

生成式人工智能时代文本与数据挖掘的著作权合理使用规则构建

汪 骋

暨南大学法学院/知识产权学院, 广东 广州

收稿日期: 2025年10月3日; 录用日期: 2025年10月29日; 发布日期: 2025年11月5日

摘 要

生成式人工智能模型在数据输入及训练阶段需依赖对海量文本、图像等作品的复制与预处理, 然而我国现行著作权法却并未为此类行为提供合法基础, 导致相关作品的复制权、演绎权等著作权侵权风险不断累积。文本与数据挖掘版权例外规则作为一项平衡技术发展与著作权人利益的制度创新, 逐渐在国际社会得到认可与广泛应用。本文以欧盟已经相对成熟的文本与数据挖掘版权例外规则为蓝本研究, 探寻这一制度的具体内容和司法实践效果并进行中国镜鉴, 最后提出把文本与数据挖掘行为纳入著作权法合理使用情形的合法路径, 同时建议配套必要的技术措施、设置开发者披露义务, 并增加著作权人权利平衡和补偿机制。

关键词

生成式人工智能, 文本与数据挖掘, 数据训练, 著作权法, 合理使用

Construction of Fair Use Rules for Copyright in Text and Data Mining in the Era of Generative Artificial Intelligence

Cheng Wang

Law School & Intellectual Property School, Jinan University, Guangzhou Guangdong

Received: October 3, 2025; accepted: October 29, 2025; published: November 5, 2025

Abstract

Generative artificial intelligence models, in the data input and training stages, rely on copying and pre-processing vast amounts of works such as texts and images. However, China's current Copyright

Law does not provide a lawful basis for such activities, leading to a continual accumulation of copyright infringement risks relating to reproduction and adaptation rights of the relevant works. The copyright exception rules for text and data mining (TDM), as an institutional innovation balancing technological development with the interests of copyright owners, have gradually gained international recognition and widespread application. This paper takes the EU's comparatively mature TDM copyright exception framework as its blueprint, investigates the specific content of this regime and its judicial practice effects, and offers a Chinese mirror assessment. Finally, it proposes a lawful path to incorporate TDM activities within the legitimate uses under copyright law, while also recommending the accompanying technical measures, the introduction of developers' disclosure obligations, and the strengthening of balance and compensation mechanisms for copyright owners.

Keywords

Generative Artificial Intelligence, Text and Data Mining, Data Training, Copyright Law, Fair Use

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. 引言

近年来,生成式人工智能技术的迅速发展及广泛运用,彻底改变了人类智力成果的传统创造模式。ChatGPT、Deepseek 等人工智能大模型借助海量的信息摄入、学习、训练与测试造就了极强的输出能力与效率。然而,生成式人工智能在运行过程中也面临了复杂的著作权法问题,尤其在数据输入及训练阶段,模型需对现有作品进行大规模复制与预处理,这些行为可能触及著作权法上的复制权等专有权利,我国现行法律却未能对这些问题进行回应,导致人工智能开发过程存在较大的侵权风险。在此背景下,文本与数据挖掘(Text and Data Mining, 简称 TDM)版权例外规则逐渐成为学界与立法机关的关注焦点。作为一项国际层面的制度创新,TDM 版权例外规则已在欧盟、日本、英国等法域得到不同程度的确立与适用,其中欧盟 TDM 版权例外规则又因规则完备程度较高和布鲁塞尔效应而拥有显著的国际影响力。

本文充分考察欧盟 TDM 版权例外规则的具体内容并分析其优势与不足之处,于扬弃的基础上提出在我国著作权法框架中将 TDM 纳入合理使用范围的路径以纾解当前的 TDM 行为合法困局,并配备对应的技术措施如水印类技术等,与此同时增加生成式人工智能开发者披露义务和相关作品著作权人的利益平衡机制。通过上述制度设计,力求为生成式人工智能的 TDM 行为提供明确的合法性依据,化解模型开发潜在的著作权侵权风险,进而激励更多主体投入生成式人工智能的研发与升级,实现利益均衡下的联动发展。

