

Intelligent transformation of archival work driven by DeepSeek: Application path and development paradigm

Qinghu Lin

China Hualu Group Co., Ltd., Dalian, Liaoning, 116023, China

Abstract

With the rapid advancement of generative AI technologies, DeepSeek—a large language model featuring open-source architecture, cost-effective deployment, and robust inference capabilities—is fundamentally transforming the technological ecosystem and operational paradigms in archival work. Addressing the practical needs of digital transformation in archival management, this paper systematically analyzes DeepSeek's application scenarios across the entire workflow, exploring its core value in enhancing operational efficiency, unlocking archival potential, and driving disciplinary innovation. Furthermore, it examines challenges faced by archival institutions in adopting large models from perspectives including technical compatibility, data governance, talent development, and ethical compliance. The study proposes a tripartite development path integrating “technology, business, and institutional frameworks,” offering theoretical insights and practical guidance for intelligent upgrading in archival practices.

Keywords

DeepSeek; large language model; intelligent file; data governance; computational archival science

DeepSeek 驱动下档案工作的智能化转型：应用路径与发展范式

林青虎

中国华录集团有限公司，中国·辽宁 大连 116023

摘要

随着生成式人工智能技术的飞速发展，DeepSeek作为具有开源特性、低成本部署及强推理能力的大语言模型，正深刻重塑档案工作的技术生态与实践范式。本文立足档案工作数字化转型的现实需求，系统分析DeepSeek在档案管理全流程中的应用场景，探讨其在提升业务效率、激活档案价值、推动学科创新等方面的核心价值。同时，从技术适配性、数据治理、人才培养及伦理规范等维度，剖析档案行业应用大模型面临的挑战，并提出构建“技术-业务-制度”三位一体的发展路径，为档案工作的智能化升级提供理论参考与实践指引。

关键词

DeepSeek; 大语言模型; 档案智能化; 数据治理; 计算档案学

1 引言：AI 浪潮下档案工作的变革契机

人工智能技术的迭代演进正引发全球产业格局的深度重构，从 AlphaFold 重塑生物分子研究范式到 ChatGPT 重构内容生产逻辑，生成式 AI 已突破技术应用的边界，向知识生产、决策支持等核心领域渗透。2025 年推出的 DeepSeek 大语言模型，以其开源策略、轻量化部署及高效推理能力，成为继 ChatGPT 之后人工智能工程化应用的里程碑成果，其“低成本 + 高性能”的技术特性为垂直行业的智能化转型提供了全新可能。

本文结合 DeepSeek 在垂直行业的应用实践，从技术适

配性、场景创新、治理体系及学科发展四个维度，系统构建大模型时代档案工作的转型框架，为档案机构的智能化实践提供理论支撑与操作指引。

2 DeepSeek 的技术特性与档案行业的适配逻辑

DeepSeek 的技术突破并非单一维度的性能提升，而是通过架构创新实现了大模型从“实验室”到“生产线”的跨越。其 R1 与 V3 版本所展现的技术特性，与档案行业的业务需求形成高度契合，为破解长期制约档案智能化的核心瓶颈提供了可能。

2.1 低成本部署与档案行业的资源约束适配

传统大模型的训练与应用依赖巨额算力投入，动辄需要数千张 GPU 组成的计算集群，这与档案部门尤其是基层

【作者简介】林青虎（1982-），男，中国辽宁盖县人，本科，档案馆员，从事电子信息工程，档案管理研究。

机构有限的资源预算形成尖锐矛盾。DeepSeek 通过知识蒸馏技术将基础模型参数压缩至原有规模的 1/10，同时采用 PTX 中间指令集架构优化计算流程，使单节点服务器即可支持模型的本地化运行，部署成本大幅降低。这种轻量化特性，使得县级档案馆、企业档案室等资源有限的机构能够突破算力壁垒，平等享受 AI 技术红利。

2.2 开源策略与档案数据的安全需求契合

档案数据包含大量敏感信息，其管理需严格遵循“涉密不上网、上网不涉密”的原则。DeepSeek 的开源模式允许档案机构在局域网内完成模型部署与微调，避免数据上传至第三方服务器带来的泄露风险。相较于闭源模型的“黑箱”操作，开源特性使技术团队能够深度审计模型代码，排查潜在的安全漏洞，这对于涉及国家机密、商业秘密的档案资源尤为重要。

2.3 强推理能力与档案业务的复杂需求匹配

档案工作不仅涉及简单的信息提取，更需要处理跨时空、多维度的关联分析。DeepSeek 通过思维链（Chain-of-Thought）技术将复杂推理过程拆解为可解释的逻辑步骤，在档案编研、历史考证等场景中展现出独特优势。例如，在抗战时期档案整理中，模型能够基于零散的电报、日记、会议记录，自动构建事件时间线、人物关系网及物资流向图，为历史学家还原战役细节提供结构化素材。这种推理能力超越了传统检索系统的关键词匹配模式，实现了从“数据查找”到“知识发现”的跃升。

