



www.jpis.az

Application of artificial intelligence in digital archiving: problems, solutions and prospects

*Elchin Aliyev*¹, *Islam Ahmadov*², *Khanmurad Abdullayev*³

^{1,3}Institute of Control Systems, B. Vahabzadeh str., 68, AZ1141, Baku, Azerbaijan

²SINAM Company, B. Vahabzadeh str., 68, AZ1141, Baku, Azerbaijan

¹elchin@sinam.net, ²islam@sinam.net, ³khan.abdullayev@gmail.com

¹ orcid.org/0000-0002-7987-2674, ² orcid.org/0000-0001-6001-1519, ³ orcid.org/0000-0002-0095-3114

ARTICLE INFO

<https://doi.org/10.25045/jpis.v17.i1.03>

Article history:

Received 16 September 2025

Received in revised form

18 November 2025

Accepted 19 January 2026

Keywords:

Electronic Archives

Digital Archiving

Artificial Intelligence (AI)

Machine Learning (ML)

GLAM (Galleries, Libraries, Archives,
and Museums)

Data Integrity

Natural Language Processing (NLP)

Cloud-Based Archiving

ABSTRACT

Digital transformation converts both legacy and newly generated archival records into electronic, i.e., digital data. As a result, automation methodologies based on artificial intelligence (AI) techniques are increasingly being implemented not only in the execution of traditional record-keeping activities but also in testing and developing new approaches for collecting, organizing, and retrieving information in various formats. Within the framework of digitalization challenges in the archival domain, this article examines the intersection of Artificial Intelligence with archival theory and practice, highlighting recent achievements in this area. Taking into account the processes occurring within the digital transformation of the economy of the Republic of Azerbaijan, the authors attempt to develop a systematic approach to the application of AI technologies in archival institutions. The purpose of the study is to analyze the practical experience of implementing technological solutions, particularly Artificial Intelligence, in archival activities and to explore the regulatory and methodological frameworks governing this field. The authors identify key tasks essential for establishing future-oriented archives based on the extensive use of AI technologies and for defining the fundamental standards for their application in archival practice.

The implementation of these tasks requires the adoption of AI technologies as well as the training of qualified professionals capable of effectively integrating these technologies into the practical work of archivists.

1. Introduction

The literature addressing archives and Artificial Intelligence (AI) encompasses several broad thematic areas, including theoretical and professional considerations, the automation of record-keeping processes, the organization and use of archives, and the emergence of new forms of digital archives. Ultimately, current research highlights emerging trends and future directions that involve applying record-keeping principles to the data and processes that power modern AI systems, as well as integrating AI into archival

systems and practice in a more structured yet critical manner. From the earliest stages of human information management, the preservation and administration of records have constituted a continuous process in one form or another. In the digital era, however, this task has become more accessible and technically feasible. These developments underscore the growing recognition of the strategic importance of information. Alongside technological advancements and the expansion of internet-based workflows and digital document management, digital archiving has evolved from an optional innovation into a fundamental necessity.

In a broader sense, what constitutes Digital Archiving? Some scholars argue that it reflects a historical process of adaptation to technological change, while others define it primarily as a form of transformation. Nevertheless, there is general consensus that electronic archiving (e-archiving) is responsible for ensuring the preservation, integrity, and accessibility of valuable information without loss, distortion, or degradation over time.

Since the establishment of the Document Management System (DMS) in 1999, its functional potential has not been fully utilized in accordance with its intended purpose. It may be noted that, in many cases where DMS platforms are currently implemented, achieving effective results requires additional attention and optimization. Therefore, replacing conventional DMS solutions with comprehensive digital archiving systems may provide a more sustainable response to long-term digital preservation challenges by enhancing accessibility and ensuring information security.

In the near future, archival professionals will face the task of managing and utilizing a significant volume of records generated without direct human participation—within information systems or through automated devices—figuratively described as “robotized” record-keeping. This development necessitates the identification of key issues to be addressed in the creation of next-generation archives based on the widespread adoption of Artificial Intelligence (AI), as well as the establishment of fundamental standards governing its use in archival practice. The issues examined and the proposed solutions are grounded in extensive professional experience in both information systems management and documentation and archival studies.

Determining the role of AI in archival operations enables a focused approach to the implementation of selected functional mechanisms, thereby facilitating the processing of the anticipated growth in records generated by information systems. It is evident that no information system or AI technology can replace human expertise; however, effective interaction with such systems requires appropriately trained specialists equipped with advanced digital competencies.

The scientific novelty of this research lies in the development of an integrated analytical framework for AI-driven digital archiving systems that simultaneously addresses technological, organizational, ethical, and sustainable development dimensions. Unlike existing studies that are often limited to describing isolated technical solutions, the proposed approach provides a comprehensive

analysis of the interconnections among archival science, artificial intelligence, information security, and digital governance. Furthermore, the framework is contextualized within the ongoing digital transformation processes of the Republic of Azerbaijan, offering a systematic approach that takes into account regional specificities.

