SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 10419-10433
Open Access. Online & Print

DOI: 10.5281/zenodo.20055240

GENERATIVE ARTIFICIAL INTELLIGENCE IN DIGITAL ARCHIVING: OPPORTUNITIES, CHALLENGES, AND FUTURE DIRECTIONS"

Hanan Ahmed Farag[1], Aljazi Mohammed Almutairi[2], Ghada taha Osman[3], Islam Babiker Elshafia[4] and Mouawia Hassa Saeed[5]

[1]Associate professor, Department of Information Science, College of Arts. Imam Abdul Rahman bin Faisal University. Email: hafarag@iau.edu.sa
[2]Assistant professor, Department of Information Science, College of Arts. Imam Abdul Rahman bin Faisal University. Email: jmalmotirey@iau.edu.sa
[3]Assistant professor, Department of Information Science, College of Arts. Imam Abdul Rahman bin Faisal University. Email: gtosman@iau.edu.sa
[4]Associate professor, Department of Information Science, College of Arts. Imam Abdul Rahman bin Faisal University. Email: ibalshafai@iau.edu.sa
[5]Lecturer- Deanship of Library Affairs. Imam Abdulrahman Bin Faisal University. Email: mhsaeed@iau.edu.sa

Received: 02/04/2026
Accepted: 23/04/2026

Corresponding Author: (Aljazi Mohammed Almutairi)

ABSTRACT

Generative artificial intelligence (GenAI) is emerging as a transformative technology in the management of digital archives, offering advanced capabilities for classification, indexing, and information retrieval. This study examines the potential of GenAI to enhance the utilization of digital archival resources, particularly in the context of increasing data volume and complexity, which often leads to the underuse of valuable information assets. The research aims to explore the benefits of adopting GenAI in digital archiving, identify key challenges associated with its implementation, and review prominent global models that demonstrate its practical applications in this field. A descriptive-analytical approach is employed to assess the impact of generative AI technologies on archival processes and services. The findings indicate that GenAI significantly improves operational efficiency by automating repetitive tasks, enhances staff productivity, and supports more effective service delivery. Additionally, it enables the provision of personalized user experiences and facilitates better access to and interaction with archival content. Despite these advantages, the study highlights the need to address technical, organizational, and ethical challenges to ensure sustainable and responsible integration of GenAI in digital archiving environments.

KEYWORDS: Generative Artificial Intelligence, Digital Archive, Data, Foundation Models, Large Language Models (LLMs).

1. INTRODUCTION

The rapid advancement of artificial intelligence (AI) technologies, coupled with the exponential growth in data volume, diversity, and computational capacity, is fundamentally reshaping digital information systems. Among these developments, generative artificial intelligence (GenAI) has emerged as a transformative paradigm that extends beyond traditional data processing by enabling machines to generate new content across multiple modalities, including text, images, audio, and video.

In recent years, particularly since late 2022, GenAI technologies have gained significant attention due to their ability to synthesize and analyze large-scale datasets from diverse sources. Unlike conventional AI systems that primarily focus on classification and prediction, generative AI is distinguished by its capacity to produce new, contextually relevant outputs. This capability makes GenAI particularly valuable in digital archiving, where it can support tasks such as metadata generation, content organization, classification, and information retrieval.

Despite the growing interest in applying AI technologies in archival contexts, a significant gap remains in understanding how generative AI can be effectively leveraged to address key challenges in digital archives. These challenges include data underutilization, increasing data complexity, and limited accessibility to archival resources. Existing studies have largely focused on traditional AI applications, with relatively limited attention given to the emerging role of generative models in transforming archival practices.

This study addresses this gap by examining the potential of generative artificial intelligence in improving digital archiving systems. It provides a comprehensive analysis of the benefits, challenges, and future implications of GenAI adoption, while also incorporating applied insights from real-world implementations.

Furthermore, this research aligns with the strategic direction of Saudi Arabia's Vision 2030, which emphasizes digital transformation and leadership in artificial intelligence. National initiatives led by the Saudi Data and Artificial Intelligence Authority (SDAIA), including the development of generative AI capabilities, reflect a growing commitment to integrating AI technologies across sectors, including digital heritage and information management.

1.1. Problem Statement

In the era of big data, digital archival resources have expanded significantly in terms of volume, variety, and complexity, resulting in a vast yet underutilized body of information. Despite their potential value, these resources often remain difficult to access and analyze due to issues related to data heterogeneity, fragmentation, and insufficient integration.

Consequently, there is a growing need for advanced technological approaches to enhance the management and usability of digital archives. In this context, generative artificial intelligence (GenAI) offers promising capabilities by enabling automated data structuring, integration, and enrichment. Through its ability to generate, organize, and interpret large-scale datasets, GenAI can significantly improve accessibility, efficiency, and knowledge extraction within digital archival systems.

This study seeks to answer the following questions:

1. How can generative artificial intelligence (GenAI) contribute to improving the management and utilization of digital archives?
2. What are the key applications and domains of generative artificial intelligence in digital archival systems?
3. How does GenAI enhance data organization, accessibility, and information retrieval within digital archives?
4. What is the main technical, organizational, and ethical challenges associated with the adoption of GenAI in digital archives, and how can they be addressed?
5. What are the future trends and implications of generative artificial intelligence for digital archiving and information institutions?

Main Objective of the study:

The primary objective of this study is to examine the potential of generative artificial intelligence (GenAI) in transforming digital archiving practices by enhancing data organization, accessibility, and utilization. The study seeks to evaluate how GenAI can support the development of institutional technical infrastructures and improve the efficiency and effectiveness of digital archive management.

Sub-Objectives of the study:

1. To analyze the conceptual foundations and operational mechanisms of generative artificial intelligence.
2. To investigate the key applications and benefits of GenAI in the development and management of digital archives.
3. To identify and evaluate the major domains in which generative AI can be applied within digital archival systems.

4. To examine the technical, organizational, and ethical challenges associated with the adoption of GenAI, and propose potential mitigation strategies.
5. To explore future trends and directions of generative AI in digital archiving and its implications for information institutions.

