

# AIGC 时代新闻作品版权价值的实现进路

张惠彬 陈 妮

**摘要:**“《纽约时报》诉 OpenAI”一案凸显了生成式人工智能技术在新闻作品应用中所引发的版权争议。这一争议的核心在于,生成式人工智能服务是一种新型媒介,新闻作品不仅具有传统版权作品的属性,还兼具数据资产的特性。单纯的合理使用制度已经难以解决技术公司与版权人之间的紧张关系。为此,首先,完善“选择进入”机制,通过建立强制协商制度,确保版权人在新闻作品被用于 AIGC 训练前享有决定权,并落实披露义务,增强透明度与可解释性。其次,在“进入之后”采取保障措施,一方面,要求 AIGC 生成内容必须注明新闻作品出处,并通过自动标注与分阶段审核确保准确性;另一方面,确立合理的公平获酬机制,综合考虑作品价值、利用程度及商业收益,确保版权人获得公平报酬。这些措施均旨在平衡版权保护与技术创新,促进新闻作品在 AIGC 时代的健康利用。

**关键词:** AIGC; 新闻作品; 数据资产; 合理使用; 公平利用

**中图分类号:** D9 **文献标志码:** A **文章编号:** 2096-5443(2025)02-0033-12

**基金项目:** 重庆市教委科学技术研究重点项目(KJZD-K202300304); 2024 年重庆市研究生科研创新项目(CYS240207)

## 一、问题及其意义

AIGC(Artificial Intelligence Generated Content)是指利用人工智能技术自动生成的内容,这些内容可以包括文字、图片、音频、视频等多种形式。与传统媒体形式相较,AIGC 展现出自动化、高效率及创新性的独特优势。它凭借算法与模型,能够深度学习海量数据,能够创造出新颖且富有创意的内容,为媒体行业开辟了全新的可能性与发展契机,在新闻报道、广告策划、影视制作、音乐创作等多个领域,均展现出了广阔的应用潜力<sup>[1]</sup>。

作为一种新型媒体形式,AIGC 的诞生与发展依托于对海量数据的学习,其中就包括了大量受版权保护的新闻作品。然而,这一行为却触及了版权法的敏感神经,引发了版权人与技术公司之间的激烈争议。2023 年 12 月 27 日,《纽约时报》在美国纽约南区地方法院对微软公司、OpenAI 人工智能公司(以下简称 OpenAI)及相关公司提起了诉讼<sup>[2]</sup>。2024 年 11 月 29 日,加拿大多家主流媒体以侵犯版权为由,对开发 ChatGPT 的美国开放人工智能研究中心 OpenAI 联合提起诉讼。参与起诉的原告方包括加拿大《多伦多星报》及该集团旗下的大都会媒体集团、邮报集团、《环球邮报》、加拿大通讯社、加拿大广播公司等加拿大具有广泛影响的媒体。它们指称,OpenAI“经常违反”版权和在线使用条款,从加拿大媒体上抓取大量内容,以帮助开发 ChatGPT 等产品。面对这一争议,我们不得不重新审视生成式人工智能时代新闻作品的版权价值。新闻作品不仅承载着传统版权作品的属性,更在数字化浪潮中具备了数据资产的特性。因此,单纯的合理使用制度已难以解决技术公司与版权人之间的紧张关系。如何在保护版权的同时,促进数据的流通与技术的创新,如今成为摆在面前的一道难题。

仔细分析上述诉讼,核心问题在于生成式人工智能语料训练中使用大量的受版权保护的作品数

据是否合法?具体而言,在生成式人工智能的训练和生成过程中可能会产生侵害版权的行为。

在生成式人工智能的训练环节,需要使用大量的数据进行模型培育,其中不可避免地会涉及使用他人享有版权的作品作为训练素材的问题。在机器学习的整个流程中,数据的聚合、准备输入以及创建数字化副本都是必不可少的环节。这样的处理使得机器学习系统能够高效地分析数据并提炼出有价值的信息<sup>[3]</sup>。然而,倘若在未经版权人明确授权的情况下,通过绕过技术保护措施等手段获取这些数据,极有可能触发版权侵权的法律风险;倘若AIGC系统所输入的数据中包含大量未获版权人许可的作品,那么这种行为便有可能侵犯到版权人的复制权。再进一步,在大语言模型的预训练及后续微调过程中,对原始数据的清洗、知识抽取以及概率引导等操作,不仅可能涉及改编、汇编、翻译等潜在的侵权行为,同时,如果在此过程中删除了作品权利管理的电子信息,同样也可能构成对版权的侵犯。

在生成式人工智能的生成环节,若AI生成的作品使用了受版权保护的原作或片段,或对原作进行演绎,则这些行为将受到版权法的限制。例如,在“《纽约时报》诉OpenAI”一案中,《纽约时报》指控OpenAI为生产GPT模型,从其网站抓取或从第三方数据集中复制其受版权保护的作品,通过存储、处理和复制包含数百万份《纽约时报》作品的训练数据集来训练该模型,并最终通过ChatGPT产品传播包含《纽约时报》作品副本和衍生品的生成输出。上述行为均可能被认定为版权直接侵权。

在实际应用中,由于各个生成式人工智能的基础技术设置和技术行为存在差异,同时对于具体指令的输入也存在个性化选择,因此在不同的案例中,判断其行为是否构成侵权存在较大的分析空间和实例佐证需求。例如,“《纽约时报》诉OpenAI”一案中,通过列举多个实践案例来论证OpenAI对其作品的侵权行为。然而,假设一个生成式人工智能的机器技术达到极高水平,其在数据筛选、信息获取、内容生成等操作中不存在复制或突破技术手段的行为,且仅学习思想而非表达,那么难以认定其构成版权侵权。

在此情境下,本研究并非在于全面剖析具体案例中的问题或给出明确的侵权判断,而是着重探讨这项新兴技术如何以空前未有的姿态挑战法律的界限,并努力在技术发展与版权保护之间探寻一种均衡之道<sup>[4]</sup>。在此过程中,值得注意的是,AIGC所使用的大部分训练数据均涵盖了受版权法保护的资料。而对这些资料的获取及使用行为,极有可能触及版权法的红线,进而对生成式人工智能的未来发展构成潜在威胁。因此,在助推技术进步的同时,必须充分考量AIGC服务提供者和版权人的利益分配问题,探寻两难困境的最优解,实现新闻作品版权价值的最大化,也确保人工智能技术与版权法的和谐共生。

