

【理论研究】

人工智能生成内容(AIGC)的 三个出版学议题

方 卿 丁靖佳

【摘 要】人工智能生成内容(AIGC)技术及其应用的发展,将彻底改变出版的内容生产范式和内容供给模式,在创新推动出版发展的同时也带来诸多风险与问题。对此,出版学需重点关注AIGC的权利归属、侵权、权益保障等著作权问题,技术伦理和学术伦理失范等伦理问题,以及进一步引发的意识形态渗透、文化价值观偏离等文化安全问题,并以此为逻辑起点探寻解决之道。

【关键词】AIGC;出版学;著作权;技术伦理;学术伦理;文化安全

【作者简介】方卿,管理学博士,武汉大学信息管理学院教授、博士生导师;丁靖佳,武汉大学信息管理学院2021级博士生(武汉 430072)。

【原文出处】《出版科学》(武汉),2023.2.5~10

对话式AI机器人ChatGPT的横空出世引发了人们对人工智能生成内容(Artificial Intelligence Generated Content, AIGC)的关注与热议,《自然》期刊发文直言这有可能彻底改变研究实践和出版实践,但其中的风险不容忽视^[1]。AIGC作为继专业生成内容(Professionally Generated Content, PGC)和用户生成内容(User Generated Content, UGC)之后的新的内容生产引擎,能够通过内容自动化编纂、智能化润色加工、多模态转换和创造性生成等方式直接改变出版的内容生产范式和内容供给模式。除此之外,AIGC还将以数字人、聊天机器人为交互媒介窗口推动出版流程的智能再造和出版服务的创新升级,从而对出版的发展产生深远影响。随着AIGC技术的革新迭代及其应用成本的下降,出版内容的生产将不可避免地从PGC、UGC转向AIGC(包括AIPGC、AIUGC,即AI协同创造内容),且AIGC的数量和内容占比将不断上升。在可预期的未来,AIGC将因其在内容易用性、可塑性、多样性、交互性、低延迟等方面的优势而成为出版活动的主要内容来源。

作为一个以知识内容选择、传播、传承为内在追

求的内容产业,出版也将直面AIGC带来的机遇与风险,从出版学视角对此加以研究与审视则将影响着AIGC以何种形式和何种路径加入人类知识成果的传播、传承序列之中,也是警惕和规制AIGC异化出版“文化构建”本质的关键。

1 AIGC的著作权问题

出版是内容产业,著作权是出版的核心资产。目前,AIGC已经能够实现完整的出版内容作品创作,《1 the Road》《阳光失了玻璃窗》等都是由AI技术创新重组而成的AIGC作品。ChatGPT等AI语言模型应用的“出圈”预示着AI协同创作以及创造性AIGC作品的生产成为可能,其中暗含的新型著作权争议则将对现有的著作权制度框架产生冲击。

一是AIGC的权利归属问题。权利归属是著作权法的核心,其可以明确作品究竟归属于谁,以及谁可以对所创作的作品享有权利并承担责任^[2]。AIGC的出现改变了传统内容的生产规律,使我们不得不对已有的著作权法的权利归属进行根本性反思。首先,是否需要AIGC作品进行著作权法上的赋权?如AIPGC、AIUGC等AIGC作品蕴含了自然人的

创作因素(思想、情感),在一定程度上符合作品构成要件的要求,该类作品能否进行赋权认定、其认定标准是什么等内容都还相对空白。其次,AIGC作品的主体资格如何确定? AIGC作品涉及AIGC技术开发者、应用所有者、功能使用者等不同利益相关主体,厘清以什么样的赋权形式将什么类型的AIGC作品归属于哪类主体是解决AIGC权利归属的关键。尤其是当AIGC技术发展至拥有自主意识,能够从组合型内容创造转向探索型和变革型内容创造,AIGC作品的权益收益和可能导致的侵权责任又该如何划分?

