国内出版领域人工智能研究主题挖掘与发展趋势

薛伟莲，太伊楠

（辽宁师范大学管理学院，大连 116082）

摘要：人工智能（AI）技术的飞速发展令出版领域各个方面得以重构。采用BERTopic动态主题建模方法，对中国知网（CNKI）831篇出版领域人工智能研究相关文献进行主题挖掘并分析发展趋势。识别出包括AI辅助出版流程重构、生成式AI伦理风险与版权治理、AI赋能文化传播与产业融合、学术期刊的AI运用、教育出版数智化与人才培养、生成式AI在出版领域的多模态生产、出版业转型与生态协同7个核心研究方向。未来研究发展趋势将在政策引领下，聚焦于伦理治理与文化融合、教育出版与多模态技术协同创新、出版流程重构与生态体系建设等重点方向。

关键词：出版；人工智能（AI）；BERTopic；主题挖掘；发展趋势

中图分类号：G231；G353.1 **文献标志码：**A **文章编号：**1671－1807(2026)05－0159－09

出版是信息传播、知识生产与国家文化软实力建设的关键环节，因此出版领域对技术革新保持着高度的敏感性与探索精神。近年来，在国家战略与产业政策的强力驱动下，出版业对于人工智能（AI）的应用与探索进入了加速期。《出版业“十四五”时期发展规划》明确提出要有效整合各种资源要素，创新出版业态、传播方式和运营模式，推进出版产业数字化和数字产业化，大力提升行业数字化数据化智能化水平[1]；2025年3月，国家新闻出版署发布的《关于组织实施2025年度出版融合发展工程的通知》进一步强调要推动出版深度融合发展，深度应用大数据、云计算、区块链等技术研发新产品，紧跟人工智能发展趋势，积极探索生成式人工智能在出版融合发展中的创新应用和落地场景[2]。这些顶层设计为人工智能与出版业的深度融合指明了方向，注入了强劲动力。

人工智能技术的介入，引发了出版业深刻的结构性变革与业态创新，推动其超越传统的“数字化”阶段，迈向“数智化”与“智慧化”的新纪元。2023年，对话生成式预训练转换器（ChatGPT）的横空出世大幅促进了出版业对生成式人工智能的思考，这不仅体现在对出版流程的效率提升上，更在于催生出全新的出版形态与服务模式。基于此，国内学术界围绕“人工智能＋出版”的研究也呈现出极为活跃的态势。期刊文献作为科研成果的核心载体与知识传播共享的重要渠道，研究其主题的分布、演变与强度，是洞察该领域发展脉络、识别研究热点与预测未来趋势的关键窗口。研究运用深度学习主题建模技术（BERTopic）主题建模法，对国内现有的出版领域人工智能相关的期刊文献进行分析，从而实现对国内出版领域人工智能研究进行多维度、多层次、动态化的主题挖掘与趋势洞察。

1 相关研究述评

通过对中国知网（CNKI）相关文献的系统梳理可知，国内出版领域的人工智能研究已形成较为丰富的学术成果，主要围绕理论探讨、场景建构、实践案例与问题反思4个层面展开，初步构建起“人工智能＋出版”的研究框架。

在理论探讨层面，早期研究主要聚焦于人工智能基本概念与技术路径的引介。王晓光和陈孝禹[3]于2011年率先在国内阐释了语义出版的概念与形式，揭示出当时国内出版传媒界在该领域研究的滞后性。汤雪梅[4]则借助国外企业实践，系统分析了人工智能在数字出版中创新应用的可能。张新新和刘华东[5]进一步从群体智能理念出发，论证了大数据如何重构出版流程并赋能多元出版形态，为该领域后续研究奠定了理论基础。随着技术演进，研究逐渐延伸至场景建构层面，覆盖出版流程再造、知识服务、教育出版、学术出版与版权治理等多个维度。夏德元[6]阐释了出版人角色从传统内容策划

收稿日期：2025-09-02

作者简介：薛伟莲（1966—），女，辽宁营口人，博士，教授，研究方向为智能信息处理；太伊楠（1998—），女，辽宁盘锦人，硕士研究生，研究方向为智能信息处理。

者向数据驱动的知识服务者与技术文化融合推动者的转型趋势；任瑞娟等[7]致力于构建“人工智能生成内容（AIGC）＋区块链”融合的内容可信生态，探索其在出版生产与流通中的赋能机制；谭春林和王建平[8]则围绕AIGC在学术出版中的使用边界展开系统分析，提出以透明度声明与人工复核为核心的双重治理路径。在实践案例层面，生成式人工智能的广泛应用推动出版业在内容生产、传播方式与业态融合方面实现深度创新。马瑞洁和王雨婷[9]基于中信出版、中少总社等案例，揭示了出版机构从内容“发现者”向“原创者”转型的生态位前移路径；沈磊和殷依娜[10]从自然语言处理技术切入，分析了其在提升学术翻译精准性与个性化、促进全球学术交流方面的实际效能。此外，《现代纺织技术》期刊借助DALL-E3辅助设计封面，亦形成具有示范性的品牌实践。随着人工智能应用的不断深入，相关问题反思也逐渐成为学界关注焦点[8]。孜里米拉·艾尼瓦尔[11]强调，既有著作权体系在合理使用与授权模式等方面难以适应“5G＋智能”时代的数字出版新业态，亟须构建更具弹性的利益平衡机制；简川和刘皓阳[12]从智能出版平台的前后端治理结构出发，提出涵盖算法过滤、透明度提升与责任规则完善的整体性应对方案；黎姣欣和李玮[13]则聚焦AIGC生成内容引发的伦理争议，倡导构建技术、制度与行业协同的多元治理框架。

