ITJ | The International Tax Journal

Digitization of the Arabic manuscript in the light of artificial intelligence technologies-from preservation to investment

Rokia KERKOURI [1]

[1] Hassiba Benbouali University of Chlef, Email: moualimakarkouri@hotmail.com

Abstract---This study aims to highlight the role of digitization and artificial intelligence technologies in the maintenance and preservation of the Arabic manuscript, and to move it from just a preserved heritage vessel to a knowledge resource amenable to scientific and cultural investment. It seeks to show how modern digital technologies have contributed to the transformation of Arabic manuscripts into indexable materials, search, and automated analysis, which opens up new horizons for their reuse in scientific research and inter-related studies. The study also highlights the most prominent challenges that hinder this transformation, such as weak digital infrastructure, lack of specialized competencies, and the dispersion of manuscripts between different institutions and groups. It proceeds from a central problem: How can digitization technologies and artificial intelligence go beyond the logic of traditional preservation, and contribute to the investment of the Arabic manuscript as an active element in contemporary knowledge. The study concludes that the adoption of clear institutional policies, the promotion of specialized training programs, and the use of artificial intelligence tools in the digital processing of manuscripts are key pillars to ensure the sustainability of this heritage and achieve a real transition from conservation to knowledge investment.

Keywords---manuscript, artificial intelligence, indexing, technology, heritage.

Introduction

Arabic manuscripts constitute an original knowledge asset in the Islamic civilization.over the centuries, they have retained the product of scientists in the fields of religion, language, medicine, philosophy, astronomy and other sciences. despite this rich asset, a large part of it is still threatened with damage

How to Cite:
KERKOURI, R. (2026). Digitization of the Arabic manuscript in the light of artificial intelligence technologies-from preservation to investment. *The International Tax Journal*, *53*(1), 13–19. Retrieved from https://internationaltaxjournal.online/index.php/itj/article/view/491

The International tax journal ISSN: 0097-7314 E-ISSN: 3066-2370 © 2026

Submitted: 30 March 2025 | Revised: 10 September 2025 | Accepted: 30 December 2025

and loss due to natural factors and the shortcomings of preservation, cataloging and documentation methods. with the accelerated digital transformation globally, digitization projects have emerged as a strategic option to sustain this heritage and turn it into a scientific material accessible for research investment.

Based on this reality, this intervention aims to explore the possibilities offered by digitization, using artificial intelligence techniques, in transforming the Arabic manuscript from a static image to a living knowledge entity that can be indexed, analyzed and linked to other sources, thus contributing to the resurrection of heritage and its reintegration into contemporary knowledge systems.

Importance of the topic:
The importance of this topic stems from the fact that it deals with one of the most precious elements of the cultural and cognitive identity of the Arab and Islamic nation, which is the written heritage that forms the basis vessel for its science and thought through the centuries. With the tremendous technological acceleration, digitization and artificial intelligence have become crucial elements in protecting this heritage from damage and loss, and in transforming it from static material into a living knowledge resource capable of research, analysis and wide dissemination. The importance of the topic is also reflected in its ability to highlight the gap between the available capabilities and the field challenges facing digitization projects in the Arab world, and to provide scientific insights and practical proposals to ensure the sustainability of the efforts exerted.addressing this topic contributes to saving the Arabic manuscript from neglect, and opens new doors for its investment in scientific research, historical studies, and cultural innovation, in a way that enhances the presence of heritage in the individual and institutional consciousness, and enables it to be transmitted to future generations in modern and effective ways.

The problematic
To what extent can the digitization of the Arabic manuscript, in the light of modern digital technologies and smart algorithms, overcome the problems of traditional preservation and cataloguing, and open new horizons for investing the manuscript heritage cognitively and culturally

Premise
The study is based on the premise that digitization of the Arabic manuscript is no longer a secondary technical option, but is a cultural necessity that allows manuscripts to be converted from paper materials that are subject to damage to organized knowledge sources that can be indexed and analyzed, provided that there is accurate scientific planning and technical understanding of the structure of the manuscript and its indexing controls.

Curriculum of study:
The study adopted the descriptive analytical approach to monitor the reality of digitizing Arabic manuscripts, analyze the challenges and possibilities, and then provide practical proposals and future scenarios to support these initiatives.

Preface:
In Algerian corners and on the shelves of ancient libraries, the treasures of our civilization that remained trapped in the pages for a long time are hidden, until the digital revolution came to give it a new opportunity for rebirth, and reconnect our present with our bright scientific past, thousands of manuscripts that carry the history of a nation and its knowledge are still unknown, subject to damage and loss, and often the civilization of a nation disappears when its written heritage is lost in the stomachs of books without the benefit of researchers and the new generation. Hence, the investment of modern technologies has become an urgent necessity to revive these treasures and ensure their transmission to future generations.