2. 生成式人工智能 TDM 行为的著作权合法困局及纾解路径

2.1. 生成式人工智能 TDM 行为的著作权侵权风险

生成式人工智能的运行过程主要分为输入-训练-输出三个阶段,输入阶段通过网络爬虫等技术收集各种文本、图像、音视频等资料,数据训练阶段由模型算法提取内在规律,最后根据用户的指令基于现有的输出能力给出相应的结果。在输入及训练阶段,大量的文本与数据是人工智能模型创作能力的养分,模型通过 TDM 行为学习并挖掘海量数据中的普遍规律,TDM 技术在这一过程中的功能主要是对搜集到的数据进行自动分析和特征提取,以得到有价值的信息或趋势[1],通过这一技术运作过程生成式人工智能才能够具有且不断提高其输出能力。正如吴汉东教授所言:“文本和数字挖掘,可视为从另一角

度对人工智能技术过程的表述” [2]。

2023 年 11 月, 小红书及其旗下 Trik 软件的主体公司, 由于未经授权使用作品训练模型并生成风格与原作类似的作品而被起诉。2024 年 6 月, 北京互联网法院审理了这一案件¹。这并非个例, 国内人工智能企业的井喷式发展已使得同类型的争议与相关诉讼案件数量剧增, 生成式人工智能在数据输入及训练阶段的合法性评价, 已成为亟待解决的问题。《生成式人工智能服务管理暂行办法》这一初步的法律规制框架对人工智能输出内容的部分侵权问题作出了回应, 但关于数据训练的输入端问题如何处理答案依然悬而未决。生成式人工智能具有较强的数据依赖性, 训练数据的数量与质量是人工智能输出能力的决定性因素[3], 极大影响了人工智能所能达到的创作高度。但在我国的著作权法体系之下, TDM 行为尚且缺乏合法依据, 可能面临多种侵权风险。

在数据输入与训练阶段, 对不同来源数据的收集与学习涉及对相关作品的复制与预处理行为, 可能侵犯的著作权主要为复制权和改编、翻译、汇编等演绎权。目前市面上生成式人工智能的数据训练库分为模型开发者自建自用的数据集和第三方建立再授权模型开发者使用的数据集两种情形, 但无论建立主体是哪一方, 都要经历两个阶段。阶段一需通过爬虫等技术获取各种网络作品信息, 再将云端作品复制存储, 此类复制行为涉嫌侵犯所使用作品的复制权。阶段二则需对收集到的信息进行预处理, 包含对信息进行翻译、删改、整理、汇总、格式转化等行为, 有侵犯他人作品翻译权、改编权和汇编权的风险。具体而言, 阶段二为了将阶段一收集到的文本与数据转换为机器可读的形式, 需要将其转化统一格式, 这一行为改变了数据和文本的外在形式, 但实质内容依然相似, 此种转换与改编行为类似, 亦可将其解释为对原始文本数据的翻译行为, 因此阶段二进行的数据预处理可能涉嫌侵犯作品的翻译、改编、汇编等演绎权。在“全球 AIGC 平台侵权第一案”的奥特曼案²中, 法院认定人工智能开发者侵犯了原告的复制权与改编权, 这一司法判决就佐证了生成式人工智能著作权侵权风险的真实存在。

2.2. 生成式人工智能 TDM 行为的合法路径探索

在我国著作权法框架中, TDM 行为可尝试的合法路径有三种, 分别为法定许可、授权许可与纳入合理使用范围, 而三者相较之下, 合理使用或能成为生成式人工智能 TDM 行为合法化的最优解。

