生成式人工智能数据训练的合规风险及解决路径

崔继一
北京化工大学文法学院，中国 ·北京 102200

摘要：在生成式人工智能快速发展的背景下，数据训练作为其核心环节，伴随数据产业扩张面临诸多合规风险，现行法律规范的滞后性加剧了合规困境。本文通过梳理数据训练的合规风险以及域外经验，结合我国司法与产业实际，从完善法律规范、明确信息透明义务、健全监管审查三个维度，提出针对性解决路径，旨在平衡技术创新与权利保护，为生成式人工智能数据训练合规发展提供支撑。

关键词：生成式人工智能；数据训练；合规风险

Compliance risks and solutions for generative artificial intelligence data training

Cui Jiyi
School of Humanities and Law, Beijing University of Chemical Technology, China Beijing 102200

Abstract: In the context of rapid generative artificial intelligence, data training is its core link. With the expansion of the data industry, it faces many compliance risks. The lag of current legal norms has intensified compliance. By sorting out the compliance risks and extraterritoriality of data training, combined with my country's judicial and industrial reality, this article proposes targeted solutions from the three dimensions of improving legal norms, clarifying information transparency obligations, and improving regulatory review, aiming to balance technological innovation and rights protection, to provide support for generative artificial intelligence data training compliance.

Keywords: Generative artificial intelligence; Data training; Compliance risk

0 引言

在数字化高度发展的今天，人工智能领域蓬勃兴起，浪潮席卷全球，以 ChatGPT 为典型代表的生成式人工智能飞速发展，给全球各个行业领域带来了深刻变革。生成式人工智能“以大数据、人工神经网络算法、算力为基础，以模拟人的思考方式与表达方式为核心，为社会公众提供新的表达输出模式”[1]。因此，海量的数据是其生成表达内容的基础，通过对数据的抓取、学习来辅助内容及决策的生成，数据的数量、质量也成为了生成表达内容的重要影响因素，数据训练的必要性也日益凸显。随着关注与需求的指数型增加，与之相关的数据要素产业市场快速扩张。“上海数据交易所预测，至 2030 年全球数据交易市场规模将达 3708 亿美元，我国数据产业规模有望增至 7.5 万亿元，形成支撑人工智能产业发展的强大数据基础[2]。”新兴产业的发展带来了新的挑战，旧有的法律难以完整囊括新兴产业行为，生成式人工智能在重塑人们生产、生活方式的同时，也在重塑着法律的边界。

“数据训练是指利用大规模数据集对人工智能模型进行训练，旨在使其能够预测数据趋势或自主做出决策[3]。”生成式人工智能通过其内部算法抓手，对网络上的海量信息进行搜集抓取，进行高质量的数据训练。信息的来源包含受著作权保护的作品，受隐私权保护的私密信息，以及难以辨别真假、来源的各类信息。虽然输出内容并非简单的机械输出，而是通过整合优化进行的转化性输出，但其数据训练仍与其他法律主体受保护性权益产生碰撞，其合规性有着极大的法律风险。现行法律对此尚未有明确条款规定，2023 年颁布的《生成式人工智能服务管理暂行办法》等规范性文件规定较为笼统，并未对其进行明确的法律界限。“训练数据是机器学习过程中的核心，直接决定模型的能力上限与实际效能”[4]，并且“高质量、大规模的训练数据往往能够造就更强大、更可靠的人工智能系统”[5]，探讨生成式人工智能数据训练的合规风险及解决路径更显得尤为重要。

