面向生成式人工智能（AIGC）的图书馆信息资源建设优化策略研究 *

丁遒劲 [1] 苏静 [2]
[1] 中国人民大学图书馆 北京 100872
[2] 陕西师范大学新闻与传播学院 西安 710069

摘 要：[目的 / 意义] 以 ChatGPT 为代表的生成式人工智能（Artificial Intelligence Generated Content，AIGC）对知识生产和服务方式带来巨大变革，图书馆不可避免地受到影响进而面临挑战，亟待充分审视现有信息资源建设模式存在的局限性，通过优化信息资源建设推动实现图书馆的智慧服务。[方法 / 过程] 从 AIGC 的典型应用入手，分析 AIGC 对图书馆服务产生的影响，进而结合图书馆信息资源建设现状和存在问题，探讨相应的资源建设优化策略。[结果 / 结论] 图书馆在适应 AIGC 发展推动服务变革的过程中，亟待加大开源信息资源建设力度、重视馆藏数字资源权益管理、探索知识库嵌入大语言模型（Large Language Model，LLM）的融合应用路径、重塑数据治理体系，以及加强与技术公司的协同合作。

关键词：信息资源建设 AIGC 智慧图书馆 知识服务
分类号：G251
DOI：10.13266/j.issn.0252-3116.2024.18.003
引用本文：丁遒劲，苏静．面向生成式人工智能（AIGC）的图书馆信息资源建设优化策略研究 [J]. 图书情报工作，2024, 68(18): 23-31. (Citation: Ding Qiujing, Su Jing. Research on the Optimization Strategies for Library Information Resources Construction Oriented to the Development of AIGC[J]. Library and Information Service, 2024, 68(18): 23-31.)

1 引言 /Introduction

生成式人工智能（Artificial Intelligence Generated Content，AIGC）被认为是继专业生产内容（Professionally Generated Content，PGC）、用户生产内容（User Generated Content，UGC）之后的一种新型内容创作方式，是自动化内容生成的技术合集。2022 年至今，各大科技巨头密集推出 AIGC 产品，从效率、质量、多样性等方面为内容生产带来新一轮变革。特别是在文本自动生成方面，OpenAI 通过引入人工反馈的强化学习（Reinforcement Learning by Human Feedback，RLHF）机制对 ChatGPT 进行训练优化，实现自动问答、诗歌创作、代码写作等。此外，AIGC 在突破文本的跨模态内容生成方面也拥有较为广阔的发展空间。

AIGC 的快速发展直接颠覆既有的知识生产和传播交流模式，催生出科学研究第五范式——科学智能范式（AI for Science）[1]，科研人员通过问答交互能够辅助完成从研究构思、方法设计、综述撰写，甚至投稿推荐等整个研究过程，为材料科学、能源、药物研发等基础科学研究领域带来全新机遇。对图书馆而言，处于上游环节的研究范式以及科研成果出版传播方式均已受到 AI 技术发展影响，进一步加剧了 AIGC 对图书馆的冲击。图书馆亟待充分把握 AIGC 对服务端产生的影响，审视现有资源建设模式存在的局限性，进而探讨相应的资源建设优化策略，推动实现图书馆的智慧服务。

2 相关研究 /Related research

2022 年 11 月，美国人工智能实验室 OpenAI 发布生成式对话机器人 ChatGPT，迅速引起全球各行各业关注。从被称为 AIGC 元年的 2022 年开始，较多研究以 ChatGPT 为切入点，以 AIGC 对信息资源管理产生的影响和发展对策展开研究。

2.1 AIGC 对信息资源管理的影响研究

在 AIGC 对信息资源管理全局影响研究方面，李

* 本文系国家社会科学基金青年项目“基于媒体融合的图书馆知识服务优化机制研究”（项目编号：19CTQ008）研究成果之一。
作者简介：丁遒劲，副研究馆员，博士；苏静，副教授，博士，通信作者，E-mail：owensujing@163.com。
收稿日期：2023-11-02 修回日期：2024-02-02 本文起止页码：23-31

白杨等[2]以数据赋值、模型赋智、空间赋能3个维度为着力点，分别探讨了AIGC的技术特征、技术要素和发展阶段，指出AIGC与技术算法的融合应用，为信息资源管理的研究与实践带来了实质性的影响，具体表现在信息组织、数据资产管理、用户研究和信息伦理4个方面。陆伟等[3]以ChatGPT为代表，从支撑算法与技术、信息资源建设、信息组织与信息检索、信息治理、内容安全与评价、人机智能交互与协同6个角度探析大模型对信息资源管理学科研究与实践带来的影响。赵杨等[4]立足于新一代AI环境下传统智慧图书馆的变革与发展，从基础设施层、算法支撑层、数据资源层、业务功能层和服务应用层5个部分构建融合AIGC技术的智慧图书馆体系框架，从转型目标制定、服务体系重构、技术设备升级、内容矩阵建设和应用生态拓展等方面提出实现路径。

