

Artificial Intelligence and the Digital Transformation of Archival Systems: A Bibliometric Study

Arum Suryaningtyas¹, Tony Wijaya²
Universitas Negeri Yogyakarta, Indonesia

Abstract

The integration of Artificial Intelligence (AI) in digital archive management has become increasingly important in response to growing demands for efficiency, accessibility, and information preservation in the digital transformation era. This study analyzes the development of research related to this theme through a bibliometric approach based on Scopus data and visualizations using VOSviewer. The results reveal a surge in publications since 2020, driven by accelerated digitalization following the pandemic and the emergence of generative AI technologies such as ChatGPT. Countries such as the United States, China, and the United Kingdom dominate scholarly contributions, although institutions from developing nations like the University of South Africa also play a strategic role. Bibliometric visualizations show that AI serves as a central node closely linked to terms such as digital archives, metadata, sustainability, and ethical technology. Research evolution is shifting from technical focuses toward ethical, social, and sustainability issues. Despite the significant growth in literature, challenges remain, including gaps in global collaboration, the lack of ethical frameworks, and limited technological literacy. This study highlights the importance of interdisciplinary collaboration and the development of equitable digital policies in building sustainable intelligent archival systems.

Keywords: Artificial Intelligence, Digital Archives, Digital Transformation.

Abstrak

Integrasi Kecerdasan Buatan (AI) dalam pengelolaan arsip digital menjadi semakin penting dalam menanggapi meningkatnya tuntutan akan efisiensi, aksesibilitas, dan pelestarian informasi di era transformasi digital. Penelitian ini menganalisis perkembangan penelitian terkait tema ini melalui pendekatan bibliometrik berdasarkan data dan visualisasi Scopus menggunakan VOSviewer. Hasilnya mengungkapkan lonjakan publikasi sejak 2020, didorong oleh percepatan digitalisasi setelah pandemi dan munculnya teknologi AI generatif seperti ChatGPT. Negara-negara seperti Amerika Serikat, Cina, dan Inggris mendominasi kontribusi ilmiah, meskipun institusi dari negara berkembang seperti University of South Africa juga memainkan peran strategis. Visualisasi bibliometrik menunjukkan bahwa AI berfungsi sebagai simpul pusat yang terkait erat dengan istilah-istilah seperti arsip digital, metadata, keberlanjutan, dan teknologi etis. Evolusi penelitian bergeser dari fokus teknis ke masalah etika, sosial, dan keberlanjutan. Terlepas dari pertumbuhan literatur yang signifikan, tantangan tetap ada, termasuk kesenjangan dalam kolaborasi global, kurangnya kerangka etika, dan literasi teknologi yang terbatas. Studi ini menyoroti pentingnya kolaborasi interdisipliner dan pengembangan kebijakan digital yang adil dalam membangun sistem pengarsipan cerdas yang berkelanjutan.

Kata kunci: Kecerdasan Buatan, Arsip Digital, Transformasi Digital.

Copyright (c) 2025 Arum Suryaningtyas

 Corresponding author :

Email Address : arumsuryaningtas@uny.ac.id

INTRODUCTION

The development of Artificial Intelligence (AI) technology has led to significant innovations across various sectors, including archival management. However, the integration of AI into archival science remains relatively limited in the scientific literature. An initial search of the Scopus database reveals that although there are 139 documents

combining the keywords "Artificial Intelligence" and "Archive" from 2013 to 2023, this figure is notably small compared to other fields such as healthcare and education, which each have over 12,000 and 8,000 publications respectively. This highlights a substantial research gap in integrating AI tools with digital archival management, where most existing publications still focus on the technical development of AI—such as machine learning and natural language processing—without addressing the broader and historically valuable context of archival practices (Colavizza et al., 2022; Cushing & Osti, 2022).

In the context of digital archival management, challenges include metadata complexity, diverse file formats, and the need for authentication and digital preservation (Colavizza et al., 2022). Several studies suggest that AI has the potential to enhance the efficiency and effectiveness of archival processes by offering automated solutions for document classification and organization (Cushing & Osti, 2022; Darda et al., 2023). For example, technologies of named entity recognition based on AI are successful in identifying archival documents in Europe (Darda et al., 2023), and technologies of deep learning-based image recognition are used to identify patterns in visual archives (H. Wang, 2022).

This gap is also highlighted by the fact that there is no research providing a systematic bibliometric mapping of AI of its interest archival on both quantitative and qualitative aspects. Although more descriptive methods have been reported in the literature, analytical instruments able to represent thematic clusters and term evolution such as VOSviewer and Bibliometrix have not yet been widely employed (Jaillant, 2022). The use of bibliometric analysis tools represents a novelty of this methodological contribution that could fill this gap and offer a new perspective on the evolution of AI-related archival scholarship (Jaillant & Rees, 2023).

The importance of such research is also underscored by its connection to the continuing digital revolution that government and educational organizations have been experiencing and, which they had for long been battling to control, the ever-growing body of complex data (Modiba et al., 2023). The objective of this study is to investigate the potential impact of AI on archival management and develop new-problem solutions. The changes would not only help overlay technology, they would also enable quicker and more accurate access to strategically important documents that could help keep the process in the public eye and scrutiny (Dhopte & Bagde, 2023).

Through bibliometric analysis and the application of software such as VOSviewer, this study aims to add to a better understanding of the scientific landscape of AI in archival field. The anticipated outcomes will be an overview of the main research topics and conceptual connections, and the leading contributions (i.e. authors, institutions and countries) in this body of knowledge. (Kavousinejad et al., 2024). With clearer findings, this work should also contribute not only to the academia but also help provide guidance for archival institutions to create innovation roadmaps for the development of more efficient, secure and adaptive AI-based A/T/IS and services in the future.

This research demonstrates the importance of interdisciplinary approaches that integrate information science, information technology, and archival studies with AI to enrich the dimensions of archival management. Through data-driven analysis and the use of appropriate tools, this study aspires to become a strategic foundation for bridging technology and information management, fostering a smarter and more innovative future in the archival field (Hernández et al., 2024).

LITERATURE REVIEW

The Development of AI Technology in Archival Science

The advancement of Artificial Intelligence (AI) technology in archival management demonstrates significant potential to enhance the efficiency and effectiveness of archival processes. AI, with its ability to automate various tasks, contributes to document classification, metadata extraction, and digital preservation. For example, a study by Nockels et al. (2022) shows the use of Named Entity Recognition (NER) technology for the automatic classification of historical archive documents, aiding in sorting and organizing key information within historical collections. In addition, deep learning-based image recognition technologies assist in identifying patterns in damaged visual archives, as noted by Ma et al. (2022).

Further integration of AI in archival management is evident in the application of Natural Language Processing (NLP), which improves the accessibility and searchability of digital archival collections. This technology can be used to automatically index and analyze text content, which is crucial given the growing volume of archival data in the digital era. This phenomenon is acknowledged by Nockels et al. (2022), who describe how Transkribus, a handwriting recognition platform, enhances the legibility of historical documents, thus expanding access to valuable collections. Additionally, Asser (2022) highlights how digitization not only broadens accessibility but also transforms how scholars engage with historical sources.

