From Digitisation to Computation: Digital History, Invented Archives, and the Epistemology of Historical Knowledge in the Age of Artificial Intelligence

Salvatore Spina

University of Catania, Catania, Italy

This article synthesises the theoretical framework of my paper "*Il documento digitalizzato, tra ricerca storica, Archivistica e intelligenze artificiali*", which was discussed at the round table entitled "Beyond the Boundaries of the Past: Innovation and Interaction in the Age of Digital Archives", held on 4 June 2025 in Modena, Italy, as part of the Seventh National Public History Conference of the Italian Association of Public History (AIPH), *Storie in cammino*. The paper deals with the current transformation of historical research through digital technologies, which cannot be adequately understood if "digitisation" denotes merely the conversion of analogue objects, signals, and documents into digital formats. Instead, it proposes a shift toward a notion of digitisation oriented explicitly toward computation, understood as the formalisation, structuring, and operationalisation of historical knowledge in ways that make it interrogable by machines without relinquishing its interpretive complexity. Taking the history of slavery as a sample phenomenon of observation, the paper situates contemporary digital projects within the long historiographical tradition of *histoire sérielle*, reinterpreted as a computational epistemology rather than a purely quantitative method. Through the analysis of *invented archives*, historical data infrastructures, and the use of artificial intelligence (AI) tools—particularly generative models—the paper aims to demonstrate that the decisive challenge for Digital History (DHy) lies not in technological adoption but in the construction of epistemically robust workflows, ontologies, and datasets. By engaging with the theoretical legacy of Richard Ennals and current debates on datafication and AI, the paper contends that only a computation-oriented approach to digitisation can ensure transparency, reproducibility, and critical control over historical knowledge in the "*i*Age of intelligent machines".

Keywords: datafication of historical knowledge, computational epistemology, historical data infrastructures, serial modelling, machine-readable historiography

Introduction

The progressive digitisation of historical sources and research practices has profoundly reshaped both the epistemological horizons and the operational tools of the historians' craft. From the emergence of the World Wide Web to the consolidation of the "Digital Ecological Niche" (Spina, 2023c), the online circulation of registers, inventories, and complex documentary corpora has enabled the creation of thematic digital archives and fully fledged "invented archives"—which no longer coincide with institutional repositories but rather constitute documentary environments designed to answer specific research questions. In this scenario, Digital

Salvatore Spina, Ph.D., Department of Humanities, University of Catania, Catania, Italy.

History (DHy) asserts itself not merely as an applied field of information technologies, but as a domain in which to reconsider the relationship among documentary seriality, the datafication of sources, and the construction of historical memory.

The methodological ideas developed in the late 20th century—from the quantitative history (Darcy & Rohrs, 1995; Dollar & Jensen, 1974) and *histoire sérielle* of Chaunu (1959; 1964) and Furet (1971; 1968) to the theoretical contributions of Chabod (1969) and Le Roy Ladurie (1973; 1968)—had already placed at the core of historical inquiry the need to select, organise, and model serial documentary collections to render long-term phenomena intelligible. The current proliferation of historical datasets and digital archives on the early modern period, for instance in the field of slavery studies, pushes this legacy further: Documentary complexes dispersed across different physical archives are reassembled in relational databases (Rosenzweig, 2003; 2001) and online platforms that function as *invented archives* (Spina, 2025a), epistemic spaces in which classifications, fields, and codifications determine what becomes searchable and what remains marginal. In this sense, Digital History is simultaneously a historiography of "digitality" (Spina, 2024)—it aims to reconstruct the processes that have created the digital "dimension", *i.e.* the Digital Ecological Niche (DEN)— and a methodology, because the design of such infrastructures constitutes a fundamental moment of historical analysis (Spina, 2022a).

At the same time, the development of artificial intelligence (AI) tools—from Turing's key question "Can machines think?" (Turing, 1950) to the Dartmouth conference (McCarthy, Rochester, Minsky, & Shannon, 1955), from symbolic AI and expert systems to machine learning, deep learning, and the contemporary foundation models based on Transformer architectures—has introduced new "algorithmic actors" into the life cycle of historical research (Spina, 2023c). The latest generation of generative models, and particularly Large Language Models (LLMs), no longer merely classify or extract information; they generate texts, narratives, data structures, code, and representations that may intervene directly in processes of reading, extracting, and formalising historical sources. The integration of tools such as ChatGPT into the daily workflow of scholars thus raises crucial questions about scientific reliability, the opacity of computational processes, embedded biases, and the redefinition of authorship and agency in the production of historical knowledge.

As early as the 1980s, Richard Ennals (1985) had recognised the transformative potential of these developments, proposing that computers be understood as epistemic historians' *ally*—provided that historians were able to formalise their inferential procedures and translate them into computable operations. Today, this requirement re-emerges with amplified urgency: The integration of LLMs and generative AI tools into research workflows compels us to rethink the relationship between source interrogation, data modelling, simulation, and explanation, thereby redefining the historian's profile as a digital scholar capable not only of designing documentary environments but also of "training" machines to recognize and handle the specificities of historical discourse.

This paper situates itself at the intersection of these trajectories and takes as a privileged case study the construction of a database on slave trading between the Mediterranean and the Atlantic (15th-16th centuries) (Spina, 2025b) to examine, from both a methodological and epistemological perspective, the use of generative AI in transforming archival documents into structured data. Its aim is twofold: first, to show how *invented archives* and *serial* datasets on slavery function as laboratories in which documentary hierarchies, analytical categories, and modes of visualizing the past are continually redefined; and second, to discuss how the use of LLMs in processes of data extraction, normalization, and modelling requires a renegotiation of the historians' critical role as mediators between algorithms and sources, between memory and computation, and between the

opacity of digital infrastructures and the demands of transparency, verifiability, and accountability that are constitutive of historical scholarship.

Digitisation. The Wrong Path of the Last 30 Years!

