

人工智能技术在高校档案信息化建设中的应用研究

王裕茹

北京联合大学应用文理学院, 北京

收稿日期: 2026年7月14日; 录用日期: 2026年7月28日; 发布日期: 2026年8月18日

摘要

人工智能技术的快速发展为高校档案信息化建设提供了新的技术路径与赋能机制。当前, 高校档案工作正面临从传统管理模式向智能化、知识化、服务化转型的关键时期。本文聚焦自然语言处理、计算机视觉、知识图谱等核心人工智能技术, 系统分析其在档案文本挖掘、图像语义解析、知识关联构建及智能服务等方面的应用场景。研究表明, 人工智能技术能够显著提升档案信息化的自动化水平、知识组织能力与服务响应效率, 推动档案资源从静态存储向动态知识服务转型。然而, 技术落地仍面临数据标准化不足、领域知识建模滞后、安全审核机制不完善及复合型人才短缺等现实挑战。为此, 高校需强化顶层设计, 推进档案数据治理体系建设, 优化开放审核机制, 培养“档案 + AI”复合型人才, 构建适应智能化发展的高校档案信息化生态。

关键词

人工智能, 高校档案, 信息化建设, 知识服务, 档案治理

Research on the Application of Artificial Intelligence Technology in University Archives Informatization Construction

Yuru Wang

College of Arts and Science of Beijing Union University, Beijing

Received: July 14, 2026; accepted: July 28, 2026; published: August 18, 2026

Abstract

The rapid development of artificial intelligence technology has provided new technical paths and

enabling mechanisms for the informatization construction of university archives. Currently, university archives are facing a critical period of transformation from traditional management models to intelligent, knowledge-based, and service-oriented systems. This paper focuses on core AI technologies such as natural language processing, computer vision, and knowledge graphs, systematically analyzing their application scenarios in archival text mining, image semantic analysis, knowledge association construction, and intelligent services. Research shows that AI technology can significantly improve the automation level, knowledge organization capability, and service response efficiency of archival informatization, promoting the transformation of archival resources from static storage to dynamic knowledge services. However, technological implementation still faces practical challenges such as insufficient data standardization, lagging domain knowledge modeling, imperfect security auditing mechanisms, and a shortage of interdisciplinary talents. Therefore, universities should strengthen top-level design, promote the construction of archival data governance systems, optimize open auditing mechanisms, cultivate “archives + AI” interdisciplinary talents, and build an intelligent university archives informatization ecosystem.

Keywords

Artificial Intelligence, University Archives, Informatization Construction, Knowledge Services, Archives Governance

Copyright © 2026 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. 引言

在“人工智能+”战略深入推进的背景下，人工智能技术正加速与教育、文化、档案等公共服务领域深度融合。高校档案作为学校历史记忆、知识资产与文化遗产的重要载体，其信息化建设水平直接影响档案资源的开发利用效率和服务能力。然而，传统高校档案信息化建设多聚焦于数字化扫描、目录录入、存储管理等基础层面，缺乏对档案内容语义的理解与知识关联的构建，导致“数据孤岛”现象突出、服务方式单一、用户满意度不高问题日益凸显。近年来，《国家档案局办公室印发〈关于推进人工智能在档案行业应用的意见〉的通知》¹明确提出要以人工智能赋能档案事业高质量发展。

在学术研究层面，学界已从多个维度对人工智能与档案工作的融合展开探讨。郑志荣系统阐述了人工智能赋能档案事业高质量发展的战略方向与政策路径，强调技术应用需与档案治理现代化协同推进[1]。黄霄羽、李沐紫从国际比较视角出发，梳理了“人工智能+”背景下国内外档案利用服务的发展态势，指出智能检索、知识发现与个性化服务正成为档案利用领域的共性趋势[2]。苏雯基于四个典型案例，考察了数智技术在非物质文化遗产档案活化利用中的实践路径，为档案资源的创造性转化与创新性发展提供了参考[3]。赵爱华则聚焦教育档案资源，探讨了面向信息时代公民素养提升的档案开发利用策略，拓展了档案服务社会教育的功能边界[4]。上述研究为人工智能与档案领域的深度融合奠定了重要基础，但多侧重于宏观政策、行业态势或特定类型档案的应用分析，针对高校档案信息化建设中人工智能技术系统性嵌入的研究仍相对薄弱。

