

出版人工智能研究：概念、技术、影响与进路^{*}

——一项系统性文献综述

唐嘉骏, 杨海林, 马子寒

南京大学信息管理学院, 210023, 南京

摘要：梳理我国出版人工智能的研究现状，有助于明确出版行业的发展方向，促进出版学科自主知识体系的构建和完善。本研究采用系统性文献综述方法，检索、筛选并分析了522篇相关文献，系统归纳了出版人工智能的核心概念、技术架构、行业影响及应对进路。学界认为智能技术正深度介入出版全流程，尤其是生成式人工智能在内容创作和知识生产中的应用，为出版行业的智能化升级提供了重要支撑。但人工智能也带来伦理和实践问题，生成内容的版权归属和原创性争论尤为突出。出版业应充分发挥在内容资源上的优势，向“知识服务提供商”和“知识把关人”角色转变，探索“出版即服务”模式，努力构建以高质量知识资源为核心的智能出版新生态。

关键词：人工智能；出版业；智能出版；系统性文献综述

文献标识码：A

现代数字技术的广泛应用提高了出版业的数字化水平，推动传统出版向数字出版转型。而在以算法为核心的人工智能时代背景下，出版的未来出现了全新的可能，随着智能技术迁移到文化产业领域，智能出版也成为传统出版业在智能时代转型升级的新方向^[1]。以ChatGPT为代表的生成式人工智能（Generative Artificial Intelligence, GAI）不仅加速了内容生产的效率，也具备了介入人类知识生产格局的能力。智能技术一方面激发了出版业界、学界对迈向深度融合发展未来图景的向往；另一方面也引起了对机器参与知识生产后不良影响的担忧。因此，充分利用人工智能等技术以推动出版的融合发展，是当下出版学研究的题中之义。

近年来出版人工智能研究迅速发展，理论和实践的问题亦随之产生，有必要对相关研究进行回顾整理，为研究者提供研究现状分析，以便系统把握其发展趋势。张新新等^[2]梳理了智能出版研究中的概念和逻辑，探讨人工智能、VR/AR、区块链等技术下出版

发展的进路，呼吁学界对智能出版研究的关注；陈丹等^[3]使用CiteSpace整理了近20年来我国数字出版相关研究，指出应在人工智能等技术背景下开展出版学研究。现有研究大多使用叙述性综述（Narrative Review）或文献计量方法对出版人工智能议题进行综述，而系统性文献综述方法（Systematic Literature Review, SLR）具有明确的原则、问题和步骤，增加了综述的透明性和客观性^[4]。因此，本研究使用该方法对国内出版人工智能研究进行梳理，试图回答下述研究问题：出版人工智能的概念和特征如何？人工智能在我国出版业的相关技术架构与应用实例有哪些？人工智能给出版行业带来了什么样的机遇和挑战？如何化解潜在风险？

1 综述过程与计量分析

1.1 综述方法

系统性文献综述第一步是全面搜集领域相关

^{*} 基金项目：国家自然科学基金面上项目“隐私政策视角下企业数据合规与风险管理研究”（项目编号：72474101）；国家自然科学基金面上项目“基于注意力机制的学术信息动态推荐研究”（项目编号：72074109）。

文献。考虑到人工智能在我国出版研究中的多种表述,本研究在中国知网、万方数据知识服务平台、维普期刊3个数据库中进行检索,检索时间为2024年12月11日。构建的检索表达式为“*出版”AND (“新质生产力” OR “人工智能” OR “AI” OR “大语言模型” OR “算法” OR “ChatGPT” OR “数智化”),限制发表时间为2014—2024年,并通过前向和后向搜索补充,共计3 818篇文献。

表1 文献筛选和纳入标准

标准	依据	说明
刊物等级	刊物收录索引	● 纳入CSCSI和北大核心收录期刊的论文 ○ 排除非CSCSI和北大核心收录期刊的论文
文章类型	标题、摘要	● 纳入经验研究或理论研究论文 ○ 排除综述、卷首语、书评等文章类型
文章相关性	标题、摘要、关键词	● 纳入将出版业或出版学和人工智能关系作为研究对象的论文 ○ 排除仅提及人工智能等技术的论文

之后确定了文献筛选纳入的标准(见表1),依照该标准对命中结果进行初步筛选。为了保证纳入文献的权威性,首先去除重复文章和非核心刊物文章,其次通过标题、摘要和关键词等信息去除非研究类文章。在此基础上,排除仅提及出版业与人工智能技术而非研究二者关系的弱相关性论文。共计522篇论文被纳入系统综述的范围,数量满足系统性文献综述的基本要求。若论文仅将人工智能与出版作为文章的核心议题,则标记为紧密相关,反之则标记为非紧密相关。在此之后,研究者对纳入文献进行了编码分析,具体研究过程见图1。

1.2 文献计量分析

在时间维度上,出版人工智能领域的论文数量呈现出明显的增长趋势(见图2)。2015—2016年仅有2篇论文发表,表明在早期,人工智能在出版领域的应用尚未受到广泛关注;2017—2019年,随着人工智能技术的快速发展及其在出版行业的逐

