

# AIGC技术赋能学术期刊数据出版的应用研究与思考

◆ 刘 娟 / 上海图书馆《图书馆杂志》编辑部

**摘 要** 文章梳理《图书馆杂志》数据出版实践,分析 AIGC 技术赋能下数据出版中的数据收集、数据检索、数据推荐、数据分析、数据论文写作、数据评审、数据编辑团队建设等具体应用,以期为推动学术期刊数据出版工作的高质量开展提供启示。AIGC 技术给学术期刊数据出版注入了活力,但也带来了一些挑战。学术期刊出版编辑人员应主动学习、开阔视野,积极应对这些挑战,不断探索 AIGC 技术在学术期刊出版中的具体应用价值,助力期刊出版的智慧化转型。

**关键词** AIGC 智慧出版 数据出版 期刊融合发展

2022 年 11 月,美国 OpenAI 公司发布的大型语言模型(Large Language Models, LLM) ChatGPT (Chat Generative Pre-trained Transformer) 上线仅 2 个月,活跃用户突破 1 个亿,以摧枯拉朽的势头席卷整个科技圈。2024 年 2 月,OpenAI 公司打造的 Sora 横空出世,意味着人工智能自身的理解能力与内容表达能力实现了新的飞跃。ChatGPT、Sora 等的爆发和迅速迭代,带来了文本、图像、视频等人工智能生成内容(Artificial Intelligence Generated Content, AIGC)的根本性变革,还激发了基于大语言模型的知识搜索、汇集、归纳、提炼、生成等知识生产力工具和智能副驾(Copilot)的日益普及。

AIGC 技术以其超大的知识储备、超强的算力,一经出现就对人类知识密集型活动产生巨大冲击。学术出版是与数据和信息紧密相连的知识密集型产业,在科学研究领域,文献检索、数据分析与查验、专门工具调用、文献翻译、文

献综述等曾经的“高素质高门槛”知识服务迅速被人工智能工具攻克,而且大语言模型已被用于提取科学论文中的数据,或者直接用于采集生成数据而取代人类数据专家,或者进行论文评审。学术出版界要充分利用人工智能、大数据、云计算、“数据要素×”等技术与方法来提升学术服务能力。

另一方面,在《关于深化改革 培育世界一流科技期刊的意见》《关于推动学术期刊繁荣发展的意见》等文件的指导下,学术期刊出版在时代变迁中迎来了转型发展的全新机遇,学术期刊进入全面融合发展阶段。在专业学科垂直领域,期刊集群积累了丰富的高价值学术资源,聚集了大量的专业用户群体,生产了许多优质的数据内容,为人工智能应用提供了肥沃土壤,人工智能被大规模地应用到学术期刊出版已经成为不可逆转的必然趋势。

## 一、科学数据与科学数据出版

科学数据是指通过基础研究、应用研究、试验开发等产生的数据,以及通过观测监测、考察调查、检验检测等方式取得并运用于科学研究活动的原始数据及其衍生数据。<sup>[1]</sup>科学数据已经成为科学研究的核心要素和国家大数据战略的基石,是推动科学创新、技术预见与经济发展的新引擎。

近年来,信息技术的变革与创新对出版业的影响愈发显著,出版业的形态不断变化,科学数据出版这一重要概念和发展模式备受学界关注。科学数据出版是指以数据作为生

产要素,把文字、图片、音频、视频等都当作数据的一种表现形式,围绕着数据的挖掘、采集、标引、存储、计算开展出版工作,通过数据模型的建构,最终上升到数据应用和数据服务的层面。<sup>[2]</sup>

科学数据出版是提高科学数据质量、规范科学数据使用、提高研究过程透明度、推动科学数据长期保存的有效方式之一。科学数据出版是技术进步的结果,也是社会发展的需要。所以学术期刊对于科学数据资源的合理收集、存储、筛选、加工、出版,形成良好的数据要素供应链,并在高质量融合发展过程中不断完善数据资源的生产分配与流通共享机制显得极其重要。在二十大报告提出要深入实施科教兴国战略和创新驱动的背景下,通过科学数据出版推动科技创新和科学研究发展,具有更加显著的时代意义。

