

Research Article

The Role of Artificial Intelligence in Restoring Ancient Texts and Analyzing Historical Documents

Ahmed Shaker Alalaq*

University of Kufa Iraq

Corresponding Author: Ahmed Shaker Alalaq, University of Kufa Iraq.**Received:** 📅 2024 Dec 27**Accepted:** 📅 2025 Jan 10**Published:** 📅 2025 Jan 30**Summary**

Anyone who closely follows the updates of the digital revolution and its most powerful outcome, artificial intelligence, will fully understand that no researcher, enthusiast, or beginner should ignore the implications, manifestations, and results of this truly remarkable revolution. The topic of "The Role of Artificial Intelligence in Restoring Ancient Texts and Analyzing Historical Documents" is a subset of the broader subject of how AI tools are modernizing the field of humanities. This specific area has aligned with and leveraged the technology and digitization revolution to create a new branch of the humanities, whose methodologies, nuances, and outcomes surpass all expectations. In our study, we will explore a series of practical steps and applied methodologies on the role of AI tools in historical studies, specifically focusing on analyzing historical documents, identifying their symbols and structure, and decoding their embedded codes, if we may say so, in an engaging and scientific manner.

Reasons for Choosing the Study Title: Artificial intelligence represents a true outcome of today's digital revolution, which has permeated almost all fields of work. Today, no one interested in research, education, or technological advancement can ignore this revolution. AI offers limitless potential for enhancing scientific research, including historical studies. When combined with digital technology, AI tools contribute to a new approach in studying the humanities, particularly history, surpassing traditional expectations with innovative methodologies and new perspectives.

Research Problem: Our study addresses the importance of technical tools in historical studies, such as deciphering cuneiform tablets, analyzing historical documents, understanding their structures, and identifying hidden codes. These tools are beneficial not only for professional researchers but also for amateurs and beginners, as they can be utilized at every stage of research, from planning to implementation.

Study Scope: This study is focused on establishing the connection between historical studies and artificial intelligence, a question we aim to answer. We will highlight key scientific strategic projects in this field and what has been achieved in practice.

Research Methodology: We employ a blend of descriptive, analytical, and inferential methods, examining various global scientific projects aimed at overcoming challenges faced by archaeologists and historians in decoding ancient symbols. We will explore the potential of AI tools in modernizing the extraction of historical texts.

Previous Studies: Our research relies on a selection of scientific studies that have previously explored the relationship between history and the digital revolution. We also focus on a range of projects and scientific experiments to examine the possibility of building a new scientific approach based on implementing and presenting historical digital data in line with modern digital advancements. Among the most notable of these studies are and other studies that showed the importance of using artificial intelligence and the digital revolution in historical studies.

Keywords: Artificial Intelligence, Digital Revolution, Historical Studies, Historical Texts, Documents, Symbols, Ancient Inscriptions

1. What is Artificial Intelligence, Its History and Origins

Historians suggest that utilizing modern computer science to understand ancient history enables them to connect distant threads of the historical record in ways previously inaccessible. This approach also helps correct distortions that arise from studying historical events in isolation. However, relying on these technologies carries risks, such

as the potential for machine learning algorithms to exhibit bias or even distort the historical record. This raises a fundamental question for historians and others who believe in understanding the present by studying the past: to what extent should we entrust the past to machines as their role in shaping our future increases?

1.1. The Early History of Digital Technology

The history of artificial intelligence dates back to ancient times, with myths, tales, and rumors about artificial beings endowed with intelligence or consciousness. In her latest book, *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology*, Stanford researcher Adrienne Mayor explores how ancient cultures imagined future technologies, leaving these imaginings behind in epics and sculptures. These accounts include stories of Ashoka fighting robots, as well as technological tales from ancient Hindu epics like the Ramayana and the Mahabharata, where deities like Vishwakarma and the sorceress Maya created humanoid robots. In Greek mythology, the god Hephaestus and craftsman Daedalus made the robot Talos, a mythical giant warrior programmed to guard the island of Crete, an example of machine learning and mechanical men [1]. Although no one knows if such robots were ever created, machine learning and artificial intelligence have long been topics of contemplation [2]. Some attribute the invention of the first computer to 1822, when Charles Babbage (1791–1871) developed his initial design for a computer, primarily paper-based, called the "difference engine" [3]. This idea of using computers and calculators grew significantly among scientists of that time. Initially, such machines were just dreams, before they were designed and transformed into reality [4]. As the mid-20th century approached, the integration of technology into human life began to take on a deeper significance, with many mediums focused on the idea of artificial humans. Scientists began to ask, "Is it possible to create an artificial brain?" Craftsmen, collaborating with scientists, even created some machines that we now call "robots," with the term being coined in a 1921 Czech play by writer Karel Čapek. Although most of these machines were relatively simple and steam-powered, some were capable of facial expressions and even walking [5].

In 1943, Warren McCulloch and Walter Pitts developed a model of artificial neurons, and in 1944, scientists J. Neumann and O. Morgenstern formulated decision theory, which provided a comprehensive framework for determining agent preferences. In 1949, Donald Hebb introduced a rule for changing the values of artificial neural connections, which allowed for learning [6]. That same year, computer scientist Edmund Callis Berkeley published a book titled *Giant Brains, or Machines That Think*, where he compared newer models of computers to human brains [5]. These theoretical advancements greatly influenced the course of artificial intelligence by providing the foundational concepts and frameworks that would be built upon in subsequent decades [7].