在信息资源管理各环节具体影响研究方面，王鹏涛等[5]探讨了以ChatGPT为代表的AIGC技术介入知识生产活动引发的信任危机，认为AIGC介入学术知识生产带来新的不确定性因素致使学术共同体信任风险增长或转向，需要从技术、人际和制度3个层面协同重构学术出版信任机制体系。方卿等[6]从出版角度出发，指出出版学需重点关注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。X. Chen[7]通过比较ChatGPT与传统图书馆聊天机器人的参考咨询服务能力，提出图书馆界应关注变革性技术带来的影响。X. Zou等[8]调查了学生群体的AIGC使用模式、动机、感知收益和风险意识，强调图书馆需从AI素养和批判性思维技能两方面加以培育。郭亚军等[9]结合内容生产方式变革视角下的图书馆服务演进历程，分析ChatGPT赋能图书馆服务的“4T特征”，构建ChatGPT应用于图书馆服务场景的“RSRC”框架图，并提出ChatGPT赋能图书馆的实现路径。

2.2 AIGC的质量评估与治理方法研究

AIGC的迅猛发展不可避免地将整个人类社会置于信息过载、信息噪声、信息安全等的负面影响之下，面向AIGC的信息质量评估和治理尤为重要。为此，国家互联网信息办公室等七部门于2023年发布《生成式人工智能服务管理暂行办法》，对AIGC进行宏观专项监管。在微观方法层面，莫祖英等[10]采用数据测试实验方法，立足于信息质量视角，根据内容生成机理与表现形式将AIGC虚假信息划分为事实性虚假信息和幻觉性虚假信息，指出虚假信息产生的根源与LLM、预训练数据集和人工标注3个要素有关。宋士杰等[11]在解析AIGC可信度概念内涵的基础上，从AI作为信息源、交互方式、社会行动者、用户算法隐喻4个方面，构建人智交互体验中AIGC可信度评价的研究框架。朱禹等[12]基于AI事故数据库，以AIGC相关事故报道为样本进行内容分析，提出由政府、企业、社会三方行动主体形成“多元+协调+制衡”的AIGC治理参与模式，并在“情境—意识—行动”的行动框架下开展治理。B. Lund等[13]认为，ChatGPT难以准确回答复杂的医疗问题，可能会提供不正确或过时的信息，这给患者护理和决策带来风险，对此，在将ChatGPT集成到医学图书馆环境中时，提出持续监控、针对可信来源的验证以及质量控制流程的实施等方法，以最大限度地减少误导或不可靠信息的风险。

2.3 针对AIGC的信息资源管理发展对策研究

具体到业务发展层面，张智雄等[14]从数据组织方式、知识服务模式、情报分析方法、文献使用方式、文献情报队伍建设要求以及文献情报工作重点6个方面，分析了ChatGPT对文献情报领域的影响，同时基于文献情报工作的特点，提出AI时代文献情报领域发展的9条建议。赵瑞雪等[15]通过分析ChatGPT的发展历程、技术特点、典型应用场景和集成应用路径，对比国内外同类技术产品，总结ChatGPT存在的技术局限和安全风险，从全文本地化建设、知识组织体系建设、深化技术应用等方面，提出ChatGPT等LLM的出现和应用对我国图书馆及相关信息机构的启示。Y. Lappalainen等[16]以阿联酋扎耶德大学图书馆基于Python和ChatGPT API接口开发聊天机器人Aisha为案例，V. Stepanov等[17]以俄罗斯国家图书馆等4家图书馆为对象，探讨了基于ChatGPT的聊天机器人在图书馆中的应用潜力。A. Adetayo[18]、X. Liu[19]也分别指出ChatGPT无法取代图书馆员，应被视为一种补充图书馆员专业知识的工具，图书馆可利用ChatGPT提供参考咨询服务，并协助研究、编目、分类、馆藏资源开发、数据集成和共享。李荣等[20]在肯定ChatGPT对开源情报的信息搜索、信息获取、信息处理环节具有一定提升作用的同时，指出现有技术缺陷使ChatGPT在开源情报全流程介入中仍面临数据可靠性、情报隐秘性、意识形态风险等问题与挑战，情报机构需要采取AIGC技术融合理论探索、生成内容可靠性评估、智能技术体系建设等积极策

略来应对本轮技术变革，更好地实现开源情报价值。针对 AIGC 如何赋能竞争情报这一问题，陈超[21]指出，科技情报工作者尤其要高度关注各类垂直专业领域 LLM 的发展态势，有可能在垂直专业领域 LLM 的基础上做进一步的优化和拓展，相应领域的诸多情报需求基本都可以满足。

综上，信息资源管理界已从整体上对 AIGC 带来的影响进行研究，信息资源建设、组织加工、服务利用各个方面均有所涉及，但是少有研究结合信息资源建设现状展开更为详尽的探析。基于此，本文在分析 AIGC 对图书馆服务带来变革的基础上，探讨信息资源建设在服务变革中的功能定位，进而形成相应的优化策略。

3 研究框架 /Research framework

图书馆的信息资源建设工作与机构自身服务使命密切相关，直接影响着信息资源建设的功能定位。因此，需要以 AIGC 对图书馆服务产生变革为中介，分析信息资源建设的局限性及相应的优化策略。研究框架如图 1 所示：