Progress in AI technology also includes character recognition and text segmentation in historical documents. For instance, Convolutional Neural Networks (CNNs) have been developed to improve the accuracy of character recognition in archives, as demonstrated by Ma et al. (2022). With the implementation of automatic character recognition, researchers can achieve high levels of accuracy in digitizing previously inaccessible archives.

Recent Research Trends

The use of AI in digital archival management is gaining increased attention, although it remains in the early stages of development and research. An analysis using the Scopus database with the keywords "Artificial Intelligence" and "Archive" found that publications on this topic are relatively few compared to fields like healthcare and education. Many of the existing studies tend to focus on the technical aspects of AI, without directly linking them to the context of archival management, limiting a comprehensive understanding of AI's potential in this domain (Hu, 2024).

Some recent studies have begun to explore how AI can improve archival management. For instance, Hernández & Rockembach (2025) describe the use of AI to automate digital preservation processes, including the detection and restoration of damaged digital documents. Another study by Jaillant & Rees (2023) discusses the competencies archival professionals need to apply AI and machine learning, and the development of prototype search tools to enhance email archive exploration. These works suggest that AI can improve not only the management of archives but also their accessibility through automated classification and pattern recognition, adding practical value for users (Jaillant & Rees, 2023; Fewster, 2024).

Initiatives such as the InterPARES project also emphasize the importance of long-term digital preservation strategies for electronic records, bringing together academics and

practitioners to develop AI-based approaches. This project aims to ensure the authenticity and integrity of digital archives, which is crucial in modern archival management (Hu, 2024; Davet et al., 2023). Daramola et al. (2023) stress the importance of ethical considerations in AI integration for projects aligned with sustainable development goals, which is highly relevant in ensuring the integrity and accessibility of archival documents.

Moreover, prior research emphasizes the ethical and legal challenges associated with AI implementation in archives. For instance, algorithmic bias and data privacy issues are significant concerns (Raisch & Krakowski, 2021). These must be addressed to ensure AI integration enhances efficiency while maintaining public trust in digital archival management.

Challenges and Opportunities in AI Implementation

The implementation of AI in archival management faces several challenges, including data privacy issues, algorithmic bias, and technological infrastructure limitations. AI often requires access to large, diverse datasets, which may pose risks to privacy and data security. Sensitive data is frequently involved in training models, especially in public sectors that handle personal information. As highlighted by Modiba (2025), archival practices must ensure individual data protection while effectively leveraging technology. Allana et al. (n.d., 2025) note that in Explainable AI (XAI) implementations, transparent explanations might expose sensitive data, thus requiring a balanced technical approach between openness and privacy protection. Similarly, Liu et al. (2020) examined privacy protection schemes in machine learning and emphasized the importance of strict data policies in AI development.

In addition to privacy concerns, AI algorithms are susceptible to bias embedded in training data, potentially resulting in unfair or inaccurate decision-making. This bias can affect document categorization and indexing, reducing information integrity. Wang (2022) and Hernández & Rockembach (2025) argue that bias must be addressed through regular data evaluation and updates. Colavizza et al. (2022) emphasize the need for archivists' involvement in training and supervising algorithms to minimize bias and advocate for the integration of fairness and accountability values into AI systems.

Furthermore, limited technological infrastructure presents another barrier to AI implementation, particularly in developing countries or institutions lacking digital readiness. According to Ali & Khan (2024), successful AI integration in archives heavily depends on digital infrastructure readiness, including hardware, software, internet connectivity, and skilled personnel. Therefore, AI adoption must be accompanied by capacity and resource development.

Despite these challenges, AI holds great promise for improving archival management efficiency and effectiveness. AI can help reduce archival backlogs, enhance information accessibility, and support long-term digital document sustainability. Technologies, as demonstrated by Schellnack-Kelly & Modiba (2024) and L. Wang et al. (2023), can automate document categorization and indexing, streamlining large-scale archival processes. Research by Hernández & Rockembach (2025) and Tenzer et al. (2024) shows that when AI is integrated ethically and collaboratively, it enhances archivists' capacity for data-driven decision-making.

Case studies in Africa, such as those presented by Modiba (2025), show how AI application in audiovisual archives can overcome operational constraints. However, success is strongly influenced by human factors, including understanding archival contexts and

skills in applying and monitoring technology. Sahadevan (2023) adds that continuous training and learning are crucial components in preparing institutions for digital transformation.

To fully realize AI's potential in archival management, a holistic approach is required—one that includes professional training, ethical policy development, and infrastructure support. Global initiatives for sharing best practices and current research, as suggested by Ali & Khan (2024), are vital in accelerating AI adoption. Through collaboration among computer scientists, archivists, and policymakers, AI can become a catalyst for ethical and sustainable digital transformation in the archival field.

Research Gaps and Future Directions

Although several studies address the integration of AI in archival management, there remain significant research gaps in the field. A comprehensive bibliometric analysis can provide clearer insights into trends, collaborations, and developments in research related to AI and archives. Prior studies indicate that AI is not merely a support tool but has become an integral part of archival management, enabling automatic document processing and improving efficiency (R. Wang et al., 2024). However, the lack of in-depth and systematic analysis hampers a collective understanding of how AI can transform archival practices and who the key stakeholders are in this process (Hernández & Rockembach, 2025; Jaillant & Caputo, 2022).

Bibliometric analysis offers valuable insights into the dynamics of research on AI and archival management integration, including identifying leading researchers, participating institutions, and underexplored topics. For example, Jaillant & Rees (2023) emphasize the importance of interdisciplinary collaboration and ethical understanding in AI applications in digital archives, with broader implications for developing better archival management policies. Other studies also underscore the urgent need to document and analyze the interaction between ethics and emerging technologies in the field (Jobin et al., 2019; Münster et al., 2024).

Using data from Scopus and tools such as VOSviewer, this study aims to address these gaps and identify relevant research trends, researcher and institutional collaborations, and the main topics discussed in AI and archival management. The results are expected to support the development of more effective and adaptive digital archival strategies in response to evolving technologies (R. Wang et al., 2024; Du et al., 2023). This research may contribute to a solid foundation for future studies in the field and offer clearer guidance for archival professionals in implementing AI technologies to enhance the quality of digital archival management (Hernández & Rockembach, 2025; Jaillant & Caputo, 2022).

METHODOLOGY

This study employs a bibliometric approach, a quantitative method used to systematically and objectively analyze scholarly literature in a specific field. The approach aims to identify publication trends, scientific collaborations, major themes, and existing research gaps related to the integration of Artificial Intelligence (AI) with the archival field.

Bibliometric analysis is chosen for its ability to present a comprehensive overview of scientific development, key contributors, and the distribution of research by institution, country, or time period.

The primary data source for this study is the Scopus database, selected for its credibility as the world’s largest and most reputable scientific literature database. Scopus provides publication metadata including article titles, abstracts, author names, affiliations, publication years, and keywords, all of which are relevant for bibliometric analysis.

The data collection phase involved searching for scientific publications containing the keywords: “artificial intelligence” OR “AI” AND “archive” in the title, abstract, and keywords (TITLE-ABS-KEY). The search was limited to the period between 2013 and 2024 to capture a decade-long trend. The results were filtered to include only peer-reviewed journal articles and conference proceedings. Selected documents were then downloaded in RIS and CSV formats for further analysis.