二是AIGC的侵权问题。AIGC的基本原理在于AI算法在海量数据中“学习”“理解”内容的生成逻辑与用户的阅读倾向,从而进行内容的再生产或再创作。在我国现行著作权法没有明确将数据爬取、数据挖掘等智能化分析行为规定为合理使用的情况下,AIGC极有可能对其他著作权人和使用用户构成侵权。一方面,AIGC对已有作品的大规模学习、模仿、组合和转化可能对原作者人享有的修改权、保护作品完整权、演绎权等权利造成侵权,国外最大艺术作品平台ArtStation上的作者们就AIGC算法中潜在的版权争议掀起了抵制AIGC的活动。另一方面,AIGC为形成符合用户需求的内容,还会收集、整理并利用用户的个人数据,存在用户隐私数据侵权和外泄的隐患。完善AIGC的侵权认定并对其侵权行为加以规制是出版业不得不探讨的现实问题。

三是AIGC的权益保护问题。虽然大家对AIGC的赋权问题还存在争议,但普遍认为应该给予满足具有独创性要求的AIGC作品著作权保护,我国“人工智能写作领域第一案”——腾讯诉上海盈讯科技侵害著作权及不正当竞争纠纷案的判决结果也体现了这样的导向。在该案件中,被告上海盈讯科技公司未经授权许可将腾讯机器人Dreamwriter生成的文章发布在其运营的“网贷之家”平台上并向公众传播。深圳南山区法院审理认为该AIGC作品“外部表达符合文字作品的形式要求,内部表达反映了对实时股票市场相关信息和数据的编排、分析和评价,结构合理,表达清晰,具有一定的独创性”,属于我国著作权法保护的文学作品范畴^[9]。然而,如何区分自然人创作作品、纯AIGC作品和AI协同创作作品;在AI

协同创作作品中,如何判定自然人在AIGC作品生成中的创造性贡献程度等问题都有待厘清。同时,在AIGC作品生成、出版、跨模态转化、传播的流通过程中,其权益相关主体之间的关系将更为复杂,AIGC作品可能成为合作作品、雇作品、职务作品或法人作品,但其界定标准以及具体的保护路径、保护限制都不清晰,将致使AIGC作品的滥用和出版内容保护机制的失调。

AIGC的发展与应用对现行的著作权制度带来了新的挑战,对此,围绕“科学研究—司法实践—制度架构”三位一体地解决AIGC的权利归属、侵权和权益保护问题,以著作权为核心资产的出版业才能有可持续发展未来。

首先,加强AIGC著作权问题的理论与实践研究。AIGC著作权研究是一个跨学科的综合议题,出版需采用多元化的研究视角凝练AIGC著作权的关键科学问题,充分发掘现有理论的解释张力,提出科学的理论优化思路和实践策略方案。在研究方法上,可侧重案例研究法和比较研究法,对中国、英国、美国、日本等国家的AIGC著作权实践进行总结与分析,如澳大利亚倾向于对计算机生成的作品视为无作者(authorless),而英国的《版权,外观设计和专利法》(Copyright, Designs and Patents Act)则规定,“对于计算机生成的文学、戏剧、音乐或艺术作品,作者应为对创作该作品做出必要安排的人”^[10]。对比探讨不同国家对AIGC作品著作权的权利归属、侵犯以及权益保障实践,厘清其背后的实践原理,能够为我国AIGC著作权的司法实践和制度架构提供理论指引和经验借鉴。

其次,深化AIGC著作权争议的司法实践,增加AIGC指导性案例的效力和数量。当前AIGC著作权争议的司法实践正处于探索阶段,理论界观点纷呈,现有法律和司法又无明确规定,司法实践面临诸多疑难问题。这一方面要求司法机关在处理不同类型、不同著作权争议的AIGC案件时,发挥集体智慧,构建法理、情理、事理逻辑合理的裁判标准。另一方面则需在结案后综合考虑案件的典型性和指导性,撰写和发布AIGC著作权纠纷判定的典型案例和指导性案例。这些案例虽然不具有正式的法源效力,

但对其他类似案件的处理具有重要的参考和指导作用,还能起到一定的教育和宣传作用,是未来AIGC著作权立法的重要支撑。

最后,完善AIGC著作权制度体系。已有著作权基本原理与制度规则无法完全解释或适用于AIGC的著作权问题,完善AIGC著作权制度体系是实现AIGC著作权保护,激励基于AIGC创新经济、文化、科技发展的关键。具体而言,需在我国现行著作权法的基础上,构建AIGC作品的署名规则、著作权归属制度、作品登记制度、合理使用制度、侵权认定制度、侵权责任制度,形成系统完整的AIGC著作权保护制度体系。

2 AIGC的伦理问题

AIGC技术嵌入多层规则后容易产生一个带有偏见的运行体系,这是AIGC技术自身缺陷以及技术的不合理使用所共同导致的。从出版学的视角来看,AIGC的应用催生了技术伦理失范和学术伦理失范的风险。