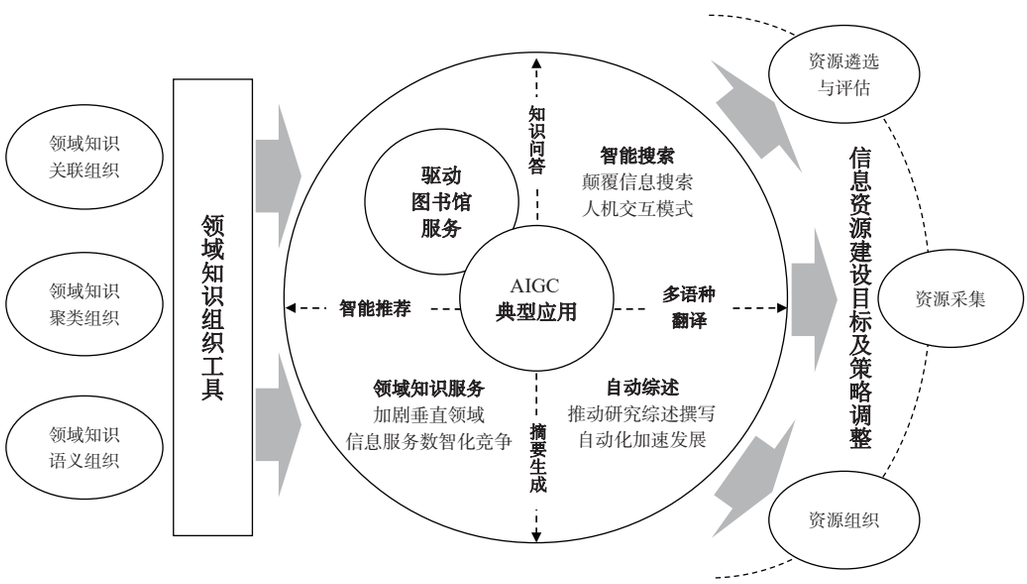

图 1 研究框架

Figure 1 Research framework

当前，AIGC 的典型应用包括聊天机器人、多语种翻译、智能推荐、摘要生成和知识问答。基于上述应用，AIGC 对图书馆主要服务模式交叉形成三方面的变革：①智能搜索中的人机交互顺畅自如，对图书馆原有基于关键词匹配的信息检索质量和用户体验提出了更高的要求；②研究综述与科研人员的研究过程密切相关，AIGC 的发展有力地推动了研究综述撰写自动化，对图书馆专题咨询服务内容和方式形成挑战；③在领域本体、知识图谱等知识组织工具支撑下，以法律、生物医药等领域为代表的领域知识服务需求旺盛，相应的信息服务产品层出不穷，图书馆在垂直领域面临的信息服务数智化竞争更加激烈。面向图书馆服务变革，现有信息资源建设模式存在较大局限性。当前，图书馆的资源建设对象仍是以付费订阅的文献资源为主，对开放资源、非文献型资源（例如科学数据、音视频等）关注有限。在文献资源建设载体方面，图书馆主要通过购买数字资源访问权满足用户需求，本地保存资源规模较小，因此也很难展开大规模的资源组织与整合利用。为此，图书馆信息资源建设需要从资源遴选与评估、资源采集和资源组织方面，逐个环节进行优化，支撑图书馆智慧服务的实现。

4 AIGC 驱动的图书馆服务变革 / Transformation of AIGC-driven library services

ChatGPT 作为具备大规模语料库训练的生成式自然语言处理模型，具备自动文本生成、智能信息处理、语义搜索与判识、智能图像生成等功能[22]，重塑了知识服务的生命形态，有可能对传统知识服务造成降维打击[23]。在实践探索方面，基于自身数据、知识产权以及技术优势，较多大型信息服务商已先于图书馆将 AIGC 应用于产品和服务之中，形成类 ChatGPT 服务平台，信息服务商与图书馆之间的服务竞争进一步加剧，驱动图书馆服务面临变革。面对图书馆服务

变革，信息资源建设作为上游环节亟待同步做出改变，推动图书馆在新一轮信息服务竞争中深入发展。

4.1 颠覆信息搜索人机交互模式

图书馆现有的检索系统，包括各类商业订购数据库、自建数据库以及联合目录系统和资源发现系统，多数都是以关键词和主题检索为主。但是以关键词检索式作为概念的组合，无法精准表达用户信息需求的完整语义，造成检索系统难以匹配，很难输出用户真正需要的文献[24]。基于海量数据、强大的逻辑推理能力和人性化的表达习惯，AIGC 能够对用户提出的个性化、差异化问题，输出更真实、更有效、更加适配的答案[25]，由此推动基于关键词匹配的信息检索服务向更加智能的学术搜索引擎发展[26]。

在 AIGC 赋能后的信息搜索中，用户可以发出指令或提供输入，机器会基于用户提供的输入或上下文来产生反馈或提问，从而进一步推动人机交互。在输入环节，LLM 更容易捕获用户意图，用户可以用自然语言多次交互，有效降低输入门槛；在信息处理环节，LLM 的自然语言理解和推理能力将用户输入数据转化为各类机器可识别读取信息，大幅度提升交互质量和效率；在输出环节，输出内容将更为自然和符合用户需求；在反馈阶段，LLM 能够根据以往的反馈进行自我调整。例如，在世界顶级法律、专利、税务服务商 LexisNexis 推出的面向法律界的 AIGC 平台——Lexis+ AI 中，用户可以直接使用类似“生产、销售伪劣商品应如何处罚”的自然语言查询相关法律问题，平台将自动生成相关详细的法律知识解读，同时附带实际发生过的案例，而来自测试用户的预加载提示问题和反馈也将进一步增强 Lexis+ AI 的功能。