本研究将以《图书馆杂志》科学数据出版为例,深入探讨 AIGC 技术融入学术期刊数据出版的应用潜力以及可能伴随的风险,并提出相应的风险管理策略,为 AIGC 技术赋能的学术期刊数据出版未来建设提供理论和实践支持。

## 二、《图书馆杂志》数据出版实践

### (一)概况

在以计算机、互联网、大数据等为技术支撑的信息社会里,需要处理的诸如多源异构文本、图像、音频、视频等信息日益复杂化、多元化,对数据管理能力的要求越来越高,对高效率信息处理技术手段的需求越来越强烈,

因而亟须智能信息处理。图书馆在智能信息处理方面具有特别的优势,《图书馆杂志》是由上海市图书馆学会与上海图书馆联合主办的全国中文核心期刊、CSSCI 来源期刊、中国人文社会科学核心期刊、中国优秀图书馆学期刊,在探索期刊创新融合发展方面一直走在行业的前列。<sup>[3]</sup>《图书馆杂志》建有成熟的网络投审稿系统,又于 2016 年开始探索数据出版,自主研发了拥有自主知识产权的《图书馆杂志》数据出版管理平台,将数据出版作为期刊的重要发展方向之一。平台定位为可公开访问的图书情报类共享数据库,所有数据均可在框架内免费下载,且每条数据均拥有永久的 DOI (Digital Object Unique Identifier, 数字对象唯一标识符)。

## (二)《图书馆杂志》数据出版成效与不足

《图书馆杂志》的数据出版主要有富媒体数据出版、支撑论文研究结论的数据出版、数据论文出版、独立数据出版四种类型,能够满足科研人员的不同数据出版要求。

《图书馆杂志》的数据出版与传统论文出版一样,具有上传数据、评审数据、发布数据、引用数据等规范的出版流程。

《图书馆杂志》数据出版又与单纯的数据仓储不同,具备同行评审、编辑加工、公开发布、查询引用等多项功能,还支持数据长久保存、数据分级管理、数据权限管理、数据版本管理、数据浏览与导航、数据检索、数据分析、数据精选、数据素养培训等多项数据服务功能。

《图书馆杂志》科学数据出版业务经过几年的探索实践,取得了良好成效,在数据出版

方面技术领先、类型多样,出版流程规范、服务功能齐全,逐步与国际数据出版接轨。另一方面,《图书馆杂志》数据出版尚处于探索阶段,数据量不足,数据论文数量占比较低,数据评审机制有待规范,数据出版团队建设有待加强。AIGC 技术的发展,给《图书馆杂志》的数据出版带来了进一步发展的良好时机,将 AIGC 技术融入数据出版,对于推动其创新和进步具有多维价值。

## 三、AIGC 技术赋能的数据出版应用集成

从应用价值的角度来看,将 AIGC 技术融入数据出版,可以显著提升数据出版的内容建设和质量管理,优化用户的服务体验,实现审稿人和数据编辑的数字化赋能,从而推动数据出版向更高层次发展。

### (一)提升数据体量 拓展数据资源供给

《图书馆杂志》多管齐下的数据政策使得平台数据的质和量有了一定的保障,平台体量初具规模。但就目前的增长速度,则难以满足高质量发展的要求。

AIGC 技术主要由算法、算力和数据构成。数据作为人工智能时代的新型生产要素,是网络化、智能化与数字化之基。大模型真正的价值是产业侧应用的落地,行业大模型不仅需要通用的数据库,还需要针对不同行业、不同场景的专业数据库。无论是文生文、文生图大模型,还是文生视频大模型,其应用的前提都离不开数据,离不开数据库的建设。

