

生成式人工智能训练数据合理使用问题研究

张陈晨

武汉工程大学法商学院(知识产权学院), 湖北 武汉

收稿日期: 2025年11月4日; 录用日期: 2025年11月25日; 发布日期: 2025年12月25日

摘要

生成式人工智能(Generative AI)作为新质生产力的重要代表,其发展高度依赖于海量数据的训练。然而,如何在保障著作权人合法权益的前提下,合法使用受版权保护的作品进行训练,已成为制约其发展的核心法律瓶颈。当前,我国《著作权法》及相关司法解释虽对合理使用制度作出框架性规定,但在面对生成式人工智能这种兼具技术创新与商业驱动的新型应用时,暴露出深刻的适用困境。一方面,合理使用制度通常适用于非营利性目的,而生成式人工智能服务提供者多为商业机构,其训练数据的使用具有明显的商业目的,因此难以直接适用合理使用制度。另一方面,合理使用制度中的“三步检验法”在生成式人工智能场景下的适用也面临较大挑战,如如何界定“影响正常使用”和“不合理损害”等问题,使得法律适用存在不确定性。此外,随着生成式人工智能技术的不断演进,其训练数据的来源和使用方式也在不断变化,传统的法律规则难以完全适应新的技术环境。因此,有必要从法律制度层面出发,系统分析生成式人工智能训练数据合理使用的法律困境,并提出相应的制度完善路径,以推动人工智能与知识产权法律体系的协调发展。

关键词

生成式人工智能, 训练数据, 合理使用

A Study on the Proper Use of Training Data in Generative Artificial Intelligence

Chenchen Zhang

Law and Business School (School of Intellectual Property), Wuhan Institute of Technology, Wuhan Hubei

Received: November 4, 2025; accepted: November 25, 2025; published: December 25, 2025

Abstract

As a significant representative of new quality productive forces, the development of Generative AI highly depends on training with massive data. However, how to legally use copyrighted works for

training while protecting the legitimate rights and interests of copyright holders has become a core legal bottleneck restricting its development. Currently, although China's Copyright Law and related judicial interpretations provide a framework for the fair use system, they reveal profound applicability challenges when confronted with new applications like Generative AI that combine technological innovation and commercial drive. On the one hand, the fair use system typically applies to non-commercial purposes, whereas Generative AI service providers are mostly commercial entities whose use of training data has evident commercial objectives, thus making it difficult to directly apply the fair use system. On the other hand, applying the "three-step test" from the fair use system in the context of Generative AI also faces significant challenges. Issues such as how to define "affecting the normal exploitation of the work" and "unreasonably prejudicing the legitimate interests of the copyright holder" create legal uncertainties. Furthermore, with the continuous evolution of Generative AI technology, the sources and usage methods of training data are constantly changing, making it difficult for traditional legal rules to fully adapt to the new technological environment. Therefore, it is necessary to systematically analyze the legal dilemmas concerning the fair use of training data for Generative AI from the perspective of the legal system and propose corresponding pathways for institutional improvement to promote the coordinated development of artificial intelligence and the intellectual property legal framework.

Keywords

Generative AI, Training Data, 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. 问题的提出

生成式人工智能(Generative AI)作为新质生产力的代表,正在深刻影响全球科技产业格局。以ChatGPT、DALL-E等为代表的大型模型,通过数据、算法和算力三大核心要素的深度融合,实现了技术的跨越式发展。因生成式人工智能对海量数据的依赖,使得训练数据成为其发展离不开的基础资源。然而,如何在保障著作权人合法权益的前提下,合法获取和使用训练数据,尤其是涉及大量受版权保护作品的使用是否构成侵权、是否可纳入合理使用范畴,已成为当前制约生成式人工智能产业发展的核心法律瓶颈。

当前,全球范围内对生成式人工智能训练数据著作权问题的政策应对呈现出显著差异。日本在其《著作权法》中明确将包含人工智能训练在内的“计算机信息处理”行为纳入合理使用范围,为产业发展扫清了法律障碍;欧盟在《版权指令》中为文本与数据挖掘(TDM)设立了例外条款,但其适用范围较窄、条件较为严格,整体仍倾向于保护著作权人;美国则主要依赖司法判例,在“合理使用四要素”框架下进行个案裁量,目前尚未形成统一的规则体系。