Digital History has solidified its standing as a primary site of methodological innovation within the historical sciences over the past several decades. Nevertheless, its progressive institutionalisation has not yet resolved a fundamental tension concerning the meaning of "digitisation" and the relationship between computational technologies, historical knowledge, and the epistemic practices of the discipline. In the realm of scholarly production and research projects, the concept of digitisation is predominantly regarded as a technical process of converting analogue documents into digital formats, often referred to as facsimiles (or pics). This approach is primarily aimed at enhancing accessibility and circulation of sources. Additionally, digitisation is frequently conceptualised as a set of tools that can accelerate and optimise conventional activities, such as archival research, cataloguing, transcription, and writing. Despite the considerable impact of these developments on the day-to-day work of historians, there has been no concomitant structural conversion of the discipline's epistemology. That is to say, the digital is treated as a support or a medium rather than as a constitutive dimension of the process of historical knowledge production. The present implementation of artificial intelligence tools and "generative language models" has made this ambiguity increasingly evident. This demonstrates that the central issue is not merely the adoption of new technologies but rather the capacity of historiography to engage with computation as a specific form of rationality, formalisation, and organisation of knowledge.

In this sense, DHy cannot be reduced to a technical specialisation or methodological fashion. It is imperative to acknowledge this domain as a fertile ground for theoretical contemplation, wherein the intricate interrelationships among sources, data, models, interpretations, and the infrastructural underpinnings of historical knowledge are undergoing a paradigm shift. The process of redefining the concept necessitates a thorough elucidation of the dual nature of Digital History. This discipline functions in a dual capacity, operating as both a (1) "*history* of digitality" (Spina, 2024) and a (2) *methodology* that is employed in the context of historical research.

In the first scenario, digital history is defined as the "history of technological developments in communication systems".

The terms "digital", "digitisation", and "digital turn" are frequently employed to denote contemporary phenomena, particularly those associated with the technological system of the past 50 years, particularly those related to the field of Computer Science. Nevertheless, if we were to extend the analytical clock further back in time, it would become evident that digitality is not a recent phenomenon. Rather, it can be seen to dissolve into the deeper currents of History. Computer Science represents a recent manifestation of an ancient human process: communication. However, nowadays, communication is not directed towards other human beings but rather towards machines, thereby enabling us to issue instructions and prompts. The development of programming languages is predicated on human-computer communication. For instance, between 1968 and 1969, Niklaus Wirth developed the "Pascal" scripting language, which was designed to compel students to produce clean, structured code that was free of common logical errors. The language incorporates rudimentary English words that function as commands, such as "Begin", "End", "Repeat", "Until", and "While", to prevent the merging of numbers and words, a common error in human communication, which cannot be present in programming languages.

The necessity to devise a language for machine control emerged in the 19th century, marking a seminal point (*terminus a quo*) in the evolution of digitality and, consequently, in the context of historiographical inquiry into the field of Digital History. The latter is defined as the *reconstruction and analysis of the theoretical and pragmatic processes that have culminated in the digital revolution that is currently being experienced*. In this sense, Digital History is regarded as a form of history that focuses on the information system revolution.

If the assertion that digital technologies have fundamentally altered our perception of life is accurate, then it follows that historians must assume the responsibility of meticulously unravelling the processes that, through the establishment of the Digital Ecological Niche (DEN) (Spina, 2023c), have metamorphosed the systems of human interaction. Within this framework, digital scholars explore the causes and consequences of this transition, focusing on the formation of the digital "dimension" and the transformation of our experience and perception of life.

From this vantage point, the designation "Digital History" is accorded the same explanatory status as other well-established subfields within the broader historical discipline—such as *economic* history, *political* history, or *religious* history. In this framework, Digital History (DHy) emerges as a new subfield dedicated to the study of digitality, the DEN, and all of their historical dimensions.

In such a context, technologies—for instance, "radio, cinema, and their programming", as Chabod once noted (1969)—assume the role of new historical sources that scholars must necessarily analyse to describe the events and actors of our age.

Digital historians who are engaged in the task of writing a historiography of Digital History will direct their attention to a wide array of technological phenomena, which include—starting from lullism, through the study of Ada Lovelace's ideas (Spina, 2024), up to AIs—techniques of calculation, machines for information processing, systems of recording, and communicative infrastructures, computers, mainframes, software and hardware, operating systems, programming languages, the Internet, the Web, websites, emails, and messages exchanged via instant messaging services such as WhatsApp, Telegram, and iMessage. Additional phenomena include artificial intelligence platforms, computational systems and software, and the technologies underpinning smartphones, tablets, and PCs. The attention of historians will also encompass digital archives and databases (Itzcovich, 1989; Derosas, 1989a; 1989b), as well as social media profiles. These sources are of paramount importance in comprehending and delineating the historical evolution of digitality, a novel system of interaction, the emerging economic landscape predicated on the Big Data market, the anthropological evolution of novel cognitive entities (e.g., Homo-Loggatus), and our progressive perception of the technological milieu and life itself.

Consequently, computers, software, and digital systems are transformed into "documentary information", disseminated through novel and varied digital media formats, contrasting with conventional paper-based methods. ICTs serve as evidence of the evolving means of human expression and of the "progress of Science"—historians, as Giuseppe Galasso observed, must adapt History and memory to these new sources (Galasso, 2010).

However, beyond this initial observation, ICTs—unlike paper documents, which require analysis by human scholars—require that technology itself analyse them. If we want to comprehend the information technology and its epistemological value for historical research, we are compelled to use "information technology" to analyse it and to dig into it and extract the historical meaning—*we need technology to analyse technology!* In order to comprehend the structural elements of a website and ascertain its epistemological value as a source for history, it is imperative to acquire a comprehensive understanding of Hypertext Markup Language (HTML).

This assumption signifies that digital history, in its capacity as an object of research, metamorphoses into a methodological framework that historians employ to comprehend it.