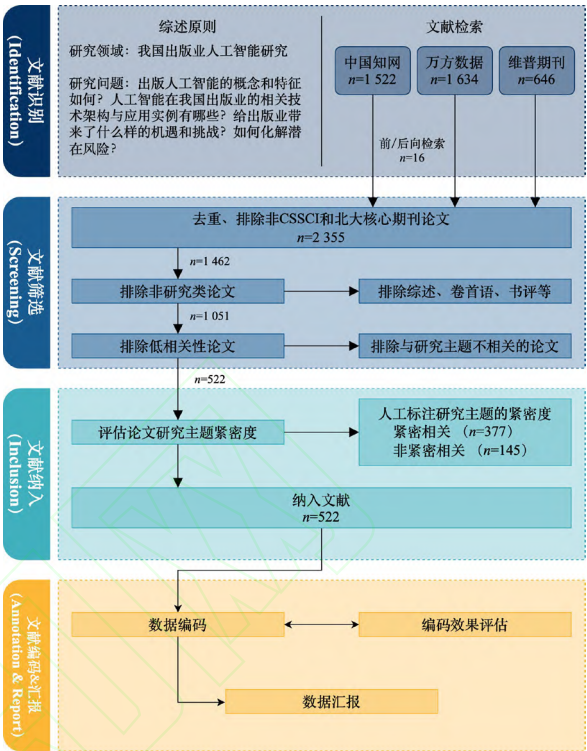


图1 研究步骤

步应用,相关研究论文的数量开始迅速增加;而至2020年,全年共发表了57篇相关选题论文,反映了学术界对这一领域的兴趣显著提升;2023年,GAI技术,尤其是ChatGPT等模型的广泛传播和应用,使出版学界对人工智能的研究热情达到高潮,当年相关论文发表数量激增至121篇,不仅数量上实现了飞跃,也表明了人工智能技术在出版领域应用的深度和广度都在不断扩展。

在期刊维度上,522篇论文分别发表在37种期刊上,《出版广角》发文最多,《科技与出版》紧随其后,《中国出版》《中国编辑》《出版发行研究》和《编辑之友》的发文量均在30篇以上(见图3)。

在作者维度上,纳入的论文涵盖了791位学者的贡献,这些作者隶属于524个不同的机构,其中高等院校(226个)、出版机构(191个)和研究中心(78个)构成了研究的主体,占比高达94.47%(见图4),反映出高等教育和专业出版机构在推动人工智能与出版业融合方面的主导作用。值得注意的是,来自出版行业的编辑和其他从业者也积极参与了学术交流,贡献了丰富的业界视角。