综上所述，现有研究已在“人工智能＋出版”的理论建构、场景拓展、实践推进与风险治理方面取得显著进展，然而，关于该领域研究主题的演变脉络、结构关系与未来趋势仍缺乏系统梳理。因此，研究采用BERTopic主题建模方法，对国内相关文献进行主题挖掘与演化分析，以揭示研究热点、结构关联与发展动向，弥补现有研究的不足。

2　研究设计

2.1　数据来源

以CNKI数据库作为数据源，以“篇关摘”为检索字段，以“TKA＝出版 AND（TKA＝AI OR TKA＝人工智能 OR TKA＝“Artificial Intelligence” OR TKA＝AIGC OR TKA＝ChatGPT OR TKA＝深度学习 OR TKA＝机器学习 OR TKA＝自然语言处理 OR TKA＝NLP OR TKA＝机器人 OR TKA＝发现系统 OR TKA＝智能系统 OR TKA＝语义网 OR TKA＝语义出版）”作为检索式。国内学界对于中文社会科学引文索引（CSSCI）来源期刊的权威性与学科代表性有一定的认可度，因此期刊来源仅选择CSSCI。此次检索于2025年6月27日进行，共检索到1 028篇文献。通过阅读文献摘要，将会议记录、专题序以及图书荐读等无关记录删除，并去除重复记录，最终得到831条文献条目。由于国内出版领域对人工智能的研究尚晚，因此未设定检索时间范围，根据筛选出符合研究要求的831条文献可以确定检索年份为2006—2025年。

2.2　研究方法

BERTopic是Grootendorest[14]于2022年提出的一种基于预训练的无监督深度学习模型，采用基于双向编码器表征转换器（bidirectional encoder representations from transformers，BERT）的深度学习预训练模型，通过Sentence-Transformers等嵌入模型和c-TF-IDF算法对句子进行编码和计算，实现了语义层面上的文档主题聚类与主题表示，相较于潜在狄利克雷分布（latent Dirichlet allocation，LDA）主题模型等主题建模技术，BERTopic在NPMI指标上能够表现出更好的主题识别效果。在使用BERTopic模型进行主题挖掘与演化趋势分析过程中，首先，将检索得到的文献数据进行文本预处理，对数据进行清洗、标准化、分词与去停用词等操作。其次，进行主题建模与优化，BERTopic通过UMAP降维与HDBSCAN密度聚类自动识别语义主题，利用c-TF-IDF加权生成主题关键词。最后，结合时间维度，利用动态主题建模来查看领域主题的研究热度以及学者对各研究主题关注度的变化。

3　结果与分析

3.1　研究主题识别

对831篇论文的摘要文本进行主题建模，通过多次实验优化参数，其中119篇摘要被标记为噪声数据，最终识别出41个研究主题（Topic0～Topic40）。主题的排序遵循文档频率分布原则，呈现正相关关系。每个主题均由一组核心特征词所表示，不同特征词的权重不同，权重较高的词语对主题的表述起着更重要的作用[15]。

图1为相关文档数量排名前8主题特征词及权重的可视化结果，通过分析模型输出的主题特征词及其对应的文献信息，可进一步确定各研究主题的具体名称。例如，Topic0主题中，核心特征词包括数据、出版业、智能等，结合该主题的相关文献内容对Topic0进行主题描述，即探讨随着人工智能（artificial intelligence，AI）在出版业中的应用，数字出版、语义出版等不同形式的出版流程发生改变。因

