面向生成式AI的高校图书馆数据资源体系构建与实践探索

——以中南大学图书馆为例

袁新辉　崔永*　周芬　王利君　戴前伟

摘要　生成式AI为图书馆服务范式重构提供了技术引擎，但高校图书馆现有数据资源存在描述粗放、标准不一、静态存储、模态单一等问题，难以支撑其深度应用，且既有研究缺乏对底层数据资源体系的系统构建。文章从生成式AI的核心能力出发，分析其数据需求，对照现有数据资源诊断问题，基于数据管理理论构建了“三层架构＋双环驱动”的数据资源体系框架，通过中南大学图书馆地球科学AI知识库和AI馆员项目的实践表明，该框架能够有效支撑生成式AI的应用，聚焦重点学科垂直领域、场景化构建是可行路径。

关键词　高校图书馆　生成式AI　数据资源体系　数据治理

分类号　G250.7

DOI　10.16603/j.issn1002－1027.2026.02.005

引用本文格式　袁新辉，崔永，周芬，等.面向生成式AI的高校图书馆数据资源体系构建与实践探索——以中南大学图书馆为例[J].大学图书馆学报，2026，44(2)：45－52.

AI技术正深刻驱动教育变革与科研范式转型[1]，高校图书馆传统的资源服务模式已难以满足师生对知识深度挖掘与创新应用的迫切需求。在此背景下，生成式AI凭借其语言理解、内容生成与逻辑推理能力[2]，为图书馆服务范式重构、服务效能跃升提供了关键技术引擎[3]。

算力、模型、数据是生成式AI的三大核心要素[4]。其中，数据作为第五大生产要素，是决定大模型效能与输出质量的关键[5]。高校图书馆虽积累了大量资源描述数据、用户行为数据、资源实体数据等数据资产，但存在描述粗放、标准不一、静态存储、模态单一等问题，其数据质量难以支撑生成式AI的深度理解与精准生成需求[6]。《“数据要素×”三年行动计划(2024—2026年)》明确指出要完善数据资源体系，推动科研、文化等领域的高质量数据集建设[7]。因此，完善数据资源体系、强化数据供给，是图书馆实现服务跃升的前提条件，也是对国家数据要素战略规划的响应。

本文旨在解构面向生成式AI的高校图书馆数据需求特征，构建一个分层闭环的数据资源体系理论框架，将分散的数据资源转化为高质量的数据资产，为生成式AI的应用提供可信数据支撑，并通过实践探索验证其可行性。

1　研究现状综述

如何构建高校图书馆的数据资源体系，国内学者的相关研究主要聚焦在以下五个方面。

(1)高质量数据集建设研究。张晓林指出面向人工智能的高质量数据集超越了静态、孤立的“数据集”概念，强调是一个迭代与融汇并进的多层次多维度的高质量数据集体系，这一体系需适应“人工智能＋”场景的复杂性，涵盖基础认知层、场景理解层、行动规划层三个递进层次[8]。徐拥军等指出，人工智能环境下，高质量数据集建设，需以其核心特征和应用需求为逻辑起点，遵循数据生命周期管理理念，依据“数据需求、数据规划、数据采集、数据预处理、数据标注、模型验证”的流程化路径展开[9]。

(2)数据资源重构研究。张谙宁等认为AI生成元数据在图书馆资源建设环境中的应用场景主要有生成描述性元数据、主题标引与分类标引、创建书目

*　通讯作者：崔永，ORCID：0000-0003-3618-5986，邮箱：cuiyong@csu.edu.cn。

记录以及元数据语义增强四种类型，已在图书馆资源建设中展现优化编目工作流程等重要的赋能效应[10]。唐睿等从数据生命周期视角，提出涵盖规划、采集、集成存储、开发、运营等八大阶段的本地化服务体系，强调数据权属可控、价值转化与服务闭环，构建“资源—资产—要素”递进式体系，推动数据从无序资源向可管理资产、可流通要素转化[11]。

(3)数据治理与管理研究。章洁等构建了智慧图书馆 AIGC 数据治理框架，提出合规治理、过程治理、协同治理、生态治理四维机制，覆盖数据采集、训练、应用、保存全生命周期，强调训练数据溯源、用户隐私保护、内容安全审核与伦理规制[12]。任妍等指出 AIGC 技术深度融入图书馆服务场景，使数据治理从传统文献资源管理转向面向生成式 AI 的全生命周期、可信可控、协同共治的新型治理体系，成为保障 AI 应用安全、可靠、高效的核心前提[13]。