1 生成式人工智能数据训练的合规风险

1.1 知识产权侵权风险

生成式人工智能在获取海量的网络数据信息之后，通过其内部的数据逻辑计算方式进行分类整合，按用户需求提供不同的场景产品，不可避免的会涉及对受知识产权保

护作品的利用，在进行数据训练的过程中面临着知识产权的侵权风险。“数据来源的合法性是判定训练数据著作权的重要基础[6]。”生成式人工智能的训练数据大多未获得著作权人授权同意，与《著作权法》中“先授权，后使用”的原则相冲突。虽然《著作权法》第二十四条列举了繁多的可以不经著作权人同意的合理使用情形，但是并未明确包含生成式人工智能数据训练情形。其中第十三款属于兜底性条款，符合法律规定的其他情形都可以通过此条款来进行合法解释，但使用生成式人工智能进行数据训练的行为是否属于著作“合理使用”的例外情形无法确定，在此过程中对知识产权进行侵权的风险极高。

1.2 个人信息与隐私保护风险

在处于大数据时代的今天，个人信息与个人的安危息息相关，个人信息、隐私的泄露不仅极大地损害他人的合法权益，而且会引发公众对生成式人工智能的恐慌。生成式人工智能在数据训练中利用爬虫技术，获取大量公开和用户形式知情同意的个人信息，包括身份内容，个人偏好等非脱敏的信息内容，在输出时极有可能泄露他人隐私信息，有极大的不确定性和不可控性。我国《个人信息保护法》要求处理个人信息须遵循合法、正当、必要原则，取得个人同意，同时要履行告知义务，但是生成式人工智能在数据训练时如果严格遵守同意与告知义务，则会造成个人信息相关干预的工作量激增，难以完成。并且，处于经济考量，生成式人工智能企业难以承担如此高昂成本。《生成式人工智能服务管理暂行办法》在其第七条中也提到了涉及个人信息的相关规定，但较为笼统，并未有细致举措。个人信息的保护处理要落实最小必要与脱敏的原则要求，但是生成式人工智能的数据训练需要海量的信息数据来帮助训练，数据的训练会在爬虫技术的帮助下尽可能的搜集更多信息来完善功能，这是一个相悖的矛盾点，在这个过程中，会产生极大地个人信息与隐私保护风险。

1.3 数据安全与网络安全风险

生成式人工智能的数据训练依赖于其通过爬虫技术来获取数据，海量的数据真伪不明，来源不明，开源数据未经核查，极有可能会违反《数据安全法》中对重要数据的保护要求。这些数据中可能包含国家秘密、商业秘密等敏感信息，通过提示词诱导等方式，易造成训练数据泄露。一旦敏感信息泄露或被非法利用，可能会对国家安全、公共利益以及企业利益造成严重损害。同时，生成式人工智能的数据训练过程也面临着网络攻击的风险，黑客可能通过攻击训练系统，篡改或窃取训练数据，进而影响人工智能模型的准确性和可靠性，甚至可能导致模型被恶意利用，对用户和社会造成危害。此外，生成式人工智能在数据训练过程中还可能产生数据偏见和歧视问题，如果训练数据中存在偏见或歧视性信息，那么生成的人工智能模型也可能继承这些偏见，从而在决策过程中产生不公平的歧视性结果。

2 生成式人工智能数据训练的域外借鉴

2.1 美国：转换性使用规则

“转换性使用概念由皮埃尔·勒瓦尔法官提出[7]。”转换性使用规则是美国在司法实践基础上由合理使用原则发展形成的，来源美国《版权法》第 107 条，共包含四项要素：“一是使用行为的目的与特征，包括是否具有商业性质或非营利教育目的；二是作品本身的属性，即作品类型及其创造性程度；三是所使用部分相对于作品整体的数量与实质性，强调其在原作品中的重要性；四是该使用行为对原作品潜在市场或价值的影响，即是否损害权利人的经济利益[8]。”以上四项要素，也被称为四要素检验法，四项要素并非孤立存在，而要结合整体把握，以此来判定数据和作品之间的关联关系，并据此决断是否符合转换性使用。