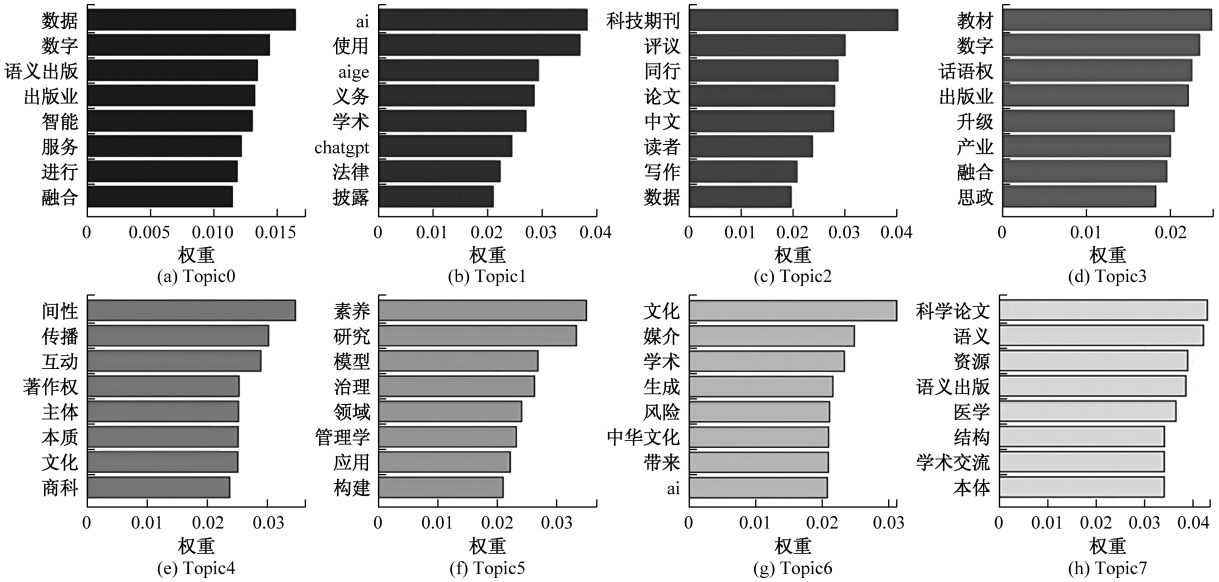


图 1　关联文献数量排名前 8 的研究主题特征词及权重分布

此，将 Topic0 主题下文档的相关信息描述为“AI 辅助出版流程重构”，其他主题的描述亦遵循此方法，确保主题命名既贴合研究内涵，又具有清晰的学术表达。

为进一步揭示出版领域人工智能的研究主题分布特征，对识别出的 41 个主题及其对应的相关文档数量进行统计分析，具体结果见表 1。从整体分布来看，36 个主题的文档数量少于 20 篇，占总主题数的 51.7%，仅有 5 个主题的文档数量超过 20 篇。这一分布特征表明，在出版领域人工智能的研究工作中，不同研究主题的受关注程度存在显著差异，部分主题的研究尚待深入。具体来说，Topic0（数据_数字_语义出版_出版业）主题的研究最为丰富，其文档数量占比高达 36.5%，是出版领域人工智能的研究工作中发文量最高的研究主题。之后是 Topic1（ai_使用_aigc_义务）、Topic2（科技期刊_

表 1　国内出版领域人工智能研究主题识别结果和相关文档数量及占比

主题（核心特征词）	文档量/篇	占比/%	主题（核心特征词）	文档量/篇	占比/%
0（数据_数字_语义出版_出版业）	303	36.5	21（共创_元宇宙_aigc_出版业）	7	0.8
1（ai_使用_aigc_义务）	27	3.2	22（数字_教育出版_世界_产业）	7	0.8
2（科技期刊_评议_同行_论文）	26	3.1	23（保护_作品_人格权_著作）	7	0.8
3（教材_数字_话语权_出版业）	26	3.1	24（翻译_尘埃落定_文学_民族）	7	0.8
4（间性_传播_互动_著作权）	20	2.4	25（信息安全_隐秘性_图像_风险）	7	0.8
5（素养_研究_模型_治理）	19	2.3	26（论文_摘要_模态_ai）	7	0.8
6（文化_媒介_学术_生成）	18	2.2	27（教育_培养_学习_产业）	7	0.8
7（科学论文_语义_资源_语义出版）	16	1.9	28（编辑_协同_知识_人机）	6	0.7
8（科技期刊_评价_论文_评议）	14	1.7	29（意大利_审核_读物_需求）	6	0.7
9（期刊_伦理_微信_水产）	13	1.6	30（应急_职业_智慧_文章）	6	0.7
10（撤稿_使用_ai_学术）	12	1.4	31（古籍_主题_互通_大众化）	6	0.7
11（企业_营销_初创_deepseek）	12	1.4	32（无障碍_模态_模型_残障）	6	0.7
12（出版业_数智化_生态_转型）	12	1.4	33（儿童_绘本_范畴_图书）	6	0.7
13（共生_有声书_人文_ai）	11	1.3	34（文化产业_文化_学术_研究）	5	0.6
14（学术期刊_科技期刊_学术_应用）	11	1.3	35（经济_现代化_出版业_社会主义）	5	0.6
15（智能网_汽车_业务_爱思唯尔）	10	1.2	36（泄密_科技期刊_aigc_年度）	5	0.6
16（语料库_实现_用户_场景）	9	1.1	37（刊群_高校_平台_科技期刊）	5	0.6
17（文章_学术_数字_中国）	9	1.1	38（质检_法律_少儿_主题）	5	0.6
18（大学生_行业_整理_古籍）	8	1.0	39（童书_行业_ar_封面设计）	5	0.6
19（科幻_生态_品牌_知识产权）	8	1.0	40（知识_古籍整理_服务_传播）	5	0.6
20（风险_侵权_生成式人工智能_版权）	8	1.0	—	—	—

评议_同行_论文)和 Topic3(教材_数字_话语权_出版业)这 3 个主题,其文档数量占比分别为 3.2%、3.1%和 3.1%,在相关研究工作中均占据重要地位。

3.2　研究方向识别

图 2 为主题文献的二维可视化分布,不同颜色表示不同的主题,在二维空间的分布反映了主题聚类、研究热点及主题间的相互关系。从密度分布来看,呈现出较为明显的多个高密度文献聚集区域。主题文献主要集中在左侧,Topic0 单独占比很高,呈现出独立于其他主题的状态,而通过局部放大图可以看出 Topic8 和 Topic2、Topic7 等主题展现出较强的关联性。从主题的空间分布来看,揭示了主题之间的潜在关联和融合。

图 3 展现了主题间的相对距离和独立性。圆圈越大表示权重越高,主题的空间分布揭示了主题相似性,距离越近说明在内容上关联性越强[16]。图 4 为基于原始高维散度计算的聚类结果,反映的是主题之间的相似度以及如何聚类。相较于图 3 而言,图 4 的层次聚类不经过降维,更能反映主题在高维空间的相似性。

在以上数据分析结果的基础上,结合余弦相似度聚类的可视化分析,对 41 个主题的相似性与差异性做归纳,将研究主题归纳为 7 个研究方向。表 2 进一步展示了各研究方向所涵盖的主题数量、相关文献数量及其占比情况,为分析出版领域人工智能的主要研究方向提供了客观依据。