算法偏见和算法歧视等技术伦理失范问题一直贯穿于AIGC技术创新与应用的始终。一是由于AIGC训练集映射着社会群体的思维偏见,AI算法通过对训练集的歧视模仿而“学会”自动输出具有价值倾向性的内容。曾有人用1.35亿条仇恨言论信息训练GPT-4语言模型,随后该AIGC模型在不到24小时内发布并传播了超过15000条充满暴力内容的帖子,且在最初没有人识别出这些内容来自AI^[5]。事实上,如果不对其输出加以控制,AIGC语言模型很容易被用来生成仇恨言论和垃圾邮件,以及可能隐含在其训练数据中的种族主义、性别歧视和其他有害联想。2022年12月,加州大学伯克利分校的一位计算神经科学家在推特上说,他要求ChatGPT开发一个Python程序,根据一个人的原籍国来判断是否应该受到酷刑。ChatGPT随后回复了代码,并邀请用户输入国家信息:而如果这个国家是朝鲜、叙利亚、伊朗或苏丹,则会输出“这个人应该受到酷刑”(OpenAI随后关闭了这种问题的回答)^[6]。二是AI算法设计师自身存在的价值偏见所导致的,如算法设计师具有明确的经济价值取向或没有及时纠正算法偏见,AIGC便无法体现合理的社会规范和伦理准则。事

实上,微软已经计划在聊天机器人Bing AI输出的AIGC中植入广告,AIGC算法的客观性和公平性难以维系。随着AIGC技术逐渐具有“类人”的智能,隐藏在技术算法“黑箱”背后的AIGC伦理失范问题更难被人为识别和察觉,再经过出版的系统化“封装”和传播,AIGC产生的影响力和破坏力无疑是巨大的。这些新的变化都对传统伦理标准生成了新的挑战,旧的出版道德规范与伦理准则是否适用,新的出版伦理规则如何构建,基于技术向善视角消解AIGC伦理风险等内容都亟待进一步研究和关注。

AIGC的学术伦理问题因ChatGPT的出现而再次受到出版界的热议。《科学》系列期刊更新了其对待AIGC的政策,表明未经编辑部明确许可,不得在期刊上发表由人工智能、机器学习或类似算法工具生成的文本,也不得使用由这些工具生成的图形、图像或图表,违反这一政策即构成科学不端行为^[7]。《自然》杂志主编则表明其不允许将ChatGPT之类的大型语言模型(Large Language Models, LLM)列为署名作者,但允许使用这些工具,并在方法、致谢等相应部分进行具体说明^[8]。这两家国际顶级期刊对AIGC的态度虽有所差异,但都表现出对AIGC学术伦理的担忧。目前,不少研究人员已经开始将ChatGPT以及其他AIGC语言模型作为研究助理,帮助他们组织思维、协助编写代码、总结研究文献、撰写和修改论文等^[9]。但基于AIGC语言模型产生的文本结果看似使人信服,实际却存在大量的事实性错误、知识盲区和虚假信息,严重损害学术成果的科学性与可靠性,也滋生了机器洗稿等学术不端行为,对学术诚信构成重大威胁。然而,从AIGC的应用及其技术发展趋势来看,AIGC甚至能够在辅助设计科学实验、开展同行评议等方面发挥效用,未来的内容创作和出版模式极有可能是协作式、算法式和自主式的,完全禁止AIGC协同学术创作、支撑出版决策或许并不现实。因此,置于出版学面前的不只有AIGC的学术伦理探析,更重要的是讨论AIGC如何在学术伦理范畴内发挥对出版的价值赋能效用。