4.2 推动研究综述的自动化撰写进程

研究综述自动生成的主要任务在于基于一系列同一研究主题的科研文档集合自动输出一篇能描述该主题研究现状的文本，能够浓缩特定主题信息，有效地提高信息获取效率[27]。研究综述自动生成一直是科技文献自动摘要领域的具体研究应用之一[28]。受限于自然语言处理等文本计算技术水平的发展，研究综述自动生成存在数据资源处理规模有限、质量不稳定、知识挖掘深度不够、展示形态单一等问题。ChatGPT 具备处理不同类型、不同领域的大规模数据能力，推动研究综述的生成质量和效率。

在文献信息服务领域，国外较多信息服务商结合 AIGC 发展探索研究综述自动生成服务。例如，Digital Science 推出的 Dimensions AI 助手基于用户提问在底层 3 300 万篇论文中进行检索，并对排名靠前的结果进行语义相关性排序，进而围绕最为相关的 4 篇论文摘要生成总结[29]。与之类似，Elsevier 发布的 Scopus AI 能够生成一段关于特定研究主题的流畅总结、引用的参考文献以及需要进一步探索的问题[30]。更有甚者，Springer Nature 接连出版三本以 AI 方式生成的图书，依托自然语言处理和 AI 技术的发展实现知识内容生产主体从人向“人 + 机器”的跨越，促进知识内容生产效率得到进一步提升。

4.3 加剧垂直领域信息服务数智化竞争

随着 ChatGPT 和 Google Bard 等通用模型的开发和应用越来越普遍，从艺术和娱乐到医疗保健和制造业，各行各业均已认识到 AIGC 的变革能力。在特定专业领域应用场景中，基于与领域更加相关的 AI 训练语料和模型，垂直领域知识服务将更加精准可靠。

在法律领域，LexisNexis 因同时拥有多融合的 LLM 和海量数据形成 Lexis+ AI，辅助法律人员自动起草商业合同、自动生成简报。在医药研发领域，生物医学研究 AI 平台 Causaly 利用医学知识图谱将论文、临床试验、专利、预印本等，实现时间线、思维图或交互式视图等可视化数据，帮助医药研发人员加速对医学知识的获取，将以往需要 2—3 年的研究过程缩短到 2—3 周。除直接提供知识服务外，Elsevier 还推出数据服务，通过 API 接口或平面文件直接支持用户进行文本和数据挖掘以及 AI 训练[31]。

对图书馆而言，目前大多数基于数据、文献或网络信息进行收集、梳理、总结的数据分析、学科咨询、情报研究等传统知识服务都有可能通过 ChatGPT 类工具与 Prompt 机制的结合来很好完成，因此，亟待利用自身专业领域科技文献数据资源以及知识组织体系研发优势，积极参与“专业和垂直”知识系统建设，针对某一具体学科和研究领域的用户需求，开发能够满足实际应用的知识服务系统[14]。

5 图书馆信息资源建设在支撑服务变革中面临的挑战 /Challenges faced by library information resources construction in supporting service transformation

除了深度学习算法模型取得突破外，AIGC 的快速发展同时归功于大规模的语料和必要的算力，海量数据语料库是 AIGC 蓬勃发展的前提、基础和底座。如果缺乏数据的“喂养和训练”，再好的强化算法

技术也无法催生类似 ChatGPT 的 AIGC 产品和服务。相比于互联网中的开源信息资源，图书馆信息资源建设的重点对象——学术期刊、会议录、科技报告等科技文献，具有结构规范、内容真实可靠、蕴含大量知识点等特征，突出了图书馆在 AIGC 应用方面的优势和价值[32]，但是现有资源建设模式仍存在一定局限性，亟待改进优化。

5.1 信息资源建设对象模态有待拓展

期刊、会议论文集、工具书等正式出版物是图书馆资源建设的重点，但是随着信息载体的多元化，其承载的信息量和知识量将有所局限，科学数据、博客、网络百科、开放获取期刊、机构知识库等新的资源形态将逐步具有更高的信息开发与利用价值。单纯以正式出版物为建设对象将无法实现对信息资源的完整保障，也难以适应当前发展需要。以 ChatGPT3 为例，ChatGPT3 的训练数据来源包括互联网爬取网页、图书和维基百科，语料库包含约 5 000 亿个标记（Tokens）[33]。此外，AIGC 处理和生成的对象涵盖从文本到图片、视频等跨模态内容。因此，图书馆需要重视正式学术交流以外的开源信息资源建设，从多模态视角构建立体化信息资源体系。

5.2 本地数据采集管理能力有待加强

图书馆在应用 AIGC 过程中，强调开源信息资源建设并不意味着学术出版物在语料库建设中作用的弱化，特别是期刊论文，具有结构规范、知识性强等优势。例如，英国开放大学知识媒介中心（The Knowledge Media Institute，KMi）的 CORE 研究中心基于 34 万篇开放获取科学论文库推出学术问答应用程序 CORE-GPT，降低生成语言模型由于训练语料库中的事实不正确内容或错误信息而产生虚假答案的危险[34]。再如，国外信息服务商拓展的 AIGC 应用均是以其自身拥有大规模数据为根本前提。当前，大多数图书馆的资源建设路径以购买数字资源网络访问权为主，本地既不掌握资源的元数据和全文，也无法把握用户使用行为数据，同时数据管理能力薄弱，严重制约图书馆服务的更新迭代。