3.3　研究方向内容分析

3.3.1　AI 辅助出版流程重构

国内学者对 AI 辅助出版流程重构的研究聚焦于利用人工智能技术对传统出版流程做智能化改造

表 2　出版领域人工智能研究方向的主题数量和文档数量

主题序号	研究方向	主题数量/个	文档数量/篇	文档占比/%
Topic0	AI 辅助出版流程重构	3	328	39.5
Topic1	生成式 AI 伦理风险与版权治理	10	123	14.8
Topic2	AI 赋能文化传播与产业融合	5	70	8.4
Topic3	学术期刊的 AI 运用	5	62	7.5
Topic4	教育出版数智化与人才培养	9	56	6.7
Topic5	生成式 AI 在出版中的多模态生产	5	42	5.1
Topic6	出版业转型与生态协同	4	31	3.7

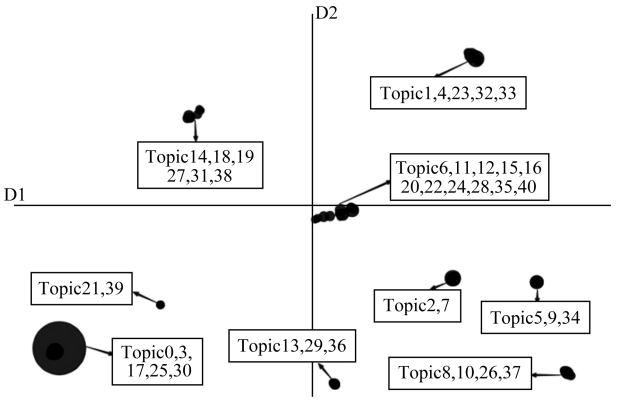


图 3　主题距离可视化

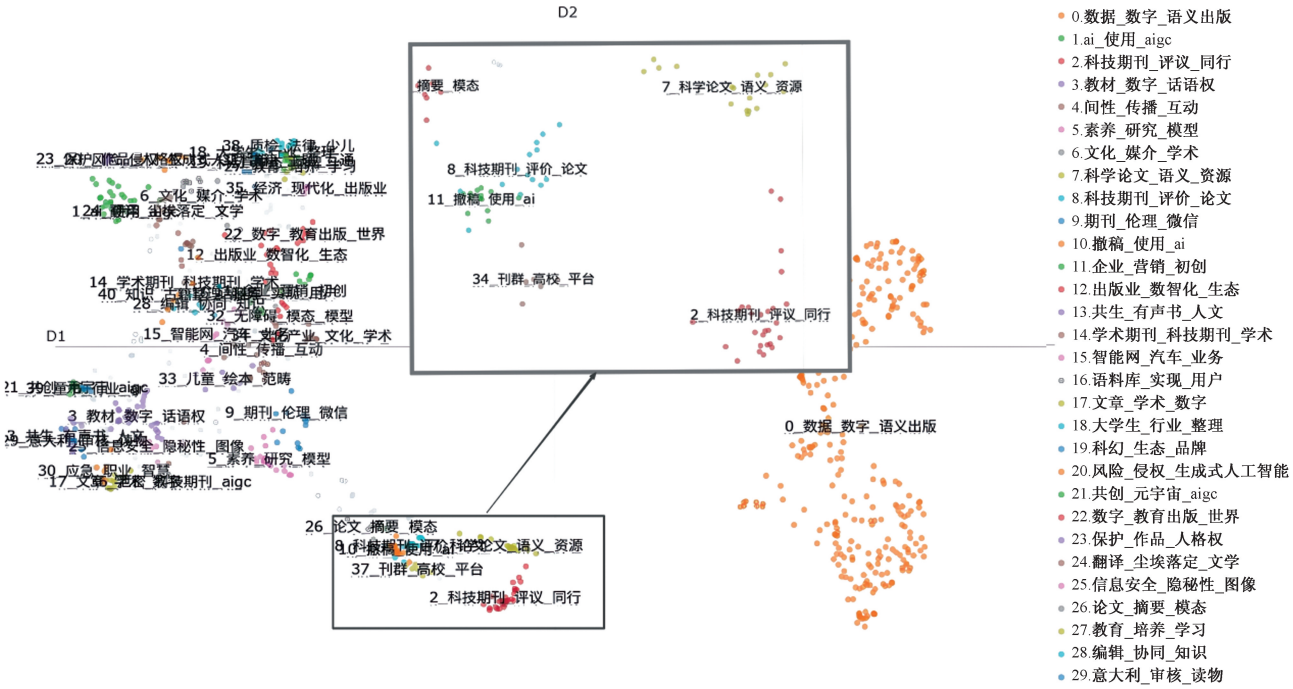


图 2　主题文献的二维可视化分布

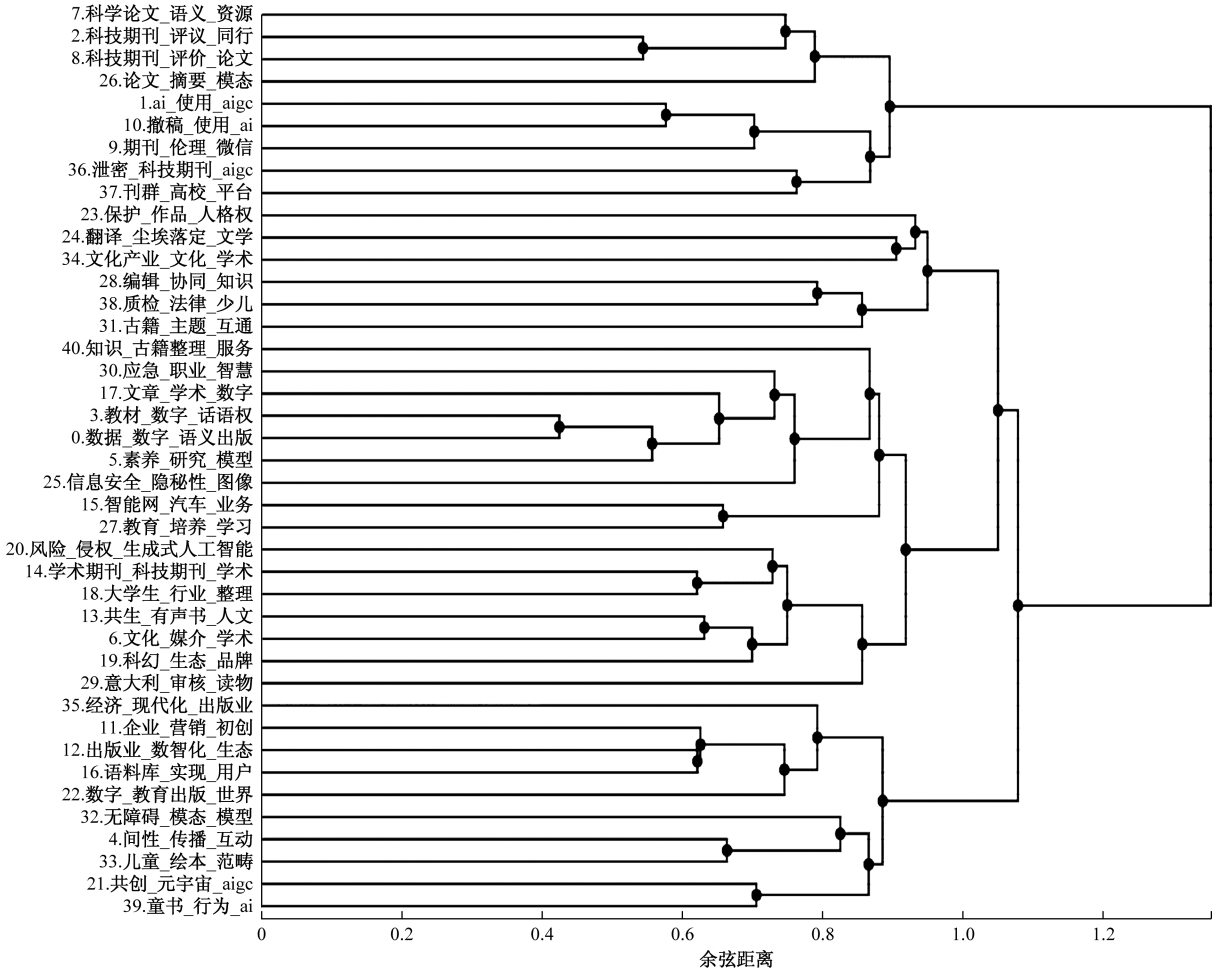


图 4　基于余弦相似度的主题聚类

及效率提升,覆盖选题策划、内容生产、编辑加工、营销传播等核心环节。在传统出版流程中选题主要依靠编辑经验和市场调研,部分出版单位会定期召开选题策划会,但仍存在主观性强、效率低等问题。运用 AI 技术能够较为准确地挖掘选题热点、预测趋势,通过多维度构建知识图谱打造智能化作者推荐机制,降低人工筛选成本以实现作者匹配与资源优化[17],还能结合历史销售数据与用户画像预测选题市场接受度,有效预判市场风险[18]。国内研究多数认为,AI 辅助出版流程重构的本质是“数据赋能”与“人机协同”的双向进化,未来可通过数据挖掘、生成式 AI、算法推荐等技术来实现出版全流程的自动化与个性化。