法定许可的适用范围相当严格, 根据当前法律规定, 主要可以被划分为教科书编写、报刊转载、录音制作和广播电视播放四种情形, 法定许可无需著作权人同意, 但使用后需要按照法律规定向权利人支付报酬, TDM 行为显然不属于现行著作权法四类法定许可情形里的任何一种。如果要突破现有范围, 为 TDM 行为专设法定许可, 也会在确权、分配等环节面临多种难题, 与之配套的规定与职能相当复杂, 可操作性较弱。

授权许可的方式对人工智能企业而言合法成本则极高, 逐一获得著作权人的许可在实践层面几乎不可能完成。数据显示, GPT 系列人工智能模型预训练使用的开源数据库中, BooksCorpus 包含超过 7000 万个句子, Wikipedia 包含 42 亿个词、The Pile 包含 825 GB 数据[4]。由于数据训练阶段涉及的作品数量和范围都极大, 因此得到所有相关权利人的授权许可绝非易事。一方面, 许多网络资源并未标明权利人, 要明晰所有权利来源不具有现实可能性, 去查证的难度和时间成本都相当大且未必能达成。此外, 在互联网数据具有跨国性特征的背景下, 如遇非本国的数据及文本, 查明具体权利人和作品其他相关信息更是难上加难。另一方面, 对于有明确权利人的作品, 若要依次进行谈判获得许可并给付报酬, 谈判与经济成本的总和会远远超出人工智能数据集开发者的承受能力。若坚持要求海量的训练数据都需获取授权后才能使用, 合法可用的训练数据将大幅缩水, 导致的只有两种后果: 一部分人工智能开发者对侵权风

¹北京互联网法院. 小红书 Trik 软件 AI 绘画大模型训练著作权侵权案[EB/OL]. 2024-06-20. <https://ipc.court.gov.cn/zh-cn/news/view-4513.html>, 2025-10-18.

²广州互联网法院, 上海新创华文化发展有限公司诉某人工智能公司著作权侵权纠纷案, (2024)粤 0192 民初 113 号。

险望而生畏，只能在公共领域的过时内容上培训他们的模型，致使技术发展停滞；还有部分开发者可能会认为违法的收益大于风险，选择采用更隐蔽的方式违反授权规则，导致“违法过剩”的局面，引发治理困难并影响法律的权威。

我国现行的著作权合理使用规则尚且无法涵摄 TDM 行为。《中华人民共和国著作权法》(简称《著作权法》)第 24 条规定了合理使用的“三步检验法”、12 种著作权具体的合理使用情形与“法律、行政法规规定的其他情形”兜底条款，人工智能数据输入及训练阶段中涉及的 TDM 行为并不在这 12 种具体情形的范围内，与已有可能相关情形进行关联的难度也较大。由于 TDM 行为的主体是自动化分析技术及背后的人工智能，多以商用目的收集并处理海量数据，所以较难符合“个人合理使用”和“教学或科研少量复制合理使用”等教育和公益性场景。此外，在其他法律法规中并没有对这一行为的版权例外豁免，所以在司法实践中援引兜底条款的难度也较大，那么 TDM 行为就不能依据现有的著作权法法律框架获得合理使用的豁免。然而生成式人工智能的数据输入及训练阶段的 TDM 行为属于非表达型机器学习，仅涉及收集的文本及数据本身，并不会使用和传播原作品的独创性表达^[5]，因而不会损害相关权利人的利益，将其认定为合理使用行为合乎我国著作权法合理使用制度的内在逻辑与价值取向，相较于法定许可和授权许可面临的理论与现实窒碍也少得多，将其作为 TDM 行为的合法路径选择是当前法律框架之下的最优解。

3. 欧盟 TDM 版权例外规则的内容及其适用评析

3.1. 欧盟 TDM 版权例外的具体规则和司法实践

欧盟的数据及人工智能立法位居全球前列，相关规则十分周详，2019 年 3 月，欧盟通过《数字化单一市场版权指令》(Directive on Copyright in the Digital Single Market, 简称 DSM 指令)，该指令将 TDM 定义为任何旨在分析数字文本和数据以便生成包括但不限于模型、趋势、相关性等信息的自动分析技术^[6]。

