生成式人工智能数据训练的著作权合规性困境和治理路径

谢黎伟 潘 圣
(福建理工大学 法学院,福建 福州 350118)

[摘 要] 在生成式人工智能数据训练环境下,传统许可使用机制实施受阻,既有权利限制条款适用不能,从而引发著作权合规性困境。对此,国内外理论与实践提出了不同规制路径,一是以"非表达性使用"理论为代表的"釜底抽薪"著作权除外方案,二是以美、日、欧盟为代表的"先进后出"著作权例外方案。基于生成式人工智能作品使用特性与权益平衡的考虑,应在借鉴日本、欧盟"类型主义"合理使用立法体例的基础上,构建生成式人工智能数据训练合理使用条款,并结合美国"要素主义"合理使用的立法体例,明确"三步检验法"的合理使用判定要素,同时辅之以信息数据披露、创作者适当补偿等配套制度,以此回应生成式人工智能数据训练导致的著作权问题,促进人工智能技术进步与著作权保护的协调发展。

[关键词] 生成式人工智能;数据训练;非表达性使用;合理使用
[中图分类号] D923.4 [文献标识码] A [文章编号] 1008-7699(2026)03-0048-12

一、问题的提出

在数字化时代,数据成为作品的无形载体。生成式人工智能的发展离不开海量数据的"投喂"。当生成式人工智能的训练数据涵盖受著作权法保护的作品时,数据训练行为可能引发著作权侵权风险。目前,国内外已发生多起此类著作权纠纷,如八位美国作家起诉谷歌公司未经许可使用其作品来训练生成式人工智能模型①;国内上海新创华文化公司起诉杭州水母智能科技公司著作权侵权,诉请被告删除其训练数据集中受著作权保护的奥特曼物料②。此类著作权侵权纠纷的集中出现,凸显了生成式人工智能数据训练所面临的严峻著作权合规性挑战,亟需法律制度对新兴技术进行规制与调整。

为有效应对此挑战,我国政府部门颁布了相应的规范性文件。以《生成式人工智能服务管理暂行办法》为例,该办法明确要求生成式人工智能服务提供者必须采用来源合法且不侵害他人知识产权的数据进行训练处理。③ 同时,《生成式人工智能服务安全基本要求》在语料安全章节中规定,若语料存在知识产权侵权问题,服务提供者不应使用相关语料进行训练。④ 但上述规定多为部门规

[收稿日期] 2026-01-06
[基金项目] 福建省社科基金重点项目(FJ2023MGCA033)
[作者简介] 谢黎伟(1972—),男,福建安溪人,福建理工大学法学院副教授、硕士生导师,法学博士。

① L. v. Google LLC, No. 3:23-cv-03440-AMO (N. D. Cal. 2023).
② 参见广州互联网法院(2024)粤0192民初113号民事判决书。
③ 国家网信办、国家发展改革委、教育部、科技部、工业和信息化部、公安部、广电总局联合发布的《生成式人工智能服务管理暂行办法》,第7条第1项、第2项。
④ 全国网络安全标准化技术委员会发布的《生成式人工智能服务安全基本要求》,第5章第2节。

章或政策文件，法律位阶较低，而且内容仅为方向性规定，缺乏具体规范。作为上位法的《中华人民共和国著作权法》（以下简称《著作权法》）也缺乏相应的明确规定，著作权限制与例外条款在适用上存在指引不足。

对此，学界将如何平衡内容创作者权益保护与生成式人工智能技术发展作为当下研究的热点问题，但相关研究尚未形成共识。现行法规下，生成式人工智能数据训练面临何种著作权规制困境？著作权除外方案与著作权例外方案，何种路径更适宜回应数据训练引发的著作权合规性问题？如何具体构建方案以平衡技术发展与创作保护的需求？在此争议背景下，本文立足于生成式人工智能的技术原理，围绕生成式人工智能数据训练的著作权困境，通过分析生成式人工智能著作权法规制的不同进路，探讨数据训练法律规制中国路径的选择与构建，以此推动人工智能技术进步与著作权保护的协调发展。

二、生成式人工智能数据训练的著作权困境

现行著作权法下，生成式人工智能模型训练使用受著作权保护的作品可能引发著作权侵权纠纷。因而，有必要对生成式人工智能数据训练过程中可能涉及的侵权风险进行分析，并从传统许可使用机制实施受阻、既有著作权限制条款适用不能的层面，深入揭示现行著作权法对生成式人工智能数据训练法律回应的不足。

（一）生成式人工智能数据训练的侵权风险

根据生成式人工智能的技术原理，其数据训练可分为数据输入阶段、模型学习阶段、模型优化阶段。

在数据输入阶段，生成式人工智能开发者存在侵犯著作权人复制权、署名权、保护作品完整权的风险。生成式人工智能在该阶段有数据收集和预处理两个环节。在数据收集环节，由于生成式人工智能训练所需数据量巨大，该环节开发者一般通过网络爬虫等技术手段，在网页中爬取数据，并采用下载或者上传云端等方式，将作品固定至服务器或本地设备，进而构建训练数据库。当爬取的数据涉及受著作权法保护的作品时，该收集行为由于再现了作品完整表达并且相对稳定和持久地固定在机器中，属于著作权法上的“复制行为”，存在复制权侵权风险。数据预处理环节包括数据转换、数据清洗、筛选、去重、切片、数据标注等步骤。在上述步骤中，模型需要对作品数据进行转码、分段，转化为机器可读格式，并多次临时或永久储存数据，该过程仍以数字化方式再现作品表达并加以固定，涉及重复复制，可能侵犯复制权。同时，模型在数据清洗中存在删除作者姓名、版权声明、来源标识等权利管理信息的技术行为，该主动去除信息的行为未经权利人许可容易导致作者身份无法被识别，涉嫌侵犯署名权。此外，数据筛选、去重、切片等技术行为涉及对作品数据的截取、删改、拼接、割裂上下文逻辑等内容的实质性改动，可能歪曲作品原意、破坏作品结构完整性，涉嫌侵犯保护作品完整权。