聚焦中国,发展人工智能已被提升至国家战略高度。我国于2017年颁布《新一代人工智能发展规划》,将人工智能发展纳入国家公共政策体系,体现了国家对人工智能产业的高度重视。然而,具体法律规制仍存在明显空白与矛盾。2020年新修订的《著作权法》未将文本与数据挖掘(TDM)行为纳入著作权合理使用范畴,导致学界与实务界对于生成式人工智能未经授权使用受著作权保护作品的行为是否构成侵权、侵犯何种权利、能否适用合理使用制度免责等问题存在诸多争议^[1]。2023年8月15日开始实施的《生成式人工智能服务管理暂行办法》虽要求服务提供者“使用具有合法来源的数据”“涉及知识

产权的，不得侵害他人依法享有的知识产权”，但并未为训练数据中对著作权作品的使用预留明确的合理使用空间，实质上将开发企业置于巨大的侵权风险之中[2]。这种政策模糊性与法律滞后性，导致中国生成式人工智能产业面临严峻挑战：一方面，企业为规避法律风险，被迫依赖来源单一或过时的公共领域数据进行训练，致使模型输出内容难以贴合现实需求，严重制约技术性能与产业竞争力。另一方面，学术界与实务界对数据训练行为合法性莫衷一是，主要存在“适用合理使用说”“法定许可/补偿金说”“权利限制说”“有阅读权即有挖掘权说”及“绝对侵权否定说”[3]等多元甚至对立的观点，缺乏权威共识与司法指引。政策模糊性导致企业难以合法获取训练数据，进而被迫依赖低质量或过时的数据源，影响模型性能与创新潜力，最终倒逼法律制度亟需解决“合理使用”制度在生成式人工智能场景下的适用性问题。由此可见，解决训练数据版权问题的核心症结，很大程度上在于“合理使用”制度能否有效适用于生成式人工智能的特定场景。

然而，正如上文所述，现行《著作权法》及相关司法解释虽对合理使用有框架性规定，却在面对生成式人工智能这种兼具技术创新与商业驱动的新型应用时，暴露出了深刻的适用困境。一方面，合理使用制度通常适用于非营利性目的，而生成式人工智能服务提供者多为商业机构，其训练数据的使用具有明显的商业目的，因此难以直接适用合理使用制度。另一方面，合理使用制度中的“三步检验法”在生成式人工智能场景下的适用也面临较大挑战，法律适用存在不确定性。此外，随着生成式人工智能技术的不断演进，其训练数据的来源和使用方式也在不断变化，传统的法律规则难以完全适应新的技术环境。因此，有必要从法律制度层面出发，系统分析生成式人工智能训练数据合理使用的法律困境，并提出相应的制度完善路径，以推动人工智能与知识产权法律体系的协调发展。

2. 生成式人工智能训练数据合理使用的法律困境

（一）合理使用制度适用的主体限制

在人工智能技术迅猛发展的背景下，传统著作权合理使用制度所显现的主体身份差异以及对适用范围限制，正逐渐暴露出其在应对生成式人工智能训练数据使用方面的结构性矛盾。具体而言，欧盟《单一数字市场版权指令》在文本与数据挖掘例外条款中，明确将适用范围限定为非商业研究机构，其要求使用主体必须具备“研究或教育目的”且仅限于“少量复制”[4]。我国《著作权法》第24条第6项亦规定，合理使用仅适用于“为学校课堂教学或科学研究”等非营利性活动，并强调“不得影响原作品的正常使用”和“不得损害著作权人的合法权益”，但并未明确允许商业机构以“教学科研”名义进行大规模数据训练。当前合理使用制度在主体资格上的严格限制，与生成式AI发展的现实需求产生了直接矛盾。无论是欧盟《版权指令》将文本与数据挖掘例外主要限定于研究机构，还是我国《著作权法》强调非营利目的，其核心思路都是将商业行为视为可能损害权利人利益的活动。然而，今天生成式AI的研发主力与核心应用恰恰集中在商业公司。

