WEP Warwick Evans Publishing

International Journal of Social Sciences and Public Administration

ISSN: 3005-9836 (Print), ISSN: 3005-9585 (Online) | Volume 10, Number 2, Year 2026
DOI: https://doi.org/10.62051/ijsspa.v10n2.03
Journal homepage: https://ijsspa.org

Generative Artificial Intelligence Empowers the Development of Archival Information Resources: A Study and Analysis based on Cases in Chinese Archival Sector

Xueqi Ren

School of Public Administration, Sichuan University, Chengdu 610065, China

ABSTRACT

Against the backdrop of digital transformation in the archival undertaking in Chinese archival sector, generative artificial intelligence (GenAI) offers a new technological pathway for the development of archival information resources. This study explores the application of GenAI in this domain, drawing upon representative cases from Chinese archival sector. It systematically analyzes the primary challenges of traditional archival development models concerning efficiency, format, and service delivery. Furthermore, it examines specific practical applications of GenAI in scenarios such as intelligent Q&A, intelligent compilation and research, archival restoration, and multimodal content creation. The research reveals how GenAI is driving transformations in archival development methodologies, service models, and the forms of outcomes. Concurrently, it identifies new challenges posed by GenAI regarding data security, technical thresholds, ethical norms, and talent development. Based on these findings, the paper proposes countermeasures and suggestions focusing on top-level design, data foundations, technical pathways, and talent cultivation, aiming to provide references for the intelligent transformation of archival institutions in China.

KEYWORDS

Generative Artificial Intelligence; Archival Information Resource Development; Application Scenarios; Case Analysis.

1. INTRODUCTION

With the comprehensive and deepening advancement of the Digital China initiative, archival work is facing a critical juncture of transformation from traditional management paradigms to digital and intelligent ones. Generative Artificial Intelligence (GenAI), as a strategic technology leading the new wave of scientific and technological revolution and industrial transformation, offers new possibilities for overcoming bottlenecks in the development of archival information resources and enhancing their utilization efficiency through its powerful content generation and semantic understanding capabilities.

In 2021, *the 14th Five-Year National Plan for Archival Development* not only prioritized enhancing archival service capabilities and accelerating the construction of digital archives, but also explicitly called for "actively exploring the application of technologies such as knowledge management, artificial intelligence, and digital humanities in the deep processing and utilization of archival information."[1] The National Archives Directors' Conference in early 2025 further emphasized the need to "leverage new-generation technologies like big data, cloud computing, and artificial intelligence," specifically identifying "AI-assisted review" as a new approach to expedite archival declassification.[2] Concurrently, the demands of the public and archival institutions at all levels for archival information services are shifting from mere availability to quality. The traditional model of archival information resource development, characterized by heavy reliance on manual labor, lengthy

processes, and homogeneous outputs, struggles to meet the higher expectations of the new era for realizing archival value. It faces multiple challenges, including low development efficiency, insufficient depth in knowledge mining, passive and lagging service models, and a lack of innovation in cultural dissemination formats.

Generative Artificial Intelligence, a forefront branch of AI, specifically refers to technologies capable of learning from data to generate novel, original content such as text, images, and audio. Leveraging its advancements in natural language processing, multimodal content generation, and logical reasoning, GenAI can not only relieve archivists from repetitive tasks like information retrieval and content compilation but also directly contribute to core processes such as knowledge distillation, content creation, and format innovation through deep semantic understanding and relational analysis of archival data. This drives the transformation of archival information resource development from information organization to knowledge generation, and from static preservation to dynamic presentation.

Against this backdrop, this paper, grounded in the application practices of GenAI within Chinese archival sector, aims to systematically outline the core dilemmas of traditional archival information resource development. By analyzing pioneering cases, it delves into the specific applications of GenAI across various scenarios, elucidates the emerging trends and challenges it introduces, and ultimately provides evidence-based and practical recommendations for archival institutions to proactively plan and implement GenAI technologies.