(4)服务场景落地和实践研究。董有明等以武汉大学图书馆为例，提出数据驱动的学科情报服务创新路径，依托机构知识库、学术成果数据、学科分析数据，支撑 AI 赋能的学术前沿追踪、投稿推荐、学科对标与成果评价服务[14]。刘江峰等依托 AI 技术对受访读者的相关数据进行深层次的分析和揭示，形成“技术－数据－服务－用户”联动关系，构建人机协同流程来赋能读者服务，提高读者满意度[15]。

(5)数据资源分类研究。杨新涯等基于智慧图书馆建设实践，提出了全数据体系的六类构成：文献元数据、文献内容数据化数据、全面信息管理系统的运行数据、读者行为数据、支持智慧图书馆系统的知识库以及零数据[16]。熊拥军等通过对 C9 联盟高校图书馆的调研，从资源利用视角将图书馆数据资产分为人员、资源和事件三大类，其中事件类数据又细分为利用数据、管理数据和服务数据[17]。王一博等以北京大学图书馆为例，将主题数据划分为用户及其行为数据、资源及其使用数据、服务业务数据、科研成果数据、长期保存数据和图书馆外部数据六类[18]，并强调了基础数据层与主题数据层的分层管理。

综观上述研究，一是研究视角偏重应用层。现有研究多从“AI 能做什么”出发，构想具体服务场景，而对“支撑这些服务需要什么数据”这一基础性问题缺乏系统解构。二是缺乏体系化的数据资源框架。现有研究或聚焦于数据治理环节，或局限于数据分类解构，尚未从生成式 AI 全流程应用的高度，构建一套覆盖数据汇聚、治理、服务的标准化资源体系框架。

2 面向生成式 AI 的数据需求分析

2.1 生成式 AI 的核心能力及其对图书馆数据资源的要求

生成式 AI 的核心在于从大规模数据中学习模式和分布，生成与真实世界内容相似的全新原创输出[19]，它具备语言理解、逻辑推理、内容生成三大核心能力。语言理解能力使 AI 能够洞察用户提问的真实意图。逻辑推理能力使得 AI 可以进行多步推理、问题拆解和因果推断。内容生成能力可使 AI 依据指令创造新内容。近年来，多模态能力的进展使得 AI 能处理文本、图像、视频等多种类型数据，进一步拓展了生成式 AI 的应用边界。这些能力对图书馆数据资源提出了以下五个方面的要求。

(1)数据构成的融合性。生成式 AI 的感知与认知建立在多源数据融合理解之上，这要求图书馆数据必须突破单一形态，将文本、图像、音频、视频、结构化数据等深度融合。

(2)数据揭示的细粒度。生成式 AI 的精准服务要求数据揭示必须从传统的资源描述下沉到内容中实体及其关系的微观揭示。

(3)数据组织的可计算性。指通过将细粒度知识单元转化为机器可识别的表示形式(如向量、知识图谱等)，使生成式 AI 能够有效理解、处理并利用数据，从而支撑语义搜索、知识推理等智能应用[20]。

(4)数据流动的实时性。生成内容的准确性受到数据时效性限制，要求数据供给从静态快照转向动态数据流，建立反馈闭环。

(5)数据质量的高标准与强规范性。数据质量是决定 AI 模型可信度与输出有效性的核心前提，且低质量数据会引发幻觉现象[21]，这要求图书馆必须建立贯穿数据全生命周期的质量控制标准。

2.2 高校图书馆数据资源现状分析

经过长期信息化建设，高校图书馆已积累了丰富的数据资源，从功能角色出发，可以将这些数据资源归纳为以下六类。(1)基础数据。描述业务核心实体的基本信息，如读者信息、馆藏基础信息、机构信息等。(2)资源描述数据。记录馆藏具体资源的外部特征，如 MARC 书目记录、电子资源元数据、机构知识库元数据等。(3)资源实体数据。资源所承载的知识信息全文，如本校学位论文、机构知识库成

果、自建特色资源等。(4)行为数据。用户与图书馆系统交互的轨迹记录，反映了用户需求偏好与使用习惯。(5)业务管理数据。图书馆日常运营产生的管理类数据，如采购清单、经费数据、空间预约数据等。(6)外部关联数据。从馆外获取的参考性数据，如分类法、主题词表、规范文档、核心期刊目录等。