5.3 资源蕴含知识单元有待关联揭示

信息资源建设不仅要充分跟踪收集高质量的信息资源，还需要借助知识组织工具，将隐藏于资源之中的知识单元及其关联关系进行充分的挖掘揭示，形成思维链，实现内容的复杂推理。例如，生物医学研究 AI 平台 Causaly 正是由于嵌入了知识图谱和本体知识系统，使其生成的答案更加精细，同时能够显示元素之间的医疗关系。反观图书馆，图书情报领域经过长期积累形成了一批成熟的知识组织工具和资源，但是由于建设应用机制相对封闭，尚未在图书馆的资源揭示与整合中得到广泛应用，多数图书馆仍然是以编目方式在品种层级对资源进行揭示，少有利用本体等知识组织工具大规模深入到知识单元层级，在很大程度上限制了图书馆开展知识服务。

6 面向 AIGC 应用的图书馆信息资源建设优化策略 /Optimization strategy of library information resources construction for AIGC application

为了适应 AIGC 带来的服务变革，图书馆信息资源建设需要同时涵盖正式出版物，以及互联网上众多的开源信息，无论是从资源规模、资源类型还是组织加工深度上，都对图书馆信息资源建设提出更高要求。因此，图书馆亟待充分认识自身信息资源建设的价值和所面临的挑战，探索如何将馆藏优势转化成多模态多粒度的大规模数据基础，将包括人、机构、项目、社区、活动在内的各类知识对象转换成可计算的知识体系，形成相应的优化发展策略。

6.1 以开源方式推动多模态信息资源建设

开源情报（Open-Source Intelligence，OSINT）是指通过收集、评估和分析公开可用的信息而产生的情报，信息来源同时包括互联网和传统的大众媒体，在文本基础上扩充科学数据、图片、音频、视频等多模态数据。对图书馆而言，为适应学术界和出版界的开放科学进程，资源建设的对象已经从商业订阅资源逐步向开放学术资源拓展。以国家科技图书文献中心为例，从 2014 年启动开放学术资源建设，覆盖开放期刊、会议论文集、学位论文、课件、图书、报告 6 种资源类型。但是，在 AIGC 和 OSINT 的双重视域下，图书馆信息资源建设对象需要向开源信息资源方向进一步拓展，同时基于 OSINT 的分析方法和技术构建智慧服务。

在付费订阅资源和开源信息资源并行的建设模式中，信息资源建设既要保持原有对学术出版物的评估优势，又要针对互联网资源建立信息资源评估遴选体系，并兼顾语料库规模和内容质量要求。相比而言，对正式出版物特别是期刊的评估遴选已有很多成熟的方法和模型，开源信息资源具有多源和多模态特征，其评估遴选则更加复杂多样，需要综合考虑资源类型、资源发布模式等多种因素，建立相应的资

源评估遴选方法和体系。从正向出发，信息资源来源的权威性是遴选开源信息资源的重要指标，例如，以宾夕法尼亚大学发布的《全球智库报告》为线索收集智库报告，就是从发布机构入手跟踪发现高质量开源信息资源的路径。从反向来看，可基于统计[35]、模型[36]和人工[37]方法测试评估AI模型幻觉，进而动态实时地进行数据去噪。

6.2 重视馆藏数字资源权益管理

无论是以全文资源直接作为训练语料还是基于全文进行深度加工，实现数据本地化是图书馆形成高质量语料的前提。随着数字化网络化的普及，数字文献资源已经成为科技领域的主流信息资源，主要科研教育单位均已将数字文献作为自己的主流科技文献资源，并不断削减纸本文献订购。目前，多数图书馆仅仅是拥有数字资源的网络访问权，无论是元数据还是全文数据都不存储在本地，信息资源建设从语料角度切入AIGC应用更是无从谈起。因此，图书馆需要重视元数据、全文的本地保存和使用权利，为后续业务拓展提供更多自主权。

在图书馆与出版商围绕数据的本地保存和使用权利展开谈判的背后，夹杂着复杂的知识产权问题。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对已有作品的大规模学习、模仿、组合和转化可能对原著作权人享有的修改权、保护作品完整权、演绎权等权利造成侵权。同时，如果训练语料中的数据涉及到版权侵犯的情况，生成的文本也可能会面临知识产权的纠纷。此外，与开源信息资源体系建设相对应，开放信息环境下的知识产权问题也更为复杂，这都对信息资源建设提出了更多要求，需要制定相应的知识产权管理策略，规避知识产权风险。

6.3 探索知识库与大语言模型的融合应用

图书情报领域已形成较多成熟的知识组织工具，包括学科分类表、主题词表、各类规范文档、领域本体等，均涵盖精良的知识体系，亟待探索与LLM融合应用焕发新的生机。各类专有名录和词典百科、机构知识库和学者专家库对扩展LLM的知识范围和推理能力具有重要意义，可作为知识库更新的词汇及其语义来源。具体而言，知识组织工具对训练语料的知识增强分为知识内化和知识外用两种路径。知识内化是直接利用知识组织工具构造训练数据，将知识学习到模型参数中。知识外用是将知识组织工具作为外部知识库，将知识库以记忆网络等形式存储供LLM调用，LLM在生成阶段从知识库中检索相关知识。此外，为了满足用户个性化需求，图书馆还需要重视构建用户画像，对用户的历史对话记录、兴趣爱好、搜索行为等信息进行建模和利用，进一步增强LLM输出的知识性，实现更加自适应和个性化的对话服务。