在模型学习阶段，模型不再是简单复制作品的外在表达，而是学习作品的思想与语言习惯。与人类大脑学习相通，生成式人工智能通过模型处理数据，再以新的表达方式展现出来。以ChatGPT3为例，此阶段Transformer大模型通过编码器中的多层神经网络架构先对数据进行拆解、抽象与重组，提炼作品表达习惯、语言逻辑等复杂特征，再通过自注意力机制在关注序列中不同位置关系的基础上，处理序列数据，最后通过解码器中的多层神经网络架构将数据特征重新转换为

表达输出。此环节虽然改动数据,但改动发生在机器内部,并不向外传播作品,因而侵权风险较小。①

在模型优化阶段,模型运用大量数据,在监督学习与无监督学习中反复调整优化参数,学习数据的潜在分布和特征,使特征逼近数据的本质结构,进而增强模型的泛化能力。具体而言,先对初始化模型设定模型结构与随机值参数,将训练数据输入初始化模型以生成预测输出,通过计算预测输出与真实标签的损失值,明确损失函数对参数的梯度,进而确定参数调整方向;以调整梯度与改进算法的方式调整参数后,再次输入训练数据重复上述步骤;最后用验证集计算模型的验证误差,当验证误差达到最小值且满足约束条件时,停止训练,输出优化后的模型。此阶段仅涉及训练数据的临时性存储以及作品参数形式的隐含存储,通常较少涉及著作权侵权问题。②

可见,生成式人工智能数据训练的著作权侵权风险主要集中在数据输入阶段,此阶段也是生成式人工智能开发的基础,是法律规制的重点所在。

(二)传统许可使用机制的不适应性

著作权是专有权,具有排他性,"许可"是著作权人利用其作品的重要方式之一,也是使作品使用行为合法化的重要前提条件。③ 根据现行著作权法,使用作品数据进行生成式人工智能数据训练,应当事先获得著作权人授权许可,但是基于生成式人工智能数据训练的特性与促进技术发展的政策考量,传统许可使用机制面临实施阻碍。

首先,生成式人工智能训练数据权利主体虚化的特性导致许可使用机制无法实施。一般而言,生成式人工智能的训练数据大多是通过网络爬虫从网络空间中抓取而来,其在数据输入阶段的预处理环节即通过数据清洗等技术措施对元数据(作者身份、授权许可等信息)实施系统性剥离,致使作品数据与版权人之间难以形成清晰、稳定的对应关系。④ 并且,由于网络空间具有匿名性与开放性的特征,作品数据的创作主体呈现出匿名化与去中心化趋势,网络空间中存在着大量的"失主作品",数据权利主体虚化问题十分严重。这使得生成式人工智能开发者难以在茫茫人海中寻找到著作权人,进而导致传统许可使用机制在实践中缺乏可行性。⑤

其次,生成式人工智能训练数据规模庞大且覆盖范围广泛的特性导致许可使用机制实行困难。生成式人工智能数据训练是一种数据密集型应用场景,其数据使用行为具有规模性与批量化的特征;而传统许可使用机制因受印刷时代作品逐件控制的理念影响,更关注逐一谈判授权的个体交易模式,与规模化的数据使用行为存在天然隔阂。⑥ 据报道,2020年ChatGPT3公布的训练数据就已达到1 TB,面对如此庞大的数据,开发者如若与著作权人进行一一协商谈判,将耗费巨大的时间和资金成本。此外,由于生成式人工智能训练数据覆盖范围广泛,其收集到的数据一般并不局限于单一的国家,而是涉及多个国家和地区,具有跨国复杂性。著作权的地域性特征使得不同国家和地区的著作权法规定存在差异。因此,面对全球化的生成式人工智能数据训练活动,传统作品许可使

① 魏远山:《生成式人工智能训练数据的准法定许可制度》,《图书馆论坛》2026年第1期,第81页。

② 参见注①,魏远山文,第81页。

③ Poorna Mysoor, *Implied Licences in Copyright Law*, Oxford University Press, 2021, pp. 17-19.

④ Shayne Longpre et al., *A Large-Scale Audit of Dataset Licensing and Attribution in AI*, Nature Machine Intelligence, Vol. 6, 2024, pp. 975-982.