## 二、认知更新:AIGC时代新闻作品的双重价值属性

AIGC时代新闻作品版权保护的困境,源于技术发展趋势触动了利益的重新分配。传统新闻作品受版权法保护,面对生存挑战,传统媒体的主要应对策略是借助其丰富的作品集向数字化方向转型,其主要营收来源于广告和新媒体流量等收入变现<sup>[5]</sup>。随着AIGC作为一种新型传播主体的崛起,其在新闻作品的复制、使用和输出方面形成了强大的技术趋势。为了实现AIGC在新闻领域的技术功能,必然需要将作品的使用价值传递给该行业。然而,这一过程中不可避免地涉及版权人的权益保护问题以及传播主体之间的利益冲突。在利益分配的背后,必须认识到,新闻作品原有价值体系的深刻变革是导致其需要向技术传递价值的核心原因。换言之,随着时代和技术的迅速演进,新闻作品被赋予了新的时代价值,这种价值的创新直接反映在与其紧密相关的版权保护上。为了缓解传统媒体与AIGC服务提供者之间的版权冲突,首先需要深刻理解AIGC时代是如何重塑新闻作品价值的。从版权价值的角度看,AIGC时代的到来改变了新闻作品的创作与传播方式,使新闻作品在填补AIGC领域空白方面具有不可替代性。同时,新闻作品的数字化形式在AIGC时代数据流动性的推动下,显然具备了数据财产的双重价值属性。

(一)AIGC 赋新新闻作品的版权作品价值

1. 创作价值的不同

在 AIGC 时代,新闻作品对 AIGC 技术而言具有极高的创作价值,其主要体现在新闻作品作为数据资源对 AIGC 训练与应用的核心支撑作用上。从技术层面看,数据系数字经济时代的“石油”,数据行为的突破是生成式人工智能区别于传统人工智能的根本因素,在生成式人工智能技术循环中处于核心地位。丹尼尔·罗德里格斯·马菲奥利(Daniel Rodríguez Maffioli)结合美国国家标准与技术研究院(NIST)提出的人工智能生命周期活动模型与博马萨尼(Bommasani)等人的研究,提出了生成式人工智能系统运作模型(图1),并大致分为以下几个技术步骤:①数据生成:用户、作者、数字共享组织等负责数据的生成。②预处理:开发人员或第三方代表从网络中收集或挖掘数据,然后将这些数据整理并组合成训练数据集。③预训练:将训练数据集输入基础模型,通过无监督学习架构,基础模型学习输入数据的有效表示。④调整/微调:在基础模型的基础上,供应链下游的人工智能服务提供商采用有监督学习和标记数据创建新模型,以执行特定任务。⑤应用:将调整后的模型嵌入消费软件产品或服务中,并进行公开部署。用户可以通过系统提示(输入),系统根据提示和从训练数据中学到的模式随机输出内容<sup>[6]</sup>。细究该流程中的各个节点,可以发现数据贯穿从数据获取到模型部署的全过程,且始终作为核心技术对象处于技术流程中的关键地位。可以说,整个生成式人工智能就是一个巨大的“数据游戏”。要想让它变得“智能”,就需要给 AI 足够的原始数据和基本数据来支持。

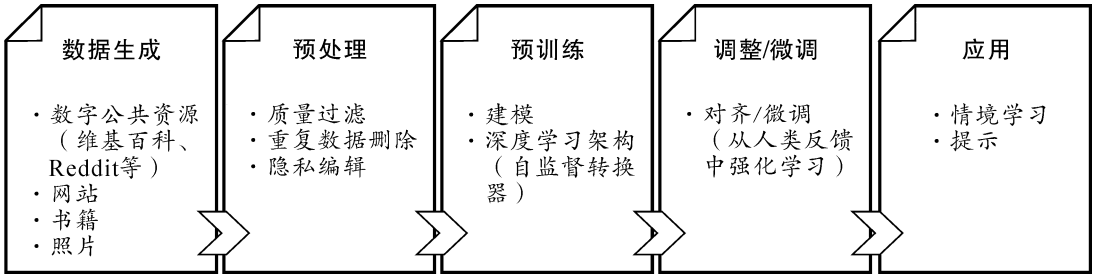


图1 生成式人工智能系统运作模型

而作为综合性或者新闻专业化的生成式人工智能类型,为充分实现其新闻信息内容生成和输出的功能,其需要大量获取新闻作品来填补其在新闻领域的信息空白。新闻作品,作为经过专业编辑和核实的信息载体,为 AIGC 提供了高质量的数据基础。这些数据不仅涵盖了广泛的社会事件、政策动态、经济趋势等,还包含了丰富的语言表达、逻辑结构和文化背景信息。这些元素共同构成了 AIGC 模型理解世界、生成内容的重要基石。以《纽约时报》为例,虽然 OpenAI 并未公布太多有关 GPT-4 的信息,《纽约时报》方就已知 ChatGPT 迭代过程中公布的信息推断出,在 ChatGPT 的训练过程中使用的 Common Crawl、WebText 和 WebText2 数据集之间,OpenAI 很可能使用了其所拥有的数百万篇受版权保护的作品来完整地训练 GPT 模型。同时,OpenAI 自身也意识到,高质量的内容(如《纽约时报》等)对于 ChatGPT 的培训更为重要,也更具价值。

2. 传播价值的不同

在 AIGC 浪潮的推动下,新闻作品的传播价值得到了显著提升和重塑。传统新闻作品的传播方式受限于物理媒介和时空限制,主要通过报纸、电视、广播等渠道进行传播,其受众范围相对固定,传播速度和效率也受到一定制约。但自人类文明步入数字时代,数字技术及其技术体系的迭代、叠加,开启了一个“加速”的新闻时代的来临。人工智能与数字技术的结合,更是使得非人类的新闻传播活动获得空前拓展<sup>[7]</sup>。然而,在享受 AIGC 带来的传播效率提升的同时,也引发了对新闻真实性和准确性这两个新闻作品核心传播价值是否得到了妥善保护的担忧。

