

[9] 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.