⑤ 张涛:《人工智能大模型训练的著作权困境及其调适路径》,《现代法学》2025年第2期,第193页。

⑥ 参见注⑤,张涛文,第193-194页。

用机制基于跨国许可的法律复杂性更加难以实行。①

最后，促进技术进步和公平竞争的政策考量导致许可使用机制实施受阻。在当前世界各国积极营造有利于人工智能技术应用和产业发展的政策法治环境背景下，我国也须积极调整法律和政策以支持新兴技术发展。在许可使用机制下，版权人对是否授权大模型使用其作品进行数据训练具有选择权。然而，基于人工智能生成物可能产生的市场替代影响，很多著作权人往往选择保留权利，对生成式人工智能开发者不予授权，这将导致训练数据收集数量减少且范围限缩，致使大模型无法得到充分优化，创造出同质化严重或质量低劣的生成物，阻碍人工智能技术进步。当人工智能训练数据支出的许可使用费过于高昂时，可能只有大型科技公司能够负担人工智能的技术研发，中小企业受制于资金等资源的欠缺，往往在技术竞争中处于劣势地位甚至放弃技术研发，进而造成人工智能行业陷入强者更强、弱者更弱的不公平竞争环境。

此外，著作权集体管理机制虽然可以在一定程度上缓解许可使用问题，但在人工智能训练场景下适用性不足，无法真正解决大模型训练的作品使用困境。从运行逻辑看，集体管理机制的核心功能是集中授权、统一收费、分配报酬，但其制度设计天然面向单次、特定、可追踪的作品使用场景，如音乐播放、文字转载等；而生成式人工智能训练具有海量性、匿名性、跨领域、不可追溯的特点，与集体管理的制度基础存在结构性冲突。具体表现为：其一，集体管理组织难以覆盖全部权利主体与作品类型。人工智能训练需要全网海量文本、图片等，而我国各类著作权集体管理组织仅覆盖部分领域、部分权利人，大量个人作者、小众作品、海外作品并未进入集体管理范畴，无法实现全覆盖许可。其二，权利核对与授权范围无法满足人工智能训练需求。集体管理通常要求明确使用方式、使用范围、使用目的；而大模型训练具有“先抓取、后学习、不特定使用”的特征，无法在训练前逐一确定作品权利归属，也无法就数据挖掘、特征提取、内部复制等新型使用方式达成合意。其三，交易成本并未实质降低。集体管理仍需完成权利核查、授权洽谈、费用核算、使用监控等流程，面对千亿级数据规模时，其制度成本、操作难度与传统一对一授权并无本质区别，依然会产生严重的市场失灵。因此，集体管理仅能解决局部、少量、特定场景的作品使用问题，无法为大模型训练提供规模化、高效率、低成本的权利保障。

（三）既有著作权限制条款的适用不能

基于保护创作者合法权利、激励创作和实现社会文化发展之间的利益平衡，我国著作权法对著作权的调整范围和独占效力进行了适当限制，具体表现为合理使用和法定许可使用。然而现行著作权限制条款并不能很好地回应生成式人工智能数据训练引发的侵权问题。

合理使用制度具体规定在《著作权法》第 24 条，该条款采用穷尽式的权利例外列举方式，在使用行为“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”的前提条件下，明确列举了 12 种作品使用可以不经许可和不用付费的情形，并设置了兜底条款，即“法律、行政法规规定的其他情形”。从法律适用的角度来看，首先，“个人学习”“适当引用”等使用条件中，作品使用的主体限于人类且为有限使用，而大模型数据训练主体为企业，且数据抓取通常对全文进行复制，这种作品完整再现的使用方式与法律规定的使用情形形成矛盾；其次，“课堂教学”“科学研究”等使用情形要求作品的使用目的具有公益性质，而生成式人工智能数据训练具有商业营利的应用目的，二者在使用目的上存在冲突；最后，“三步检验法”受限于该条款所明确列举的 12 种使用情形，而现

① Niklas Maamar, *Urheberrechtliche Fragen Beim Einsatz Von Generativen KI-Systemen*, ZUM, 2023, p. 481.

行著作权法并没有对数据训练合理使用情形进行明文规定。因而现行著作权法并不能为生成式人工智能数据训练行为提供有效的侵权豁免。

根据《著作权法》的规定,法定许可使用制度限于第25条"教科书编写许可"、第35条"报刊转载许可"以及第42条"制作录音制品许可"三种情形。只有使用行为符合上述规定的三种情形,使用人才可以不经著作权人的许可,仅在支付合理报酬的条件下使用作品。而生成式人工智能数据训练应用场景众多,人工智能对数据的搜集、存储、分析等,包括了公共领域的自由使用、专有领域的合理使用、侵权使用等各种情形,这些多样化的数据使用场景难以进行单一的法定归类。① 现行法定许可的立法旨在为公共文化产品提供供给保障,具有突出的公共性;而由商业主体主导的生成式人工智能数据训练,其使用目的在于将大模型应用于商事交易活动,具有明显的营利性。此外,即使将生成式人工智能数据训练情形强行纳入法定许可范围,如何计算许可使用费仍是绕不开的一大问题,且许可使用费同样会面临因前述作品权利主体虚化而支付不能的问题。

三、生成式人工智能数据训练的既有著作权法规制路径检视

生成式人工智能数据训练具有巨大的侵权风险,如何平衡好技术发展与创作保护,回应生成式人工智能数据训练引发的著作权合规性问题,国内外理论与实践提出了不同规制路径。最具代表性的有两类,一是"釜底抽薪"的著作权除外方案;二是"先进后出"的著作权例外方案。