This duality indicates that Digital History is not merely a history composed with the aid of digital instruments; rather, it is a critical examination of the manner in which Information and Communication Technologies reconfigure the relationship between historians and the past, between documentation and interpretation, and between tacit knowledge and its explicit formalization. The increasing centrality of born-digital sources, in conjunction with the proliferation of traces produced within a digital ecological niche, renders this transformation particularly evident, insofar as contemporary documentation is not only more abundant but also intrinsically structured, relational, and potentially computable. Nevertheless, despite this potential, most digitisation practices continue to reproduce the traditional methodology (close reading), treating the digital document as a functional equivalent of the paper document, thereby relegating the computer to an instrumental role of logistical support.

Digital archives, online libraries, and thematic collections (*invented archives*) have undoubtedly expanded access to archival sources and accelerated research processes, but they have rarely affected the deeper structures of historical knowledge, which remain almost exclusively entrusted to the interpretive competence of historians and to the discursive form of written text (Spina, 2025a). In this context, documentary digitisation appears as a necessary but insufficient condition for a genuine methodological transformation, since it does not render historical knowledge truly operable by computational systems, nor does it allow for the systematic interrogation of relations, patterns, and inferences that remain implicit within historiographical discourse. This issue becomes even more urgent with the entry of generative artificial intelligence (such as ChatGPT) tools into historiographical debate, since such systems, while demonstrating a remarkable capacity to produce coherent and historically plausible texts, expose the structural limits of a digitisation process that is not oriented toward computation. Large Language Models (LLMs) are predicated on statistical correlations learned from vast textual corpora, predominantly derived from the Web (e.g., accessible websites, Wikipedia). These models lack direct access to archival sources in their complexity and to the epistemic criteria that guide the formation of historical knowledge. Consequently, they produce a form of simulated knowledge that risks being confused with historical knowledge proper precisely because it lacks a clear distinction between source and interpretation. The automated text generation reveals that the issue does not originate from the utilisation of artificial intelligence systems per se. Rather, it stems from the dearth of adequate historical data infrastructures capable of supporting authentic computational interaction between machines and historical knowledge.

It is within this framework that the theoretical contribution of Richard Ennals acquires relevance, as he had already identified, in a phase preceding the development of generative AI, the necessity of making the cognitive practices of historians explicit so that they could be supported by computational systems. His insistence that the fundamental problem does not reside in the computer, but in historians' capacity to describe what they do, brings to light a central epistemological issue: Computation cannot operate effectively on a form of knowledge that remains implicit, tacit, and embedded exclusively in individual scholarly practice. The three areas identified by Ennals—(1) intelligent information retrieval, (2) modelling and simulation, and (3) explanation and advice—do not aim to replace human interpretive labour, but rather to create the conditions under which the computer can operate on structured representations of historical knowledge, rendering questions, criteria, relationships, and inferences interrogable. This approach entails a crucial shift from the mere digitisation of sources to the formalisation of what historiography asserts about the Past, that is, toward the datafication of historical knowledge.

The process of datafication should not be construed as a simplification of historical complexity, reducing the intricacies of the Past to a series of numbers or isolated variables. Datafication aims to construct structured representations of events, actors, relationships, and processes that make inferences explicit and enable systematic interrogation. From this scenario, datafication concerns not only sources, but also Historiography itself, understood as a set of assertions, interpretations, and models that can be rendered partially machine-readable without losing their historical character. The transition, however, is laden with epistemological implications, as it compels historians to reflect on which levels of knowledge can be formalised and which must remain entrusted to interpretive discourse, acknowledging that not all hermeneutic complexity can be translated into computational structures, while recognising that a significant portion of cognitive operations can be made explicit without impoverishing historical meaning.

Many of the so-called digital *invented archives* that have emerged in recent decades clearly reveal the limits of an approach that privileges documentary accumulation over conceptual modelling, since they facilitate access to sources but do not allow those sources to be operated upon computationally, nor do they integrate heterogeneous sources, multiple interpretations, and levels of uncertainty coherently. In the absence of *ontologies*, controlled codelists, and relational structures, such archives remain epistemologically weak and difficult to interoperate, thereby limiting the possibilities for comparative analysis, simulation, and the verification of inferences.

By contrast, projects aimed at constructing genuine historical data infrastructures demonstrate how it is possible to integrate sources, metadata, interpretations, and models within coherent systems capable of supporting complex queries and rendering the epistemic presuppositions of research transparent. Within this framework, digitisation oriented toward computation must be understood as an articulated and stratified process that does not end with document conversion but instead implies a true pipeline of "digital knowledge production", encompassing the reliable transcription of texts through automatic recognition technologies (Handwritten Text Recognitions) (Erwin, 2020; Kahle et al., 2017; Milioni, 2020; Muehlberger et al., 2019; Romein et al., 2025; Seaward & Kallio, 2017; Spina, 2020; 2025b; 2021; 2022a; 2022b; 2023b; 2023a), semantic enrichment through metadata and ontologies, the formalisation (Orlandi, 2010; 1998; 2007; 1999; 1996) of historical assertions, and the construction of interrogable infrastructures.

Only through such a process can the conditions be created for AI tools to be used critically and in a controlled manner within historical research, not as a source of epistemic authority, but as a tool supporting analysis, comparison, and hypothesis formation. The use of AI in the absence of such a computational infrastructure instead risks producing a new form of opacity (*hallucinations*), in which generated answers appear authoritative yet are not anchored to a verifiable system of sources and inferences (Maleki, Padmanabhan, & Dutta, 2024; Salvagno, Taccone, & Gerli, 2023).

The issue of computation is inextricably linked to the concepts of transparency and reproducibility in historical research. The process of rendering a portion of knowledge computable entails the explicit articulation of interpretive steps and the facilitation of more rigorous critical comparison between alternative hypotheses. In this sense, computation does not pose a threat to the interpretive dimension of History; rather, it presents an opportunity to expand its critical capacity. By leveraging computational processes, we can illuminate relationships, correlations, and gaps that often remain imperceptible within conventional discursive analyses.