AIGC 技术凭借其在大规模、多模态数据的收集和生成方面的能力,为数据出版的

数据资源建设注入了新的活力。在数据的结构和量级方面,人工智能语言大模型可以通过分析和抓取网络大数据,将网络上的开放获取资源完整地整合到《图书馆杂志》的数据出版管理平台中,实现资源的聚合,丰富数据内容,并完善数据资源体系。在促进数据资源融合方面,人工智能语言大模型能够通过强化学习,建立知识链接、知识结构和知识框架,实现对互联网资源与期刊自身数据资源的深度融合。这有助于构建一个一站式的资源获取平台,增强用户与数据出版之间的互动以及资源的利用效率。

## (二) 实现跨模态检索,精准匹配用户需求

《图书馆杂志》数据出版管理平台所有用户不用注册均可检索平台中所有已发布的数据资源,也可以在单个数据集中检索数据资源。已发布数据集中的受限制数据会出现在检索结果中,用“锁”图标标注;未发布的数据集中的数据资源不在检索范围内。支持基本检索与高级检索,提供基本检索框输入关键词,高级检索能够对特定字段进行检索。

在 AIGC 技术加持下,平台用户可以更准确高效地检索跨类型信息资源。例如,利用多模态技术,基于用户需求及用户检索/问答历史等,自动关联多模态信息资源,进行交互检索,结合自动化内容生成,自动生成图像标题、视频字幕等,进一步生成多模态内容融合的推荐,从而更好地匹配用户需求和满足用户偏好。

AIGC 技术可以提供基于语义理解的搜索功能,帮助用户更准确地找到所需信息。与传统的关键词搜索相比,语义搜索通过对用

户提问的语义和上下文理解分析,能够更好地捕捉用户的查询意图,从而提供更为精准的搜索结果。可以在智能问答、自然语言搜索和智能提示等场景下应用,与搜索引擎技术的结合将大力提升搜索服务的效率和质量。

## (三) 优化数据推荐,满足用户潜在需求

《图书馆杂志》数据出版管理平台可以基于数据下载热度、数据搜索热度、自定义数据精选等,将优质数据通过图片滚动等方式进行展示,用户通过点击图片可以直接进入该数据在数据平台的位置。

基于 AIGC 类技术的智能推荐系统可以根据用户的阅读历史、浏览历史、阅读收藏和阅读需求等记录内容,分析用户数据,识别用户的兴趣、行为模式和潜在需求,为用户推荐合适的论文、数据、数据论文和其他资源,提供更精准前瞻性的学术服务。还可以根据用户的发文类别、研究方向、合作研究中的作者贡献等相关数据分析,精准提供更加个性化的论文推荐和服务。

通过应用信息检索和信息过滤的方法,了解当前学术研究的热点和前沿问题,还可以推测出可能的作者和读者群体,基于内容的推荐算法、协同过滤推荐算法和基于知识的推荐算法,找到可能的目标用户,然后通过差别化服务,满足用户的个性化数据需求。<sup>[4]</sup>

## (四) 智能分析数据,建立可视化学术研究体系

《图书馆杂志》数据出版管理平台提供一种有效的基于 R 语言的计算统计工具,并提供广泛的统计分析模型。支持在线分析工具的集成,支持对数据多层次、多角度可视化分析。

出版业依托 AIGC 技术,融入或搭建出版数据共享平台,可以更大程度发挥 AIGC 的出版作用。<sup>[5]</sup>融合出版平台的竞争优势取决于对大数据的获取能力和利用效率。<sup>[6]</sup>《图书馆杂志》数据出版可以通过应用 AIGC 技术,利用强化学习算法进行深入的文本与数据挖掘和进行聚类整合。文本与数据挖掘(Text and Data Mining,TMD)是对科学数据进行深度利用,从而发现知识关联的新型研究技术与方法,在当下已经成为知识发现与数据分析技术的研究热点,是专业性极高的数据分析流程,是从“海量数据中提取有价值信息的工具”,主要包括文本与数据资源获取阶段、样本处理阶段和结果表达阶段。