(一)"釜底抽薪"的著作权除外方案

数字技术的发展使得复制成为机器阅读的前提,但数字复制技术区别于传统复制行为。对此,马修·萨格教授认为,依赖复制的技术既无法理解或欣赏受著作权保护的作品,也不能向公众直接提供这些作品,但是它们必须复制这些作品作为算法技术的原料,因此这种使用具有"非表达性",不应被视为著作权侵权。② 此后的相关研究中,詹姆斯·格里默尔曼对人类阅读与机器阅读进行了区分,并指出传统版权立法保护的是以人类欣赏为目的进行的创作,若仅将作品作为数据,为了便于输入而进行复制,则该过程不涉及对作品的表达性使用,应当受到合理使用的豁免。③ 国内学者刘晓春虽未直接使用"非表达性使用"的表述,但提出了与其具有内在一致性的"非作品性使用"概念,并从数据训练指向对象的非特定性与训练过程的中间性,论证生成式人工智能数据训练不受著作权法规制。④ 王诗童、杨利华等学者以生成式人工智能对特定作品表达的再现效果对作品使用是否构成市场替代的效果为判定机制,将生成式人工智能划分为"非表达性使用型"机器学习和"表达性使用型"机器学习,并对此设置分层分级规制路径。⑤

在生成式人工智能数据训练的著作权法规制方案构建中,"非表达性使用"理论受到广泛关注,但其理论仍具有不完善之处。一方面,该理论可能对著作权制度根基造成解构性冲击。著作权法的立法宗旨之一是鼓励作品的创作与传播,其通过赋予创作者专有权来刺激作品创作,形成"创作—保护—再创作"的良性循环格局。⑥ 如果把人工智能数据训练归入"非作品性使用"范畴,则可

① 吴汉东:《数据信息分析合理性认定的版权规则》,《中国版权》2024年第3期,第8页。

② Matthew Sag, *Copyright and Copy-Reliant Technology*, Northwestern University Law Review, Vol. 103, No. 4, 2009, pp. 1607-1609.

③ James Grimmelmann, *Copyright for Literate Robots*, Iowa Law Review, Vol. 101, No. 2, 2016, p. 657.

④ 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》,《法学论坛》2024年第3期,第68页。

⑤ 王诗童、杨利华:《生成式人工智能机器学习的版权分层规制模式——以"表达性使用"为视角》,《编辑之友》2025年第2期,第82页。

⑥ William Patry, *How to Fix Copyright*, Oxford University Press, 2012, p. 8.

能大幅降低著作权法对数字技术应用场景的规制效力，打破良性循环格局。创作者无法从作品的数据化使用中获得收益，而人工智能开发者却能无偿获取海量作品的“非表达要素”，这会导致著作权制度的激励创作功能异化为激励技术滥用，形成创作动力因收益丧失而衰减、作品供给量减少——人工智能开发者依靠免费原料形成垄断优势——进一步挤压创作者生存空间的恶性循环趋势，触发制度刚性下的“创作萎缩—技术垄断—公共利益受损”的系统性风险。①

另一方面，该方案过于倾斜保护生成式人工智能开发者，忽视了创作者的权益保护，可能引发权利配置的结构性失衡。著作权法的立法宗旨之一是通过赋予创作者获益权进而激励创作。若将“非表达性使用”理论适用于生成式人工智能，则会导致作者失去创作收益，而生成式人工智能开发者却能在免费获取作品的基础上实现商业盈利，这有违权利义务对等原则。创作者投入时间、精力成本创作作品，理应获得作品使用带来的正当收益；使用者利用作品获得收益，则应当承担支付对价的义务。“非表达性使用”理论允许人工智能开发者无偿使用作品，本质上是一种“单向掠夺”，其理论下“创作者承担成本，开发者独占收益”的权利配置逻辑完全背离权利义务对等原则。

（二）“先进后出”的著作权例外方案

1. 美国转换性使用规则的局限

美国作为“要素主义”合理使用认定规则的代表，通过司法实践中的判例将数据训练行为归入合理使用范畴，其合理使用认定的裁判根基是“转换性使用”理论。② 1994 年坎贝尔案正式明确了“转换性使用”，改变了合理使用认定规则中四个构成要件间的均衡地位，突出了涵盖使用目的、使用功能、使用内容的转换性使用要件的重要性。2015 年谷歌数字图书馆案中，谷歌公司未经授权大量扫描受著作权保护的图书并向公众提供数字化图书的搜索及部分内容展示服务，同时旗下的搜索引擎向用户提供图书中词语使用频率等查询服务。法院认为，谷歌公司的复制目的具有高度转换性，文本的公开展示有限制性，并且展示不对原作的市场形成替代，谷歌的商业特性和整体盈利目的不能作为否定合理使用的正当理由，因而法官判决谷歌公司的复制行为构成合理使用。③ 上述数字化使用作品的判例表明，转换性使用规则已为生成式人工智能数据训练合理使用预留了解释适用的空间。近年来，关于美国人工智能训练数据侵权的待决案例，其争议焦点都集中在机器对作品数据的使用行为是否具有转换性上。

但是，在 2023 年沃霍尔案中，联邦法院判决认为，作品使用行为是否有进一步目的或不同性质是一个程度问题，需结合合理使用的其他构成要素来进行综合判断。如果原作品与二次利用有相同或高度相似的目的，且二次利用属于商业性质，在缺乏其他正当理由的情况下，第一项要素的评估不倾向于认定合理使用。④ 该判决在一定意义上将引发转换性使用要素地位的下降，导致人工智能数据训练的合法性问题须结合其他三要素综合判断，这进一步加剧了人工智能数据训练合理使用抗辩的复杂性，甚至开发者可能面临着抗辩不能的结果。

美国运用转换性使用规则来应对人工智能数据训练的合法性问题，该规则具有开放性与灵活性，但容易陷入主观化困境，导致法律适用不可预测。转换性使用的核心为使用目的要素，但这一标准缺乏客观量化指标，完全依赖个案判断，法官的自由裁量权过大，容易导致同案不同判的结果。

① 徐龙、郑冠宇：《论机器学习之著作权困境与应对》，《台大法律论丛》2023 年第 2 期，第 430 页。

② 吴汉东：《数据信息分析合理性认定的版权规则》，《中国版权》2024 年第 3 期，第 10 页。

③ Authors Guild *v*. Google, Inc., 804 F. 3d 202 (2015).