2. RESEARCH METHODOLOGY

This study adopts a descriptive–analytical approach to examine the impact of generative artificial intelligence (GenAI) on digital archiving practices. The descriptive component provides a comprehensive overview of generative AI technologies, their conceptual foundations, and their emerging applications in digital archives.

The analytical component focuses on evaluating the potential benefits and challenges associated with the adoption of GenAI. It also includes a qualitative analysis of selected real-world applications and case studies derived from existing literature, institutional reports, and documented digital projects. These cases were purposefully selected to represent diverse applications of GenAI in digital archives, including metadata enhancement, preservation, content generation, and user engagement.

The study further analyzes how these applications contribute to improving data organization, accessibility, and retrieval processes within digital archival environments.

3. LITERATURE REVIEW

Recent studies have increasingly highlighted the growing role of artificial intelligence, particularly generative AI, in transforming digital archives and information institutions. The reviewed literature was collected through a systematic search of major academic databases, including Scopus, Web of Science, and Google Scholar, in addition to relevant institutional reports and conference proceedings. These sources were selected to ensure comprehensive coverage of recent developments and diverse perspectives on the application of generative AI in digital archives.

Hawkins (2021) examined the integration of linked data and semantic web technologies in digital archives, emphasizing the role of advanced technologies—including generative approaches—in producing structured, interoperable, and shareable semantic data. The study identified key barriers limiting user access to digital archival resources and proposed the integration of AI-driven tools into semantic data workflows to enhance accessibility and utilization.

Similarly, Ajani et al. (2022) explored the awareness and readiness of academic librarians in Nigeria to adopt artificial intelligence technologies in library operations. Using a survey-based approach across six academic libraries, the study found that while awareness of AI applications is relatively high, institutional readiness remains uneven. Key challenges identified include limited funding, lack of technical expertise, and insufficient training. Despite these constraints, the study emphasized the potential of AI to improve efficiency and reduce human error in repetitive tasks.

In a broader organizational context, **Luca, Narayan, and Cox (2022**) analyzed the evolving role of artificial intelligence in library and information professions. Their study highlighted the alignment of library activities with national AI strategies and examined the ethical and regulatory dimensions of AI adoption, including issues related to data protection, algorithmic bias, and automated decision-making. The authors emphasized the importance of developing ethical, transparent, and trustworthy AI systems within information environments.

Additionally, the **CILIP report (2021) provided** a comprehensive analysis of the impact of artificial intelligence, machine learning, and automation on information professions. The study focused on the implications of these technologies for workforce transformation, highlighting the need for developing new competencies and skill sets to support users in increasingly AI-driven environments. It also underscored the importance of addressing ethical considerations and preparing future professionals for rapid technological change.

3.1. Relation Of the Current Study to Previous Research

Building on these studies, the current research aims to bridge the identified gap by providing a comprehensive analytical perspective on the role of generative artificial intelligence in digital archiving. Unlike previous studies that address AI in general terms, this study specifically focuses on generative AI as an emerging paradigm with the potential to transform archival practices.

The study contributes by examining both the technological capabilities of generative AI and its practical implications for improving data organization, accessibility, and utilization in digital archives. It also integrates technical, organizational, and ethical dimensions into a unified analytical framework, thereby offering a more holistic understanding of the opportunities and challenges

associated with adopting generative AI in archival environments.

4. THEORETICAL FRAMEWORK

1. Generative Artificial Intelligence (Genai)

The first decades of the 21st century have witnessed rapid advancements in artificial intelligence technologies, particularly in the ability of systems to process large-scale data, identify patterns, and generate predictive insights. Within this context, generative artificial intelligence (GenAI) has emerged as a transformative paradigm, enabling machines not only to analyze data but also to produce new and original content.

Generative AI refers to a class of artificial intelligence systems that leverage machine learning techniques, particularly deep neural networks, to generate novel outputs such as text, images, audio, and video based on patterns learned from training data (Spennemann, 2024). Unlike traditional AI systems that focus primarily on classification or prediction, GenAI is characterized by its capacity to create new data that resembles or extends existing datasets. This distinction is particularly significant in digital archiving, where the ability to generate metadata, summaries, and structured content can enhance the accessibility and usability of archival resources.

From a conceptual perspective, generative AI represents a shift toward systems that simulate aspects of human cognitive processes. As noted by Joshi (2019), the goal of AI is not necessarily to develop universally intelligent machines, but rather to design systems capable of performing tasks that reflect human-like behavior, such as natural language understanding, problem-solving, and decision-making in complex data environments. In digital archival contexts, this capability supports more advanced interaction with archival materials, including semantic search, automated description, and knowledge discovery.

At the core of generative AI are foundation models, which are large-scale machine learning systems trained on vast and diverse datasets, including text, images, and audio. These models are highly adaptable and can be fine-tuned for a wide range of applications. Notable examples include models such as BERT, DALL·E, and GPT-4. Within this framework, large language models (LLMs) represent a specialized category of foundation models designed specifically for processing and generating human language. For instance, GPT-3 was trained on hundreds of billions of words from diverse sources, while its successor, GPT-4, extends these capabilities by supporting multimodal inputs, including both text and images (Luca et al., 2022). These models enable practical applications in digital archives, such as automated cataloging, intelligent retrieval systems, and conversational interfaces.

Overall, generative AI can be understood as an umbrella concept encompassing advanced AI systems capable of producing original and contextually relevant content across multiple modalities. This capability positions GenAI as a key enabling technology in various domains, particularly digital archiving, where it supports data organization, knowledge discovery, content generation, and enhanced user interaction.

Figure No. (1): Shows Models for the Application of Generative Artificial Intelligence (Saudi Authority for Data and Artificial Intelligence, 2023).

Data is the cornerstone of generative AI—not only for training the foundation models themselves but

also for fine-tuning them to perform specific tasks. In an enterprise context, examples may include everything from legacy code to real-time operational data and customer insights. Generative AI technologies depend on multiple technology layers to support the operation of generative applications and services. These layers can be divided into four main layers: applications, development and operation tools, generative models, and technical infrastructure (Saudi Data and AI Authority, 2023).