1.2. The Early Stages of Artificial Intelligence Technologies

Before 1949, computers lacked a fundamental requirement for intelligence—they could execute commands but could not store them. In other words, computers could be told what to do, but they could not remember what they had done. Additionally, computing was extremely expensive; in the early 1950s, renting a computer cost up to \$200,000 per month, a price only prestigious universities and major

tech companies could afford. To secure funding, notable advocates were needed to convince financial sources that artificial intelligence was worth pursuing [8]. In 1950, British mathematician Alan Turing published a seminal paper titled "Computing Machinery and Intelligence," speculating on the possibility of creating machines that could think. He proposed what is now known as the Turing Test, a method for determining whether a computer can think intelligently like a human [9,10]. Turing predicted that by the year 2000, a computer could "play the imitation game so well that the average interrogator would have no more than a 70% chance of making the correct identification (machine or human) after five minutes of questioning" [11]. The first functioning AI program in the United States was a checkers game program written by Arthur Samuel in 1952. This program, developed on the IBM 701 prototype, was the foundation of Samuel's checkers-playing software, which he expanded significantly over the years. By 1955, Samuel added features that allowed the program to learn from experience, leading to his program winning a match against Connecticut's checkers champion in 1962 [12].

Four years later, proof of concept was demonstrated by Allen Newell, Cliff Shaw, and Herbert Simon through a program called *Logic Theorist*, designed to simulate human problem-solving skills. Funded by the RAND Corporation, this program was considered by many to be the first AI program. It was presented in 1956 at the Dartmouth Conference in Hanover, New Hampshire, organized by Dartmouth mathematician John McCarthy, who coined the term "artificial intelligence" at the event. Although the conference fell short of McCarthy's expectations—with attendees coming and going as they pleased and failing to agree on standardized methods for the field—there was a shared enthusiasm that AI was achievable [8]. The search for advances in computing and digital technology continued. In 1957, Dr. Frank Rosenblatt introduced the *Perceptron*, a single-layer artificial neural network designed to simulate brain neurons. Initially implemented on an IBM 704 computer in the Cornell Aeronautical Laboratory, the Perceptron could capture an image of a triangle and transmit it through a random series of lines to response units. This was the first time in history that a machine could perceive, recognize, and identify its environment without human control [13]. In 1966, the first chatbot, *ELIZA*, was developed by MIT computer scientist Joseph Weizenbaum. *ELIZA* simulated therapeutic conversations by reusing user inputs in questions to prompt further conversation. Weizenbaum intended to demonstrate the simplicity of machine intelligence, but instead, many users believed they were speaking to a human professional. In a paper, Weizenbaum noted, "It was difficult to convince some people that *ELIZA*... was not human" [14].

During the mid-1980s, Japan announced its ambitious "Fifth Generation Project" aimed at conducting pioneering AI research with practical applications. The Japanese preferred the programming language PROLOG, a European counterpart to the American LISP, which had emerged in the 1970s [1]. Project leaders described it as a "modern" leap in computer technology, intended to position Japan as a technological

leader for years to come. Rather than relying on standard microprocessors, this new generation of computers would utilize multiprocessor machines specialized in logic programming. The aim was for these high-powered logic machines to revolutionize information processing and achieve artificial intelligence [15].

1.3. The Maturity Phase of AI Applications

In later years, extensive studies and literature began to populate search engines and libraries, especially after Tom Mitchell published his book *Machine Learning* in 1997, where he defined machine learning as a "computer program that learns from experience." Another study, titled *Deep Learning*, by researchers Ian Goodfellow, Yoshua Bengio, and Aaron Courville, built on Mitchell's definition, attributing deep learning—a type of machine learning credited with many modern AI advancements—to the ability to "experience datasets" [16]. By the end of the 1990s, a robust foundation for machine learning concepts and their integration into scientific institutions had been established, introducing terms such as deep learning and artificial intelligence. In 1997, IBM's chess computer *Deep Blue* defeated world champion Garry Kasparov in a six-game match [12]. Then, in 2000, the social robot *Kismet* was developed at MIT's AI Lab under Dr. Cynthia Breazeal, equipped to recognize and simulate human emotions through sensors, a microphone, and programming for processing human emotions [14].

In 2015, AI faced a major ethical hurdle when Elon Musk, Stephen Hawking, Steve Wozniak, and 3,000 others signed an open letter urging governments to ban autonomous weapon development. By 2016, robots could mimic human emotions thanks to Hanson Robotics' human-like robot, Sophia. In 2017, Meta developed two AI chatbots to practice negotiation skills, and Stanford researchers published a paper on diffusion models, pioneering unsupervised deep learning using thermodynamics. Google researchers introduced the groundbreaking concept of transformers in their paper "Attention Is All You Need," inspiring future research on large language models [17]. In 2019, Microsoft released its *Turing Natural Language Generation model*, which contained 17 billion parameters. Google AI and Langone Medical Center's deep learning algorithm surpassed radiologists in detecting potential lung cancers. From 2020, Oxford University developed the AI test *Curial* for rapid COVID-19 detection in emergency rooms. The same year, OpenAI launched *GPT-3 LLM* with 175 billion parameters, producing human-like text, and Nvidia announced the beta release of its *Omniverse* platform for 3D modeling in the physical world. In 2021, OpenAI introduced DALL-E, a multimodal AI system that could create images from textual prompts. The University of California, San Diego, also developed a soft quadrupedal robot powered by compressed air rather than electronics. In 2023, OpenAI released *GPT-4 LLM*, a multimodal model capable of processing both text and image prompts [17].