2. Related work

At present, Artificial Intelligence (AI) has become an integral part of everyday life, with new solutions and innovative approaches emerging almost daily in this field (Abbasov & Rzayev, 2023). The central issue is no longer whether archival systems should incorporate AI technologies, but rather how information technology specialists and archivists can effectively leverage the full potential of AI within the digital archiving process. Contemporary research increasingly focuses on identifying the points of intersection between AI theory and archival science, as well as exploring the practical applications of these technologies.

In the field of digital archiving, deeper and more practice-oriented understanding has been developed not only through theoretical contributions from mathematicians and linguists but also through real-world implementations of advanced AI-based technologies. Over the past three decades, significant efforts have been undertaken to establish effective practices for the preservation of digital materials within the GLAM (Galleries, Libraries, Archives, and Museums) sector (Deegan & Tanner, 2006; Delve & Anderson, 2014; Corrado & Sandy, 2017; Harvey & Weatherburn, 2018).

To ensure the protection of data and intellectual property rights, AI technologies have been widely integrated into many digital archival systems (Jaillant & Caputo, 2022; Jaillant, 2022a, 2022b). Examples of AI-based tools used to enhance the accessibility and efficiency of digital archives include: automated sensitivity review and classification of government records within the framework of Freedom of Information requests prior to publication or transfer to national archives (The National Archives UK, 2016; McDonald et al., 2020a; Baron et al., 2020); the application of computer vision technologies to improve the discovery and analysis of visual archival collections (Abbasov et al., 2001; Angelova et al., 2020); and the deployment of algorithmic search mechanisms (Nix & Decker, 2021).

As awareness of AI technologies increases, public scrutiny also intensifies, bringing ethical concerns—

including transparency and algorithmic bias—into broader societal debate (Fast & Horvitz, 2017; Cave et al., 2019; Kelley et al., 2021). There is a growing call for the development of widely agreed-upon standards and governance mechanisms regulating the algorithms that underpin many AI systems. Digital archives face distinct ethical challenges arising from the application of AI. Recent studies indicate that ethical guidelines currently exert only limited influence on software engineering practices (Biegelbauer et al., 2022; Hagendorff, 2020). Ethical questions related to emerging technologies increasingly intersect with professional codes of conduct governing practice within the GLAM sectors.

In addition to ethical standards embedded in civil service, GLAM institutions, and academic professions, AI technologies are also subject to various ethical codes and guidelines developed by private companies, research institutions, and public sector bodies—although these frameworks are often less formally standardized or internationally harmonized (Jobin et al., 2019; Mittelstadt, 2019; Hagendorff, 2020).

Most existing research focuses on isolated technical or functional aspects of AI implementation in digital archiving. For example, some studies emphasize automated classification and search systems, while others concentrate primarily on ethical or legal considerations. In contrast, the present study proposes a unified conceptual framework that analyzes the integration of AI into archival systems across technological, organizational, security, and governance dimensions. By synthesizing existing scholarship and addressing its limitations, this article offers a comprehensive and systemic approach to AI-driven digital archiving.

Relevance of the Research Problem. In the context of digital transformation, the growing volume of archival collections and the increasing diversity of data formats have exposed the limitations of traditional archiving approaches. The rapid expansion of born-digital records, multimedia content, and dynamically generated data has significantly complicated processes related to classification, preservation, access, and long-term management.

In this regard, the application of Artificial Intelligence (AI) technologies to archival activities emerges as a necessary scientific and practical solution for ensuring systematic information management, safeguarding data security, and providing efficient and timely access to records. AI-

driven tools offer the potential to automate metadata generation, enhance searchability, detect sensitive content, and support decision-making processes within digital archival systems.

The relevance of this article is therefore determined by the need to investigate the role of AI in improving the efficiency, sustainability, and strategic development of archival practice within an increasingly digitalized society. By addressing these challenges, the study contributes to the broader discourse on modernizing archival infrastructures in alignment with contemporary technological and governance frameworks.

3. Materials and methods

3.1. Research methodology

In this study, the primary materials consist of existing scholarly literature on the application of Artificial Intelligence (AI) in digital archiving, international analytical reports, and documented practical projects. The research employs comparative analysis, a systematic literature review, and a qualitative analytical approach. In addition, practices related to the management of digital archives in various countries, as well as examples of AI-based technological implementation, were examined in order to identify key challenges, potential solutions, and future development trajectories.

Conceptually, long before the emergence of what is now referred to as “big data,” archival collections already existed in the form of extensive physical files and folders stretching over kilometers of shelving. As a result of large-scale digitization initiatives worldwide, at least part of these collections has been successfully converted into digital data. Furthermore, since the 1990s, governments and other institutions with archival responsibilities have increasingly operated in digital formats.