Digital History, understood in this strong sense, is thus not merely a form of history that makes use of digital tools, but a reflection on how historical knowledge is constructed, represented, and transmitted in the "*i*Age" of

intelligent machines, and it requires a profound rethinking of digitisation practices so that they do not merely reproduce analogue models in digital form, but consciously orient themselves toward computation as a constitutive dimension of contemporary historical research, accepting the epistemological challenge of rendering historical knowledge partially formalizable without relinquishing its interpretive complexity.

Computation and AI: New Aims and Perspectives for the Digital History

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think". In 1950, Alan Turing posed this seminal question that initiated a novel paradigm for the interpretation of life. This inquiry has contributed to the progressive consolidation of history as a pivotal domain for AI experimentation in recent decades.

Let's start from the beginning. In a multitude of academic and project-based contexts, the term "digitisation" is predominantly employed to denote the technical process of photographing a document and storing it on a server. This process is undertaken with the objective of enhancing accessibility, preservation, and the dissemination of sources. Alternatively, "digitisation" can be understood as a suite of tools designed to augment the efficiency of established practices, such as archival research, transcription, and writing. While these developments have had a significant impact on historians' everyday work, they have not generated a substantial transformation of the discipline's epistemology, insofar as the digital has typically been treated as a medium or support rather than as a constitutive dimension of the knowledge-making process. The emergence and rapid diffusion of artificial intelligence systems—especially generative models—have rendered this ambiguity increasingly visible, showing that the central problem lies less in the adoption of new technologies than in historiography's capacity to engage with computation as a specific mode of formalisation, organisation, and interrogation of historical knowledge.

From this perspective, Digital History should be understood as a dimension of theoretical and methodological reflection in which the relationships among sources, data, interpretive models, and infrastructures of knowledge are being reconfigured.

As we highlighted in the previous paragraph, Digital History is at the same time, the object and the methodology to analyse it. This duality is crucial: It prevents DHy from being confined to a purely technical domain and instead places it at the centre of a broader reflection on the relations between Past, documentation, and knowledge in the contemporary world.

In the second scenario, DHy is the method that is based on the application of ICT tools to interrogate how computers, software, and all kinds of technology can be considered historical sources.

The growing centrality of born-digital sources and the proliferation of traces produced within digital environments make this transformation particularly apparent. Emails, databases, system logs, online platforms, and software are not merely instruments; they are fully fledged historical objects that generate specific documentary forms and demand new conceptual tools for analysis.

Nonetheless, the prevailing digitisation practices tend to adhere to conventional historical methodologies. Consequently, the majority of digitisation initiatives are constrained in their capacity to reproduce analogue documents, perceiving the digital version as a functional equivalent of the paper document. This relegates the computer to a predominantly logistical function. Digital archives, online libraries, and thematic collections have significantly expanded access to sources, but they have rarely affected the deeper structures of historical knowledge, which remain almost exclusively entrusted to historians' interpretive expertise and to the discursive form of scholarly writing. In this sense, documentary digitisation appears as a necessary but insufficient condition

for genuine methodological transformation: It does not render historical knowledge truly operable by computational systems, nor does it enable systematic interrogation of relations, patterns, and inferences that remain implicit in historiographical discourse. To clarify, I invite you to consider "The September 11 Digital Archive"[1]. Within it, no documents relating to the activities of the terrorists are brought together. There are no records concerning the political idea or motivations that led Al-Qaeda to carry out the tragedy. The archive expresses only the perception that American citizens have of the event. It preserves "Memory", not History— also, there is no data to process.

We can consider, in the same way, the "Portale delle Fonti della Repubblica Italiana"[2]. The website was created by the Consiglio Nazionale delle Ricerche to build a "historical information infrastructure" dedicated to describing, interlinking, and representing online documents relating to the political and institutional history of the Italian republican period.

All documents—from the Presidency of the Republic archive, the Historical Archive of the Senate, the Historical Archive of the Chamber of Deputies, and the Central State Archive—were scanned, but there are no data for the computational process.

For this reason, when the "Database on the Slave Trade Between the Mediterranean and the Atlantic (15th-16th Centuries)" started, the research team decided to create a historical instrument to let scholars describe the phenomenon of slavery, creating a conceptual bridge between the idea of slavery before and after the discovery of America.

From another perspective, my initial objective in joining the team was twofold: first, to elucidate my theoretical position, and second, to demonstrate the necessity of adopting a novel perspective on the application of ITCs in the domain of historical research.

Slavery is a structural, transnational, and *long-duration* phenomenon, characterised by extreme fragmentation of sources and by the systematic redacting of subaltern subjects. For precisely these reasons, it has proven particularly suitable for the application of the *histoire sérielle approach* (Chabod, 1969), understood as a method grounded in the construction of "documentary series" capable of rendering complex historical phenomena intelligible through repetition, comparison, and structural analysis. The genealogy of *histoire sérielle*, elaborated by Pierre Chaunu and developed by scholars such as François Furet and Fernand Braudel, should not be read simply as a precursor of quantitative history (Darcy & Rohrs, 1995; Dollar & Jensen, 1974; R. Ahnert, S. E. Ahnert, Ryan, & Computational Humanities Research (CHR), 2020; Cavalli-Sforza & Feldman, 1981; Dollar & Jensen, 1974; Fogel, 1975; Franzosi, 2017; Michel et al., 2011; Turchin, Currie, et al., 2018; Lemercier & Zalc, 2008; Morrissey, 2015; Ryan, S. E. Ahnert, & R. Ahnert, 2020), but as an effort to shift the centre of gravity of historical research from the singular event (*événementiel*) to *longue-durée* processes, recognising *seriality* not as a constraint but as an interpretive resource.