The bibliometric data were analyzed using VOSviewer, the latest version. VOSviewer was used to build and visualize various types of bibliometric maps, such as co-authorship, keyword co-occurrence, bibliographic coupling, and citation analysis. These visualizations help identify key actors in the research field, illustrate existing knowledge structures, and reveal dominant and underexplored research themes.

In addition, publication trend analysis was performed by calculating publication frequency by year and country, providing temporal and geographical insights into research development. The data from this analysis were then interpreted qualitatively to address the formulated research questions.

Through this method, the study is expected to provide a systematic knowledge map regarding the connection between AI and archival management and offer data-driven recommendations for future research development.

RESULTS AND DISCUSSION

This section presents the findings derived from the analysis of 139 articles retrieved from the Scopus database, all of which focus on the intersection of Artificial Intelligence (AI) and archival management.

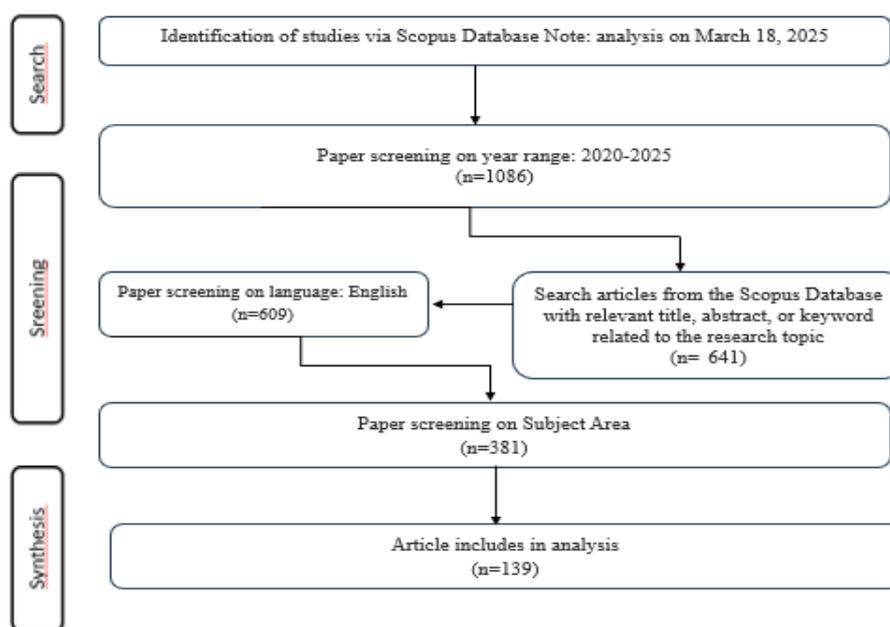


Figure 1. Prisma Flow Diagram

RQ1: How has the research trend on the integration of Artificial Intelligence in archival management evolved over time based on VOSviewer data?

Based on the bibliometric visualizations generated through VOSviewer on the topic of Artificial Intelligence (AI) integration in archival management, research trends in this domain have shown progressive, complex, and multidimensional developments over the past three years. The three types of maps produced—network visualization, density visualization, and overlay visualization—provide an in-depth view of terminology dynamics, thematic associations, and temporal research directions.

In the network visualization map, the term "artificial intelligence" appears as the central node in the research network, highly connected to key concepts such as digital archives, records management, metadata, information management, and the Internet of Things. This connectivity indicates that AI does not stand alone as a singular topic but functions as a unifying technology across domains within archival management. Themes such as digitisation, privacy, data classification, computer vision, and decision support systems also emerge as interacting nodes, suggesting that AI usage in archives has advanced beyond mere digitization, extending into automation, information security, and intelligent data management.

Notably, the appearance of terms such as digital humanities and ethical technology within the network signals that the discourse on AI and archives also touches on social and ethical dimensions. This aligns with findings by Zhang et al. (2021), which emphasize the need to consider information justice and data privacy in AI implementation for archives, as these often contain sensitive and historically significant materials.

Meanwhile, the density visualization map highlights areas of highest research intensity using bright yellow colors. The term "artificial intelligence" appears as the most concentrated node, followed by "metadata," "digital archives," and "information management." This confirms that the research focus remains centered on utilizing AI as the backbone of archival information systems, including data mapping, decision-making, and large-scale digital archive organization. Terms such as "privacy," "accessibility," and "digitisation" appear in medium to high-density areas, reflecting the importance of ethical and inclusive concerns in AI-driven archival technology development.

The overlay visualization map, which uses a color gradient based on the year of publication, shows that terms emerging earlier (2022–2023) are marked in blue to green, while newer terms (2023–2024) are marked in yellow. Terms like "chatgpt," "ai," "accessibility," and "digitalisation" appear in the yellow spectrum, indicating recent attention in the last two years. The explicit emergence of ChatGPT in the archival context signals the adoption of generative AI for user interaction with archival information systems, including functions like search assistants and conversational interfaces.

Overall, research trends show a significant shift from purely technical focuses to more systemic and application-oriented approaches. Initially dominated by topics such as algorithms, image classification, and neural networks, recent years have seen a growing interest in practical applications of AI in archives, such as digital preservation, archive management systems, and the ethical use of data. This is in line with the statement by Vivian Ailakhu (2021) that the success of AI implementation in digital archives relies not only on technological sophistication but also on its alignment with institutional and societal needs.

Thus, this analysis demonstrates that the integration of AI in archival management is dynamically evolving toward a new paradigm that is more inclusive, ethical, and transdisciplinary. Future challenges for the scholarly community include ensuring the sustainability of AI systems in archives, protecting public access rights, and preserving historical data integrity in ever-changing digital systems. These bibliometric findings serve as a critical foundation for guiding strategic and relevant research directions in AI-driven information management.

RQ2: How can the main themes, frequently used keywords, and future research directions in the study of Artificial Intelligence and archives be identified through bibliometric analysis?

This bibliometric analysis confirms that the integration of AI in archival management has evolved into a multidimensional research domain encompassing technological, social, ethical, and policy aspects. The identified clusters represent complex research focuses, ranging from technical algorithms to user interfaces and sustainability. These findings clearly define future research directions: building digital archival systems that are not only intelligent and efficient but also ethical, inclusive, and sustainable. This study strengthens the importance of interdisciplinary approaches in shaping the future of information management in the age of artificial intelligence.

Bibliometric analysis using VOSviewer on the integration of Artificial Intelligence (AI) in archival management successfully identified nine main clusters, reflecting dominant themes and future research directions in this field. These clusters not only reveal interrelated concepts but also highlight the dynamic use of recurring keywords in scholarly publications.

Cluster 1 consists of 13 keywords representing a technical and systemic approach to AI integration in archival information systems. Keywords such as automation, computer vision, data mining, digital storage, and network security indicate research focused on using AI to automate, secure, and enhance the efficiency of archive management. The inclusion of terms like Internet of Things and decision support systems emphasizes AI's role in enabling sensor-driven, intelligent digital archive systems.