制度适用主体身份的限制与人工智能技术的本质特征之间存在根本性冲突。一方面，商业公司作为人工智能技术的主要推动者，其训练行为往往涉及全面、系统、非定向的数据采集，这与传统合理使用制度中“少量使用”“特殊情形”等要件相悖；另一方面，由于商业主体身份被预设为“可能损害著作权人利益”，其在“三步检验法”框架下难以获得司法认可。尤其在生成内容与原作存在实质性相似的情况下，司法实践中往往采取举证责任倒置的规则，即要求服务提供者自证其训练数据来源合法，否则将面临直接侵权指控。此外，我国司法机关在处理AI训练数据侵权案件时，普遍对“转换性使用”理论持谨慎态度，甚至在某些判例中完全不予采纳，从而进一步压缩了商业主体援引合理使用的空间[5]。

因此，当前著作权合理使用制度在面对生成式人工智能时，不仅未能有效回应其技术特性与商业模式，反而因制度设计的滞后性与适用范围的局限性，导致人工智能企业陷入“要么面临侵权风险，要么

遭受创新受阻”的双重困境。这一困境不仅反映了法律制度与新兴技术之间的不匹配，也凸显了在当前法律框架下，合理使用制度难以成为生成式人工智能产业发展的制度保障。

(二) “三步检验法”的适用障碍

在生成式人工智能训练数据的合理使用问题中，“三步检验法”作为我国《著作权法》中判断合理使用的法律标准，其适用性受到多重挑战。“三步检验法”要求合理使用须满足：(1) 限于“特殊情形”；(2) 不与作品“正常利用”相抵触；(3) 不得不合理损害权利人合法权益。然而，在 GenAI 训练数据的使用中，由于数据来源广泛、使用方式多样，这一标准在实践中变得难以操作。

其一，“特殊情形”要件难以满足，生成式 AI 的训练行为本质上是高度自动化、非定向的，其使用作品的方式和目的并非基于特定的“特殊情况”，而是为了实现模型的深度学习与泛化能力。这种“常态性”使用方式与“三步检验法”中对“特殊情况”的要求相悖，使得该行为在司法实践中难以被认定为合理使用[6]。

其二，“不得影响作品的正常使用”这一要件在人工智能场景下难以操作。根据《伯尔尼公约》及 WTO 解释，“正常使用”被界定为“所有具备或可能获取经济利益的使用行为”，而生成式 AI 的训练行为因其海量性、潜在商业性，必然被纳入“正常使用”的范畴。因此，权利人可主张任何未经许可的训练使用均可能剥夺其未来数据许可市场的可期待利益，从而形成与潜在著作权市场的竞争关系[7]。这种对“正常使用”的宽泛理解，使得第二项要件在人工智能训练场景中缺乏明确的操作边界。这种逻辑如果被绝对化，合理使用制度在 AI 领域将毫无用处，因为它等于承认权利人对所有未来的、未知的使用方式都拥有绝对控制权。

其三，“不得不合理地损害著作权人的合法权益”这一要件在 AI 训练数据使用中也面临适用难题。问题的关键在于，是否必须通过“一对一”的事前授权来避免损害？市场化授权机制在此场景下存在结构性失灵。一方面，必须承认，未经许可的训练确实会对创作者造成影响。在经济层面，它可能侵蚀了本可通过集体管理或法定许可等方式实现的潜在收益。在非经济层面，损害更为深刻：创作者会感到其心血被技术无情“吞噬”，其个人风格沦为模型参数，导致身份模糊、创作独特性被削弱；另一方面，若要求企业为每一项训练数据获取授权，将极大增加交易成本，甚至阻碍技术发展。简单地要求“先授权后使用”看似保护了权利，实则可能因为无法操作而让所有使用都变成“侵权”，最终既无法让创作者获得实际补偿，又阻碍了技术创新。因此，如何在保护著作权人利益与促进技术创新之间取得平衡，成为当前法律适用中的核心争议点。