Within this theoretical framework, the notion of "invented archives" acquires particular significance. It does not designate "artificial" archives in an arbitrary sense, but rather digital environments designed to collect, normalise, and render interrogable a body of documentation originally dispersed across multiple physical

---

[1] Link to the website: https://911digitalarchive.org.
[2] The project was coordinated by Roberto Palaia and by a Project Committee appointed by the President of the CNR, Maria Chiara Carrozza, composed of Roberto Palaia, Marina Giannetto, Paolo Massa, Giampaolo D'Andrea, Andrea De Pasquale, Giuseppe Vacca, Andrea Giardina, Claudia Villani, Alessio Gagliardi, Paolo Acanfora, the Director of IMATI, the Director of ILIESI, the Director of ISTC, and Maurizio Gentilini. Link to the website: https://portalefontirepubblicaitaliana.cnr.it/portal/home_page.php.

repositories. Invented archives in slavery studies—such as the Transatlantic Slave Trade Database, Enslaved: Peoples of the Historical Slave Trade, and analogous projects—may be interpreted as a computational reactivation of *histoire sérielle*, insofar as they materialise seriality through the construction of structured datasets capable of supporting comparative, demographic, social, and relational analyses. In these contexts, digitisation does not merely reproduce the document; it translates it into data, entities, and relations, enabling a form of knowledge that exceeds individual reading and opens onto modelling.

The crucial transition, however, concerns not only sources but historiography itself. If digitisation is to be genuinely oriented toward computation, then what historiography asserts about the past—interpretations, hypotheses, explanatory models—must also become, at least in part, machine-readable. This is where the concept of datafication assumes central epistemological importance. The datafication of History does not mean reducing the complexity of the Past to a series of numbers or isolated variables; rather, it entails constructing structured representations of events, actors, relations, and processes that make inferences explicit and enable systematic interrogation. From this standpoint, datafication pertains not only to sources but to historical knowledge itself, understood as a set of claims that can be formalised without forfeiting their interpretive character.

Richard Ennals's theoretical contribution is particularly useful for clarifying the implications of this shift.

Even before the development of generative AI tools, Ennals had identified the need to make historians' cognitive practices explicit so that they could be supported by computational systems. His model—articulated around intelligent information retrieval, modelling and simulation, and explanation—did not aim to replace historians with the machine. Rather, it sought to create the conditions under which the computer could operate on structured representations of historical knowledge. In this sense, computation is not an end in itself but a means of rendering interrogable questions, criteria, and relations that often remain tacit within historiographical discourse.

The introduction of generative artificial intelligence systems into historical work makes this need even more pressing. LLMs can produce coherent and historically plausible texts, yet they operate on statistical regularities learned from predominantly web-based corpora, which rarely include structured archival sources or specialist historiographical literature. The result is a form of simulated knowledge—plausible but epistemologically fragile—that risks being mistaken for historical knowledge precisely because it lacks a clear distinction between source, interpretation, and automated text generation.

To clarify: If we prompt ChatGPT the question "What was Louis (of France) XVI's last wish?", the outcome is that the King asked to meet his family.

The day before his execution, Louis XVI asked to see his family for a final farewell. He was allowed to meet with his wife, Marie Antoinette, and their children. The meeting was extremely emotional and painful, with the family embracing and weeping together. Louis XVI tried to comfort his family, urging them to be brave and to trust in divine providence.

In addition to this, Louis XVI expressed the desire to confess and receive communion as part of his spiritual preparation for death. He also asked to have his chaplain, Abbé Edgeworth de Firmont, by his side to provide spiritual support until the final moment.
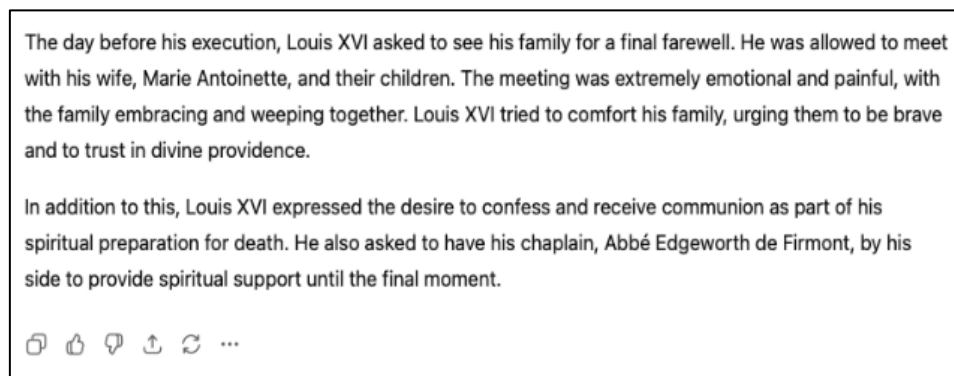
Figure 1. Outcome about Luis XVI's last wish.

Statistics say that persons sentenced to death, in most cases, ask to meet with family members. From this point of view, based on statistics, the AI necessarily responds that the King of France asked to meet his wife, Marie Antoinette. No one will ever know whether this was his last wish, but, certainly, through reading Henry Essex Edgeworth de Firmont's work (Firmont, 1815), historians know that King Louis, after his arrest, never met Marie Antoinette again.

As we can assume, in the absence of adequate historical data, "AI-open-access" infrastructures, the use of this technology thus risks producing a new opacity rather than greater transparency.

Nevertheless, experiments conducted in the context of building slavery databases—as described by Spina (Spina, 2025b)—show with particular clarity that AI can become a powerful instrument only if embedded within a computational workflow designed with methodological rigour.

On the other side, Handwritten Text Recognition technologies such as Transkribus and eScriptorium can help historians transform handwritten documents into digital text (datafication)—and this is only the first step in a broader pipeline that must include named-entity extraction, ontology design, data structuring in interoperable formats, and relation-building. Within this process, Named Entity Recognition tools and generative models can accelerate the extraction and organisation of information, but they require sustained human oversight: Historians must design prompts, verify outputs, and correct inevitable distortions (González-Gallardo et al., 2023).