对照上述生成式AI的要求，图书馆现有数据资源主要存在以下四个方面的不足：一是数据停留在描述层面，缺乏内容理解能力，难以支持构建新型服务模式，如“语义理解—检索增强—生成式推荐”[22]的检索服务。二是标准不一，如来源于不同信息系统的行为数据、不同厂商的资源描述数据，其数据结构不同。三是数据以静态存储为主，缺乏实时交互能力，用户反馈渠道有限且数据难以利用[23]，无法形成数据优化服务的良性循环。四是数据以单一模态为主，缺乏多模态融合能力，难以支撑生成式AI的多模态理解与生成需求。

2.3 面向生成式AI的图书馆数据资源体系变革

基于上述差距分析，要满足生成式AI的应用需求，高校图书馆的数据资源体系必须在以下两个层面进行系统性变革。

2.3.1 现有数据的治理

在持续开展传统数据治理工作的基础上，面向生成式AI的应用场景，现有图书馆数据资源以下三个方面的治理尤为迫切。

在资源实体数据方面，受版权限制，图书馆不直接存储大规模商业数据库的全文，主要通过元数据管理访问入口，这限制了生成式AI对全文内容的深度利用。图书馆只能通过版权协商、机构知识库建设等方式，逐步扩大可本地化存储的资源实体数据范围。

在行为数据方面，由于数据分散在不同业务系统中，需要建立统一的行为数据采集标准，打通借阅、检索、下载、咨询等各环节数据，形成完整的用户交互轨迹。

在外部关联数据方面，依赖外部采购或开放获取，与馆藏数据的关联需要人工维护，数据时效性和一致性难以保证。需要建立自动映射机制，并探索与数据商、开放数据平台的规范合作。

2.3.2 面向生成式AI的增强数据建设

为满足生成式AI对数据细粒度、可计算性等高阶需求，图书馆必须在现有数据基础上，通过语义标注、知识抽取、实体对齐等深度加工，构建全新的面向AI的增强数据。这类数据与原始数据的核心区别在于：原始数据是对资源的描述(如题名、作者、)或原始记录(如借阅日志、检索词)，而增强数据是对知识的组织和表达(如“这篇论文的创新点是什么”)。具体而言，增强数据包括细粒度语义标注数据、知识图谱、用户画像数据以及FAQ知识库等。增强数据使AI能够“读懂”资源内容、理解用户需求、进行知识推理，是实现生成式AI可信服务的核心“燃料”。

3 面向生成式AI的高校图书馆数据资源体系构建

3.1 理论依据

(1)数据生命周期理论。该理论强调数据从采集到销毁的全过程管理，治理应贯穿整个生命周期，支撑图书馆数据体系构建[24]。

(2)数据中台理论。该理论的核心是“分层解耦”，通过数据汇聚、治理、服务、应用的层次化架构，实现多源异构数据接入统一数据资产池，提供高效的数据服务[17]。

(3)数据安全与合规理论。该理论强调数据治理必须贯穿隐私保护、知识产权合规、访问控制等安全要素，形成安全管理、运营、规范保障等多系统融合的运行机制[25]。

3.2 数据资源体系框架构建

基于以上的需求分析以及相关理论指导，本文提出“三层架构＋双环驱动”的数据资源体系框架，如图1所示。框架自下而上分为数据汇聚层、知识组织层、智能服务层三个层次。

3.2.1 数据资源汇聚层

数据汇聚层是数据资源体系的“原材料仓库”，负责实现全域数据的广泛接入与本地化存储，汇聚了六类数据资源。数据汇聚层的关键能力在于连接的广泛性与接入的灵活性。需要支持数据库同步、接口调用、文件导入、网络爬虫等多种采集方式，并兼顾实时采集与离线批处理两种时效性需求。同时，本层需遵循数据安全与隐私保护原则，对敏感信息(如读者身份信息)进行脱敏处理。

3.2.2 知识组织层

知识组织层是数据资源体系的“核心加工厂”，负责将原始数据转化为高质量、语义化的数据资产，核心任务是加工面向AI的增强数据。各功能模块遵循的逻辑顺序如图1所示，各步骤的主要任务如下。

(1)数据清洗与融合。解决多源数据的一致性问题，包括格式转换、字段映射、重复记录合并、缺失值处理等。例如，将不同系统中同一读者的记录合并；将不同数据库商的元数据格式转换为统一标准。