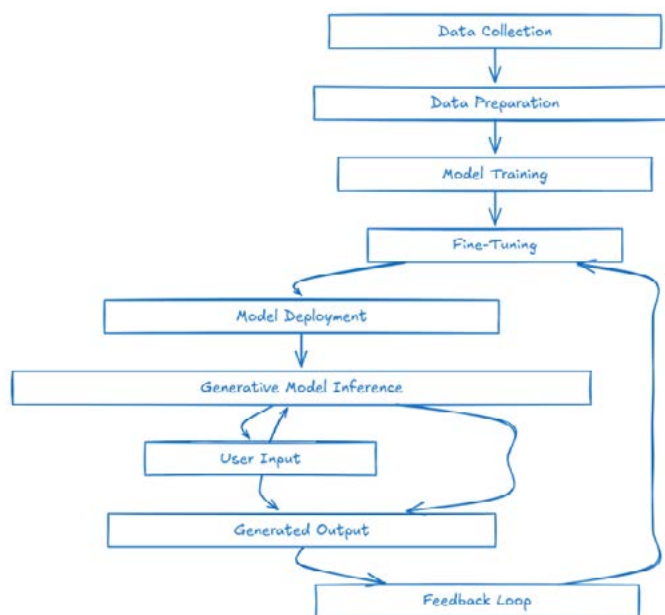


Figure 1: Shape Number (1) How Generative Artificial Intelligence Works

2. Projects for Reconstructing Ancient Texts and Languages

When an AI model is trained, it processes massive amounts of data that flow through layers of neural networks. These layers adjust their weights based on the difference between the model's predictions and actual outcomes. Over time, the model becomes proficient at providing accurate predictions. When a user requests information from the AI model, they are essentially querying the knowledge acquired during its training. The user's request is processed through the same neural network layers, which now use their adjusted weights

to produce relevant outputs [18]. Historians, therefore, often rely on conjectures, which are sometimes inaccurate and based on unknown variables. To assist in this area, Yanniss Assael, a researcher at the British AI company Deep Mind, collaborated with The Sommerschildt to develop a neural network called *Ithaca*. This model can reconstruct missing parts of historical inscriptions and identify the dates and locations of these texts. The researchers claim that their deep learning approach, trained on a dataset of over 78,000 inscriptions, is the first to tackle restoration and attribution simultaneously by learning from vast amounts of data [19].

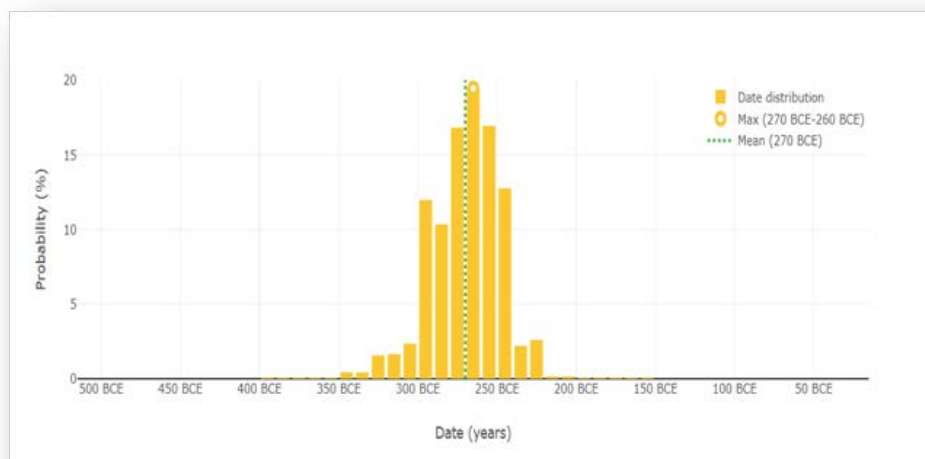


Figure 2: Represents the Extent of Ithaca's Knowledge of Ancient Texts

According to DeepMind's official data and information published on the *Ithaca* project site, named after the Greek island long awaited by the hero Odysseus, *Ithaca* can achieve an accuracy rate of up to 62% in reconstructing damaged texts. When historians use *Ithaca*, their performance jumps from 25% to 72%, underscoring the significant impact of this research tool. *Ithaca* can also attribute inscriptions to their original find locations with 71% accuracy and date them within a margin of 30 years from actual historical ranges. This capability aids in re-dating key texts of Classical Athens and contributes to thematic discussions in ancient history [19].

Ithaca uses non-sequential beam search, where each beam begins with the prediction that scores the highest confidence and then iteratively proceeds to recover the characters with the highest certainty at each time step. In an experiment I personally conducted by inputting a Greek text on the platform, *Ithaca* was able to recognize the text with 66% accuracy, successfully dating it to its original period and deciphering its code. The figure below shows *Ithaca's* temporal attribution hypotheses, visualized as a categorical distribution spanning decades, highlighted in yellow between 800 BCE and 800 CE. This approach allows for more effective handling of historical periods and aids in interpreting hypotheses. In a study conducted by a group of researchers titled "Restoring and Attributing Ancient Texts Using Deep Neural Networks," published in *Nature* in 2022, the detailed exploration of decoding remnants of symbols and writings from a collection of ancient Greek inscriptions and artifacts using AI techniques was discussed. The researchers relied on the *Ithaca* project to decode these texts. The study concluded with concrete findings regarding some artifacts, clay tablets, inscriptions, and languages inscribed on these stones.