DSM 指令中的 TDM 版权例外规则的条件可以从不同的维度综合概括。在行为类型上，可被豁免的行为限于 TDM 过程中涉及的“复制”和“提取”两种行为，但不得超出合理范围或构成对作品的实质性替代。在适用目的上，《DSM 指令》第 3 条确立了以科学研究目的 TDM 的强制性版权例外规则，保障研究机构和文化遗产机构的合法使用；而第 4 条则进一步规定了出于一般性目的的 TDM 非强制性版权例外，允许包括商业公司在内的更广泛的主体开展 TDM 行为^[7]。此外，在作品客体方面，TDM 版权例外规则面向的作品客体需由合法渠道获取。最后，在法律效力上，需注意任何版权人或数据库权利人都不得用合同条款排除适用以科学研究为目的的 TDM 版权例外规则，该规定保障了这一规则的有效实施；但对于不以科学研究为目的的一般性 TDM 版权例外规则，权利人可以通过机器可读的方式声明“保留权利”予以限制。

2024 年 3 月 13 日，欧洲议会批准通过了《人工智能法案》(Artificial Intelligence Act)，这是世界首部针对人工智能领域的全面监管法规，法案规定了通用人工智能模型的提供者在责任主体、TDM 许可与例外等维度所需履行的合规义务^[8]。法案还对 TDM 增设了更高的透明度义务要求，要求使用受著作权保护的作品时记录并公开摘要，摘要需包含对作品充足详细的使用情况。《人工智能法案》除了进一步强调对 TDM 行为合法性的认可，其对人工智能模型透明度和开发者公开义务的规定亦为 TDM 在合法轨道上运行提供了指引。

欧盟 TDM 的相关规定将绝对豁免主体的范围局限于科研机构与文化遗产机构，在德国汉堡地区审理的 Kneschke & LAION 一案³中法院又承认非营利组织 LAION 以人工智能的数据训练为目的自动下载图片的行为具有合法性，属于非商业性的 TDM 合理使用。这一司法判决标志着 TDM 版权例外规则突破

³Hamburg Regional Court. Robert Kneschke v. LAION e.V. Case No. 310 O 227/23.

了原有规定中以科学研究为目的的桎梏，开始向人工智能数据训练场景拓展，然而这一规则的适用范围在实践中并未达成统一。对 Google 公司的数据训练行为，法国法院就拒绝引用 TDM 规则给予豁免，声明不符合该豁免规则的非商业性使用前提。司法实践证明，欧盟的 TDM 版权例外规则在科学研究与商业应用之间尚未实现平衡，还需进一步探索如何调整相关规则。

3.2. 欧盟 TDM 版权例外规则的优劣评析与中国镜鉴

欧盟的 TDM 版权例外规则具有前瞻性与突出优势，通过专门立法的方式为 TDM 行为提供了法律基础，同时明确了 TDM 的合法性边界，降低了人工智能活动在版权合规上的不确定性，降低知识的传播成本进而提高了社会公共知识的数量与质量，产生了巨大的社会效益。此外，立法对训练数据透明度的重视则有助于建立权利人监督与问责的制度前提，促成人工智能规范发展与著作权人权益有效保障的双赢局面。

然而主要由 DSM 指令建立起的欧盟 TDM 版权例外规则，在司法适用中也产生了诸多不同的理解与争议。其一，一般目的的 TDM 行为涉及的作品允许其著作权人选择权利保留，但海量作品的权利主张与退出声明追踪实际操作起来难度较大，成本与技术要求都很高。其二，在欧盟 TDM 版权例外制度下，著作权人拥有选择退出权的强保护模式可能加剧技术创新与著作权保护之间的张力，从而导致人工智能开发者无法真正放心、大规模、无障碍地使用所需数据，阻碍公共领域的作品使用[9]。其三，司法实践中各成员国法院在 TDM 行为的“科研”与“商业”主体及目的判定上存在分歧，导致商业主体的合规成本与面临的法律风险不够明确，不能拥有稳定的合法预期。