在新传播技术建构的场景中,传统的新闻业态遭遇挑战,新的新闻业态尚未成型,新闻传播尤其是新闻从业者创作作品的传播场景变得异常复杂<sup>[8]</sup>。在当今的信息生态系统中,错误信息充斥着互联网、电视和其他媒体,公众更难分辨事实与虚构。由于AIGC技术依赖于大量的数据和算法模型,如果数据存在偏差或算法模型存在缺陷,就可能导致生成的新闻内容出现错误或误导性信息。这种错误信息的快速传播,不仅可能损害受众的知情权,还可能对社会稳定和公共安全造成潜在威胁。一个突出的案例是在“《纽约时报》诉OpenAI”的案件中,《纽约时报》控诉OpenAI的生成式人工智能技术产生了大量虚假信息和错误内容。具体而言,《纽约时报》指出,OpenAI的ChatGPT和必应聊天等产品在生成回复时,错误地将虚假信息归于《纽约时报》,甚至捏造了《纽约时报》从未发表过的文章和观点。这种行为不仅损害了《纽约时报》的声誉和公信力,还可能误导广大受众,对社会造成不良影响。

相较而言,传统媒体在新闻采编的全过程中,始终秉持着严格的审核与校对原则,以切实保障新闻内容的真实性和准确性。记者们通过深入实地采访、广泛搜集并多方验证信息,获取最为直接的一手资料。在此基础上,通过对资料细致入微的分析和客观公正的梳理,最终产出具有深度和广度的新闻报道。这种以事实为依据、以证据为支撑的新闻报道模式,构成了传统媒体长期以来积累的核心竞争优势,也是其在新闻传播领域始终保持权威性和公信力的坚实基础。为了对抗生成式人工智能带来的虚假信息洪流,作为新闻事实的直接创作源头的传统媒体兼具原始作品数据生产者的地位,应当充当健康新闻业的守望者,坚持发展独立和专业的新闻业<sup>[9]</sup>。

## (二) AIGC 赋新新闻作品的数据资产价值

### 1. 作为一种数据资产的新闻作品

在数字化经济的浪潮下,企业的商业生态和价值创造模式正在经历由数字化技术驱动的深刻变革。以媒体数据为例,传统媒体长期以来专注于发掘和创作高质量的独立新闻报道,投入大量人力物力进行新闻实地报道,以获取第一手新闻事实,从而形成独特的新闻资源。然而,随着电子产品的快速更新换代和信息生态系统的数字化趋势,传统的纸质媒体正面临着用户流失的严峻挑战,制作独立且优质的新闻作品变得愈发困难。传统媒体深刻认识到,其困境的根源在于用户连接的失效。为了应对这一挑战,传统媒体开始积极探索转型之路,意识到数字经济的重要性<sup>[10]</sup>。传统媒体积极利用大数据技术和资源,形成了集新闻文章、深度调查、观点文章、评论、操作指南等作品于一体的高质量媒体数据资源集,并受到版权保护。同时,这些媒体重塑了商业模式,将数字订阅、许可、广告流量和联盟收入作为主要营收来源,从而实现了在数字化浪潮中的转型与发展。可以说,在新闻领域,数字传播技术的出现与发展,以及其对新闻生产、分发和消费流程的全面渗透,已经在很大程度上重塑了新闻、新闻业和新闻学的面貌。这一变革不仅为传统媒体带来了新的发展机遇,也为新闻行业注入了新的活力<sup>[11]</sup>。而随着大数据技术的不断发展,新闻数据已然成为传统媒体的核心资产之一。

### 2. 数据资产价值与新闻作品保护的态度

生成式人工智能能够通过与自然语言的交互实现指令指引下的非特定内容的输出,其背后的技术体系将深刻地影响到数据的使用方式、范围和收益。尤其是在大规模使用过程中,数据表现出逐渐由资源化到资产化和资本化的转变<sup>[12]</sup>。当受版权保护的数据具有了突出的数据资产属性,显然对该类作品的版权保护不应当完全局限于传统版权法保护私权的理念之下,而应当兼顾数据资产的规模化、流通性和市场性,综合考虑数据效益的转化成效和权利保护的法律限度。

作品价值争议的本质,其实是两大传播媒介在市场竞争中的地位博弈。传统媒体依靠版权法来捍卫其新闻作品的权益,但这种做法也引起了人们对其是否为了保持竞争优势而限制作品流通的疑虑。对此,《著作权法》开篇明义,其立法宗旨在于捍卫作者的著作权及其相关权益,同时推动那些对社会主义精神文明和物质文明建设有益的作品的创作与流通,从而为社会主义文化和科学事业的进步与昌盛提供助力。这表明,这一制度设计的出发点是为了实现社会公益,确保在保护私人权益

和鼓励作品创作与传播之间实现微妙的平衡,而非相互背离。因此,我们既要保障作者的合法权益,也要确保作品能够广泛传播,以推动文化和科学的繁荣发展<sup>[13]</sup>。

从经济学的角度来看,版权制度在激励创作方面扮演着不可或缺的角色,它确保了作者的权益得到充分保障,对源头作品及数据的生成具有不可替代的重要作用。如果新闻机构无法自主生产和保护其独特的新闻报道,将会出现一个计算机或人工智能难以填补的空白,因此,对版权的制度性保护不应被忽视或废弃。然而,也必须认识到,过度的知识垄断可能会阻碍知识的更新和发展。版权保护不应成为抑制文化事业进步或创新的借口。过度关注单一或小规模数据产权的归属问题,可能会对未来数字化技术的发展造成阻碍,进而影响到我国新技术产品的迭代和竞争优势。因此,对于 AIGC 使用新闻作品的问题,不应采取一刀切的禁止态度。版权人也不应过分拘泥于权利的表象,而应顺应作品在 AIGC 时代进一步传播的必然趋势,避免过度保护作品而限制其传播和二次利用的可能性。

### 三、“逃向”合理使用:AIGC 时代新闻作品版权保护的困境

在 AIGC 引爆社会的背景下,以版权为代表的私权利益保护和数据价值流通的争端逐渐暴露。虽然 AIGC 训练过程中的版权侵权风险基本毋庸置疑,但是甚嚣尘上的数字决定论、传播效率论等观点,将注意力集中在数据自身的价值和在流通中的价值增值上,强调数字作品传播使用所带来的正外部性影响,提倡数据分享和数据流动,这对长期以来被人们所重视的版权保护造成了冲击<sup>[14]</sup>。在当前的媒体与 AIGC 服务提供者的版权之争中,可以看到一种趋势,即为了避免版权风险并扩大 AIGC 的发展空间,很大一部分舆论和实践思维倾向于寻求合理使用的规则作为出路。然而,这种倾向却忽视了现行合理使用制度在 AIGC 领域从解释论到立法论角度所存在的根本适用障碍,也未能充分认识到数据资产化的私有属性对于持有者自身的重要性。

