

Article

Zoë (Abbie) Teel*

Artificial Intelligence's Role in Digitally Preserving Historic Archives

<https://doi.org/10.1515/pdtc-2023-0050>

Received December 26, 2023; accepted January 31, 2024;

published online February 27, 2024

Abstract: The term “Artificial Intelligence” (AI) is increasingly permeating public consciousness as it has gained more popularity in recent years, especially within the landscape of academia and libraries. AI in libraries has been a trending subject of interest for some time, as within the library there are numerous departments that serve a role in collectively contributing to the library’s mission. Consequently, it is imperative to consider AI’s influence on the digital preservation of historic documents. This paper delves into the historical evolution of preservation methods driven by technological advancements as, throughout history, libraries, archives, and museums have grappled with the challenge of preserving historical collections, while many of the traditional preservation methods are costly and involve a lot of manual (human) effort. AI being the catalyst for transformation could change this reality and perhaps redefine the process of preservation; thus, this paper explores the emerging trend of incorporating AI technology into preservation practices and provides predictions regarding the transformative role of Artificial Intelligence in preservation for the future. With that in mind, this paper addresses the following questions: could AI be what changes or creates a paradigm shift in how preservation is done?; and could it be the thing that will change the way history is safeguarded?

Keywords: Artificial Intelligence (AI); libraries; digital preservation; historical collections; technological advancements

1 Introduction

The significance of delving into the intersection of Artificial Intelligence and the digital preservation of historical records

lies in its dual importance: safeguarding cultural heritage and recognizing the evolving landscape of archives/preservation departments. It underscores the necessity for these institutions to adapt and leverage emerging technologies to fulfill their crucial roles in preserving the past for the benefit of the present and future generations.

How exactly would AI make preserving documents digitally easier? Well, consider the binary system or the binary code – the famous 1s and 0s. Digital objects, 1s and 0s, are what underlie most computer systems – they are the information “bits.” These components, or “bits,” serve as the building blocks for conveying various forms of information, encompassing text, numbers, audio, images, and more. It is a language and this language handles the storage, processing, and transmission of digital information; in this way, it is an “important element in modern technology, being employed as a means of calculation in digital computers and as a means of controlling a considerable variety of machine tools” (Heath 1972, 6). Binary code can look like a structure that stores data in a hierarchical manner (AI for Anyone n.d.), and AI makes this data easier to search and retrieve, as well as to update and delete (AI for Anyone n.d.). So, why does this matter when it comes to preserving historical records?

Essentially, it has been established that AI makes the digitization process more efficient because it makes processes quicker and easier. But why does it need to be quicker and easier, when it has been such a human-manual process for years? Consider The National Library of Norway, who have been entrusted with the monumental task of digitizing all printed materials published in Norway since 2006 (Takle 2009, 2). This mission is undeniably enormous, to say the least. So, how can AI contribute to making this task more manageable? AI possesses the capability to analyze concepts and apply them effectively, which is good for metadata processes in creation of preservation: “morphological analysis, part-of-speech tagging named entity recognition, word sense disambiguation, machine learning, text processing and e–governance services, etc.” (Das et al. 2022, 93). These automated various tasks could help preserve historical materials, by simplifying and expediting the digitation process. This, in turn, could help ensure the longevity of

*Corresponding author: Zoë (Abbie) Teel, College of Information-Student, University of North Texas, 3940 N Elm St, 76203-1277 Denton, TX, USA, E-mail: abigailteel@gmail.com. <https://orcid.org/0009-0007-0569-6114>

these collections and prevent them from physical damage (because they would be digitized). Therefore, technology not only aids in the restoration of lost heritage but also holds the potential to recreate it in remarkable ways (Das et al. 2022, 95). With this in mind, the approach behind this paper is to explore the synergy of Artificial Intelligence and archival research, examining how these two domains interact.

Preservation came about as a result of wanting to help keep materials that were deteriorating, for example, the Brittle Book Problem in the 1950s and the 1960s (NEDCC n.d.). However, a lot of preservation is deciding what to preserve: “to assess the subjects’ strengths and weaknesses of any particular library’s collection” (NEDCC n.d.). Utilizing AI assistance in making such decisions, while acknowledging the necessity of streamlining, can aid in retaining materials, specifically, the decision of which materials to preserve – the materials that might otherwise be solely dismissed based on the library or museum’s physical storage space.