在此背景下，如何将人工智能技术系统性地嵌入高校档案信息化建设，推动档案管理从“被动保管”向“主动服务”转型，成为亟需研究的重要课题。本文立足技术赋能视角，借鉴农业类高校图书馆信息

¹<https://www.saac.gov.cn/daj/tzgg/202604/12aa9fab60c4dbeab5444f03e78bdb9.shtml>

化建设与高校档案编研智能化转型的研究成果,聚焦自然语言处理、计算机视觉、知识图谱三大核心技术,系统探讨人工智能在高校档案信息化建设中的核心应用场景、现存问题及实施路径,以期为高校档案工作的智能化转型提供理论与实践参考。

2. 人工智能技术概述

人工智能是计算机科学、信息科学、认知科学等多学科交叉的前沿领域,其核心是通过算法模型使机器具备感知、学习、推理、决策等智能行为。当前,深度学习、自然语言处理、计算机视觉、知识图谱等技术已日趋成熟,并在图像识别、语义理解、知识组织等方面取得显著成效[5][6]。特别是近年来,大语言模型的突破性进展,进一步拓展了人工智能在文本生成、语义理解与知识问答等领域的应用边界,为档案资源的智能化开发利用提供了新的技术可能[7]。

在高校档案信息化建设中,人工智能技术的价值主要体现在三个层面。一是通过自然语言处理实现档案文本的自动分类、实体识别与主题标引。研究表明,命名实体识别技术能够精准捕捉档案中的人名、机构名、地名等关键实体,结合语义分析构建实体间的关联网络,显著提升档案文本的深度挖掘能力[8]。二是借助计算机视觉完成图像档案的文字识别、场景分类与破损修复。计算机视觉技术通过光学字符识别自动提取图像中的文字信息,人脸识别辅助标注人物身份,图像分类技术按主题自动归档,有效激活图像与视频档案的利用价值[9]。三是利用知识图谱构建档案实体之间的语义关联,支撑智能化检索与知识服务。知识图谱技术通过语义化组织档案资源,将人物、机构、事件、成果等实体及其关系转化为网状知识模型,突破传统档案管理的线性叙事局限,实现从碎片化信息到结构化知识的跃迁[10]。

需要指出的是,高校档案工作具有专业性强、语义密度高、安全性要求严等特点,通用人工智能模型难以直接迁移应用,必须在领域知识建模、数据治理、安全审核等方面进行针对性优化[11]。例如,在档案分类与标引任务中,通用文本分类模型难以准确区分“教学研究”与“教学管理”等语义相近的档案类别;在命名实体识别中,通用模型对高校特有的机构名称、项目类型、职称序列等实体的识别准确率较低。因此,构建面向高校档案领域的本体体系与知识图谱,加强档案就绪资源建设,是实现人工智能技术深度赋能的基础前提[12]。此外,人工智能在档案领域的应用还需兼顾效率与安全的平衡,在智能检索、自动编研等服务场景中嵌入可解释性与合规追溯机制,确保技术应用符合档案工作的职业伦理与安全要求[13]。

3. 高校档案信息化建设的特点与现状

3.1. 建设特点

高校档案信息化建设呈现出鲜明的特征。一是档案类型多样,涵盖行政公文、教学档案、科研档案、学生档案、人事档案、校史资料等多模态资源。二是服务对象广泛,既包括校内师生、行政管理人员,也包括社会研究者、校友等外部用户。三是知识密度高,档案内容涉及学科发展、学术传承、人物关系、制度变迁等深层语义信息。四是安全要求严格,部分档案涉及个人隐私、科研成果、保密信息等敏感内容。

3.2. 发展现状

当前,大多数高校已完成档案数字化基础建设,建立了档案管理系统或档案信息平台,实现了目录级检索与原文浏览功能。然而,正如李秀华所指出的,当前高校档案工作仍以数字化扫描和目录录入等基础建设为主,信息化水平尚处于初级阶段,尚未实现内容的深度语义解析与知识关联[14]。廉谨嫣也强调,数据治理滞后于技术发展,跨系统数据壁垒严重制约了档案资源的融合利用[7]。具体而言,存在以下瓶颈:一是档案资源多处于“扫描入库”状态,缺乏内容的语义解析与知识标引;二是检索方式以关