值得注意的是，司法实践中对“实质性相似”问题的处理也加剧了合理使用制度的适用困境。例如，北京互联网法院 2023 年“AI 文生图”著作权案¹中，法院认定 AI 生成内容具有独创性，但并未深入探讨训练数据的合法性问题，仅以“实质性相似”为由初步认定侵权。这种倾向表明，在当前司法实践中，法院更倾向于从结果出发判断侵权责任，而忽视了训练过程是否构成合理使用这一关键问题。

此外，我国司法实践中对“转换性使用”理论的适用也存在明显分歧。部分法院在审理 AI 生成内容案件时，倾向于采用“四要素法”，即美国判例中的合理使用标准，而忽视了《著作权法》中“三步检验法”的适用[8]。然而，值得注意的是，作为该理论源头的美国司法界，对“转换性使用”的界定也正处于调整和限缩之中。例如，美国最高法院在 2023 年的“沃霍尔诉戈德史密斯案”中[9]，强调转换性需考察使用的具体目的而非仅风格改变，这为 AI 训练数据“转换性”的认定带来了新的不确定性。这种司法实践的不统一，使得商业机构在面对 AI 训练数据合法性问题时，缺乏明确的法律指引和裁判预期，进一步加剧了法律适用的不确定性。

¹(2023)京 0491 民初 11279 号。

因此,“三步检验法”在生成式人工智能训练数据合理使用问题中的适用障碍,不仅体现在法律条文的模糊性上,更体现在司法实践中的不一致性与适用困境。因此,有必要从法律制度层面出发,系统分析生成式人工智能训练数据合理使用的法律困境,并提出相应的制度完善路径。

(三) 法律与技术发展的不平衡

生成式人工智能,特别是大型语言模型(LLM)的崛起,必然会运用海量、高质量的训练数据。然而,这种技术与现有法律体系之间,在数据层面有显著的不平衡,是当前一大突出的法律困境。

生成式 AI,尤其是大模型的预训练阶段,本质上是利用统计方法从海量数据中学习模式、关联和概率分布,以生成新的内容。其技术效能与模型规模、数据量及质量呈高度正相关。生成式 AI 在训练阶段需要无差别、大规模、高效率地获取和利用一切可用数据,其追求的是数据的广度、深度和利用效率。生成式 AI 的训练过程倾向于将数据视为“同质燃料”,不区分其来源、类型或权利状态。这种处理方式在技术上是高效的,但与法律对数据权利的精细化管理存在根本性冲突。

然而,现行法律体系强调对数据权利的精细化管理,要求对数据进行分类,并为每一类数据设定明确的权利主体和权能。法律将数据世界切割为不同的类别,如个人信息/非个人信息、公开数据/非公开数据、版权作品/公有领域作品,并为每一类别设定了清晰的权利主体,如数据主体、版权人和具体的权能,如知情同意权、删除权、复制权等。法律要求数据处理的每一步都需符合特定的合法性基础、履行告知义务、尊重个体授权、清晰界定使用目的和范围。这种根本性的范式冲突,即技术的“效率导向”与法律的“权利精细保护导向”之间的冲突,导致了一系列具体法律规则在生成式 AI 训练数据场景下的实际失效或执行困境。

3. 生成式人工智能训练数据合理使用的制度完善路径

(一) 明确合理使用制度的适用主体范围

为破除生成式人工智能发展中的制度适用困境,重构合理使用制度的适用主体框架已刻不容缓。第一,应突破“非营利性主体”的传统限制,将企业等营利性组织明确纳入合理使用制度的适格主体范围。当前产学研融合日益深化,营利性组织已成为技术创新的重要力量,继续以主体性质作为适用标准已不合时宜。核心在于将审查重点从“谁在使用”转向“为何使用及如何使用”。为功能性、非表达性的数据训练行为提供法律空间。第二,构建“商业目的”与“非商业目的”双轨授权机制。在允许商业主体实施数据挖掘的同时,应设定透明度要求、权利保留等限制条件,确保其行为在法律可控范围内。该机制有助于平衡著作权人与开发者之间的利益,为 AI 技术发展提供制度支持。第三,借鉴欧盟《单一数字市场版权指令》经验,通过增设专门条款,明确承认人工智能开发者的主体地位。此举既符合产业发展现实需求,也有助于构建更加开放和灵活的著作权生态体系。