Dataset construction thus emerges as an epistemically crucial moment, because the categories, relations, and structures adopted embody specific interpretive choices. The CSV file, the database, or the ontology are not neutral containers; they are forms of formalised historiography in which historical knowledge is translated into computational structures. In this way, digitisation oriented toward computation makes explicit what often remains implicit in discursive writing, fostering greater transparency and contestability of historical inferences.

In conclusion, Digital History stands today before a decisive epistemological choice. Limiting oneself to documentary digitisation and to a superficial use of artificial intelligence amounts to renouncing a deep transformation of the discipline. By contrast, taking computation seriously entails rethinking the relationship among sources, data, and interpretation, accepting both the risk and the opportunity of rendering historical knowledge partially formalizable. From this perspective, digitisation is not an end in itself but a process oriented toward building computational infrastructures—see, for example, the SESHAT[3] database (Francois et al., 2016; Turchin, Whitehouse, et al., 2018; Turchin et al., 2021; 2023; 2015)—that make it possible to interrogate the past in new ways, without relinquishing the hermeneutic complexity that characterises historians' craft. Only under these conditions can Digital History be configured as a genuine methodological frontier, capable of redefining historical work in the *i*Age of artificial intelligence systems.

Conclusion

Digital History now stands at a decisive crossroads, one that compels historians to choose between two fundamentally different futures for the discipline. On the one hand lies a path in which digitisation remains largely documentary, additive, and instrumental, enhancing access to sources and accelerating established scholarly practices without altering the epistemological foundations of historical knowledge. On the other hand lies a more demanding but intellectually generative path, in which digitisation is reconceived as a process oriented toward computation, requiring the formalisation of historical knowledge, the explicit modelling of interpretive choices,

---

[3] Link to the website: https://seshatdatabank.info.

and the construction of infrastructures capable of supporting meaningful interaction between historians and machines. The analyses developed in this article suggest that only the latter path can address the profound transformations introduced by digital environments and artificial intelligence systems and do so without sacrificing the interpretive and critical dimensions that define historical inquiry. The history of slavery, examined here as a privileged field of experimentation, makes this choice particularly visible because it exposes both the limitations of traditional documentary digitisation and the epistemic potential of computational approaches grounded in seriality, data modelling, and infrastructural thinking.

The proliferation of digital archives and collections devoted to slavery has undeniably transformed access to sources and contributed to the recovery of voices and experiences long marginalised within historical narratives. Yet access alone does not suffice to transform knowledge. When digital archives merely reproduce the logic of analogue repositories, the computer remains confined to a logistical role, while interpretation continues to reside exclusively in the historian's discursive practice. In such cases, digitisation increases volume and speed but leaves untouched the implicit nature of historical reasoning, the opacity of inferential steps, and the difficulty of systematic comparison across large corpora. Invented archives, understood not as arbitrary constructions but as deliberately designed epistemic environments, point toward an alternative model. By aggregating dispersed documentation into structured datasets and by enforcing explicit categories, relations, and constraints, they reactivate the logic of *histoire sérielle* in a computational register, transforming seriality from a heuristic intuition into an operational framework.

Reframing *histoire sérielle* in this way reveals its enduring relevance for contemporary Digital History. Far from being a relic of quantitative history, seriality emerges as a foundational epistemology for computation, one that privileges repetition, comparability, and structure as conditions of intelligibility. In digital environments, seriality is no longer merely inferred by the historian through prolonged engagement with sources, but materially instantiated in databases, schemas, and ontologies. This materialisation forces interpretive decisions to be made explicit and renders them open to critique, revision, and reuse. The construction of datasets, therefore, becomes a central epistemic act, not a technical afterthought. Categories, relationships, and data models encode historical assumptions and theoretical commitments, transforming files, tables, and ontologies into forms of formalised historiography.

The emergence of artificial intelligence, and particularly of generative language models, intensifies the urgency of this epistemological shift. AI systems do not possess historical understanding in any meaningful disciplinary sense; they operate by detecting statistical regularities within textual corpora and generating plausible continuations. Without access to structured historical data and without explicit modelling of sources, interpretations, and uncertainties, their outputs risk collapsing distinctions that are foundational to historical scholarship. The danger, therefore, is not that AI will replace historians, but that it will produce a new form of epistemic opacity, in which authoritative-looking narratives circulate without traceable evidentiary foundations. This risk can only be mitigated by embedding AI within computation-oriented workflows, where its role is clearly delimited and subordinated to infrastructures designed for transparency, verification, and critical control.

The relevance of Richard Ennals's theoretical insights becomes particularly clear in this context. His insistence that historians must describe what they do for computers to assist them remains a powerful corrective to both technological determinism and disciplinary defensiveness. Computation cannot operate on tacit knowledge alone; it requires explicit representations of questions, criteria, and inferential moves. This requirement does not imply the reduction of historical knowledge to formal logic or numerical abstraction, but

rather the partial externalisation of cognitive practices that have long remained implicit. By making these practices explicit, computation-oriented digitisation enhances, rather than diminishes, the reflexivity of historical research.

Ultimately, the challenge facing Digital History is not technological but epistemological. The tools required for computation-oriented digitisation already exist, from text recognition systems and entity extraction tools to database platforms and ontology frameworks. What remains uncertain is whether historians are willing to embrace the methodological consequences of using them seriously. Doing so entails accepting that historical knowledge can be partially formalised, that interpretive choices must sometimes be encoded, and that scholarly authority must be exercised not only through narrative eloquence but through the design of models, datasets, and infrastructures. This shift does not eliminate interpretation; it redistributes it across new material and conceptual forms.

If Digital History is to become more than a technologically enhanced version of traditional historiography, it must fully confront the implications of computation. Digitisation, in this stronger sense, is not an endpoint but a means of constructing environments in which historical knowledge can be interrogated, compared, and challenged at scales and levels of complexity previously unattainable. The history of slavery demonstrates both the necessity and the promise of this approach, showing how computational seriality can illuminate structures of domination and resistance that remain invisible within isolated documents. By orienting digitisation toward computation, historians can transform digital technologies from passive tools into active partners in critical inquiry, ensuring that the discipline remains both epistemologically rigorous and socially relevant in the age of artificial intelligence.