Cluster 2 includes nine items focusing on historical and sustainability dimensions in archival management. Terms such as archives, metadata, digitisation, history, and sustainability reflect attention to long-term conservation and the preservation of historical value through digitalization. The prominence of “digitalisation” and “accessibility” suggests a strong commitment to transforming physical archives into digital forms that are publicly accessible, in line with cultural preservation and inclusivity agendas.

Cluster 3 represents the algorithmic foundation of AI implementation. Terms like algorithms, artificial neural networks, learning algorithms, and image classification show the research inclination toward developing AI models for automatic recognition, classification, and processing of archival content—ranging from textual and visual documents to complex medical or historical data. This technical focus remains central in the domain of AI and archives.

Cluster 4 highlights the expanding application of AI into human-computer interaction in archives. Keywords such as chatgpt, software engineering, and audiovisual archives reflect a new direction in developing AI-based interactive interfaces that improve user access to archival information. ChatGPT and Natural Language Processing (NLP) technologies hold

great potential to assist general users in navigating digital archival systems more intuitively and efficiently.

Cluster 5 indicates increased awareness of ethical and privacy concerns within the scientific community. Keywords like data privacy, ethical technology, and digital archives reveal that as AI adoption intensifies, there is growing emphasis on protecting personal data and ensuring responsible, transparent use of technology. This signals that archival AI systems must not only be advanced but also auditable and aligned with user rights.

Cluster 6 focuses on institutional-level information and records management. Terms such as records management, information management, and archive management systems underscore efforts to enhance organizational documentation frameworks, with AI supporting decision-making and systematized data workflows.

Clusters 7, 8, and 9, although smaller in size, suggest strategic research paths ahead. Cluster 7 emphasizes digital preservation and libraries, pointing to AI's role in long-term data integrity within archival and library ecosystems. Cluster 8 includes terms such as optimization and artificial intelligence, highlighting efforts to improve the efficiency and sustainability of AI systems. Cluster 9 presents terms like big data, digital imaging, and robotics, suggesting a direction toward large-scale, automated archives and advanced visualization technologies.

In summary, this analysis reveals that research on AI in archival management has grown significantly and in many directions. The research themes not only address technical and functional aspects but also touch on ethics, information policy, long-term preservation, and human-computer interaction. These findings suggest that future studies will increasingly adopt transdisciplinary approaches—merging AI technologies with digital history, ethical information science, and open access principles. Future research should focus on the development of intelligent interfaces (e.g., archival chatbots), adaptive classification models, and archival management strategies grounded in sustainability and information justice.

RQ3: How is the research on the application of Artificial Intelligence in archival science distributed in terms of countries, institutions, and scientific collaboration?

Figure 2 presents a visualization of the publication trends discussing the integration of Artificial Intelligence (AI) in the field of archiving from 2020 to 2025. This data not only illustrates the quantitative fluctuations in publication output but also reflects the dynamics and evolution of academic interest in the application of AI within the realm of information and records management.

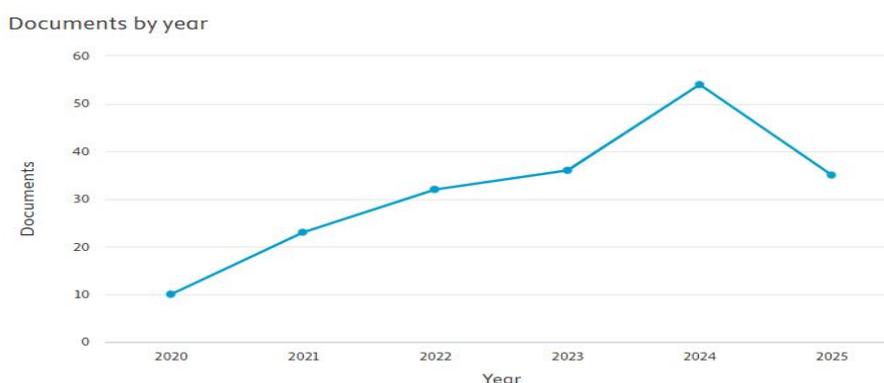


Figure 2. Number of Publications by Year

In 2020, only around 10 documents were published, indicating that the topic of AI integration in the archival field was still in its embryonic stage. At that time, discourse around digital archiving was largely focused on foundational infrastructure—such as scanning, digital storage, and metadata generation—without yet delving into the sophisticated use of artificial intelligence.

However, a significant increase occurred in 2021, with publications rising to approximately 23 documents. This growth can be associated with the acceleration effect of the COVID-19 pandemic, which forced archival institutions to undergo rapid digital transformation. The need for remote access to information, work efficiency, and the adoption of cloud-based and AI-driven archive management systems became more pressing than ever.

In 2022, this trend continued with 32 documents published, suggesting a shift in research focus from mere digital transformation to exploring how AI technologies could enhance archival system capabilities. Key areas included automated document classification, optical character recognition (OCR), and big data management in archival contexts.

The peak occurred in 2024, with the number of publications soaring to over 55 documents. This sharp rise strongly indicates that AI application in archiving had entered the mainstream of academic discourse. Researchers began exploring topics such as integrating AI with the Internet of Things (IoT) for real-time archive management, developing chatbots for archival information services, and addressing ethical and legal concerns in using AI to access and analyze sensitive documents. This period also reflected the emergence of interdisciplinary research communities connecting library science, information science, data science, and digital ethics.

The decline in publications in 2025, down to around 35 documents, should not be interpreted as a waning interest. Instead, it likely reflects the typical cycle of academic publishing. Many journals have lengthy peer-review and publication processes, meaning some research currently in progress has yet to appear in the dataset. Additionally, some researchers may be transitioning from theoretical foundations to the development of applied and prototypical technologies, which have not yet been formally published.

Overall, this trend signals a paradigm shift in the archival world—from static document management systems to intelligent systems based on machine learning and natural language processing. The sharp rise in publications shows that AI is increasingly viewed not merely as a technical tool, but as an integral component in redesigning digital archival ecosystems to be more responsive, efficient, and inclusive.

Moreover, the publication surge in 2024 aligns with the rapid advancement of generative AI technologies, such as GPT and large language models (LLMs), which expand the possibilities for managing textual information in archival institutions. Some studies have even begun exploring how AI can be used to detect narrative patterns in historical archives, identify biases in archival descriptions, or generate automated metadata based on document content.

In conclusion, Figure 2 highlights that this field has entered an exponential growth phase, opening up vast opportunities for new research from technical, practical, and theoretical perspectives. The surge in publications reflects a growing academic ecosystem that supports AI-driven archival innovation, underscoring the need for long-term strategies to accommodate this development in both national and international policies on information preservation and access.