Digitization and artificial intelligence today play a pivotal role in bringing the Arabic manuscript back to life, not only as a means of preservation, but as tools capable of transforming the manuscript into a living knowledge resource capable of analysis, research, scientific and cultural investment. Through this intervention, we will address the reality of digitization in the Arab world, highlight the most important challenges and possibilities, in order to present practical proposals to ensure the sustainability of this heritage and its integration into contemporary knowledge systems.

1. conceptual framework :

1.1. Definition of the manuscript:

Due to the importance that manuscripts occupy in the revival of heritage, they have received wide attention in the methods of historical, educational and scientific research, and this interest has contributed to the push towards digitization, using modern technologies and artificial intelligence techniques, in order to preserve them, facilitate access to them, and invest them in the service of scientific research.

According to the dictionary, a manuscript is everything that is written by hand, whether it is a book or a document,which is " a book written in calligraphy, not in a printing press and collected by manuscripts 1, therefore, a manuscript is every handwritten book, unlike a printed book, and it carries intellectual, scientific, religious or literary productions in the pre-print era or the results of subsequent investigation.this manual skill gives it a cultural and historical value; it is not just a text carrier, but a historical document reflecting the identity of the Arab society at that time through knowledge of the conditions of copying, the quality of the paper used, the type of calligraphy, as well as ink, approved decoration, etc.

1.2. definition of artificial intelligence:

It is one of the branches of computer, and one of the main pillars on which the technology industry is based "relies on computer simulation of the qualities of human intelligence, such as the ability to think, learn or understand"[2]. artificial intelligence is not just digital tools or mechanical programs, but is an attempt to make a machine capable of simulating basic human mental abilities, allowing computer systems to perform functions that require human-like intelligence; what is meant is the ability of a machine to simulate the human mind through computer programs that are designed, where it refers to the ability of a computer or any other machine to "perform those activities that require intelligence, he is interested in developing machines and represent knowledge for use in making Inferences artificial intelligence can also be viewed as an attempt to model aspects of human thinking on computers"[3].

1.3. definition of indexing:

Indexing is an essential step in working on manuscripts, because it represents a scientific input for organizing data, and contributes to facilitating digitization and automated work on them, and the index: a book in which science chapters and others are collected"[4], this definition reveals that the concept of index has been associated since ancient times with the collection and arrangement of Sciences, which highlights the organizational nature of the term, and this meaning establishes the subsequent terminological understanding, where indexing remains a process of collection and arrangement, but is evolving today to become an input for building digital databases of manuscripts.

By definition, indexing is the process of describing documents in a systematic way that allows them to be easily identified and retrieved, according to specific scientific rules and standards."this concept 5 provides us with a precise methodological definition of the function of indexing, as a structured description and not just a narrative of information. This definition is fully consistent with the requirements of modern digitization, because digital systems rely mainly on unified scientific standards to retrieve documents easily and accurately, because indexing is not just a description of a manuscript, but a "scientific analysis of its structure, content, and physical characteristics, allowing it to be built

cognitively within an information environment"6.this perception makes indexing a cognitive step, not just technical, which provides a rich database that allows artificial intelligence to deal with manuscripts as multidimensional cognitive entities, not just illustrated texts.

2-the relationship of indexing with the digitization of the manuscript

In one of its reports on manuscript heritage, the Arab educational, cultural and scientific organization (alecso) emphasizes that"digitization of manuscripts cannot succeed without high-quality indexing, because metadata is the basis on which digital databases are based"7.this institutional statement reflects the fundamental fact that digitization is not only a process of imaging, but it is the construction of a digital system managed by metadata. The absence of good indexing leads to Digital Chaos, while scientific description provides a solid basis for creating searchable and analyzable databases.

Micro indexing opens the way for the use of artificial intelligence technologies in automatic text recognition and classification[8]. this text links micro indexing with artificial intelligence applications, which shows that the quality of automated results, such as OCR or objective classification, is highly dependent on the quality of metadata. And here it becomes clear that indexing is not only an introductory step, but also part of the smart infrastructure of the digital manuscript.

3-indexing and its role in facilitating the digitization of the Arabic manuscript

Before digitizing manuscripts, it is important to index thoroughly. indexing is a systematic description and bibliographic process that includes several elements such as the title of the manuscript, the name of the author or copyist, the date of copying (if any), as well as the number of sheets or pages, the type of Font, a description of the condition of the copy (complete, incomplete, damaged), the place of preservation and for example, in Algeria, studies have indicated that accurate indexing was the basis for the project "Algerian treasury of heritage"[9], and is a necessary prelude to any Therefore, it can be said that indexing is the necessary preliminary stage before digitization, as it provides the structured data on which digitization software and artificial intelligence algorithms rely in:

- ✓ Optical Character Recognition (OCR)
- ✓ Content classification
- ✓ Linking between different versions
- ✓ Automatic indexing of text within databases.