2. *Specifications Of Generative Artificial Intelligence (Genai)*

Generative artificial intelligence (GenAI) is characterized by a set of advanced capabilities that distinguish it from traditional artificial intelligence systems. These capabilities enable the generation of diverse types of data, support adaptability across various domains, and enhance the efficiency of data-driven processes. In the context of digital archives, these characteristics contribute to improving data organization, accessibility, and knowledge extraction. The following sections outline the key specifications that define generative AI technologies and their operational characteristics.

4.1. Data Generation

Generative artificial intelligence models are designed to produce various forms of data, including text, images, and audio. The development of these models involves a series of structured stages, including data preparation, model construction, testing, deployment, and continuous refinement. This iterative lifecycle contributes to improving the efficiency, accuracy, and quality of generated outputs over time. In digital archival environments, this capability supports automated metadata generation, content summarization, and data enrichment.

4.2. Pre-Trained Foundation Models

Pre-trained foundation models serve as a fundamental basis for developing generative AI applications. These models are trained on large-scale and diverse datasets, enabling them to capture complex patterns across multiple domains. They can be further adapted to specific tasks through a process known as fine-tuning, which allows the model to specialize using domain-specific data. This approach significantly reduces training time and computational cost while enhancing the relevance and precision of the generated outputs.

For example, a model initially trained on general textual data can be fine-tuned using specialized archival datasets to support classification, indexing, and retrieval tasks.

4.3. Creativity And Originality

Generative AI systems are characterized by their ability to produce novel outputs that are derived from, yet not identical to, the data on which they were trained. This capability enables the generation of creative and contextually relevant content, which can be applied in digital archives for tasks such as reconstructing incomplete records or generating descriptive content for archival materials.

4.4. Versatility

Generative AI demonstrates a high degree of flexibility, as it can be applied to a wide range of creative and analytical tasks across various domains (Saudi Data and Artificial Intelligence Authority, 2023). In digital archives, this versatility supports multiple functions, including metadata enhancement, semantic organization, and user interaction.

4.5. Types Of Outputs

Generative AI models are capable of producing multiple types of outputs. Advanced systems, particularly multimodal models, can process and generate different data formats simultaneously, such as text and images. These capabilities are particularly valuable in digital archiving, where diverse content formats must be managed and analyzed.

The main categories of outputs include:

- **Text Generation:**

The creation or modification of textual content, including automated descriptions, summaries, and metadata for archival materials.

- **Image Generation**:

The production or enhancement of visual content, including restoration and reconstruction of archival images.

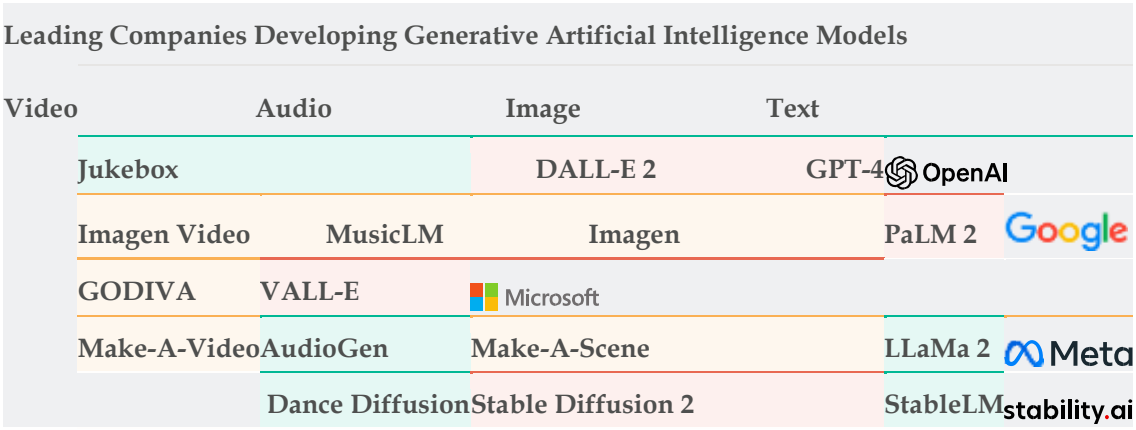
- **Audio Generation:**

The generation or modification of audio content, such as speech-based archival records.

- **Video Generation:**

The creation or editing of video content and animations, which can support digital storytelling and interactive archival experiences (Saudi Data and Artificial Intelligence Authority, 2023).

Figure No. (2): Highlights the Leading Companies Developing Generative AI Models (Saudi Authority for



Data and Artificial Intelligence, 2023).

3. Generative AI And Its Role in the Development of Digital Archives

Generative artificial intelligence (GenAI) has increasingly become a central component of institutional awareness, given its advanced capabilities in extracting, synthesizing, and summarizing information from large and diverse datasets. Within this context, an important research question emerges: to what extent will generative AI reshape the management and development of digital archives in the future? (Spennemann, 2024).

Scholars widely acknowledge that generative AI is poised to play a transformative role across various professional domains, including digital archiving. GenAI models are capable of delivering comprehensive and contextually relevant responses to complex queries within seconds, thereby enhancing access to knowledge. Despite concerns related to inclusivity, accuracy, and potential biases in training datasets, the widespread adoption of generative AI reflects its strong cross-sector applicability and user acceptance (Münster et al., 2024).

From an operational perspective, generative AI is expected to significantly enhance archival workflows by automating repetitive and time-consuming tasks such as classification, data extraction, summarization, aggregation, and large-scale content generation. In addition, these systems can uncover hidden patterns and relationships within archival data that may not be readily identifiable by human users, thereby supporting deeper knowledge discovery.

Furthermore, generative AI represents a critical advancement in the field of Natural Language Processing (NLP), enabling a wide range of language-based applications, including text classification, sentiment analysis, and the generation of contextually appropriate responses. These capabilities extend to producing structured and unstructured textual content, such as reports, emails, and even programming code, which can support various archival and administrative functions.

Technically, most generative AI models are based on Transformer architectures, which are designed to capture statistical, linguistic, and semantic patterns within textual data. Through training on extensive corpora—such as books, academic articles, and web content—these models learn to predict sequential language patterns, enabling them to generate coherent, human-like responses. This combination of large-scale data and advanced modeling techniques underpins the effectiveness of generative AI in digital archival environments (Rolan et al., 2018).