Using Artificial Intelligence for preservation and digitization is not a futuristic concept. Often, people react defensively or skeptically when they hear terms like Artificial Intelligence or machine learning. They may associate these technologies with dystopian visions from movies like *iRobot*. However, the aim of this paper is to inspire and promote the use of AI for preservation and digitization.

The purpose of embracing the usage of Artificial Intelligence in archives and preservation is not to advocate for the complete human replacement in the preservation process, but rather to advocate for utilizing these tools to assist the professional by improving performance on certain tasks and allowing the professional to allocate more of their valuable time to tasks that cannot be effectively performed by existing technology. Most of the issues revolving around preservation have to do with cost-effectiveness and proper staffing, and AI can improve cost-effectiveness and reduce strain. In other words, it could mean more time with the documents and less time with the system.

2 Literature Review

The significance of AI in historic preservation becomes evident when reflecting on historical catastrophes, such as the 1966 flood in Florence. This event marked a turning point for preservation and conservation efforts, prompting global attention and collective action to salvage cultural artifacts (Gerbracht and Baden-Württemberg 2020, 200). AI extends the concept of salvaging beyond physical preservation, offering the potential to create digital duplicates enriched with various elements from the original, such as written text, seals, symbols, and features specific to the first copy or make of an item.

Recent advancements in technology, specifically within digitizing, have significantly eased the challenging task of preserving cultural heritage. As Goudaroli, Sexton, and Sheridan (2019, n.p.) state, “the arrival of new information and communication technologies radically alters the ways in which people create, access, and engage with information and archival records.” Technology has continued to advance over the past few decades at an unprecedented rate; innovations available today would have been unheard of only a short time ago. Artificial Intelligence (AI), as the future architect, has now emerged as a pivotal force in safeguarding our unique cultural legacy for generations to come (Ibaraki 2019). Recognizing the potential of AI in digital preservation prompts an exploration of its relevance in historic preservation. While Ibaraki’s article dates back to 2019, the past year has witnessed a remarkable surge in the integration of AI into our daily lives; notable examples include the widespread adoption of technologies such as ChatGPT, face ID for phones, digital voice assistants, and smart home devices. Therefore, it can be inferred that the integration of AI into various domains, such as preservation, is expanding.

One intriguing application of AI in digital preservation involves the use of a classical model for handwriting recognition referred to as a “CRNN (Convolutional Recurrent Neural Network) with image line-level data” (Ferro et al. 2023 560). This innovative approach employs image extraction capabilities coupled with sequence modeling, and the model not only recognizes handwriting but also captures crucial features and context within the input data (Ferro et al. 2023). Consequently, it transforms handwritten content into a more readable format, facilitating easier comprehension and conversion through a searchable and accessible form. Imagine the world of opportunity that opens when old, even ancient text is able to be scanned, processed, and searched – quickly – as one would do with a regular research inquiry.

Additionally, take, for instance, the preservation efforts for the Great Wall of China – a formidable undertaking for architects and historians. Construction began, on the wall, in c. 220 B.C. and the wall stretches around 20,000 km long (UNESCO 2018), with some sections of the wall inaccessible, making manual examinations laborious (Ibaraki 2019). To address this, Intel collaborated with the China Foundation for Cultural Heritage Conservation, employing cutting-edge drone technology, with thousands of photos collected and analyzed using AI to identify precise areas requiring restoration (Ibaraki 2019). Alyson Griffin, Intel’s Vice President of global brand and thought leadership marketing, emphasizes that this approach enables quicker, more efficient, and cost-effective restoration by providing accurate information

(Ibaraki 2019). The technology powering the drone, the AI, stands as a testament to the remarkable manifestations of cutting-edge technology at work in terms of aiding in preservation processes.

