

Research on Digital Archiving Applications in Higher Education Institutions Empowered by DeepSeek Intelligent Technology

Juan Li

Archives, Zhongyuan University of Technology, Zhengzhou 450007, CHINA
Corresponding Author: Juan Li

ABSTRACT: Under the guidance of China's national "AI+" initiative strategy, the domestically developed large-scale model DeepSeek leverages its unique advantages—such as open-source compatibility, lightweight architecture, and localized deployment capabilities—to provide a novel technical pathway for addressing the traditional pain points in university archives management. Based on real-world operational scenarios in university archives work, this study constructs a comprehensive technical framework encompassing multimodal intelligent classification of archives, dynamic intelligent retrieval via knowledge graphs, and full-layer security protection. Notably, the multimodal knowledge graph technology applied in the framework has been verified by practical patent achievements: DarkMatter AI applied for a patent for a multimodal knowledge graph construction method in 2024, which can improve the integrity and accuracy of knowledge graphs through standardized data extraction and storage processes; Tencent also obtained an authorized patent for multimodal knowledge graph construction in 2024, realizing rapid construction of refined multimodal knowledge graphs. On this basis, the study further develops the DeepSeek Intelligent Archives Management Solution, which is fully tailored to the operational needs of universities, thereby driving a high-quality transformation of university archives work from "passive preservation" to "active knowledge services."

Date of Submission: 02-09-2026

Date of acceptance: 12-09-2026

I. INTRODUCTION

University archives, as core resource hubs responsible for documenting the institutional development history, supporting teaching and research activities, and preserving the university's historical culture, have long faced challenges. With the rapid development of information technology, university archive resources have developed in multiple directions, with increasing data volume and big data characteristics, resulting in low efficiency in processing such multi-source heterogeneous data and insufficient retrieval precision. Meanwhile, they also face weak compatibility with security protection measures, as there are hidden security risks such as imperfect infrastructure, lack of professional talents, and absence of security systems in university archives, and the archive information management system is in a dilemma of security and overload [1]. According to empirical statistics show that the traditional OCR recognition accuracy for most university archives consisting of damaged handwritten paper documents is only 62%, which is consistent with the performance of traditional OCR when dealing with complex handwritten documents like those with scribbled handwriting, faded ink, as its recognition accuracy often drops below 60% as noted in relevant technical practices; furthermore, the average manual response time for retrieving records across different departments exceeds two working days, and similar to the situation where staff spend a lot of time manually proofreading due to low OCR accuracy, such inefficient manual retrieval also causes great work pressure. The traditional approach can no longer meet the governance requirements of universities in the digital and intelligent era [2]. With the maturation and practical application of domestic large language models Technologies such as DeepSeek [3–4], which are open-source, free-to-use, and highly efficient, can significantly enhance the quality and efficiency of archival management when applied to this field, which aligns with the working principle of the National Archives Administration to "safely explore AI-enabled archival management" [5]. To this end, our archive department has implemented an intelligent transformation of archival management using DeepSeek (as shown in Fig. 1), achieving excellent results.

