

生成式人工智能赋能数智档案馆建设： “人工智能 + 档案”十六大应用场景展望

杨 荀¹ 杨冬权^{2,3}

(1. 人民交通出版社股份有限公司, 北京 100101;

2. 国家档案局, 北京 100037;

3. 中央档案馆, 北京 100095)

【摘要】生成式人工智能正深度重塑档案工作发展范式, 是推动档案事业高质量发展的重要动能。文章立足国家档案局“人工智能 + 档案”专项行动政策导向, 以杨冬权“五全”档案工作理念为理论支撑, 结合“人工智能 + 档案”行业发展趋势, 剖析生成式人工智能与档案工作的内在契合性, 梳理十六大“人工智能 + 档案”应用场景, 同时结合行业现状, 研判技术落地过程中数据、技术、安全、人才、标准等现实困境, 针对性提出优化策略, 为新时代档案工作高质量发展提供参考。

【关键词】生成式人工智能; 人工智能 + 档案; 数智档案馆; 应用场景

【作者简介】杨荀, 人民交通出版社股份有限公司党务及档案主管, 助理馆员, 研究方向为文献信息服务与方法、党史党建, 邮箱: yimucaocao@163.com; 杨冬权, 国家档案局原局长、中央档案馆原馆长, 中共十七大、十八大党代表, 第十二届全国政协委员, 研究馆员, 研究方向为档案事业管理、档案数字转型、档案文化建设、党史党建。

1 引言

档案是机关、团体、企业事业单位及其他组织和个人在各种活动中直接形成的、具有保存价值的各种形式的历史记录。伴随数字信息技术迭代演进, 我国档案行业先后经过存量数字化、增量电子化两轮转型, 当前行业整体迈入智慧化建设关键窗口期。生成式人工智能依托大模型与多模态融合架构, 在自然语言解析、非结构化信息萃取、关联知识生成等方面实现技术突破, 为破解档案行业长期存在的海量资料沉睡、查阅利用效率偏低、档案隐性价值难以深挖等现实堵点提供全新技术解法。

“十五五”时期, 国家档案局将拓展“人工智能 + 档案”, 推进数智技术赋能档案工作^[1]。2026年全国档案局长馆长会议提出要系统谋划推进档案工作现代化建设, 稳妥探索运用人工智能、大数据等新技术驱动治理效能提升^[2]。2026年4月, 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》^[3], 并开展人工智能应用试点工作, 从顶层设计层面明确档案领域拓展智能应用场景、培育档案领域新质生产力的发展方向, 标志着生成式人工智能在档案领域从学术研讨阶段转入规模化试点落地阶段。现阶段国内档案界对生成式人工智能的价值共识持续加深, 但落地实践仍普遍存在应用边界模糊、技术适配方案缺失、智能系统与传统档案业务割裂等现实问题。基于上述行业现状, 本

文从档案“收、管、存、用”全业务链条出发, 细化梳理十六项落地性较强的人工智能应用场景, 客观梳理技术融合进程中的现实挑战并提出适配性对策, 以期助推智能技术与档案业务深度融合, 助力档案事业高质量发展。

2 生成式人工智能与档案工作

2.1 生成式人工智能概述

生成式人工智能是基于大模型预训练迭代的新型智能范式, 区别于传统规则化智能的浅层处理模式, 可自主完成语义理解、多模态解析与知识生成, 实现档案治理从机械存储检索向智能赋能的层级跨越。从现有的生成式人工智能模型来看, 其功能特征主要体现为三个方面: 一是具备强大的自然语言认知与交互性能, 可精准识别人类日常口语化、场景化指令, 实现高效、自然的人机智能对话与互动协作。二是拥有良好的场景自适应能力, 模型依托海量数据完成预训练学习后, 可结合不同行业场景进行参数微调, 灵活适配各类专业化、精细化业务场景需求。三是具备开放式多模态内容创作优势, 能够自主输出文本、影像、数字虚拟形象等多种形式的內容, 真正完成从传统决策式智能到生成式智能、从用户原创内容到人工智能原创内容、从时空信息变革到知识体系革新的全方位技术跨越^[4]。

2.2 生成式人工智能与档案工作的内在契合性

2024 年施行的《中华人民共和国档案法实施条

例》增设“档案信息化建设”章节^[5],明确提出推进档案智能识别与资源深度开发利用,为生成式人工智能在档案领域的规范落地与创新应用提供了制度支撑。生成式人工智能依托算力运算、语义学习、多模态融合等技术优势,可全面升级档案服务模式与服务效能,与档案“收、管、存、用”全生命周期业务高度适配。其多模态能力可实现图文、音视频等异构档案标准化处理,支撑全域建档归档;知识推理技术能够挖掘碎片化档案的隐性价值,深化资源开发利用;智能检索可精准匹配多元查档需求,改善传统服务短板;智能审核与风险溯源可构建全周期安全防护体系,筑牢档案安全底线,助力新时代档案服务提质增效。