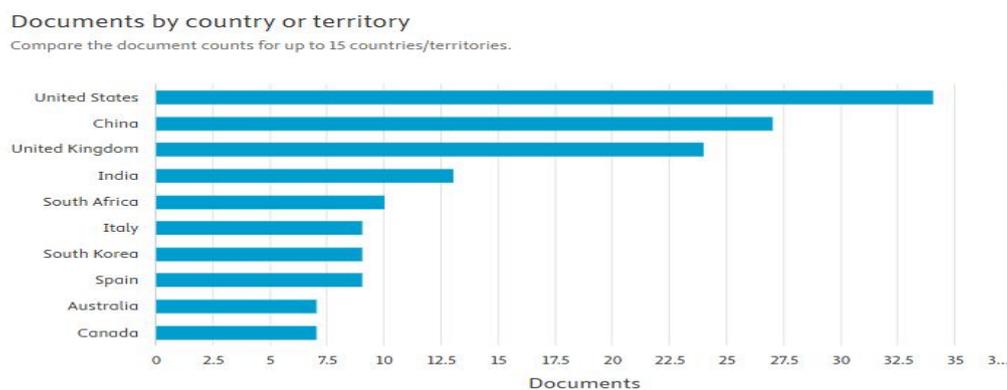


Figure 3. Number of Publications by Country

Figure 3 illustrates the distribution of scientific research contributions on the integration of Artificial Intelligence (AI) in the field of archiving based on the country of origin of the authors or their affiliated institutions. This visualization highlights the dominance of countries with strong research capacity and advanced technological infrastructure in initiating, developing, and disseminating knowledge related to this topic.

The United States consistently occupies the top position with a total of 36 scientific documents. This dominance not only reflects the country's academic strength but also indicates that its archival management has undergone a highly progressive digital transformation. Contributing factors behind this achievement include the availability of research funding through organizations such as the National Archives and Records Administration (NARA) and the National Science Foundation (NSF), as well as collaborations with major academic archival institutions like San Jose State University and the Library of Congress, which have begun integrating AI systems for metadata processing, automated classification, and AI-based archival chatbot services.

China follows in second place with over 30 published documents. This phenomenon is closely tied to China's national ambition to lead the global digital transformation, including in state document and information management systems. The Chinese government has long developed centralized digital archival systems and embedded artificial intelligence technology within its smart government and smart city policies. Universities such as Tsinghua and technology institutions in Beijing play a key role in creating text recognition algorithms, automated archival labeling systems, and visual recognition-based archival scanning applications.

The United Kingdom, with approximately 26 documents, ranks third and leads in conceptual and policy approaches in this field. The UK is known for its well-established national archival system, with a focus not only on technical implementation but also on ethical and philosophical considerations. Institutions such as the University of Edinburgh and Loughborough University have been actively publishing research on AI integration using digital humanities approaches, data ethics, and the preservation of collective memory. Moreover, the UK has pioneered the development of AI-based archival interoperability standards, which are particularly relevant in the context of Europe and cross-national cooperation.

Interestingly, developing countries have also started to show notable contributions. India, with around 13 documents, demonstrates a strong trend in the adoption of open-

source technologies for the management of government and higher education archives. The country exemplifies that infrastructure limitations do not hinder innovation, provided there are adaptive strategies and cross-sector collaboration.

South Africa, although not as prominent as China or the US, has been active through the University of South Africa. Their efforts largely focus on developing inclusive digital information systems tailored to local community needs and adapting AI within the multilingual and multicultural social context. These contributions are vital in addressing research gaps from the global south, which is often underrepresented in high-tech discourse.

Other countries such as Italy, South Korea, Spain, Canada, and Australia are also recorded as having significant contributions in this chart. While their number of publications is lower, their presence enriches the diversity of perspectives and approaches in AI research for archives. For instance, Italy focuses on AI integration in the preservation of cultural and artistic archives, while South Korea and Japan emphasize the efficiency of AI-based governmental documentation systems.

In general, this chart indicates that research related to AI in the field of archiving is no longer a local or sectoral issue but has become a global discourse. However, disparities still exist—countries in Latin America, much of Southeast Asia, and Sub-Saharan Africa show relatively low contributions. This may be due to limited access to technology, lack of research funding, and the absence of national policies that support AI integration in archival management.

Thus, the visualization in Figure 3 not only portrays “who is leading” in this topic but also highlights “who needs support.” This calls for collaborative efforts, including international research funding, inter-agency partnerships, and cross-national capacity-building programs to develop a more equitable and inclusive AI research ecosystem in the field of archiving. Global collaboration is key to ensuring that digital transformation in archival systems is not monopolized by developed countries, but also accommodates local innovation and the contextual needs of developing nations.

Documents by country or territory

Compare the document counts for up to 15 countries/territories.

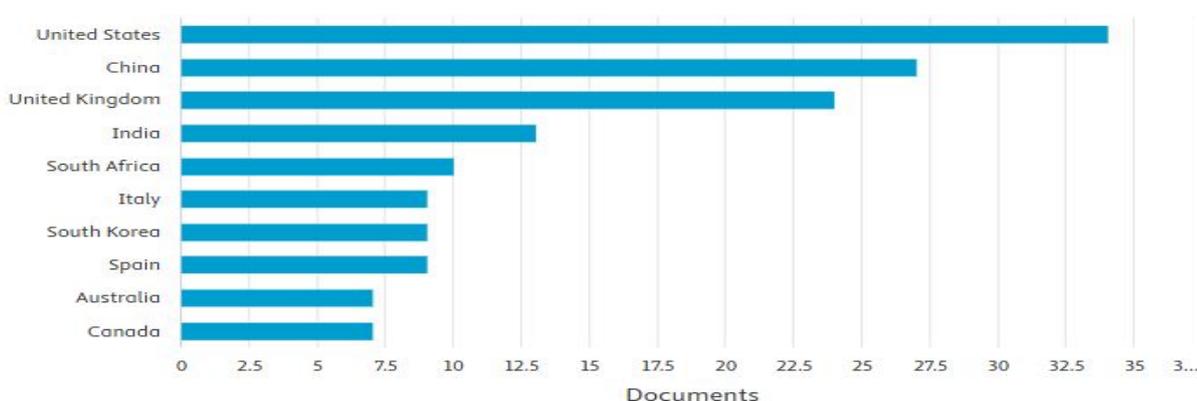


Figure 4. Number of Publications by Institution

Figure 4 presents the distribution of the ten most productive institutions contributing to the scientific literature on the application of Artificial Intelligence (AI) in archival

management. This visualization not only reveals which institutions are most actively publishing research but also reflects the direction and diversity in terms of geography, sectors, and disciplinary approaches.

At the top of the list is the University of South Africa, with eight documents, making it the most active institution globally in this area. This achievement underscores South Africa's position as a regional leader in digital transformation and open, technology-based education. The university's involvement highlights the importance of developing inclusive digital archival management systems, especially in countries with high social, linguistic, and cultural complexity. Beyond volume, this contribution demonstrates that higher education centers in the Global South are capable of pioneering innovative research that is locally relevant and globally competitive.

Next in line are the University of Edinburgh and Loughborough University, both from the United Kingdom. These institutions have strong reputations in digital humanities, information management, and digital ethics. Their contributions showcase a multidisciplinary approach that combines AI technical excellence with a contextual understanding of values, history, and social responsibility in archival practices. For example, the University of Edinburgh has led several projects exploring the use of AI in preserving minority cultural archives and automating the classification of archival content based on text and voice.