Generative artificial intelligence (GenAI) plays a significant role in advancing the digitization of manuscripts, particularly through its ability to recognize and interpret handwritten texts that are often difficult to decipher due to aging, damage, or deterioration. In addition, these systems can translate such texts into modern languages, thereby enhancing accessibility and usability. A notable example is Fabricius, a generative AI tool developed by Google Arts and Culture, which is capable of decoding ancient Egyptian hieroglyphs and demonstrates the broader potential of AI to interpret other ancient or lost languages (Plecher & Ela, 2020).

Despite these advancements, current generative AI tools remain limited in their ability to reliably construct predictive scenarios based on strategic foresight methodologies. However, they can effectively support the aggregation and synthesis of diverse information sources related to future opportunities. By generating structured summaries, these tools help clarify the current landscape of emerging trends, thereby assisting organizations in identifying potential pathways for future scenario development (van der Duin & Ligtvoet, 2019).

Furthermore, the analytical capabilities of

generative AI are reshaping research practices by shifting the focus from structured bibliographic data toward full-text analysis and large-scale data mining. Rather than retrieving individual information objects—such as books or reports intended for manual reading—GenAI enables knowledge discovery across entire collections of data. This transition reflects a paradigm shift from traditional information retrieval to more advanced, data-driven knowledge exploration.

At a broader level, generative AI operates within the framework of Big Data, significantly expanding the range of inputs and outputs that computational systems can process. This includes not only structured data but also unstructured formats such as text, audio, and images. Consequently, the concept of "data" itself has evolved to encompass a wider spectrum of information types that can be processed, analyzed, and generated by intelligent systems.

As highlighted by Kamran (2023), generative artificial intelligence has evolved beyond processing unstructured textual data to effectively handle multiple forms of information. These capabilities include:

Processing human speech inputs, such as voice-based search queries, and generating corresponding audio outputs, thereby enhancing human–computer interaction.

Identifying and classifying visual content with high accuracy, enabling advanced applications such as image-based retrieval and facial recognition systems that match human faces with existing visual datasets.

In addition to these capabilities, generative AI has begun to influence knowledge production and organization. Notably, the first fully computer-authored scientific text was published by Springer in 2019, reflecting the growing potential of AI in academic content generation. Emerging tools are increasingly capable of describing both individual texts and large-scale text collections, thereby supporting content organization and reuse.

Users benefit significantly from generative AI applications, particularly through automated summarization, which provides rapid access to key information, as well as recommendation and personalization systems that This development includes ensuring the delivery of relevant and tailored content.

Furthermore, generative AI enhances natural language processing for information retrieval and research within digital archives by transforming unstructured text and audio into structured and analyzable information. These systems can identify meaningful patterns within data, even when such patterns are not explicitly predefined.

From a methodological perspective, generative AI operates through different learning paradigms. In supervised learning, models rely on labeled datasets to learn pattern recognition, whereas in unsupervised learning, algorithms autonomously detect underlying structures without the need for pre-labeled data (Rolan et al., 2018).

4. Applications Of Generative Artificial Intelligence (Genai) In Digital Archives:

4.6. Generative AI As Interfaces for Existing Knowledge Discovery Systems

Generative artificial intelligence (GenAI) is increasingly used as an advanced interface in knowledge discovery systems, enhancing how users interact with and explore digital library content. A notable example is Yewno, integrated with ExLibris, which provides an innovative approach to information discovery through visual concept mapping. This enables users to navigate knowledge domains more intuitively by revealing relationships between concepts, moving beyond traditional keyword-based search.

In addition, GenAI supports the improvement of metadata creation and management by analyzing usage patterns and content characteristics. These capabilities help identify underutilized resources that may require reevaluation, thereby support more efficient collection management and enhance the effectiveness of digital archival systems.

4.7. Utilizing Generative AI For User Experience Management

Generative AI can support user experience management by analyzing interaction data, such as chat logs, to better understand user needs. These systems can guide users—such as students—toward relevant content and personalize materials based on prior usage patterns.

However, this process may face organizational challenges, including limited interoperability, as well as ethical concerns related to monitoring user interactions. Such practices may discourage users from exploring certain topics, potentially restricting freedom of expression.

According to Mishra (2021), the use of social and technical practices to collect, aggregate, and analyze user interactions and intellectual interests may conflict with core library values related to privacy and intellectual freedom. This highlights a central ethical dilemma associated with the application of AI

technologies.

4.8. Image Analysis and Retrieval

Generative AI and AI-driven algorithms play a significant role in the analysis, retrieval, and restoration of visual and audio materials, including images, paintings, photographs, and sound recordings that may be aged, damaged, or degraded. These technologies enhance image quality, remove noise and distortions, and can reconstruct missing elements, thereby supporting the preservation of cultural heritage assets.

In addition, such systems can identify stylistic features, classify artistic forms, attribute works to specific creators, and detect forged or counterfeit items. These capabilities contribute to improving the authenticity, reliability, and integrity of digital archival collections.

4.9. Source Recognition and Classification

Generative AI, particularly through computer vision techniques, enables the automated recognition and classification of cultural and archival sources. By analyzing visual characteristics and underlying patterns, these systems can accurately categorize artifacts, sculptures, and architectural elements, thereby supporting efficient organization, indexing, and cataloging of museum and archival collections.

Practical applications include automated metadata generation and prediction, such as detecting color patterns in textile artifacts, determining technical specifications, estimating historical periods, identifying materials, and tracing object origins. For example, such systems can support the analysis of historical materials, including textile collections like European silk artifacts.

In addition, generative AI can be used to recognize and classify symbols and inscriptions in ancient papyri, enhancing the accessibility and interpretability of historical documents (Haliassos et al., 2020).

4.10. Translation And Transcription

Generative AI models demonstrate advanced capabilities in translating ancient texts, inscriptions, and manuscripts into modern languages, thereby enhancing accessibility and cross-cultural understanding. These systems also support the translation of metadata and full-text heritage content into multiple languages, facilitating cross-linguistic resource sharing and broader dissemination of archival materials.

In addition, AI-powered transcription tools enable the automatic conversion of handwritten texts into machine-readable formats, allowing researchers to efficiently access, interpret, and analyze historical documents (Nockels et al., 2022).