2.3 生成式人工智能对数智档案馆的赋能逻辑

“人工智能+档案”是技术与业务的深度融合与范式重构,能进一步打破传统数字档案馆“重保管、轻利用”的发展局限和选择性管护的“选时代”模式,构建智能治理新格局,催生数智档案馆业态。相较于传统数字档案馆与初级智能档案馆,数智档案馆以知识赋能为核心、精准服务为导向、全域安全为底线,具备数据驱动、智能自主、价值深挖的特质,能够有效破解传统档案管理发展瓶颈,助力档案工作融入国家治理现代化。

3 “人工智能+档案”应用场景解析

智能化技术是实现档案全量留存、智慧开发的核心支撑。本文提出十六大生成式人工智能赋能档案场景,覆盖了数字化归集、数据治理、编研开发、公共服务、安全保障、文化传播六大领域。

3.1 文字档案自动识别转换

纸质古籍、手写稿件、历史公文、老旧报表等文字类档案,普遍存在字迹褪色、手写潦草、版式杂乱等问题。生成式人工智能在OCR基础上叠加NLP语义校正、上下文推理能力,能够批量完成印刷体档案识别,还可精准辨析模糊字迹、异体字、方言手写体、破损残缺文本,自动修正识别乱码、错别字等问题,并按照档案著录规范完成结构化拆分与标注,输出可检索、可编辑、可统计的标准化文本数据。该技术模式识别准确率可达99.9%,整体作业效率提升20倍以上,将大幅降低存量文字档案数字化改造成本^[6]。国内多家省级档案馆已将该技术应用于民国档案、革命历史文献修复归集工作;国外普林斯顿大学联合多国科研团队,依托智能识别转换技术完成埃塞俄比亚、厄立特里亚和埃及的《圣母奇迹》手稿数字化加工,搭建PEMM专项档案数据平台,为古籍类档案利用提供技术借鉴^[7]。

3.2 录音档案自动识别转换

口述史料、会议录音、访谈音频等音频档案,是承载鲜活历史细节的重要资源,也是口述档案工作的核心载体。生成式人工智能搭载档案领域语音模型,内置政务专业术语、地方方言、行业黑话词库,一站式完成音频降噪、杂音剔除、多发言人分离、时间戳精准标注、双语种同步转写等全套作业,最终生成音文一一对应的结构化索引文件。目前北京、四川、广东等地市级档案馆,已运用该技术完成大批量红色口述史料、老干部访谈录音的转写工作,建设红色档案专题数据库、启动长征口述历史档案采集计划等,有效盘活音频档案资源,完善红色档案资源体系。

3.3 照片与录像档案自动识别转换

影像档案作为直观化的历史记录形式,广泛涵盖政务活动影像、民生纪实照片、历史专题录像等资源。依托生成式人工智能多模态语义解析与计算机视觉技术,系统可自主提取影像档案的拍摄时间、地点、参与人物、核心事件等多维元数据,自动拆分长时录像的关键片段、剔除无效冗余画面,并生成适配档案业务的内容摘要;同时可针对影像画面人物情绪、场景氛围开展情感分析,丰富影像档案的标签体系,实现从“静态存储”向“动态智能利用”转型,全面提升影像档案资源利用率。从行业实践来看,浙江省档案馆率先落地多模态影像智能治理系统,针对馆藏海量政务视频、历史老照片开展智能化编目、片段拆解与主题标签生成;上海市档案馆开发了“AI影像档案修复系统”,对近现代城市纪实影像进行语义化加工,补齐老旧影像缺失元数据,有效提升档案利用效果。

3.4 档案数据挖掘与知识图谱构建

数字化档案普遍存在碎片化、关联性弱的问题。生成式人工智能基于流的生成模型(Flow-based Model)、扩散模型(Diffusion Model)等在档案馆内部进行模型训练^[8],针对文本、音频、影像等多模态档案数据开展深度治理:一方面完成数据清洗、冗余内容剔除、元数据补全,统一馆藏档案数据标准;另一方面自动抽取人物、机构、事件、时间、地域五大核心实体,梳理实体之间的从属、关联、因果关系,搭建层级化、主题化的全域档案知识图谱。该模式可实现档案资源纵向时序串联、横向主题关联,打破单份档案的信息孤岛,依托链式检索模式深度挖掘隐性史料价值,是“档案大脑工程”全国一体化档案知识图谱建设的重要底层技术。