As is well known, archaeologists rely on access to vast repositories of information to determine the origin and history of artifacts. These repositories primarily consist of the researcher's own memory bank. Recently, however, digital

datasets have been used for search processes. Yet, variations in search queries and results can exclude or obscure true findings, making it nearly impossible to accurately estimate the distribution and actual dates of these inscriptions. When key dating elements and useful search tools are missing from such tablets, historians must find alternative criteria to attribute the place and time of writing, such as letter shapes and dialects. In such cases, all these obstacles can be overcome using the latest machine learning research. Deep neural networks, inspired by biological neural networks, are capable of discovering and harnessing complex statistical patterns in vast amounts of historical data. [20]. The study referenced above relied on *Ithaca* to perform text restoration, geographical attribution, and temporal attribution simultaneously. It was trained on inscriptions written in ancient Greek between the 7th century BC and the 5th century AD. This choice was made for two main reasons: first, the diversity of content and contexts in Greek records, making it a distinctive challenge for language processing; and second, the availability of digital datasets for ancient Greek, which is an essential resource for training machine learning models.

In the figure above, there is an image of an ancient Greek inscription that researchers processed using the *Ithaca* project, and the result was truly amazing, as the remaining missing letters and texts from the ancient text were discovered. The study concluded that *Ithaca* is the first computational model for restoring inscriptions and attributing them to the era in which they were written. By significantly improving the accuracy and speed of epigraphists' work, it can help restore and attribute newly discovered or uncertain inscriptions, increasing their value as historical sources and helping historians reach a more comprehensive understanding of the distribution and nature of physical tablets throughout the ancient world [21]. Through text restoration, historians can uncover hidden patterns of influence, relationships, and social dynamics that might otherwise remain obscure.



Figure 3: Model of Decoding the Ancient Greek Language Using Artificial

The collaborative results between historians and intelligent technologies were not limited to the Ithaca project alone, but went beyond the idea of reviving written text to a project of reviving an entire city with all its features, characters, and its material and intellectual history. The Venice Time Machine project is one of the largest and most profound technical projects globally, launched by the École Polytechnique Fédérale de Lausanne and Ca' Foscari University in Venice in 2012. It aims to build a multi-dimensional collaborative model of Venice by creating an open digital archive of the city's cultural heritage covering more than 1,000 years of development. The project aims to track the circulation of news, money, commercial goods, migration, and artistic and architectural patterns to create digital data. The first reconstruction attempt of Venice showing the city's development in 1750 was presented in September 2018 during an exhibition at the Grand Palais in Paris. If successful, the project will pave the way for a more ambitious project linking similar time machines in Europe's historical centers of culture and commerce, revealing unprecedented details about how social networks, trade, and knowledge evolved over centuries worldwide [21].

Perhaps among the most prominent and important twinning projects between artificial intelligence technologies and historical studies is The Vesuvius Challenge, a machine learning and computer vision competition to restore analysis of some ancient historical documents using highly intelligent technical networks with precise interpretation, analysis, and diagnosis. The project's goal was to revive manuscripts from the coastal city of Herculaneum, located near modern Naples. When Mount Vesuvius erupted in 79 CE, it buried the coastal resort of Herculaneum. The city was rediscovered in the eighteenth century, and excavators found more than 1,000 papyrus manuscripts, the largest collection of its kind. Archaeologists believe thousands more remained buried in parts of that region but were carbonized by volcanic heat, and many were damaged during initial reading attempts. Only about 600 intact manuscripts remained from those initially discovered, appearing like lumps of coal or burned tree trunks. Based on this, the Vesuvius Challenge came to revive and read these manuscripts again [22].

The Vesuvius Challenge was launched in March 2023 to the whole world to read the Herculaneum scrolls, with a grand prize for the first team to recover 4 passages of 140 characters from the manuscripts after a year of notable progress. The research began in 2015 when researchers led by Brent Seales at the University of Kentucky discovered how to use X-ray tomography and computer vision to "unwrap" an ancient manuscript. The technology created computational images of what the pages would look like, but distinguishing letters required more progress. In March 2023, Seales announced the Vesuvius Challenge, offering substantial monetary prizes for crucial steps toward reading the Herculaneum scrolls. By year's end, a research team consisting of Youssef Nader, Luke Farritor, and Julian Schilliger managed to decode about 2,000 characters, thus winning the grand prize of \$700,000. This team couldn't have reached this result without technology experts, and although only about 5% of the text has been read so far, this is enough for scientists to determine the manuscript's perspective and subject matter [23].

The fruitful collaboration projects between artificial intelligence technologies and historical research have not stopped. Among the most prominent archaeological exploration projects using AI technologies was the Cultural Landscape Classifier (CLC) project, implemented by the Italian Institute of Technology (IIT) and the European Space Agency (ESA) from August 2022 to August 2023. Project leaders say that undiscovered cultural heritage sites beneath the earth's surface (such as buried ancient structures and artifacts) can be identified through remote sensing data from various sensors (multispectral and hyperspectral data from satellite platforms, radar data, etc.). Such datasets have enormous potential and are already widely used within the global cultural heritage community, thanks particularly to the availability of time-series imagery. Therefore, the new challenge in remote sensing studies of cultural heritage is developing or improving tools that can facilitate automatic detection of significant objects. The automation of these analyses through artificial intelligence will achieve significant benefits in terms of detecting archaeological objects in satellite images and represents a major advancement in this field, replacing current procedures based on subjective observation [24].