(2)知识抽取与标注。利用自然语言处理、计算机视觉等技术，从非结构化内容中抽取实体、关系、事件，形成细粒度知识单元。

(3)实体对齐与消歧。识别并合并来自不同数据域的同一实体(如图书、学者、机构)，建立统一标识符。例如，将“莎士比亚”与“Shakespeare，William”识别为同一学者；将同一种文献在不同系统中的记录关联起来。本步骤依赖于知识抽取的成果。

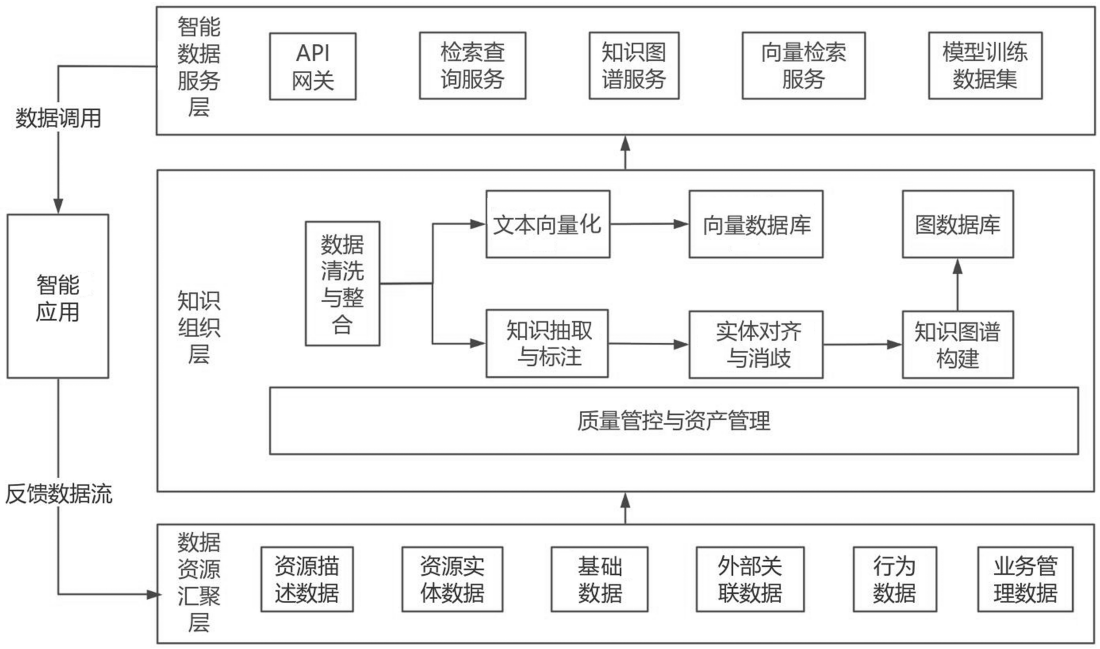

图 1　面向生成式 AI 的高校图书馆数据资源体系总体框架示意图

(4)文档向量化。在知识抽取后，将资源实体数据全文、语义标注数据等转化为向量表示，构建向量索引库。这一处理支撑检索增强生成(RAG)，使用户提问能够通过语义相似度匹配最相关的知识片段。

(5)知识图谱构建。将抽取的知识进行关联、融合，形成机器可理解、可推理的语义网络。知识图谱是支撑逻辑推理的核心数据形态，通过“实体－关系－实体”的三元组结构，使 AI 能够在不同知识节点之间进行关联推理。本步骤依赖于实体对齐与消歧的成果。

(6)质量管控与资产管理。对数据资产进行分级分类、质量稽核、价值评估，建立数据资产目录。

3.2.3　智能数据服务层

本层是数据资源体系的价值输出端口，提供易于调用、安全可控的数据服务，核心数据服务功能如表 1 所示。

表 1　智能数据服务层核心服务功能表

数据服务类型	功能描述
统一 API 网关	提供统一的 API 访问入口，负责请求路由、身份认证、权限校验、流量控制、调用监控等功能，保障数据服务的安全性与可管理性
检索查询服务	支持基于元数据的传统检索(如题名、作者、关键词的精确/模糊匹配)以及基于倒排索引的全文检索，满足结构化查询需求
知识图谱服务	提供知识图谱的查询接口，支持实体检索、关系路径发现、关联推理等功能，赋能智能问答、学科前沿分析等场景
向量检索服务	提供基于嵌入模型的语义向量检索能力，通过计算向量相似度实现语义层面的内容匹配，是检索增强生成(RAG)技术的核心支撑
模型训练数据集	提供经过清洗、标注、脱敏的高质量数据集，供上层 AI 模型进行微调或预训练，提升模型在垂直领域的表现