3.3.2　生成式 AI 伦理风险与版权治理

生成式 AI 的伦理与版权问题始终是领域内研究热点,学者们围绕风险识别和治理路径分别搭建了研究体系。伦理风险研究主要聚焦于 3 个层面:内容真实性争议、学术诚信危机和意识形态偏差。而版权治理困境的相关研究同样聚焦于 3 个方面:AI 生成内容的著作权归属问题模糊、侵权责任界定的争议性和侵权追溯难度大。针对以上所面临的困境,学者们从技术、制度、行业协同 3 个层面提出解决办法。技术层面,采用区块链开展版权存证事务[12],采用数字水印跟踪 AI 生成内容的最初来源;制度层面,建议修订《中华人民共和国著作权法》以厘清 AI 生成物的权利归属问题,制定问责制确立人类作者最终责任[19];行业层面,呼吁建立 AI 内容审核规范,开展“人工复核 + 技术检测”的双重把控机制。

3.3.3　AI 赋能文化传播与产业融合

人工智能技术为文化传播与相关产业的深度融合提供了新动能,国内学者对此的研究主要围绕技术驱动的文化创新、跨文化传播效能提升及产业生态重构 3 大维度展开,初步构建了“技术-内容-场景”协同的学术框架。在文化传播领域,AI 技术正推动传播范式革新。生成式 AI 的应用已

从辅助工具层面，逐渐演进为内容生产与分发的关键驱动力量之一，显著提升了文化内容的创作效率与传播广度。从产业融合视角看，研究表明“平台范式”正在重塑出版产业链生态。数字出版依托算法与数据技术，对内容生产、分发与消费全链条进行系统性整合，构建了“趣缘-地缘”双轨驱动的市场体系[20]。在全球化进程中，既要平衡好嵌入国际平台跟建设自主平台的关系，又要推动把AI版权溯源、数据脱敏等标准加入国际出版协议，以此抵御文化霸权侵害[21]。当前研究认为AI不仅是技术工具，更是文化传播的“关系媒介”。未来需要进一步强化生成式AI的文化解释力，建立跨境数据合规框架，为中华文化深度“走出去”提供学术支撑。