AIGC的伦理问题决定着AIGC融入出版活动的伦理边界和价值底线。只有匡正AIGC介入出版活动的目的性,以技术向善视角看待并规制AIGC的发

展,才能最大程度地发挥AIGC的正价值效用。

对AIGC技术伦理失范的规制,应从全球治理格局出发,建立AIGC技术伦理规则框架。AIGC技术及其应用都具有极强的跨国流动性,没有一个国家或者政府能够完全独立地实现AIGC技术及其模型的技术伦理监管,其引发的伦理追责和管制问题也通常需要跨国合作来加以应对,全球治理视野由此显得十分重要。结合联合国教科文组织发布的《人工智能伦理问题建议书》(Recommendation on Ethics of AI)和中共中央办公厅、国务院办公厅印发的《关于加强科技伦理治理的意见》,我们认为,出版业可以构建AIGC的共性伦理框架和出版业AIGC伦理框架。AIGC共性伦理框架的构建旨在为全球AIGC技术的开发与应用提供一个集体性的伦理框架,该框架应适用于各国法律体系、道德规范和价值理念,能够为AIGC的技术伦理纠纷、技术开发过程中的技术伦理嵌入、社会个人技术应用过程中的技术伦理原则引导等提供系统性的参考。出版业AIGC伦理框架则旨在集中解决出版场景中的AIGC技术伦理问题,该框架需明晰出版业采纳AIGC技术的原则和规范,为出版业商业化应用和创新转化AIGC技术提供行为准则,以警惕AIGC技术逻辑和资本逻辑对出版文化建构逻辑的“蚕食”。

对AIGC学术伦理失范的规制,应从协同治理理念出发,实现AIGC学术伦理的内外部规约。AIGC引发的学术伦理失范问题涉及科研人员、高校、出版机构、科研项目资助机构、政府部门等多类利益相关主体,仅依靠出版单一主体的力量难以可持续地规制AIGC的学术伦理示范问题。因此,多元利益主体的协同治理理念应贯穿于AIGC学术伦理治理的始终。在外部规约上,出版应协同科研项目资助机构、政府部门等管理机构探索构建AIGC协同学术创作或AIGC参与学术出版流程的学术伦理规则、学术诚信体系和出版伦理规范,从标准、政策、法律法规等维度完善AIGC学术伦理的约束、问责和反馈机制。在内部规训上,出版则需协同高校、学协会等利益主体开展AIGC学术伦理的宣传、教育、培训工作,提高和规范科研人员使用AIGC的学术伦理意识和行为,强化期刊编辑、审稿专家对科学成果的学术把关能

力,形成AIGC学术伦理内化于心、外化于行的科研活动、出版活动。

3 AIGC的文化安全问题

出版长期以来承担着发挥意识形态价值、教育价值、文化价值和科学价值的基本功能^[10],对维护以国家和社会主流意识形态为核心的国家文化安全具有重要作用。而AIGC作为一种完全开放的内容生成范式,对我国的文化安全造成了诸多外在威胁和内在挑战,也对出版的意识形态把控、文化价值构建功能提出了新的要求。

人工智能已经成为人类现今为止最具开放性的创新,以此为算法支撑的AIGC也能够透过网络空间实现无障碍的信息接触与意识形态输出。目前以西方国家为主导的AIGC技术及其应用则极有可能成为意识形态渗透的强力工具。受训练集中数据分布偏差以及国外算法开发人员价值理念的影响,AIGC可能“天然的”拥有意识形态立场。从已有的AIGC应用来看,其在内容互动中更倾向于输出西方的主流意识观点乃至虚假意识形态信息,而对我国优秀传统文化、革命文化和社会主义先进文化又出现认知匮乏甚至认识错误的现象,无疑将影响不同国家用户对中西方文化的态度和价值倾向。因此,在这种新形态的技术霸权以及新形式的意识形态渗透下,我国文化安全面临着外来文化侵入、本土文化价值观念偏离等文化安全威胁,如何捍卫我国的文化安全、构筑文化保护屏障、实现我国先进文化的向外输出与传播显得日益紧迫。

AIGC对我国文化安全形成的内在挑战主要表现在文化价值观的认同危机和出版意识形态的把握危机上。AIGC“温水煮青蛙”“和平演变”式的意识形态渗入使我国民众在不知不觉中接受了多元文化的冲击,对已有的文化价值理念产生动摇。同时,对AIGC的恶意使用和滥用易产生大量低俗化、粗浅化、媚俗化等不良信息资源,为错误文化价值观的滋生提供了温床,折射出我国文化安全存在的内在隐忧。与此相对应的,则是出版对意识形态的把控危机。面对数量庞杂和质量良莠不齐的AIGC,面对几乎没有准入门槛而可以广泛传播的各类意识形态信息,出版对内容意识形态的把控变得尤为艰难又尤为重要。