(二) 完善“三步检验法”的实施细则

由于生成式人工智能的训练行为具有高度的自动化、非定向性和海量性,传统“三步检验法”中的“特殊情形”“不得影响作品的正常使用”“不得损害著作权人的合法权益”等要件在实践中难以操作。因此,有必要对“三步检验法”进行细化和重构,以适应生成式人工智能技术发展的现实需求。

第一,明确“特殊情形”的认定标准。生成式人工智能对训练数据的常态性使用,与现行“三步检验法”中“特殊情形”的要求存在冲突,导致实践中难以被认定为合理使用。为此,建议在《著作权法实施条例》中增设专门条款,将“生成式人工智能数据训练”明确列为一种“特殊情形”,并引入“转换性使用”理论,为解决此问题,不应简单照搬美国合理使用的“四要素法”而应致力于将“转换性使用”的核心精神——即对作品的使用旨在实现与原作不同的、新的目的或功能——有机地融入“三步检验法”的既有框架之中。以“目的转换性”重塑“特殊情形”的认定标准。其核心是将“为非表达性目的进行的生

成式人工智能模型训练”明定为一种“特殊情形”。所谓“非表达性目的”，旨在强调其使用功能在于机器学习与性能优化，而非向公众传播作品表达。这一本质区别，从法理上将其与传统作品使用行为区分开来，为其通过“第一步”检验提供了正当性基础。

第二，细化“不得影响作品正常使用”的判断标准。在生成式人工智能训练中，由于数据使用规模庞大且可能涉及商业目的，这类行为常被视为影响了作品的“正常使用”。为解决这一问题，建议引入“比例原则”和“市场替代效应”等标准，对“正常使用”进行量化评估。例如，若 AI 训练行为未实质性替代原作品的市场价值，即可认定为不影响其正常使用。此外，还可借助大数据分析和市场替代效应模型，科学评估 AI 训练行为的潜在影响，从而提升司法裁判的公正性与科学性。

第三，建立“不得不合理地损害著作权人的合法权益”的评估机制。在判断是否构成“不合理损害”时，必须认识到，针对 AI 训练采取完全市场化的“一对一”事前授权机制，并非更优或可行的解决方案。从可行性看，生成式 AI 训练涉及海量作品，其搜寻、谈判与清算成本高到无法承受，这种市场本身因交易成本过高而无法有效形成。强行要求授权，只会导致法律在普遍违法中失去效力，或彻底扼杀产业创新。为解决这一问题，建议通过建立“预期利益损失”评估机制，结合大数据分析和市场替代效应模型，对 AI 训练数据使用行为的潜在影响进行量化评估。例如，可以将“预期利益损失”作为判断是否构成“不合理损害”的标准，即如果 AI 训练数据的使用行为不会对著作权人的预期利益造成实质性损害，则可以认定为不构成不合理损害[10]。

第四，推动“三步检验法”与合理使用制度的动态调整。“三步检验法”作为合理使用制度的一般条款，其适用性受到国际公约的约束，但其在具体案件中的适用仍需结合实际情况进行动态调整。为适应生成式人工智能技术的快速发展，建议在《著作权法实施条例》中增设“生成式人工智能数据训练”合理使用条款，并通过司法解释或行政法规进一步细化适用条件，确保合理使用的认定标准具有可操作性。