3.3.4 学术期刊的AI运用

学术期刊的AI应用研究方面，主要聚焦于出版流程优化、质量控制及学术伦理治理3大方向。在出版流程方面，AI技术深度渗透出版各个环节，通过智能选刊、学术不端检测、审稿人匹配等功能显著提升稿件处理效率[22]。同时，生成式AI正推动期刊内容的形态革新。随着技术从AIGC向AIGV演进，动态视频摘要技术应运而生，极大地提升了科研成果的可视化传播效能[23]。然而AI在学术质量控制与学术伦理治理亦存在较大挑战，同行评议环节，AI难以替代人工对学术创新性的客观评判[24]，且AIGC可能引发数据伪造、图像失真等学术不端问题。对此，学者们提出“人机协同”的治理路径。在未来，AI技术进一步深化融合，其应用可以实现从单点工具扩展至全流程智能重构，推动科技与学术期刊向“智慧知识服务”转型。

3.3.5 教育出版数智化与人才培养

在教育出版数智化与人才培养的相关研究中，国内学者主要聚焦于两大方向：教育出版的智能化升级和出版教育的体系化重构，体现了“技术赋能”与“人才反哺”的双向互动逻辑。在教育出版领域，研究集中在AIGC技术驱动教材形态、内容生态与服务模式变革。在出版教育领域，相关研究围绕“智能出版人”的能力重塑与培养路径展开。有学者提出“一联、二融、三通”模式：通过校企联动衔接产业需求；推动跨学科资源融合与多元素养培育融合；将AI技术贯通于课程设计、资源供给及教学理念全环节[25]。整体来讲呈现两大趋势：①教育出版正从静态资源供给转向动态智能服务生态，注重人机协同的沉浸式学习体验；②出版教育强调人才培养的“多元能力复合”，通过学科交叉与产教协同重构人才知识体系，以响应出版业数智化进程中的伦理治理、技术应用与业态创新需求。

3.3.6 生成式AI在出版中的多模态生产

国内学者探讨生成式AI驱动出版内容从单模态向多模态演化的相关研究，主要集中在技术融合、场景创新与生态构建3个维度。就技术融合层面而言，利用文本生成、语音合成加文生视频技术协同赋能跨模态内容生产[26]。场景创新侧重于元宇宙与AR/VR驱动的沉浸式出版范式。实证研究表明，多模态编辑器可以把碎片化资源整合成交互式知识单元。生态构建阶段以垂类模型与算力协同为核心。当前研究提出三级技术路径：依托通用大模型强化出版垂直领域、以迁移学习为主导改造垂类大模型和通过全架构构建出版垂类大模型。研究倡导搭建“云边端”算力协同体系，提高多模态内容生产效率，并通过培育首席数据官(CDO)与提示词工程师等复合人才来强化数据治理与技术适配能力[27]。

3.3.7 出版业转型与生态协同

针对出版业转型与生态协同，国内学者强调技术、组织及价值链的深度耦合。研究表明，出版机构正把自身从传统的内容中介转型成原创生产者与生态主导者，凭借开发自有知识产权资源达成生态位前移。在生态协同层面，研究揭示出版业需突破数据孤岛与组织壁垒，建立动态适配机制。此外，生态价值重构体现为两大趋势：①数据资产化，出版企业通过数据确权、入表及开发增值服务，将数据资源转化为新型生产要素[28]；②版权运营体系升级，结合区块链技术构建全链条版权保护机制，并通过跨媒介IP开发延展价值链[29]。未来出版生态的核心竞争力在于“技术嵌入深度”与“生态协同广度”。通过强化垂直领域大模型研发和完善创新联合体机制，推动出版业从封闭生产系统转向开放赋能的生态平台，最终实现“技术-内容-场景”三位一体的高质量发展。

4 研究方向的演化趋势分析

BERTopic的动态主题建模方法能够直观呈现研究领域内各主题的研究热度与学者关注度的变化趋势，便于对国内出版领域人工智能的各研究方向的主题演化趋势进行分析。研究将文献数量视为各研究主题的热度指标，图5展现7个研究方向

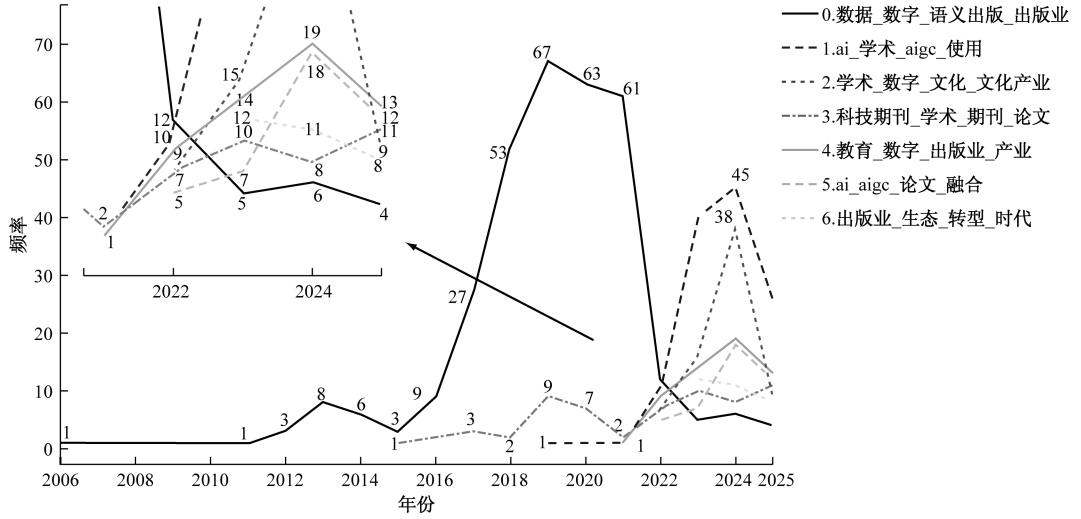


图 5　2006—2025 年国内出版领域人工智能研究主题热度演变趋势

自 2006 年以来的研究热度变化趋势。结合政策发布与技术突破等关键节点,国内出版领域人工智能研究可大致划分为 3 个阶段:技术认知与初步探索期(2006—2018 年)、应用拓展与场景深化期(2019—2021 年)、融合创新与问题反思期(2022—2025 年)。