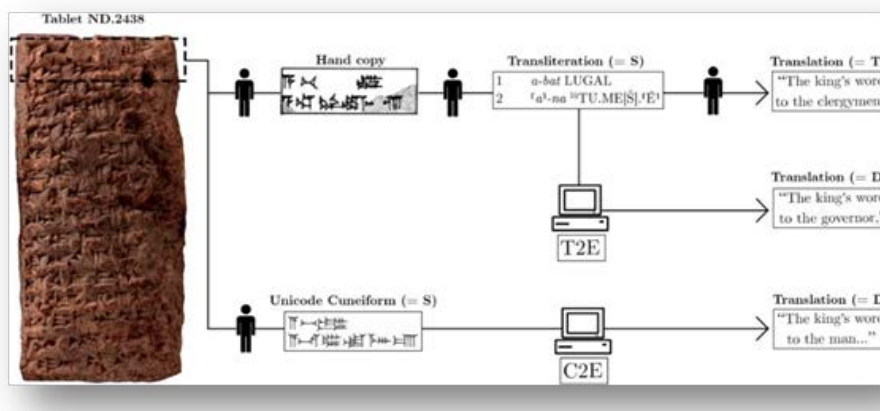


Figure 4: The Mechanism of Artificial Intelligence in Decoding the Ancient Akkadian

Archaeologists' use of remote sensing tools, such as satellite imagery, aerial photography, and geographic information systems, are now standard parts of the archaeological toolkit for activities such as landscape analysis, site discovery, and monitoring cultural heritage properties. This helps in understanding landscape archaeology and cultural heritage management. The use of remote sensing has clear advantages, as it allows for documenting cultural heritage and mitigating threats in areas that are difficult or impossible to access, and accelerating the pace of assessment that cannot be achieved in the field. Openly available remote sensing tools like Google Earth enable easy research and facilitate monitoring of these archaeological sites and thus their international management [25]. Among other strategic scientific projects that highlight the close collaboration between humans and artificial intelligence is a study published in 2023 by a research team titled:

"Translating Akkadian to English with Neural Machine Translation." This study utilized natural language processing (NLP) techniques, such as convolutional neural networks (CNN), to automatically translate Akkadian from cuneiform script directly into English. The research team stated that high-quality translations can be obtained when translating directly from cuneiform to English, achieving scores of 36.52 and 37.47 in the best binary language evaluation, which indicates that the model performs best with short to medium-length sentences (approximately 118 characters or fewer) [26]. As the number of digital texts grows, the model can be further improved through additional training as part of a human-in-the-loop system that corrects the results [27].

3. Practical Experience of Artificial Intelligence in Recognizing Ancient Inscriptions

Artificial intelligence tools can recognize ancient languages and analyze the content of certain ancient texts, such as cuneiform and hieroglyphics, among others. For instance, I personally presented an image of a clay tablet inscribed in cuneiform to the Gemini model and asked it to identify the language of the tablet. Its response was accurate. Moreover, it provided me with a summary of information about the tablet, including details about its writing style, its creator, and how it was preserved.

Below is the response from the AI model regarding the recognition of the tablet in the image above:

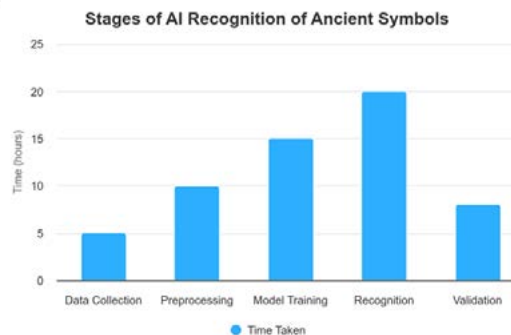
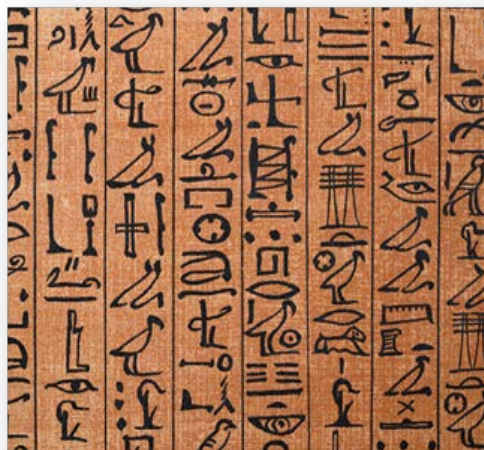
Here are some important points about cuneiform tablets that might help you:

- Cuneiform writing is an ancient writing system that was used in ancient Mesopotamia (present-day Iraq).
- Languages Used: Many languages were written in cuneiform, such as Sumerian, Akkadian, Babylonian, and Assyrian.
- Content: Cuneiform tablets were used to record a variety of information, including literary, religious, legal, and commercial texts.
- Significance: Cuneiform tablets are an important source for understanding ancient civilizations in Mesopotamia.