3.5 跨模态与语义智能搜索

生成式人工智能重构档案检索底层逻辑，搭建全新跨模态语义检索体系，支持文字、语音、图片、短视频等多种检索形式；同时系统可深度解析用户口语化、碎片化、模糊化的检索意图，跳出字面匹配局限，基于语义关联匹配全馆藏多模态资源，秒级输出检索结果，整体查准率稳定维持在 91% 以上。从行业落地实践来看，北京市昌平区档案馆自主研发跨模态智能检索利用系统，实现了检索过程的精准化与智能化，并获得国家版权局软件著作权，在本地化部署大模型后打通多类型档案数据壁垒，将用户平均查档时长缩短 90%，显著降低检索门槛。该检索模式既能满足党政机关专题资政、史料溯源等专业化查档需求，也能够简化普通民众民生档案查询流程，优化普惠型公共档案服务体验，为全国档案馆检索服务智能化升级提供可复制范式^[9]。

3.6 档案开放智能审核与鉴定

档案开放审核与涉密审核是档案利用的前置关键环节。生成式人工智能依托保密与开放政策构建规则库，通过全文语义分析识别涉密、隐私等敏感信息，智能标注开放等级并生成审核报告。当前，国家档案局持续健全数字档案安全治理制度体系，稳步推进电子文件安全管控与智能审查试点落地工作，浙江、江苏等多地档案馆在智慧化建设过程中，积极引入区块链溯源、信创安全管控等技术机制，强化档案数据访问全程留痕、行为可追溯，进一步规范档案数据流转流程，夯实档案数字化建设的合规安全基础。

3.7 档案专题资料智能编写

档案专题编研是资政服务的重要载体，传统专题编研需要工作人员自主检索史料、梳理内容框架、整合多方资源，一份高质量专题编研成果往往需要数周甚至数月时间，编研成本高。生成式人工智能可根据指定主题自动检索馆藏资源、梳理逻辑框架、提炼核心内容，快速生成专题简报、研究报告、史料汇编等成果，不仅能适配党政决策、学术研究、社会科普需求，大幅提升档案资政效能，同时适配新时代新媒体传播趋势，生成轻量化、碎片化的科普类微编研内容，辅以人工优化内容细节、校验史实真伪即可定稿，大幅缩短编研周期，提升档案资政赋能效能。

3.8 档案展览大纲智能策划与展品选配

档案主题展览是档案资政育人价值外化的重要途径。生成式人工智能可贯穿策展前期筹备、中期布展、后期宣发全业务链条，实现全流程智能化赋能：工作人员录入展览主题、目标受众、展陈场地、

预算规模、展出周期等基础参数后，系统可结合馆藏资源储备情况智能划分展陈单元、搭建层级化、逻辑化展览大纲；依托多模态检索与语义匹配技术，自动筛选适配单元主题的文、老照片、实物影像等展品，并按照展陈节奏完成排序；最后一键生成展板文案、展品解说词、前言结语等全套配套文字素材，极大简化策展作业流程。四川省档案馆打造了全国首个档案展览专属 AI 智能体“档小江”“档小河”，依托红色档案专属知识库，打破档案馆的物理围墙，提供了覆盖更广、触达更易的档案服务。

3.9 大事记与年鉴智能编撰

大事记与年鉴编撰需梳理海量史料，人工工作量极大。生成式人工智能依托档案垂直大模型与语义抽取技术，可实现史料全自动归集、过滤与分级筛选；按照官方年鉴、大事记编撰规范，从时间、主题、层级多维度筛选重大政务活动、民生惠民项目、专项业务成果与突发事件，智能梳理时序脉络、统一行文体例与表述话术，快速生成完整编撰初稿。此外，生成式人工智能还可发挥二次校验作用，自动排查条目内容重复、时间逻辑冲突、史实表述错误等问题，进一步降低校对压力。国内多地档案部门已完成规模化试点：温州市鹿城区档案馆本地化部署 DeepSeek 垂直大模型，一键生成大事记、机构沿革等编研初稿，编研整体效率提升 70% 以上。