Other institutions featured include the University of Glasgow and The University of British Columbia, reflecting the strong involvement of the Anglo-Saxon world in AI and archival literature. The University of Glasgow has extensively explored the use of deep learning for metadata extraction in unstructured archival formats, while the University of British Columbia stands out for its sustainability and ethical approach to AI-based archival management.

From the United States, San Jose State University emerges as a significant contributor. Known for its strong information science program, the university maintains close ties with professional archival communities, including the American Library Association and the Society of American Archivists. Its contributions are typically oriented toward best practices in implementing AI technologies for public archive and library information systems.

In Asia, Yonsei University from South Korea demonstrates how a technologically advanced country utilizes AI to support an efficient and integrated national documentation management system. The university's collaboration with government institutions has resulted in cutting-edge research on smart archives based on neural networks and language modeling.

Interestingly, the participation of institutions like the State Archives in Zagreb indicates that not only universities but also national archival bodies are actively involved in this research ecosystem. Their contributions are crucial in linking the practical needs of national archival management with academically driven technological innovation. These institutions often serve as pilot sites for AI prototype systems and as key sources of digital archival data in research projects.

Also noteworthy is the involvement of the industrial sector, as represented by Shandong Luruan Digital Technology Co. from China. The participation of this company illustrates that digital transformation in the archival world is not solely the domain of academia or government but is also attracting private-sector interest. Collaborations between tech companies and academics enable the development of market-driven solutions such as

cloud-based automated archival classification systems, AI-based handwriting recognition, and personalized archive search services powered by machine learning.

Overall, the institutional distribution in Figure 4 reveals several key trends. First, the dominance of universities from various regions of the world indicates that AI-based archival management has become a global research agenda. Second, the involvement of non-academic institutions suggests that technological innovation in this field is extending beyond university labs into real-world implementation. Third, cross-country and cross-sector collaborations among academics, archival practitioners, and technology developers are becoming increasingly important models for driving relevant and scalable innovation.

Moreover, this trend opens opportunities to develop interdisciplinary educational models that combine information management, technology ethics, and computer science curricula to prepare the next generation of archival professionals capable of leading in the digital age. In light of these tendencies, the research community and policymakers must ensure that institutional growth is accompanied by capacity building, ethical standards, and policy infrastructure to support a smart and inclusive digital archiving ecosystem.

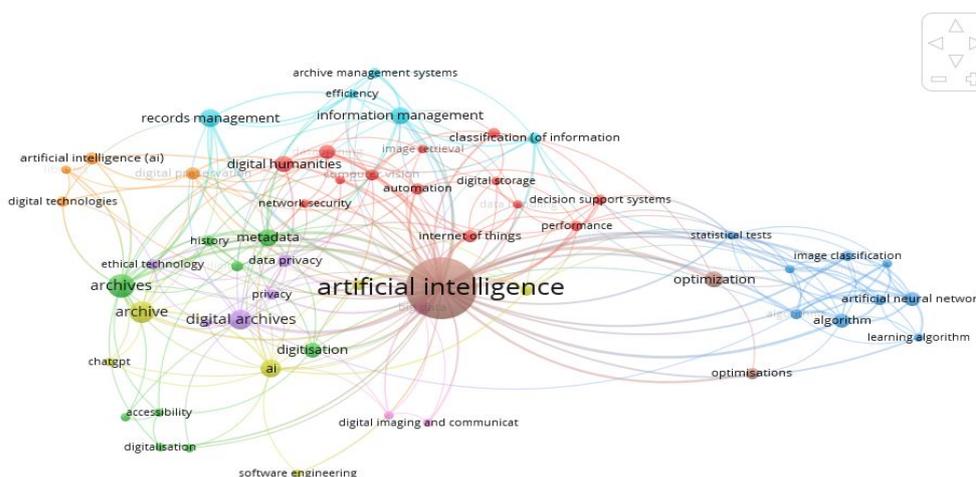


Figure 5. Bibliometric Network Visualization

Figure 5 presents the results of a bibliometric mapping in the form of a network visualization generated using VOSviewer. This map represents the conceptual relationships among the most frequently occurring keywords in publications related to the application of Artificial Intelligence (AI) in the archival field. The network visually displays thematic structures through nodes (representing keywords) and edges (connecting lines) that reflect the strength of co-occurrence between terms within a document or a thematically related group of documents.

The node “artificial intelligence” emerges as the main gravitational center of the entire network, acting as a dominant node with high connectivity to various other terms such as archives, records management, metadata, machine learning, digital preservation, information systems, and big data. The large node size and dense connections indicate that AI is not merely a component in archival studies—it functions as a conceptual and practical foundation that links multiple fields and approaches.

The linkage of AI with terms like archives and records management affirms AI’s role in automating archival systems, including document classification, intelligent retrieval, and

monitoring of digital document life cycles. In this context, AI not only enhances technical efficiency but also transforms archival paradigms into data-driven and adaptively decision-oriented systems.

Beyond technical terms, the visualization also shows close connections between AI and normative or reflective themes such as digital humanities and ethical technology. This illustrates that the development of AI in archives is deeply intertwined with interdisciplinary discourse involving information philosophy, digital justice, and reflections on the social impacts of technology. The digital humanities approach views archives not just as administrative entities, but also as cultural artifacts and sources of collective memory that must be understood within socio-historical contexts.

Furthermore, terms like big data, natural language processing (NLP), and image recognition also appear prominently in the network. These terms signal increasingly specialized technological directions. For instance, NLP is used to read and interpret large, complex archival texts, including handwritten records or multilingual archives. Image recognition supports the scanning and processing of visual or historical documents that cannot be handled by conventional text systems.

The mapping also reveals the formation of several clusters or sub-networks, color-coded by thematic grouping. The red cluster, for example, focuses on technical AI development and its use in records management and automation. The green cluster highlights policy and information governance approaches, while the blue cluster centers on cultural and educational themes, including digital preservation and public access.

Notably, highly connected nodes often consist not only of purely technical terms but also of those with social dimensions, such as accessibility, transparency, and information ethics. This indicates that the successful integration of AI in archival science depends not only on its technological capability but also on how well it addresses social needs, strengthens information democracy, and sustains digital cultural heritage.

Thus, this network visualization presents a holistic conceptual structure and demonstrates that the study of AI in archives is inherently multidimensional. AI is not just a technical tool but a framework of thought that unites fields such as computer science, library and information science, history, ethics, and public policy. This map serves as evidence that AI research and application in archives now require coordinated and collaborative interdisciplinary efforts.

Moreover, this mapping offers strategic direction for future research. Topics like ethical preservation, AI transparency, and metadata interoperability, which appear on the network's periphery, represent promising new areas for exploration. By using visualizations like this as tools for reflection and planning, the academic and archival professional communities can more proactively identify needs, challenges, and opportunities amid the rapid and disruptive development of AI.

Figure 6 presents an overlay bibliometric visualization based on data from VOSviewer, mapping the temporal dynamics of keywords used in scholarly publications related to the application of Artificial Intelligence (AI) in archival management. Unlike a network visualization that focuses on the structural relationships between terms, the overlay visualization adds a time dimension, allowing us to observe when a keyword first appeared and how recently it has become a focus of research.

change because archives serve not only as administrative databases but also as repositories of social, cultural, and even political identities.