Then, I presented an image of a tablet inscribed in hieroglyphics to the same model (Gemini) and asked it to compare the two scripts (cuneiform and hieroglyphics), and the result was:

Feature	Cuneiform Tablet	Hieroglyphic Writing
Material Used	Wet clay tablets	Papyrus, stone, wood
Writing Tool	Stylus (pointed, beveled)	Fine pen or sharp tool
Shape of Symbols	Wedge-shaped and cylindrical marks	Images of animals, people, objects, or abstract concepts
Writing Direction	Left to right, sometimes top to bottom	Right to left, sometimes vertically
Type of Writing System	Represents both syllables and letters	Represents words and ideas with phonetic elements
Development	Evolved from a pictographic system to a phonetic system	Evolved from a pictographic system to a syllabic phonetic system
Civilization	Mesopotamia (Sumerian, Akkadian, Babylonian, Assyrian)	Ancient Egypt

Uses	Recording laws, accounts, literature, history, religion	Recording religious, literary, historical texts, and biographies of pharaohs
Accuracy	Can be very accurate in conveying information, but requires deep knowledge of the language and symbols	Can sometimes be ambiguous, depending on the context of the text



Stages of AI Recognition of Ancient Symbols

In the Copilot model, I presented it with an image of the hieroglyphic tablet and asked it to identify elements of the symbols contained within. Within seconds, it accurately interpreted the meanings of some of the signs and symbols as follows:

The hieroglyphic texts in the second image consist of symbols that represent forms of daily life and nature, such as animals, human figures, and objects. These symbols can be read as images representing specific things or as symbols for sounds or groups of sounds. For example, the following symbols may represent:

- The bird: Often represents the letter "A" or "Aleph."
- The eye: Known as the "Eye of Horus," it symbolizes protection and health.
- The seated man: Represents the letter "M" or "Mem."
- The lotus flower: A symbol of beauty and purity.
- The snake: May represent the letter "J" or "Ji."
- The lion: Symbolizes strength and courage.
- The hand: Often symbolizes action or work.
- The water: Represented by wavy lines, symbolizes life and purity.

When I presented an image of an ancient Greek inscription, as shown above, and asked it to recognize and discover its

content and meaning, the result was:

The image shows an ancient stone inscription written in Greek letters. The text is arranged in lines, and the inscription appears to be well-preserved, with the letters clearly visible and distinct from the stone background. This inscription represents a piece of historical writing that can provide insights into languages, cultures, or ancient events.

The stone inscription written in Greek contains words that may be names or ancient phrases. I will attempt to interpret some of the words found in the text:

- YAHIKANOAH: May be a name or title.
- EPTOMIEHEOAH: May be a phrase or part of a religious or philosophical text.
- DYKTOE KEPPEINIG: Could be a phrase referring to a place or person.
- OYYIOE. OYATINIA: May be part of a larger text, perhaps referring to an event or person.
- AOZEBAE: Could be a name or a word referring to something specific.
- DE AILO: May be a phrase referring to a place or person.
- ENOETP: Could be part of a religious or philosophical text.
- IMMAT: May be a name or a word referring to something specific.

Item	Description
Understanding Ancient Civilizations	Texts reveal customs, traditions, and social and cultural patterns.
History of Writing	Illustrates the evolution of writing, the invention of writing systems, and their impact on communication and documentation.
Economic History	Provides records of trade, agriculture, and industry, offering insights into economic life.
Legal Texts	Clarify legal systems, individual rights, and the evolution of legal concepts.
Myths and Literature	Offer insights into beliefs, cultures, and literary influences among peoples.
Cultural Translation	Highlights the exchange of ideas and mutual influences between civilizations.
Academic Development	Fuels research fields such as anthropology and archaeology.
Reconstructing History	Serves as a primary source to help historians accurately reconstruct historical events.

Table 2: The Importance of a Historian's Knowledge of Ancient Linguistic Texts According to the Copilot Model

The careful study of ancient languages offers numerous benefits to historians and aids them in their scientific research journey. Students of classical languages gain an appreciation for all world literature when they return to original sources. Over the centuries, poets and literary giants of their respective eras have demonstrated the allure of ancient literature in Rome and Greece, for example. Their vocabulary significantly increases, as does their knowledge of synonyms, nuances of meaning, and etymology (the origin and history of words). They also acquire rhetorical skills in using literary devices and stylistic diversity in their writings. Thus, understanding the impact of ancient civilizations on our current civilizations comes through learning those languages. Researchers learn concepts related to governance, legal systems, medicine, currency, astronomy, architecture, music, science, mathematics, and more, gaining a crucial historical link between the present and the past that provides timely perspective on recurring events in the modern era. Another benefit arising from the study of classical language is the training in the fundamentals of the scientific method of induction, which includes observation, comparison, generalization, and continuous exchange with inference [28]. Learning an ancient language helps open

the door to a past and history that many modern languages cannot provide. It teaches us historical understanding that fundamentally aids us in learning effectively from the past. It also grants us access to the thoughts and concepts of humans from hundreds or thousands of years ago, allowing us to hear their voices and learn from their wisdom.

4. Text Analysis and Recognizing Historical Patterns

Artificial intelligence can analyze historical texts to understand the historical and social context of the period in which we are writing. Natural Language Processing (NLP) techniques can be used to extract important information from texts and analyze it.

4.1. What is Natural Language Processing (NLP)?

Natural Language Processing (NLP) is a branch of computer science and artificial intelligence that uses machine learning to enable computers to understand and communicate in human language. NLP allows computers and digital devices to recognize, understand, and generate text and speech by combining computational linguistics — rule-based language modeling — with statistical modeling, machine learning (ML), and deep learning.

