

人工智能嵌入电子文件归档全流程的路径研究

王琳洪¹, 尹雪云^{2,*}

¹北京联合大学应用文理学院, 北京, 中国

²北京联合大学学生处, 北京, 中国

*通讯作者

【摘要】在数字治理体系持续推进与档案资源体量快速增长的背景下, 传统人工主导的电子文件归档模式已难以适配海量数据高效管理的现实需求。电子文件归档是档案工作数智化转型的核心环节。本文立足人工智能与档案业务深度融合的现实需求, 系统梳理电子文件形成、办理、归档、保管四阶段的业务特征, 提出 AI 全流程嵌入的理论框架, 设计各环节的智能应用路径, 并从技术实现、制度保障、风险防控三个维度给出落地建议, 为推动电子文件归档管理范式升级提供参考。

【关键词】人工智能; 电子文件归档; 全流程嵌入; 数智化转型

1. 引言

1.1 研究背景与问题提出

数智技术赋能下的电子文件归档模式创新, 是当前国内档案学界与实务界共同关注的核心研究议题。电子文件归档作为连接业务系统与档案系统的关键枢纽, 其效率与质量直接影响档案资源体系建设成效。然而, 当前电子文件归档仍面临元数据著录依赖人工、四性检测手段单一、归档质量参差不齐等突出问题, 制约了档案管理从“数字化”向“数智化”的跃迁。人工智能技术的快速发展为破解上述难题提供了新的可能, 但如何将 AI 系统性嵌入归档全流程, 尚缺乏清晰的路径设计与理论支撑。

1.2 研究意义

数字环境下, 非结构化档案信息呈现体量大、类型多、处理速度快且蕴含价值高等特征, 其收集、鉴定、保管及利用等环节日益繁琐复杂。这一现实矛盾正驱动各类 AI 技术加速嵌入档案管理全流程, 推动档案工作实现由繁入简、由难变易的效能跃升[1]。针对各级档案馆(室)在推进电子文件归档智能化过程中面临的路径不清、标准缺失、风险难控等现实问题, 提出可操作的嵌入路径与治理方案, 为推动国家档案工作数字化转型战略部署落地提供技术与制度层面的参考支撑, 助力以数智化转型驱动档案工作高质量发展。

2. AI 嵌入电子文件归档全流程的理论框架

2.1 电子文件归档全流程业务解析

电子文件归档全流程可划分为四个阶段:

形成阶段, 即电子文件在业务系统中产生并附带初始元数据; 办理阶段, 即文件经流转、审批、会签等业务处理; 归档阶段, 即文件经鉴定、整理、移交进入档案系统; 保管阶段, 即档案的长期保存、利用与处置。每个阶段均产生大量结构化与非结构化数据, 构成 AI 介入的基础。

2.2 AI 技术与归档业务的融合逻辑

当前, 电子文件归档工作的重心, 需由事后“被动接收”转向事前“主动介入”。电子文件归档应嵌入数字化业务全过程, 实现对电子文件的全生命周期管理[2]。AI 与归档业务的融合遵循“数据驱动—智能增强—人机协同”的递进逻辑。第一层, 利用自然语言处理(NLP)和光学字符识别(OCR)技术实现非结构化数据的结构化转化; 第二层, 利用大语言模型(LLM)和知识图谱实现智能著录、分类与检索; 第三层, 利用规则引擎与机器学习模型实现四性自动检测与风险预警[3]。三层逻辑逐层深入, 最终实现归档全流程的智能化升级。

3. 各环节 AI 嵌入路径设计

3.1 形成环节: 智能捕获与元数据自动生成

在形成环节, AI 的核心任务是解决“元数据从哪来”的问题。传统模式下, 元数据完全依赖业务人员手工逐条填写, 不仅效率低下, 而且遗漏率高、填写标准不统一, 给后续检索利用埋下隐患。引入 AI 后, 这一瓶颈得到有效突破: 一方面, 通过 NLP 自然语言处理技术, 系统可智能识别并结构化抽取文件正文中的关键信息, 包括发文机关、文

号、日期、主题词等核心字段，再结合业务系统接口自动补全其余元数据字段，真正实现“文件生成即著录”，大幅降低人工操作成本。另一方面，针对纸质文件，利用 OCR 光学字符识别技术完成数字化扫描后，AI 可自动从中提取文本元数据，替代原本繁琐的人工录入流程，从而进一步减少人工干预，提升整体归档效率[4]。

