人工智能大模型训练数据的版权合理使用边界研究

代　巍
（佛山市银星智能制造有限公司）
DOI：10.65285/RWYSHKX.V2I5.002

摘　要：生成式人工智能大模型的迭代升级高度依赖海量文本、图像、音频等数字化数据资源，而网络公开数据大多受著作权法保护，模型训练过程中的批量复制、解析、学习行为，与传统版权保护规则形成显著冲突。当前我国著作权合理使用制度的封闭式列举模式，难以适配人工智能训练的新型使用场景，存在法律适用模糊、边界界定不清、权责划分不明等问题。本文立足我国《著作权法》规范体系，结合“三步检验法”与四维度判定标准，剖析大模型数据训练的版权使用争议核心，厘清合理使用的适用边界，区分科研非商用与商业落地、原始作品复制与算法学习的行为差异，并结合域外司法实践与国内产业发展需求，提出适配人工智能时代的版权合理使用规则完善路径，以期平衡版权权利人合法权益与人工智能产业创新发展需求[1]。
关键词：人工智能大模型；训练数据；著作权；合理使用；边界界定

Research on the Boundaries of Copyright Fair Use of Training Data for Large Artificial Intelligence Models

Wei Dai
Foshan Yinxing Intelligent Manufacturing Co., Ltd.

Abstract: The iterative upgrading of generative large artificial intelligence models relies heavily on massive digital data resources including texts, images and audios. Most publicly available online data are protected by copyright law, and the batch copying, parsing and learning behaviors involved in model training conflict significantly with traditional copyright protection rules. The current closed enumeration mode of China’s copyright fair use system is difficult to adapt to the new scenarios of artificial intelligence training, leading to ambiguous legal application, vague boundary definition and unclear division of rights and obligations. Based on the normative system of the Copyright Law of the People’s Republic of China, this paper adopts the three-step test and four-dimensional judgment criteria to analyze the core copyright disputes concerning data training of large AI models. It clarifies the applicable boundaries of fair use, distinguishes between non-commercial scientific research and commercial application scenarios, and differentiates between original work duplication and algorithm-based learning behaviors. Combining foreign judicial practices and domestic industrial development demands, this study proposes optimization paths for improving fair use rules applicable in the AI era, so as to balance the legitimate rights and interests of copyright owners and the innovative development of the artificial intelligence industry.
Keywords: large artificial intelligence models; training data; copyright; fair use; boundary definition

前言：

随着生成式人工智能技术的快速普及，大语言模型、多模态生成模型已广泛应用于生产生活各个领域，成为数字经济创新发展的核心驱动力。大模型的智能能力迭代，本质是依托海量公开数据进行深度学习、特征提取与规律归纳的过程，数据的规模、质量与多样性直接决定模型的训练效果与应用价值。相较于传统信息技术应用，人工智能大模型训练具有数据使用海量性、批量复制自动化、内容转化隐匿性、应用场景商业化的显著特征，其在未经授权情况下使用海量版权作品的行为，打破了传统版权法“授权使用、有偿流转”的核心逻辑。

版权合理使用制度作为著作权权利限制的核心规则，旨在平衡权利人专有权利与社会公共利益，为非典型、公益性的作品使用行为提供合法依据。但我国现行《著作权法》规定的合理使用情形均基于传统作品使用场景制定，无法直接适配人工智能自动化、规模化的数据训练行为。司法实践中，国内外关于 AI 训练数据版权侵权纠纷频发，各地裁判标准不一、认定尺度差异较大，既导致人工智能企业面临极高的合规风险，也可能不当损害版权权利人的合法权益。在此背景下，精准界定大模型训练数据的版权合理使用边界，构建适配人工智能技术特征的判定体系，成为化解产业发展与版权保护矛盾、规范行业发展秩序的核心议题[2]。

一、人工智能大模型训练数据的版权使用争议与制度困境

（一）大模型数据训练的版权行为属性

人工智能大模型的训练流程主要分为数据采集、预处理、模型学习、参数迭代四个环节，全过程均涉及对版权作品的数字化复制与使用。从著作权法视角来看，模型训

练首先需要将网络公开的文字、图片、音频等作品进行批量爬取、存储，形成训练数据集，该行为属于典型的复制行为，落入著作权专有权利的控制范围。与传统个人学习、合理引用的少量使用不同，大模型训练具有规模化、自动化、无差别化的特征，单次训练往往需要使用数百万乃至上亿份版权作品，且使用过程不区分作品类型、授权状态与使用场景。

同时，大模型训练的核心价值在于算法对作品特征、逻辑、语义、风格的抽象学习，而非对原始作品的直接传播与复用，训练完成后原始数据会被剥离，模型输出内容不会直接复刻原作品，具备显著的转换性使用特征。这种“输入端批量复制、输出端抽象转化”的特殊使用模式，区别于传统版权侵权行为，也与传统合理使用场景存在本质差异，成为版权规则适用争议的核心根源。