However, these transformations did not immediately lead to a fundamental restructuring of archival practice and workflows. The substantial scale of collections and records meant that archival processes continued to rely predominantly on manual procedures. This was largely due to the fact that appraisal, selection, and review activities were traditionally grounded in human expertise and professional judgment. Nevertheless, this challenge proved to be transitional. An increasing number of archival collections are now digitized, and vast volumes of born-digital records are being transferred to archival institutions on an unprecedented scale.

Consequently, manual archival processing has become progressively less feasible. At the same time,

high-quality and reliable archiving—core principles of archival science—continues to require rigorous appraisal of records. As a result, archivists increasingly depend on machine agents to support the processing and analysis of large-scale archival datasets. This development reshapes the professional role of archivists, who must learn not only to employ machine-assisted reasoning in appraisal and selection processes but also to critically evaluate algorithmic outputs.

In this context, archival institutions are evolving into organizations that operate within a big data environment. Like other data-intensive institutions, archives must therefore rely, at least partially, on AI tools—particularly machine learning technologies—to effectively manage ongoing digital transformation processes.

3.2. Problem solution

Archives worldwide are increasingly restructuring their workflows around the large-scale datasets they manage, while simultaneously adopting AI-driven capabilities to provide controlled access to their data for external institutions and stakeholders. This shift reflects the growing recognition that archival institutions operate within a data-intensive ecosystem requiring scalable, automated, and intelligent processing mechanisms.

A decade ago, activities related to Artificial Intelligence in archival environments were largely experimental in nature. At that time, AI applications primarily demonstrated potential by proposing innovative methods for working with specific segments of archival holdings, such as digitized newspaper collections and other discrete digital corpora. These initiatives illustrated the feasibility of computational approaches but remained limited in scope and institutional integration.

This situation has fundamentally changed. A new trend has emerged in which AI technologies are increasingly applied not merely to access and discovery functions, but also to core record-keeping processes that characterize archival institutions. AI is now being incorporated into appraisal, classification, metadata generation, sensitivity review, and large-scale information management tasks. As a result, the intersection between AI technologies and archival processes is no longer peripheral but structural in nature.

The present study therefore examines recent research addressing the convergence of AI and archival practice, with particular attention to how intelligent systems are reshaping archival

workflows, professional roles, and institutional strategies. By analyzing this evolving intersection, the article aims to clarify the conceptual, technological, and organizational implications of AI integration within contemporary archival systems.

3.3. Proposed solution

Digital Archiving. If we begin by defining the fundamental principles of digital archiving, it becomes evident that it represents a structured process of preserving and organizing digital information to ensure long-term access and retrieval. In essence, digital archiving may be understood as a distinct evolution of traditional paper-based archives, adapted to electronic records and data management environments.

Digital archiving involves not only the secure storage of digital records but also the implementation of a systematic and structured approach to their preservation. This includes organizing information in a manner that maintains authenticity, integrity, accessibility, and contextual coherence over time. Unlike simple digital storage solutions, digital archiving requires deliberate strategies designed to safeguard records against technological obsolescence, data corruption, and loss.

The effectiveness of digital archiving depends on the comprehensive application of tools and strategies that meet the demands of the modern digital ecosystem. These may include standardized metadata frameworks, secure repository infrastructures, preservation policies, and intelligent retrieval mechanisms.

In simple terms, a digital archive does not merely upload and store files on electronic devices; rather, it functions as an integrated system that responds to the technological, organizational, and informational requirements of the digital age. It ensures that digital records remain usable, verifiable, and accessible across time, despite continuous technological change.

Objectives of Digital Archiving. Electronic archiving serves several critical objectives whose importance has become increasingly evident in the digital age. Among these objectives, one of the most fundamental is ensuring the long-term preservation of information by preventing data loss, corruption, or distortion over time. In rapidly evolving digital environments, safeguarding records against technological obsolescence and system failures is essential for maintaining their evidential and historical value.

Data integrity represents a core principle of electronic archiving. Ensuring integrity means guaranteeing that records remain complete, accurate, and unaltered from their original state. One widely adopted technical mechanism for protecting data integrity is the use of verification codes, commonly known as checksums. A checksum is a cryptographic or algorithmic value generated from a file's contents; it functions as a digital fingerprint. If even a minor alteration occurs in the file, the checksum value changes, thereby signaling potential corruption or unauthorized modification. Through regular checksum validation procedures, archival systems can continuously monitor and verify the authenticity and stability of stored records.

A second major objective of electronic archiving is to ensure efficient accessibility and searchability of digital materials. Unlike traditional paper-based archives, electronic systems enable advanced indexing, metadata structuring, and intelligent search capabilities that significantly improve retrieval speed and accuracy.

The digitization of physical assets also contributes to cost efficiency. By converting paper-based collections into digital formats, institutions can reduce expenses associated with physical storage space, preservation materials, and manual handling processes. Furthermore, disaster recovery and data restoration capabilities constitute another strategic priority. Digital archiving systems often incorporate backup infrastructures, redundancy mechanisms, and recovery protocols that allow institutions to restore records in the event of system failures, cyber incidents, or natural disasters.

