



## OPEN ACCESS

## EDITED BY

Heidi Kloos,  
University of Cincinnati, United States

## REVIEWED BY

Heidi Kloos,  
University of Cincinnati, United States  
Werner Scheltjens,  
University of Bamberg, Germany  
Lajos Somogyvari,  
University of Pannonia, Hungary

## \*CORRESPONDENCE

Nikhil Gangwar  
✉ [nikhil.gangwar90@gmail.com](mailto:nikhil.gangwar90@gmail.com)

RECEIVED 15 June 2025

ACCEPTED 16 September 2025

PUBLISHED 30 September 2025

## CITATION

Gangwar N (2025) Historians: from manuscript to machine learning.  
*Front. Educ.* 10:1647282.  
doi: 10.3389/feduc.2025.1647282

## COPYRIGHT

© 2025 Gangwar. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Historians: from manuscript to machine learning

Nikhil Gangwar \*

Department of History, Daulat Ram College, New Delhi, India

## KEYWORDS

digital humanities, machine learning, artificial intelligence, digital innovation (DI), technological innovation, historians and archivists

## 1 Introduction

Historians have consistently adapted their methodologies to the available technological tools, progressing from manual examination of manuscripts to the use of advanced machine learning-based instruments. Early modern historians employed manual techniques to create, collect, and analyze handwritten manuscripts and literary sources found in archives, libraries, private collections, and government records, thereby reconstructing historical events. The invention of the printing press in the fifteenth century fundamentally transformed historical study and writing. Subsequent innovations, including typewriters and microfilm, significantly altered the processes of recording, preserving, and disseminating historical information.

In the late twentieth century, the advent of computers initiated a digital revolution that redefined the sources, techniques, and epistemologies of historical research. Computers allowed historians to organize and analyze source materials with unprecedented precision (Alves, 2014). Recent advances in machine learning have further accelerated this transformation, enabling the analysis of large-scale digital data. This study examines the evolving relationship between technological innovation and historical practice, from handwritten manuscripts to machine learning. The findings contribute to a deeper understanding of how technological advancements shape historical thought and methodology.

## 2 Sub-sections

### 2.1 Pre digital age

#### 2.1.1 Printing press

Before the advent of the printing press, historians manually copied manuscripts, which limited access to historical documents and confined research to a select, elite audience primarily focused on religious subjects. For instance, the Venerable Bede's monastic chronicles reinforced prevailing religious views, while the Epic of Gilgamesh demonstrated the scarcity and variety of preserved accounts. Consequently, historical narratives largely described and supported established beliefs, with minimal critical analysis. The arrival of the printing press democratized access to books and significantly changed the nature and scope of historical scholarship (Curtius, 1990).

The printing press expanded historical research by facilitating the production of sources. While handwritten documents were exclusive, printed works became more affordable and accurate due to reduced errors (Eisenstein, 1980). This change promoted secular and comparative perspectives. Increased accessibility enabled historians like Edward Gibbon (Gibbon et al., 1776) and François Guizot (Cave, 2000) to approach history with greater analytical rigor, fostering the development of professional historiography.

Print technology improved source criticism and citation. Scholars debated document reliability and separated authentic records from forgeries. Print made critical analysis more accessible, as with the work of Scaliger and Bodin (Grafton, 1993). Texts by Bede and Biondo inspired interest in classical narratives (Kelley, 1998). The scale of book production led to the establishment of professional historians and archives, despite some resistance. These innovations set the stage for the development of scientific historiography.

Despite enhanced accessibility and methodological advancement achieved through print technology, the exclusion of diverse groups persisted. Marginalized perspectives—including those of women, indigenous peoples, and colonial subjects—were frequently omitted. This exclusion highlights the continuing imperative for inclusivity in historical scholarship. Published narratives that foregrounded established viewpoints often reflected prevailing social biases, thereby limiting the representation of alternative voices. However, collective efforts gradually increased the visibility of some previously marginalized accounts, such as those associated with slaves, colonial subjects, and the women's suffrage movement.

When more records were printed and standardized, historians could compare the same editions more easily. This improvement led to enhanced accuracy and consistency in research (Eisenstein, 1980; McKeown, 2021). Evidence-based analysis and reliability became more common. Earlier methods, such as oral traditions, still offer valuable information that print may miss. The print era introduced new tools, such as pagination and indexes, which made research more systematic and efficient. These developments created the foundation for today's critical approaches to history (Serjeantson, 2006).

