

大模型时代“人机耦合”的数据出版业发展图景

孙蔚

北京印刷学院新闻传播学院，北京

收稿日期：2025年9月5日；录用日期：2025年9月30日；发布日期：2025年10月11日

摘要

随着数据成为关键生产要素，数据出版这一融合出版新业态也呼之欲出。本文围绕“数据出版业转型”，明确其逻辑起点、演进机理以及实践路径。遵循“人机耦合”的理论框架，提出数据出版媒介形态变革的演变机制，揭示大模型时代生产关系的演进机理，明确出版数据化建设的思路。根据媒介形态发展演变的基本逻辑，明确数智化时代变革中的非变量，提出人机耦合视域下开放的生产边界以及全新的阅读感知，对推动出版数据要素的挖掘和价值实现，催生数据出版新业态，纵深推进出版深度融合发展驱动新质生产力具有启发意义。

关键词

数据出版，大模型，人工智能

The Development Landscape of “Human-Machine Coupling” in the Data Publishing Industry in the Era of Large Models

Wei Sun

School of Journalism and Communication, Beijing Institute of Graphic Communication, Beijing

Received: September 5, 2025; accepted: September 30, 2025; published: October 11, 2025

Abstract

As data has become a key factor of production, data publishing—a new integrated form of publishing—

has also emerged as a necessity. Focusing on the “transformation of the data publishing industry”, this study clarifies its logical starting point, evolution mechanism, and practical path. Guided by the theoretical framework of “human-AI coupling”, it proposes the evolution mechanism of media form transformation in data publishing, reveals the evolution mechanism of production relations in the era of large models, and defines the approach for the construction of publishing dataization. In accordance with the basic logic of the development and evolution of media forms, it is essential to identify the invariant factors amid changes in the digital-intelligent era, and propose an open production boundary and a new reading experience from the perspective of human-AI coupling. This study holds enlightening significance for promoting the excavation of publishing data elements and the realization of their value, fostering new forms of data publishing, and advancing the in-depth integrated development of publishing to drive new-quality productive forces.

Keywords

Data Publishing, Large Models, Artificial Intelligence (AI)

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 数据出版业转化趋势

(一) 数字基建到智能引擎：技术代际驱动出版流程重构

出版业的数字化转型初期，主要聚焦于数字资源库与在线平台的搭建，实现“资源电子化”。而大模型技术的兴起，则推动行业发生质的飞跃，进入“流程智能化”的数智化新阶段(表 1)。这种转变的本质，是从单纯的数据驱动逐步进阶为算法驱动，并最终形成知识驱动的螺旋式上升发展路径。垂直领域大模型凭借对行业数据和特定场景的深度挖掘，大幅提升了生产效率，例如蜜度公司则聚焦智能审校赛道，其“文修智能校对大模型 3.0”新增事实性校对、图文一致性检查等能力，将 20 万字书稿校对时间压缩至 90 秒，视频审核效率提升至 5 分钟/90 分钟，重塑了知识服务的范式，催生出“一书一模型”等创新模式[1]。在未来的出版行业竞争中，核心竞争力将体现在人类脑力创意与机器智能的深度耦合，后续将会把数据资产转向知识服务价值转化效率的高低上。

Table 1. Technical iteration logic of large models

表 1. 大模型技术迭代逻辑

基础数字化(2000~2020)	单点智能化(2020~2023)	全流程数智化(2024-)
资源电子化、二维码应用	AI 审校/自动排版	“一书一模型” + 人机协同 workflows

通用垂直类大模型因专业度不足难以适配出版需求，许多出版业通过训练行业专属大模型突破瓶颈，使得大模型应用场景深化，实现全流程数智化 workflow。例如，数传集团基于 10 年积累的 3 亿读者数据和 300 家出版社资源，构建出版业首个大模型 BOOKSGPT，其针对古籍数字校对、多模态内容生成的准确率比通用模型提升 37% [2]。

出版数字化是未来发展的必由之路。数智时代，出版业的数字化进程不断加速，已成为出版机构的重要营收来源，也是出版业发展的必然趋势。出版流程中的编辑、校对、发行等环节广泛应用数字化技术，能显著提高出版效率和质量。在出版数据产品研发领域，出版数据产品化遵循“采集 - 存储 - 加工 - 治理”