In summary, electronic archiving fulfills not only preservation and accessibility objectives but also enhances operational efficiency, institutional resilience, and long-term sustainability in digital information management.

The Significance of Digital Archiving. In the contemporary era, the preservation and management of information have become issues of critical importance. The present analysis serves as a conceptual framework outlining the definition, significance, objectives, and key components of electronic archiving. Despite continuous technological advancements, data remain inherently vulnerable. Society constantly faces the risk of losing small fragments of information; however, even minor data losses may lead to substantial damage to professional, institutional, or strategic information resources.

The significance of digital archiving cannot be overstated or underestimated. The following key

considerations further emphasize the importance of electronic archiving.

1. *Preservation of the Digital Environment.* When conducting searches on the Internet, it is possible to access information from previous years, which is made available as a result of digital archiving processes.

2. *Data Integrity.* Digital archiving systems are characterized by robust and resilient mechanisms designed to prevent data loss, distortion, or unauthorized modification. Therefore, ensuring data integrity primarily involves safeguarding the accuracy, completeness, and reliability of information throughout its lifecycle.

3. *Efficient Retrieval.* Owing to the digitization of files, information retrieval has become significantly more efficient than ever before. This capability enhances productivity and supports informed decision-making processes.

The Evolution of Archival Practice and the Implementation Mechanisms of Digital Archiving.

It is well known that information may exist in various forms and formats. As long as data exist, there will always be a need for archiving. In the modern digital world, electronic archiving has gained significant importance. Both personal and business records play a crucial role in the storage and preservation of digital information. To fully understand the evolution of data storage, it is essential to examine its fundamental components and operational mechanisms in depth.

Throughout history, the preservation of important information and documents has been one of humanity's fundamental needs. From parchments, pebbles, and clay tablets to paper-based media, this developmental trajectory ultimately led to a major transformation that resulted in the global expansion of electronic archiving. Electronic archiving represents one of the key contributions of the modern digitalization process. The gradual transfer of reality into virtual environments, the emergence of cryptocurrencies as new forms of currency, and the prospective development of Web 3.0 and metaverse concepts are expected to further strengthen the governing role of the digital world.

At the same time, the replacement of paper documents with electronic archiving has significantly facilitated the storage, security, and long-term preservation of critical information. Understanding the mechanisms underlying digital archives and their operating principles is essential; consequently, trusting and effectively utilizing digital archiving systems without concern for data security becomes both feasible and advantageous.

The primary significance of digital archives lies in ensuring the accessibility of digital information, which can be achieved through specialized software solutions and modernized storage infrastructures.

Digital archiving involves the implementation of the following stages:

- The first step is to ensure the accurate collection of data and its categorization for improved usability. The extraction of metadata and relevant information from datasets is particularly important, as it enhances searchability and accessibility.
- The next step is data storage. All necessary security measures must be implemented to preserve data integrity, including the use of advanced algorithms. This stage also involves the regular verification of data completeness and consistency.
- To ensure security validation, stored data are encrypted during transmission from the sender to the recipient and subsequently decrypted upon receipt. Such encryption mechanisms allow only authorized users (administrators with appropriate privileges) to modify stored data when necessary.
- To prevent data loss, it is essential to continuously monitor stored information and periodically migrate it to upgraded and reliable storage systems.

Core Components of Digital Archiving. In many cases, data loss may lead to undesirable consequences. Data loss remains a critical issue that must be carefully considered even in the era of advanced technologies. Regardless of how intelligent or sophisticated technology becomes, it remains vulnerable to threats such as data corruption or manipulation, hardware failure, cyberattacks, and accidental deletion. Digital archiving helps mitigate these risks while ensuring the reliability and sustainability of information. Furthermore, through regular monitoring and system adjustments, archiving processes can be strengthened and updated in accordance with evolving requirements.

In this regard, the following critical components of digital archiving should be emphasized.

1. *Data Repository* – ensures the storage and management of digitized information.

2. *Archiving Software* – enables the effective administration and control of the archiving process.

3. *Metadata* – facilitates efficient indexing and retrieval following the archiving process.

4. *Access and Security Controls* – prevent unauthorized users from accessing archived information.

5. *Data Integrity Mechanisms* – utilize algorithms and validation procedures to prevent data corruption and unauthorized modification.

6. *Compliance Verification* – ensures adherence to legal and regulatory requirements.

7. *Data Migration Strategy* – guarantees the updating of data formats and storage environments in response to evolving technologies.

From specialized archiving software to the transmission of data between two devices, each of the listed components plays a critical role.

Data Preservation. Over time, the preservation of information in digital form has become a matter of fundamental importance. It ensures the continuity of daily life and enables organizations to operate without interruption while safeguarding their institutional heritage. The following considerations highlight the critical importance of data preservation.