4.11. Automated Text Analysis

Generative AI supports a wide range of automated text analysis techniques, including semantic indexing of structured and semi-structured historical content. These approaches enable researchers to efficiently process large volumes of data, uncover hidden patterns, and gain deeper insights into archival sources.

Such capabilities enhance large-scale textual exploration and facilitate knowledge extraction within digital archival environments.

4.12. Virtual Reality (VR) And Augmented Reality (AR)

Generative AI and AI-driven technologies play a key role in enabling immersive experiences through virtual reality (VR) and augmented reality (AR) applications in cultural heritage contexts. These technologies allow users to explore historical sites, ancient ruins, and museum collections in interactive virtual environments.

Moreover, AI-generated elements, such as virtual characters and objects, enhance user engagement and support a deeper understanding of cultural and historical contexts, particularly within digital archival and heritage platforms (Muenster, 2022).

4.13. Recommendation Systems for Personalized Experiences

Generative AI algorithms play a significant role in developing personalized user experiences by analyzing user preferences, historical interaction data, and contextual information. These systems enable the delivery of tailored cultural heritage (CH) content, such as recommending relevant exhibitions, generating customized virtual tours, and providing personalized informational materials, thereby enhancing user engagement and overall satisfaction.

The effectiveness of these systems has been further strengthened by the advancement of Large Language Models (LLMs), including GPT-based conversational agents and intelligent chatbots. These technologies support interactive and adaptive communication with users, enabling more dynamic and context-aware recommendation processes. Practical applications include the integration of chatbots within museum environments and the use of AI-driven recommendation systems in digital archival collections.

However, despite their advantages, these systems

also raise concerns related to information filtering and algorithmic bias, which may limit users' exposure to diverse content. Therefore, careful system design and ethical considerations are essential to ensure balanced and inclusive user experiences.

4.14. Artificial Intelligence Technologies

Artificial intelligence technologies, particularly Natural Language Processing (NLP), play a crucial role in analyzing large-scale cultural content, including literature, music, and artistic works. These approaches enable the identification of patterns, themes, and cultural influences, thereby providing deeper insights into historical contexts and artistic movements. In addition, they support metadata enrichment and integration with open data sources, enhancing the accessibility and interoperability of cultural heritage information (Muenster, 2022).

These analytical capabilities are supported by a range of advanced computational techniques, including:

1. Nlp Techniques

These include entity recognition, part-of-speech tagging, sentiment analysis, and topic modeling. Recent advancements in deep learning—particularly Convolutional Neural Networks (CNNs) and Transformer-based architectures—have significantly improved feature extraction and reduced errors, even in scenarios involving unsupervised or minimally supervised pre-training.

2. Text Classification Algorithms

Common approaches include Naive Bayes, Support Vector Machines (SVM), and Random Forests, which are widely used for categorizing textual data and supporting large-scale content organization.

3. Sequence Models

These include Hidden Markov Models (HMMs), Conditional Random Fields (CRFs), and Recurrent Neural Networks (RNNs), which are effective in modeling sequential data and capturing contextual relationships.

4.15. Digitization And Preservation of Cultural Heritage

Generative artificial intelligence (GenAI) plays a vital role in the digitization and preservation of cultural heritage by automating digitization and knowledge extraction processes. This enhances the preservation of rare artifacts and enables remote access to cultural collections for both researchers and the public.

Key applications include:

- **3D Modeling and Scanning:**

GenAI supports the creation of accurate digital replicas of physical objects, reducing the need for direct handling and minimizing deterioration risks. It also contributes to the restoration of damaged artifacts and texts. Examples include ArchAIDE for identifying archaeological fragments and AI-based reconstruction of fragmented objects using 3D technologies.

- **Virtual Reconstruction and Immersive Experiences:**

GenAI enables the reconstruction of historical sites and cultural artifacts by integrating data from multiple sources, supporting the creation of interactive and immersive environments.

- **Audiovisual Materials Processing:**

GenAI supports the digitization, restoration, and analysis of audiovisual content. Key functions include content summarization, metadata enhancement, and transcription and translation.

- **Intelligent Search and Retrieval:**

GenAI enhances search capabilities by enabling semantic and context-aware retrieval of archival materials, improving access to large and complex datasets.

- **Artistic Expression and Language Preservation:**

GenAI facilitates creative production by analyzing cultural content and generating new artistic outputs. It also supports endangered languages through translation, text generation, and speech synthesis.

- **Semantic Data and Access to Unstructured Content:**

GenAI enhances metadata and supports semantic data development, improving access to unstructured archival data and enabling more efficient retrieval.

- **Future Directions and Research Networks:**

The increasing adoption of GenAI is reflected in the emergence of research initiatives such as AEOLIAN, AURA, and HAIRA, which highlight ongoing advancements in AI applications within digital archives.

5. APPLIED CASE STUDIES OF GENERATIVE AI IN DIGITAL ARCHIVES

To further strengthen the practical dimension of this study, the analysis incorporates both international and regional (Arab) case studies. This combination allows for a more comprehensive understanding of how generative AI is applied across different archival contexts, while also reflecting

emerging developments within the Arab digital heritage ecosystem.

To complement the conceptual and analytical discussions presented in the previous sections, this study incorporates an applied perspective that examines real-world implementations of generative artificial intelligence (GenAI) in digital archives and cultural heritage contexts. The inclusion of applied case studies serves to bridge the gap between theoretical potential and practical deployment, providing empirical insights into how GenAI technologies are currently being utilized across different archival domains.

These case studies have been purposefully selected to reflect a diversity of applications, including metadata enhancement, multilingual access, knowledge organization, digital preservation, creative content generation, and user experience design. Collectively, they illustrate the evolving role of generative AI in transforming digital archives from static repositories into dynamic, intelligent, and user-centered knowledge environments.

Building on the descriptive overview of the selected projects, Table (2) provides an analytical expansion that links each case study to its underlying generative AI techniques, core archival functions, and the specific types of value generated within digital archival systems.