的标准化流程。这一体系化运作模式，为出版业从数字化迈向数智化奠定了坚实的数据基础。

(二) 数据资产化：从资源沉淀到价值创造的范式革命

在技术代际后，出版业的核心竞争力从内容储备转向数据资产运营能力的过程中，大模型成为释放数据价值的关键枢纽。数据驱动已成为数智化时代商业模式的核心资源、基础资源，数据要素与算力、算法一起形成具有数智时代特征的数字生产力。传统出版的数据应用表现为数据资产体系化建设滞后、数据与业务应用场景结合不紧密、数据产品与服务开发滞后等。尽管出版企业积累了大量的内容数据和用户数据，但深度开发数据的能力较弱，数据驱动的商业模式尚未形成[3]。而大模型则可以实现对读者行为的深度挖掘，实现内容数据资源的定制化产品。例如数传集团通过 AIRAYS 平台为每本书构建独立知识库，基于读者扫码行为生成个性化语料库。教辅书《快乐 5+2》接入后，AI 依据学生答题数据动态生成针对性习题，扫码量提升 10 倍，成为读者数据驱动的服务升级的价值创造。

此外，知识图谱与大模型的双向赋能也在推动价值创造的范式革命，中信出版社提出“大模型 + 知识库”模式，通过知识图谱矫正大模型幻觉(如历史事实错误、编造性知识等)，同时利用大模型更新图谱结构，形成“数据 - 算法 - 知识”闭环。而在古籍数字化领域，中华书局联合高校研发“荀子”古籍大模型，将 OCR 识别、自动标点等传统技术与大模型结合，实现典籍的智能标引与语义重构[4]。

2. 人机耦合理论体系：数据出版媒介形态的变革

(一) 人机耦合理论的核心要义

人机耦合理论强调人类智能与机器智能的协同共生，并非简单的人机交互，而是通过深度融合实现优势互补、效能倍增。人类具备独特的创造力、情感理解、价值判断和复杂问题解决能力，能够从宏观视角把握内容的文化内涵与社会价值；机器智能则凭借强大的数据处理、高速运算和模式识别能力，可快速完成海量数据的分析与处理任务。在数据出版领域，人机耦合体现为人类策划选题、把控内容方向，机器则辅助进行数据挖掘、内容生成与优化，二者相互协作，形成有机整体，共同推动数据出版业务的发展。

人与人工智能交互(Human-AI Interaction)是一个快速发展的领域，20 世纪 60 年代，美国计算机科学家约瑟夫·利克莱德(Joseph Licklider)提出“人机共生”(Man-Computer Symbiosis)概念，该理念的核心主张在于：当人与计算机构建起共生协作关系时，能够充分发挥人类智慧与人工智能各自的优势，最终达成远超单一主体独立运作的效能[5]。尽管当前 AI 已初步具备独立完成创作任务的能力，但其在语境理解、导向把控、创新引入等方面尚难以完全替代人类。人机耦合体关注的是一种全新的主体观，传统认知中“人机二元对立”的思维模式，往往要么过度强调人类的主导地位，要么片面放大技术可能带来的威胁。而“人机耦合体”则打破了这种对立，提出了一种交融共生的新型主体形态，人不再是孤立的个体，AI 也不是单纯的工具，二者通过认知、情感、意义建构等方面的交织融合，形成新一代智能系统发展趋势。

在数据出版阶段，DeepSeek 作为代表性多模态 AI 大模型，其先进的数据处理架构和认知智能特性，以及语义理解、知识图谱构建及自动化推理能力，在出版数据要素供给、出版数据产品研发、出版数据高效流通等方面为数据出版提供了革命性解决方案[6]。大模型驱动出版媒介从“静态容器”向“自适应智能体”进化，首先体现在媒介本体重构，传统范式的纸质和电子书作为封闭数据载体，其内容固化不可交互。而智能媒介通过“二维码 + 边缘计算 + 大模型”技术三角，将书籍升级为“活态知识接口”，例如中信出版社《中国植物志》：扫码激活 AI 植物识别系统，实时比对 10 万种物种数据库，准确率高达 96.3%，这种从底层进行人机耦合的现象，充分体现了知识成为活态数据的优势。

(二) 数据出版的定义与生产关系重构

数据出版的核心内涵，可界定为“以数据作为核心要素的出版形态：其将文字、图像、音视频、游戏、动漫等各类内容均视作数据的具体呈现形式，围绕数据的挖掘采集、标引分类、存储管理与计算分析等关键环节展开出版全流程工作，并借助数据模型的搭建，最终实现出版数据应用的落地与出版数据服务的拓展”[7]。传统数据出版的生产关系是典型的“中心化线性控制”模式，其核心特征体现为角色固化和单向价值传递，编辑作为知识守门人，在知识筛选与审核过程中遵循“三审三校原则”，即为编辑初审、专家复审、总编终审，生产关系呈单向线性链条。