转换性使用实质上是指相对于原作品，人工智能生成内容已经成为了一个有显著区别的新作品，与原作品之间虽然有着相似与共同之处，但其是在原作品基础上添加了新的元素或有了新的意图，两者在价值上无趋同之处。2025 年，美国加州北区法院在 Bartzv.AnthropicPBC[9] 案中作出标志性判决，进一步明晰了生成式人工智能数据训练中数据使用的合法边界，法院考量了其数据的来源及生成成果对原作的影响，最后认定被告使用其下载的电子书籍训练大语言模型属于合理使用。由此，也可以看出转换性使用规则尚无具体明确规定，需要法官发挥自由裁量权，结合个案具体分析。

2.2 日本：信息处理例外规则

针对生成式人工智能，日本在法律领域不断跟随修订，呈现出了更为开放的态度，规定也更为明确细致，起到良好的指引作用。2018 年修订《著作权法》，增设第 47 条第 5 款，明确在计算机信息处理中，为产出新信息而在必要限度内附随使用作品，并且不构成不当损害著作权人利益的，可认定为合理使用。需要明确的是，其中但书条款排除了明知侵权仍提供、不当损害权益的情形，指出“若存在证据表明训练行为在类型、目的或使用方式上侵害了著作权人的利益，则该行为将被排除在所取消的限制之外，并可能进一步被判定为侵犯著作权”[10]，强调使用必

要性与利益平衡。

2.3 欧盟：临时复制例外规则

欧盟针对生成式人工智能数据训练的版权规制，以《欧盟版权指令》中的临时复制例外规则为核心，同时搭配《人工智能法案》的合规要求，形成了版权例外与 AI 全生命周期监管并行的双重规制体系，强调版权保护、数据合法获取与算法安全的多重平衡，区别于美国司法裁量模式与日本法定例外模式，更注重成文法层面的清晰界定与强监管落地。欧盟的临时复制规则，核心法律依据为《欧盟版权指令》第 5 条第 1 款的内容，该条款是欧盟层面统一各成员国版权制度的关键规定，其核心内容为：对作品的临时复制行为，若具备暂时性或附随性，属于技术过程中不可或缺的组成部分，且唯一目的在于实现合法使用，同时不具有独立的经济价值，不损害著作权人的合法权益，则可以不经过权利人许可，不构成版权侵权。

3 生成式人工智能数据训练的解决路径

3.1 完善法律规范

法律规范是生成式人工智能数据训练合规的引路石和护城河，发挥着底线性的作用。在当前人工智能迅速发展的时代，法律规范要迅速反应，紧追时代发展，勉力弥补法律滞后性的不足。首先，要加快相关立法进程，及时制定暂行条例，修改补充《著作权法》第二十四条的列举内容，“增加‘数据训练目的’‘机器学习’等类似表述”[11]，将法律规则细化，增设不同环节的侵权认定标准，区分商业性与非商业性使用、授权数据与非授权数据的不同法律后果，消除司法实践与企业合规的模糊地带。其次，学习美国等西方国家在针对以生成式人工智能为代表的新型科技方面的法律变通，探索合理使用与许可制度的结合，借鉴避风港规则，形成具有中国特色的转换性使用规则，通过多要素检验，综合分析把握是否形成一个有显著区别的新作品，并不影响原作品著作权人的权利。规定明确司法规则，规范法官自由裁量权，统一裁判尺度，避免裁量权过大导致的同案不同判事件发生，“并逐步总结典型案例与可复制做法，以实证经验反哺制度完善”[12]。最后，针对公民个人隐私问题，“《个人信息保护法》所规定的‘匿名化’标准在实践中缺乏具体的认定细则”[13]，会导致相关企业在进行数据处理时难以判定是否符合法定标准。故此，要结合《个人信息保护法》，细化生成式人工智能训练场景下个人信息收集、脱敏、存储、跨境传输的具体规则，明确敏感个人信息的禁止使用情形与豁免条件。