2. DILEMMAS FACING TRADITIONAL ARCHIVAL INFORMATION RESOURCE DEVELOPMENT

Although national policies consistently promote the in-depth exploration of archival resource value, at the practical level, the traditional model of archival information resource development is constrained by technological means and management paradigms. It encounters bottlenecks across multiple dimensions-efficiency, format, and service-which hinder the depth of development and the breadth of dissemination of archival information resources, making it difficult to fully meet the demands of the new era.

2.1. Efficiency Bottlenecks

The process of traditional archival information resource development remains heavily reliant on manual operations. Key stages such as theme extraction, material arrangement, and information verification require substantial investments of manpower and time, exhibiting a "labor-intensive" nature that limits the scale and speed of development. Core tasks like thematic compilation and research or declassification review depend profoundly on the professional expertise and manual efforts of archivists. A single in-depth compilation project, from topic selection, retrieval of records, and research to final drafting, can easily take several months, rendering it incapable of providing rapid responses to emergent decision-making needs or social hotspots. A vast amount of foundational, repetitive work-such as metadata cataloging, index creation, and sensitive information screening-consumes the majority of professionals' energy, leading to a misallocation of resources where "highly skilled personnel are mired in low-complexity tasks."

2.2. Format Limitations

In terms of output forms, traditional development primarily yields static products such as printed compilations, archival collections, and thematic exhibitions. This unidirectional mode of information delivery lacks interactivity and appeal, failing to satisfy the public's demand for visualization, interactivity, and immediacy in the information society. While some archival institutions have attempted multimedia formats like digital exhibitions, the content often remains limited to the simple

presentation of images and text, lacking multimodal capabilities and deep narrative power. The historical and cultural values embedded within archives tend to be flattened by traditional narrative techniques, struggling to capture the interest of the public, particularly younger generations. This, to some extent, constrains the reach and impact of archival cultural dissemination.

2.3. Service Lag

Traditional archival service models predominantly operate on a passive, reactive basis-"user queries, archivist retrieves." This model makes it difficult to construct user profiles, thereby hindering the delivery of precise knowledge delivery and proactive services. Consequently, it falls short of meeting users' diverse and personalized knowledge acquisition needs. Particularly in cross-departmental or cross-regional usage scenarios, insufficient interoperability between archival information systems leads to "information silos," obstructing the correlation and integration of knowledge. Users often receive only fragmented information rather than systematic knowledge, leaving the task of deep knowledge mining and synthesis to themselves. This situation stands in contrast to the goal of building a convenient, knowledge-oriented service system.

In summary, the traditional model of archival information resource development reveals shortcomings in its operational mechanisms, output forms, and service models. The root cause lies in the dual effects of lagging technological updates and the inertia of management practices, which prevent the full realization of the knowledge and social value of archival resources. The application of emerging technologies like Generative Artificial Intelligence provides a novel technological pathway to address these dilemmas.

3. SPECIFIC APPLICATION SCENARIOS OF GENERATIVE ARTIFICIAL INTELLIGENCE IN ARCHIVAL INFORMATION RESOURCE DEVELOPMENT

In recent years, the application of Generative Artificial Intelligence (GenAI) in archival information resource development has gradually transitioned from the experimental exploration phase to practical implementation. Based on its technical characteristics and industry trends, four representative application scenarios have emerged, each reshaping the working methods of archival development from the dimensions of service efficiency, knowledge mining, resource restoration, and innovative dissemination.

3.1. Intelligent Q&A and Reference Inquiry

Traditional archival retrieval heavily relies on keyword matching from manual cataloging, which limits both recall and precision rates and places high demands on users' professional search skills. The introduction of GenAI, through the construction of intelligent Q&A systems based on natural language interaction, has optimized and upgraded reference inquiry services.

The core of such applications is the Retrieval-Augmented Generation (RAG) technical framework. The system first builds a local knowledge base from digitized archival holdings, including full texts, catalogs, and metadata. When a user poses a question in natural language, the system retrieves relevant content from the knowledge base based on semantic understanding. This content is then passed to a large language model (LLM) for in-depth analysis, synthesis, and generation, ultimately outputting structured answers, analytical conclusions, and accurate source references. The system is capable of understanding complex user queries, correlating relevant policy documents and meeting minutes, and generating comprehensive analytical reports, rather than merely listing documents.