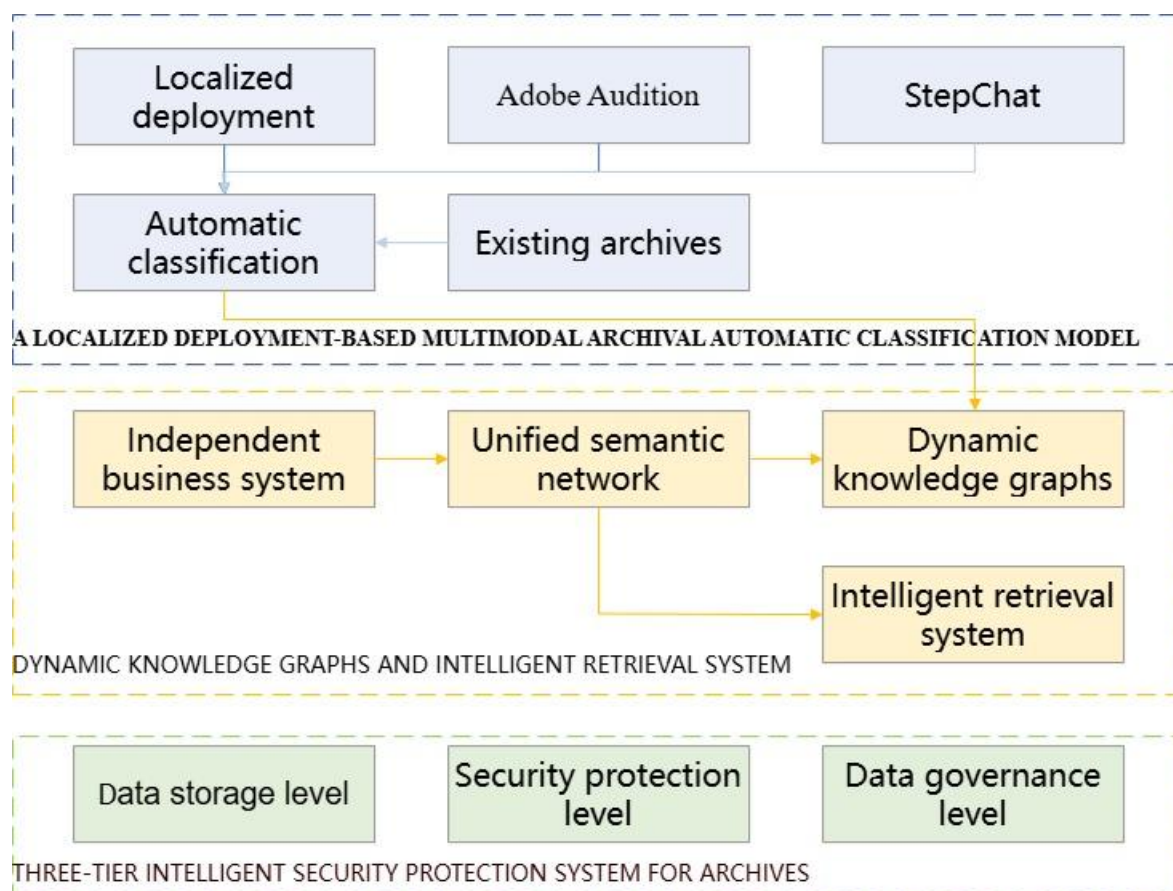


Fig. 1 Research flowchart

II. A LOCALIZED DEPLOYMENT-BASED MULTIMODAL ARCHIVAL AUTOMATIC CLASSIFICATION MODEL

By leveraging the existing 16-card GPU servers available at the university, DeepSeek-V4 is deployed; through model distillation, the memory footprint per card is controlled within 16GB. As revealed by relevant tests, the official version of DeepSeek-V4 has achieved a technical breakthrough of reducing video memory usage by 10%, and institutional research reports predict that this model will effectively break through the 'memory wall', eliminating the need for additional procurement of expensive computing clusters and demonstrating high cost-effectiveness in deployment. In conjunction with Adobe Audition, the system performs automatic noise reduction and human voice separation on audio archives—including oral histories of senior professors and recordings of academic conferences; combined with Yuewen StepChat's interactive verification capabilities, it can perform contextual logic correction and automatic classification on aged handwritten archives and traditional school history materials, thereby increasing the accuracy of traditional OCR recognition to over 97%, reducing the manual verification workload by 60%, and completing the digitization of existing archives—a task previously estimated to take 18 months—merely 6 months. The system incorporates a built-in rule library based on the "Archival Entity Classification System for Higher Education Institutions," supports custom extensions to classify dimensions specific to various disciplines, and achieves an annual archive collection completeness rate of up to 99.3%.