As time progresses, the essential role of data in structuring everyday life becomes increasingly evident. Nevertheless, it is important to recognize that information in any form remains vulnerable. Hardware failures, software errors, cyberattacks, and even natural disasters may pose serious risks to data security. Proper data storage not only protects information against such threats but also ensures its integrity and immutability.

Data integrity constitutes the cornerstone of reliability and accuracy. It is crucial that information be preserved in its original form over time. Data preservation mechanisms are designed to prevent corruption, unauthorized access, and unintended modification. Ensuring integrity also involves controlling access to data. Only authorized users or processes should be permitted to access, modify, or delete information. Access control mechanisms—such as user authentication and authorization—play a vital role in maintaining data integrity. This is directly related to the reliability and security of original records. At the same time, preservation mechanisms must guarantee that the origin and source of data remain authentic and unaltered, as original data are considered accurate and trustworthy.

Data are preserved primarily to ensure ease of retrieval. However, if preservation methods become obsolete or access becomes difficult, the accessibility and searchability of information may be significantly reduced. Therefore, preservation

frameworks must provide a balance between robust protection and operational accessibility, enabling effective retrieval and reliable use of data.

In this context, efficiency and search capability are closely linked to metadata management. Metadata facilitates indexing and ensures that search processes are both fast and accurate. With properly structured indexing systems, users can easily locate specific information within digital archives.

Types of Digital Archiving. Digital archiving consists of a set of interrelated approaches, each designed to address specific objectives and requirements. Some of the main types of digital archiving include the following.

- *Institutional Archiving* – involves the digitization of cultural and scientific heritage

materials, such as books, manuscripts, and historical documents.

- *Email Archiving* – entails the preservation of email communications for compliance and legal purposes. It ensures that each message is securely stored, searchable, and accessible when required.
- *Digital Media Archiving* – requires the archiving of digital content beyond textual data, including audio, video, and image files.
- *Web Archiving* – focuses on preserving websites, blogs, and social media content, taking into account evolving digital trends, and contributing to the systematic preservation of online data environments.

Tables 1 and 2 present the corresponding types of digital archiving classified according to active (passive) and cloud-based (local) characteristics.

Table 1. Active and Passive Digital Archiving

Type of archiving	Description	Advantage
Active	Frequently accessed or regularly updated data are managed through active archiving. This approach enables fast, real-time access to information, supports frequent modifications, and ensures operational efficiency through the use of high-performance storage solutions.	Accessible and up-to-date information, efficient data management, and real-time updates.
Passive	Data that are retained solely for compliance, legal, historical, or reference purposes and are not actively used are preserved within the framework of passive archiving. This approach focuses on the secure long-term storage of information and is intended for data that are accessed infrequently.	Efficient storage capacity, long-term preservation, and reduced management costs.

Table 2. Cloud-Based and Local Digital Archiving

Type of archiving	Description	Advantage
Cloud based	Cloud-Based Archiving – involves the storage and management of data on remote servers provided by an external cloud service provider.	Remote accessibility from any location, scalability, reduced storage costs, and automatic updates.
Local	Local Archiving – involves the digitization of data and assets within an organization’s own infrastructure and their storage on local storage devices.	Predictable costs, enhanced security, low latency, and full control over data.

Media Supporting the Storage of Archival Data. Archiving is not limited to textual content. Each supported media type has its own specific requirements for preservation and accessibility, and the use of digital archiving media provides an effective solution for accommodating each of these data types. The supported media types include the following:

- *Textual Data:* Textual data consist of reports, documents, emails, and other forms of text-based content and information. As technology increasingly assumed a dominant role in information environments, graphical data soon

became equally important alongside textual materials. Graphical data include images and video content.

- *Images:* Data in the form of photographs, illustrations, and scanned documents.
- *Videos:* Personal or professional video recordings.
- *Audio Data:* Includes music recordings and other audio or voice-based materials.

As *databases and datasets* have gained prominence, structured data have become another significant category of media types. Consequently:

- *Software code and applications* require the

preservation of historical versions of software systems;

- Archiving of *websites and web content* has become a key mechanism for preserving online information environments.

An analysis of the existing literature indicates that various methods and approaches have been developed in the field of digital archiving. However, these methods typically focus on isolated technical or functional aspects, and their

advantages and limitations are not systematically compared. This makes it difficult to identify optimal solutions for specific application domains. Therefore, a comparative and critical analysis of existing approaches is considered necessary.

Table 3 below provides a summarized overview of the principal methods used in digital archiving, presenting their characteristics, advantages, and limitations in a structured format.