我国关于合理使用的立法体例与欧盟相似，但是在欧盟已经赋予人工智能 TDM 行为涉及的作品使用合法性的背景下，我国面对这一问题依然未置一词。这说明我国的著作权制度已与蓬勃发展的生成式人工智能技术脱节[10]，亟需对相关问题作出法律上的回应，可以参考欧盟法的相关规定，与国际接轨从而适应人工智能时代的创新需求，并以立法和司法实践贯彻“激励创新、包容审慎”的人工智能治理理念。

4. 生成式人工智能 TDM 行为合理使用规则的中国方案

4.1. 明确 TDM 行为作为合理使用具体情形的要件

如上文所述，将生成式人工智能的 TDM 行为划入著作权法的合理使用范围有理论基础支撑和实践呼唤，亦有国际经验参考。由于我国在 2020 年已经对著作权法进行了第三次修订，为了维护法律的稳定性，可考虑先以司法解释或者著作权法实施条例的形式先行调整，根据《著作权法》第 24 条第 1 款第 13 项“法律、行政法规规定的其他情形”把 TDM 行为纳入著作权合理使用的范围中，为司法实践遇到的相关著作权问题争议提供裁判依据，待日后条件成熟之后再通过修订著作权法的方式把 TDM 行为加入著作权合理使用的具体情形中。司法解释或实施条例先行可以极大地提升相关制度的调整效率，也能够根据其适用效果再决定修法的具体事宜。至于 TDM 行为合理使用规则的具体要件，可从主体、客体、行为、目的四个方面进行规定。

4.1.1. 主体应具备开放性

欧盟的 TDM 版权例外规则的收益主体限定为研究组织和文化遗产机构，然而生成式人工智能模型的研发机构多为商业主体而非科研机构，如果将主体限定为科研机构，作为我国当前人工智能模型研发主力军的大型科技和软件公司会十分受限，而本身可利用资金资源就十分有限的初创公司和中小型公司更会失去创新激励，最终阻碍人工智能产业的持续发展[11]。我国宪法规定应保护公民进行科学研究、文艺创作和其他文化活动的自由。所以 TDM 版权例外的适用主体应当突破科研机构的范围，保持主体的

开放性,让这一制度便利亦能惠及其他组织与个人,给予人工智能初创企业、互联网科技企业以及致力于科学研究的机构和个人适当的 TDM 版权例外豁免权利,促进科学研究与创新,保证 TDM 技术的充分应用与发展。

4.1.2. 行为限定为临时复制及演绎行为

合理使用的行为类型应当涵盖临时复制行为和翻译、改编、汇编等演绎行为。为复制和演绎行为赋予合法性是推动 TDM 使用的必要保障,但应注意这些行为应当限定在必要限度内。若 TDM 实施过程中对作品的复制和演绎行为超出了学习范畴,进行了实际的传播,就可能引发对原有作品潜在或实际的替代效应。所以应只给予数据输入及训练阶段因技术过程需要产生的临时性、附带性的复制与演绎行为,为验证模型生成结果亦可对复制件进行一定时间的保存,但需在尽可能短的时间内删除复制件。

4.1.3. 客体应为合法渠道获取的作品

为了实现生成式人工智能开发者与数据训练相关作品著作权人之间的利益均衡,应当要求 TDM 行为只能对通过合法渠道获取的作品实施。由于实践中文本及数据种类繁多、来源复杂,溯源并判定其合法性也较为困难,所以可以考虑建立数据治理沙盒并鼓励人工智能开发者参与沙盒实验,通过观察沙盒中的实际情况辅助合理使用的判断。比如可以在沙盒中向人工智能开发者提供训练数据,按照风险等级将这些数据划分为高风险(如未经授权的专有作品)、中低风险(如用户生成内容)与无风险数据(如公共领域的免费开源数据),并观察各类数据在收集、转化和使用过程中遇到的问题。通过沙盒中对各类型数据的合法风险评估,在司法实践中更游刃有余地判断 TDM 行为涉及作品来源的合法性。