在内外文化安全威胁的背景下,我们首先要探索建立“国家主导、多元共治”的文化安全防御和管理机制,形成文化安全保障体系。这既是国家治理体系和治理能力现代化的整体要求,也是解决AIGC文化安全问题的内在需求。在文化安全防御机制方面,可开发一套适用于中国文化情境的文化安全指标体系和监测程序,对AIGC平台及其内容信息进行动态监测、评测和风险趋势把控,建立AIGC文化风险评估、审查、预警、处理一体化的防御机制。在文化安全管理机制方面,则要建立以政府文化安全管理部门为主导,AIGC市场相互牵制、AIGC平台自治管理、社会公众参与监督的管理网络,对AIGC中涉及意识形态安全、违背社会公德、激化民族矛盾和宗教冲突、侵犯个人权利等文化安全风险进行分层、分类管理。

其次,开发具有竞争力的AIGC语言模型,构筑主流意识形态阵地。AIGC的文化安全风险一定程度上是AIGC的技术霸权所导致的,因此有必要自主开发具有国际竞争力的AIGC语言模型,构筑主流意识形态阵地。从AIGC的技术开发逻辑来看,该模型既需要在合理合法的前提下,将大量蕴含中华优秀传统文化、革命文化和社会主义先进文化的内容文本纳入训练数据集进行“学习”,“培养”AIGC语言模型的中国立场和中国理念。同时,也要以大国的开放包容心态,给予其他优秀文化意识形态内容的生存空间,促使AIGC成为推动本土文化和外来文化友好交流互鉴的载体。

此外,还需加快推进文化产业和文化事业的繁荣发展,发挥其维护国家文化安全的基石作用。唯有“强起来”的文化产业和文化事业,才能为社会源源不断地输送高质量且丰富多样的精神文化食粮^[11],实现中华文化的自觉、自信和自强。为此,包括出版在内的文化产业和文化事业需将AIGC数智赋能、多维感知、认知交互、多模态融合的技术特点融入文化内容生产、流程运作和服务升级上,通过提高自身的生产力和文化服务力,增强中华文化的凝聚力和传播影响力。同时,文化产业和文化事业还要提升对AIGC技术的控制能力,强化对AIGC及相关著作的内容过滤能力和风险识别能力,形成文化安全保障底线,以

正确的价值导向实现AIGC不良信息内容的净化、优化与引导,推动高质量AIGC的创新生产与传播。

4 结语

AIGC技术及其应用无疑将引导出版走向一场深刻的变革,其中可能衍生的著作权、伦理以及文化安全风险是出版不容忽视和回避的关键问题,亟须我们加以深入探讨与分析。应当说,目前AIGC的发展仍处于起步阶段,AIGC背后的人工智能计算科学正在以我们难以想象的速度学习和汲取现实世界的数据、信息和知识,并实现算法的创新更迭,而出版则处于这种变革的起点。出版只有谨慎地对待AIGC的发展与应用,不断追问、揭示AIGC对出版带来的负向价值及其风险消解路径,才能最大程度地焕发AIGC赋能出版创新发展的生机与活力。

参考文献:

- [1]van Dis E A M, Bollen J, Zuidema W, et al. ChatGPT: Five priorities for research[J]. Nature, 2023, 614(7947): 224-226.
- [2]刘云.论人工智能的法律人格制度需求与多层应对[J].东方法学,2021(1):61-73.
- [3]最高人民法院.2020年度人民法院十大案件[EB/OL]. [2023-03-09].<https://baijiahao.baidu.com/sqid=1688375184809561932&wfr=spider&for=pc>.
- [4]McCutcheon J, The vanishing author in computer-generated works: A critical analysis of recent Australian Case Law[J]. Melbourne University Law Review, 2013, 36(3): 915-969.
- [5]邵文.“最邪恶”AI? 由一亿多条仇恨言论喂养,很难与人类区分[EB/OL]. [2023-03-09].https://www.thepaper.cn/news-Details_forward_18501825.
- [6]Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science[J]. Nature, 2023, 614(7947): 214-216.
- [7]Science. Science Journals: Editorial Policies[EB/OL]. [2023-03-09].<https://www.sciencemag.org/authors/science-journals-editorial-policies>.
- [8]Nature. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use[EB/OL]. [2023-03-09].<http://www.nature.com/articles/d41586-023-00191-1>.
- [9]Hutson M. Could AI help you to write your next paper[J]. Nature, 2022, 611 (7934): 192-193.
- [10]方卿,徐丽芳,许洁,等.出版价值引导研究[M].北京:商务印书馆,2018:19-23.
- [11]吴腾飞.新时代维护国家文化安全的三个维度[J].延边大学学报(社会科学版),2020,53(2):85-92+143.