### 2.1.2 Typewriter

The typewriter and microfilm improved efficiency and helped establish archival standards. The widespread adoption of the typewriter allowed for faster dissemination and clearer presentation of scholarly ideas, influencing both the teaching and understanding of history (Adler, 2023).

The typewriter's mechanical process increased productivity by generating legible text more quickly. Typed manuscripts enabled clearer peer evaluation and scrutiny, which strengthened scholarly standards (Roy, 2021). The introduction of the typewriter thus improved the speed, readability, and consistency of historical writing, further formalizing the field.

These technological developments redefined not only historians' workflows but also the perceived authority and objectivity of historical scholarship. For example, historians such as Charles Beard produced comprehensive analyses with greater efficiency, marking a transition in the economic interpretation of history.

Standardizing text with the typewriter encouraged professionalism and empirical study, but could obscure subjective interpretation. Documents seen as objective are influenced by historians' choices; narrative emphasis or selective quoting may sway interpretation. Thus, personal bias persists within seemingly objective writing.

However, in both colonial and postcolonial contexts, the standardization imposed by the typewriter often reinforced existing class and gender hierarchies (Fine, 1993; Chung, 2022). Access to typewriters was generally confined to privileged social strata due to its high cost (Gitelman, 2000) and their professional use frequently mirrored prevailing power dynamics, notably in contexts where women were predominantly relegated to clerical positions (England and Boyer, 2009).

In addition to technological developments, the social norms associated with writing evolved in response to the typewriter. Initially, mechanical writing was perceived as less personal than handwriting, which was seen as conveying sincerity and refinement. Over time, the typewritten word became a marker of professionalism and formal expertise. This evolution illustrates that historical knowledge relies not only on empirical evidence but also on the changing relationship between scholars and their tools. The typewriter fostered an aura of objectivity and authority, while its standardization sometimes obscured interpretive subjectivity, thereby shaping the boundaries of historiographical practice.

## 2.2 Early digital age

### 2.2.1 Microfilm

In the early twentieth century, microfilm technology, a precursor to mass digitization (Milligan, 2022), fundamentally restructured preservation and access to sources (Binkley, 1936). Developed by John Dancer (Luther, 1959) and further improved by Rene Dagrón (Luther, 1996), microfilm reduced the size of materials for efficient storage and aimed to stabilize fragile documents. It played a crucial role in debates concerning archival access, as it enabled unprecedented opportunities for historians to access rare and geographically dispersed materials.

Microfilm enabled researchers to access documents without the need for travel, making rare records more widely available. In the 1930s, institutions began microfilming projects to increase global access to manuscripts (Foster, 1985).

Microfilm and foundational digital practices greatly expanded access and transformed historical research. By making previously inaccessible materials available, microfilm widened research possibilities. Digitization amplified this shift, enabling analysis on a larger scale with greater complexity. Contemporary digitization, with frameworks like International Image Interoperability Framework (IIIF), prioritizes metadata to systematically enhance access and engagement. These developments reshaped historical research by supporting data-driven methods and highlighting technology as an active force in the evolution of the discipline.

Nevertheless, despite these advancements in access, microfilm was mainly limited to select institutions and required specialized equipment. Navigating collections was slow, highlighting the eventual need for digital transformation. Digital databases later addressed these issues by simplifying document search and retrieval, vastly improving research efficiency and access. However, this phase of digitization is not without its challenges. Issues such as digital preservation, data loss risks, and access inequalities have emerged, prompting ongoing discussions on

how to ensure equitable access while securing digital records for future generations.

## 2.3 Digital age

Punch cards (early data storage cards) revolutionized data storage and enabled quantitative analysis in the 1950s, leading historians and social scientists to adopt computational methods. Roberto Busa pioneered the application of digital technology in the humanities with the creation of Index Thomisticus, a computational analysis of Thomas Aquinas's works (Fantoli, 2023; Sula and Hill, 2019; Rockwell and Passarotti, 2019). These early innovations marked the beginning of a profound transformation: digital technologies would fundamentally reshape how historians access information, conduct research, and interpret the past.