在人机耦合理论体系下，重塑业务链条形成“机器标准化 + 人类创造力”的新型数字分工体系。AI接管可程序化环节，蜜度智能审校系统的多项功能，可处理语法纠错、覆盖 1.2 亿知识实体的事实核查、版权冲突检测系统。而由 AI 接管基础核实功能后，人类工作则转向高阶跃迁，编辑核心能力聚焦为“AI 训练师”，其工作内容包括领域语料标注、提示词工程优化等，或是创作模式转为创作者与 AI 间的人机接力写作流程，AI 生成初稿后，人类强化情感张力的部分内容。在未来，人机交互的范式迁移应体现为以下方向(图 1)，而衡量编辑工作量应需要一定的量化公示方能使人机协作有更好的衡量标准，实现上下游资源协同创新，从而提升融合效能，量化公式：人机协同效能比 = 人类创意贡献量/AI 标准化处理量。



Figure 1. Paradigm shift in human-AI interaction

图 1. 人机交互范式迁移

(三) 技术驱动下数据出版媒介形态的多维变革

物质实体赋予出版以“形”，承载的符号赋予其“意”。但出版物是具有社会性的传播媒介，还需以“用”将其从文本层面带至现实生活中，通过可视界面与读者真正建立联系。安德鲁·霍斯金斯(Andrew Hoskins)在谈及数字记忆时提出“联结性转向”(connective turn)，在他看来，数字媒介使得处于不同网络节点中的人相互联结，进而重塑时空和记忆[8]。传统出版无法处理动态的信息，导致出版的可视界面仅能与人实现互联，界面与界面之间以及界面与其他介质之间无法联结。智能技术是一个新的交转系统，多模态符号在此聚合，让出版从单一的媒介形态转为“出版+”，而现在大模型时代的数据出版业正经历从“零和供应链”向“共生知识共同体”的生态跃迁。这一变革的核心在于构建去中心化的价值创造网络，通过技术协同与制度创新解决数据孤岛、权属争议与伦理失范等痛点，实现知识生产的社会化扩展。出版业实践目前案例有古籍数字化联盟，中国国家图书馆联合 28 家古籍馆建立联邦平台，使《永乐大典》残卷的 OCR 识别错误率从 8.3%降至 1.7%，且各馆珍本图像无需离开本地服务器[9]。这样成功地破解数据孤岛的方法在未来可以以更多样化的形式呈现。

出版从“平面”向“立体”的转变，也意味着过去出版受制于界面的“边界”被打破，真正成为无边界、可融合的形态。智能出版的“破界”是对阅读界面的“打破”，只有打破才能实现联结。出版数据产品媒介形态的研发方向可在产品、服务和模型三个方面展开。在出版数据产品维度，协同推进单一型出版数据产品和集合型出版数据产品研发进程。单一型数据产品研发是基于单个数据单元，应用相应的音频、视频、动漫、三维模型等技术，制作生产相应的出版数据产品；集合型数据产品研发是基于若干数据单元、数据单元集或海量数据单元，来研发数据集合型产品。在数据出版服务维度，提供面向大众群体、专业群体、个人用户或机构用户的出版数据服务，包括数据查询、检索、标记、核验、定制化等各种数据服务解

决方案。在出版数据模型维度,可以考虑在海量出版数据语料库的基础上,根据用户提问,提供文生文出版垂直大模型服务;或者基于文字描述,提供文生音频、文生视频的多模态出版垂直大模型服务。

3. 出版媒介形态变革的发展基本逻辑

(一) 数智化时代变革中的非变量

面对人工智能技术对出版产业全流程、全产业链应用的深刻影响,出版业既需要积极求变,应用新技术,开发新技术;也需要守正,坚守自身的核心价值与责任担当,于这场深刻变革之中,仍存在诸多稳固的非变量因素,它们是出版业坚守本质、实现可持续发展的基石,是出版业数智化变革的基本逻辑。内容的核心价值是数智化时代始终不变的关键要素。尽管大模型等技术能够辅助内容生产,提升效率与丰富表现形式,但优质内容生产仍是数智时代出版企业竞争的“红海”[10]。内容的质量不仅体现在文字、图片、音频、视频等多媒体元素的融合创新,更体现在内容的独特性、专业性和实用性。这些始终依赖人类的智慧与创造力,就像经典文学作品历经岁月沉淀仍被广泛传诵,正是因其蕴含的人文精神与深刻思想。出版业无论技术如何更迭,都应以挖掘、呈现优质内容为根本,满足人们对知识、审美与精神世界探索的需求。