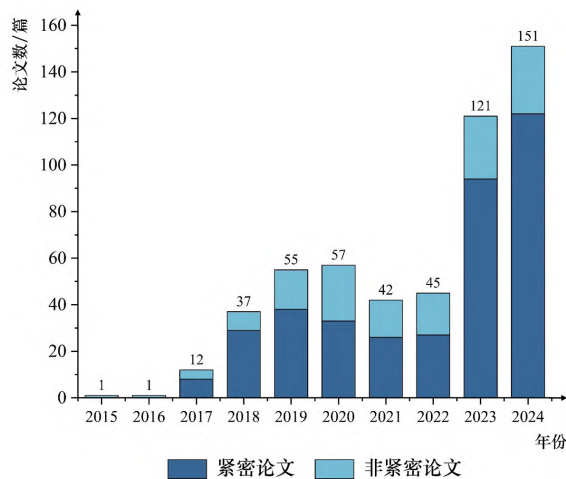


图2 论文发表年份分布

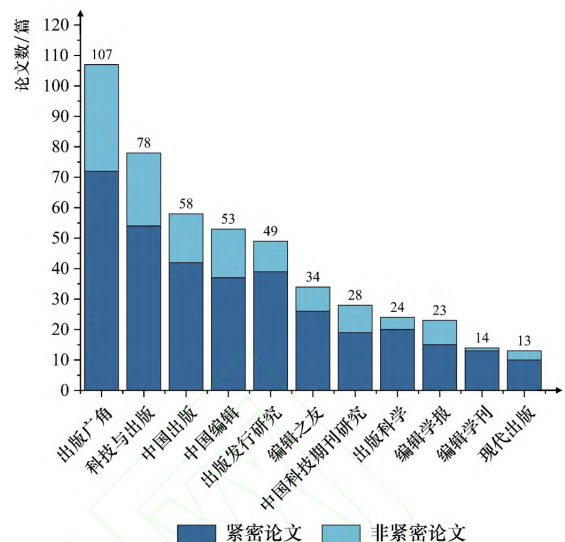


图3 论文发表期刊分布

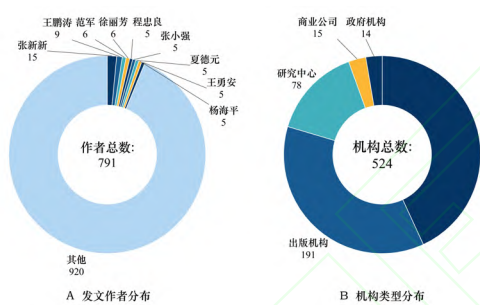


图4 研究主体分布

于出版业在采纳人工智能技术过程中可能遇到的伦理和安全挑战，包括版权保护、内容真实性、数据安全等关键问题。

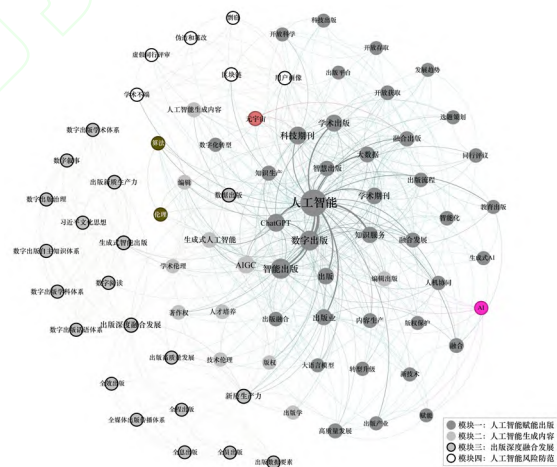


图5 关键词共现网络

本研究使用Gephi软件将纳入论文的关键词及其共现关系转化为可视化共现网络图谱，从而更直观地展示研究主题的分布和联系。根据图5，研究关键词主要形成了4个模块，模块一以“人工智能赋能出版”为核心，与“数字出版”和“智能出版”等关键词形成了紧密的联系，揭示了人工智能技术在推动出版业向数字化和智能化转型方面的研究趋势和重要性；模块二以人工智能生成内容（Artificial Intelligence Generated Content, AIGC）为核心，围绕人工智能在内容创作领域的应用及其对出版业可能带来的变革性影响进行了深入探讨；模块三以“出版深度融合发展”为核心，聚焦人工智能和其他技术对出版业深度融合的促进作用，并从新质生产力的角度探讨了这些技术对出版学科自主知识体系建设的潜在贡献；模块四专注

2 出版人工智能相关概念

学界提出或使用了诸多概念用于描述出版与人工智能等技术结合的业态，例如智能出版、智慧出版、融合出版、出版人工智能等，但尚未对相关概念的定义达成一致。“智能出版”是目前被最多论文采纳的概念（105篇），宋伟等^[5]于2018年明确提出“智能出版”的概念，并指出其是基于数字

出版流程再造的“人工智能+数字出版”新业态。伴随着智能技术在出版业的进一步扩散,“融合出版”(34篇)和“智慧出版”(28篇)概念逐渐兴起,除此之外,一些研究使用“智能化”(142篇)来描述人工智能对出版全流程的改造。然而,上述概念之间存在着较大的差异,“融合出版”是出版业中内容、平台、渠道、推广、销售与技术的相互渗透,出版物通过新的形态和渠道触达消费者;“智慧出版”是“数字出版”发展的新阶段,重点在于5G、区块链和人工智能等数字技术在出版行业应用后,出版物个性化、互动化和互联化程度的提高;“智能出版”强调出版全流程的编辑、创作、分发、反馈环节的智能化改进,例如基于大数据实现智能选题策划、精准图书推荐、智能排版审核等。上述概念在文献中仍然存在分歧,张新新等^[2]认为“智能出版”是“智慧出版”的高级阶段,是在后者满足读者个性化基础上的出版全流程智能化;但亦有学者认为“智能出版”与“智慧出版”均是以技术作为基础的新产物,但侧重有所不同,前者以技术作为驱动力,而后者以知识作为驱动力。通过对文献的梳理,研究发现所述概念之间的先后衍生或上下从属关系难以确定,各个概念既有其独特的关注视角,又存在相互交叠的范畴。作为前提,研究者应当厘清其研究所涉及的概念定义和内涵。

本研究纳入文献所指“人工智能”概念同样存在差异,概念演变实际上反映了技术快速迭代进步所造成的人工智能与出版业融合演变态势,包括早期简单的基于规则自动排版、编校等自动化技术或计算机辅助技术的运用,以及如今复杂的AIGC等智能技术与出版业的深度融合。

从时间维度来看,可将出版研究中的人工智能划分为3个阶段:第一阶段研究对象仍处于智能的初级阶段,人工智能以基于规则的自动化专家系统为主要形态,在出版业中的应用主要为出版流程的部分自动化,如自动排版、拼写检查和语法校对。这一阶段的研究大多集中在如何利用计算机工具辅助出版工作,目标是提升出版效率、减少人力成本

和规范出版物的质量。

随着大数据和机器学习技术的兴起,人工智能在出版业的研究进入第二阶段。在这一阶段,研究的核心转向了数据驱动的智能系统。通过对用户行为数据、阅读偏好数据和市场趋势数据的大规模分析,出版机构能够实现个性化推荐系统、智能排版和编辑系统及内容优化策略。研究的核心问题从如何自动化出版流程转向如何智能化数据和内容的交互,学界和业界的研究者开始关注如何利用用户数据优化内容创作、分发和营销。

GAI(如GPT等)的出现,标志着出版研究中人工智能的第三阶段。这一阶段的研究不再局限于出版流程的优化和自动化,而是开始尝试将人工智能生成能力嵌入内容创作环节,成为出版业发展的“新质生产力”之一。在这一阶段,人工智能不只是出版的工具,而是内容创作者的“合作者”,它不仅具有早期基于规则生成新闻报道、诗歌、小说的能力,甚至还能进行一定程度的创作。出版过程中的编辑、排版和审稿等环节,也因深度学习技术的进步而被进一步智能化和自动化。

每一阶段的研究侧重点不同,但都体现了出版业对技术进步的快速响应和深度融合。在这一背景下,可以总结出版人工智能的5个核心特点:

(1) 智能化。人工智能深度卷入出版流程,选题策划、编辑审校、发行营销等环节智能化水平全面提高;(2) 主体化。人工智能推进内容由专家生产(Professional Generated Content, PGC)到机器生产(Machine Generated Content, MGC),以GAI为代表的智能体(Agent)逐渐参与人类沟通过程,演变为意义共同创造者;(3) 知识化。人工智能突出知识在出版业的核心地位,改变知识封装方式,创新知识产品和知识服务形式^[6],拓展数字出版的要素构成;(4) 融合化。VR、AR等技术与人工智能在出版业融合互补,创新出版内容消费媒介形式,提升沉浸化体验;(5) 个性化。出版内容实现精确创作、精准营销、个性消费和实时反馈,进一步满足按需出版(Publishing on Demand, POD)的需求。

3 出版人工智能技术架构

纳入文献共提及104种人工智能支撑或赋能技术，出现频次前20名的技术名称见图6。根据这些技术之间的关系，可将其归纳为数据处理层、硬件支撑层、通用算法层和应用实现层4个维度：

- (1) 大数据、自然语言处理、互联网等技术构成了人工智能的数据处理层底座，机器学习范式的人工智能依赖大量数据的训练，而大数据等技术提供海量“数据养料”。对出版选题、编校、生产、销售和反馈等环节产生的结构化或非结构化的文本、图像等形式的数据加以收集、处理、存储和利用，是建构出版人工智能的关键路径；
- (2) 云计算、5G等技术组成了人工智能的硬件支撑层，为出版智能化发展奠定了算力基础；
- (3) 神经网络等深度学习算法构成了通用算法层，人工智能获得从数据中学习、发现规律和模式，从而实现决策的强大能力；
- (4) VR、AR、区块链等技术与人工智能相辅相成，创造出全新的智能化、安全性和个性化出版消费场景，推动智能出版发展。

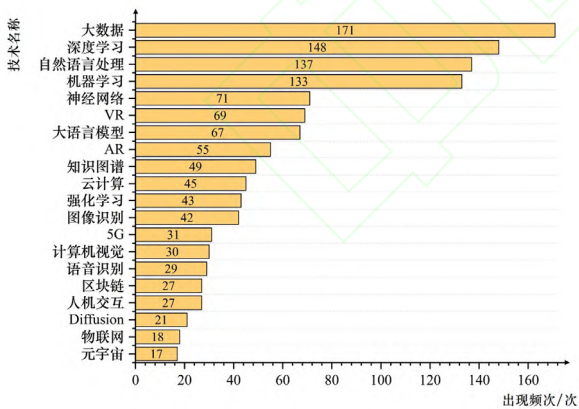


图6 人工智能相关技术出现频次前20名

出版企业在推进智能化革新过程中，面临的挑战不仅仅是技术层面的复杂性，还包括业务流程的再造和组织管理的变革。要实现从传统出版到智能出版的转型，企业需要将新技术的潜力与出版业务场景的实际需求深度融合。

首先，从业务流程的视角看，出版企业需要重

新审视和设计选题策划、编辑校对、排版印刷、市场营销和用户反馈等环节的工作流程。在传统出版中，这些环节大多以线性方式串联，但在智能化环境下，各环节的边界逐渐模糊，信息流动更加灵活。人工智能的介入，使得智能选题预测、自动化校对、个性化内容生成和精准营销成为了可能。

其次，从技术架构的视角看，数据的获取、管理和使用已成为出版业智能化的核心任务。与传统的“作者—编辑—读者”单向链式生产模式不同，智能出版更像是一个“数据驱动的动态反馈系统”。在这一系统中，出版企业需要建设自己的数据中台，整合来自不同来源的海量数据，并在统一的架构下实现数据的清洗、标注、建模和管理。此外，区块链技术的引入使得数字内容的版权保护和交易管理更加透明和可追溯，增强了用户的信任感，降低了数字内容盗版的风险。

最后，从用户体验的视角看，VR、AR、MR技术的赋能，推动了“沉浸式阅读”体验的普及。传统的纸质出版物受限于二维空间，而VR技术则能让用户“进入”书中场景，增强了阅读的沉浸感和交互感，其带来的沉浸式阅读体验也为教育出版和学术出版创造更多可能性，使“教、学、考、练”一体化的智能化学习平台逐步成型。

出版企业的智能化转型是一场“技术—业务—用户”多维联动的系统工程。这场转型不仅依赖于人工智能、大数据、区块链等前沿技术的协同作用，也依赖于出版企业对传统业务的流程重塑和对用户需求的深刻洞察。未来，随着5G、云计算和GAI的加速发展，出版企业的业务形态将进一步突破“内容提供者”的角色，逐渐转变为“内容+技术+平台”三位一体的综合服务提供者。

4 出版人工智能行业影响

人工智能技术在出版业已得到广泛应用，从出版类型来看，研究者普遍认为专业出版（206次）受人工智能影响最大，其次是大众出版（94次）和教育出版（80次）。专业出版，尤其是学术出版处于技