It would also be remiss not to acknowledge the transformative impact of digital preservation on storage paradigms, even before the era of AI and now with AI and its ongoing influence. Colavizza et al. (2021, 2, 4) emphasize that this evolution commenced well before the advent of big data, as archives quantified their extensive collections in physical terms, spanning kilometers of files and folders, with the choice of storage medium inherently linked to size considerations, with physical collections dictating the allocation of tangible space. In contrast, the integration of digitization has introduced new considerations for digital collections, such as cloud storage and gigabyte capacity. The global emphasis on extensive digitization efforts has transformed substantial portions of archives into digital data, presenting a potential area where AI could offer assistance in the future. Consequently, this shift has posed fresh challenges for archivists dealing with historical materials; Aangenendt (2022, 5) notes that archivists are now tasked with finding innovative methods to collect and preserve the diverse array of born-digital media formats and content types that encapsulate contemporary information about society and public life.

Interestingly, amidst this landscape, major institutions are increasingly turning to AI for the management of older materials rather than focusing solely on born-digital media.

Aangenendt (2022) study, which involved interviews with professionals in the Swedish archival sector, revealed a noteworthy trend in this regard. The study found that the Swedish National Archives, the Stockholm City Archive, the Popular Movements' Archive in Uppsala, and the Centre for Business History are presently not actively exploring AI for born-digital archival material. Instead, there is a predominant focus on prioritizing the digitization and accessibility of older analogue materials (Aangenendt 2022, 35).

More recent work by Marchello et al. (2023) delved into the use of AI-driven robotics for cultural heritage digital preservation. The concept revolves around creating digital twins – 3D copies or scans – of cultural heritage entities. This approach not only enhances accessibility to a broader audience but also introduces the concept of self-maintenance. Digital twins can facilitate predictive maintenance of physical assets by continuously monitoring data from sensors, thus preventing costly downtime and repairs (Marchello et al. 2023, 995–996).

Drawing insights from Ibaraki's work in 2019 and the more recent findings by (Marchello et al. 2023), it becomes

evident that AI-driven tools are revolutionizing the preservation of historical artifacts. Yet, this marks only the beginning. Another dimension of historical preservation involves data that goes beyond models or robotics; instead, it entails the assignment of data to objects, items, pictures, etc., aiding preservationists in classifying their collections – also known as assigning metadata: “AI can support metadata creation for images by generating descriptions, titles, and keywords for digital collections in libraries” (Reiche 2023, n.p.). This offers a potential reduction in time and resource requirements for hands-on tasks, as, traditionally, assigning metadata to preserved content has been a time-consuming process. This not only enhances efficiency in the preservation workflow, but also allows preservationists to channel their efforts toward their other duties.

Cushing and Osti (2022) conducted a comprehensive study exploring the perspectives of international archives and digital preservation professionals on the impact of Artificial Intelligence (AI) on digital archive expertise and its likely future applications. The findings revealed a widespread expectation among participants that AI would play a central role in archival work. As emphasized in PD203, participants expressed the belief that AI would become an integral tool for librarians, information professionals, and archivists across various roles (Cushing and Osti 2022, 19).

However, the authors also highlighted some concerns raised by participants, including the potential for AI to introduce additional tasks and the need to carefully review AI-generated data before granting access to patrons, while questions also arose about how to seamlessly integrate AI into existing workloads (Cushing and Osti 2022, 19). Notably, the authors did not delve into specific benefits resulting from the implementation of AI in day-to-day practices but did acknowledge the perceived value of AI technology in enhancing archival processes.

Additionally, the study by Cushing and Osti (2022, n.p.) highlighted a prevailing theme of “optimism” among participants concerning the role of AI in the digitization of archives.

Thus, it is understood that AI can take many different forms and impact various aspects of digitizing historical archives. As the European Parliament (2023, 1) eloquently noted in the summary of their briefing, “the results [of AI] are both promising and surprising: reconstructing a piece of art, completing an unfinished composition of a great musician, identifying the author of an ancient text, or providing architectural details for a potential reconstruction of the Notre Dame de Paris cathedral would have seemed like science fiction just a few years ago.”

3 Discussion

The principle underlying digital preservation aligns with the preservation of access to digital heritage, ensuring its ongoing accessibility to the public (UNESCO 2003, para. 4). The main purpose for digitization in the realm of Artificial Intelligence (AI) is apparent – the facilitation of ease and increased accessibility. However, the discussion also opens avenues for exploring the inherent complexities in preserving digital artifacts, emphasizing the trade-offs between the value of originals and the imperative of prioritizing access.