3.2 办理环节：流程智能监控与辅助决策

办理环节的 AI 应用重点在于流程合规性监控。传统模式下，归档流程的合规性主要依赖人工逐环节核查，不仅耗时费力，且容易出现疏漏。引入 AI 后，通过构建归档流程知识图谱，系统可实时比对文件流转状态与预设归档规则，自动识别超期未办、缺失签字、附件不全等异常情况，并第一时间向相关责任人推送提醒，确保问题及时处理[5]。此外，AI 还能基于历史归档数据的学习与分析，辅助档案人员更加客观地判断文件是否具备归档条件，有效降低因个人经验差异导致的主观判断偏差，从而提升归档工作的整体质量与效率。

3.3 归档环节：四性智能检测与自动分类

人工智能在电子档案分类阶段的核心价值，在于通过多维度智能分类模型，突破传统人工分类“单一维度、经验依赖”的局限[6]。归档环节作为 AI 嵌入的核心场景，其重要性源于该阶段是电子文件从“业务流”向“档案流”转化的关键枢纽，需同时满足真实性、完整性、可用性、安全性四大核心要求，且涉及大量非结构化数据处理，AI 的介入能显著提升效率与精准度。

在真实性检测上，区块链存证技术为文件来源提供不可篡改的“数字指纹”，AI 则通过哈希值校验、语义相似度分析等比对，验证文件内容与来源的一致性[7]。例如，某政策文件归档时，AI 可自动比对区块链存证的原始版本与当前文件，识别是否存在篡改痕迹，确保档案“来源可信”；完整性检测依赖哈希校验与 AI 内容比对的双重机制。哈希校验可快速判断文件是否缺页、缺件，而 AI 通过图像连续性分析或文本逻辑校验，进一步识别“隐性缺失”，避免因人工疏漏导致档案不完整；在可用性检测中，格式识别引擎是核心工具。AI 可自动识别 PDF、DOCX、JPG 等文件格式，并对照长期保存标准判断是否需转换。例如，某老旧系统生成的 DOC 文件，AI 会提示需转换为 PDF/A 格式，以确

保未来可读取；安全性检测通过敏感信息识别模型实现。AI 可自动标注密级程度与开放范围，基于关键词匹配、实体识别等技术，识别文件中的涉密内容如人名、地名、敏感政策术语等，避免涉密档案误开放。

自动分类则是 LLM 与预训练模型的协同成果。LLM 通过语义理解提取文件核心主题，结合预训练分类模型实现自动归类，准确率可以大幅提升。例如，某会议纪要文件，AI 可自动识别其为“行政管理”类，归类至对应目录，极大程度的减少人工分类的工作量。

3.4 保管环节：智能检索与长期保存预警

保管环节作为电子档案生命周期中的核心阶段，其 AI 应用的核心目标是在保障档案安全的前提下，最大化释放档案价值，同时预防潜在风险。智能检索是价值释放的关键抓手，传统关键词检索依赖字面匹配，常因表述差异导致漏检，而 AI 通过向量数据库将文件内容转化为高维向量，结合语义检索技术理解用户意图，实现“以意搜档”[8]。例如用户输入“高校科研项目中期检查的材料清单”，AI 能识别“科研项目中期检查”“材料清单”的语义关联，快速定位到《XX 大学科研项目中期检查实施细则》《XX 学院科研项目中期考核材料提交指南》等具体文件，帮助用户快速获取所需的材料要求，显著提升检索效率，减少人工筛选的时间成本。

长期保存预警则是风险预防的核心手段，AI 通过实时监控存储介质的状态，建立故障预测模型，提前识别介质老化或损坏风险，发出迁移建议；同时跟踪文件格式的演变趋势，当检测到某格式的兼容性下降，AI 会提示将档案转换为更稳定的格式，避免因软件过时或格式淘汰导致“读不出、打不开”的风险。

4. 实现路径与关键技术

4.1 大语言模型与规则引擎协同架构

归档全流程涉及著录、分类、检测、合规校验等多个环节，需求差异大，单一技术很难同时兼顾灵活性与规范性。为此，本文提出“LLM+规则引擎”的协同架构。LLM 擅长处理语义理解、智能著录、自动分类等柔性任务，能应对表述多样、上下文复杂的场景；规则引擎则负责四性检测、合规校验等刚性任务，依据预设规则逐条判定，结果确定且可追溯。二者通过 API 接口实现联动，LLM 的输出结果需经规则引擎校验通过后

才能正式入库。这一设计的核心逻辑在于：LLM 解决“能不能理解”的问题，规则引擎解决“合不合规矩”的问题，两者互为补充，既保留了 AI 处理非结构化数据的灵活性，又确保了归档成果符合档案管理的刚性要求。