2.4 多模态处理与档案资源的形态多样性适配

档案资源已从单一的文本形态发展为包含图像、音频、视频的多模态集合，传统 OCR 技术在手写体识别、破损文档修复等方面存在局限。DeepSeek 集成的多模态认知引擎，通过生成对抗网络（GAN）修复霉变档案图像，利用 Transformer 架构实现语音档案的语义转写，将不同形态的档案资源转化为统一的知识表征。

3 DeepSeek 驱动的档案应用场景创新

DeepSeek 与档案工作的融合通过重构业务流程实现价值再造。其应用场景可分为管理效率提升与知识价值激活两大方向，覆盖档案生命周期的各个环节，形成全方位智能化解决方案。

3.1 管理效率导向的业务流程革新

档案管理中的重复性劳动长期制约工作效能，DeepSeek 通过自动化处理与智能辅助，推动业务模式从“人力密集型”向“技术驱动型”转变。

在归档环节，模型基于预训练的档案分类体系，自动识别文件类型、确定保管期限并提取著录项。针对机关文书档案，可解析红头文件的发文字号、主送机关、成文日期等元数据，生成标准化目录；

档案鉴定工作方面，DeepSeek 通过学习《各级国家档案馆馆藏档案解密和划分控制使用范围的暂行规定》等法规，结合历史鉴定案例，对档案开放等级进行初步判定。模

型不仅能识别显性敏感信息，还能通过语义分析发现隐晦表述，提升开放审核效率，减少人工复核工作量。

档案借阅服务的智能化转型则体现在智能客服与流程自动化。用户通过自然语言提问即可获取档案借阅指引，模型基于馆藏目录自动判断档案存放地点、利用限制及预约流程。对于符合条件的借阅申请，系统自动生成审批单并流转至相关部门，实现“查询-申请-审批”的全流程线上化。

3.2 知识激活导向的价值挖掘创新

DeepSeek 通过 AI 技术深度挖掘档案的历史价值，将其转化为决策支持的“活字典”。

智能编研：改变“剪刀加浆糊”模式。以 DualPipe 引擎整合散落资料，自动识别串联关键事件与核心特征，高效生成结构化初稿（如《企业三十年发展史》），大幅缩短周期，释放人力进行深度分析。

精准审核：构建多层次审核模型，综合关键词匹配、MLA 语义增强技术（识别隐晦表达）及知识图谱（排查跨文档关联风险），高效平衡开放利用与信息安全。

知识图谱化：运用实体识别与关系抽取技术，将碎片化档案要素（人物、事件、机构等）关联为可视化知识网络（如构建革命时期党组织图谱、企业技术关联网络），揭示深层联系，支撑研究与管理。

智能检索：突破关键词局限，实现语义理解。用户输入自然语言查询（如“查找 2000-2010 年员工考核办法”），系统精准解析意图，结合多维筛选与相关性排序返回结果。

DeepSeek 显著提升了档案资源的历史智慧沉淀、安全利用效率和知识发现能力。

3.3 知识激活导向的行业价值延伸

DeepSeek 通过深度挖掘档案中的规律与经验，拓展档案服务的应用边界，形成跨领域的价值辐射。

在应急管理领域，档案中的灾害应对记录可为突发事件处置提供历史借鉴。DeepSeek 通过分析灾害档案，提取预警响应时间、物资调配方案、人员转移路线等关键信息，构建应急决策模型。

企业档案的商业价值通过智能分析得以释放。模型对生产档案、销售数据、客户反馈的深度挖掘，可识别产品质量波动规律、市场需求变化趋势，为企业战略调整提供依据。

4 档案数据治理体系的重构与挑战

DeepSeek 在释放档案价值的同时，也对数据治理提出了更高要求。高质量的数据是模型发挥效能的基础，而档案数据的敏感性又决定了治理过程必须兼顾安全与利用，形成全流程的管控体系。

4.1 高质量档案数据集的构建路径

训练数据的质量直接影响模型性能，档案数据集的构建建立标准化的处理流程。数据清洗环节需剔除重复归档、残缺不全的档案，通过实体对齐技术解决同一对象的名称差异问题；数据标注则需结合档案行业规范，定义事件、人物、时间等实体标签，标注实体间的隶属、因果、协作等关系类型。

数据集的构建模式呈现多元化趋势。“通用模型+领域语料”的混合训练模式，既保留大模型的通用能力，又通过档案专业数据增强领域适配性；“联邦学习”模式则适用于跨机构的数据协作，各档案馆在不共享原始数据的情况下共同训练模型，解决“数据孤岛”问题。

4.2 数据的安全与隐私保护的体系构建

档案数据的敏感性使安全防护成为重中之重，需建立技术、制度、管理多维度的防护体系。在技术层面，采用差分隐私技术对训练数据进行匿名化处理，通过同态加密实现模型训练过程中的数据加密；在访问控制方面，基于角色的权限管理（RBAC）确保不同岗位人员只能获取权限范围内的数据，结合区块链技术记录数据操作日志，实现全程可追溯。