术创新的前沿，在“人工智能驱动科学创新 (AI for Science)”的背景下，众多学术期刊出版社和编辑部积极开展智能技术在学术出版领域的探索，通过人工智能进行审稿专家匹配推荐、多模态学术成果推广等。人工智能在科技期刊出版的投稿、审稿、生产和传播阶段均得以应用。投稿阶段，人工智能应用可为研究者提供写作选题和思路，其超越了传统学术检索引擎，能够匹配和研究更相关的文献，帮助学者进行文献综述；审稿阶段，人工智能应用可用于学术不端检测、审稿专家精准匹配，助力推进同行评议进程；生产阶段，人工智能可用于辅助论文自动排版、机器翻译、论文校对；传播阶段，人工智能可用于精准化论文推送和学术推广。以图书出版为主的大众出版是传统出版业的重要分支，人工智能可帮助完成图书内容制作、编辑加工和出版发行3个出版核心流程的工作。在教育出版中，人工智能丰富教育出版物的内容和形式、丰富教育出版物的“用户主体”。人工智能教材赋予传统教育新形态，提高教材教辅资料的可互动性，促进教材编写者和使用者间的双向互动，在与教育出版融合中主要体现出学习个性化、评价过程化、监督实时化的特征。

本研究基于出版的“多环节”论视角^[7]，探讨人工智能技术对出版流程各环节的多维度渗透。

首先，在分析的文献中，内容创作环节被认为是人工智能介入最深的领域（见表2）。作为出版流程的起点和核心环节，内容创作环节汇集选题策划、创意生成和内容生产等关键任务。人工智能的深度学习算法和自然语言生成技术，使“人机共创”成为出版创作的新范式。人工智能不仅能够对读者的阅读偏好、市场热点和流行趋势进行动态捕捉，还可基于大数据自动生成多模态内容（如文本、图片和视频），为出版物的内容创作提供源源不断的“机器叙事”。由此，AIGC不再仅仅是人类创作的补充，而是成为新型出版资源的重要来源之一。人工智能的崛起正在重塑从创作到消费的内容供应链，为出版企业提供前所未有的灵活性和响应能力。

其次，编辑校对环节的自动化升级是出版业的另一大变革焦点。在这一环节，传统出版的人工

表2 人工智能影响出版环节

出版环节	提及频次/次
内容创作	362
编辑校对	328
发行宣传	170
市场反馈	161
销售交易	122
生产复制	115

编辑方式以高频、重复和规则化操作为主，而人工智能的介入使这些任务的自动化水平大大提升。基于自然语言处理和语法纠错算法，人工智能可以高效执行文本的自动校对和智能审核，包括语法检测、敏感内容筛查、格式排版和术语一致性检查等任务。此外，通过算法生成的视觉方案不仅能够符合出版行业的规范要求，还能实现个性化的美学创新，从而加快出版物的排版与设计流程。这种“编辑—校对—排版—设计”的一体化智能 workflows，将原本分散的多步操作转变为一个可视化的动态流程，使出版企业能够以更低的成本、更快的速度和更高的质量交付图书和部分出版物。

再次，人工智能技术在发行宣传、市场反馈和用户体验环节的深度应用，标志着出版业从“产品导向”向“用户导向”转型。在发行和销售环节，人工智能对于控制出版企业的库存成本、优化市场投放策略具有重要价值。例如，人工智能可以根据图书的销售潜力和市场反馈数据，动态调整印刷批次和库存规模。智能推荐算法（如电子书平台中的“千人千面”推荐系统）也可以基于用户的阅读历史、兴趣标签和消费习惯，为每个用户精准推送个性化的图书推荐，这一策略已经在数字阅读平台中得到了大规模应用。此外，智能聊天机器人在用户交互中的作用也日益突出，不仅能够提供即时的售前咨询和售后反馈，还能充当“阅读顾问”，为读者推荐符合其兴趣偏好的书籍。“产—销—消”的智能交互机制，一方面增强了读者的品牌忠诚度，另

一方面使读者的反馈直接反哺内容创作，实现了一个动态循环的出版反馈系统。

最后，人工智能使出版行业发生了转型，从“出版商”向“知识服务商”进化。传统的出版企业以“内容提供者”自居，而在智能出版和知识服务的转型浪潮中，出版业的角色已经发生重大变化。出版的本质是知识的生产、流动和传播，但在知识泛化和多元知识生产的背景下，出版企业不再是唯一的知识提供方，GAI作为“知识生产者”的身份正在被广泛认可。它不只是作为工具被动使用，而是成为一种主动的“社会参与者”，不仅能对既有知识资源进行再组织和再利用，还能在数据驱动的模式中迭代生成新内容和新视角。从操作逻辑来看，GAI的运作机制与人类的知识生产活动有着某种相似性。具体而言，AIGC大多是对“已有知识的再加工和重构”^[8]，通过深度学习模型对海量数据的“吸收—总结—重组”，人工智能能够模拟人类的写作过程，并在生成的过程中生成新的语言表达形式和知识形态。由此，GAI的运作逻辑与人类的知识生产逻辑在一定程度上趋同，这种拟人化生产方式使其成为一种知识服务平台的核心生产力。在这一新生态中，出版不再局限于“内容生产—成品出售”的线性业务模式，而是转向“内容生产—知识流动—服务再生”的非线性生态。出版的逻辑从“物的销售”向“知识的服务化销售”转变。