具体来看,Topic0(AI 辅助出版流程重构)作为基础性研究方向,贯穿整个研究历程,其文献数量在 2015 年后显著增长,2019 年研究数量到达顶峰,表明该阶段 AI 在出版流程中的技术认知逐步深化。2019 年起,伴随《关于促进文化和科技深度融合的指导意见》等国家层面政策的出台,生成式人工智能伦理风险与版权治理(Topic1)等议题开始成为学界关注焦点。进入 2022 年后,以大语言模型为代表的 AIGC 技术实现突破性进展,进一步推动了多模态内容生产(Topic5)、文化传播和产业融合(Topic2)、教育出版数智化(Topic4)等新兴研究方向热度的快速上升。相比之下,学术期刊的 AI 应用(Topic3)与出版业生态转型(Topic6)虽整体发文量有限,但作为支撑行业发展的结构性议题持续存在。

从最新演进态势看,截至 2024 年,除 Topic3 与 Topic6 外,多数研究方向仍保持活跃,文献数量呈总体上升趋势。值得注意的是,Topic0 虽在总文献量中占比最高,但其增长速度已趋于放缓,显示出研究重点正从基础性流程重构,逐步向垂直场景的深入应用与综合治理等方向扩散。基于上述分析,未来国内出版领域人工智能研究预计将呈现系统化、融合化的发展路径。首先,生成式 AI 伦理治理(Topic1)与文化传播和产业融合(Topic2)构成双核主线。国际层面上,欧盟于 2025 年 7 月发布了《AI 实践准则》,这将为中国本土化版权治理体系的建设提供参考,实现技术创新与学术诚信风险相平衡。在文化传播和产业融合方面,出版领域将依托“数据要素×文化旅游”等政策,通过非遗“数字分身”、入境游大模型等场景实现文化资源的生态化重构,推动出版从内容输出转向沉浸式交互服务。其次,教育数智化(Topic4)与多模态内容生产(Topic5)形成协同引擎,学术期刊的 AI 应用(Topic3)有待深度挖掘。教育部于 2025 年 7 月发布的《数字化赋能教师发展行动》将数字素养纳入教师资格考核,促进教育出版向“数据驱动型服务”转型,而多模态技术在文化产业融合中起到了核心作用。例如,广东出版集团的教材数字化专利通过图像识别提升了审核精度,印证技术落地的产业价值。此外,2025 中国科技期刊发展论坛发布了“学术科协”平台,该平台实现了编辑出版业务全流程一体化运行和文献数据本地安全存储,将为期刊提供出版传播和学术诚信监测等服务,为科技工作者提供智能化知识服务。最后,流程重构(Topic0)与生态转型(Topic6)渗透全产业链,构成研究领域的底层支撑。通过图书企业资源计划(enterprise resource planning,ERP)动态优化、人机协同运营以及区块链确权机制等,构建“数据-场景-服务”闭环生态,最终推动出版业从技术应用向范式创新跃迁。

5　结论

基于 BERTopic 模型,对 2006—2025 年中国知

网收录的831篇出版领域人工智能研究文献进行主题挖掘与演化分析，系统揭示了该领域从技术认知探索、应用场景深化到融合治理反思的三阶段演进路径，识别出包括AI辅助出版流程重构、生成式AI伦理风险与版权治理、AI赋能文化传播与产业融合、学术期刊的AI运用、教育出版数智化与人才培养、生成式AI在出版领域的多模态生产、出版业转型与生态协同在内的7大热点主题。未来研究应重点关注以下方向：①加强生成式AI在出版中的伦理规范与版权治理体系构建，推动技术、制度与行业实践的协同发展；②深化垂直领域大模型与出版业务的融合机制，尤其是在学术期刊、教育出版等典型场景中的落地路径与效能评估；③拓展多模态、沉浸式出版形态与用户交互机制，探索AIGC＋VR/AR、数字人等新型内容范式；④关注数据安全、算法可解释性与人机协同编辑等底层支撑问题，构建可信、可解释的智能出版系统；⑤推动出版教育体系与人才能力结构的数智化转型，强化“出版＋技术＋伦理”复合型人才培养。

研究虽初步勾勒出出版领域人工智能研究的热点结构与演变轮廓，但在数据来源的广度与技术迭代的跟进方面仍存在局限。后续研究可进一步拓展语料库范围、纳入国际对比视角，并加强对技术前沿出版应用的跟踪分析，以全面把握智能出版的发展趋势与理论前沿。

参考文献

[1] 出版业“十四五”时期发展规划[J]. 中国出版，2022(3)：8-20.

[2] 国家新闻出版署. 关于组织实施2025年度出版融合发展工程的通知(国新出发〔2025〕3号)[EB/OL]. (2025-03-21)[2025-09-30]. https：//www. nppa. gov. cn/xxfb/tzgs/202503/t20250321_888045. html.