④ Andy Warhol Foundation for the Visual Arts, Inc. *v*. Goldsmith, 598 U. S. 508 (2023).

且美国转换性使用规则适用基础为“遵循判例+个案创新”,法官可通过裁判调整规则的适用范围,无需受限于僵化的法律条文;而我国是成文法国家,司法审判强调以法律为准绳,法官的自由裁量权受到严格限制,难以通过个案裁判创设或调整规则,若直接引入该规则,则可能与现行《著作权法》“封闭列举+兜底条款”的合理使用立法模式产生冲突。

2.日本柔性合理使用规则的局限

日本是采用合理使用“类型主义”立法体例的代表国家,相较于“要素主义”,该立法体例下合理使用局限于法律规定的具体情形,由法官自由裁量的空间较小,面对社会生活中的新型问题可能导致既有法律规定无法调整。但在世界范围内,日本被称为机器学习的天堂,原因就在于其类型化的立法开放取向。日本《著作权法》第30-4条为非欣赏性使用条款,即不是为了让自己或他人欣赏作品中表达的思想、情感时,可以在必要范围内以任何方式使用该作品,但不得对著作权人的合法利益产生不当损害。该条款采用“概括+列举+兜底”的立法模式,其中“用于信息分析的情形”符合生成式人工智能数据训练中的使用行为。具体而言,该条款要求信息分析行为基于非欣赏性使用目的,且使用程度限于必要限度,其规定相对宽泛。此外,为更好地促进本国人工智能技术的发展,其参照美国转换性使用规则,基于《著作权法》第47-5条规定“为了提供新的知识或信息”,允许人工智能开发者在必要限度内复制他人作品,并将复制副本向公众提供。①

相较于美国的转换性使用规则,日本的柔性合理使用规则局限于计算机分析等信息通信领域,其适用范围的领域限制与技术发展存在脱节,规则适用性不足。生成式人工智能的技术应用已突破信息通信领域,延伸至设计、工业等多个产业,其领域限制的要求难以适应人工智能发展迭代,无法长远有效应对技术发展带来的新问题,具有陷入规则僵化的适用困境。此外,虽然该规则在较大程度上减轻了数据信息分析合理使用的证明责任,但也较大幅度地倾向于保护人工智能技术方,允许将“复制副本向公众传播”的规定忽视了创作者的权益保护,直接冲击了著作权的核心权利保护。“复制副本向公众传播”的行为,本质上是对原作品的二次传播,不仅有损于创作者的复制权,更侵害了创作者的信息网络传播权。

3.欧盟文本与数据挖掘例外的局限

欧盟也是适用合理使用“类型主义”立法体例的代表之一,但与日本完全开放的立法态度不同,其采取有限开放的立法态度。欧盟《单一数字市场版权指令》确立了文本与数据挖掘例外规则。该法第3条规定了科研及文化遗产机构基于科学研究目的的文本与数据挖掘合理情形;第4条规定了以数据分析为目的的文本与数据挖掘情形,与第3条存在区别的是该条没有对适用主体和非商业性使用目的进行限制,但该条明确限定了侵权豁免的前提条件,即必须通过合法来源获取大模型训练所使用的作品数据,因而当训练数据来自商用文献数据库时,生成式人工智能开发者仍需支付相应许可费用。此外,第4条明确限定了使用行为的方式为作品复制与提取,并赋予创作者对其作品的选择退出权。

欧盟的文本与数据挖掘例外规则较之前二者在条款要求规定上更为谨慎,但其赋予创作者的声明退出权可能导致人工智能企业无法安心地使用大量数据,因为人工智能企业必须采取有效措施使得创作者能够行使退出权,而训练数据的海量性将导致该技术措施实施困难。此外,我国著作权限制条款规定的情形通常具体且明确,而生成式人工智能数据训练面向包罗万象的场景,因此若

① 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》,《图书馆建设》2025年第3期,第53页。

直接参照欧盟文本与数据挖掘例外规则进行引入，则可能导致数据使用方式超出权利限制条款预设的适用范围，进一步引发促进应用技术进步与著作权人权益的利益衡量标准模糊问题。

四、生成式人工智能数据训练著作权风险治理的路径选择

对于生成式人工智能训练数据的著作权合规性问题，各国的学者提出了多种可供参考的解决方案。本文在借鉴前人文献资料和学者观点的基础上，提出以著作权合理使用为核心，并辅之以相关配套制度的治理路径，以期在促进人工智能产业发展的同时，实现人工智能开发者与著作权人之间利益的平衡。

（一）合理使用的正当性分析

生成式人工智能数据训练引发的著作权合规性问题，实质上是新兴技术发展与内容创作保护之间的利益衡量问题。具备一定灵活性的合理使用制度，可凭借其开放的利益平衡机制对动态发展的数字技术进行适当调整；同时，大模型数据训练对作品的使用特性也在一定程度上为合理使用制度的适用提供了技术基础。①