Table 1: Analytical Overview of Selected Generative AI Case Studies in Digital Archives

Project	GenAI Technique Used	Core Archival Function	Type of Value Added
Transcribathon (Europeana)	NLP + Handwriting Recognition	Digitization and transcription of manuscripts into machine-readable text	Improved accessibility and metadata enrichment
Europeana Translate	Machine Translation + NLP	Translation of metadata and archival content	Enhanced multilingual access to archives
Finto (Finland)	Semantic AI + Topic Modeling	Intelligent subject indexing and classification	Improved knowledge organization and information retrieval
Rekrei	Computer Vision + 3D Modeling	Digital reconstruction of damaged cultural artifacts	Supports digital preservation and cultural heritage restoration
The Next Rembrandt	Generative Models (GANs)	Creation of new artistic content based on historical styles	Enables creative reuse of cultural heritage data
Hidden Florence	AI + AR/VR + Storytelling Systems	Interactive and immersive user experience	Enhances user engagement and experiential access

This table provides an analytical overview on how generative AI techniques are operationalized within selected digital archive projects. The analysis reveals that these applications extend across multiple functional domains, including digitization, metadata enhancement, knowledge organization, preservation, and user experience.

Notably, different generative AI techniques contribute to distinct value dimensions. For example, NLP-based applications primarily enhance accessibility and metadata quality, while computer vision and 3D modeling technologies play a critical role in digital preservation and reconstruction. In contrast, generative models enable new forms of cultural production, shifting the role of digital archives from passive repositories to active creators of content.

Furthermore, user-centered applications, such as immersive storytelling and augmented reality environments, demonstrate a transformation in how users interact with archival materials, moving from traditional information retrieval toward experiential and exploratory engagement. This transformation highlights the evolving nature of digital archives as dynamic, intelligent, and user-driven knowledge environments.

5.1. Transformational Dimensions of Generative AI In Digital Archives

To further deepen the applied analysis, this table examines the transformative impact of generative artificial intelligence (GenAI) across selected digital archive projects. Rather than focusing on functional or technical aspects, the table highlights how these applications contribute to fundamental shifts in the nature and role of digital archives. This perspective emphasizes transformation as a key dimension in understanding the broader implications of GenAI adoption within archival environments.

Table (2): Transformational Impact of Generative AI Applications.

Project	Level of Impact	Type of Transformation in Digital Archives
Transcribathon (Europeana)	High	From unreadable content → searchable digital content
Europeana Translate	High	From local archives → globally accessible multilingual archives
Finto (Finland)	Medium–High	From traditional indexing → intelligent semantic indexing
Rekrei	High	From physical loss → digital reconstruction
The Next Rembrandt	Medium	From content preservation → content creation

Hidden Florence	High	From passive use → immersive interactive experience

The transformation-oriented analysis presented in Table (2) reveals that generative AI is driving a paradigm shift in digital archival systems across multiple dimensions.

First, at the level of accessibility, applications such as Transcribathon and Europeana Translate demonstrate a clear transition from limited and localized access toward inclusive and globally accessible archival ecosystems. By converting handwritten or language-restricted content into searchable and multilingual formats, these systems significantly reduce structural barriers and improve information accessibility.

Second, in terms of knowledge organization, projects such as Finto demonstrate a shift from static and rule-based indexing toward adaptive and semantically enriched systems. This transformation enhances the efficiency of information retrieval and supports more advanced forms of knowledge navigation

Third, the case of Rekrei illustrates a fundamental shift in preservation practices. Rather than focusing solely on safeguarding existing materials, digital preservation becomes an active reconstructive process, where lost or damaged artifacts can be digitally reconstructed and reintroduced into the cultural record.

Moreover, generative AI introduces a new dimension in which digital archives are no longer confined to preservation functions. As demonstrated by The Next Rembrandt, digital archives can function as sources for generating new cultural outputs, thereby expanding their role into creative and knowledge production domains.

Finally, user interaction experiences a significant transformation through applications such as Hidden Florence, where traditional passive consumption of archival materials evolves into immersive and interactive experiences. This shift reflects a broader transition toward user-centered and experience-driven archival environments.

Overall, these transformations indicate that generative AI is not merely enhancing existing archival processes but is fundamentally redefining both the epistemological and functional boundaries of digital archives.

5.2. Arab Applications of Generative AI In Digital Archives

In the Arab context, emerging initiatives reflect an increasing adoption of artificial intelligence technologies in digital archiving, with gradual integration of generative AI capabilities. For example, the Saudi Digital Library (SDL) provides large-scale access to digital resources and has begun incorporating intelligent retrieval systems to enhance user experience. Similarly, the Qatar Digital Library offers advanced digital heritage collections supported by metadata enrichment and multilingual accessibility. These initiatives suggest that, although the adoption of generative AI in Arab archival institutions remains at an early stage, there is a clear shift toward more intelligent, accessible, and user-centered digital knowledge environments.

In comparison to international initiatives, these applications remain at relatively earlier stages of generative AI integration, highlighting the need for further technological investment and research in the Arab region.

Table (3): Selected Arab Applications of AI In Digital Archives.

Project	GenAI Technique	Function	Value
Saudi Digital Library	AI-based retrieval / NLP	Access & retrieval	Improved access
Qatar Digital Library	Metadata enrichment + NLP	Digital preservation	Multilingual access

This table demonstrates that Arab digital archive initiatives are increasingly incorporating AI-driven techniques, particularly in enhancing accessibility and multilingual capabilities. However, the integration of advanced generative AI applications remains relatively limited, indicating significant opportunities for further development and innovation in the region.

6. CHALLENGES AND RISKS OF GENERATIVE AI IN DIGITAL ARCHIVES

Despite the considerable potential of generative artificial intelligence (GenAI) in enhancing digital archives, several challenges and risks must be carefully addressed to ensure effective and responsible implementation.

6.1. Data Quality and Availability

The performance of generative AI systems largely depends on the quality and diversity of training data (Münster et al., 2024). However, archival data is often heterogeneous, incomplete, and challenging to standardize. In addition, legal restrictions and institutional ownership limit access to high-quality

datasets, particularly those containing primary and visual materials (Tan et al., 2018).

6.2. Model Limitations and Lack of Transparency

Generative AI models, particularly large-scale neural networks, often operate as "black boxes", producing outputs without providing clear explanations of how results are generated (Spennemann, 2024). This lack of transparency raises concerns regarding interpretability, accountability, and trust in AI-generated outcomes.