Following these foundational developments, mainframe computers—such as the IBM 1,401 and UNIVAC—enabled large-scale data analysis. This advance sparked the cliometric revolution, which used quantitative, or numerical, techniques in the study of history. Tools like SPSS, a statistical analysis program, let scholars reinterpret historical narratives using empirical, data-driven approaches. Owsley (1990) used quantitative methods to study Southern social patterns. Building on this, from the 1960s, computers reshaped global historical research. Researchers such as Robert Fogel, Stanley Engerman, and Emmanuel Le Roy Ladurie applied digital techniques to historical records. Since 2000, digital libraries—collections of digitized resources online—have further enhanced access and research methods.

As the field advanced, historians witnessed transformative effects. Technology provided access to millions of documents and artifacts online, enabling remote research (Gomes and Costa, 2014). Geographic Information Systems (GIS) and data mining uncover patterns and encourage collaboration with programmers. Notably, the Valley of the Shadow project (1993) applies GIS to examine two communities during the American Civil War (Thomas and Ayers, 2003). Interactive media and digital storytelling—exemplified by the Digital History Project—bolster public engagement and revitalize historical narratives.

This shift also unfolded in phases as computers integrated into historical research. In the initial stage, historians converted source information into structured datasets by encoding facts, dates, names, and events from documents, such as census records, into databases. For instance, population studies digitized census data for easier analysis.

With digital foundations established, historians entered a second phase, where they analyzed datasets, identified patterns, tested hypotheses, and uncovered language changes using linguistic software (Hendrickx and Marquilhaes, 2011). Computers produced new insights—graphs, tables, dictionaries. These methods fostered specialized fields.

As these methods became central, digital technologies have significantly altered the nature, accessibility, and volume of historical sources. As Timothy Hitchcock states, “The digital archive has changed the nature of the historical record”

(DHLU 2013 Symposium (Luxembourg, 5 December 2013)—Keynote: Tim Hitchcock-CVCE Website, 2025). Previously inaccessible manuscripts, newspapers, and records are now available online, allowing historians to use vast datasets. Projects like the Old Bailey Online exemplify how digitisation reshapes the materiality of sources by offering searchable access to 197,000 trials, a scale previously unimaginable (Old Bailey Proceedings Online, 2025). Similarly, the Transcribe Bentham project demonstrates how collaborative digital initiatives have expanded participation in both the creation of and access to historical documents; volunteers transcribe manuscripts, making primary sources widely available and accelerating research progress (Causser and Terras, 2014).

Despite enhanced access, the use of digital surrogates, while offering accessibility, raises questions about authenticity and the material context of original documents. As Johanna Drucker notes, “Digitized materials are not facsimiles but interpretations” (Drucker, 2013), highlighting how the transformation of physical sources into digital formats introduces a new layer of mediation.

Alongside these challenges, technologies influence methods and methodology, especially through digital humanities tools for quantitative analysis, text mining, and GIS mapping. For example, historians use Voyant Tools for textual analysis (Bradley, 2012; Nyhan et al., 2023) and GIS for spatial history (Da Silveira, 2014). Yet, Ian Milligan argues the “methodological shift is secondary to the fundamental change in the source base itself” (Milligan, 2019). Methodology changes in response to the transformed materiality and scale of digital sources, not as a standalone transformation. Digital tools enable quantitative analysis on an unprecedented scale. Network analysis and visualization tools offer new ways to examine patterns, relationships, and trends across vast corpora. For example, the “Mining the Dispatch” project utilizes topic modeling to explore themes in Civil War newspapers, uncovering insights that are difficult to find manually. Digital history also encourages interdisciplinary methods, blending historical inquiry with data science, linguistics, and geography. This convergence challenges traditional methodology, pushing historians to engage with new epistemologies and reconsider notions of evidence, interpretation, and narrative.

This global digital progression exhibits stark regional contrasts, reflecting differences in priorities, strategies, and challenges. In India, digitization efforts focus on preserving diverse community records in multiple languages, aiming to address the country's significant linguistic diversity; however, these efforts are often constrained by limited infrastructure and resources (Chowdhary, 2024). In Southeast Asia, by contrast, state-run projects in countries like Vietnam and Indonesia often prioritize official narratives, restricting alternative or community perspectives. Latin America offers a contrasting, community-driven approach: projects in Peru and Argentina, such as Archivo Memoria Abierta, prioritize indigenous and social histories, representing a “bottom-up” model distinct from the state-centered strategies seen elsewhere.