3.3 双环驱动的优化机制

智慧图书馆应具备自适应学习与持续优化的能力，因此，数据资源体系不是静态的存储架构，而是具有生命力的动态系统。其核心驱动力在于两个层次分明、相互支撑的闭环。

3.3.1 治理闭环（数据质量保障环）

治理闭环旨在保障数据质量的持续提升。原始数据从汇聚层进入知识组织层后，经过清洗融合、知识抽取、实体对齐等加工步骤，形成初步的数据资产。随后进入质量审计环节，系统依据预定义的质量规则，自动检测数据中的格式错误、实体歧义、关键属性缺失等问题。将审计发现的问题反馈至数据采集和清洗环节，触发对数据接入标准、采集策略或清洗规则的优化调整。通过“采集－加工－审计－反馈－优化采集”的闭环迭代，使数据质量得以持续改善。

3.3.2 价值闭环（服务优化环）

价值闭环旨在实现服务的持续优化。经过治理层加工的高质量数据资产，通过智能数据服务层提供检索、图谱、向量等数据服务，支撑上层的智能应用（如AI馆员、学科知识库）。在应用运行过程中，系统实时采集用户交互行为并将反馈数据回流至数据汇聚层。随后，这些数据进入知识组织层，用于分析知识库覆盖盲点、回答偏差、用户偏好等，并据此优化知识抽取规则、更新语义标注数据或调整知识图谱结构。通过“治理－服务－应用－反馈－优化治理”的闭环迭代，使服务效果越用越精准。

4 中南大学图书馆的实践探索

生成式AI的应用对数据的需求广泛且复杂，涉及多个层面的数据关联，构建完备的数据资源体系需要长期的投入与积累。鉴于此现实情况，中南大学图书馆的探索实践遵循“场景切入、技术试点、逐步沉淀”的路径，主要基于两方面的考量：其一为数据需求驱动，生成式AI对语料库的规模和质量有着极高要求，集中资源优先攻克重点学科垂直领域相较于全面推进更为可行；其二是应用场景驱动，智能咨询、语义化检索等场景与生成式AI的对话交互、语义理解能力高度适配，能够迅速验证技术效果、积累实践经验。

4.1 基础建设

长期以来，中南大学图书馆高度重视数据资源建设，在组织机制、数据资源积累以及学科资源保障等方面已奠定一定基础，为后续开展生成式AI应用探索提供了必要支撑。

4.1.1 组织保障与元数据规范

（1）成立跨部门大数据工作小组。组建由技术服务、资源建设、读者服务等部门骨干构成的图书馆大数据工作小组，建立常态化协同机制，推动元数据标准建设以及在业务流程中落地。

（2）构建图书馆特色元数据标准。启动核心元数据规范制定工作，以国际通用标准为基础，结合本馆特藏资源、机构知识库及学科服务需求，对图书、期刊论文、学者、机构等核心实体的元数据元素进行扩展与细化，特别加强了对细粒度知识单元描述与关联的标准要求，以提升数据的语义可计算性。

（3）融入学校信息化整体。图书馆积极参与学校“一张表工程”等信息化项目建设，推动相关元数据标准、机构知识库等数据纳入学校数据交换共享平台统一规范管理，依规范从学校数据交换共享平台获取师生信息、学科信息等基础数据，为跨系统数据融合奠定基础。

4.1.2 数据基础

针对数据主权与安全可控的需求，图书馆构建了数据资源本地化服务实施策略体系，积极开展本地化数据中心建设。通过整合服务平台中分散的各类数据，将高价值的用户行为数据、资源利用日志和本地化数字资源进行统一汇聚，并存储于本地数据中心。例如通过电子资源统一服务平台、新一代图书馆服务平台等，广泛、合规地采集资源使用数据；依托机构知识库、查收查引一站式服务平台，持续收集中南大学师生的论文、被引频次、期刊分区等数据。

为确保电子资源元数据的高价值与针对性，中南大学图书馆构建了以学科导向为原则的电子资源保障体系。通过期刊学科分类关联、学科发文趋势分析及师生推荐等途径，优化资源采购决策，重点提升本校优势学科的资源保障率，并在合同中要求数据库商提供元数据。

目前，图书馆本地化数据中心已初步汇聚了包括馆藏书目数据、电子资源元数据、用户借阅与检索日志、机构成果元数据、中南大学学位论文元数据等在内的多源数据，形成支撑后续数据治理与智能应用的基础数据集。