#### (一)AIGC 背景下合理使用规则的解释论困境

##### 1. 美国“四要素”标准下合理使用的适用困境

为了确保版权法律权益得到充分保护的同时,不妨碍新闻媒体传播信息的自由,多数国家的版权法都明确规定了在新闻报道领域中的合理使用规则<sup>[15]</sup>。《美国版权法》第 107 条将“新闻报道(News Reporting)”作为合理使用的情形之一,还规定了“四要素”作为判定标准,即使用的目的与性质、该版权作品的性质、所使用的部分的质与量与版权作品作为一个整体的关系、该使用对版权作品潜在市场或价值所产生的影响。

为深化人工智能著作权合理使用制度的研究,有学者主张可借鉴美国的“四要素分析法”作为补充参考<sup>[16]</sup>。但实际上,“四要素”本身只是在特定情形下的参考因素,并非必要条件,对于各要素的考量与权衡,存在较大的裁量空间,尤其需要关注“度”的把握。以“使用的目的与性质”为例,需明确是否涉及商业性使用及转化性使用的可能性;在“版权作品的性质”方面,需区分作品类型以匹配合理使用的适度性;针对“所使用的部分的质与量”,需结合个案从定量与定性角度分析合理使用与侵权使用的边界;对于“使用对版权作品潜在市场或价值所产生的影响”,则需界定该使用是否逾越法律界限。

此外,亦有学者提出,人工智能训练数据原则上不与公众接触不损害著作权人市场利益,因此通过在现有框架内引入“四要素”从而将人工智能训练作品解释为合理使用具有必要性和可行性<sup>[17]</sup>。该观点实际上忽视了人工智能训练阶段的复杂性和使用目的的可转化性,也未关注到“四要素”在不同阶段的生成式人工智能中也呈现出不同的合理使用评估可能性。在预先训练基础模型阶段,预训练过程的独立性使其目的与行为性质尚不确定。若受版权保护的数据仅用于此阶段的预训练,其技术变革性质及对原始数据非竞争性替代使用的特性,可能使合理使用的抗辩得到支持。然而,在微调阶段,针对专门用于内容生成的模型进行的微调增加了侵犯版权的可能性,并降低了合理使用的适用空间。这是因为微调过程使用更具体的数据,增强了模型能力,并可能产生与原始数据相关的

人工智能生成内容,进而可能影响版权作品的市场。最后,在模型部署阶段,蚕食数据市场的风险显著增加。此时,合理使用的判定变得尤为复杂,主要依赖于模型所针对的特定任务和产出,很可能超出合理使用原则的保护范围。因此,面临的挑战在于如何界定市场以及确定模型的转化性和非商业性使用是否构成对数据所有者潜在市场的侵占。

## 2. 我国“三步检验法”标准下合理使用的适用困境

我国《著作权法》第二十四条及《著作权法实施条例》第二十一条完整呈现了“三步检验法”。具体而言,三步检验标准包括:合理使用需限于《著作权法》第二十四条规定的十三种特例;使用方式不与作品正常使用冲突,且结果不得不合理侵害版权人权益;应指明作者姓名或名称、作品名称。在实践中,适用“三步检验法”的难点往往在于如何界定“影响作品正常使用”和“不合理损害著作权人权益”。对此,法院通常会考量引用目的和性质、被引用作品性质、引用部分比例及使用是否影响原作潜在市场或价值等因素。

然而,就立法技术本身而言,“三步检验法”存在其固有缺陷。一方面,我国《著作权法》对何为正常使用、不合理的损害、合法权益目前尚无司法解释加以明确和细化。即便在《伯尔尼公约》与TRIPS协定中,“三步检验法”的判断标准依然被认为是开放的、可能导致不确定性的。另一方面,我国“三步检验法”中对概括性合理使用法定情形的列举和兜底性条款突破的不易不可避免地导致了面对新技术的滞后性。有学者提出主张兜底条款为接口构建生成式人工智能合理使用条款并辅以相关的配套措施<sup>[18]</sup>。但实际上,当“三步检验法”落入生成式人工智能领域时,其与以开放共享为基本特征的生成式人工智能技术的内在冲突确难以调和。

首先,概括性合理使用情形的列举本质上体现了立法者对于版权的限制持保守态度,虽然存在兜底性条款,但是其特定情形与条件仍应定性为“例外的”“狭窄的”,而生成式人工智能对作品的获取与使用通常是“普遍的”“海量的”“多元的”<sup>[19]</sup>,两者特质存在根本冲突。若要想生成式人工智能版权侵权落入合理使用范围,则首先应当确定生成式人工智能技术行为落入法定合理使用情形的范畴,这无疑难以具有普遍性和可操作性。结合生成式人工智能基础阶段的划分与《著作权法》第二十四条规定的法定情形,就人工智能模型开发的整体目的看,经过数据输入的预训练基础模型和在预训练基础模型上根据应用需求进行改编训练的生成模型仅止步于模型研发阶段,若尚未进入部署阶段,其目的可能止于个人学习研究或科学研究目的,但也依旧存在如OpenAI的开源转商业的主观目的转换的可能性;就生成式人工智能模型与具体作品之间的关系看,其根据技术设置与指令输入或调整的不同,可能达到为介绍、评论、说明某一作品或某一问题而适当引用他人作品的效果,但在法院受理的各类版权侵权案件中,适当引用通常是被诉侵权的被告经常采用的抗辩意见,真正被法院采纳作为适当引用情形的并不多<sup>[20]</sup>;就生成式人工智能获取的数据类型不同,若其使用的是媒体在公众集会上发表的未禁止刊登播放的讲话也可能归入法定合理使用情形,但此类讲话也仅占新闻内容产出的极小部分。可见,生成式人工智能归入法定合理使用情形与否则仍需根据使用内容、作品的适用目的和使用形式来个案讨论。鉴于不同案件的情况和应用程度各不相同,对于这类案件如何仔细审查事实,并作出合理的归入判断,依旧是审判实务中一个令人困惑的问题。另外,2020年修订的《著作权法》第二十四条新增了兜底性条款,即“法律、行政法规规定的其他情形”也可作为法定合理使用情形。虽然该兜底性条款确为生成式人工智能有朝一日顺理成章走入合理使用框架留下了看似光明的希望,但目前并无相关法律和行政法规对该条款作出回应,且我国司法实践中对于兜底性条款突破始终慎之又慎,难言从司法层面解决个案性从而进行类型归入<sup>[21]</sup>。