3.2. Intelligent Compilation & Research and Knowledge Generation

Archival compilation and research constitutes a professional task involving the collection, selection, and processing of archives possessing research or practical value, transforming them into compiled outputs in various formats. The traditional model is time-consuming, labor-intensive, and heavily reliant on expert experience. The application of Generative Artificial Intelligence (GenAI) in this domain is shifting archival compilation and research towards an intelligent-driven process of "knowledge reproduction."

Specifically, Large Language Models (LLMs) can automatically extract entities and relationships from unstructured full-text archival data, thereby facilitating the construction of archival knowledge graphs. Building upon this foundation, GenAI can assist in several key functions: Firstly, intelligent topic selection and outline recommendation. By analyzing high-frequency themes and their interrelationships within archival collections, it can identify research gaps and suggest potential directions for compilers. Secondly, automated content generation. Based on a selected topic, it can produce initial drafts of chronicles, organizational histories, or thematic reports, significantly reducing the manual writing workload. Thirdly, deep relational analysis. It can uncover historical connections hidden across different fonds or categories of archives, providing novel perspectives for compilation and research.

3.3. Archival Restoration and Information Enhancement

When dealing with aged or physically damaged archival carriers, Generative Artificial Intelligence (GenAI) offers unique advantages in digital restoration and information enhancement. Current applications extend beyond simple image repair to achieve in-depth restoration and enhancement of the informational content within archives.

In processing photographic archives, deep learning-based technologies can perform super-resolution reconstruction, scratch repair, and color restoration on blurry, damaged, or faded historical photographs and documents, thereby revealing details with greater clarity.[3] For audio archives, AI techniques, particularly those utilizing audio diffusion models, can effectively separate background noise and restore damaged audio tracks. More importantly, in the domain of text recognition and enhancement-addressing common issues in historical documents such as traditional Chinese characters, mixed fonts, handwritten script, and ink bleed-through or fading-the combination of GenAI with specialized OCR models leverages contextual semantics for intelligent recognition and correction, significantly improving accuracy. This lays a crucial foundation for subsequent data processing and knowledge extraction. As a result, a vast quantity of previously "machine-unreadable" valuable archives can undergo a "digital rebirth," providing robust technical support for their development and utilization.

3.4. Multimodal Content Creation and Dissemination

To overcome the challenges of monotonous formats and limited interactivity in archival cultural dissemination, Generative Artificial Intelligence demonstrates powerful cross-modal capabilities for automated content creation and innovative storytelling. It can transform standalone textual archives into immersive, multimedia cultural products.

Key applications include: First, cross-modal retrieval methods such as "search by image" and "audio retrieval." Leveraging computer vision and audio processing technologies, systems can automatically identify and link figures, scenes, and speech within photos and videos, allowing users to search related content using visual or auditory inputs. Second, virtual digital narrator guides. Generating virtual personas based on archival content to provide interactive guided tours and narration for online exhibitions. Third, dynamic knowledge graph visualization. Presenting complex relationships between individuals and events through interactive, visual diagrams, making historical contexts and

connections immediately clear. Fourth, intelligent content creation. Utilizing "text-to-video" and "image generation" technologies to automatically convert suitable archival resources into integrated media products such as promotional short videos and social media graphics.

These applications significantly diversify the methods for developing and utilizing archival resources. They empower archival institutions to engage with the public in more dynamic and contemporary ways, effectively boosting the influence and reach of archival culture and facilitating the revitalized use of archival resources.

4. PRACTICAL ANALYSIS OF CASES IN CHINESE ARCHIVAL SECTOR

4.1. Zhejiang Provincial Archives: The Smart Archives Model Featuring Systematic Development and Regional Collaboration

Guided by policy and top-level design, the Zhejiang Provincial Archives has progressively advanced from a digital archives to a smart archives model through systematic development and regional collaboration. Its construction, spanning nearly two decades, has unfolded in four phases: starting with initial exploration in 2005 by integrating into the "Digital Zhejiang" initiative, moving through standardized development during the "12th Five-Year Plan" period, achieving the milestone of becoming the nation's first provincial-level demonstration digital archives in 2019, and finally, since 2021, leveraging digital reform as the driving force to establish a provincial Archival Data Sharing Center and integrate large language model technologies like DeepSeek. This evolution marks an upgrade from digitalization to digital intelligence, demonstrating the synergistic advancement of policy, technology, and operational practice, thereby providing a replicable pathway for the intelligent transformation of provincial-level archives.[4]