人工智能在出版业已得到广泛运用，522篇文献提及了815个实例（见图7）。其中OpenAI及旗下的ChatGPT深受学界关注，诸多学者对其技术原理和特征进行介绍，并探讨其与出版业的融合路径。回顾出版人工智能的应用，研究者讨论了国外的实例以期为我国出版人工智能提供参考。亚马逊（Amazon）在图书推荐中引入人工智能算法开辟了人工智能介入出版营销领域的先河^[9]。数字出版时代，越来越多的国家和地区在新闻出版业采用人工智能技术，数字出版与人工智能技术愈发紧密地连接。我国亦有较多出版相关企业探索人工智能落地出版的尝试，例如人民卫生出版社利用人工智能挖掘医学数据中广泛存在的或隐性或显性知识，从

而支撑更智慧的知识服务。

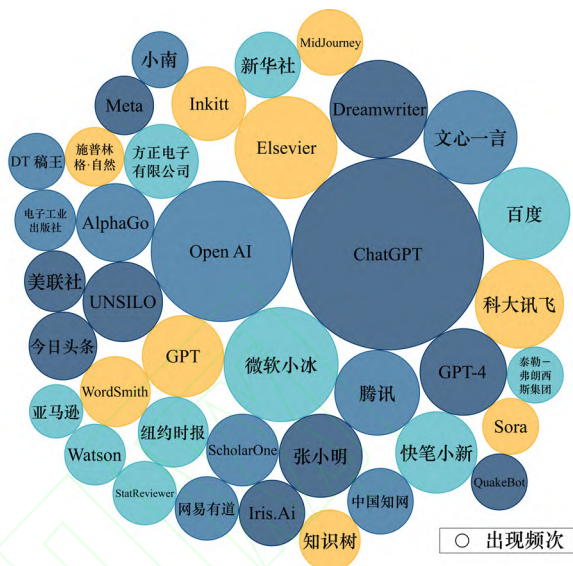


图7 出版人工智能主要实例

诸多学者提出，人工智能对出版而言是一柄双刃剑，既带来技术进步的机遇，也对传统出版生态造成深刻冲击。传统编辑首当其冲，智能技术不仅对出版从业者的职业素养提出新要求，还使原属于编辑的工作也逐渐被算法接管，加剧出版业就业的竞争态势。同时，出版企业在训练人工智能模型所需的海量数据方面面临困境，数据资源的不足成为行业智能化升级的瓶颈。

另外，人工智能带来了著作权等方面的争议。第一，公开训练语料中常涉及作家和出版社的内容，导致权利人无法获得应有的经济回报；第二，AIGC难以确定著作权的归属，无法界定原创和版权所属范围，极易陷入版权争议；第三，AIGC的真实性、准确性难以保障，且存在偏见、歧视内容，存在造假的可能。最重要的是，人工智能介入人类知识生产传播格局之后，人机伦理问题凸显，算法与人类在出版流程中的角色边界变得模糊，出版过程的规则和生产流程被算法形塑，伦理责任分配不清，出版行业面临更高的伦理监管压力。因此，出版业亟须重构出版编辑的伦理规约，更新行业的制度保障^[10]，共同努力设计、构建和部署“负责任的AI”。人工智能所代表的自动生成技术对出版业的

深远影响远不止于此,传统的线性生产模式被超文本的非线性知识链接所取代,出版与社会的关系发生重构,个性化的内容生成与快速知识获取削弱出版企业的“中介”角色,消费者的知识获取路径从“出版—阅读”转向“平台—搜索”,出版企业的市场地位受到平台化知识消费模式的冲击。平台化的内容生产还解构了出版业对印刷知识的垄断,读者注意力逐渐转向数字平台,深度阅读被即时搜索替代。出版企业的专业边界也在不断消解,知识的生产主体从精英化扩展到个体和大众。生成式智能不仅实现个性化的知识输出,还使人类对知识的生产垄断地位受到挑战,甚至颠覆出版业的“知识再加工”功能,知识的定义及其边界随之发生变化,人工智能的生成能力不断拓展知识的外延,促使学界重新审视“知识”的本质。

5 出版人工智能应对进路

本研究将纳入文献对出版人工智能的态度划分为积极、中立和消极3类,并对其进行线性拟合。根据图8,在时间维度上,积极态度文献占比显著减少($t=-4.88, p<0.005$),中立态度文献显著增加($t=4.96, p<0.005$),而消极态度文献占比变化并不显著($t=3.02, p>0.01$),表明随着技术发展和研究深入,研究者对出版人工智能等技术的态度逐渐回归技术理性,客观探讨出版人工智能的创新赋能和风险冲击,从而提供更为全面的见解。

在GAI的迅猛发展中,出版业的角色定位正从传统的“内容提供商”向“知识服务提供商”转型。这一转型不仅是商业模式的变化,更是知识资源管理和技术创新的深度重塑。过去,出版企业的核心使命是为人类读者提供高质量的内容,而在生成式智能的语境中,出版物的“读者”不再局限于人类,还包括人工智能等“机器读者”。这种用户边界的拓展为出版业带来前所未有的创新机遇。出版企业正在从“内容提供商”向“知识把关人”^[11]角色转变。与互联网平台依赖开放式的网络数据不同,出版企业拥有深厚的专业知识资源和严密的内容生产流程。