文化传承与版权保护是出版业不可动摇的责任使命。出版作为文化传播的重要载体,保护好原创版权就是保护出版产业价值链的核心。数智化时代,内容的复制与传播变得异常便捷,AI技术虽然可以极大提升内容生产的效率和自动化程度,但对原创性思想和内容的创意表达,是AI尚无法达到的。古籍数字化出版项目中,虽借助智能技术实现古籍的高效整理与广泛传播,但仍需建立版权激励与保护机制,这样才能维护创作者的合法权益,不断推动创作者的驱动力。出版企业通过区块链技术,实现版权信息的可追溯、不可篡改,为版权保护提供强有力的技术支持仍是出版工作的核心,避免在技术发展中迷失方向。

(二) 人机耦合开放内容生产边界

传统出版模式下,内容生产主要依赖专业作者、编辑团队,生产主体相对单一。而人机耦合模式引入智能机器作为新型“生产主体”,极大地拓展了内容生产参与者的范畴。智能出版媒介形态变革,使得人人都成为内容创造者和传播者,个体可利用人工智能系统自动生成一部小说,并根据需求形成不同的开放式结局。媒体写作时大模型能够依据预设指令,快速生成新闻稿件、科普短文等内容,大幅提升新闻时效性。这样的专业人员、机器、大众用户共同参与的多元生产格局,使内容生产从“精英化”走向“大众化”。人机耦合的视域下,机器并非人类内容生产的外化工具,而是与人合二为一,在颠覆原有以人类为主导的知识生产的基础上,开放大模型时代出版内容生产的边界。

(三) 人机耦合重塑阅读感知

从“界面”转向“场景”,体现了出版从平面化向立体化的转变,也意味着过去出版受制于界面的“边界”被打破,真正成为无边界、可融合的形态。智能出版的“破域”是对阅读界面的“打破”,只有打破才能实现联结。《机械复制时代的艺术作品》中,本雅明提出,“灵晕”是艺术品中的独特性、距离感和本真性的存在,是艺术之所以为艺术的关键。机械复制时代,大规模的批量生产让作品本身与作者的劳动实践和情感相隔离,导致“灵晕”消失[11]。智能出版对场景的复现让消失的“灵晕”以一种新的方式回归,国内各大出版社推出的定制化出版在为读者们服务。未来的智能出版以“融合”为创新路径,阅读不再是单一的读,而是以多版本融合、多模态符号、多场景联结等再造阅读感知,阅读成为调动全身感官在场的全新体验。

但艺术的机械复制逻辑和虚拟场景复现的逻辑并不一样,本雅明指出的艺术本身的“灵晕”显然已在时间的流轉中一去不复返,物质实体与现实时空的关系无法复原,但借助人工智能技术,虚拟的时空

场景以个性化体验的方式被想象和重建。智媒时代的读者经历了“从置身事外到参与其中”的转变，用户不再满足于被动获取模式化的知识内容，而是依托大数据驱动主动寻找和获取个性化的偏好内容。例如华文出版社与抖音合作推出的“抖音定制版”图书，用限量版藏书票吸引用户眼球，搭配定制化营销取得良好销售额。从这个层面来说，已经消逝的“灵晕”得到了复现，在算法机制的加持下，出版作品的内容价值得以被精准触达偏好用户，随着智能出版媒介形态的变革，阅读在将来会更加充满想象。

4. 出版数据智能基建驱动新质生产力

随着《关于促进数据产业高质量发展的指导意见》面向社会公开征求意见，发展数据产业已成为深化数据要素市场化配置改革、构建以数据为关键要素的数字经济的重要举措[12]。理念是行动的先导。出版数据思维优化，是出版数据化建设的前提，也是出版数据要素潜能激发、要素价值实现的前置性条件。出版数据是劳动对象，是新生产要素。从劳动对象来看，劳动对象的质变及其与劳动者、劳动资料优化组合的质变，是新质生产力的基本内涵之一。对于国有出版单位而言，传统出版的劳动对象，是知识、作品、版权素材；数字出版的劳动对象，主要是已经出版的图书产品，是对图书产品进行简单数字化或深度数字化加工，转化为电子书、条目数据、数据库、知识库等数字出版产品。然而，数据出版的劳动对象则是数据，是出版业的内容数据、用户数据、交互数据和治理数据。数据出版的过程，就是数据出版编辑利用数智工具改造多种类、多元化出版数据以适应人们学习和阅读需要的过程；对出版数据改造的结果，是形成多种出版数据产品、出版数据服务。