生成式 AI 的“幻觉”问题可能导致错误信息传播，需建立输出内容的校验机制。通过 RAG（检索增强生成）技术将模型生成结果与原始档案进行比对，验证信息的真实性；引入人工复核环节，对涉及重大决策的生成内容进行专业审核。

4.3 “人智协作”生成内容的可信性保障

AI 参与生成的档案内容，其可信性保障需突破传统电子档案“四性”的框架，建立新的评估体系。真实性方面，需记录模型的训练数据来源、算法版本及开发团队背景，通过技术审计确保生成过程的可追溯；完整性则要求保存人机交互的全流程记录，包括输入指令、参数调整、中间结果等元数据；可用性需考虑模型的鲁棒性，评估不同软硬件环境下的运行稳定性；安全性则要防范恶意指令操控模型生成虚假信息，建立异常输入检测机制。

5 计算档案学的理论演进与教育转型

DeepSeek 的应用推动档案学研究从传统范式向计算范式转变，形成理论与教育的协同革新，为学科发展注入新动能。

5.1 计算档案学的理论重构

计算档案学通过引入 AI 理论与方法，构建新的学科体系。其研究重点包括：档案数据的计算表征模型，探索如何将档案信息转化为适合机器处理的形式；人机协同的决策机制，明确人类与 AI 在档案鉴定、编研等环节的权责边界；档案知识的推理规则，建立基于逻辑与经验的混合推理体系。中南大学档案学团队提出的“档案知识图谱构建理论”，通过融合符号主义与连接主义的优势，解决了历史事件关联分析中的歧义问题，为计算档案学提供了核心理论支撑。

理论研究需回应技术带来的伦理挑战。AI 应用可能放大历史偏见，需建立算法公平性评估指标；模型的“黑箱”特性与档案工作的透明性要求存在矛盾，需探索可解释 AI 在档案领域的适配路径。

5.2 档案教育体系的智能化转型

人才培养模式需适应 AI 时代的能力需求，构建“档案业务+AI 技术+伦理素养”的复合型知识结构。课程体系中新增《大语言模型原理与应用》《档案数据标注技术》等课程，通过项目式教学培养学生的实操能力；与档案馆、科技企业合作建立实习基地，开展真实场景的模型微调、知识

库构建等实践训练。

6 面临的挑战与应对策略

DeepSeek 在档案领域的应用仍处于探索阶段，面临技术、制度、人才等多方面的挑战，需通过系统施策实现可持续发展。

6.1 技术层面的瓶颈突破

中小档案机构的数字化基础薄弱，难以直接应用先进模型。需推动“轻量化”解决方案，开发适用于低配置设备的简化版模型；建立区域档案算力中心，通过云服务为基层机构提供模型调用接口，降低应用门槛。

6.2 制度与标准的完善路径

行业标准的缺失导致技术应用碎片化。需加快制定《档案大模型训练数据规范》《智能编研成果质量评价标准》等文件，统一数据格式、接口协议与评估指标；推动档案管理系统模块化改造，实现不同厂商产品的互联互通。

6.3 人才队伍的结构优化

需建立跨学科的人才培养机制，档案院校与计算机学院联合招生，培养交叉学科研究生；引进 AI 企业的技术专家担任产业教授，将实战经验融入课堂教学。

行业人才结构需动态调整。随着 AI 替代重复性工作，档案人员的角色将向“AI 训练师”“知识工程师”转变，需通过职业技能认证引导转型方向；

7 结论与展望

DeepSeek 的出现标志着档案工作进入“数智融合”新阶段。其低成本、开源化特性打破技术垄断，推动行业从“差异发展”迈向“普惠智能”。在应用层面，模型通过重构流程提升管理效率，激活知识价值拓展服务边界，形成“效率-价值”双轮驱动。

未来档案智能化将呈现三大趋势：一是从“单点应用”向“全域智能”演进，实现全流程智能化；二是从“辅助工具”向“协作伙伴”转变，人机协同中历史洞察力与 AI 算力互补；三是从“行业应用”向“生态构建”升级，通过跨域数据融合形成知识服务体系。

档案工作者需平衡技术理性与人文关怀：既拥抱效率提升，又坚守真实性底线；既推动理论创新，又传承历史使命。通过技术与制度创新协同，DeepSeek 等 AI 技术将成为档案事业高质量发展的引擎，使档案资源在记录历史、服务当下、启迪未来中释放更大价值。

参考文献

- [1] 刘越男,钱毅,王平等.挑战与展望: DeepSeek 对档案工作的影响及应用前景[J].浙江档案,2025(02).
- [2] 杨建梁,郑梦霏,杨格秀.DeepSeek在档案管理领域中的应用与展望[J].机电兵船档案,2025(01).
- [3] 刘洋.“大模型+RAG”技术在档案工作中的应用探析[J].中国档案,2025(03).
- [4] 刘华英,扈桂红.基于知识图谱的我国档案管理热点及趋势研究[J].资源信息与工程,2019(10).