关键词匹配为主,难以支持自然语言查询与语义关联发现;三是编研工作依赖人工整理,效率低下且成果同质化严重;四是档案服务以“人找信息”为主,缺乏个性化、主动化的知识推送能力。

这些问题的深层原因在于:档案信息化建设缺乏人工智能技术的系统性嵌入,数据治理滞后于技术发展,领域知识建模不足,安全审核机制尚未实现智能化转型。

4. 人工智能技术在高校档案信息化建设中的核心应用场景

4.1. 自然语言处理赋能档案文本深度挖掘

自然语言处理(Natural Language Processing,简称 NLP)技术通过语义解析与模式识别,显著提升了档案文本的自动化处理能力。传统档案编研与整理中,非结构化文本如会议记录、总结报告、教学评估材料等,需依赖人工逐篇阅读与标引,耗时耗力且主观性强。NLP 技术通过文本分类、关键词提取、命名实体识别及聚类分析,可自动化完成档案的主题归类、核心内容标引及隐含关联发现。例如,系统可自动将教学档案按“课程改革”“学科建设”“教学质量”等主题分类,或从科研项目中提取关键技术术语与研究方法,形成结构化知识单元。命名实体识别技术能精准捕捉档案中的人名、机构名、地名、项目名称等实体,结合语义分析构建实体间的关联网络,揭示学术合作脉络与学科发展轨迹。在服务层面,基于 NLP 的智能检索系统突破传统关键词匹配局限,支持自然语言查询(如“某教授主持的国家级科研项目”),通过意图解析与语义关联返回精准结果,显著提升档案利用效率[5]。

4.2. 计算机视觉技术激活图像与视频档案价值

计算机视觉(Computer Vision,简称 CV)技术为图像与视频档案的智能化编研开辟了新路径。历史照片、校史影像等多媒体档案常因元数据缺失而难以利用。CV 技术通过光学字符识别自动提取图像中的文字信息(如横幅标语、题词内容),人脸识别辅助标注人物身份,图像分类技术则按主题(如校园建筑、学术活动、历史事件)自动归档。例如,系统可识别老照片中的横幅文字并关联相应历史事件,或通过场景分类将照片归入“校庆典礼”“毕业合影”“学科竞赛”等专题库。

针对视频档案,CV 技术通过镜头分割、关键帧提取实现内容结构化,结合目标检测与场景识别自动标注画面中的人物、物体及背景。此外,图像修复技术(如去噪、超分辨率重建)能有效修复破损档案,恢复历史影像的清晰度,为档案编研提供高质量视觉素材[13]。

4.3. 知识图谱构建档案关联知识网络

知识图谱技术通过语义化组织档案资源,实现从碎片化信息到结构化知识的跃迁。该技术将档案实体(人物、机构、事件、成果)及其关系(如“任职于”“主持”“参与”“发生于”)转化为网状知识模型,突破传统档案管理的线性叙事局限。例如,校史人物档案经知识图谱关联后,可直观呈现其学术合作网络、师生传承关系或职务变迁轨迹;学科发展档案通过时序分析与关系挖掘,可揭示专业建设的阶段性特征与关键转折点。在知识服务层面,结合自然语言处理与图谱推理的智能问答系统,能理解用户复杂的查询意图(如“某学科近十年国家级项目的负责人及其合作者”),并生成精准的答案与可视化知识图谱[10]。

4.4. 智能编研与个性化服务

人工智能技术还推动了档案编研工作的智能化转型。基于文本生成技术的智能编研辅助系统,可自动生成专题汇编、校史年鉴、人物志等编研成果初稿,大幅缩短编研周期。同时,通过用户画像与行为分析,系统能够实现档案服务的个性化推荐。例如,面向科研人员推送相关科研项目档案与学术成果,

面向教学管理人员推送教学改革档案与评估材料, 面向社会研究者推送校史专题资源[9]。

5. 人工智能赋能高校档案信息化建设面临的挑战

5.1. 数据标准化缺失与“数据孤岛”突出

当前, 高校档案数据来源广泛, 涵盖教务、科研、人事、学生管理等多个业务系统, 数据格式多样且元数据标准不统一, 导致“数据孤岛”问题突出。以教学档案为例, 不同部门的课程编码、教师工号、学期标识等关键字段定义不一致, 难以实现跨系统的数据关联与融合。此外, 档案数字化过程中缺乏对语义元数据的规范标注, 使得人工智能模型难以有效理解档案内容, 制约了智能服务的深度[11]。