It is crucial to acknowledge that duplicates, models, or copies of original items may never attain the exact same value or reverence as some historical artifacts. Nevertheless, the emphasis should shift towards prioritizing access to these items, as deterioration of originals is at times unavoidable. Consider the dilemma of entirely losing an item versus having access to an almost identical twin – which holds greater significance? This prompts a reconsideration of the traditional notions of preservation, urging a paradigm shift towards prioritizing accessibility over losing something completely.

The barriers to achieving seamless digital preservation are numerous, with technological obsolescence standing out as a prominent challenge. The rapid pace of technological evolution introduces a formidable barrier, questioning the sustainability of preservation efforts, and a comprehensive solution to coping with ever-evolving technologies remains elusive. The fundamental question surfaces: how do we determine what to preserve when faced with constant advancements?

Preservationists grapple with this question extensively, recognizing the subjective nature of significance. What holds importance for one individual, based on personal experiences, nostalgia, or cultural relevance, may not carry the same weight for another. However, the introduction of AI into the preservation landscape introduces a potential avenue for overcoming this barrier, as AI's ability to optimize storage, discern patterns, and facilitate decision-making could potentially alleviate the burden of subjective choices. Could AI play a role in preserving more, while necessitating fewer decisions on what to discard?

As Fisher (2017) notes, “another set of barriers to digital preservation, and perhaps the most frequently discussed, relates to resource limitations.” Resource constraints further compound the challenges associated with digital preservation, necessitating a judicious approach to allocation. Integrating AI could potentially address resource limitations by streamlining processes.

Despite the existing barriers at both the individual and broader professional levels, opportunities tend to prevail, often overshadowing these challenges. As previously outlined – resource limitations, the dilemma of what to retain, duplicative value, and the quick pace of advancements, among others – these barriers also harbor a positive side. As articulated by Das, Maringati, and Dash (2022, 1), “Technology not only helps to restore the lost heritage but also does wonders in recreating it.”

The clear objective is to leverage AI for improving accessibility to cultural heritage that may be at risk of loss, a concern shared by numerous preservationists. Barlindhaug (2022, 12) underscores this goal in reference to the use of AI to archive a multitude of documents in Norway's National Library, expressing the aspiration that it will provide future generations with an unbiased collection of historical material that is readily available for further research.

4 Recommendation

After reviewing the research, individuals involved in preservation, archiving, or digitization should undertake their own investigations to understand how this emerging technology could enhance their work. The findings strongly indicate that Artificial Intelligence can play a significant role in preserving historical materials. However, its relevance to specific assignments ultimately depends on the preferences of the individuals utilizing it.

When contemplating the possible incorporation of AI into preservation and archiving procedures, it is crucial to carefully assess the ethical implications. AI, a frequently debated subject, carries both advantageous and detrimental consequences, particularly concerning issues such as bias and privacy (Lund and Wang 2023), and it is imperative for archivists to be mindful of these potential challenges when deciding on the application of AI in their processes. Further research into the critical topic of AI in the preservation of materials is warranted as, presently, very little literature exists on this topic. Archival researchers should leverage their expertise to investigate these issues and create a body of literature needed to support practitioners in decision-making surrounding AI for preservation purposes.

A suggested initial move in incorporating AI into the preservation process could be to task it with assigning metadata at a smaller scale, such as year, relative condition, type, and so on. This approach would allow professionals in the field to acclimate to the workings of this technology, experiencing its benefits and assistance in a reversible, less permanent manner.

5 Conclusions

The integration of Artificial Intelligence technologies into the archival and preservation process presents both challenges and opportunities. This paper has highlighted a clear desire within the profession for AI assistance, with professionals expressing optimism about its potential benefits. On the flip side, as is customary with any emerging technology, there is a sense of anticipation regarding how it will unfold and the potential influence it might have on the established processes within the field.

Many individuals, particularly those employed in the work of history and preservation, possess a natural curiosity about new technologies. This inherent inquisitiveness should be extended to embrace AI, recognizing it not as a barrier but as a significant opportunity to improve the quality and efficiency of archival work. While it is widely acknowledged that challenges exist in the realm of AI, especially during these early phases of development in generative Artificial Intelligence technology, it also holds the potential to serve as a valuable aid in safeguarding history and supporting the efforts of preservationists and archivists alike.