4.2 实践探索

中南大学图书馆优先选择需求明确、数据基础较好的场景和领域进行试点，主要开展了地球科学 AI 知识库和 AI 馆员两个项目，初步验证了体系框架的可行性。地球科学 AI 知识库项目重点验证了数据汇聚层、知识组织层、智能服务层三层框架与治理闭环的可行性，而 AI 馆员项目则验证了价值闭环可行性。

4.2.1 地球科学 AI 知识库

地球科学 AI 知识库项目是面向生成式 AI 在垂直学科领域的深度应用，由中南大学图书馆与学校二级学院深度合作共同构建。该知识库通过高质量的垂直领域语料库与 RAG 技术相结合，为通用大模型注入领域知识，有效缓解生成式 AI 在专业问答中的幻觉问题，确保输出的可信度与准确性。

首先，在数据汇聚层收集了超过 1000 种地球科学相关图书、超过 3 万篇中文期刊论文以及近 6000 篇外文期刊论文，涵盖了资源描述数据和资源实体数据。

其次，在知识组织层形成了地球科学知识库。重点开展了以下工作：一是对学术论文全文进行细粒度知识抽取，识别研究方法、实验数据、核心观点、技术路线等关键知识单元；二是对多模态资源进行语义化处理，包括地质图、岩心照片、地理信息数据等，初步构建了跨模态知识的理解与关联能力；三是构建贯穿全流程的质量监督机制，对数据准确性、一致性进行校验。

最后，在智能服务层采用 RAG 机制抑制生成式 AI 的幻觉问题。当用户通过自然语言提问，系统首先在知识库中进行语义检索，得到最相关的知识片段，然后将这些片段与问题一起提交给大模型进行总结生成。这种“先检索，后生成”的模式，确保了答案有据可循，实现了可信、可控的智能问答。

地球科学 AI 知识库已对接中南大学网络教学平台，教师在建设智慧课程时可直接调用，用于知识梳理、辅助教学、题库构建等，学生可以直接利用该知识库开展学习。

4.2.2 AI 馆员

在引入 AI 馆员之前，中南大学图书馆主要借助智能问答系统、电子邮箱等途径开展线上参考咨询服务，在智能化水平与时效性方面表现欠佳，亟待改善。AI 馆员具备自然语言理解与多轮对话交互能力的显著优势，可针对开放性问题生成个性化解答，其服务模式更契合用户的沟通习惯，能够有效提升服务体验。与地球科学 AI 库比较，其数据需求更容易得到满足，能迅速彰显生成式 AI 的应用成效。

AI 馆员的初始知识库主要来源于两类数据：一是馆务数据。通过自动采集图书馆官方网站发布的内容、对接馆藏书目系统等方式，获取开放时间、借阅规则、空间预约流程、常见业务办理指南等数据。数据经人工审核入库，进行知识组织处理，作为基础知识库。二是 FAQ 问答对数据。从历史参考咨询记录中提取高频问题及标准答案，经过馆员筛选、分类、归集后形成 FAQ 知识库，为 AI 馆员提供高质量的训练素材。

AI 馆员依托图书馆门户网站、微信公众号、线下大屏提供 7×24 小时咨询服务，同时对接了馆藏书目系统，提供语义化的书目检索服务。在服务运行过程中，系统自动记录每一次用户交互数据，为优化知识库提供基础。每周由馆员团队对交互记录进行统计分析，识别高频但回答不准确的问题、知识库覆盖盲点、用户满意度偏低的问题，补充或修改知识库内容；同时，对于用户反复提问但当前知识库无法覆盖的问题，形成新的 FAQ 问答对数据。这一机制形成了“数据驱动服务－服务产生数据－数据反哺优化”的持续迭代，体现了框架中的价值闭环。AI 馆员自 2025 年 3 月上线以来，服务师生 2 万余人次，知识库的命中率大幅提升。后期将探索使用 AI 技术实现自动筛选、分类、归集、优化，最后由馆员审定的自动化优化闭环。

5 总结与展望

生成式 AI 的语义理解、内容生成、逻辑推理与多模态处理等核心能力，为图书馆服务范式重构、服务效能跃升提供了技术引擎，现有数据资源存在描述粗放、标准不一、静态存储、模态单一等问题，需要通过完善数据资源体系、加强数据供给支撑其深度应用。针对既有研究“重应用、轻基础”的缺口，本文系统解构了面向生成式 AI 的数据需求，提炼出细粒度、可计算性、融合性、实时性、高标准强规范性的数据需求特征，构建了“三层架构＋双环驱动”的数据资源体系框架，为生成式 AI 的应用提供可信、可控的高质量数据支撑，并结合中南大学图书馆的实践进行了验证。研究表明，数据资源体系能够有效支