5.2. 领域知识建模滞后, 智能适配性不足

通用人工智能模型在高校档案领域的直接应用, 往往暴露出领域知识适配性不足的问题。例如, 在档案分类与标引任务中, 通用文本分类模型难以准确区分“教学研究”与“教学管理”等语义相近的档案类别; 在命名实体识别中, 通用模型对高校特有的机构名称、项目类型、职称序列等实体的识别准确率较低。究其原因, 是高校档案领域缺乏系统构建的本体体系与知识图谱, 导致人工智能模型无法充分利用领域知识进行语义推理[10]。

5.3. 安全审核机制尚未实现智能化转型

档案开放审核是档案流通的关键环节, 直接关系到信息安全与知识共享的平衡。当前, 大多数高校仍采用“人工逐份阅卷、凭经验划级”的审核模式, 效率低下且难以应对日益增长的数字化档案规模。虽然自然语言处理技术可辅助进行隐私实体识别与敏感度判定, 但高校档案中涉及的多重权益关系(如科研成果归属、学生个人信息、合作项目保密条款)尚未形成系统化的风险建模与审核规则, 智能化审核应用仍处于探索阶段。

5.4. 复合型人才短缺, 技术落地受阻

人工智能技术在高校档案信息化建设中的应用, 需要既懂档案业务又具备人工智能素养的复合型人才。然而, 当前高校档案队伍普遍存在“懂档案不懂技术、懂技术不懂档案”的两分格局。档案专业人员缺乏数据思维与算法思维, 技术人员则难以理解档案价值与职业伦理, 导致业务需求与技术实现之间的错位, 严重制约了人工智能技术的深度应用与可持续迭代。

6. 推进人工智能赋能高校档案信息化建设的实施路径

6.1. 强化顶层设计与数据治理体系建设

面对数据标准化不足与“数据孤岛”问题, 高校应从战略高度统筹规划, 系统推进档案数据治理体系建设。首先, 应制定覆盖档案全生命周期的数据标准规范, 明确元数据字段定义、格式要求、语义标签规则等, 确保不同业务系统之间的数据可互操作。其次, 构建统一的高校档案数据资源池, 打破教务、科研、人事等系统间的壁垒, 实现元数据、主题数据、引文数据的关联融合。此外, 应引入数据溯源、质量评估、权限管理等机制, 借助区块链等技术实现数据全生命周期可追溯与可信流转, 为人工智能模型训练提供高质量语料支撑。

6.2. 深化档案资源的知识化组织与语义化建设

档案信息化的真正价值不仅在于数字化扫描, 更在于完成“可被机器理解”的信息再造。高校应推

动档案资源从“数据化”向“知识化”升级。具体路径包括：一是构建面向高校档案的领域本体与知识图谱，从人员、机构、事件、成果、时间等维度梳理核心概念及其语义关系，形成规范化的知识框架；二是利用自然语言处理技术从海量档案文本中自动抽取实体与关系，并通过知识推理获取隐性关联，最终形成多维度、多粒度的档案知识图谱；三是引入检索增强生成等大语言模型技术，实现档案语义检索、智能问答、自动摘要等知识服务功能。

6.3. 推进档案鉴定与开放审核的智能化转型

为提升档案开放审核的效率与精准度，高校应从三个维度推进智能化转型。一是构建风险敏感度模型，利用自然语言处理对档案文本进行隐私实体识别、情感倾向判定与主题敏感度打分，结合知识图谱进行“人物-事件”关系的风险传播研判，自动触发高级人工复核。二是实行“分级分类+双轨审核”机制，对安全风险较低、社会需求度较高的档案采用“机器初审+抽样复核”的简化流程，对涉密或涉及多方权益的复杂档案则走“专家会审+技术支撑”路线。三是引入可解释性与合规追溯机制，任何档案的开放级别判定都应自动生成决策链条，记录触发规则与人工意见，形成可回溯的审计日志，为政策动态调整提供案例支撑。