3.10 地方志与专业志智能编纂

地方志、行业专业志编纂资料来源广、内容体系复杂、体例要求严苛、编纂周期长。生成式人工智能可自动归集多源史料，完成去重纠错与标准化处理，严格遵循志书体例撰写章节内容、理顺行文逻辑，核查内容疏漏与逻辑矛盾，进一步提升志书编纂效率与质量。上海市档案馆及相关研究机构推出了名为“申知·沪志”的上海方志大模型，对海量的方志资源及城市纪实影像进行语义化分析，激活历史智慧，为超大城市的运行治理提供历史启示与数据支撑；杭州市地方志办公室上线“杭小志”数智编纂系统，将生成式 AI 深度嵌入年鉴全流程编纂环节，有效破解基层史志编研人员不足的难题。

3.11 档案安全隐患智能预警

档案安全涵盖实体安全与数据安全两大核心维度。生成式人工智能联动库房安防、环境监测与数据管理系统，实时监测库房温湿度、火情、入侵等实体风险，动态预警数据篡改、泄露、非法访问等数字风险，筛查开放档案敏感内容，推动档案安全从事后处置转向事前预警、事中管控，筑牢全留存、全备份安全防线。

3.12 档案利用场所智能知识服务

传统查档服务形式单一、被动滞后。AI 智能服务终端可提供语音、文字双向咨询,解答查档流程、开放政策等常见问题,结合用户检索习惯个性化推送关联档案资源,搭配虚拟导览、智能讲解功能,丰富公共服务形态,提升档案惠民服务质量与资源利用率。比如,上海市运用分布式数据库整合等技术,实现民生档案全种类在线出档,并将长三角地区民生档案异地查档、便民服务接入移动端,实现查档“零跑动”。

3.13 老照片、老录音、老录像智能修复

老旧声像档案普遍存在模糊、破损、杂音、失真等问题,属于濒危珍贵史料。生成式人工智能专项修复模型可实现老照片去黄补损、高清上色,老录音降噪修音,老录像优化画质、去除划痕,批量完成濒危声像档案抢救修复,还原真实历史视听记忆。

3.14 电子文件自动归档与接收监管

海量原生电子文件人工归档易出现错归、漏归、著录不规范等问题。生成式人工智能可对接各类办公业务系统,实时捕获原生电子文件,自动完成分类、著录、标引、格式转换与批量归档,同步核验文件合规性,留存全流程操作日志,实现电子文件全程可溯、规范管理。

3.15 文生图、图生视频、文生视频等智能创作

档案文化宣传、展览文创亟需优质可视化素材。生成式人工智能的出现能够进一步降低元宇宙空间等的基建成本^[10],依托多模态创作能力,可根据文字脚本生成档案主题海报、历史场景插画、科普短视频、宣传影像,将静态老照片转化为动态纪实视频,大幅降低创作成本,丰富档案文化传播载体,增强档案文化感染力,赋予档案馆“生命力”。

3.16 AI 生成档案智能识别与标记

AI 合成内容泛滥对档案原始性、权威性形成冲击。生成式人工智能可通过文本、图像、音频特征,批量筛查归档资源,精准识别 AI 生成内容并添加专属标记、建立专项台账,严格区分原生档案与合成档案,从源头保障档案史料的真实性与规范性。

4 “人工智能+档案”应用的现实挑战与优化框架

4.1 现实挑战

当前“人工智能+档案”规模化落地仍存在多重现实瓶颈。一是数据治理体系不完善。馆藏档案数字化质量参差不齐,元数据缺失、标准不一,异构数据混杂存储,高质量训练数据集匮乏,同时数

据壁垒突出、标注难度大,制约智能应用落地。二是技术适配性不足。通用大模型缺乏档案专业语料训练,存在模型幻觉、史实偏差等问题,国内档案垂直大模型研发滞后,难以适配档案求真求实的专业要求。三是安全风险多维叠加。生成式人工智能应用存在数据泄露、算法权责模糊等隐患,AI 生成档案权属、归档标准尚不明确,冲击档案原始权威底线。四是人才结构失衡。行业从业人员普遍缺乏 AI 技术与数据治理能力,懂档案、懂技术、懂安全的复合型人才缺口显著。五是标准体系滞后。全国暂无统一的 AI 档案应用规范、数据标准与评价体系,行业建设碎片化、无序化问题突出。