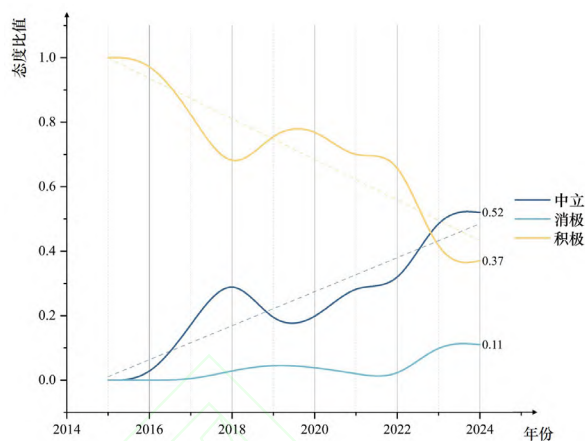


图8 研究对出版人工智能态度变化

大模型的预训练数据大多来源于互联网的开放数据,这些数据虽广泛但质量不均,导致模型“幻觉”现象的出现,即模型输出的内容可能是无意义的或不准确的。而出版企业的内容则经过专业编辑和审校,其专业性和权威性远高于网络数据。德国出版商Springer与OpenAI的合作便是典型案例,后者将出版商的高质量内容作为模型的训练数据。通过这种合作,出版企业不仅能为模型提供更高质量的数据输入,还能在产业链中扮演关键的“知识把关人”,确保生成的内容具有更高的可靠性和权威性。

出版企业的知识资源从“隐性知识”向“显性知识”转化。传统出版物中的知识多以文本或图片的形式呈现,具有一定的“隐性”特征,无法直接被人工智能模型高效读取和调用,而GAI的学习过程依赖于大规模、结构化的显性数据。为此,出版企业可以通过数据仓库和知识图谱等工具将传统出版物的内容进行结构化和语义化处理,将出版物中的专业知识转化为易于机器理解的“显性知识”,不仅可提升出版企业的内生竞争力,还能为大模型的训练提供更高质量的输入数据来源,进一步增强人工智能的预测能力和生成质量。

与此同时,出版企业应当积极开发和推出多种形式的知识服务产品,为企业、科研机构、教育部门等不同行业用户提供定制化的知识服务。与传统的“卖书”逻辑不同,出版企业需要探索更为多样的服务模式,如专家系统、嵌入式知识库、社交媒

体模式等。专家系统模式的典型应用是在科研和教育领域中提供精细化的知识问答服务,帮助科研人员和学术用户高效获取专业领域的前沿知识。嵌入式知识库模式则是将出版物内容以API或知识插件的方式嵌入大模型的内部,使得人工智能在回答特定领域的问题时,能够调用出版企业的高质量知识库,改善GAI的输出效果。社交媒体模式则结合网络社交平台的特性,出版企业可通过社交平台提供的实时更新机制,将最新的出版内容嵌入平台的知识流中,以满足用户的实时信息需求。

出版企业的转型不仅体现在服务模式的多样化上,也体现在内部运营的智能化变革上。通过建立智能中台,出版企业能够将内容生产的各个环节智能化和平台化。智能中台的核心是基于大模型、数据仓库和知识图谱的智能运营系统,它具有内容加工、知识转化和智能服务的多重功能。在内容加工方面,智能中台可自动执行文本的智能排版、语义分析、关键词抽取和知识标注等操作,从而提升出版物的生产效率。在知识转化方面,中台可将出版物内容转化为可供人工智能调用的结构化数据,并将这些数据输入到大模型中,支持问答和生成任务。在智能服务方面,出版企业的智能中台还可以作为服务平台向外部客户提供知识服务,覆盖科研机构、企业客户和政府机关等多种用户场景。

出版企业的这场转型不仅是自身业务的变革,也是对GAI技术生态的深度嵌入。如前文所述,出版企业在人工智能产业链中具备不可替代的“数据话语权”,如果其能将自身的内容数据化、结构化,并以API或“软件即服务”(Software as a Service, SaaS)的方式提供给人工智能企业使用,那么其业务模式将从传统的内容销售转变为知识服务,进一步扩大收入来源。

此外,GAI的发展促使出版业的竞争优势从“控制内容”向“控制知识”延伸。与网络数据相比,出版企业掌握的知识资源更加权威、系统、稳定,这为其提供了与人工智能企业合作的独特议价权。出版企业不必局限于传统的书籍销售,而可以与人工智能企业、搜索引擎、在线教育平台等多方

合作,成为这些平台的“内容输入者”和“知识把关人”,以嵌入式、订阅式的方式提供高价值的知识服务。这不仅能提高人工智能大模型的学习效果,还能帮助出版企业实现“出版即数据、数据即服务”的创新路径。

出版企业的这场“知识创新”转型还催生一种新的业务模式——“出版即服务(Publishing as a Service, PaaS)”。在这一模式下,出版企业不再简单销售纸质出版物,而是提供一整套的智能出版解决方案,面向企业、政府、科研机构提供的内容不再是“静态的书”,而是可搜索、可调用的“动态知识库”。这意味着,出版企业可借助大模型技术,将自己打造成一个智能知识平台,通过API、数据服务、嵌入式知识库等多种方式为外部客户提供知识服务产品。在这种新模式中,出版企业的收入不再局限于出版物销售的单一收入,而是通过提供嵌入式的智能知识服务,获取数据授权、订阅收入和平台服务收入。出版业的这一业务转型与SaaS模式类似,未来可能形成“内容即服务(Content as a Service, CaaS)”的新业态。