6.4. 培养“档案+AI”复合型人才

人才是人工智能赋能高校档案信息化建设的关键支撑。高校应从理念更新、课程重塑与实践驱动三方面同步发力。在理念更新层面，档案馆应与信息学院、计算机学院联合开展跨学科沙龙，增强档案人员的数据思维与算法思维，同时让技术人员理解档案价值与职业伦理。在课程重塑层面，档案学专业可引入“自然语言处理概论”“知识图谱构建”“数字人文导论”等交叉课程，计算机专业则可设置“档案信息组织”“数字档案管理”等选修课。在实践驱动层面，可设立“智能档案工作室”，鼓励师生在真实数据环境中开发小型应用，如自动生成会议摘要、标注历史照片等。对于在职档案人员，应采取分层培养策略：初级阶段侧重工具使用，中级阶段强调数据治理与需求分析，高级阶段则具备算法选型与项目管理能力。

6.5. 构建产学研用协同创新生态

高校档案智能化转型涉及技术研发、系统开发与场景落地等多环节，仅凭高校自身难以独立完成。高校应积极寻求与本地IT企业、科研院所的合作，构建“产学研用”协同创新生态。初期可通过试点项目开展合作，如校史档案智能检索、历史照片自动修复等；随着合作深入，可共建“智能档案联合实验室”，聚焦多模态档案融合分析、跨年代档案语义关联等共性技术攻关。此外，合作还应延伸至人才培养领域，企业提供实习岗位与技术培训，高校为企业输送具备档案业务知识的AI人才，形成良性循环。

7. 结论

档案人工智能技术为高校档案信息化建设带来了革命性机遇。自然语言处理、计算机视觉、知识图谱等核心技术的应用，正深刻改变着档案资源的组织方式、知识挖掘的深度以及服务呈现的形式，推动高校档案工作从传统“保管利用”向“知识服务”与“智能治理”转型。然而，技术落地仍面临数据标准化不足、领域知识建模滞后、安全审核机制不完善及复合型人才短缺等现实挑战。高校档案部门需主动拥抱变革，通过强化顶层设计、深化数据治理、推进知识化组织、优化开放审核、培养复合型人才、构建协同创新生态等系统性举措，充分释放人工智能的赋能潜力，最终实现高校档案事业的高质量发展与智能化转型升级。

参考文献

- [1] 郑志荣. 以人工智能赋能档案事业高质量发展[J]. 中国档案, 2026(4): 18-19.
- [2] 黄霄羽, 李沐紫. “人工智能+”背景下国内外档案利用服务的发展态势[J]. 北京档案, 2026(4): 27-34.
- [3] 苏雯. 数智技术赋能非物质文化遗产档案的活化利用——基于四个典型案例的研究[J]. 山西档案, 2026(5): 170-176.
- [4] 赵爱华. 面向信息时代公民素养提升的教育档案资源开发利用研究[J]. 山西档案, 2026(2): 161-163.
- [5] 罗宝勇, 陈诗篇. 用户需求导向下档案公共服务的 AI 赋能路径研究[J]. 北京档案, 2026(5): 7-12.
- [6] 沈萍. 人工智能驱动的档案价值理论重构: 从静态保管到动态赋能[J]. 山西档案, 2026(5): 59-62.
- [7] 黎梓芄, 曾纪钧, 安然, 等. 大语言模型赋能档案服务场景及实现路径[J]. 中国档案, 2025(11): 64-65.
- [8] 陈文烨. 人工智能赋能高校档案管理流程优化与效率提升研究[J]. 情报理论与实践, 2025, 48(S1): 218-223.
- [9] 邢变变, 刘阳. 生成式人工智能背景下档案馆智慧编研的应然态势、实然困境和使然策略[J]. 档案与建设, 2025(2): 71-76.
- [10] 安新宇, 潘洁敏. “+AI”与“AI+”: 人工智能赋能档案业务应用场景的要素结构与模式分化[J]. 北京档案, 2025(11): 11-18.
- [11] 廉谨嫣. 面向“人工智能+”行动战略的档案数据治理模式优化探讨[J]. 档案管理, 2025(6): 38-45.
- [12] 钱毅, 武晓璇. 面向人工智能的档案就绪资源建设研究[J]. 北京档案, 2025(11): 4-10.
- [13] 卢燕. 人工智能技术在高校档案编研中的应用研究[J]. 兰台世界, 2026(4): 50-54.
- [14] 李秀华. 论人工智能赋能高校档案工作的应用路径及其优化策略[J]. 档案管理, 2024(4): 106-110.