其次,关于“不得与作品的正常使用相冲突”“不得损害著作权人的合法权益”目前尚无司法解释加以明确和细化,国际条约规定和理解亦存在争议,这也造成了司法认定标准缺乏可预期性。例如,对于前者问题,《伯尔尼公约》研究小组认为,其认为对作品的一切形式的利用,凡具有或可能具有显著经济或实用价值的,都必须保留给作者。而世贸组织讨论得出,除了对目前产生大量或有形

收入的开发形式的保护外,还有在一定程度上具有可能性和可信性,可能获得相当大的经济或实际重要性的开发形式,也即此处的正常使用应考虑作品目前和未来潜在的市场<sup>[22]</sup>。对于后者问题,关于“合法权益”的理解仍然存在狭义的版权所赋予的专有权的经济价值利益以及广义的著作权权益乃至民法权益之争。在生成式人工智能版权与合理使用之争中,版权方往往试图确立其对作品享有广泛的利用保护和获得经济收益的权利,而在实际中,若以生成式人工智能的长远发展以及未来数据流通信息共享为视角,并着眼于版权方和生成式人工智能开发者的各自诉求与立场,不难发现,数据、作品流通的壁垒往往并不在于版权方对其作品极强的控制欲与封闭要求,相反,若生成式人工智能的使用并不会冲击版权方自身的作品市场且能给予版权方一定的经济回报或作品、作者流量,版权方或对其作品的被获取、被使用可能也是乐见其成。

最后,关于“指明作者姓名或者名称、作品名称”,一方面,若对上述“不得损害著作权人的合法权益”采广义解释,则此处特别规定作者姓名名称或作品名称的标注似乎有与之重复之疑;另一方面,生成式人工智能的深度学习特性使得其在对原始数据进行识别和贡献度的标注方面面临较大挑战,因此,建立统一且规范的标注体系显得尤为必要。

## (二) AIGC 背景下合理使用规则的立法论争议

为顺利将 AIGC 行为平滑归入合理使用,有学者提出应当将生成式人工智能数据训练作为整体纳入合理使用法定情形<sup>[23]</sup>,也有学者建议在合理使用制度框架下增设“文本与数据挖掘”的合理使用情形且不论使用主体是否以营利为目的<sup>[24]</sup>。两种观点均认识到了现有“三步检验法”和兜底条款并不能完全满足裁判需要,并提出应当将 AIGC 数据行为以合理使用情形或者其他法规条例予以明确。具体来看,其论据主要在于数据挖掘作为人工智能时代的基础性技术,合理使用规则的缺失必然会限制信息自由流动与创新发展。且基于现实层面考虑,在我国人工智能尚处于初创阶段之时,基于合理使用制度下的免费使用无疑能为人工智能技术的研发节约更多的成本,满足人工智能开发主体的研发需求,故许可主体对数据加以利用,并不禁止以营利为目的的使用方式是当下最好的选择。

根据合理使用规则的立法意图来看,其作为一项侵权抗辩,是版权保护法的例外情形,它突破了版权人享有的专有权利状态,允许在特定情况下对作品进行使用或改编,从而实现了权利的再分配和利益的共享。其制度目的在于平衡版权人对作品享有的专有权利与公众获取作品的需求,促进创新和文化多样性发展,保障公众基本利益。上述观点看似综合考虑了版权人及人工智能开发主体的利益,也提供了一个理想的方案。但从其陈述来看,不免看出其对于技术的偏爱。该观点认为现有著作权法对版权人的过度保护将使得受版权保护的数据处于封闭环境下,认为此种制度构造不利于数据的流通和 AIGC 技术的发展,但实际上忽视了数据尤其是受数据保护的作品之于版权人本身的重要价值,尤其是当数据规模化后一举成为企业赖以生存的核心数据资产,其私有属性应当受到充分的尊重与保障。

就新闻领域的作品而言,传统媒体兼具原始数据生产者和传播者一职,新闻作品的数据化、集成化显然作为类似其产业“逃生通道”一般的存在,尤其是当数字订阅、许可、广告收入作为主要营收来源时,新闻产业的循环完全依赖于其作品创造的价值。同为传播者的生成式人工智能出现以后,在宏观、中观和微观三个层面上分别对现有信息体系进行了重塑<sup>[25]</sup>。这必然会出现一些版权权益上的让渡,但“完全免费”的合理使用制度完全可能在盘活 AIGC 行业的同时,在某种意义上抹杀数据创建者尤其是版权人之于数据、作品所享有的权益。以“《纽约时报》诉 OpenAI”案中体现的生成式人工智能技术应用为例,其对于《纽约时报》作品内容的逐字摘录和详细摘要,难言其变革性价值。而在此背景下,AIGC 技术出现所谓的趋势化和创造性不断被强调,能够通过简单的指令获知新闻信息或作品对于用户显然更具有诱惑力,并且必然会侵蚀原有的媒体市场,分走原本属于传统媒体的那一部分蛋糕,新闻作品的创作积极性将被极大削弱,而当生产的源头受到阻塞,各类主体的新闻传

播行为都会面临内容的真空。

而作为后来居上者和既得利益者,目前的AIGC并不能很好地实现其传播功能。在“《纽约时报》诉OpenAI”案中,令人在意的地方在于ChatGPT在吸收消化了大量新闻作品数据后仍然产生了虚假新闻并将其归之于传统媒体。当作为新闻作品新型传播者的AIGC挤走了一线独立且高质量新闻作品生产者的传统媒体,独占市场鳌头,AIGC引以为傲的信息传播效率乃至其背后产出作品的新闻业本身也被打上虚假的标签,将会导致正常的社会传播系统秩序失衡,“新闻像谣言,谣言像新闻”的怪象出现,传统媒体的公信力一降再降,最后造成新闻数据源头之水的逐渐干涸。显见的是,AIGC的应用潜力一方面源于其数据整合行为对于人类信息获知成本的节约与效率的提升,而同时很大程度上也体现在人类对其后输出内容创造性的期望,在此过程中人工智能机器学习产生的知识增量,应当是在作品原有价值之外发现新价值的过程<sup>[26]</sup>。在被喂食了大量的新闻数据之后,如果人工智能的最终产品不是完全独立的,或者说最终产品与原产品基本一致,甚至是出现有虚构事实和虚假信息,那么人工智能仍将失去其智能的特性,甚至助推数字时代的虚假性,成为“可靠的谣言缔造者”。