6.4 重塑图书馆数据治理体系

在印本资源建设时期，图书馆基于书目数据构建联合目录系统，结合馆藏借阅服务，在用户发现和使用信息资源方面发挥重要作用。但是，在“以用为主”的数字信息时代，多数图书馆编目业务外包，面对大规模、多模态的数据资源治理需求，图书馆的数据管理能力严重缺位，数据治理体系亟待重构。具体而言，在机制层面需要强化图书馆数据协同治理体系建设，充分发挥不同类型、不同规模图书馆的主观能动性，良好的多元协同治理模式有助于提升数据产出的高质量、数据使用的活跃度、数据价值的创造性。在技术层面，需要充分重视元数据在数据治理中的作用，通过建立统一的元数据标准和清洗规范流程，实现人机协同的多实体、多模态信息资源的规范描述和有序关联，针对特定应用场景进行数据计算和知识挖掘，为服务决策提供更加准确和可靠的支持。同时，用户对话等使用行为数据也在图书馆数据治理的范围内，治理形成的用户画像通过融入模型参数或外部知识库方式，将进一步推动图书馆向数字化、智能化服务升级。

6.5 加强与技术公司的协同合作

AIGC的发展包括基础设施层、模型层和应用层，对应其不同的产业链环节。就图书馆而言，无论是从技术力量还是资金投入角度来看，几乎不可能独立开发和训练大模型，因此，需要基于自身资源优势，加强与AI研发公司合作，融入AIGC的发展，这是图书馆适应当前AI技术快速发展的必然选择。特别是对本地已存储大规模信息资源的图书馆而言，可以以高质量文献资源和科研实体的多维关联抽取为优势，借助技术公司的算力和算法实现共赢。例如，中国科学院文献情报中心在2023年10月24日发布的“科技文献大模型—星火科研助手”正是其与科大讯飞股份有限公司合作的产物，是科技文献资源与通用大模型之间的有效结合。

7 结语 /Conclusion

图书馆的信息资源建设一直与前端的知识内容

生产模式息息相关，无论是数字出版时代带来的资源载体和服务方式变革，还是近几年备受关注的开放科学，都在影响着图书馆的信息资源建设模式，甚至是带来颠覆性的变革。例如，大量开放获取期刊出版之后以采购为主的资源建设模式何去何从，本地数据和使用权益的匮乏导致知识服务成为空中楼阁，这都是针对图书馆信息资源建设提出的时代之问。本文认为，图书馆参与 AIGC 的优势仍然在于资源遴选评估和知识组织，并形成高质量的信息资源甚至是知识资源，只有这样，图书馆才能更好地拓展知识服务，并在火热的 AIGC 应用中彰显自身价值。同时，本文也存在一定不足，不同类型图书馆在资源采集、资源组织方面的能力差异较大，不同类型的图书馆如何分门别类地融入到 AIGC 基础设施层、模型层或应用层，仍需要进一步探讨。

参考文献 /References：

[1] 王飞跃，缪青海 . 人工智能驱动的科学研究新范式 : 从 AI4S 到智能科学 [J]. 中国科学院院刊，2023, 38(4): 536-540. (WANG F Y, MIAO Q H. Novel paradigm of AI-driven scientific research: from AI4S to intelligent science[J]. Bulletin of Chinese Academy of Sciences, 2023, 38(4): 536-540.)

[2] 李白杨，白云，詹希旎，等 . 人工智能生成内容 (AIGC) 的技术特征与形态演进 [J]. 图书情报知识，2023, 40(1): 66-74. (LI B Y, BAI Y, ZHAN X N, et al. The technical features and aromorphosis of artificial intelligence generated content[J]. Documentation, information & knowledge, 2023, 40(1): 66-74.)

[3] 陆伟，刘家伟，马永强，等 . ChatGPT 为代表的大模型对信息资源管理的影响 [J]. 图书情报知识，2023, 40(2): 6-9, 70. (LU W, LIU J W, MA Y Q, et al. The influence of large language models represented by ChatGPT on information resources management[J]. Documentation, information & knowledge, 2023, 40(2): 6-9, 70.)

[4] 赵杨，张雪，范圣悦 . AIGC 驱动的智慧图书馆转型：框架、路径与挑战 [J]. 情报理论与实践，2023, 46(7): 9-16. (ZHAO Y, ZHANG X, FAN S Y. AIGC-driven intelligent library transformation: framework, pathways and challenges[J]. Information studies: theory & application, 2023, 46(7): 9-16.)

[5] 王鹏涛，徐润婕 . AIGC 介入知识生产下学术出版信任机制的重构研究 [J]. 图书情报知识，2023, 40(5): 87-96. (WANG P T, XU R J. Reconstruction of trust mechanism in academic publishing under the intervention of AIGC in knowledge production[J]. Documentation, information & knowledge, 2023, 40(5): 87-96.)