At the level of technological implementation and application, the core characteristic of the Zhejiang Provincial Archives lies in its deeply integrated and converged platform based on a cloud-middleware architecture. The Provincial Archival Data Sharing Center, established in 2023, integrates various AI technologies including Optical Character Recognition (OCR), Natural Language Processing (NLP), and machine learning, creating a fertile ground for the deployment of DeepSeek. Specifically, its applications focus on four core scenarios: First, it optimized the AI-assisted declassification review process using NLP, enhancing the transparency and adjustability of review procedures. Second, it introduced Retrieval-Augmented Generation (RAG) technology to build an archival business knowledge base, enabling semantic-based natural language intelligent Q&A. Third, it utilized DeepSeek to extract entities and relationships from unstructured archives, automatically constructing thematic knowledge graphs to transform archival resources into machine-understandable and processable structured knowledge. Fourth, it implemented intelligent archival categorization through deep semantic analysis, significantly improving the efficiency of archival resource management. These applications are not isolated functionalities but are interconnected, collectively forming an intelligent ecosystem that covers the entire archival workflow of "acquisition, management, preservation, and use."

Regarding interconnectivity, the Zhejiang Provincial Archives has broken through the limitations of single repository collections, pioneering a new model for province-wide archival resource interconnection and collaborative service. Using the "Zheli Archives" application as a breakthrough point, it established an archival data sharing mechanism, realizing "one-network search, hundred-archives linkage." It has also connected data interfaces with public service departments such as human resources and social security, and civil affairs. Furthermore, it has achieved integration with platforms like the Zhejiang Government Service Network and the Yangtze River Delta's "cross-regional online service" system. This promotes the cross-regional and cross-departmental sharing of archival resources on a broader scale. This construction philosophy, oriented towards sharing and centered on

service, has transformed it from a static "information warehouse" preserving history into an active data hub that empowers government governance and public services.[5]

4.2. Zhuhai Municipal Archives: Building an Intelligent System Based on "Full-Stack Domestic Technology" with Security and Controllability

The practice of the Zhuhai Municipal Archives highlights a distinct characteristic of intelligent transformation under specific policy and security constraints: a foundation of full-stack domestic technology with a core focus on security and controllability. Since being designated as a "National Demonstration Digital Archives" in 2015, it has continuously deepened its application effectiveness, driving the comprehensive completion of its digital archives system. During the comprehensive system upgrade initiated in 2021, the Zhuhai Municipal Archives resolutely adopted a full-stack domestic technology roadmap. From the Kunpeng 920 chip and the KylinOS operating system to databases, middleware, and security protection systems, it achieved comprehensive autonomy and controllability across hardware, software, and data. This foundational choice established a secure and trustworthy technological base for all subsequent intelligent applications.

In the domain of data security, through in-depth communication with the administrative body overseeing the government cloud, the archives successfully established a physically isolated, independent partition within the government cloud's administrative server room. This formed a three-tier network architecture comprising the "Internet-Government Network-Isolated Local Area Network." The domestically produced server clusters deployed here not only compensated for the initial shortcomings in domestic computing power through load balancing but also achieved high availability and disaster recovery via cluster deployment. This unique "air-gapped" architecture eliminated the risk of archival data exposure to external networks, ensuring robust data security.