4.2 优化框架

针对上述发展瓶颈,结合学界研究与行业实践,可从五方面推进优化升级。第一,健全全域数据治理体系。统一档案元数据与归档标准,常态化开展存量档案提质、增量档案规整,构建标准化档案资源池;实行数据分级管控,涉密数据闭环留存、全程备份,非涉密数据合规共享,夯实智能应用数据根基。第二,深耕档案垂直模型研发。推动产学研协同攻关,训练专属档案语料库,优化算法逻辑、修正模型幻觉,强化史实校验能力,实现技术与档案专业深度适配。第三,筑牢全维度安全防线。推行 AI 系统本地化、私有化部署,建立算法备案、合规审查、全程溯源机制,明确 AI 生成档案的归档标准与权属边界,实现闭环安全管控。第四,完善复合型人才培育机制。构建高校学科培养、行业在岗培训、项目实操锻炼的培育模式,推动档案专业与数字技术交叉融合,补齐人才短板。第五,健全行业标准规范。由国家档案局牵头完善生成式人工智能档案应用的数据、技术、安全、评价标准,统一行业建设口径,破解碎片化发展难题。

5 结论与展望

生成式人工智能与档案业务深度融合,是档案事业培育新质生产力、实现高质量发展的必然趋势。未来,随着档案垂直大模型持续迭代、行业标准不断完善,生成式人工智能在档案治理、资源开发、文化传播、公共服务领域的赋能价值将持续凸显。各级档案馆室应主动拥抱技术变革,抢抓发展机遇,尽快对接生成式人工智能技术企业,结合本单位实际优先落地可实现的应用场景,逐步构建高维、新质的智慧档案管理体系,让档案资源活起来、用起来、价值显出来,推动我国档案工作从数字化、智能化向智慧化跃迁,为传承文明、服务社会、助力国家治理现代化作出更大贡献。

注释及参考文献

- [1] 中共中央办公厅、国务院办公厅印发《全国档案事业发展“十五五”规划》（全文）[EB/OL]. (2026-06-11)[2026-06-12]. <https://www.saac.gov.cn/daj/yaow/202606/c75591cb82564d669635c859041512cf.shtml>.
- [2] 全国档案局长馆长会议在京召开 [EB/OL]. (2026-01-17)[2026-06-11]. <https://www.saac.gov.cn/daj/yaow/202601/58cca44a5edc400faa13dc5607053238.shtml>.
- [3] 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 [EB/OL]. (2026-04-10)[2026-06-11]. <https://www.saac.gov.cn/daj/tzgg/202604/12aa9fab60c4dbeab5444f03e78bdb9.shtml>.
- [4] 郭小东. 生成式人工智能的风险及其包容性法律治理 [J]. 北京理工大学学报 (社会科学版), 2023, 25(6): 93-105+117.
- [5] 《中华人民共和国档案法实施条例》[EB/OL]. (2024-01-12)[2026-06-11]. https://www.gov.cn/zhengce/zhengceku/202401/content_6928164.htm.
- [6] 许闯. 生成式人工智能赋能档案数据要素化：机理、困境与治理路径 [J]. 兰台世界, 2026(4): 55-59.
- [7] 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.
- [8] 张笑玮, 吴建华. 生成式人工智能赋能档案馆服务质量的优化路径研究 [J]. 档案与建设, 2025(2): 62-70.
- [9] 昌平区档案馆“跨模态检索系统”获国家软件著作权认证 [EB/OL]. (2026-01-22)[2026-06-11]. <http://www.chinaarchives.cn/mobile/category/detail/id/46534.html>.
- [10] 解学芳, 高嘉琪. AIGC 模式赋能数字文化创新的逻辑与善治：基于 ChatGPT 热潮的思考 [J]. 江海学刊, 2023(3): 86-95.

Empowerment of Generative Artificial Intelligence in the Construction of Digital-Intelligent Archives: Prospect on Sixteen Application Scenarios of Artificial Intelligence plus Archives

YANG Xun¹, YANG Dongquan^{2, 3}

(1.China Communications Press Co., Ltd., Beijing 100101;

2.National Archives Administration of China, Beijing 100037;

3.Central Archives, Beijing 100095)

Abstract: Generative artificial intelligence is profoundly reshaping the development paradigm of archival work and serves as a key driving force for the high-quality development of the archival undertaking. Guided by the special-action policy of "AI+Archives" issued by the National Archives Administration of China, supported by Yang Dongquan's "Five-Comprehensiveness" archival work philosophy, and combined with the industrial development trend of "AI+Archives", this paper analyzes the internal compatibility between generative artificial intelligence and archival work, and sorts out sixteen application scenarios of "AI+Archives". Considering the current status of the industry, it identifies practical dilemmas in technology implementation involving data, technology, security, talent and standards, and puts forward targeted optimization strategies, so as to provide references for the high-quality development of archival work in the new era.

Keywords: Generative artificial intelligence; AI+Archives; Digital-intelligent archives; Application scenarios