4.2 数据标准与接口规范建设

AI 要真正嵌入归档全流程，前提是数据能够顺畅流通。当前业务系统与档案系统之间往往存在数据壁垒，元数据字段不统一、接口协议不兼容、数据格式不一致，导致 AI 模型难以获取高质量的训练与推理数据。为此，建议推动制定电子文件归档 AI 应用相关数据接口规范，从三个维度统一标准：一是统一元数据字段标准，确保业务端产生的文件属性信息能被档案端准确识别；二是统一接口协议与数据交换格式，降低系统对接成本；三是建立数据质量评估机制，为 AI 模型提供可靠的数据基础。只有打通数据壁垒，AI 才能在著录、分类、检测等环节真正发挥作用，否则再先进的模型也只是“无米之炊”[9]。

5. 风险防范与治理建议

5.1 算法偏见与元数据失真风险

AI 模型在档案管理中的应用虽能显著提升工作效率，但也潜藏不容忽视的风险。由于训练数据本身可能存在偏差，模型在著录过程中容易产生错误，或在分类环节出现失准，这些问题将直接影响档案的真实性与后续可用性。为有效应对这一挑战，应建立“AI 初检+人工复核”的双重校验机制——由 AI 完成初步筛选与处理，再由专业人员进行二次审核，从而在效率与准确性之间取得平衡。加强对人工智能技术的伦理审查和监管力度，防止技术滥用和误用导致的伦理风险，尤其在密级判定、长期保存策略等高风险操作中，必须保留人工终审权，避免因算法偏差导致关键决策失误，确保档案管理的严谨性与可靠性[10]。

5.2 制度配套与人机协同机制

技术的有效落地，离不开完善的制度保障作为支撑。当前 AI 技术在档案管理领域的应用正逐步深入，但相应的制度规范尚显滞后，亟需从顶层设计层面加以补齐。建议在《电子文件管理办法》现有框架下，专门增设 AI 应用章节，系统梳理并明确 AI 辅助决策的法律效力边界——即哪些环节可由 AI 自主处理，哪些必须经人工确认；同时清晰界定责任归属，当 AI 辅助决策出现偏差时，

如何追溯责任链条；并提出具体的审计追溯要求，确保每一项 AI 操作均可查证、可回溯。通过上述制度设计，最终构建起“制度管人、技术管事”的协同治理格局——用制度规范人的行为与权责，用技术保障事务处理的效率与标准，二者相互配合、相互约束，方能推动 AI 在档案管理中安全、规范、可持续地发挥作用。

6. 结语

人工智能深入电子文件归档全过程，是档案工作由以往“数字化”向“数智化”的一次飞跃发展，这是技术上的提升也是档案管理工作理念及方式上的一场革命，是档案工作由被动记录到主动智能化管理的过程。而本文正是为了满足这种需要，提出一种“四个环节嵌入方式+LLM 与规则引擎相结合的方式+人机协同治理”的总体思路，在这四个环节中包括从文件创建、传输到归档以及长期保存整个过程；LLM 与规则引擎结合的方式既可以实现对文本的理解也可以施加一定的硬性限制；人机协同治理可以在效率与安全性之间达到一个良好的协调。这三个方面互相配合形成相对完整的技术和管理结合的方法。

此方案是当下档案信息化工作的一个有益探索，具有较好的实用性和借鉴意义，在未来可以进一步针对政府、卫生、银行等行业进行实际调研以验证其可行性和实用性，促进研究成果真正落地生根，服务于档案事业的发展。

参考文献

- [1] 陈慧，罗慧玉，张凯，等.AI赋能档案：AI技术在档案管理中的赋能模式探究[J].山西档案，2020，（4）：76-83+131.
- [2] 王抢，高朕栋，邹绍辉.数字化转型背景下电子文件归档方法与策略[J].山西档案，2025，（4）：160-163.
- [3] 陈会明，史爱丽，王宁，等.人工智能技术在档案工作中的应用与发展刍议[J].中国档案，2020，（3）：72-74.
- [4] 杨靖，尚小雨.人工智能赋能档案治理现代化：情境转向、体系创新与路径构建[J].浙江档案，2024（12）：23-27.
- [5] 于英香，赵倩.人工智能嵌入档案管理的逻辑与特征[J].档案与建设，2020，（1）：4-8.
- [6] 程昕烨.人工智能在电子档案分类归档流

- 程中的应用[J].办公自动化, 2026, 31(12): 110-112.
- [7] 鲜春燕.人工智能驱动的电子文件智能鉴定与归档优化研究[J].办公自动化, 2026, 31(12): 64-66.
- [8] 王金兰, 张晓艳.数字化时代电子文件单套制归档与长期保存策略[J].山西档案, 2026, (6): 161-164.
- [9] 王宇哲, 宋凌云.面向电子文件长期保存的云原生架构设计与关键技术[J].山西档案, 2025, (12): 130-132.
- [10] 陈研.人工智能在电子档案分类与检索中的应用研究[J].兰台世界, 2025, (1): 79-81.