Regarding the integration and application of AI technologies, the Zhuhai Municipal Archives demonstrates a "multi-layered and deeply embedded" approach. Long before the local deployment of the DeepSeek large language model, it had already extensively utilized various traditional AI technologies to address operational pain points. For instance, in intelligent-assisted archiving, it combined OCR, NLP, and Hidden Markov Models to achieve automatic classification of electronic records and prediction of retention periods. In news video processing, it employed speech recognition, facial recognition, and automatic segmentation technologies to enhance daily video processing efficiency. For declassification review, it utilized the DFA algorithm to improve the efficiency of sensitive term screening. These technologies were not isolated but were tightly integrated with business processes, establishing efficient mechanisms combining machine preliminary review with human re-inspection. The successful deployment of DeepSeek in 2025 further established a heterogeneous computing platform based on "Chip Server Clusters + Inference Cards." Its capabilities were then infused into every stage, from intelligent compilation and research to multimodal intelligent retrieval, providing a valuable reference for other archives regarding domestic and intelligent system development.[6]

4.3. Changping District Archives of Beijing: Grassroots Practice of Lightweight Deployment and Business Process Integration

The practice of the Changping District Archives in Beijing offers a pragmatic and replicable pathway for grassroots archival institutions to introduce Generative Artificial Intelligence (GenAI). It emphasizes substantial improvement in the efficiency of archival information resource development by deeply integrating large language model technologies into existing workflows through rational resource allocation, without pursuing extreme computing power or massive parameters.

Its deployment solution exhibits distinct "lightweight" characteristics. At the hardware level, the institution did not procure top-tier computing clusters; instead, it utilized two NVIDIA 3090 graphics cards as the core computing support, demonstrating that medium-level hardware can adequately

sustain model operations at a certain scale. For model selection, it opted for the DeepSeek R1 version with 7 billion parameters. This model offers a balanced performance in semantic understanding and multimodal processing while presenting relatively lower requirements for the hardware threshold and operational costs associated with local deployment. The technical architecture is based on the mature J2EE framework and Java programming language, ensuring system stability and scalability. This pragmatic resource allocation strategy enables district/county-level archives to adopt cutting-edge AI technology within relatively constrained budgets. Its core logic lies in achieving the intelligent transformation of key business processes with minimal viable technological investment, laying a solid foundation for subsequent business applications.

Leveraging a knowledge base trained on selected fonds, the Changping District Archives concentrated resources on developing several functional modules. The multimodal retrieval module achieved a leap from catalog to full-text search. It not only supports OCR recognition of both printed and handwritten text but also enables capabilities such as search by image, video search by image, and voiceprint retrieval, thereby expanding the dimensions of archival search. The AI-powered knowledge Q&A module can answer user queries based on the knowledge base content and provide references with archival identifiers. The AI-assisted declassification review module utilizes semantic analysis technology, combined with a pre-defined sensitive term database, to provide reference suggestions for manual review.

The practice of the Changping District Archives has achieved notable results summarized as "Three Increases and Three Decreases": namely, increases in management efficiency, service quality, and security effectiveness; and decreases in labor costs, error rates, and barriers to access. This case illustrates that for grassroots archival departments, adopting a strategy of "lightweight deployment" can also successfully drive the digital transformation of archival information resource development work.

4.4. Changping District Archives of Beijing: Grassroots Practice of Lightweight Deployment and Business Process Integration

The practice of the Changping District Archives in Beijing offers a pragmatic and replicable pathway for grassroots archival institutions to introduce Generative Artificial Intelligence (GenAI). It emphasizes substantial improvement in the efficiency of archival information resource development by deeply integrating large language model technologies into existing workflows through rational resource allocation, without pursuing extreme computing power or massive parameters.

Its deployment solution exhibits distinct "lightweight" characteristics. At the hardware level, the institution did not procure top-tier computing clusters; instead, it utilized two NVIDIA 3090 graphics cards as the core computing support, demonstrating that medium-level hardware can adequately sustain model operations at a certain scale. For model selection, it opted for the DeepSeek R1 version with 7 billion parameters. This model offers a balanced performance in semantic understanding and multimodal processing while presenting relatively lower requirements for the hardware threshold and operational costs associated with local deployment. The technical architecture is based on the mature J2EE framework and Java programming language, ensuring system stability and scalability. This pragmatic resource allocation strategy enables district/county-level archives to adopt cutting-edge AI technology within relatively constrained budgets. Its core logic lies in achieving the intelligent transformation of key business processes with minimal viable technological investment, laying a solid foundation for subsequent business applications.