Thus, cataloging contributes to the transition from the preservation of the manuscript to its digital operation, and from its documentation to its scientific research and investment.

4-digitization and artificial intelligence technologies

When micro-indexing is combined with digitization, that is, photographing and saving the manuscript in digital format and associating it with metadata — the manuscript turns into a digital cognitive resource, and with the development of artificial intelligence technologies specializing in Arabic font processing (OCR / HTR), it became possible to convert scanned pages into searchable texts or automatic indexing, with font processing, modulation, margins, motifs... This would make the manuscript more valuable for research and scientific accessibility.

The use of artificial intelligence technologies in digitizing and documenting historical sources represents a great opportunity to enhance historical research and studies, which contributes to the preservation of cultural heritage for future generations, and calls for the development of institutional policies and practices that support this trend and ensure its sustainability"[10]

5-the possibilities and opportunities offered by digitization and artificial intelligence

The use of digitization in Arabic manuscripts today has become one of the most important living means of preserving this heritage and ensuring its continuity across generations.a study conducted on manuscripts indicates that the publication of digital copies of Arabic manuscripts in digital environments has become a necessity to preserve heritage from loss, and to make manuscripts easily available to researchers and readers"[11] , therefore, digitization is no longer a luxury option, but a practical response to the challenges facing traditional preservation, which lays the foundation for comprehensive digital cataloguing projects.

Along with digital preservation, digitization is not limited to just scanning, but goes beyond that to making available advanced analysis and indexing capabilities. Artificial intelligence technologies have opened up wider horizons in transforming the manuscript from a static image into a cognitive entity that can be analyzed and used academically. As a study suggests:
"Artificial intelligence applications, such as image analysis, document classification, and data mining, enable systematic archiving and documentation of manuscripts"[12], which confirms that the integration of artificial intelligence and digitization is not only limited to the preservation of texts, but also makes manuscripts living knowledge resources usable for research on a larger scale, and allows linking them to other sources within an organized database.

Moreover, modern algorithms13, help improve the clarity of ancient texts, restore the content of damaged documents, and analyze them automatically, and this reflects the ability to resurrect the manuscript heritage, as ancient texts can be re-read and extract their knowledge even if the original copy is damaged, which enhances scientific research and allows the use of manuscripts in a contemporary way, so the combination of digitization and the capabilities provided by artificial intelligence:

- Indexable and analyzable
- Live knowledge resources
- Amenable to scientific publication and academic use

This opens up prospects for launching new research and cultural initiatives, and puts the Arab manuscript heritage as an effective partner in contemporary knowledge, not only for preservation, but for creativity and scientific investment.

6-the challenges facing the digitization of Arabic manuscripts

6.1. lack of technical infrastructure and equipment
Many conservation institutions in Algeria lack professional photographic equipment, secure storage technologies, and stable information networks. Some studies have also indicated that effective digitization requires high processing investments, without this structure, digital images become vulnerable to damage or loss during storage or transportation, which reduces the value of digitization and weakens sustainable accessibility.

6.2. lack of specialized competencies in manuscripts and digital cataloging
Cataloging is a delicate process that requires knowledge of manuscript science (codecology), calligraphy, bibliographic classification and international standards for the characterization of manuscripts. In the analysis of the reality of digitization of manuscripts in Algeria it was pointed out that the lack of training of specialists is a major obstacle.Because the absence of these competencies leads to incomplete indexing, weak metadata, or errors in linking the digital image and data, which weakens the credibility of the digital manuscript and hinders its academic use.

6.3. scattered manuscripts and poor institutional coordination
Manuscripts are distributed among national libraries, corners, research centers, private vaults, often without a unified national database or a coordination policy. This dispersion hinders the construction of

a unified digital archive, leads to duplication of effort or loss of copies. Research on the digital libraries of manuscripts in Algeria has proved that this dispersion is one of the most prominent difficulties facing cataloguing and digital accessibility projects.

6.4. The complexity of Arabic calligraphy and the physical qualities of the manuscript

Even with the presence of digital images, the challenge remains in the automatic recognition of handwritten texts, especially in ancient manuscripts that include motifs, composition, margins, sockets. A recent study has shown that modern artificial intelligence models, such as" Qalam " achieve low error rates in OCR/HTR, but still have difficulty with texts associated with ancient manuscripts.

This means that digitization + indexing alone is not enough to make a manuscript a searchable and analyzable digital text; rather, specialized algorithms and training on large data sets are needed.