#### 四、超越合理使用:AIGC时代构建新闻作品公平利用体系

由上观之,无论是现有的以“三步检验法”为核心的合理使用,还是倡导的引入AIGC专门性情形的变革型合理使用制度,在缺乏制度解释以及对新闻作品价值正确认知的情况下,目前的合理使用制度并不具备应对本文反复讨论的复杂性和细微差别的能力,也无法较好解决目前生成式人工智能所带来的利益权衡问题,且从根本上与数据资产以及版权的保护需求相冲突。要想为这些错综复杂的挑战找到潜在的解决方案,必须避免简单的二元对立。在数据共享的发展趋势下,若要维持创新的动力,则势必要在兼顾利益平衡性的要求下打破原有的数据利益分配格局,既不能忘记版权法激励社会创新的初衷,也不能忽视数据持有者,尤其是其中版权所有者的权利需求,构建新闻作品公平的利用体系。

##### (一)“选择进入”的机制完善

###### 1. 强制协商:选择进入机制的建立

“选择进入”机制,指的是赋予了新闻作品版权人决定是否将其作品数据用于大型模型训练的决定权。为此,应当通过建立前置协商机制,以纠正人工智能开发者因权利人“不知情同意”而过度使用的行为,确保数据获取的合法性,为人工智能的后续开发与应用奠定坚实基础。有学者对“选择进入”策略表示关切,认为这可能会导致版权作品数据库规模偏小、优质训练数据稀缺,进而使得生成式人工智能模型的效果难以得到保障<sup>[27]</sup>。然而,作为新闻作品版权集中管理方的媒体机构,其具备类著作权集体管理组织的功能,在与AIGC平台的协商中拥有天然的组织优势。对于人工智能公司而言,作为版权规制的对象,应基于自身数据需求和技术实际,主动向目标数据作品的版权所有者提出协商授权,特别是那些拥有大量版权的权利方。事实上,私法自治下的协商已经成为人工智能公司解决版权风险的重要途径。例如,面对版权诉讼的压力,OpenAI已经与美联社等内容出版业者达成了内容授权协议;试图在生成式人工智能领域追赶OpenAI的苹果公司,也正在与内容出版业者就使用内容的协议进行磋商。

然而,仅凭自愿选择,可能难以全面保障其权益。在《纽约时报》的案例中,虽然他们最初尝试与OpenAI和微软进行友好磋商,但并未能达成解决方案。这恰恰说明了在缺乏强制性协商机制的情况下,权利人可能难以在谈判中取得应有的话语权。因此,为了构建一个能够促进谈判的平台,可以借鉴法国的强制谈判政策。为给出版商、新闻机构和数字平台之间的平衡谈判创造条件,法国于2019年将欧盟《数字单一市场版权指令》(以下简称《版权指令》)第15条转化为国内的版权法新规,要求数字平台在利用新闻作品数据之前,必须与版权方进行协商并获得正式授权。这一政策的实施,不仅赋予了新闻机构与数字平台平等谈判的权利,还通过实际案例展示了其在AIGC应用场景下的有

效性。例如,法国竞争管理局曾对谷歌发出禁令,要求其与新闻机构和出版商就使用受版权保护的内容进行谈判并支付相应的费用。最终,谷歌与出版商达成了和解协议,承诺进行诚信谈判,并向出版商提供评估报酬所需的信息。

综上所述,选择进入与强制协商是构建生成式人工智能数据使用合理机制的关键要素。选择进入机制赋予权利人决定作品是否被使用的主动权,而强制协商机制则确保这一权利在实际操作中得到有效保障。

## 2. 披露义务:透明度原则的体现

随着生成式人工智能技术的迅猛发展,新闻作品被广泛应用于模型训练中,在这一过程中,生成式人工智能服务往往被视作“黑匣子”,其内部运作机制对于外界而言充满了神秘感,版权人则处于信息不对称的弱势地位,难以知晓自己的作品是否被使用,更无法评估其使用情况和潜在影响<sup>[28]</sup>。因此,为了确保新闻作品版权人能够及时发现其作品被 AIGC 服务提供者使用,作品披露义务的落实显得尤为重要<sup>[29]</sup>。

从理论层面出发,透明度原则是作品披露义务的重要支撑。对此,《欧盟人工智能法案》中关于透明度和可解释性的相关规定也为理解透明度原则提供了深刻的洞见,共同组成了这一原则的实际内容。具体来说,“透明度”强调人工智能企业应公开其数据获取渠道、数据处理手段以及数据使用意图,从而确保用户和监管机构能够获得详尽的信息,深入了解人工智能体系的运行机制、可能存在的偏见及其局限性。同时,“可解释性”强调人工智能体系的操作和决策流程应具有清晰易懂的特性。这要求人工智能企业阐释其算法和模型的工作机制,使得用户和受影响的个体能够理解人工智能体系做出决策的逻辑。这一原则的实施,不仅有助于提升用户对人工智能体系的信赖度,还可以有效减弱人工智能行为的无序性,从而有力地保障版权人的合法权益。

在实践层面,作品披露义务的落实需要明确义务主体、界定披露范围,并采取合适的披露方式。义务主体显然应是那些利用新闻作品进行模型训练的生成式人工智能开发者。他们作为技术应用的主体,应承担起披露作品的责任,确保版权人的知情权得到充分保障。披露范围则应涵盖所有被用于模型训练的作品信息,这些信息应足以使版权人能够识别出自己的作品是否被使用,并了解作品被使用的具体情况。此外,需要明确的是,信息披露并不等同于强制要求人工智能开发者公开其训练数据集或采取开源策略。仅仅要求生成式人工智能公开其数据集或允许数据提供者访问庞大的数据集,虽然在形式上满足了透明度的要求,但实际上忽略了数据提供者是否具备确认自己的数据或作品被纳入训练数据集的能力,以及评估其贡献度和是否存在侵权问题。因此,人工智能开发者无需提供完整的数据集,而是应根据模型类型、使用阶段或建立系统的任务,对数据或作品的使用进行有针对性的披露。