Leveraging a knowledge base trained on selected fonds, the Changping District Archives concentrated resources on developing several functional modules. The multimodal retrieval module achieved a leap from catalog to full-text search. It not only supports OCR recognition of both printed and handwritten text but also enables capabilities such as search by image, video search by image, and voiceprint retrieval, thereby expanding the dimensions of archival search. The AI-powered

knowledge Q&A module can answer user queries based on the knowledge base content and provide references with archival identifiers. The AI-assisted declassification review module utilizes semantic analysis technology, combined with a pre-defined sensitive term database, to provide reference suggestions for manual review.

The practice of the Changping District Archives has achieved notable results summarized as "Three Increases and Three Decreases": namely, increases in management efficiency, service quality, and security effectiveness; and decreases in labor costs, error rates, and barriers to access. This case illustrates that for grassroots archival departments, adopting a strategy of "lightweight deployment" can also successfully drive the digital transformation of archival information resource development work.[7]

5. NEW TRENDS AND EMERGING ISSUES

5.1. New Trends

(1) The Intellectualization of Development Methods. Generative Artificial Intelligence (GenAI) is shifting archival information resource development from reliance on human expertise to technology-driven processes. Tasks traditionally dependent on the experience of archivists are being upgraded by intelligent technologies such as natural language processing, knowledge extraction, and generative models, significantly enhancing production efficiency and content quality. Research indicates that the semantic search capabilities of models like DeepSeek facilitate deeper semantic understanding, focusing not only on keywords but also on underlying intent and contextual information, outperforming traditional methods.[8] Future archival development will increasingly rely on the synergistic generation enabled by algorithms and data.

(2) The Proactive Transformation of Service Models. The application of GenAI enables a shift in archival services from "passive response" to "proactive push." While traditional services rely on user queries and manual retrieval, AI-powered systems can predict potential information needs based on user profiles, semantic understanding, and behavioral data, actively generating personalized service solutions. Furthermore, GenAI endows archival services with greater interactivity and visual capabilities. New formats such as virtual digital human narration and immersive exhibitions allow archival information dissemination to transcend temporal and spatial constraints, transforming into an intelligent, experiential form of knowledge service.[9]

(3) The Diversification of Outcome Forms. GenAI is driving the evolution of archival information products from static and homogeneous formats to dynamic and diversified ones. While past archival outcomes primarily consisted of compilations, research papers, and physical exhibitions, they are now expanding into multidimensional forms such as videos, images, visualizations, and interactive experiences. Scholars argue that GenAI achieves a "secondary representation" or revitalization of archives for audiovisual purposes through "content restructuring," promoting the integration and reuse of archival content with film and television resources.[10] This diversification of outcomes not only strengthens the dissemination power of archival culture but also fosters the integrated development of archival resources with industries such as education, tourism, and public cultural services. Archival information resource development is moving from "single-format output" towards "multi-dimensional knowledge presentation," creating a virtuous cycle between archival dissemination and societal understanding.

5.2. Emerging Issues

(1) Data Security and Privacy Risks. Generative Artificial Intelligence (GenAI) may introduce security risks such as data breaches and fabricated content when processing archival data. Given that archival information often involves personal privacy, state secrets, and internal institutional matters,

any unauthorized model training or data access could lead to severe consequences. Research indicates that GenAI presents potential hazards related to "data security" and "fabricated generation" during both the data collection and output stages.[11] Therefore, archival institutions must establish rigorous security protocols and data usage review mechanisms when implementing AI systems.

(2) Technical Thresholds and Cost Pressures. The deployment and maintenance of GenAI require substantial investment in computing resources and technical expertise, posing a significant challenge for grassroots archival departments. In the process of adopting GenAI, attention must be paid not only to the direct costs of hardware and software procurement but also to indirect costs associated with subsequent personnel training and system maintenance.[12] A lack of long-term support may result in systems that are "built but not used" or "used but not optimized," undermining the sustainability of digital initiatives. Balancing technological innovation with budgetary constraints remains a considerable and unavoidable challenge for grassroots archival institutions.