（二）现行合理使用制度的适用困境

我国《著作权法》第二十四条采用“封闭式列举 + 三步检验法”的规范结构，明确划定了十二种合理使用情形，涵盖个人学习、科研教学、新闻报道、适当引用等传统场景。但现有列举情形均无法直接覆盖人工智能大模型的数据训练行为，存在明显的制度滞后性。首先，个人学习、课堂教学等合理使用情形仅适用于自然人非商业行为，而大模型训练多由企业、机构主导，具备产业化、商业化属性，无法适用公益性合理使用规则。其次，传统合理使用强调“少量使用、适当引用”，而大模型训练的海量数据使用特征与该要求完全相悖。最后，现有规则未区分 AI 训练的技术属性，无法界定“算法学习”与“作品侵权复用”的边界，导致法律适用陷入空白 [3]。

此外，作为兜底判定标准的“三步检验法”在司法适用中存在模糊性。“特定情形使用、不冲突正常利用、不损害合法权益”的原则性规定，缺乏针对 AI 场景的细化判定标准，司法实践中法官自由裁量空间过大，容易出现同案不同判的现象。部分裁判案例将商业性 AI 训练直接认定为侵权行为，忽视其技术创新价值；也有案例过度放宽合理使用边界，弱化了版权权利人的财产权益，不利于文化创新与内容创作产业发展。

（三）产业发展与版权保护的利益冲突

从产业发展视角来看，若严格适用版权授权规则，要求 AI 企业对海量训练数据逐一获取授权，将极大增加技术研发成本，大幅延缓大模型迭代速度，制约我国人工智能产业的创新发展。海量数据的逐一授权在实操层面不具备可行性，会形成行业发展的制度壁垒。但从版权保护视角而言，无边界的合理使用会导致创作者的劳动成果被无偿商用，弱化著作权的激励作用，打击内容创作积极性，最终造成优质训练数据枯竭，反向制约 AI 产业高质量发展。二者的利益冲突，亟需通过精准的合理使用边界界定实现动态平衡。

二、大模型训练数据版权合理使用的核心判定标准

结合大模型训练的技术特征与我国著作权法核心原则，摒弃单一的形式化判定模式，以“转换性使用”为核心，结合使用目的、作品属性、使用程度、市场影响四大维度，依托“三步检验法”构建层级化、类型化的合理使用判定体系，是界定版权使用边界的核心路径。

（一）使用目的：区分公益科研与商业落地

使用目的是判定合理使用的首要标准，核心区分维度为非商业公益科研与商业化产业应用。高校、科研机构开展的非盈利性大模型技术研究、算法迭代、学术实验，不以市场盈利为目的，不面向公众提供商业化服务，符合著作权合理使用的公益属性，应当纳入合理使用范畴。该类使用行为旨在推动技术创新与学术进步，不会损害版权人的市场利益，契合合理使用制度的立法初衷。

而企业主导的商业化模型训练、产品迭代、场景落地，以盈利为核心目的，依托训练数据实现商业价值变现，具备明确的市场逐利属性，原则上不适用完全合理使用。但需避免一刀切的认定方式，对于商业模型研发过程中纯算法优化、技术测试的数据使用，未形成市场替代效应的，可适度放宽认定标准，兼顾产业创新需求 [4]。

（二）作品属性：区分事实素材与原创作品

版权保护的核心是独创性表达，而非事实、数据、通用逻辑等公共素材。在数据集筛选与使用中，需根据作品独创性高低划分合理使用边界。对于新闻资讯、公开数据、科普知识等独创性较低的事实性作品，其公共属性更强，AI 训练过程中的使用行为可适用宽松的合理使用标准。对于文学作品、艺术画作、原创音频等独创性极高的创作作品，权利人的财产权益与人身权益更值得保护，应当严格限制合理使用范围，禁止无授权规模化商用。

（三）使用程度：区分整体复制与特征提取

传统版权侵权的核心是对作品独创性表达的整体或实质性复用，而大模型训练的核心是对作品语义、逻辑、风格等抽象特征的提取与学习，并非对原始作品的直接使用。判定合理使用需区分“输入端复制”与“内容复用”的差异：训练过程中为实现算法学习进行的暂时性、技术性复制，数据仅用于模型参数迭代，训练完成后原始数据不再留存，未向公众传播、展示原始作品，属于技术性、附随性使用，具备转换性特征，可认定为合理使用。若模型训练过程中直接留存原始作品、复用核心表达，或输出内容与原作品实质性相似，则超出合理使用边界，构成版权侵权。