4.1.4. 目的包括科学研究等公益目的和其他合理的商业目的

针对 TDM 行为的版权例外情形,欧盟对科学研究目的和一般目的进行了区分,采用强制许可和一般例外的双轨制。我国的 TDM 合理使用规则不应排除出于商业目的的 TDM 行为,出于科学研究等公益目的与出于商业目的的 TDM 行为均有值得保护的社会效益。

为科学研究、公共安全、医疗健康等公益目的允许生成式人工智能大模型通过 TDM 行为展开数据训练与模型进化,其研究成果能够提升全社会的知识质量并增加社会知识总量,进而更大限度地发挥作品的社会价值^[5],为实现公共利益作出贡献。欧盟禁止商业性 TDM 的核心法理基础为防范市场替代效应,而非否定技术中立价值。在我国著作权法的“三步检验法”已能作为风险控制避风港的前提下,可以利用这一方法严格审查出于商业目的的 TDM 行为的市场替代效应。当 TDM 行为满足“署名义务”“不影响作品正常使用”“不损害权利人合法权益”三个要件时,TDM 的商业属性并不会形成合法障碍,因此在这种情况下应当认定该类商业目的的 TDM 行为是合理使用,构建涵盖范围更广的 TDM 合理使用机制。

4.2. 完善保障 TDM 行为合理使用规则实施的相关技术措施

当前著作权权利人使用技术措施保护作品的情况屡见不鲜,但部分权利人采取的技术保护措施已经不仅仅限于确保数据库的作品的安全性和完整性,保护程度已大大超出风险水平,阻碍了 TDM 的实施^[12]。在此背景下 TDM 行为要顺利实施必然涉及合法规避技术措施,这一措施需要相关法律及时完善技术规避例外情形认定予以配合^[13]。

此外,为平衡著作权风险与生成式人工智能模型性能,理论与实践正积极探寻能够干预人工智能模型数据输入及训练阶段的技术工具,借助合规技术手段保障人工智能最大限度符合现有的著作权法规定。循此思路可尝试通过水印类技术和分析数据归因类技术的介入规范人工智能的 TDM 行为,水印类技术通过在文本及数据中嵌入不易察觉的信号以追踪相关作品的来源及许可情况,分析数据归因类技术则能

够通过事后分析的方法衡量特定作品对输出内容的贡献，从而判断 TDM 行为对该作品的使用是否超出合理使用范围或者授权界限。

4.3. 设置人工智能开发者披露义务并建立著作权权利人利益共享机制

我国人工智能产业采用包容审慎的治理理念，2023 年 8 月颁布的《生成式人工智能服务管理暂行办法》强调统筹人工智能发展与安全、平衡监管与创新。为了保障文本与数据挖掘的合理使用条件真正有效实施，确保 TDM 行为收集的文本和数据通过合法渠道获取，避免人工智能开发者滥用技术目无法度地收集作品，要求 TDM 的应用者承担一定的披露义务，以保证监管行为顺利展开，平衡著作权人、技术拥有者和社会公众等多方利益。具体而言，生成式人工智能的数据输入及训练阶段，所使用的数据文本的来源、处理与使用必须具有高度的透明度和可追溯性，在进行合理使用时应当公开作品名称及著作权人的姓名，减少算法黑箱带来的不确定性^[14]。也需对人工智能开发者提出在合理限度内公开技术原理的要求，让著作权人拥有知晓其作品存在侵权风险的可能性。