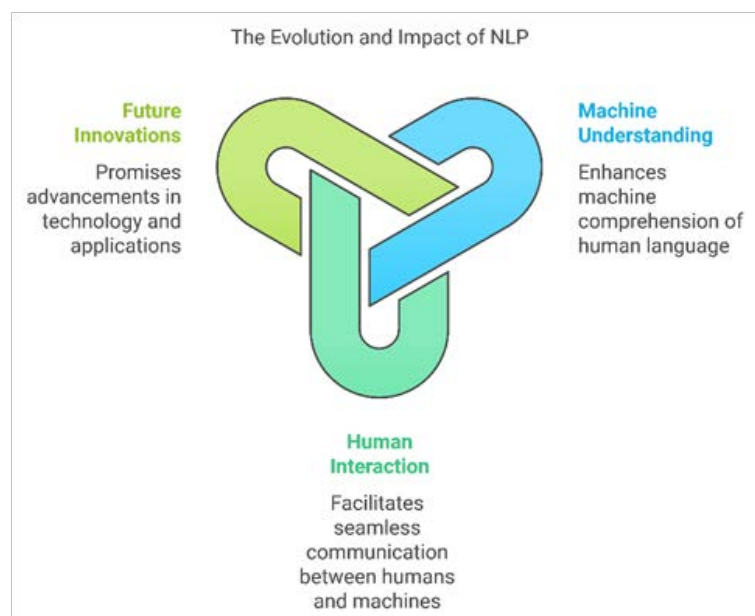


Figure 5: Shows how Natural Language Processing Works in Artificial Intelligence

NLP research has enabled the era of generative AI, ranging from communication skills in large language models (LLMs) to the ability of image generation models to understand requests. NLP is already part of everyday life for many, powering search engines, driving voice-activated customer service chatbots, voice-guided GPS systems, and digital assistants on smartphones. NLP also plays an increasing role in enterprise solutions, helping to streamline and automate business processes, boost employee productivity, and simplify essential business operations [29].

4.2. How Natural Language Processing Works

- **Text Analysis:** The NLP process typically begins with text analysis, where the text is broken down into smaller units, such as words and sentences. Techniques include:
- **Tokenization:** Dividing the text into words or sentences.
- **Parsing:** Understanding the syntactic structure of sentences.
- **Meaning Comprehension:** After text analysis, language models are used to understand the various meanings of words and sentences. These models include:
- **Word Embeddings:** Converting words into numerical representations that facilitate understanding relationships between them.
- **Neural Networks:** Used to understand context and complex meanings.
- **Text Generation:** After understanding the text, systems can use text generation techniques such as:
- **Automatic Text Generation:** Creating new text based on input information.
- **Machine Translation:** Converting text from one language to another.

In recent years, initiatives have been launched to digitize as many historical documents as possible, including library catalogs containing millions of pages of newspapers, archival records, documentary documents, and manuscripts, some dating back to ancient historical periods.

For researchers and historians specifically, this material accumulation of knowledge (in paper form) represents both a problem and an opportunity. Yes, it is now possible to subject this intellectual wealth to theoretical frameworks that can be processed by computer technology. Now, we can capture every piece of paper with written information, regardless of its scientific value, as computing machines have overcome all obstacles that historians faced when flipping through archival pages full of information. However, there is still no established method to filter them.

In a project called **Sphaera**, the aim was to understand the mechanisms of homogeneity and diversity in scientific knowledge from the late Middle Ages to the end of the early modern period, and to reconstruct the broader transformation process—and its mechanisms—at work in hundreds of historical sources considered in the study. Thus, the project explores the evolutionary path of the scientific system between the 13th and 17th centuries [30]. In the context of such projects, researchers from the German **BIFOLD Institute** trained a neural network to detect, classify, and cluster (based on similarity) illustrations from early

modern texts; this model is now available to other historians through a highly significant web service for historians called **Cor Deep**. They also adopted a new approach to analyzing other data. For instance, it was not possible to visually compare the various tables in hundreds of books in the collection because “the same table could be printed in a thousand different ways.” Therefore, the researchers developed a neural network architecture that detects and groups similar tables based on the numbers they contain, ignoring the design [31].

4.3 What is the Cor Deep Project?

Cor Deep is a machine-learning-based web application for extracting visual elements from historical sources and classifying pages containing numerical and alphanumeric tables. It identifies and classifies visual elements into the following categories: “content illustrations,” “initial letters,” “decorations,” and “typographic marks.” The application was trained on a dataset of 359 early modern manuscripts containing approximately 78,000 pages, 30,000 visual elements, and 10,000 pages with tables. Visual elements were manually annotated using bounding boxes and semantic tags, while pages containing tables were semi-automatically identified by a progressively refined model overseen by a human expert. Cor Deep achieves an average accuracy of 98% in detecting visual elements and 94% in classifying pages with tables. These values may vary depending on the style, content, and quality of the input images. Various data and documents in diverse formats can be used to conduct comprehensive content analysis [32].

I personally experimented with the application by subjecting several scientific files to its testing, and the result was that it was able to identify the data within those files, including “content illustrations,” “decorations,” “typographic marks,” and “tables.” You can easily upload your files processed by the application. The impact and importance of pattern recognition and keyword recognition techniques in historical studies through artificial intelligence are rapidly increasing in many areas of our daily lives. Among the most affected fields is the recognition of historical patterns of cultural heritage. There is a growing need for new solutions to document and manage sites, artifacts, and any type of data containing information about the past. However, the information found in digital representations of physical objects is not easy to exploit, requiring intelligent analysis and innovative methodologies. These issues and questions push historians to leverage recent technological advances to develop new applications aimed at better analyzing available data or generating new data in a smarter ways’ [33].