References

Ahnert, R., Ahnert, S. E., Ryan, Y., & Computational Humanities Research (CHR). (2020). Networking archives: Quantitative history and the contingent archive. In *Proceedings of the Workshop on Computational Humanities Research (CHR 2020)* (pp. 397-419). Queen Mary University of London. Retrieved from https://qmro.qmul.ac.uk/xmlui/handle/123456789/68412

Braudel, F. (2003). Scritti sulla storia. *Bompiani.* Retrieved from http://www.bompiani.it/catalogo/scritti-sulla-storia-9788845254123

Braudel, F. (2015). *Storia, misura del mondo*. New York: Il Mulino.

Cavalli-Sforza, L. L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton: Princeton University Press. Retrieved from https://www.jstor.org/stable/10.2307/j.ctvx5wbt8

Chabod, F. (1969). *Lezioni Di Metodo Storico*. (L. Firpo, Ed.). Rome: Laterza.

Chaunu, P. (1959). *Séville et l'Atlantique, 1504-1650*. Paris: SEVPEN.

Chaunu, P. (1964). *Histoire quantitative ou histoire serielle*. Malakoff: Armand Colin.

Darcy, R., & Rohrs, R. C. (1995). *A guide to quantitative history*. San Francisco: Praeger.

Derosas, R. (1989a). A database for the study of the Italian population registers. *Historical Social Research, 14*(4), 59-65

Derosas, R. (1989b). Un Database per Lo Studio Delle Fonti Anagrafiche. *Bollettino Di Demografia Storica, 9*, 89-93.

Dollar, C. M., & Jensen, R. J. (1974). *Historian's guide to statistics; quantitative analysis and historical research*. Malabar: R. E. Krieger Pub. Co.

Ennals, R. (1985). *Artificial intelligence: Applications to logical reasoning and historical research*. New York: Ellis Horwood.

Erwin, B. (November 6, 2020). Digital tools for studying empire: Transcription and text analysis with transkribus. *Not Even Past*. Retrieved from https://notevenpast.org/digital-tools-for-studying-empire-transcription-and-text-analysis-with-transkribus/

Firmont, H. E. E. (1815). *Mémoires de M. l'abbé Edgeworth de Firmont, dernier confesseur de Louis XVI*. Paris: Gide.

Fogel, R. W. (1975). The limits of quantitative methods in history. *The American Historical Review, 80*(2), 329-350. Retrieved from https://doi.org/10.2307/1850498

Francois, P., Manning, J. G., Whitehouse, H., Brennan, R., Currie, T., Feeney, K., & Turchin, P. (2016). A macroscope for global history. Seshat global history databank: A methodological overview. *Digital Humanities Quarterly, 10*(4). Retrieved from https://researchprofiles.herts.ac.uk/en/publications/a-macroscope-for-global-history-seshat-global-history-databank-a-

Franzosi, R. (2017). A third road to the past? Historical scholarship in the age of big data. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 50*(4), 227-244. Retrieved from https://doi.org/10.1080/01615440.2017.1361879

Furet, F. (1968). Sur quelques problèmes posés par le développement de l'histoire quantitative. *Social Science Information, 7*(1), 71-82.

Furet, F. (1971). L'histoire quantitative et la construction du fait historique. *Annales. Histoire, Sciences Sociales, 26*(1), 63-75.

Galasso, G. (2010). *Nient'altro che storia: Saggi di teoria e metodologia della storia*. New York: Società editrice il Mulino. Retrieved from https://doi.org/10.978.8815/144676

González-Gallardo, C.-E., Boros, E., Girdhar, N., Hamdi, A., Moreno, J. G., & Doucet, A. (2023). Yes but can ChatGPT identify entities in historical documents? Retrieved from https://doi.org/10.48550/arXiv.2303.17322

Itzcovich, O. (1989). Lo Storico e Il Database. *Quaderni Storici, 24*(70(1)), 321-325.

Kahle, P., Colutto, S., Hackl, G., & Mühlberger, G. (November 2017). Transkribus. A service platform for transcription, recognition and retrieval of historical documents. In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (pp. 19-24). Kyoto, Japan. Retrieved from https://doi.org/10.1109/ICDAR.2017.307

Le Roy Ladurie, E. (1968). La Fin Des Érudits. L'historien de Demain Sera Programmeur Ou Ne Sera Pas. *Nouvel Observateur*.

Le Roy Ladurie, E. (1973). *Le territoire de l'historien*. Rezé: Gallimard.

Lemercier, C., & Zalc, C. (2008). Méthodes Quantitatives Pour l'historien. *La Découverte.* Retrieved from http://www.quanti.ihmc.ens.fr/-Le-Reperes-.html

Maleki, N., Padmanabhan, B., & Dutta, K. (June 2024). AI hallucinations: A misnomer worth clarifying. In *2024 IEEE Conference on Artificial Intelligence (CAI)* (pp. 133-138). Singapore, Singapore. Retrieved from https://doi.org/10.1109/CAI59869.2024.00033

McCarthy, J., Rochester, N., Minsky, M. L., & Shannon, C. E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine, 27*(4), 12-14.

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, … Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science, 331*(6014), 176-182. Retrieved from https://doi.org/10.1126/science.1199644

Milioni, N. (2020). Automatic transcription of historical documents. Transkribus as a tool for libraries, archives and scholars. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-412565

Morrissey, R. M. (2015). Archives of connection. *Historical Methods: A Journal of Quantitative and Interdisciplinary History, 48*(2), 67-79. Retrieved from https://doi.org/10.1080/01615440.2014.962208

Muehlberger, G., Seaward, L., Terras, M., Oliveira, S. A., Bosch, V., Bryan, M., … Zagoris, K. (2019). Transforming scholarship in the archives through handwritten text recognition. Transkribus as a case study. *Journal of Documentation, 75*(5), 954-976. Retrieved from https://doi.org/10.1108/JD-07-2018-0114

Orlandi, T. (1996). Formalizzazione Dei Dati, Semiotica e Comunicazione. *Archeologia e Calcolatori, 7*, 1247-1258.