[6] 方卿，丁靖佳 . 人工智能生成内容 (AIGC) 的三个出版学议题 [J]. 出版科学，2023, 31(2): 5-10. (FANG Q, DING J J. Three publishing topics concerning artificial intelligence generated content[J]. Publishing journal, 2023, 31(2): 5-10.)

[7] CHEN X. ChatGPT and its possible impact on library reference services[J]. Internet reference services quarterly, 2023(27): 121-129.

[8] ZOU X Z, SU P, LI L X, et al. AI-generated content tools and students' critical thinking: insights from a Chinese university[J]. IFLA journal-international federation of library associations, 2023(1): 1-14.

[9] 郭亚军，郭一若，李帅，等 . ChatGPT 赋能图书馆智慧服务：特征、场景与路径 [J]. 图书馆建设，2023(2): 30-39, 78. (GUO Y J, GUO Y R, LI S, et al. ChatGPT empowers library smart service: characteristics, scenarios and realization paths[J]. Library development, 2023(2): 30-39, 78.)

[10] 莫祖英，盘大清，刘欢，等 . 信息质量视角下 AIGC 虚假信息问题及根源分析 [J]. 图书情报知识，2023, 40(4): 32-40. (MO Z Y, PAN D Q, LIU H, et al. Analysis on AIGC false information problem and root cause from the perspective of information quality[J]. Documentation, information & knowledge, 2023, 40(4): 32-40.)

[11] 宋士杰，赵宇翔，朱庆华 . 从 ELIZA 到 ChatGPT：人智交互体验中的 AI 生成内容 (AIGC) 可信度评价 [J]. 情报资料工作，2023, 44(4): 35-42. (SONG S J, ZHAO Y X, ZHU Q H. From ELIZA to ChatGPT: AI-generated content (AIGC) credibility evaluation in human-intelligent interactive experience[J]. Information and documentation services, 2023, 44(4): 35-42.)

[12] 朱禹，陈关泽，陆泳溶，等 . 生成式人工智能治理行动框架：基于 AIGC 事故报道文本的内容分析 [J]. 图书情报知识，2023, 40(4): 41-51. (ZHU Y, CHEN G Z, LU Y R, et al. Generative artificial intelligence governance action framework: content analysis based on AIGC incident report texts[J]. Documentation, information & knowledge, 2023, 40(4): 41-51.)

[13] LUND B D, KHAN D, YUVARAJ M. ChatGPT in medical libraries, possibilities and future directions: an integrative review[J]. Health information & libraries journal, 2024(1): 1-12.

[14] 张智雄，于改红，刘熠，等 . ChatGPT 对文献情报工作的影响 [J]. 数据分析与知识发现，2023, 7(3): 36-42. (ZHANG Z X, YU G H, LIU Y, et al. The influence of ChatGPT on library& information services[J]. Data analysis and knowledge discovery, 2023, 7(3): 36-42.)

[15] 赵瑞雪，黄永文，马玮璐，等 . ChatGPT 对图书馆智能知识

服务的启示与思考 [J]. 农业图书情报学报 , 2023, 35(1): 29-38. (ZHAO R X, HUANG Y W, MA W L, et al. Insights and reflections of the impact of ChatGPT on intelligent knowledge services in libraries[J]. Library and information science in agriculture, 2023, 35(1): 29-38.)

[16] LAPPALAINEN Y, NARAYANAN N. Aisha: a custom AI library chatbot using the ChatGPT API[J]. Journal of web librarianship, 2023, 17(3): 37-58.

[17] STEPANOV V K, MADZHUMDER M, BEGUNOVA D D. Exploring the potential of applying the artificial intelligence language model ChatGPT-3.5 in library and bibliographic activities[J]. Scientific and technical information processing, 2023, 50(3): 166-175.

[18] ADETAYO A J. Artificial intelligence chatbots in academic libraries: the rise of ChatGPT [J]. Library hi tech news, 2023, 40(3): 18-21.

[19] LIU X C. Smart library transformation research empowered by AIGC technology[J]. The frontiers of society, science and technology, 2023, 5(8): 34-38.

[20] 李荣 , 吴晨生 , 董洁 , 等 . ChatGPT 对开源情报工作的影响及对策 [J]. 情报理论与实践 , 2023, 46(5): 1-5. (LI R, WU C S, DONG J, et al. Study on the impact of ChatGPT on open source intelligence work and countermeasures[J]. Information studies: theory & application, 2023, 46(5): 1-5.)

[21] 陈超 . 基于大语言模型的生成式 AI 如何赋能 CI[J]. 竞争情报 , 2023, 19(4): 1. (CHEN C. How generative AI based on large language models empowers CI[J]. Competitive intelligence, 2023, 19(4): 1.)

[22] 叶鹰 , 朱秀珠 , 魏雪迎 , 等 . 从 ChatGPT 爆发到 GPT 技术革命的启示 [J]. 情报理论与实践 , 2023, 46(6): 33-37. (YE Y, ZHU X Z, WEI X Y, et al. Enlightenment from the explosion of ChatGPT to the technological revolution of GPT[J]. Information studies: theory & application, 2023, 46(6): 33-37.)

[23] 张晓林 . 从猿到人：探索知识服务的凤凰涅槃之路 [J]. 数据分析与知识发现 , 2023, 7(3): 1-4. (ZHANG X L. From ape to man: exploring the challenges and possible responses of knowledge service[J]. Data analysis and knowledge discovery, 2023, 7(3): 1-4.)