一方面，生成式人工智能数据训练行为的特性基本符合合理使用的认定核心要素。从训练方式角度来看，人工智能大模型训练主要是通过深度神经网络的多层次表征学习，对训练数据集进行去标识化的特征萃取，其本质是对作品思想内核和表达范式的抽象建模，而非对具体表达形式的机械再现②。从训练目的角度来看，人工智能大模型训练呈现出显著的“技术性使用”特征，③旨在通过数据要素的算法熔炼，建构具有通用认知能力的智能基座，这有别于传统著作权侵权中直接攫取作品表达价值的商业化利用④。从市场效果角度来看，训练后的生成式人工智能应用场景多样，涵盖科学、教育、文化、卫生等多种类型，其开发的市场具有技术工具性，无法对内容创作市场形成替代效应。

另一方面，生成式人工智能数据训练行为的社会价值契合合理使用制度推动文化发展的价值取向。训练成功后的人工智能大模型在提高生产效率、提升生活质量、优化社会治理等方面具有重大意义，其产生的社会价值远远超出了开发者的利益范围，不仅有利于提升社会福利水平，而且有利于推动社会文化发展。而合理使用制度的适用可进一步为生成式人工智能数据训练提供“能源供给”，促进生成式人工智能泛化能力的提高，进而助推文化繁荣发展的著作权立法目的的实现。

（二）增设人工智能数据训练的合理使用条款

类别化合理使用条款的增设一定程度上可以防止法官自由裁量权过大，为司法审判提供明确的法律依据，降低案件审判难度，并且有助于为生成式人工智能开发者提供行为指引。生成式人工智能对数据的训练行为一般为数据信息分析，因而对于生成式人工智能数据训练引发的著作权合规性问题，我国可以在既有的合理使用规则中增设数据信息分析的例外条款，即“为数据信息分析的目的，在必要的限度内使用合法接触的作品”。该条款不对使用主体加以限制，营利性或非营利性等任何主体都可适用；在使用方式上，对作品的使用不限制方式类型，但限定在必要范围内且不得损害著作权人的合法权益。此外，该条款对使用对象予以限缩，只有合法接触的作品才具备讨论合理使用制度适用的必要，对于他人销售或者预售的数据库信息，生成式人工智能开发者不得未经

① 张涛：《人工智能大模型训练的著作权困境及其调适路径》，《现代法学》2025年第2期，第200页。

② Mark A. Lemley & Bryan Casey, *Fair Learning*, Texas Law Review, Vol. 99, No. 4, 2021, pp. 772-773.

③ Edward Lee, *Technological Fair Use*, Southern California Law Review, Vol. 83, 2010, pp. 797-874.

④ 张新宝、卞龙：《生成式人工智能训练语料的著作权保护》，《荆楚法学》2024年第5期，第85-86页。

许可使用;对于侵权复制品同样不得使用。需要强调的是,增设的例外条款可与既有的合理使用兜底条款保持协调,前者明确规定数据信息分析行为;后者可基于其类型开放的特点涵摄数据分析以外的人工智能作品使用情形。①

(三)明确人工智能数据训练的合理使用判定要素

即使是开放的类型化立法,亦不能穷尽合理使用的所有法外情形,也不能明确说明合理使用的判断标准。因而须结合要素主义的概括式规定,在类型列举规定的基础之上发挥补充解释兜底开放性条款和提供合理使用一般性判定标准的作用。我国合理使用的司法判断标准为"三步检验法",其在人工智能数据信息分析领域的判定要素有以下三点:

首先,对使用目的要素进行分析。生成式人工智能使用作品的表达性内容以训练算法,因而其生成物是具有内容转化性的新表达。针对生成式人工智能的表达性使用,可根据作品是否特定以及生成物内容将其划分为一般性表达使用与个性化表达使用。前者输入的作品数据不特定,并且基于统计分析的概率,调整参数和特征向量权重,进而训练得到一般通用型表达模型,通过该模型输出的内容一般不与在先作品构成实质性相似,因而符合合理使用目的要素的判定条件。后者输入的作品数据具有特定性,以达到原创作者的个性表达能力为目的,因此训练后该大模型输出的内容往往再现在先作品的表达特征,具有侵权风险。

其次,对使用方式要素进行分析。作品的使用方式涉及使用程度和被使用作品的性质两方面,合理使用制度要求其都不得影响著作权人对其作品的正常使用。在使用程度方面,生成式人工智能对数据的使用不得为公众接触。数据使用仅供机器学习,即数据使用行为只发生在生成式人工智能机器内部,不得向公众传播作品数据。在被使用作品的性质方面,生成式人工智能对数据的使用须为非欣赏性表达使用。生成式人工智能对数据的使用是通过统计概率学,提取数据结构特征,学习人类语言表达规律,即对作品构成的抽象要素进行学习,进而形成自己通用表达的模型,在此过程中其并没有欣赏作品所表达的思想、情感。总的来说,如果生成式人工智能再现与在先作品相同或类似的表达,则其对数据的使用属于欣赏或享受类的使用,不在合理使用的范围之内。②