（四）市场影响：判断是否形成替代效应

市场影响是合理使用判定的核心兜底标准，也是“三步检验法”的核心要件。判定关键在于 AI 训练行为是否挤占原作品的市场空间、损害权利人的合法收益。若模型训练仅用于算法优化，生成内容不会替代原作品的传播、售卖与授权使用，未对权利人的现有市场与潜在市场造成负面影响，则符合合理使用要求。反之，若大模型生成内容可直接替代原创作品，导致原作品市场价值贬损、授权收益减少，则超出合理使用边界，应当认定为侵权行为。

三、人工智能大模型版权合理使用的边界界定与规则完善

（一）明确合理使用的合法边界范围

结合上述判定标准，可明确大模型训练数据合理使用的正向边界：一是非商业性学术科研场景，高校、科研机构为技术研究、学术交流开展的模型训练，可无偿使用公开版权作品；二是技术性转换使用场景，商业机构训练过程中仅进行特征提取、算法学习，无原始作品留存与复用，未形成市场替代效应；三是公共素材使用场景，对无独创性的事实数据、公开资讯、通用内容的规模化使用。

同时划定合理使用的负面边界，明确绝对禁止的侵权情形：一是直接复用原作品核心独创性表达，模型输出内容与原作品实质性相似；二是利用训练数据替代原作品市场供给，损害权利人商业利益；三是非法爬取付费、私密、未公开作品，突破数据合法获取边界；四是将训练数据用于侵权、低俗等违规场景，损害公共利益与权利人权益[5]。

（二）优化合理使用的司法认定规则

针对现有法律规范滞后问题，司法实践中应突破封闭式列举的局限，以“转换性使用”为核心细化“三步检验法”适用标准，建立类型化裁判体系。对于非商用科研训练，直接认定为合理使用；对于商用模型训练，综合考量使用目的、作品属性、转换程度、市场影响四大因素进行个案裁量，摒弃单纯以“商业属性”判定侵权的单一标准。同时确立“宽进严出”的规制思路，对模型输入端的技术性数据学习适度放宽，降低产业合规成本，对输出端的内容生成严格管控，杜绝作品侵权复用，实现技术创新与版权保护的平衡。

（三）构建分层的数据授权与使用机制

为兼顾产业发展与版权保护，应建立分层分类的版权使用机制，弥补合理使用制度的局限性。对于公共领域作品、无独创性素材，实行完全开放使用规则；对于独创性较低的公开作品，设立法定许可规则，允许 AI 企业使用并支付适度报酬，无需逐一获取授权；对于高独创性原创作品，坚持授权使用原则，鼓励行业建立版权数据共享平台，实现批量授权、集约授权，降低交易成本。同时建立利益分配机制，将商业模型的部分收益反哺内容创作者，形成“创作 – 训练 – 创新”的良性产业生态。

（四）完善行业合规监管体系

行业层面应建立 AI 训练数据合规标准，明确数据爬取、存储、使用、清理的全流程规范，要求企业建立数据溯源机制、侵权筛查机制与留存清理机制，从技术层面规避版权侵权风险。监管部门应强化事中事后监管，严厉打击恶意爬取付费作品、复用原创内容、损害版权人权益的违规行为，同时避免过度监管制约技术创新，为人工智能产业发展预留合理的制度空间。

结语：

人工智能大模型训练数据的版权合理使用边界界定，本质是数字时代下版权专有权利与社会公共创新利益的动态平衡问题。传统版权合理使用制度难以适配人工智能规模化、自动化、转换性的数据使用模式，存在明显的制度滞后性。在产业创新与版权保护的双重需求下，既不能过度收紧版权规则、制约人工智能技术迭代，也不能无限制放宽合理使用边界、损害创作者合法权益。

通过构建“目的 – 属性 – 程度 – 市场”四维度判定体系，明确公益科研与商业应用、技术学习与内容复用的边界差异，结合司法规则优化、授权机制完善、行业合规监管，能够形成适配人工智能时代的版权合理使用规则体系。未来需持续结合技术发展与司法实践，动态优化合理使用认定标准，在保护知识产权、激励内容创作的同时，赋能人工智能产业高质量创新发展，实现数字文化产业与人工智能产业的协同共赢。

参考文献：

[1] 姬蕾蕾 . 人工智能模型训练中数据利用的治理义务 [J]. 苏州大学学报 (法学版),2026,13(4):87−100.

[2] 刘文杰 . 人工智能数据训练的行为边界与规制路径 [J]. 中外法律评论 ,2026(2):142−158.

[3] 姬蕾蕾 . 人工智能模型训练中数据利用的治理义务 [J]. 苏州大学学报 (法学版),2026,13(4):87−100.

[4] 耿蕊 . 生成式人工智能训练数据的合理使用限度研究 [N]. 南方科技报 ,2026−08−01(006).

[5] 王惠敏 , 古剑 . 生成式人工智能训练数据的风险及其协同治理 [J]. 中国信息安全 ,2026(7):59−63.

作者简介：代巍（1988−）、男、汉族、湖南常德人、硕士、研究方向：从事企业数智化转型规划与建设工作。