6.3. Bias And Hallucination Risks

Generative AI systems may produce biased or inaccurate outputs due to biases or limitations in their training datasets (Spennemann, 2024). In certain cases, these systems may generate misleading or fabricated information, commonly referred to as hallucinations, which necessitates careful human validation and critical evaluation.

6.4. Technical Challenges in Cultural Data Processing

Processing cultural and historical data presents specific technical challenges, including the degradation of archival materials such as faded manuscripts and damaged images, variability in formats and resolutions, and the limited availability of high-quality datasets for training AI models (Münster et al., 2024).

6.5. Ethical And Legal Concerns

The adoption of generative AI raises several ethical and legal issues, including copyright infringement, data privacy and security, and algorithmic bias (Spennemann, 2024). Furthermore, the absence of clear regulatory frameworks complicates the effective governance of AI applications in digital archives.

6.6. Over-Reliance On AI And Loss of Human Expertise

Excessive reliance on AI technologies may contribute to the erosion of critical human skills, including interpretation, analytical thinking, and contextual understanding (Popova, 2023). However, human expertise remains essential to ensure the accuracy, ethical integrity, and contextual relevance of AI-generated outputs (Spennemann, 2024).

6.7. Complexity Of Outputs and Need for Oversight

Generative AI systems can produce complex and potentially unpredictable outputs that may be difficult to interpret and evaluate (Popova, 2023). Therefore, transparency, clear documentation of processes, and continuous human oversight are essential to ensure responsible and reliable use.

6.8. Customization And Infrastructure Challenges

The effective deployment of generative AI requires flexible and adaptable systems that can be tailored to specific research and institutional needs. This includes advanced technical infrastructure, skilled professionals, and collaborative environments to support model training and refinement (Ferrara et al., 2022).

6.9. Data Security and Privacy Risks

Generative AI raises significant concerns related to data security and privacy, including the risk of unauthorized access to sensitive information (Spennemann, 2024). Institutions must implement comprehensive data governance policies, robust access controls, and continuous monitoring mechanisms to mitigate these risks.

7. FUTURE PERSPECTIVES OF GENERATIVE AI IN DIGITAL ARCHIVES

In view of the challenges and limitations discussed above, the future of generative artificial intelligence (GenAI) offers significant opportunities and strategic directions for transforming digital archives and information institutions.

Although Artificial General Intelligence (AGI) remains in its early stages, the rapid advancement of artificial intelligence technologies signals the emergence of a new digital paradigm. This paradigm is expected to reshape fundamentally how information is accessed, content is created, services are delivered, and organizations are managed (Russell & Norvig, 2021; Zhang & Lu, 2021). Within this evolving landscape, generative AI, large language models (LLMs), and foundation models are expected to play a central role in enhancing productivity, augmenting human capabilities, and enabling new pathways for innovation and growth (Bommasani et al., 2021; OpenAI, 2023).

In the context of digital archives, future developments of generative AI are expected to support the creation of intelligent and adaptive archival systems capable of automating complex processes such as metadata generation, semantic organization, and knowledge discovery. These advancements will contribute to transforming digital archives from static repositories into dynamic and

interactive knowledge environments (Wang et al., 2023; Münster et al., 2024).

However, realizing these opportunities requires a comprehensive rethinking of archival practices, including the development of advanced information technology infrastructures, flexible organizational models, and adaptive cultural frameworks. Institutions must also address critical challenges related to transparency, accountability, and ethical governance to ensure responsible AI adoption (UNESCO, 2021; Luca et al., 2022).

Furthermore, the successful implementation of generative AI depends on the balanced integration of technological innovation and human expertise. This includes investing in advanced digital infrastructures, redesigning workflows, and equipping information professionals with interdisciplinary skills in AI engineering and data management (Ray, 2023; McKinsey Global Institute, 2023).

Looking ahead, generative AI is expected to expand its applications within libraries and information institutions, particularly in areas such as advanced data analytics, personalized information services, remote access to archival resources, and the integration of immersive technologies. These developments will position digital archives as active knowledge ecosystems that support research, learning, and innovation (Brown et al., 2020; Raffel et al., 2020).

Overall, the future of generative AI in digital archives lies in its potential to create more accessible, intelligent, and user-centered information environments, while maintaining a careful balance between technological advancement and ethical responsibility.

8. CONCLUSION

Over the past two decades, substantial investments in digitization and digital transformation have led to the creation of extensive digital archival resources. These resources have evolved from mere digital substitutes into dynamic and valuable digital assets that can be effectively leveraged through advanced technologies.

This study demonstrates that generative artificial intelligence (GenAI) represents a transformative approach in digital archiving, with significant potential to enhance data organization, accessibility, and utilization. It highlights how GenAI contributes to improving productivity, automating repetitive processes, enriching metadata, and enabling more efficient and intelligent information retrieval.

Furthermore, the study shows that GenAI extends beyond operational efficiency by supporting knowledge discovery, content generation, and personalized user experiences within digital archival environments. However, its adoption is accompanied by significant technical, organizational, and ethical challenges that require careful consideration.

Overall, this study contributes to the growing body of knowledge by providing a comprehensive analytical and applied perspective on the role of generative AI in digital archives. It emphasizes the transformative potential of GenAI in reshaping archival practices and supporting the future development of intelligent, user-centered information institutions.

9. RECOMMENDATIONS

- Develop a clear strategic framework for the adoption of generative AI in digital archives, aligned with institutional goals and measurable outcomes.
- Establish and continuously update legal and intellectual property policies to address challenges associated with generative AI applications.
- Ensure the availability of high-quality, diverse, and well-structured datasets to support accurate and reliable AI outputs.
- Implement mechanisms to detect and mitigate algorithmic bias by utilizing diverse datasets and applying fairness and accountability standards.
- Invest in continuous professional development programs to equip information professionals with interdisciplinary skills related to generative AI technologies.
- Strengthen data protection and privacy policies by ensuring secure access to sensitive information and promoting user awareness.
- Continuously evaluate and fine-tune generative AI models to maintain output quality, reliability, and alignment with institutional requirements.