Table 3. Comparative Analysis of Digital Archiving Methods

Method	Description	Advantages	Limitations	AI Integration Capabilities
Raditional DMS Systems	Conventional electronic document management	Simple structure, widespread adoption	Limited automation and constrained analytical capabilities	Low
Metadata-Based Archiving	Structured indexing	Efficient search capability	Risk of manual input errors	Medium
Cloud-Based Archiving	Storage on remote servers	Scalability and flexibility	Security vulnerabilities	Medium
AI-Based Archiving	Application of machine learning and NLP	Automation and high accuracy	High costs and ethical concerns	High
Blockchain-Based Archiving	Immutable registry system	Data integrity	Scalability challenges	Limited

The conducted comparative analysis demonstrates that most existing archiving methods possess limitations either in terms of technical capabilities or in relation to security and governance aspects. Traditional and metadata-based systems do not ensure a high level of automation, while cloud-based solutions introduce potential security risks. AI-based systems offer advanced functionality; however, they are accompanied by ethical and organizational challenges. Although blockchain technologies ensure data immutability, they are not yet sufficiently optimized for large-scale implementation. For this reason, the development of an integrated model based on the combination of existing approaches emerges as a relevant scientific problem.

The effective implementation of Artificial Intelligence in digital archival systems is grounded in the use of specific algorithmic models. One of the most widely applied approaches in contemporary archiving solutions is the use of Convolutional Neural Networks (CNNs). These models are primarily employed for the automatic recognition, classification, and indexing of image and video materials.

Another significant direction in deep learning involves Transformer-based models. These models demonstrate high efficiency in processing

sequential data and identifying semantic relationships, and they are particularly applicable in the analysis of large-scale textual archives.

In the field of Natural Language Processing (NLP), models such as BERT, GPT, and similar architectures expand the possibilities for automatic document summarization, semantic search organization, and contextualization of content. These approaches enable the generation of more precise and targeted responses to user queries.

Support Vector Machines (SVM) and Random Forest algorithms are also applied for the structuring and automatic clustering of large-scale archival datasets.

Thus, the targeted integration of various AI models significantly enhances the functionality, accuracy, and adaptability of digital archival systems.

4. Discussion

Future Trends in Digital Archiving. Digital archiving technology continues to evolve on a daily basis, similar to other advanced technologies such as the Internet of Things, Big Data Analytics (Rzayev, 2016), and Blockchain technologies. Regardless of the variety of hypotheses proposed, a

foreseeable list of trends has consistently remained within reach.

Currently, the role of Artificial Intelligence (AI) tools and methods in business and the global economy is widely and actively discussed among experts across various fields (Rzayev, 2012). This interest is not accidental, as AI has the potential to introduce fundamental—and in some cases unprecedented—changes in the way people work and interact with technology.

Although the development level of AI technologies has progressed far beyond its initial stages, the synergy and long-term benefits resulting from their widespread implementation have not yet been fully realized. Today, AI is deeply integrated into everyday life, contributing to enhanced security, increased productivity, and the creation of new transformation opportunities for consumers, business entities, and society as a whole in response to emerging challenges.

AI possesses significant potential to influence the development of the global economy. When combined with foundational technologies such as the Internet of Things, Big Data Analytics, and Blockchain, Artificial Intelligence can become a core driver of economic growth, strengthening competitiveness and generating new high-technology employment opportunities. Based on these prerequisites, studying the prospective impact of AI technologies on the digital economy becomes both relevant and essential.

In the short term, AI technologies may significantly accelerate the comprehensive development of the digital economy in the following five key areas:

1) *Computer vision*; 2) *Natural language technologies* – including software agents capable of solving problems and/or providing services based on commands given in natural language; 3) *Data analysis*; 4) *Process automation with a high degree of robotization* – as a form of workflow automation technologies based on bots (metaphorical software robots) or digital workers; 5) *Advanced machine learning*.

In the near future, AI will automate data analysis, data classification, and precise data storage processes. As a result, organizations will be better equipped to manage errors (including detection and correction) and ensure data security within digital environments.

According to future trends, Artificial Intelligence in electronic archives will be capable not only of automatically recognizing and classifying text-based documents, but also of processing visual and

multimedia content such as images, videos, and audio files. For example, when traveling to a city with your family, a digital archiving system will not merely store your photographs but will also automatically record contextual information such as location, date, environment, and related metadata.

From the perspective of preserving data integrity and reliability, Blockchain technology holds particular significance, especially in addressing the following issues:

Immutable Records. Blockchain technology is distinguished by its decentralized structure across peer-to-peer (p2p) ledger systems, creating tamper-resistant data registries for digital archival systems. These immutable records ensure data integrity, prevent human and system errors, and consequently enhance both organizational efficiency and trustworthiness.

Verified (Authentic) Digital Records. Blockchain technology remains a focal point for institutions and individuals seeking to verify digital records. It can be concluded that within archival systems supported by peer-to-peer (p2p) networks, altering stored information is virtually impossible. Therefore, verification mechanisms play a crucial role in maintaining the integrity and authenticity of digital records.

To ensure enhanced search and retrieval capabilities, AI-driven Natural Language Processing (NLP) tools are employed. NLP enables archiving systems to better understand, process, and respond to human commands by taking into account contemporary demand trends and user-provided contextual cues. As a result, digital archiving systems become more user-friendly and efficient.