3.2 明确信息透明义务

生成式人工智能数据训练在进行数据抓取时，由于各企业的算法黑箱，难以了解其运用爬虫技术进行数据信息抓取的具体方式、范围、方法等各项内容，对其合规性难以考量，处于公众的视野盲区。对此，一方面，可以设立分层分类的信息披露机制，针对不同类型的数据主体与监管主体，设定差异化的披露内容与披露方式。对于著作权人、个人信息主体相关运营企业应当以清晰易懂的方式，告知其数据被用于人工智能数据训练，保障其知情权，也便于拒绝被使用训练的主体及时提出异议，降低侵权风险；对于监管部门，要完整披露其训练数据来源，数据算法，获取方式等核心信息，及时备案形成相关工作日志。另一方面，强化训练数据公开透明要求，适度公开训练数据的来源构成、授权比例等，主动接受社会监督，在保障公众知情权与企业商业利益之间寻求平衡。

3.3 健全监管审查

针对当前生成式人工数据训练领域野蛮生长现状，监管部门应当加强行业监管，发挥主导作用。第一，加强培训。生成式人工智能属于数字领域新兴内容，通过培训提高监管部门人员技术水平，强化监管能力。第二，可以“探索建立由国家网信主管部门或知识产权主管部门负责的专门监管平台”[14]，便利统一尺度监管，拓展监管平台，监管平台可以面向社会公众开放渠道，进行交互，既便利查询了解企业数据训练规则情形，又便利及时群众监管投诉。第三，充分发挥行业协会作用，发挥其行业熟悉优势，及时监管技术漏洞，建立行业内的投诉举报机制。多管齐下，健全生成式人工智能数据训练的监管审查。

4 结语

生成式人工智能的健康发展，离不开合规数据训练的坚实支撑。数据训练的合规风险既有时代的特性，又有法律的滞后性。未来，需持续优化规制路径，推动技术创新与合规发展良性互动，助力我国生成式人工智能产业高质量发展。

参考文献：

[1] 施雪森. 人工智能生成内容数据训练行为的规范路径——基于信息网络传播权限制的视角[J]. 西部法学评论，2025(06).

[2] 参见《〈2024 年中国数据交易市场研究分析报告〉全面剖析数据流通市场的发展态势 | 重磅成果》，载“上海数据交易所”微信公众号 ,https://mp.weixin.qq.com/s/PdUvrxuCFBayi5cB3kVh8Q，2026 年 1 月 28 日访问.

[3] 邓皓 . 生成式人工智能训练数据的著作权合理使用研究[J]. 传播与版权，2025(23).

[4] 赵立冬 . 合理使用还是法定许可？生成式人工智能训练数据著作权规制例外路径研究[J/OL]. 图书馆建设，2025：1-15[2025-10-10]. https：//link.cnki.net/urlid/23.1331.G2.20250917.1316.002.

[5] 刘水美 . 人工智能数据训练著作权合理使用法律规则路径探究[J]. 暨南学报（哲学社会科学版），2024(11):60-73.

[6] 聂洪涛，侯景译 . 生成式人工智能数据训练著作权侵权风险及规制路径[J]. 西华大学学报（哲学社会科学版），2026(01).

[7] LEVAL P N. Toward a fair use standard[J]. Harvard Law Review, 1990, 103(5): 1105-1136.

[8] 陈咏梅，郝悦彤 . 著作权视角下“合理使用”在生成式人工智能场域的适用：以美国《版权法》所涉相关案例为分析中心[J]. 国际经济法学刊，2025(03):87-101.

[9] Bartz v.Anthropic PBC,No.C 24-05417 WHA,2025 WL 1741691(N.D.Cal.June 23,2025).

[10] 刘禹 . 机器利用数据行为构成著作权合理使用的经济分析[J]. 知识产权，2024,34(3):107-126.

[11] 徐小奔 . 论人工智能生成内容的著作权法平等保护[J]. 中国法学，2024(1):166-185.

[12] 再论生成式人工智能的侵权风险及其应对[J]. 广东社会科学，2026(1).

[13] 林北征 . 个人信息匿名化概括式立法的困境与完善[J]. 行政法学研究，2024(6).

[14] 叶胜男 . 生成式人工智能训练数据合理使用的制度完善[J]. 数字法治，2025(6).