数据智能基建和出版业务融合是关键环节，出版社要实现编辑从“内容把关人”到“数据驱动创新者”的成功转型，这一过程需要多方面的努力。一方面，出版社应积极引进或着力培养既懂技术又懂出版业务的复合型人才[13]。这类复合型人才如同连接技术与业务的桥梁，能够在编辑团队中发挥至关重要的作用。他们凭借自身对数字技术(如大数据、人工智能、多媒体制作等)的深入理解，以及对出版业务流程(包括选题策划、内容编辑、排版设计、发行营销等环节)的熟悉掌握，能够将数据驱动的创新理念与具体的技术应用巧妙地结合起来。另一方面，出版社要大力鼓励编辑与技术人员开展广泛而深入的合作项目。通过实际项目运作，为双方创造更多相互学习、相互协作的机会，从而逐步消除技术与业务融合的障碍。在合作项目中，编辑可以深入了解各种数字技术的实际应用场景和操作流程，亲身体验数字技术如何为出版业务带来创新和变革。在数据分析工具使用上，确保编辑熟练掌握 Excel、SPSS 等，Excel 培训从基础函数到高级功能逐步深入，通过案例让编辑学会整理分析稿件数据，SPSS 培训则聚焦其统计分析功能，指导编辑运用其处理市场调研数据，为出版策略提供依据；在数据挖掘基础技能方面，编辑要了解概念、方法及应用场景，通过项目演练挖掘读者数据中的价值信息；数据可视化技能方面，可通过培训教会编辑将销售等数据可视化呈现，助力其在各项工作中作出明智决策，确保技术应用能够更好地服务于出版业务。进而实现技术与业务的无缝对接，使出版社能够推出更多既符合市场需求又富有创新性的数字出版产品，有力推动编辑向“数据驱动创新者”转型。

最后，建立数据发展与安全理念是数据出版的底线和保障。出版数据资产，涉及内容数据、文化数据以及意识形态领域的的数据，关涉内容安全、文化安全和意识形态安全；出版数据资源，往往能直接反映我国最新的科技动态、科技成果和重大攻关项目成果。由此，其衍生数据的开发还涉及科技安全、信息安全、自然资源安全、基因安全等众多安全范畴。只有确立科学的发展理念，拥有先进的数据思维，方可推进出版数据要素发掘、数据产品研发、数据高效流通以及数据价值实现。

参考文献

- [1] 邱毅. 蜜度垂直大模型赋能交通与物流数智化建设应用[J]. 中国物流与采购, 2025(1): 70.

-
- [2] 武汉理工数传公司首创出版行业大模型 全国超百亿册图书变身“现代纸书” [EB/OL]. 湖北日报. https://news.hubeidaily.net/pc/c_2742531.html, 2024-06-12.
- [3] 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.
- [4] 聂慧超. 中版集团深入大模型打造文化新质生产力[N]. 中国出版传媒商报, 2024-01-12(031).
- [5] Licklider, J.C.R. (1960) Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, 1, 4-11. <https://doi.org/10.1109/THFE2.1960.4503259>
- [6] 康亚飞. 虚实“破域”: 智能出版的媒介形态变革与未来构想[J]. 现代出版, 2024(11): 23-29.
- [7] 张煜琦. 数据驱动、内容为王, 图书出版探寻选题策划新方向[J]. 现代阅读, 2025(7): 10-12.
- [8] Hoskins, A. (2011) 7/7 and Connective Memory: Interactional Trajectories of Remembering in Post-Scarcity Culture. *Memory Studies*, 4, 269-280. <https://doi.org/10.1177/1750698011402570>
- [9] 张新新. “数据要素 × 出版”: 从数字出版走向数据出版[J/OL]. 图书情报知识: 1-10. <http://kns.cnki.net/kcms/detail/42.1085.G2.20250214.1500.002.html>, 2025-06-01.
- [10] 周慎, 莫菲菲, 张佳伟. 人机耦合体与人机双螺旋: AIGC 时代数字内容创作关系与方法重塑[J]. 数字出版研究, 2024, 3(4): 121-129.
- [11] 本雅明. 机械复制时代的艺术作品[M]//阿伦特. 启迪: 本雅明文选. 张旭东, 王斑, 译. 北京: 生活·读书·新知三联书店, 2014: 234.
- [12] 六部门印发《关于促进数据产业高质量发展的指导意见》[J]. 中国外资, 2025(2): 7.
- [13] 张新新, 孙瑾. 要素·结构·功能: 出版业高质量发展经济维度分析——基于提高出版经济活动质量的视角[J]. 数字出版研究, 2023, 2(4): 47-56.