III. DYNAMIC KNOWLEDGE GRAPHS AND INTELLIGENT RETRIEVAL SYSTEM

The system is equipped with a universal heterogeneous data adapter and unified semantic network, which enables seamless integration with existing DeepSeek integrates with independent business systems—such as university academic affairs, research, and personnel management—without requiring any secondary modifications; it automatically extracts decentralized data, including records of student status changes and final reports for research projects. DeepSeek can automatically perform entity extraction and relationship mapping for newly archived records, updating knowledge graph nodes in real time, thereby completely resolving the longstanding issue of delayed data synchronization and knowledge graph updates. Upgrades to traditional knowledge graphs. Users can initiate queries using natural language to obtain results within seconds; for example, when searching for "the overview of all Henan Provincial Key Laboratories led by our university since

its establishment," the system can automatically link project initiation and acceptance archives from different eras to generate a complete timeline of development. Both the completeness and precision rates of archival retrieval have exceeded 95%. The implementation The practice of our university's "AI-powered Archives" demonstrates that this model can support 24/7 uninterrupted service, reducing the standard archival retrieval response time from 2 days to merely a few seconds.

IV. THREE-TIER INTELLIGENT SECURITY PROTECTION SYSTEM FOR ARCHIVES

Establish an intelligent security protection system for archives from three levels, including data storage, security protection and data governance. At the data storage layer, a multi-node distributed blockchain architecture is adopted, ensuring that all writing and retrieval operations involving sensitive archives are fully recorded on the blockchain for evidentiary purposes preservation—thereby achieving immutability and millisecond-level traceability. Combined with a three-tier storage strategy consisting of "local all-flash storage + cloud object storage + cold tape archiving," this approach improves storage resource utilization by 65%. As relevant researches show, tape archiving is a cost-effective method for long-term data preservation, and according to IDC's prediction, about 60% of data will be cold data, which further highlights the necessity of such a three-tier storage strategy. The security protection layer incorporates an anomaly detection model that identifies suspicious activities—such as high-frequency batch downloads and unauthorized access—in real time, automatically triggering circuit breaker alerts; it also implements differentiated dynamic desensitization strategies for archives of different classification levels, automatically obfuscating sensitive fields such as ID numbers and classified technical parameters. This dynamic desensitization can carry out real-time processing based on identity permissions while maintaining the original data features, business rules and data relevance, avoiding impact on subsequent business use. The data governance layer encompasses functions including archive classification and grading, automatic desensitization, and cross-departmental collaborative auditing. With multi-factor access control and sensitive data access permission management capabilities, it can effectively prevent illegal access and data disclosure risks, and enables the generation of compliant audit reports with a single click, thereby fully meeting the compliance requirements under regulations such as the Personal Information Protection Law and the Data Security Law, compliance management requirements stipulated by the Archives Law.

V. CONCLUSION

This intelligent archival management solution for universities, built upon DeepSeek, fully aligns with China's strategic focus on developing self-owned and controllable technologies. Upon implementation, it significantly enhances archival processing efficiency, reduces manual archival processing costs, transforms dormant archival resources into dynamic knowledge assets that can be easily accessed, and establishes a comprehensive security protection framework covering the entire workflow. This solution features low investment costs and a short implementation cycle, making it perfectly compatible with the varying information technology infrastructure of universities at different levels. It provides a replicable practical model for the intelligent transformation of university archives nationwide, thereby helping universities truly realize the core value of their archival work—"preserving history, supporting governance, and fostering talent."

REFERENCES

- [1]. Duan Hongping. "A Study on How Artificial Intelligence Technology Can Empower the High-Quality Development of Archival Management in Higher Education Institutions" *LanTai WaiNei*, Issue 1: pp.32–34.
- [2]. Yang Jianyun, Li Ruomeng. "Optimization of University Archives Information Services Driven by DeepSeek Large Language Model Technology" *Academic Research*, Issue 2: pp.24–29.
- [3]. Wei Yan. "Research on the Application of DeepSeek Large Language Model in Archival Management in University Financial Management" *Shandong Textile Economics*, Issue 1: pp.49–53.
- [4]. Chen Weidong, Yang Dan. "An Exploration Study on the Application of DeepSeek in University Archives Work" *Aviation Archives*, pp.68–71.
- [5]. Qian Yi, An Xinyu. "Steadily Implementing AI-Enhanced Archival Initiatives and Actively Promoting the Intelligent Transformation of Archival Work: Interpretation of the 'Opinions on Promoting the Application of Artificial Intelligence in the Archival Sector'" *Beijing Archives, Regulations and Standards*, Issue 6: pp.41–46.