Within the NLP framework, semantic search mechanisms provide a more intuitive and faster environment for users compared to simple keyword matching. The ability of archiving systems to interpret “keyword cues” in a contextual and meaningful manner represents a significant breakthrough in the field of digital archiving.

The application of AI methods also allows greater emphasis to be placed on security in digital archiving. For example, advanced encryption techniques can be implemented to enhance data protection. Once data integrity is ensured, their security must be further reinforced. Sophisticated mathematical algorithms help maintain data in an unreadable (encrypted) form for all parties except the sender and the authorized recipient. Advanced encryption methods function as a protective layer within archiving systems against cyberattacks, phishing attempts, and unauthorized access.

Additionally, AI-based multi-factor authentication technologies strengthen digital archiving systems by providing an extra layer of protection through confidential PINs, passwords, or biometric verification. This significantly reduces the risk of unauthorized access and protects archived data from cyber threats.

The use of Artificial Intelligence in digital archiving also generates two further significant advantages. First, modern archiving solutions contribute to the reduction of electronic waste (e-waste) and support sustainable data management by optimizing energy consumption. Low-energy storage devices and environmentally sustainable data centers minimize the ecological impact of data preservation. At the same time, there is a growing emphasis on the use of renewable energy sources within archival infrastructures.

Second, through the implementation of disciplined governance mechanisms, organizations can regularly eliminate obsolete or redundant data from archival systems. This enables more efficient data volume management, prevents unnecessary resource consumption, and contributes to the development of a sustainable digital information environment.

Thus, one of the most powerful applications of AI in digital archiving is its ability to automate data contextualization. Contextualization refers to the process of linking data with metadata that provide meaning, such as the time of creation, source, and relevance to other related information.

AI-driven systems are capable of analyzing data at scale and automatically tagging documents with relevant metadata based on content. For healthcare organizations, for example, this means that complex datasets—such as clinical trial data or laboratory results—can be intelligently indexed, facilitating the tracing of data provenance and improving the understanding of their significance. AI algorithms can process large volumes of structured and unstructured data, identifying patterns, relationships, and trends that human operators might overlook. The key advantages include:

- *Improved accuracy:* Automated metadata tagging reduces the risk of human error and ensures more precise data representation;
- *Time efficiency:* By automating the process, AI reduces the time required to contextualize and archive large datasets.

In addition to enhancing search capabilities, Artificial Intelligence plays a crucial role in optimizing the information retrieval process.

Machine learning algorithms can be trained to predict which types of information are most relevant within a particular context. This enables proactive recommendations and significantly reduces search time. By analyzing user behavior and previous data queries, AI-based archiving systems can anticipate user needs and make data access simpler and more efficient.

This approach offers particular advantages during audits, inspections, and regulatory assessments, where rapid and accurate access to specific information is critical for demonstrating compliance.

5. Conclusion

Archivists are increasingly becoming custodians of large-scale datasets and complex data clusters. This development creates significant challenges, particularly in an environment where clearly established approaches for addressing emerging issues are still lacking. In general, we analyzed several emerging strands of research concerning the application of Artificial Intelligence technologies in archives. The present study offers one of the first comprehensive and distinctive reviews of such work. We also continued to explore approaches frequently encountered in current record-keeping practices that may have a substantial impact on the field.

Our analysis began with the question of how the integration of Artificial Intelligence reshapes the theoretical foundations of archival practice. We then examined newly applied AI techniques relevant to this domain. Our findings indicate that there is considerable interest within archival communities regarding the potential of AI technologies. From an organizational perspective, AI assists archives in moving beyond the limitations of traditional archival units while simultaneously preserving their core strengths—such as the principle of provenance. Archival provenance, for example, requires specific methods for its identification and preservation.

AI methods are also widely employed to generate new insights from digital collections, to enable innovative search forms, and to support distant reading techniques that facilitate direct engagement with emerging communities, including those within the digital humanities. Furthermore, we described the role of AI technologies in enabling new types of archives—such as social media collections or digital diaries—that expand existing archival concepts related to contemporary institutions. Due to their

scale, such collections pose distinct challenges in terms of organization and access. These challenges can only be addressed through further development of AI technologies and through the establishment of specialized access mechanisms, including commercial-type Application Programming Interfaces (APIs). Only in this way can archives fulfill their core mission—maintaining trust in records of the past.

Despite significant progress, further research and direct engagement with archival institutions remain necessary. We observe persistent gaps in transforming experimental AI applications into sustainable archival practices and infrastructures. Therefore, greater efforts are required to strengthen ethical frameworks and to enhance trust in AI methodologies by better understanding their impact on research and archival processes. At the same time, additional work is needed to update the theoretical foundations of archival practice in light of recent advancements in Artificial Intelligence.