#### (五)模板化数据论文写作,创新出版服务模式

《图书馆杂志》数据论文影响力大,但出版的数据论文数量占比不高,数据出版与学术期刊融合发展的深度和广度还有待进一步提高。

数据论文结构包括数据集基本信息简介、概述/背景、数据采集和处理方法、数据字典和数据样本、数据质量控制、数据价值、数据使用方法和建议等,不同于传统学术论文主流的 IMRaD(Introduction 前言,Methods 研究方法,Results 研究结果以及Discussion 讨论)逻辑写作结构,研究人员要完全熟悉数据论文写作,需要投入大量的时间及精力。

在生成式写作技术路线下,可以设置数据论文写作模板,实现从无到有生成内容,减少研究人员论文写作的时间和成本。接入 AIGC 技术的数据论文写作服务可以为用户

提供辅助写作与编辑服务。根据用户提供的关键词推荐下一个可能的词汇,也可以检查和识别文本中的语法和拼写错误并提供修改方案,还可以根据用户提供的关键词和主题,自动生成连贯、清晰、逻辑的内容,这有助于提高用户的写作效率和质量,为用户带来极大的便利。<sup>[7]</sup>

#### (六)规范数据评审,提高数据评审效率

数据评审是数据质量控制的重要方法,同行评议审查数据的标准也与传统学术论文存在差异,《图书馆杂志》数据以及数据论文的评审标准还不够细化,要探索规范、多元的评审标准,从而规范数据论文的审稿流程,确保数据集的可靠性。

数据的多源异构特性以及多模态,导致数据的评审异常复杂,对审稿人的要求很高。在 AIGC 知识生产体系下,将会更加依赖学术期刊对知识生产的创新性、严肃性和可靠性的把关。

从技术方面看,利用 AIGC 技术进行辅助审稿可以自动分派稿件,简化审稿流程,有效提高评审效率。还可以帮助评审专家从数据内容的完整性、科学性、可信度、可重用性等方面提出准确、有效的具有个性化的审稿意见,为作者提供有益的反馈。

#### (七)赋能数据编辑,组建数据出版团队

《图书馆杂志》设立了数据编辑岗位,对平台进行日常维护,并对投稿数据从形式上进行审核,对数据的时效性、客观性、真实性等进行初步评估。平台维护与数据初审工作量大,技术要求高,但编辑部的数据编辑人员不足,不利于数据出版的进一步发展。



在数据出版的环境中,编辑是提升服务质量和效率的关键因素。随着 AIGC 技术的集成,数据出版的环境将更加智能化,并对编辑进行数字化赋能。在这种技术的支持下,数据编辑人员结构将经历显著变化,逐步演变成由“数据编辑”“外审专家”以及“虚拟数字人”等组成的数据出版团队,并通过他们的协同作用,不断优化数据出版服务系统。

#### 四、失范风险与防控

AIGC 技术可以为数据出版注入新的活力,可融合图像、音频、视频等各类信息数据,并整合到一个统一的语义空间中,造就了 AIGC 技术在图像、音频、视频、多模态问答等方面的优势,可以更好地理解和描述复杂现实世界的信息。这一技术的成功应用对数据处理的分析、提取、生成和决策带来巨大的效率提升,从而带来智能数据处理的颠覆性变革。


我们在积极拥抱 AIGC 技术带来的革命浪潮时,也要清醒认识到此类技术的不完美。它还存在一定的局限性:

AIGC 生成数据的真实性和准确性存在不确定性,可能受到指令攻击和提示注入的影响,导致生成数据的不可靠。还存在捏造参考文献的情况,导致数据来源存疑。对 AIGC 技术参与下数据生产合规化问题的认识尚不统一,生成数据的版权归属仍存在争议。训练数据中包含大量的个人信息和隐私,如何保护这些数据的安全和隐私,规避数据泄露和滥用以及防止虚假数据、避免抑制创