[24] 陈博立 , 鲜国建 , 赵瑞雪 , 等 . 科技文献问答式智能检索总体设计与关键技术探析 [J]. 中国图书馆学报 , 2023, 49(3): 92-106. (CHEN B L. XIAN G J, ZHAO R X, et al. Overall design and key technology of Q&A style intelligent retrieval for scientific and technical literature[J]. Journal of library science in China, 2023, 49(3): 92-106.)

[25] 张新新 , 丁靖佳 . 生成式智能出版的技术原理与流程革新 [J]. 图书情报知识 , 2023, 40(5): 68-76. (ZHANG X X, DING J J. Technical principles and process innovations of generative intelligent publishing[J]. Documentation, information & knowledge, 2023, 40(5): 68-76.)

[26] ChatGPT-like AIs are coming to major science search engines[EB/OL]. [2024-06-25]. https://www.nature.com/articles/d41586-023-02470-3.

[27] 马浩 , 崔运鹏 . 基于混合深度学习模型的科技文献自动综述模型构建研究 [J]. 情报理论与实践 , 2021, 44(9): 176-182, 168. (MA H, CUI Y P. Research on the construction of model for automatic review of scientific literatures based on hybrid deep learning[J]. Information studies: theory & application, 2021, 44(9): 176-182, 168.)

[28] NANBA H, KANDO N, OKUMURA M. Classification of research papers using citation links and citation types: towards automatic review article generation[J]. Advances in classification research online, 2011, 11(1): 117-134.

[29] A new way of exploring dimensions data[EB/OL]. [2024-06-25]. https://www.dimensions.ai/discover-dimensions-ai-assistant/.

[30] Scopus AI: trusted content. powered by responsible AI[EB/OL]. [2024-06-25]. https://www.elsevier.com/products/scopus/scopus-ai.

[31] Elsevier introduces authoritative scientific datasets to fuel innovation and business-critical decisions in life sciences, chemicals and other research-intensive industries [EB/OL]. [2024-06-25]. https://www.elsevier.com/about/press-releases/.

[32] 《ChatGPT 对文献情报工作的影响》研究报告 (简版) 公开发布 [EB/OL]. [2024-06-25]. http://www.las.cas.cn/zhxw/202302/t20230228_6685890.html. (The research report on “the impact of ChatGPT on documentary information work”(short version) was publicly released[EB/OL]. [2024-06-25]. http://www.las.cas.cn/zhxw/202302/t20230228_6685890.html.)

[33] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners[EB/OL]. [2024-06-25]. https://arxiv.org/abs/2005.14165.

[34] CORE-GPT: combining open access research and AI for credible, trustworthy question answering[EB/OL]. [2024-06-25]. https://blog.core.ac.uk/2023/03/17/core-gpt-combining-open-access-research-and-ai-for-credible-trustworthy-question-answering/.

[35] SHUSTER K, POFF S, CHEN M, et al. Retrieval augmentation

reduces hallucination in conversation[C]// Findings of the association for computational linguistics: EMNLP 2021. Punta Cana: Association for Computational Linguistics, 2021: 3784-3803.

[36] HONOVICH O, CHOSHEN L, AHARONI R, et al. Q2: evaluating factual consistency in knowledge-grounded dialogues via question generation and question answering[C]// Proceedings of the 2021 conference on empirical methods in natural language processing. Online and punta cana: association for computational linguistics, 2021: 7856-7870.

[37] LEE N, PING W, XU P, et al. Factuality enhanced language models for open-ended text generation [EB/OL]. [2024-06-25]. https://arxiv.org/abs/2206.04624.

作者贡献说明 /Author contributions：

丁道劲：提出论文选题，设计研究结构，收集资料并撰写；

苏静：讨论研究结构，修改论文。

Research on the Optimization Strategies for Library Information Resources Construction Oriented to the Development of AIGC*

Ding Qiujing[1] Su Jing[2]

[1]Renmin University of China Libraries, Beijing 100872

[2]School of Journalism and Communication, Shaanxi Normal University, Xi'an 710069

Abstract: [Purpose/Significance] The rise of generative AI, represented by ChatGPT, has significantly transformed the model of knowledge production and knowledge service. Libraries are also affected by the transformation and face significant challenges. It is urgent to examine the limitations of the information resource construction and to support the realization of smart services in libraries. **[Method/Process]** From the typical applications of AIGC, this article analyzed its impact on library services. It then explored strategies for optimizing information resource construction in libraries, based on the current situation and existing problems of information resource construction. **[Result/Conclusion]** To adapt to the development of AIGC and promote service change, libraries urgently need to strengthen the construction of open-source information resources, manage the rights and interests of digital collections, explore the path of embedding the knowledge base into LLM, reshape the library data governance system, and strengthen cooperation with technology companies.

Keywords: information resource construction AIGC smart library knowledge service

*This work is supported by the youth program of the National Social Science Fund of China titled "Research on the Optimization Mechanism of Library Knowledge Service Based on Media Convergence" (Grant No. 19CTQ008).

Author(s): Ding Qiujing, associate research librarian, PhD; Su Jing, associate professor, PhD, corresponding author, E-mail: owensujing@163.com.

Received: 2023-11-02 Revised: 2024-02-02 Pages: 23-31