5. Conclusion

The subject may seem unappealing to some, especially to historians and old writers who do not believe in the connection between digitization and technology in historical work. Many consider that dealing with ancient papers, document texts, and manuscripts requires a sharp, attentive research mind. Yes, we all undoubtedly believe that. However, what prevents the technological revolution from entering the study of certain models in the humanities,

including history? What is the problem with blending the past with the present? Historical research before the era of the digital revolution was slow and arduous compared to today, involving significant time consumption in the process of researching archives and libraries. The primary research tools at that time were cumbersome microfilm reels. The steps were somewhat primitive and slow, beginning with the challenge of selecting a title and its ramifications, followed by exploring the relevant primary and secondary literature related to the research topic. The researcher would contact a specific writer and arrange an interview with another. Imagine how much time that takes! They would search for a translator to translate some texts from languages other than the one in which they were writing, copy some documents and articles, which sometimes incurred substantial financial costs. Now, all of these pathways have changed and have become part of the past with the rise of digital warfare and the revolution against everything old and primitive. In today's world, all research tools are available to historians while they are at home, with the exception of a few things that may require them to leave their houses to obtain, such as a calculator or a high-resolution digital camera, and so on.

References

- Mijwel, M. M. (2015). History of Artificial Intelligence Yapay Zekânın T arihi. *Computer Science*,(April 2015), 3-4.
- Saluja, S. (2019). E1 : AI Evolution and Definition.
- Andrzej Grzybowski, Katarzyna Pawlikowska-Łagód, W. Clark Lambert, Andy, Phippen. (2024). Artificial Intelligence.
- Marquis, P., Papini, O., & Prade, H. (2020). Elements for a history of artificial intelligence. *A Guided Tour of Artificial Intelligence Research: Volume I: Knowledge Representation, Reasoning and Learning*, 1-43.
- Tableau. (n.d). What is the history of artificial intelligence (AI)?.
- Benko, A., & Lányi, C. S. (2009). History of artificial intelligence. In *Encyclopedia of Information Science and Technology*, Second Edition (pp. 1759-1762). IGI global.
- Guerrero, J. M. (2023). Chapter 5—The history of modern artificial intelligence. *Mind Mapping and Artificial Intelligence; Academic Press: New York, NY, USA*, 131-161.
- Anyoha, R. (2017). The history of artificial intelligence. *Science in the News*, 28.
- Kaul, V., Enslin, S., & Gross, S. A. (2020). History of artificial intelligence in medicine. *Gastrointestinal endoscopy*, 92(4), 807-812.
- Turing, A. M. (2009). *Computing machinery and intelligence* (pp. 23-65). Springer Netherlands.
- Manhattan Rare Book Company.(n.d). Computing Machinery and Intelligence.
- Copeland, B. (2024). History of artificial intelligence (AI). *Encyclopedia Britannica*.
- Basuru, Vikas. (2020). The Evolution of Artificial Intelligence.
- Coursera, S. (2024). The History of AI: A Timeline of Artificial Intelligence.
- Smith, C. (2006). The History of Artificial Intelligence. University of Washington.
- Sternfeld, J. (2023). AI-as-Historian. *The American Historical Review*, 128(3), 1372-1377.
- Karjian, R. (2023). The history of artificial intelligence: complete AI timeline.
- Sigurdson, P. (2023). The Engines of Excellence: From Ancient Secrets to AI Mastery.
- Assael, Y., Sommerschild, T., Shillingford, B., Bordbar, M., Pavlopoulos, J., Chatzipanagiotou, M., ... & de Freitas, N. (2022). Restoring and attributing ancient texts using deep neural networks. *Nature*, 603(7900), 280-283.
- Assael, Y., Sommerschild, T., Shillingford, B., Bordbar, M., Pavlopoulos, J., Chatzipanagiotou, M., ... & de Freitas, N. (2022). Restoring and attributing ancient texts using deep neural networks. *Nature*, 603(7900), 280-283.
- Abbott, A. Digital history. *Nature* 409, 556–557 (2001).
- Resurrect an ancient library from the ashes of a volcano. (n.d).
- Postrel, V. (2024). The Future of AI Is Helping Us Discover the Past.
- ESA.(n.d).Cultural Landscapes Scanner (CLS).
- Fisher, M., Fradley, M., Flohr, P., Rouhani, B., & Simi, F. (2021). Ethical considerations for remote sensing and open data in relation to the endangered archaeology in the Middle East and North Africa project. *Archaeological Prospection*, 28(3), 279-292.
- But T. (2024). The Latest AI Innovations in Archaeology.
- Gutherz, G., Gordin, S., Sáenz, L., Levy, O., & Berant, J. (2023). Translating Akkadian to English with neural machine translation. *PNAS nexus*, 2(5), pgad096.
- Klousia, William L. (n.d). The Importance of Classical Language Study.
- Holdsworth, J. (2024). What is NLP?.
- Sphaera. (n.d).
- Donovan, M. (2023). How AI is helping historians better understand our past. *MIT Technology Review*.
- What is CorDeep? (n.d).
- F. Fontanella, F. Colace, M. Molinara, A. Scotto Di Freca, F. Stanco, Fagelson, Doron. (2024). Bridging History and Innovation: The Impact of Digital Twins on Cultural Preservation.