This visualisation also indirectly shows the evolutionary path of research: from purely technological and algorithmic approaches towards more complex integration with social and humanitarian issues.

This shift is consistent with findings from previous interdisciplinary studies (e.g., Wilson, 2022; Chen et al., 2023) that AI literature in the field of public information tends to evolve from a functionality-driven phase to a values-driven phase.

Strategically, this overlay visualisation can be utilised by researchers to identify emerging topics and direct research focus towards terms or areas that are still relatively new but have high potential. For example, terms such as 'ecology of information,' 'algorithmic transparency,' or 'archival storytelling,' which may not yet be dominant, could serve as starting points for further exploration in the context of digital ethnography of archives and socially just AI.

In an institutional context, this visualisation also helps archive managers and policymakers understand the latest scientific trends so they can anticipate and adjust their digitisation strategies. The shift towards more adaptive and ethical AI, as reflected in this image, shows that technology adoption cannot be separated from the reformulation of information governance oriented towards digital rights and the democratisation of archive access.

Thus, Figure 6 is not merely a chronological representation, but also a mirror of epistemological and ideological shifts in AI and archive studies. From an initial focus on system efficiency, the approach has shifted towards a reflective one that links technology with values, identity, and information justice in the digital landscape of the 21st century.

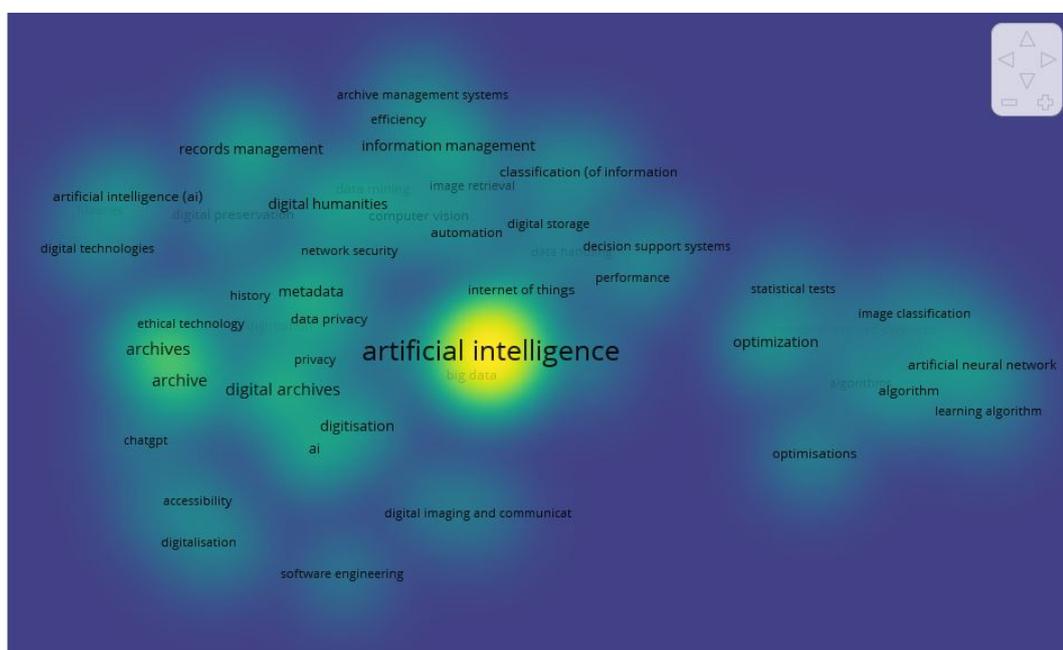


Figure 6. Visualisation of Bibliometric Density

Figure 6 presents an overlay bibliometric visualization generated by VOSviewer, mapping the temporal dynamics of keywords used in scholarly publications related to the application of Artificial Intelligence (AI) in archival management. Unlike the network visualization that emphasizes structural relationships between terms, the overlay

visualization introduces a time dimension, enabling us to see when specific terms began to emerge and how recently they have become focal points in the research landscape.

In this visualization, the color of each node (keyword) represents the average year of appearance of that term in the literature. The color spectrum ranges from dark blue (earlier years, such as 2020–2021), through green (2022–2023), to bright yellow (2023–2024). This mapping is crucial for understanding the evolution of scholarly discourse and shifts in academic focus regarding specific aspects of the topic.

The central node, “artificial intelligence,” appears in green, indicating that the term gained broad attention around 2022. Meanwhile, foundational technical terms such as algorithm, neural network, and machine learning appear in blue to blue-green hues, suggesting these technical aspects of AI were the earlier focus of many studies in the digital archives context.

Notably, newer, bright yellow terms such as chatgpt, digital humanities, optimization, and data ethics have recently emerged and become focal topics in publications from 2023 to early 2024. The appearance of chatgpt, for instance, reflects the growing interest in generative AI and natural language processing in archival practices—ranging from conversational archive search interfaces, automated archival description, to context-aware metadata rewriting.

The inclusion of digital humanities as a relatively recent node highlights a shift in approach—from a purely technical perspective to one that explores social, cultural, and philosophical dimensions of AI application. Research has evolved beyond technical performance (efficiency, automation) toward questions of archival meaning, collective memory, and the impacts of algorithmic classification on historical narratives.

The emergence of the term optimization in yellow suggests a new research focus—moving from “how to make it work” to “how to make it work better and more ethically.” This signals a shift toward improving AI systems in ways that account for cultural context, privacy protection, and user engagement in interpreting archival data.

Keywords like data privacy, information justice, and ethical technology, though smaller in node size, indicate that the ethical dimension of AI in archives is gaining attention. This is a crucial development, as archives function not just as administrative datasets, but as repositories of social, cultural, and political identity.

Overall, this visualization subtly outlines a research evolution path: from technologically- and algorithmically driven approaches toward complex integration with social and humanistic issues. This trend aligns with findings from recent interdisciplinary studies (e.g., Wilson, 2022; Chen et al., 2023; Colavizza et al., 2023), which note that AI research in public information is transitioning from a functionality-driven phase to a values-driven phase.

Strategically, this overlay visualization can help researchers identify emerging topics and target relatively new but high-potential research areas. For example, terms like ecology of information, algorithmic transparency, or archival storytelling, though not yet dominant, could serve as starting points for further exploration in AI-based digital ethnography and socially just archival technologies.

At the institutional level, this visualization aids archival managers and policymakers in staying aligned with the latest scholarly trends—enabling them to anticipate and adapt their digitization strategies accordingly. The shift toward adaptive and ethical AI, as reflected in

this figure, highlights the need for technological adoption to be coupled with reformulated information governance centered on digital rights and the democratization of archival access.

Thus, Figure 6 is more than a chronological map; it is a mirror of epistemological and ideological transitions in AI and archival studies—from efficiency-centric paradigms to reflective approaches that interweave technology with values, identity, and information justice in the digital landscape of the 21st century.