生成式人工智能 TDM 行为合理使用规则的设计与实施还需兼顾著作权人的利益，若完全放任 TDM 行为，可能削弱权利人对其作品的控制权，影响市场收益。因此可以考虑引入著作权集体管理组织参与许可环节，考虑要求使用人为合法来源的作品支付合理的许可费，同时配套补偿金制度允许著作权人参与收益分配，在降低交易成本的同时达到利益均衡。集体管理组织可以帮助著作权人将其作品授权人工智能开发者使用，亦可借助技术手段实时获得著作权费用，保证权利人能够享受作品带来的收益，实现数据共享与开放。

5. 结语

TDM 是生成式人工智能时代的核心技术之一，对于促进人工智能产业繁荣、推动经济发展有着关键作用。欧盟为了适应技术跃迁作出的立法与制度创新值得我国借鉴，我国应当将生成式人工智能的 TDM 行为纳入著作权法合理使用的范围，可采用解释论的方法，通过司法解释或著作权法实施条例的形式对《著作权法》第 24 条合理使用的兜底条款进行解释确立 TDM 版权例外规则，日后再修订《著作权法》正式将 TDM 行为纳入合理使用的具体情形。此外，还需为 TDM 合理使用规则的实施配备必要的技术措施与开发者义务，并设计利益共享机制，以平衡著作权人权利与公共利益，进一步完善人工智能领域的战略布局。生成式人工智能时代著作权法的传统利益平衡状态被打破，在新时代处理好著作权人权利与技术发展之间的关系具有重大意义。只有构建适应我国国情的 TDM 合理使用新型规则，筑建起 TDM 行为的合法基础，才能为人工智能产业发展提供坚实的制度保障，帮助我国在人工智能科技的大国竞争中占据优先地位。

参考文献

- [1] 包赛君, 肖冬梅. 生成式人工智能训练数据的著作权法因应: 欧盟版权例外规则及其对我国的启示分析[J]. 图书馆论坛, 2025, 45(6): 123-132.
- [2] 吴汉东. 著作权合理使用制度研究[M]. 北京: 中国人民大学出版社, 2020: 236.
- [3] 王振硕. 生成式人工智能时代数据安全法益的保护模式塑造[C]//《新兴权利》集刊 2023 年第 2 卷——生成式人工智能法律问题研究文集. 天津: 南开大学法学院, 2024: 94-106.
- [4] 哈尔滨工业大学自然语言处理研究所. ChatGPT 调研报告[EB/OL]. 2025-09-24. <https://cloud.tencent.com/developer/news/1022565>, 2025-10-18.
- [5] 包赛君, 唐思慧. 生成式人工智能模型训练中的作品合理使用问题研究[J]. 图书馆建设, 2025(3): 48-59.
- [6] 樊桔玺. 数字环境下文本与数据挖掘版权例外规则构建[J]. 图书馆工作与研究, 2024(9): 26-35.
- [7] EUR-Lex (2019) Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright

-
- and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC (Text with EEA Relevance. <https://eur-lex.europa.eu/eli/dir/2019/790/oj/eng>
- [8] 刘诚. 生成式人工智能数据训练的版权保护及其国际经验[J]. 重庆社会科学, 2025(7): 6-17.
- [9] 张涛. 人工智能大模型训练的著作权困境及其调适路径[J]. 现代法学, 2025, 47(2): 189-208.
- [10] 马一德, 汪婷. 人工智能训练数据版权侵权风险规制: 欧盟实践、本土困境与解决路径[J]. 德国研究, 2025, 40(1): 82-99, 150-151.
- [11] 尹待龙. 人工智能时代文本与数据挖掘著作权例外规则研究[J]. 武汉交通职业学院学报, 2024, 26(2): 51-57.
- [12] 张惠彬, 肖启贤. 人工智能时代文本与数据挖掘的版权豁免规则建构[J]. 科技与法律(中英文), 2021(6): 74-84.
- [13] 沈廖佳, 龙燕燕. 文本与数据挖掘的著作权风险及其层次化应对——基于生成式人工智能的展开[J]. 中国政法大学研究生学报, 2025(1): 22-45.
- [14] 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.