References

- Abbasov, A.M., Rzayev, R.R. (2023). Artificial intelligence and digital economy: development prospects. *Lecture Notes in Networks and Systems*, 610: 147-153 DOI: 10.1007/978-3-031-25252-5_23
- Abbasov, A.M., Mamedova, M.H., Orujov, G.H., Aliyev, H.B. (2001). Synthesis of the methods of subjective knowledge representations in problems of fuzzy pattern recognition. *Mechatronics* 11(4): 439-449. [https://doi.org/10.1016/S09574158\(00\)00027-1](https://doi.org/10.1016/S09574158(00)00027-1)
- Angelova, L., Collomosse, J., Smillie, R., Cubey, R., Norman, J. (2020). Deep discoveries: Interim Report. Foundation Projects, Towards a National Collection, Arts and Humanities Research Council. <https://www.nationalcollection.org.uk/sites/default/files/2021-02/Deep%20Discoveries.pdf> (accessed 5 June 2025).
- Baron, J. R., Sayed, M. F., Oard, D. W. (2020). Providing more efficient access to government records: a use case involving application of machine learning to improve FOIA Review for the deliberative process privilege. *Journal on Computing and Cultural Heritage*, 15(1): 1–19. <https://doi.org/10.1145/3481045>
- Biegelbauer, P. et al. (2022). Engineering trust in AI: the impact of debates on AI regulation on the work of software developers, in 20th STS Conference Graz 2022, Critical Issues in Science, Technology and Society Studies.
- Cave, S., Coughlan, K., Dihal, K. (2019). Scary robots: Examining public responses to AI. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, Association for Computing Machinery Honolulu, HI, USA. <https://doi.org/10.1145/3306618.3314232>
- Corrado, E.M., Sandy, H.M. (2017). *Digital Preservation for Libraries, Archives, and Museums*. London: Rowman & Littlefield Publishers.
- Deegan, M., Tanner, S. (eds). (2006). *Digital Preservation*. London: Facet.
- Delve, J., Anderson, D. (eds) (2014). *Preserving Complex Digital Objects*. London: Facet.
- Fast, E., Horvitz, E. (2017). Long-term trends in the public perception of artificial intelligence. *AAAI (Association for the Advancement of Artificial Intelligence) Conference*, pp. 963–69.
- Hagendorff, T. (2020). The ethics of AI ethics: an evaluation of guidelines. *Minds and Machines*, 30(1): 99-120, doi:10.1007/s11023-020-09517-8.
- Harvey, R., Weatherburn, J. (2018). *Preserving Digital Materials*. Washington DC: Rowman & Littlefield
- Jaillant, L. (2022a). How can we make born-digital and digitised archives more accessible? Identifying obstacles and solutions. *Archival Science*. <https://doi.org/10.1007/s10502-022-09390-7>
- Jaillant, L. (ed.) (2022b). *Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitized Archival Collections*. Bielefeld, Germany: Transcript.
- Jaillant, L., Caputo, A. (2022). Unlocking digital archives: cross-disciplinary perspectives on AI and born-digital data, *AI & Society*. <https://doi.org/10.1007/s00146-021-01367-x>
- Jobin, A., Ienca, M., Vayena, E. (2019). Artificial Intelligence: the global landscape of ethics guidelines. *Nature Machine Intelligence*, 1: 389–99. <https://doi.org/10.1038/s42256-019-0088-2>
- Kelley, P. G., Yang, Y., Heldreth, C., et al. (2021). Exciting, useful, worrying, futuristic: public perception of artificial intelligence in 8 countries. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, Association for Computing Machinery. <https://doi.org/10.1145/3461702.3462605>
- Mcdonald, G., Macdonald, C., Ounis, I. (2020a). How the accuracy and confidence of sensitivity classification affects digital sensitivity review. *ACM Transactions on Information Systems (TOIS)*, 39(1): 1–34. <https://doi.org/10.1145/3417334>
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1: 501–07. <https://doi.org/10.1038/s42256-019-0114-4>
- Nix, A., Decker, S. (2021) Using digital sources: the future of business history? *Business History*, 1–24. <https://doi.org/10.1080/00076791.2021.1909572>
- Rzayev, R.R. (2012). *Neuro-Fuzzy Modeling of Economic Behavior*. Lambert Academic Publishing, Saarbrücken (in Russian).
- Rzayev, R.R. (2016). *Analytical Decision Support in Organizational Systems*. Palmerium Academic Publishing, Saarbruchen (in Russian).
- The National Archives UK. (2016). The application of technology-assisted review to born-digital records transfer, inquiries and beyond. www.nationalarchives.gov.uk/documents/technology-assisted-review-to-born-digital-records-transfer.pdf (accessed 5 June 2025).

How to cite: Elchin Aliyev, Islam Ahmadov, Khanmurad Abdullayev (2026). Application of artificial intelligence in digital archiving: problems, solutions and prospects. *Problems of Information Society*, 1, 24–34. <https://doi.org/10.25045/jpis.v17.i1.03>