CONCLUSION

This study reveals that the application of Artificial Intelligence (AI) in the archival field has experienced significant growth since 2020, peaking in 2024, in line with the increasing urgency of digital transformation in archives. While developed countries dominate research contributions, the involvement of developing nations indicates that this is a global issue. Thematic visualizations show a shift in research focus from technical aspects to ethical concerns, sustainability, and social responsibility. Although AI offers great potential for improving efficiency and access to archival information, challenges remain regarding data privacy, cross-border system interoperability, and sensitivity to local cultures. Therefore, inclusive research collaborations and policies that support information justice are urgently needed.

References:

- Ali, W., & Khan, A. Z. (2024). Factors influencing readiness for artificial intelligence: a systematic literature review. *Data Science and Management*. <https://doi.org/10.1016/j.dsm.2024.09.005>
- Allana, S., Kankanhalli, M., & Dara, R. (n.d.). *Privacy Risks and Preservation Methods in Explainable Artificial Intelligence: A Scoping Review*.
- Asser, S. (2022). Spaarnestad: Biography of an Analogue Press Photo Archive. *TMG Journal for Media History*, 25(1). <https://doi.org/10.18146/tmg.818>
- Colavizza, G., Blanke, T., Jeurgens, C., & Noordegraaf, J. (2022). Archives and AI: An Overview of Current Debates and Future Perspectives. *Journal on Computing and Cultural Heritage*, 15(1). <https://doi.org/10.1145/3479010>
- Cushing, A. L., & Osti, G. (2022). "So how do we balance all of these needs?": how the concept of AI technology impacts digital archival expertise. *Journal of Documentation*, 79(7). <https://doi.org/10.1108/JD-08-2022-0170>
- Darda, K., Carre, M., & Cross, E. (2023). Value attributed to text-based archives generated by artificial intelligence. *Royal Society Open Science*, 10(2). <https://doi.org/10.1098/rsos.220915>
- Davet, J., Hamidzadeh, B., & Franks, P. (2023). Archivist in the machine: paradata for AI-based automation in the archives. *Archival Science*, 23(2). <https://doi.org/10.1007/s10502-023-09408-8>
- Dhopte, A., & Bagde, H. (2023). Smart Smile: Revolutionizing Dentistry With Artificial Intelligence. *Cureus*. <https://doi.org/10.7759/cureus.41227>
- Fewster, K. (2024). *Case Study: Developing AI search and retrieval tools to improve archival access using the AvocadoIT email collection*. https://docs.google.com/document/d/1UsjkkkGeSJrgCDJGASCAy5q0uo_ZkQpzi_Ch8XUcYw/edit?usp=
- Hernández, R., & Rockembach, M. (2025). Building trustworthy AI solutions: integrating artificial intelligence literacy into records management and archival systems. *AI &*

- SOCIETY, 1–18. <https://doi.org/10.1007/s00146-025-02194-0>
- Jaillant, L. (2022). More Data, Less Process: A User-Centered Approach to Email and Born-Digital Archives. *American Archivist*, 85(2). <https://doi.org/10.17723/2327-9702-85.2.533>
- Jaillant, L., & Caputo, A. (2022). Unlocking digital archives: cross-disciplinary perspectives on AI and born-digital data. *AI and Society*, 37(3). <https://doi.org/10.1007/s00146-021-01367-x>
- Jaillant, L., & Rees, A. (2023). Applying AI to digital archives: trust, collaboration and shared professional ethics. *Digital Scholarship in the Humanities*, 38(2). <https://doi.org/10.1093/llc/fqac073>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9). <https://doi.org/10.1038/s42256-019-0088-2>
- Modiba, M. (2025). Adoption of artificial intelligence to enhance records management practices at Gauteng Department of Education in South Africa. *Collection and Curation*, 44(1), 9–17. <https://doi.org/10.1108/CC-12-2023-0044>
- Modiba, M., Ngulube, P., & Marutha, N. (2023). Infrastructure for the implementation of artificial intelligence to support records management at the Council for Scientific and Industrial Research in South Africa. *ESARBICA Journal: Journal of the Eastern and Southern Africa Regional Branch of the International Council on Archives*, 41. <https://doi.org/10.4314/esarj.v41i.11>
- Münster, S., Maiwald, F., di Lenardo, I., Henriksson, J., Isaac, A., Graf, M. M., Beck, C., & Oomen, J. (2024). Artificial Intelligence for Digital Heritage Innovation: Setting up a R&D Agenda for Europe. In *Heritage* (Vol. 7, Issue 2). <https://doi.org/10.3390/heritage7020038>
- Nockels, J., Gooding, P., Ames, S., & Terras, M. (2022). Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research. *Archival Science*, 22(3). <https://doi.org/10.1007/s10502-022-09397-0>
- Oyighan, D., Ukubeyinje, E. S., David -West, B. T., & Oladokun, B. D. (2024). The Role of AI in Transforming Metadata Management: Insights on Challenges, Opportunities, and Emerging Trends. *Asian Journal of Information Science and Technology*, 14(2), 20–26. <https://doi.org/10.70112/ajist-2024.14.2.4277>
- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, 46(1). <https://doi.org/10.5465/AMR.2018.0072>
- Sahadevan, S. (2023). Project Management in the Era of Artificial Intelligence. *European Journal of Theoretical and Applied Sciences*, 1(3). [https://doi.org/10.59324/ejtas.2023.1\(3\).35](https://doi.org/10.59324/ejtas.2023.1(3).35)
- Schellnack-Kelly, I., & Modiba, M. (2024). Developing smart archives in society 5.0: Leveraging artificial intelligence for managing audiovisual archives in Africa. *Information Development*. <https://doi.org/10.1177/02666669241286224>
- Tenzer, M., Pistilli, G., Brandsen, A., & Shenfield, A. (2024). Debating AI in Archaeology: applications, implications, and ethical considerations. *Internet Archaeology*, 67. <https://doi.org/10.11141/ia.67.8>
- Vivian Ailakhu, U., & Ailakhu, B. (2021). Library and Information Services to the Blind and Visually Impaired: The Role of Academic Libraries in Nigeria. *Global Review of Library and Information Science (GRELIS)*, 17(2).
- Wang, H. (2022). Providing public access to archival materials and the problem of fact-checking in the post-truth era. *Social Science Quarterly*, 103(7). <https://doi.org/10.1111/ssqu.13232>
- Wang, L., Zhang, X., Su, H., & Zhu, J. (2023). *A Comprehensive Survey of Continual Learning: Theory, Method and Application*. <http://arxiv.org/abs/2302.00487>
- Wang, R., Huang, S., Wang, P., Shi, X., Li, S., Ye, Y., Zhang, W., Shi, L., Zhou, X., & Tang, X.
- Artificial Intelligence and the Digital Transformation.....**

- (2024). Bibliometric analysis of the application of deep learning in cancer from 2015 to 2023. *Cancer Imaging*, 24(1). <https://doi.org/10.1186/s40644-024-00737-0>
- Wilson, C. (2022). Public engagement and AI: A values analysis of national strategies. *Government Information Quarterly*, 39, 101652. <https://doi.org/10.1016/j.giq.2021.101652>
- Zhang, Y., Wu, M., Tian, G. Y., Zhang, G., & Lu, J. (2021). Ethics and privacy of artificial intelligence: Understandings from bibliometrics. *Knowledge-Based Systems*, 222. <https://doi.org/10.1016/j.knosys.2021.106994>