(三) 构建训练数据“安全港”与合理使用例外制度

在当前生成式人工智能的训练中，常需使用包含版权内容、个人信息及公开数据等在内的数据集。为促进 AI 技术发展、产业创新和业务优化，建立数据合理使用制度十分必要。尽管欧盟《通用数据保护条例》(GDPR)与我国《个人信息保护法》在立法时未专门针对机器学习训练场景作出规定，但欧美已逐步将现有数据法规延伸至 AI 领域，并特别强调数据来源合规与处理透明。然而，过于严格的信息保护规则，在一定程度上会影响通用大模型的研发与应用。因此，在加强训练数据来源和处理合规性的同时，也应合理设定适用于生成式 AI 训练数据的保护标准，适度调整监管力度，以在激励创新与保护权益之间取得平衡。目前生产式人工智能针对海量数据资料的训练性使用存在法律规定上的障碍，建议未来制定的《人工智能法》区分研发训练和商用提供阶段，借鉴早期互联网搜索引擎建设时应用的“安全港”制度，建立训练数据“安全港”制度，即使用人不知道数据来源是否合法，但是可以利用数据进行研发或应用，日后有人来提出相应权利主张，要按照法律的规定付费或补偿[11]。进一步完善我国个人信息保护法律，可以借鉴欧盟 GDPR、英国 ICO 的合法利益评估标准和新加坡 PDPA 项下的《关于在 AI 推荐和决策系统中使用个人数据的咨询指南》创设的科研例外和业务改进例外制度，即对于为了实现公共利益、科学或历史研究或统计目的处理，成员国的法律可以对访问、更正、限制处理和反对等部分个人信息权利进行克减，并可以成为拒绝删除的抗辩理由[12]。建议在《中华人民共和国著作权法实施条例》中增加“文本或数据分析、训练、挖掘明确规定为著作权法的权利限制或例外的法定情形”的条款，为人工智能模型训练获取高质量数据集扫清法律障碍[13]。

4. 结语

生成式人工智能的迅猛发展，不仅推动了技术变革，也对传统法律体系提出了深刻挑战。本文围绕

生成式人工智能训练数据的合理使用问题，系统分析了现行著作权制度在应对这一新兴技术时所面临的困境，重点探讨了合理使用制度在主体适用、目的限制及“三步检验法”适用等方面的局限性。研究表明，当前法律框架难以有效回应生成式人工智能在数据获取与使用方面的特殊需求，导致企业在技术创新与法律合规之间陷入两难境地。文章指出，合理使用制度的适用范围与商业目的的冲突、法律适用标准的模糊性以及技术演进带来的不确定性，构成了当前法律规制的主要障碍。因此，有必要从立法、司法与技术协同推进的角度，构建适应人工智能发展的新型数据使用法律体系。未来的研究可进一步探索“合理使用”在生成式 AI 场景下的具体适用路径，推动相关法律制度的完善与创新，为人工智能产业的健康发展提供坚实的法律保障。

参考文献

- [1] 王雪蕾. 人工智能数据挖掘适用著作权合理使用制度的审思[J]. 河北法学, 2025, 43(3): 185-200.
- [2] 张伟君. 论大模型训练中使用数据的著作权规制路径[J]. 东方法学, 2025(2): 79-92.
- [3] 施小雪. 重塑复制权: 生成式人工智能数据训练的合法化路径[J]. 东方法学, 2024(6): 70-83.
- [4] 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.
- [5] 知产财经(吴子芳). 生成式人工智能发展中值得关注的著作权问题[EB/OL]. <https://mp.weixin.qq.com/s/2KEwCqmOTKG2WZwdQj1dww>, 2024-07-12.
- [6] 刘祖兵. 生成式人工智能使用在先作品数据的适法路径、梗阻与制度完善[J]. 西华大学学报(哲学社会科学版), 2025, 44(2): 17-29.
- [7] 熊琦. 著作权合理使用司法认定标准释疑[J]. 法学, 2018(1): 182-192.
- [8] 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.
- [9] 金海军. 演绎作品创作的专有权与合理使用抗辩[J]. 中国版权, 2022(6): 52-62.
- [10] 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.
- [11] 张平. 生成式人工智能实现突破创新需要良法善治——以数据训练合法性为例[J]. 新经济导刊, 2023(8): 26-28.
- [12] 傅宏宇. 生成式人工智能的治理模式与风险辨析[J]. 数字法治, 2023(4): 191-206.
- [13] 丁道勤. 生成式人工智能训练阶段的数据法律问题及其立法建议[J]. 行政法学研究, 2024(6): 16-28.