最后,对使用后果要素进行分析。即对于作品的使用,不得对著作权人的合法权益造成不合理损害。在合理使用制度中,作品的使用后果指向客观市场影响分析,而评价这一影响最重要的标准就是著作权人的利益损害。具体而言,利益损害认定的条件为:损害须达到一定程度,即作品的使用对著作权人造成的损害超出法律容忍的范围;损害须客观真实存在,即损害是客观实在的,已经发生或是阻碍权利的行使。而生成式人工智能对数据的使用,不管是形成一般通用型表达模型还是个性化表达模型,即使是前者在合理使用容忍范围内造成的损害,使用人也有必要给予著作权人恰当的补偿。

(四)人工智能数据训练著作权风险治理的配套制度

信息数据披露制度与著作权适当补偿机制的构建,是在著作权风险治理框架内实现权利平衡、防止制度异化为侵占创作成果工具的必要配套规则。二者之间是"合理使用划定行为边界+披露与补偿实现权益矫正"的层级式、互补式法律关系。

1.配套制度的必要性分析

① 吴汉东:《数据信息分析合理性认定的版权规则》,《中国版权》2024年第3期,第14页。

② 日本《著作权法》第30条之4规定了"非享受性合理使用"的情形。

生成式人工智能训练所依赖的海量作品，本质上是由全体创作者共同贡献的数据生产要素。若仅依靠合理使用提供侵权豁免，而不附加任何披露与补偿义务，将导致权利义务严重失衡。一方面，生成式人工智能开发者可以无偿利用他人创作成果获得巨大商业利益；另一方面，创作者既无法知晓自身作品是否被使用，也无法从人工智能产业增长中获得任何收益。这种权益格局既违背著作权法激励创作的立法目的，也会削弱创作供给，最终损害数字内容生态与人工智能产业的长期健康发展。同时，生成式人工智能数据训练具有典型的黑箱属性，训练来源、使用范围、提取方式均不对外公开，导致著作权人即使遭受潜在损害也难以举证、无法救济。因此，信息披露是实现权利保障的程序前提，补偿机制是实现权益平衡的实体保障，二者共同构成人工智能数据训练得以正当化、可持续化的关键支撑。

具体而言，合理使用解决使用行为的合法性问题，即通过法律限定生成式人工智能数据训练的合法边界，豁免未经许可使用作品的侵权责任，为技术创新提供稳定预期。信息披露解决使用过程的透明性问题，其法律效果并非赋予著作权人禁止权，而是通过数据来源公示、训练范围说明等方式，确保著作权人能够知晓其作品被使用的事实，实现合理使用从“单方受益”向“透明可信”的转变。著作权适当补偿机制解决权益分配的公平性问题，即在生成式人工智能开发者大规模使用某一著作权人作品等特定情形下，对商业性人工智能开发者课以适当的补偿义务，矫正免费使用带来的权益失衡。质言之，合理使用是侵权豁免规则，披露与补偿是平衡与矫正规则。

2. 建立信息数据披露制度

生成式人工智能训练数据的披露制度，不仅可以为著作权人和社会公众获取信息搭建有效途径，而且体现了对著作权人人身利益的尊重。对于生成式人工智能数据披露制度，欧盟与美国已经先后在其相关法案中作出相应规定。欧盟《人工智能法》第53条明确了通用人工智能模型开发者负有公开通用人工智能模型训练内容详细摘要的义务。美国加利福尼亚州《生成式人工智能训练数据透明度法案》规定，人工智能大模型服务提供者应当在其网站上公布以人工智能大模型开发数据集摘要为主要内容的数据训练文档。此外，由于生成式人工智能训练数据的地域覆盖范围广泛、数据权利主体虚化等原因，信息达不到全部披露的程度。就美国加利福尼亚州《生成式人工智能训练数据透明度法案》而言，公布的摘要必须包括以下内容：(1)数据集的来源、权利人；(2)数据集是否涵盖任何受知识产权保护的数据，或数据集是否完全属于公共领域；(3)数据集是否由开发者购买或得到许可；(4)人工智能服务在其开发过程中是否使用或持续使用合成数据等。而根据欧盟《人工智能法》第53条的规定，该国学者将训练数据的数据集和数据源详细情况、数据处理情况等作为重要组成部分纳入披露的摘要内容。①

鉴于我国《生成式人工智能服务安全基本要求》只在第5章“语料安全要求”中规定了“宜公开语料中涉及知识产权部分的摘要信息”，未对训练数据信息披露给予充分关注，更没有明确信息披露的具体内容，因此我国有必要建立健全生成式人工智能训练数据披露制度以配套合理使用制度的有效实施。具体而言，应要求生成式人工智能开发者记录和公开其训练数据的内容摘要，摘要内容应至少包括：(1)数据的来源。包括开源训练数据、商业训练数据及用户输入信息等，并且对于作品数据应当指明作品名称以及作者姓名、名称。(2)数据集构成和使用范围。即生成式人工智能训

① Zuzanna Warso, Maximilian Gahntz & Paul Keller, *Towards Robust Training Data Transparency*, at https://openfuture.eu/publication/towards-robust-training-data-transparency/(Last visited on December 22, 2025).