撑生成式AI的应用，聚焦重点学科垂直领域、场景化构建是可行路径。需要指出的是，本文所提到的实践尚处于探索验证阶段，案例仅覆盖地球科学学科和AI馆员场景，更大数据规模及更多场景应用有待进一步检验。

基于生成式AI对数据需求的规模性要求以及图书馆的实践探索，建议未来可从以下方向深化探索。

(1)赋能科技查新与知识产权信息服务。融入学校科研创新全流程服务体系，通过生成式AI的语义解析与知识推理能力，为科研项目从立项查新、专利布局到成果转化提供智能化支撑。

(2)构建数据共建共享联盟。单个图书馆数据资源有限，难以支撑大规模AI应用。未来应探索区域性或行业性图书馆数据联盟，推动元数据、特色资源、语义标注数据等核心资产的共建共享。

(3)融入学校AI整体布局。大模型训练对算力、平台的要求远超图书馆自身能力。未来应主动对接学校AI规划，将数据资源体系嵌入学校平台，聚焦图书馆最擅长的数据治理与场景应用环节，形成“学校建平台、图书馆供数据、师生享服务”的分工协作模式。

总之，面向生成式AI的高校图书馆数据资源体系建设是一项需要持续探索的长期工程。本文只是一个起点，期待更多同行共同推动图书馆从“资源供给”向“智慧赋能”的范式跃迁。

参考文献

1 国务院. 国务院关于印发新一代人工智能发展规划的通知[EB/OL]. [2026－03－10]. https://www.gov.cn/gongbao/content/2017/content_5216427.htm.

2 曹树金. 生成式AI在情报领域的应用及效果[J]. 情报资料工作，2023，44(5)：5.

3 童云海，陈建龙. DeepSeek热潮下的双重变革：大模型的技术革新与高校图书馆服务范式的重构[J]. 大学图书馆学报，2025，43(1)：66－70.

4 世界互联网大会人工智能工作组. 发展负责任的生成式人工智能研究报告及共识文件[R/OL]. https://cn.wicinternet.org/static/pdf/发展负责任的生成式人工智能－中文版.pdf.

5 朱文凤. 信通院王志勤：三要素螺旋迭代，转动AI增长“飞轮”[EB/OL]. [2026－03－10]. https://www.iitime.com.cn/html/10187/10592225.htm.

6 曹树金，石佳，张君宜. “十五五”时期图书馆的生成式AI战略[J]. 图书馆论坛，2025(10)：1－10.

7 国家数据局，中央网信办，科技部，等. 十七部门关于印发《“数据要素×”三年行动计划(2024—2026年)》的通知[EB/OL]. [2026－03－10]. https://www.cac.gov.cn/2024－01/05/c_1706119078060945.htm.

8 张晓林. “人工智能＋”背景下的高质量数据集建设：图书馆的机遇与挑战[J]. 中国图书馆学报，2025，51(6)：4－17.

9 徐拥军，张群群，傅予，等. 哲学社会科学高质量数据集的核心特征、应用需求与建设进路[J]. 图书情报知识，2025，42(6)：6－15，27.

10 张谙宁，叶兰，周文琦，等. AI生成元数据赋能图书馆资源建设的实践与启示——基于国内外案例调查[J]. 大学图书馆学报，2025，43(4)：90－104.

11 唐睿，罗孟儒，崔永. 基于数据生命周期的高校图书馆数据资源本地化服务策略研究[J]. 图书馆学研究，2025(7)：104－114.

12 章洁，洪芳林. 迈向有序AI：智慧图书馆AIGC数据治理机制与策略[J]. 图书馆建设，2025(2)：119－131.

13 任妍，段涛，付玉，等. AIGC时代信息资源管理领域可信数据空间的构建路径：主体比较与策略研究[J]. 情报资料工作，2026(2)：62－68.

14 董有明，马浩琴. 数据驱动的高校图书馆学科情报服务创新——武汉大学图书馆的实践探索[J]. 大学图书馆学报，2025，43(3)：15－23.

15 刘江峰，胡志伟，董焕晴，等. 生成式AI赋能图书馆读者服务体验的影响因素分析[J]. 图书馆学研究，2025(12)：100－107.