新等成为 AIGC 应用中不可避免的问题。另外还有偏见和歧视导致的学术研究的客观性问题,算法、技术成熟度问题,成本问题等。

面对 AIGC 技术的急速发展与广泛应用,需要各方共同努力,加强监管和规范,在推动技术进步和应用的同时,积极应对相应的潜在风险,以实现更加安全、可靠的发展。在实际应用中,要结合 AIGC 技术特征与数据出版的创新发展,强化数据编辑素养和技能,加强人工监控,对内容进行审查,过滤虚假数据。厘清生成数据的著作权归属,加强权利保护。树立算力意识,强化算法监管,以主流价值驾驭算法程序。研制 AIGC 环境下数据出版标准,提升数据治理效能。将失范风险置予可控之范围内,使 AIGC 技术更好地服务于数据出版的可持续发展,共同营造一个健康有序的人机共存环境。

数据出版的核心是数据服务。AIGC 技术赋能的学术期刊数据出版将大幅提高服务质量和效率,让用户获得更佳的经验。但同时,也必须正视一系列挑战。

数字化时代,AIGC 作为一种新质生产力,它具有提高经济发展效率、优化升级经济结构、增强经济韧性等优势,充分发挥科技创新的主导作用,以科技创新推动产业创新。在数据作为新生产要素的基础上,AIGC 应用于数据出版的“出版+人工智能”的前景广阔,未来出版新模式与新形态还将不断涌现。

#### 注释

<sup>[1]</sup> 姜恩波,裴玉香.科学文献与科学数据的融合方法与

# 时尚期刊多元化发展及对传统图书出版的启发

◆ 朱凌云 / 上海译文出版社

**摘要** 文章在对“时尚期刊”的概念、特点及其发展历程做梳理的基础上,重点探讨了时尚期刊在新媒体时代的市场地位和影响力,即纸质期刊的权威性无可置疑、时尚期刊的品牌本身具有正向溢出效应、时尚期刊的读者群体日益变化。随后,文章分析了上海译文出版社旗下时尚期刊矩阵的多元化发展路径,即守住纸质期刊的品牌根基、构建全媒体传播生态圈、重视私域用户群的打造和维护、尝试产品流量变现的各种方式。最后,文章阐述了时尚期刊多元化发展经验对传统大众图书出版的启发。

**关键词** 时尚期刊品牌 基于用户群流量变现 大众图书出版

1994年,中国雅虎正式登录中国互联网,成为我国第一个门户网站。彼时,大概所有从事时尚期刊的媒体人没有想到,互联网会对纸质媒体产生革命性影响。随之产生的期刊生存方式的改变——读者的重新定位、信息的传播方式、双向互动模式、市场信息的反馈、盈利模式等,都颠覆性改写了纸

实例研究[J]. 知识管理论坛,2019,04(02):69-79.

[2] 张莉婧,张新新.论5G环境下的数据出版[J].出版广角,2019(23):6-9,63.

[3] 刘娴.学术期刊科学数据出版实践研究——以《图书馆杂志》数据出版管理平台为例[J].中国科技期刊研究,2022,33(08):1081-1087.

[4] 杨志辉.学术期刊数字化出版到智慧出版的变革[J].

编辑之友,2019(01):36-41.

[5] 谢泽杭,李武.从赋能到融合:生成式AI出版的价值、困境与发展图景[J].编辑学刊,2023(06):13-19.

[6] 赵宏源.融合出版的平台化路径[J].编辑学刊,2024(01):24-29.

[7] 刘喜球,杨亚非.AIGC技术赋能智慧图书馆:应用、风险及其策略[J].图书馆研究,2023,53(06):19-28.