(3) Ethical and Copyright Challenges. GenAI's content generation capabilities may impact the authenticity inherent in archival work. First, AI models possess an inherent "generative" nature, where outputs are often based on probabilistic reasoning rather than absolute facts. This may lead to the generation of texts containing a degree of fabricated detail. Without a robust human review mechanism, AI-generated archival products risk disseminating inaccuracies and undermining the evidential value of archives. Second, the use of archival data for model training may raise questions of originality, and the copyright ownership of AI-generated content remains contentious.[13] Consequently, archival institutions should establish ethical review frameworks for AI applications, clarify accountability for ensuring the authenticity of generated content, strengthen human oversight, and ensure that AI technology serves the faithful representation of archival materials rather than producing "AI hallucinations."

(4) Adaptation Pressure on the Professional Workforce. The proliferation of GenAI is reshaping the archival profession. On one hand, some responsibilities within traditional roles such as compilation, research, and arrangement are being automated. On the other hand, demand is rapidly growing for interdisciplinary professionals proficient in both technology and archival science. Without systematic training, archival staff may struggle to utilize new technologies effectively, hindering the realization of genuine "human-AI collaboration." Therefore, strengthening team development and cultivating digital literacy are essential to ensuring the successful implementation of AI applications.

6. COUNTERMEASURES AND SUGGESTIONS

6.1. Strengthen Top-Level Design and Normative Guidance

The application of Generative Artificial Intelligence (GenAI) is not merely a technical issue but a systematic project involving policy, law, and ethics. At the national level, consideration should be given to formulating dedicated guiding documents for AI applications in the archival domain, clarifying its scope of use in processes such as acquisition, management, development, and utilization. Referencing relevant provisions of the Interim Measures for the Management of Generative Artificial Intelligence Services and considering the specific characteristics of archival work, guidelines for GenAI application in archives should be issued. These guidelines must address critical issues such as data training permissions, accountability for generated content, and user privacy protection. Concurrently, mechanisms for algorithm filing and explainability assessment should be promoted to prevent the "algorithmic black boxes" of AI from compromising the authenticity of archival information. Local archival departments can subsequently formulate regional implementation rules, establishing a coherent, tiered policy system with clear responsibilities.

6.2. Consolidate the Data Foundation

The efficacy of GenAI is highly dependent on the quality and scale of training data. Currently, archival data still suffers from inconsistent quality and the phenomenon of "data silos," which constrains the training and application of AI models. Regarding data quality, a quality control system covering the entire data lifecycle should be established. To address issues like non-standard archival data, unified digitization standards and metadata specifications need to be developed, with strict validation mechanisms implemented during data collection and entry phases. Simultaneously, AI-assisted data cleaning and annotation tools can be introduced to standardize historical stock data, systematically building a high-quality archival data repository. Concerning data sharing, regional or thematic archival data-sharing platforms should be established under the premise of ensuring data security and privacy, promoting the interconnection and interoperability of data resources.

6.3. Explore Inclusive and Flexible Technical Pathways to Lower the Application Barrier

Facing the significant cost pressures associated with GenAI applications, archival institutions need to adopt intensive and efficiency-oriented technical pathways to achieve a balance between cost control and performance optimization. Actively exploring lightweight AI deployment models focused on the intelligent transformation of core business scenarios can achieve "low investment, quick results."

At the technical level, actively adopting cloud computing services and lightweight models is advisable. A hybrid cloud architecture can keep core data within a local private cloud while utilizing the elastic computing power of public clouds for non-sensitive tasks, effectively reducing hardware investment and operational costs. At the management level, a gradual strategy of "pilot projects first, phased implementation" should be adopted. Priority should be given to conducting small-scale pilots in scenarios with clear demand and significant benefits, such as declassification review and intelligent reference inquiry, to quickly validate effectiveness and accumulate experience, thereby controlling initial investment risks. Subsequently, AI technology can be gradually extended to deeper applications like intelligent compilation & research and knowledge mining, ensuring that each stage of investment yields corresponding returns and achieving sustainable, high-quality development of archival AI applications.

6.4. Strengthen the Development of an Interdisciplinary Talent Team

Generative Artificial Intelligence does not replace archival professionals but reshapes their roles and skill sets. Future archival talent will require composite capabilities integrating "archival expertise + AI technology." On one hand, systematic in-service training programs to enhance AI literacy should be implemented. Through specialized training, hands-on exercises, and inter-institutional exchanges, staff's understanding and application skills of AI tools can be strengthened, transforming them from "operators" to "collaborative managers." On the other hand, archives should be encouraged to establish practical training bases and joint laboratories with universities and enterprises, conducting innovative research on AI applications in archives to form a closed loop of "industry-academia-research-application."