## (二)“进入之后”的保障措施的保障

### 1. 输出的内容必须注明出处

在 AIGC 技术广泛应用的背景下,新闻作品的版权保护面临着前所未有的挑战。为了确保新闻作品在 AIGC 生成内容中的版权,尤其是署名权等权益得到妥善保护,输出内容必须注明出处成为一项基本而重要的保障措施。同时,这一要求通过外化数据来源信息以及对来源显示的动态核查、矫正,实现了对于数据出处的进一步把控,不仅符合我国《生成式人工智能服务管理暂行办法》对于数据合法来源的要求<sup>[30]</sup>,也可以作为 AIGC 服务提供者风险防控的重要手段。具体来看,该项措施的执行可以从以下几个层面展开:

首先,AIGC 服务提供者应建立有效的自动标注机制,该机制应能够自动识别并标注生成内容中所使用的新闻作品元素,包括但不限于新闻作品的标题、作者姓名、出处(如报纸名称、网站链接等)以及使用日期。另外,为了保障标注工作的统一性和标准化,AIGC 服务提供者需制定详尽的标注准则,其内容应涵盖标注的格式规定、位置要求、内容规范等细节,从而确保标注行为能够有效满足版

权权利管理信息保护等相关标准。

其次,在自动标注的基础上,为了确保标注的准确性和真实性,AIGC 服务提供者应设立一套高效的审核流程,可以将其细化为如下分阶段、有重点的审核机制:

第一,初始自动审核阶段。这一阶段的目的在于通过系统的高速性自动完成标注行为的核查与矫正,实现系统内部的良性运作。在这一阶段,AIGC 系统将对自动标注的结果进行初步筛查,主要检查标注的真实性、完整性、格式规范以及是否存在明显的错误,也即首先确定标注信息的真实性与完整性,其次快速过滤掉大部分不符合规范或存在明显问题的标注,而后重新完成识别与标注行为。

第二,定期抽检与反馈机制。除了自动审核外,AIGC 服务提供者还应建立定期抽检机制,通过人工手段对已经发布的生成内容进行随机抽检,以验证自动标注效果。同时,应设立用户反馈渠道,鼓励用户就标注内容的准确性提出意见和建议,并就用户反馈的问题进行及时的核查和处理,不断提升标注工作的质量和效率。

## 2. 合理的公平获利机制确立

在版权许可领域内,市场失灵现象时有发生,即当交易成本超出公众支付意愿的界限时,版权使用申请者在寻求版权所有者许可的过程中可能遭遇阻碍,导致作品无法得到充分有效的利用与开发,进而阻碍社会福利的增长<sup>[31]</sup>。因此在权利人与 AIGC 系统提供者就新闻作品的使用达成授权合意后,许可费条款将成为保障权利人经济权益的核心内容。为了确保权利人的公平获利,需要构建一套科学、合理、透明的许可费机制。

首先,在 AIGC 技术广泛应用的今天,确立公平合理的许可费标准显得尤为重要。对此,可以借鉴欧盟《版权指令》中的“公平报酬”机制,根据作品的实际使用情况和商业价值来确定应得的报酬:①报酬与实际价值相匹配:新闻作品版权人的合理报酬应与其版权许可的实际价值或潜在经济价值相匹配。在确定报酬时应综合考虑多种因素,包括但不限于作品的知名度、影响力、创作难度、目前或潜在的市场价值,以及 AIGC 系统对作品的利用程度、商业收益等,从而确保报酬的合理性。②使用情况披露义务:被许可方有义务积极披露作品的后续使用情况,并定期向版权人提供详细信息,包括使用方式和情况等。这样,版权人能够充分了解作品的实际使用情况,从而根据商业价值来确定应得的报酬。③合同调整机制:为平衡双方报酬约定的不均衡,引入合同调整机制。若版权人在签订合同时对未来作品商业价值的预估不准确,导致最初约定的报酬与实际收益不符,他们可以要求被许可方根据市场实际价值支付附加报酬。同时,考虑到新闻作品的时效性和传播特性,许可费标准还可以设置动态调整机制,以适应市场变化和技术进步带来的新情况。

其次,明确许可费支付方式和周期也是保障新闻作品版权人权益的重要环节。支付方式应简便快捷,可采用预付费、按次计费或按流量计费等多种灵活方式,以适应不同权利人的实际需求。同时,支付周期的明确也至关重要,无论是按月、按季度还是按年支付,都应确保权利人能够定期、稳定地获得应有的经济回报。为了防范未来可能出现的支付纠纷,双方应在协议中详细规定支付方式和周期的变更条件与程序。

最后,建立许可费审计和争议解决机制对于维护新闻作品版权人与 AIGC 服务提供者之间的合作关系同样至关重要。通过引入许可费审计机制,可以定期对支付情况进行深入核查,从而确保每一笔费用的公平性和合理性。同时,完备的争议解决机制如协商、调解、仲裁或诉讼等,为双方提供了多元化且高效的纠纷解决途径。在整个过程中,应当始终尊重并保障权利人的知情权和参与权,力求每一个争议都能得到公正、及时且有效的处理。

## 五、结论

本文通过对 AIGC 时代新闻作品版权价值的深入探讨,全面分析了新闻作品在生成式人工智能背景下所面临的版权保护困境及实现价值的路径。笔者从 AIGC 技术发展的背景出发,强调其对新

闻作品版权保护带来的挑战,特别是版权侵权问题与合理使用规则的适用困境。随后,阐述 AIGC 时代新闻作品的双重价值属性,即版权作品价值和数据资产价值,并探讨这种双重价值属性对新闻作品创作与传播方式的影响。针对 AIGC 版权保护问题,笔者分析了现有合理使用制度的不足,并提出了增强透明度和可解释性、构建公平利用体系等具体措施。在措施方面,笔者建议通过强制协商机制和作品披露义务的落实给予版权人“选择进入”的自由,而后通过要求生成内容标注出处、建立合理的公平获酬机制充分保障版权人“进入之后”的权益保护问题,力求在此过程中实现权利人与 AIGC 服务提供者之间的合作共赢。

当然,尽管本研究有一定贡献,仍存在不足。第一,AIGC 技术快速发展,相关法律和监管框架尚不完善,讨论版权问题时存在一定局限性。第二,具体措施虽基于现有法律框架和技术趋势推断,但实际效果需通过实践验证。AIGC 时代新闻作品版权价值的实现进路是一个复杂而重要的问题,需要在尊重版权、促进数据流通和技术创新之间找到平衡点。未来研究可进一步关注 AIGC 技术的发展趋势和法律法规的完善,探讨不同国家和地区的 AIGC 版权保护差异,加强实证研究验证措施的有效性。同时,平衡版权保护与技术创新关系,推动 AIGC 技术健康发展,实现多方共赢。通过研究,希望能够为新闻作品权利人和 AIGC 服务提供者之间的合作提供有益的参考和启示。