[3] 王晓光，陈孝禹. 语义出版的概念与形式[J]. 出版发行研究，2011(11)：54-58.

[4] 汤雪梅. 人工智能与数字出版的创新应用[J]. 编辑之友，2015(3)：15-18.

[5] 张新新，刘华东. 出版＋人工智能：未来出版的新模式与新形态：以《新一代人工智能发展规划》为视角[J]. 科技与出版，2017(12)：38-43.

[6] 夏德元. 出版业的数智化转型与出版人的职业升维[J]. 编辑学刊，2025(2)：19-26.

[7] 任瑞娟，归亦辰，濮德敏. AIGC与区块链服务网络架构下的新型内容生产研究[J]. 出版发行研究，2024(2)：10-14，22.

[8] 谭春林，王建平. AIGC在学术研究和出版中的使用边界、透明度与伦理[J]. 编辑学报，2024，36(6)：661-666.

[9] 马瑞洁，王雨婷. 知识产权视角下出版机构生态位重塑[J]. 中国出版，2025(6)：22-27.

[10] 沈磊，殷依娜. 基于NLP技术的学术著作翻译策略研究[J]. 上海翻译，2025(3)：56-62.

[11] 孜里米拉·艾尼瓦尔. 聚焦“5G＋智能”时代：数字出版著作权法治理困境及应对[J]. 科技与法律(中英文)，2022(2)：88-97.

[12] 简川，刘皓阳. 智能出版平台版权侵权治理困境及应对机制[J]. 现代传播(中国传媒大学学报)，2025，47(1)：151-160.

[13] 黎姣欣，李玮. AI大模型辅助网文数字出版的伦理争议与治理框架[J]. 出版广角，2025(5)：72-79.

[14] GROOTENDORST M. BERTopic：neural topic modeling with a class-based TF-IDF procedure[EB/OL]. (2022-03-11)[2025-07-14]. https：//arxiv. org/abs/2203. 05794.

[15] 程越欣，杨峰，郭剑明. 基于BERTopic模型的国内政府数据开放研究主题挖掘及内容分析[J]. 图书馆学研究，2025(5)：26-37，65.

[16] 陈江燕，寇龙，段晓聪，等. 基于BERTopic的中国生成式人工智能研究主题挖掘与分析[J]. 计算机时代，2025(6)：39-44.

[17] 邢颐，马萌. 人工智能视域下的学术出版：流程革新、发展困境与优化策略[J]. 中国出版，2024(13)：35-39.

[18] 李强辉. 出版产业视域下的AI赋能与冲击研究[J]. 科技与出版，2024(4)：71-79.

[19] 徐拥军，卢思佳，殷名. 国内外学术出版机构AIGC使用政策研究[J]. 图书馆建设，2025(4)：11-20.

[20] 姬德强，白彦泽. 平台范式：数字出版推动文化走出去的路径选择[J]. 中国编辑，2024(6)：44-49.

[21] 李华君，雷月秋. 智能出版的技术特征、伦理风险与编辑把关[J]. 中国编辑，2024(1)：33-37，48.

[22] 刘冰. 人工智能在科技期刊出版中应用的机遇与挑战[J]. 出版广角，2024(23)：51-56.

[23] 王萌，张加洪，李瑜. 从AIGC到AIGV，科技期刊与人工智能生成技术的融合发展[J]. 编辑学报，2025，37(S1)：137-141.

[24] 黄茜，蒋琳灵，张百灵，等. AI技术在期刊出版中应用边界的研究与探索[J]. 编辑学报，2025，37(S1)：145-150.

[25] 陈丹，李奕佳. 人工智能时代出版教育数智化转型方向与路径[J]. 中国出版，2025(1)：18-24.

[26] 匡文波，姜泽玮. 生成式人工智能推动数字出版智能化发展路径研究[J]. 出版广角，2024(14)：15-21.

[27] 张新新，游恒飞. 从ChatGPT到DeepSeek：多模态大模型与出版业未来：兼论走向“十五五”的出版智能化转型[J/OL]. 中国编辑，1-11[2025-08-30]. https：//link. cnki. net/urlid/11. 4795. g2. 20250409. 1500. 00.

[28] 周青. 出版业数智化的逻辑、路径与对策[J]. 中国编辑，2025(4)：50-57.

[29] 郭全中，佟雨欣，曾婉琪. GAI背景下智能化出版的变革逻辑与实践路径[J]. 出版广角，2025(5)：36-43.

Research Themes and Development Trends of Artificial Intelligence in Domestic Publishing Field

XUE Weilian, TAI Yinan

(College of Management, Liaoning Normal University, Dalian 116082, China)

Abstract: The rapid advancement of artificial intelligence (AI) technology leads to the reconstruction of various aspects of the publishing field. The BERTopic dynamic topic modeling method was employed to analyze 831 AI-related research articles in the publishing domain from the China National Knowledge Infrastructure (CNKI) database, aiming to identify research topics and development trends. Seven core research directions have been identified, including AI-assisted publishing process reconstruction, ethical risks and copyright governance of generative AI, AI-enabled cultural dissemination and industrial integration, AI applications in scientific and academic journals, digital-intelligent transformation and talent cultivation in educational publishing, multimodal production of generative AI in publishing, and publishing industry transformation and ecological collaboration. Future research trends, guided by policy, will center on ethical governance and cultural integration, collaborative innovation in educational publishing and multimodal technologies, as well as publishing process reconstruction and ecological system development, with a particular emphasis on key areas such as intelligent transformation and talent development.

Keywords: publishing; artificial intelligence(AI); BERTopic; topic mining; development trend