7. SUMMARY

Generative Artificial Intelligence (GenAI) is reshaping the concepts and pathways of archival information resource development, rendering archival work more intelligent and service-oriented. Multiple practical cases demonstrate that the application of GenAI technology in the archival field is gradually forming replicable and scalable models, providing a practical pathway for building smart

archives. The introduction of GenAI enables archivists to comprehend the value of archival data from a broader perspective, effectively shifting the focus from "archiving" to "utilizing" archives.

Concurrently, GenAI introduces new problems and challenges. Data security, ethical norms, copyright ownership, and the updating of professional competencies are all unavoidable issues in the process of archival intellectualization. Archival institutions must maintain a balance between innovation and regulation, seizing technological opportunities while guarding against the risks of over-reliance. Looking ahead, the development of archival information resources should be grounded in high-quality data and guided by a balanced approach that prioritizes both technology and institutional frameworks. This will propel the construction of smart archives towards a future characterized by safety, standardization, and sustainable development, ultimately enabling archives to play a greater role in serving social governance and cultural heritage.

REFERENCES

[1] National Archives Administration of China. The "14th Five-Year" Plan for National Archival Development [EB/OL]. (2021-06-08)[2025-10-15]. https://www.saac.gov.cn/daj/yaow/202106/899650c1b1ec4c0e9ad3c2ca7310eca4.shtml.

[2] National Archives Administration of China. Upholding Integrity and Innovation, Working Pragmatically to Comprehensively Realize the "14th Five-Year" Archival Development Goals and Tasks [EB/OL]. (2025-01-23)[2025-10-15]. https://www.saac.gov.cn/daj/yaow/202501/9cb357d9822c44d69d9fb6c4c9926f25.shtml.

[3] Jianming Zhang, Lijuan Gao. Research on the Application of AI Deep Learning-based Super-resolution Reconstruction Technology in Image Archive Restoration [J]. China Archives, 2022, (11): 36-37.

[4] Le Gao. Deployment, Application, and Security Protection Practice of DeepSeek in Provincial Archives: A Case Study of Zhejiang Provincial Archives [J]. Zhejiang Archives, 2025, (03): 5-7.

[5] Zhejiang Provincial Archives. Continuously Iterating and Upgrading Digital Archives to Actively Create "Three Directions" [EB/OL]. (2025-04-03)[2025-10-15]. https://www.saac.gov.cn/daj/szdagjsdxal/202504/c6e7557493cb4b2cb634b5017056d5fb.shtml.

[6] Renfang Luo. Practice of Iterative Upgrading of Zhuhai Digital Archives [J]. China Archives, 2022, (03): 46-47.

[7] Aihua Li, Ying Wei, Jun Liu. Practice and Reflection on Localized Deployment of DeepSeek Large Model in Changping District Archives [J]. Beijing Archives, 2025, (04): 61-64.

[8] Yuenan Liu, Yi Qian, Ping Wang, et al. Challenges and Prospects: The Impact and Application Prospects of DeepSeek on Archival Work [J]. Zhejiang Archives, 2025, (02): 5-13.

[9] Jingjing Yang. Research on the Application of Generative Artificial Intelligence in Digital Archival Scenarios [J]. Zhejiang Archives, 2024, (01): 45-47+54.

[10] Min Liu, Yuanying Li. Activation of Archival Resources Based on Film and Television Creation [J]. Shanxi Archives, 2025, (08): 151-153+157.

[11] Shang Wu. Security Issues of DeepSeek Application in Archival Development [J]. Shanxi Archives, 2025, (08): 93-95.

[12] Manru Han. Research on Cost and Efficiency of Archival AI Application [J]. Shanxi Archives, 2025, (05): 137-140.

[13] Wei Miao. Ethical Issues of DeepSeek Application in the Archival Field [J]. Shanxi Archives, 2025, (08): 89-92.