References

- AI for Anyone. n.d. Binary Tree. "AI for Anyone." <https://www.aiforanyone.org/glossary/binarytree#:~:text=Binary%20trees%20are%20a%20data,easy%20to%20update%20and%20delete> (accessed February 8, 2024).
- Aangenendt, G. 2022. "Archives in the Digital Age. The Use of AI and Machine Learning in the Swedish Archival Sector." *Young* 2761: 2765.
- Barlindhaug, G. 2022. "Artificial Intelligence and the Preservation of Historic Documents." *DOCAM* 9 (2). <https://doi.org/10.35492/docam/9/2/9>.
- Colavizza, G., T. Blanke, C. Jeurgens, and J. Noordegraaf. 2021. "Archives and AI: An Overview of Current Debates and Future Perspectives." *ACM Journal on Computing and Cultural Heritage (JOCCH)* 15 (1): 1–15.
- Cushing, A.L., and G. Osti. 2022. "So How Do We Balance All of These Needs?: How the Concept of AI Technology Impacts Digital Archival Expertise." *Journal of Documentation* 79 (7): 12–29.
- Das, B., H.B. Maringanti, and N.S. Dash. 2022. "Role of Artificial Intelligence in Preservation of Culture and Heritage." In *Digitalization Of Culture through Technology*, Vol. 92. Taylor & Francis.
- European Parliament. 2023. "Artificial Intelligence in the Context of Cultural Heritage and Museums: Complex Challenges and New Opportunities." *Think Tank*. [https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI\(2023\)747120](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)747120) (accessed February 8, 2024).
- Ferro, S., M. Pelillo, and A. Traviglia. 2023. "AI-Assisted Digitalisation of Historical Documents." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 48: 557–62.
- Fisher, K. 2017. "Barriers to Digital Preservation in Special Collections Departments." *Preservation, Digital Technology & Culture* 45 (4): 180–5.
- Gerbracht, J., and L. Baden-Württemberg. 2020. "Flood in Florence, 1966: A Fifty-Year Retrospective." *American Archivist* 83 (1): 199–201.
- Goudarouli, E., A. Sexton, and J. Sheridan. 2019. "The Challenge of the Digital and the Future Archive: Through the Lens of the National Archives UK." *Philosophy & Technology* 32: 173–83.
- Heath, F.G. 1972. "Origins of the Binary Code." *Scientific American* 227 (2): 76–83.
- Ibaraki, S. 2019. "Artificial Intelligence for Good: Preserving Our Cultural Heritage." *Forbes*, March 28. <https://www.forbes.com/sites/cognitiveworld/2019/03/28/artificial-intelligence-for-good-preserving-our-cultural-heritage/> (accessed February 8, 2024).
- Lund, B.D., and T. Wang. 2023. "Chatting about ChatGPT: How May AI and GPT Impact Academia and Libraries?" *Library Hi Tech News* 40 (3): 26–9.
- Marchello, G., R. Giovanelli, E. Fontana, F. Cannella, and A. Traviglia. 2023. *Cultural Heritage Digital Preservation through Ai-Driven Robotics: Copernicus GmbH*.
- NEDCC. n.d. "Introduction to Preservation." *Northeast Document Conservation Center*. <https://www.nedcc.org/preservation101/session-1/1what-is-preservation#:~:text=Preservation%20involves%20keeping%20a%20balance,are%20of%20te%20n%20more%20easily%20understood> (accessed February 8, 2024).
- Reiche, I. 2023. "The Viability of Using an Open Source Locally Hosted AI for Creating Metadata in Digital Image Collections." *Code4Lib Journal* 56.
- Takle, M. 2009. "The Norwegian National Digital." *Ariadne* 60.
- UNESCO. 2003. "Charter on the Preservation of Digital Heritage." <https://www.unesco.org/en/legal-affairs/charter-preservation-digital-heritage> (accessed February 8, 2024).
- UNESCO. 2018. "The Great Wall." *World Heritage Convention*. <https://whc.unesco.org/en/list/438/> (accessed February 8, 2024).