6 总结与启示

本研究使用系统性文献综述方法对出版人工智能相关论文进行了检索、筛选,并从元数据和内容2个维度分析了522篇论文。在文献计量分析的基础上,整理出版人工智能相关概念、技术架构、行业影响和应对进路4方面的内容,总结出版人工智能的研究现状与业界发展。

国内出版人工智能研究已形成了较为成熟的研究图景,然而也存在一些研究空白。首先,实证证据较为匮乏,研究方法较为单一。人工智能对出版业所造成的影响缺乏数据支撑,理论与实际应用之间的联系不够紧密,因此未来的研究应当坚持在实证研究的基础上建构理论框架^[12],从而促进出版学学科自主知识体系的完善。其次,过于强调技术,造成出版研究“本末倒置”。无论技术如何变革,出版业始终以内容为核心竞争力,因此出版学

研究应当着重探讨如何利用技术更好地满足社会需求,人工智能介入出版业后用户的接受度和市场适应性是一个重要因素,这方面的研究需要进一步扩展。再次,人工智能所带来的风险和挑战仍需进一步阐释,如版权侵权、内容确权和学术不端等,特别是对于这些问题的预防和解决方案的探讨。这些问题不单涉及社会文化层面,应当结合社会历史、经济等背景综合讨论。最后,出版业如何在人工智能背景下维持并增强可持续性,即如何在推动出版业智能化转型的同时,确保这一转型的可持续性,包括经济、社会和环境的可持续性,以及文化的延续性,是一个值得探究的问题。

人工智能对出版业的影响体现在各个环节,从内容创作到市场反馈,均发生深刻的变革。同时,GAI的崛起进一步加剧了这种影响,推动出版业向知识服务平台转型。然而,人工智能技术的应用也带来诸多挑战,如版权侵权、内容确权和学术不端等问题。出版业需积极探索技术与伦理的平衡,提升技术使用的规范性与可控性,开发具有优势的知识资源和“出版即服务”的商业模式。

作者简介

唐嘉骏,男,南京大学信息管理学院硕士研究生。研究

方向:数字人文和计算社会科学。

杨海林,男,南京大学信息管理学院硕士研究生。研究方向:数字出版。

马子寒,南京大学信息管理学院硕士研究生。研究方向:数字阅读心理与行为。

参考文献

- [1] 胡玉玺,王雨薇,程海威.智能出版:智媒时代传统出版业务的转型升级方向[J].科技与出版,2020(11):56-63.
- [2] 张新新,齐江蕾.智能出版述评:概念、逻辑与形态[J].出版广角,2021(13):21-25.
- [3] 陈丹,郑泽钊,张敏.我国数字出版研究综述(2003—2023年)[J].现代出版,2024(1):36-46.
- [4] LITTELL J H, CORCORAN J, PILLAI V. Systematic reviews and meta-analysis[M].Oxford: Oxford University Press,2008:5.
- [5] 宋伟,刘禹希,王鑫鑫.智能出版:开启后数字出版新业态[J].传媒观察,2018(8):24-29.
- [6] 易龙.从数字出版到智能出版:知识封装方式的演进[J].出版科学,2023,31(1):81-90.
- [7] 李新祥.出版定义的类型研究[J].出版科学,2011,19(1):43-48.
- [8] 张新新.生成式智能出版:知识生成原理、沿革与启迪:从智慧驱动到数据驱动[J].编辑之友,2023(11):36-44.
- [9] 苏磊,杨晓新.美国出版业人工智能应用研究[J].中国出版,2019(24):65-68.
- [10] 匡文波,姜泽玮.ChatGPT在编辑出版中的应用、风险与规制[J].中国编辑,2024(1):72-77.
- [11] 杨晓新,杨海平.人工智能背景下的出版业知识服务模式研究[J].科技与出版,2019(11):61-65.
- [12] 王鹏涛.出版业智能化发展研究的学术构想:关键维度与可用视角[J].新闻界,2018(11):95-100.

Concepts, Applications, Impacts and Measures of Artificial Intelligence in Publishing Industry: A Systematic Literature Review

TANG Jiajun, YANG Hailin, MA Zihan

School of Information Management, Nanjing University, 210023, Nanjing, China

Abstract: Combing through the current state of artificial intelligence (AI) research in publishing industry helps to clarify the development direction of the industry and promote the construction and improvement of the independent knowledge system of publishing discipline. This study employed a systematic literature review method, retrieved, screened, and analyzed 522 related literatures, and systematically summarized the core concepts, technical architecture, industry impacts, and response strategies of AI in publishing. The academic community believes that intelligent technology is deeply intervening in the entire publishing process, especially the application of generative AI (GAI) in content creation and knowledge production, which provides important support for the intelligent upgrading of the publishing industry. However, AI also brings ethical and practical issues, among which the copyright ownership and originality of AI generated content (AIGC) is particularly debated. The publishing industry should give fully play to its advantages in content resources and shift its roles to “knowledge service provider” and “knowledge gatekeeper”. By implementing “publishing as a service” model, the publishing industry is expected to achieve intelligent upgrading and build a new intelligent publishing ecosystem centered on high-quality knowledge resources.

Keywords: Artificial intelligence; Publishing industry; Intelligent publishing; Systematic literature review