16 杨新涯，文佩丹，卓应忠. 智慧图书馆的全数据体系研究[J]. 图书情报工作，2023，67(13)：29－35.

17 熊拥军，白瀚祯，张廷成. 基于数据中台的图书馆数据资产管理架构[J]. 图书馆学研究，2023(8)：36－47.

18 王一博，刘丹，张俊娥，等. 高校图书馆数据增值服务体系构建研究——以北京大学图书馆为例[J]. 图书馆杂志，2024，43(8)：62－71.

19 Mladan J，Mark C. Generative artificial intelligence：trends and prospects[J]. Computer，2022，55(10)：107－112.

20 成舒云，朱宏智，柯岚馨. AIGC赋能高校图书馆数据资源开发路径研究[J]. 情报科学，2025，43(11)：157－165.

21 付跃安. 大语言模型在图书馆应用中的幻觉问题与应对策略[J]. 四川图书馆学报，2026(1)：9－15.

22 郭利敏，刘悦如，付雅明. 从OPAC到GPAC：生成式人工智能重构图书馆目录系统的路径研究[J/OL]. 图书馆杂志，1－18[2026－03－13]. https://link.cnki.net/urlid/31.1108.g2.20251216.1710.002.

23 Li Y，Erjiang E，Tian X. The user experience of university library：a text mining analysis of a Q&A platform in China[J]. Library & Information Science Research，2024，46：101326.

24 卢小宾，洪先锋，蒋玲. 智慧图书馆数据标准体系研究[J]. 图书情报知识，2021(4)：50－61.
25 张春春，孙瑞英. 智慧图书馆用户数据合规治理机制研究[J]. 图书情报工作，2024，68(4)：15－26.

作者贡献说明：
袁新辉：论文选题、框架设计及修改
崔永：论文框架设计、撰写及修改
周芬：文献调研及部分内容撰写
王利君：文献调研及部分内容撰写
戴前伟：组织及论文审核

作者单位：中南大学图书馆，湖南长沙，410083
手稿日期：2026年3月5日
修回日期：2026年3月15日
（责任编辑：支娟）

Construction and Practical Exploration of a Generative AI-Oriented Data Resource System in Academic Libraries: A Case Study of Central South University Library

YUAN Xinhui　CUI Yong　ZHOU Fen　WANG Lijun　DAI Qianwei

Abstract: Generative artificial intelligence (AI) offers a powerful technological engine for reconstructing library service paradigms, yet existing data resources in academic libraries remain constrained by coarse-grained description, inconsistent standards, static storage, and limited modality, which hinders their capacity to support advanced AI applications. Moreover, the systematic construction of underlying data resource systems remains unaddressed. This study addresses this gap by proposing a comprehensive framework for generative AI-oriented data resource systems, providing high-quality data infrastructure for academic library services. To achieve this, the research first analyzes the core capabilities of generative AI—language understanding, content generation, logical reasoning, and multimodal processing—and identifies five key data requirement features: fusion, fine granularity, computability, real-time responsiveness, and high quality with rigorous standardization. These requirements are then systematically mapped against six categories of existing library data resources, including basic data, resource description data, resource entity data, behavioral data, operational management data, and external data. The study diagnosed four major limitations: inadequate content-level understanding, heterogeneous and inconsistent standards, static data structures lacking real-time interaction, and insufficient multimodal integration. Based on data lifecycle theory, data middle platform theory, and data security and compliance theory, the study designs a "three-layer architecture with dual-loop drive" framework. The three layers consist of a data convergence layer for integrating multi-source heterogeneous data, a knowledge organization layer for processing raw data into AI-enhanced data through cleaning, knowledge extraction, entity alignment, document vectorization, and knowledge graph construction, and an intelligent data service layer that provides API gateways, retrieval services, knowledge graphs, vector-based retrieval, and training datasets. The dual-loop mechanism includes a governance loop for continuous data quality improvement and a value loop for service optimization via user feedback. The framework is validated through two practical projects at Central South University Library—the Earth Science AI Knowledge Base, which demonstrates the functionality of the three-layer architecture, and the AI Librarian System, which exemplifies the operation of the value loop. The results confirm that the proposed data resource system effectively supports generative AI applications. The study concludes that a data resource system can effectively support generative AI applications, and focusing on key disciplinary domains and adopting scenario-based approaches is a feasible path for advancing AI-oriented data infrastructure in academic libraries.

Keywords: Academic Libraries; Generative AI; Data Resource Systems; Data Governance