练所用数据库包含的数据种类及使用范围。(3)数据信息分析情况。包括生成式人工智能系统对原始数据的修改和预处理情况,如数据清洗等。①

3.探索特定情形下著作权适当补偿机制

合理使用制度的完善对人工智能数据训练的发展具有重要意义。但是对于数据训练的特殊情况,如大规模使用某一著作权人的作品,适用合理使用制度就难谓公平。因此,建立特定情形下创作者适当补偿机制以保障著作权人合法权益,是为人工智能大模型制定人权友好型著作权框架的重要内容。② 具体有以下两种方案:一是由著作权人或其代表与训练者协商确定。为防止协商陷入僵局,著作权行政管理部门可介入以促成谈判。可参考专利开放许可制度的经验,在著作权行政管理部门的指导下,由相关行业协会、知识产权保护中心、作者协会、训练者等协商确定作品使用费支付标准。二是直接由著作权行政管理部门确定作品使用费支付标准。可参考标准必要专利许可费计算中的"自上而下法",由著作权行政管理部门确定一定比例的生成式人工智能收益作为作品使用费,并根据作品数量计算出每部作品使用费,再以"作品强度系数"(即考虑作品类型及同类作品中的显著高价值作品)微调。③

这一补偿机制不同于现有法定许可制度,其特殊之处在于:第一,补偿机制的制度逻辑是开发者利用作品获得收益,需按比例向创作者返还利益,本质是对人工智能产业"数据红利"的公平分配——人工智能大模型的商业价值(如营收、市场份额)高度依赖海量作品的训练支撑,创作者作为数据贡献者,理应参与收益分配。其制度设计围绕人工智能收益展开,不局限于特定使用场景,而是覆盖所有基于作品训练的商业性人工智能开发行为,核心目标是解决开发者独占收益、创作者无法获得对价的权利失衡问题。法定许可的制度逻辑是为保障特定公共利益(如教育、新闻传播),简化授权流程但保留报酬支付义务,本质是对著作权人许可权的有限限制。我国《著作权法》规定的法定许可(如报刊转载、教科书编写、录音制品翻录)均限定于特定场景,这些场景具有公共属性——旨在促进作品的广泛传播、降低公共文化产品的供给成本,而非针对商业性技术开发的利益分配。第二,补偿机制的报酬确定具有动态化、市场化、关联收益的特性。计算基数是生成式人工智能开发者的实际收益:通过自上而下法先确定补偿总额(如人工智能营收的5%—10%),再按作品数量、作品强度系数(类型、价值)分摊至单个著作权人,实现"收益越高、补偿越多"的动态匹配;其核心考量为作品贡献度。法定许可的报酬确定具有固定化、标准化、与收益脱钩的特点。计算基数是作品的使用量,报酬标准通常由国家版权局制定(如文字作品的稿酬标准、录音制品的许可费标准),与使用者的最终收益无关,亦不考虑作品对作者收益的贡献度。

五、结语

生成式人工智能数据训练引发的著作权问题,本质上是创作者权益保护与人工智能技术发展之间的利益平衡问题。如何在保障著作权人合法权益的基础上促进生成式人工智能技术的发展,是我国现行著作权法律制度面临的重要挑战。以"非表达性使用"理论为代表的方案,从根本上否定了生成式人工智能数据训练问题进入著作权法规制的空间,可能对我国著作权制度根基构成解构性冲击。而以美国、日本、欧盟为代表的合理使用制度方案亦具有局限性,并不能完全适配我国

① 张涛:《人工智能大模型训练的著作权困境及其调适路径》,《现代法学》2025年第2期,第206页。

② Christophe Geiger, *Elaborating a Human Rights-Friendly Copyright Framework for Generative AI*, International Review of Intellectual Property and Competition Law, Vol. 55, 2024, pp. 1151-1154.

③ 魏远山:《生成式人工智能训练数据的准法定许可制度》,《图书馆论坛》2026年第1期,第88页。

的法律环境，但可以为我国著作权法律制度构建提供相应参考。因此，在结合日本、欧盟“类型主义”与美国“要素主义”立法体例经验的基础上，对于生成式人工智能数据训练引发的著作权问题，我国可以在增设生成式人工智能数据训练合理使用条款的基础上，明确“三步检验法”的合理使用判定要素，并辅之以相关配套制度。随着生成式人工智能的快速发展，技术对法律制度仍会发起挑战，如何维系著作权人与技术发展之间的权益平衡，仍需归纳总结实践经验，深入开展理论研究。

Copyright Compliance Dilemma and Governance Path of Data Training in Generative Artificial Intelligence

XIE Liwei, PAN Sheng

(*School of Law, Fujian University of Technology, Fuzhou* 350118, *China*)

Abstract: Data training in generative artificial intelligence has hindered the implementation of traditional licensing mechanisms, making it impossible to apply existing rights restrictions, thus leading to a dilemma in the legitimacy of copyrights. In response to this, different regulatory approaches have been proposed in both domestic and international theories and practices. One is the drastic “fundamental negation” copyright exclusion scheme represented by the “non-expressive use” theory, and the other is the “first in, last out” copyright exception scheme applied by the United States, Japan and the European Union. Considering the usage characteristics and rights balance of generative artificial intelligence works, on the basis of the “type-based” legislative system for fair use in Japan and the European Union, fair use clauses for the data training in generative artificial intelligence should be constructed. At the same time, in combination with the “essentialist” legislative system for fair use in the United States, the fair use determination elements of the “three-step test method” should be clearly defined. In addition, supplementary systems, such as information and data disclosure, and appropriate compensation for creators, should be established to address the copyright issues caused by data training in generative artificial intelligence, and promote the coordinated development of artificial intelligence technology progress and copyright protection.

Key words: generative artificial intelligence; data training; non-expressive use; fair use

(责任编辑:董兴佩)