REFERENCES

Ajani, Y. A., Tella, A., Salawu, K. Y., & Abdullahi, F. (2022). Perspectives of librarians on awareness and readiness of academic libraries to integrate artificial intelligence for library operations and services in Nigeria. *Internet Reference Services Quarterly, 26*(4), 213–230.

https://doi.org/10.1080/10875301.2022.2086196
Anichini, F., Fantini, F., & Gattiglia, G. (2021). The automatic recognition of ceramics from only one photo: The ArchAIDE app. *Journal of Archaeological Science: Reports, 36*, 102788. https://doi.org/10.1016/j.jasrep.2020.102788
Babl, F. E., & Babl, M. P. (2023). Generative artificial intelligence: Can ChatGPT write a quality abstract? *Emergency Medicine Australasia, 35*(5), 809–811. https://doi.org/10.1111/1742-6723.14233
Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., … Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv Preprint*. https://arxiv.org/abs/2108.07258
Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901. https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html
Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*. https://doi.org/10.18653/v1/N19-1423
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In *Advances in Neural Information Processing Systems, 27*. https://papers.nips.cc/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html
Hagendorff, T., & Fabi, S. (2023). Why we need biased AI: How including cognitive biases can enhance AI systems. *Journal of Experimental & Theoretical Artificial Intelligence*. https://doi.org/10.1080/0952813X.2023.2178517
Haliassos, A., Antonacopoulos, A., & Karatzas, D. (2020). Classification and detection of symbols in ancient papyri. In *Visual Computing for Cultural Heritage* (pp. 121–140). https://doi.org/10.1007/978-3-030-37191-3_7
Hawkins, A. (2021). Archives, linked data and the digital humanities: Increasing access to digitised and born-digital archives via the semantic web. *Archival Science, 22*(3), 319–344. https://doi.org/10.1007/s10502-021-09381-0
Joshi, A. V. (2019). Introduction to AI and ML. In *Machine learning and artificial intelligence* (pp. 3–7). Springer. https://doi.org/10.1007/978-3-030-26622-6_1
Kaplan, A. M., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? *Business Horizons, 62*(1), 15–25. https://doi.org/10.1016/j.bushor.2018.08.004
Koutsoudis, A., Arnaoutoglou, F., & Chamzas, C. (2007). On 3D reconstruction of the old city of Xanthi. *Journal of Cultural Heritage, 8*(1), 26–31. https://doi.org/10.1016/j.culher.2006.08.003
Liang, P., et al. (2022). Holistic evaluation of language models. *Transactions on Machine Learning Research*. https://openreview.net/forum?id=iO4LZibEqW
Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., … Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv Preprint*. https://arxiv.org/abs/1907.11692
Luca, E., Narayan, B., & Cox, A. (2022). Artificial intelligence and robots for the library and information professions. *Journal of the Australian Library and Information Association, 71*(3), 185–188. https://doi.org/10.1080/24750158.2022.2104814
Lucchi, N. (2023). ChatGPT: A case study on copyright challenges. *European Journal of Risk Regulation*. https://doi.org/10.1017/err.2023.59
Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
McKinsey Global Institute. (2023). *The economic potential of generative AI: The next productivity frontier*. https://www.mckinsey.com
Mishra, M. (2021). Machine learning techniques for heritage buildings. *Journal of Cultural Heritage, 47*, 227–245. https://doi.org/10.1016/j.culher.2020.09.005
Münster, S., et al. (2024). Artificial intelligence for digital heritage innovation. *Heritage, 7*(2), 794–816. https://doi.org/10.3390/heritage7020038
Nockels, J., et al. (2022). Handwritten text recognition in heritage contexts. *Archival Science, 22*(3), 367–392. https://doi.org/10.1007/s10502-021-09390-z
OpenAI. (2023). GPT-4 technical report. *arXiv Preprint*. https://arxiv.org/abs/2303.08774
Periti, F., et al. (2022). Semantic shift detection. In *Proceedings of the Workshop on Computational Approaches to Historical Language Change*. https://doi.org/10.18653/v1/2022.lchange-1.5
Plecher, D. A., et al. (2020). Learning Egyptian hieroglyphs. In *IEEE ISMAR Adjunct*.

https://doi.org/10.1109/ISMAR-Adjunct51615.2020.00039
Popova, D. (2023). Digital reconstruction of lost heritage. *Digital Presentation and Preservation of Cultural and Scientific Heritage, 13*, 117–126. https://doi.org/10.55630/dppcsh.2023.13.10
Raffel, C., et al. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research, 21*(140), 1–67. https://www.jmlr.org/papers/v21/20-074.html
Ray, P. P. (2023). ChatGPT: A comprehensive review. *Internet of Things and Cyber-Physical Systems, 3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003
Rolan, G., et al. (2018). Artificial intelligence in the archive. *Archives and Manuscripts, 47*(2), 179–203. https://doi.org/10.1080/01576895.2018.1512669
Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
Schreur, P. E. (2020). Linked data and AI in technical services. *Cataloging & Classification Quarterly, 58*(5), 473–485. https://doi.org/10.1080/01639374.2020.1792323
Smith, A. (2020). Emerging technologies in libraries. *Library Hi Tech, 38*(3), 567–579. https://doi.org/10.1108/LHT-09-2019-0173
Spennemann, D. H. (2024). AI and cultural heritage management. *Heritage, 7*(3), 1453–1471. https://doi.org/10.3390/heritage7030078
Tan, C., et al. (2018). Deep transfer learning. In *Artificial Neural Networks and Machine Learning* (pp. 270–279). https://doi.org/10.1007/978-3-030-01424-7_27
Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433
UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000381137
van der Duin, P., & Ligtvoet, A. (2019). Foresight processes. *Futures, 105*, 187–198. https://doi.org/10.1016/j.futures.2018.10.005
Wang, P., et al. (2023). AI in digital heritage preservation. *Heritage Science, 11*(1), 1–15. https://doi.org/10.1186/s40494-023-00891-2
Zhang, C., & Lu, Y. (2021). Artificial intelligence: State of the art and future prospects. *Journal of Industrial Information Integration, 23*, 100224. https://doi.org/10.1016/j.jii.2021.100224