6.5. financing and continuity:

Digitization projects require large material investments: equipment, maintenance, storage, backup copies, updating, data management. Without sustainable funding, projects may stop after a short time, and libraries lose the effort they have monitored.

7-proposals and future visions for the revival of Arabic manuscripts

Based on the reality of Arabic and Algerian manuscripts in particular and the challenges facing their preservation and investment, several strategic paths can be proposed to effectively employ digitization and artificial intelligence:

7.1-establishment of a comprehensive digital national archive of manuscripts:

The state, in cooperation with universities and research centers, can create a central database of Arabic manuscripts, including high-quality digital copies with accurate descriptive information, so that they are available to researchers, students and academics, because this archive contributes to unifying research efforts and avoiding the repetition of similar projects at the local and regional level[14].

7.2-the use of artificial intelligence in indexing and analysis:

Tools should be developed based on artificial intelligence techniques to automatically recognize, classify and analyze ancient texts, allowing a deeper understanding of the content of manuscripts and reconstruct damaged texts, for example, deep learning algorithms can be used to improve the clarity of symbols, extract cognitive entities (names, dates, topics), making manuscripts searchable and analyzable in accurate and flexible ways.

7.3-preparation of training programs for specialists and researchers:

The success of any digitization project depends primarily on the presence of specialists capable of handling digital manuscripts and applying artificial intelligence tools. Therefore, training programs should be developed for those interested in manuscript heritage, including digital scanning skills, documentation, cataloging, and artificial intelligence techniques in analyzing historical texts4. Digitization is not just converting paper copies into digital formats, it is an integrated knowledge process that requires specialists in the field, as it is not enough to have hardware or software, but it is necessary to enable specialists with the necessary technical and knowledge skills to ensure the sustainability of the project and the quality of its outputs. This training makes digital manuscripts a living research resource that can be used academically and culturally on a large scale, as well as reduces technical and scientific errors that may arise when dealing with sensitive historical manuscripts, encouraging regional and international cooperation: opening channels of cooperation with international institutions specialized in digitizing manuscripts allows the exchange of experiences, access to modern technologies, and the exchange of databases, which contributes to strengthening the status of the Arabic manuscript in the global academic field

7.4. promoting regional and international cooperation:
The opening of cooperation channels with international institutions specialized in the digitization of manuscripts allows the exchange of experiences, access to modern technologies, and the exchange of databases, which contributes to strengthening the position of the Arabic manuscript in the global academic field.

7.5. development of open access programs for digital manuscripts :
Digital initiatives should allow researchers and students to access manuscripts in an open and secure manner, while ensuring the intellectual property rights of the custodians, thereby enhancing the possibility of scientific and cultural benefit[15].

---

[1] Ghazi Hussein Enaya, preparation of scientific research, Beirut, House of Galilee, 1992, p.95.
[2] Wafa Deif Allah, the impact of human resource diversity on artificial intelligence: an applied study, Department of Jordanian status or passports, Arab Journal for scientific publishing, issue 24, 2020, P.219.
[3] Nevin Farouk, the machine between natural intelligence and artificial intelligence, Journal of scientific research in the Arts, Faculty of Arts and Educational Sciences/ vol.03, No. 13, Ain Shams University, Egypt 2012, p. 481.
[4] Ibn Manzoor, the tongue of the Arabs, material: index, Sadr House, Beirut, 1997.
[5] Mohammed Al-Shaibani, Arabic indexing: its concepts and history, Dar Al-marefa, Rabat, 2014, P.28.
[6] Faisal Al-hafyan, introduction to Arabic manuscript science, King Faisal Research Center, Riyadh, 2016, P.15.
[7] Alecso, manuscript heritage conservation and digitization report, Tunisia, 2019, P.42.
[8] IFLA, Guidelines for Digitization Projects, 2017, p. 11
[9] https://www.google.com/search?q
[10] Naziha abulqasem Al-rajibi, modern practices in digitizing and documenting historical documentary sources: applications of artificial intelligence as a model., Asmariya University of Applied Sciences , 2024, P. 13.
[11] The role of digital media in the preservation and retrieval of Arabic manuscripts. (2024). iKNiTO Journal for Technology and Disability Studies, 6(2), 141 174. DOI: 10.21608/skje.2024.378117
[12]Al-Rajabi, Naziha Abu Al-Qasim. (2024). Modern practices in digitizing and documenting historical documentary sources: applications of artificial intelligence as a model. Asmari University Journal of sharia sciences, 37, special 5.
[13] **Algorithmics:** is a set of rules that express a specific sequence of operations that would include all computer programs, considered: the basics of algorithms, translation, localization and preparation of Mustafa Obaid
[14] Naziha abulqasem Al-rajibi, modern practices in digitizing and documenting historical documentary, P. 5.
[15] Ibid., p. 10