#### 参考文献:

- [1] 何强,杨崧,曾学真等.智能技术引领新闻样态革新——新华社全国两会报道的虚实融合实践.传媒,2024,9:19-20.
- [2] The New York Times Company V. Microsoft Corporation,1:23-cv-11195.[2025-02-01]<http://www.courtlistener.com/docket/6811749/the-new-york-times-company-v-microsoft-corporation/>.
- [3] 王利明.生成式人工智能侵权的法律应对.中国应用法学,2023,5:27-38.
- [4] 张惠彬,王怀宾.版权优先还是技术优先?——法国应对 AIGC 版权风险的趋势及启示.编辑之友,2024,5:103-112.
- [5] 陈国权.主流媒体经济支持体系的新构成——基于 136 家媒体的调研报告.现代传播(中国传媒大学学报),2022,4:1-10.
- [6] M. W. Carroll. Copyright and the Progress of Science: Why Text and Data Mining Is Lawful. UC Davis Law Review, 2019,53(2):893-964.
- [7] 谢湖伟,简子奇,沈欣怡.认知框架视角下 AIGC 对媒体融合的影响研究——对 30 位媒体融合从业者的深度访谈.新闻与传播评论,2023,6:5-18.
- [8] 胡泳.人工智能驱动的虚假信息:现在与未来.南京社会科学,2024,1:96-109.
- [9] 单宇,许晖,周连喜等.数智赋能:危机情境下组织韧性如何形成?——基于林清轩转危为机的探索性案例研究.管理世界,2021,3:84-104+7.
- [10] 刘翠敏.数字经济时代传统媒体融合发展路径分析.传媒,2023,6:28-31.
- [11] 袁曾.数据规模化利用的法律规制.地方立法研究,2023,5:71-84.
- [12] 董彪.二次创作短视频合理使用规则的适用与完善.政治与法律,2022,5:141-149.
- [13] 冯晓青,马彪.数字版权下合理引用制度的价值取向和制度完善——以用户生成内容为研究对象.法律适用,2022,4:53-62.
- [14] P. Goold. Artificial Authors: Case Studies of Copyright in Works of Machine Learning. Journal of the Copyright Society of the USA, 2020,67(3):427-470.
- [15] 林秀芹.人工智能时代著作权合理使用制度的重塑.法学研究,2021,6:170-185.
- [16] 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.河北法学,2025,3:185-200.
- [17] 刘云开.人工智能训练作品的著作权合理使用进路.东北大学学报(社会科学版),2025,1:117-126.
- [18] 关春媛.生成式人工智能训练版权合理使用探究:国际趋势、本土发展与规则构建.出版发行研究,2024,12:91-97.

- [19] 杜灵燕. 适当引用的判定. 人民司法, 2022, 2: 96-99.
- [20] 高阳, 胡丹阳. 机器学习对著作权合理使用制度的挑战与应对. 电子知识产权, 2020, 10: 13-25.
- [21] M. Senftleben. Compliance of National TDM Rules with International Copyright Law: An Overrated Nonissue? International Review of Intellectual Property and Competition Law, 2022, 53(10): 1477-1505.
- [22] 詹爱岚, 田一农. 生成式人工智能机器学习中的著作权风险及其化解路径. 电子知识产权, 2023, 11: 4-14.
- [23] 王一蓉. 文本与数据挖掘行为的版权风险及其化解. 科技传播, 2023, 20: 137-139.
- [24] 张文祥, 沈天健, 孙熙遥. 从失序到再序: 生成式人工智能下的信息秩序变局与治理. 新闻界, 2023, 10: 41-51.
- [25] 徐小奔, 杨依楠. 论人工智能深度学习中著作权的合理使用. 交大法学, 2019, 3: 32-42.
- [26] D. Eastman. Data Dignity: Developers Must Solve the AI Attribution Problem, 2023-05-27. [2025-01-20] <https://thenewstack.io/data-dignity-developers-must-solve-the-ai-attribution-problem/>.
- [27] 蒋舸. 论著作权法的“宽进宽出”结构. 中外法学, 2021, 2: 327-345.
- [28] 董青岭. 人工智能时代的算法黑箱与信任重建. 人民论坛·学术前沿, 2024, 16: 76-82.
- [29] 王影航. 人工智能创作物信息披露问题的著作权法规制. 中国出版, 2022, 21: 38-43.
- [30] 冯晓青, 沈韵. 生成式人工智能服务提供者著作权侵权责任认定. 法治研究, 2025, 1: 46-58.
- [31] 张吉豫, 汪赛飞. 大模型数据训练中的著作权合理使用研究. 华东政法大学学报, 2024, 4: 20-33.

## The Progress of Realizing the Copyright Value of Journalistic Works in the AIGC Era

Zhang Huibin, Chen Ni (Southwest University of Political Science & Law)

**Abstract:** “The New York Times v. OpenAI” case highlights the copyright controversy arising from the application of generative artificial intelligence technologies to journalistic works. At the heart of this controversy is the fact that generative AI services are a new type of medium, and journalistic works not only have the attributes of traditional copyrighted works, but also the characteristics of data assets. A simple fair use system has been difficult to resolve the tension between technology companies and copyright holders. Therefore, this paper proposes the following countermeasures. Firstly, improve the “opt-in” mechanism, through the establishment of a mandatory consultation system, to ensure that copyright holders have the right to decide on the use of journalistic works for AIGC training, and implement the disclosure obligation to enhance transparency and interpretability. Secondly, safeguard measures should be taken after “access”. On the one hand, AIGC-generated content should be required to indicate the source of journalistic works, and ensure accuracy through automatic labeling and phased auditing. On the other hand, a reasonable and fair remuneration mechanism should be established, taking into account the value of the works, the degree of utilization and commercial returns, to ensure that copyright holders are fairly compensated. All these measures aim at balancing copyright protection and technological innovation, and promoting the healthy utilization of news works in the AIGC era.

**Key words:** AIGC; news works; data assets; fair use; equitable utilization

■ 收稿日期: 2024-04-09

■ 作者单位: 张惠彬, 西南政法大学民商法学院; 重庆 401120

陈 妮, 西南政法大学民商法学院

■ 责任编辑: 刘金波