Orlandi, T. (1998). Formal methods in the humanities and their teaching. The future of the humanities in the digital age, University of Bergen, September 25.

Orlandi, T. (1999). Linguistica, sistemi, e modelli. *Il ruolo del modello nella scienza e nel sapere (Roma, 27-28 Ottobre 1998)*. Accademia Nazionale dei Lincei. Retrieved from http://www.cmcl.it/~orlandi/pubbli/modello.html

Orlandi, T. (2007). Teoria e prassi di una edizione computazionale. In A. Ciula and F. Stella (Eds.), *Digital philology and medieval texts. Proceedings of the Arezzo Seminar 2006, 19-21 January*, Vol. 4. Arti Spazi Scritture. Filologia latina medievale e umanistica. Pacini.

Orlandi, T. (2010). *Informatica testuale, teoria e prassi*. Rome: Laterza.

Romein, C. A., Kırmızıaltın, S., Reshef, R., Schneider, C., Agostini, G., Charlton, A., … Wilbrink, H. (2025). From research proposal to project management. A guide from the transkribus community on planning and executing workflows for researchers and GLAM-professionals. *International Journal of Digital Humanities, 7*(1). Retrieved from https://doi.org/10.1007/s42803-025-00107-7

Rosenzweig, R. (2001). The road to Xanadu: Public and private pathways on the history web. *Journal of American History, 88*(2), 548-579.

Rosenzweig, R. (2003). Scarcity or abundance? Preserving the past in a digital era. *American Historical Review, 108*(3), 735-762.

Ryan, Y., Ahnert, S. E., & Ahnert, R. (2020). Networking archives: Quantitative history and the contingent archive. In *Proceedings of the Workshop on Computational Humanities Research (CHR 2020)* (pp. 385-396). Queen Mary University of London. Retrieved from https://qmro.qmul.ac.uk/xmlui/handle/123456789/68412

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Artificial intelligence hallucinations. *Critical Care, 27*(1), Article 180. Retrieved from https://doi.org/10.1186/s13054-023-04473-y

Seaward, L., & Kallio, M. (2017). Transkribus. Handwritten text recognition technology for historical documents. Paper presented at *ADHO 2017—Montréal*. Retrieved from https://dh-abstracts.library.cmu.edu/works/4193

Spina, S. (2020). *Archivi nell'era delle Digital Humanities, dei Big Data e della Genetica*. Viagrande: Algra.

Spina, S. (2021). Datificazione Delle Fonti Storiche per La Digital History Delle Pandemie. *Umanistica Digitale*, (10), 1-18.

Spina, S. (2022a). *Digital history. Metodologie informatiche per la ricerca storica*. Napoli: Edizioni Scientifiche Italiane.

Spina, S. (2022b). Historical network analysis & htr tool. Per un approccio storico metodologico digitale all'archivio Biscari di Catania. *Umanistica Digitale*, (14), 163-181. Retrieved from https://doi.org/10.6092/issn.2532-8816/15159

Spina, S. (2023a). Artificial intelligence in archival and historical scholarship workflow: HTR and ChatGPT. *Umanistica Digitale*, ahead of print. Retrieved from https://doi.org/10.48550/arXiv.2308.02044

Spina, S. (2023b). Handwritten text recognition as a digital perspective of archival science. *Aidainformazioni*, (1-2), 115-132.

Spina, S. (2023c). Homo-loggatus. The anthropological condition of historians in the digital world. *Journal of Mathematical Techniques and Computational Mathematics, 2*(10), 431-437.

Spina, S. (2024). Digitality as a Longue Durée Historical Phenomenon. *Umanistica Digitale*, (18), 1-25.

Spina, S. (2025a). Invented archives. Un Problema Teoretico per La Ricerca Storica o per l'archivistica? *Digitalia*, (2), 203-226.

Spina, S. (2025b). Schiavitù e Invented Archives. Intelligenza Artificiale Generativa Nella Costruzione Del Database on the Slave Trade. *AIDAinformazioni, 1*(2), 85-119.

Turchin, P., Brennan, R., Currie, T., Feeney, K., Francois, P., Hoyer, D., … Whitehouse, H. (2015). SESHAT: The global history databank. *Cliodynamics, 6*(1), 77-107. Retrieved from https://doi.org/10.21237/C7clio6127917

Turchin, P., Currie, T., Collins, C., Levine, J., Oyebamiji, O., Edwards, N. R., … Whitehouse, H. (2021). An integrative approach to estimating productivity in past societies using SESHAT: Global history databank. *The Holocene, 31*(6), 1055-1065. Retrieved from https://doi.org/10.1177/0959683621994644

Turchin, P., Currie, T. E., Whitehouse, H., François, P., Feeney, K., Mullins, D., … Spencer, C. (2018). Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization. *Proceedings of the National Academy of Sciences, 115*(2), E144-151. Retrieved from https://doi.org/10.1073/pnas.1708800115

Turchin, P., Whitehouse, H., François, P., Hoyer, D., Alves, A., Baines, J., … Xie, L. (2018). An introduction to SESHAT: Global history databank. *Journal of Cognitive Historiography, 5*(1-2), 115-123. Retrieved from https://doi.org/10.1558/jch.39395

Turchin, P., Whitehouse, H., Larson, J., Cioni, E., Reddish, J., Hoyer, D., … François, P. (2023). Explaining the rise of moralizing religions: A test of competing hypotheses using the SESHAT databank. *Religion, Brain & Behavior, 13*(2), 167-194. Retrieved from https://doi.org/10.1080/2153599X.2022.2065345

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, LIX*(236), 433-460. Retrieved from https://doi.org/10.1093/mind/LIX.236.433