Kenya and other African countries present another contrast: their digitization initiatives emphasize building digital infrastructure and safeguarding historical materials. Despite limited funding, these efforts expand access and focus on long-term preservation. For example, the Digitization of Africa's Cultural Heritage (DACH) project targets the preservation of

Kenyan archives for future generations. This strategy, centered on infrastructure and heritage protection, contrasts with regions that prioritize content diversity or dominant narratives (Ndegwa et al., 2022; Aldirdiri, 2024; Musembe et al., 2025).

Europe and North America provide a further contrast, characterized by robust infrastructure, advanced analytics, and handling of complex datasets. These regions typically prioritize major languages, sometimes at the expense of less widely spoken ones (Ahnert et al., 2023). Their strategies focus on maximizing analytical capability and data access, but risks persist: lost context, technological obsolescence, and training gaps remain, necessitating sustained collaboration. Juxtaposing these regions with others reveals how policy orientation, language priorities, and resource allocation create distinctly state-driven, community-centered, or analytics-focused approaches to digital history.

Despite differences in approach, these regional digitization strategies have a direct impact on historical research by shaping the accessibility and interpretation of information. Advanced infrastructure enables high-level analysis and diverse narratives, whereas areas with resource constraints may face challenges in accessing and preserving information. This global diversity enriches historical research and showcases regional adaptation to technological opportunities.

In addition to research practices, digital media have also transformed the communication of historical research. Open-access platforms now reach broader audiences, while projects feature interactive maps, timelines, and videos. Social media and online forums expedite scholarly communication and foster academic communities.

These global developments demonstrate that digital methods now define historical research by reshaping sources, interpretation, and engagement. Technology's influence is not just technical but methodological, fundamentally changing how historians access information and construct knowledge about the past. Ongoing adaptation to technological shifts is crucial, as the transformative impact of digital tools remains central to the historian's practice.

## 2.4 Machine learning era

Machine learning (ML) marks a turning point in historical scholarship, enabling historians to process vast datasets, discover patterns, and classify records at scale. This transforms both the scope and limits of analysis, signaling a deeper shift in how historical evidence and interpretation are shaped in the digital era.

ML introduces new approaches to historical study. These ML based tools enable large-scale tests of narrative patterns and the exploration of how historical accounts are constructed, a task once considered challenging. For example, custom language models—computer programs trained to understand, generate, or analyze language, often based on advanced neural network architectures—can generate multiple versions of the same event, allowing researchers to study the relationships between different genres and interpretations.

Projects like the Sphaera initiative and the Sacrobosco Collection utilize deep learning models—these networks of

interconnected data processing layers—for clustering historical illustrations, comparing astronomical tables, or identifying temporal-geographic trends in scientific knowledge (Valleriani et al., 2023; Eberle et al., 2024; Zamani et al., 2023). By employing these approaches, historians have uncovered subtle trends, such as the convergence of scientific communities during periods of political and religious division—insights that are nearly invisible through traditional analysis.

Another striking example is the use of deep learning models to decipher and date ancient inscriptions. Ithaca, a deep learning model trained on classical Greek epigraphy (the academic study of engraved inscriptions), can restore missing portions of inscriptions and propose alternative dating—sometimes contesting established interpretations and aligning with new historical breakthroughs (Assael et al., 2022). Deep learning models have also been applied to Latin, Akkadian, and Egyptian Hieroglyphic texts, enabling a detailed study of ancient languages once regarded as nearly inaccessible.

The Venice Time Machine project utilized machine learning to digitize centuries of records and map social networks, thereby creating digital models of past communities (Kaplan, 2020; Donovan, 2023). By tracing relationships in many documents, new social patterns became visible.

However, these tools also revealed the challenge of algorithmic inference, which can mix meaningful links with irrelevant ones. This relates to the machine learning “black box” problem and the risk of amplifying biases in sources (Kansteiner, 2022). These issues are evident when AI models, such as Google's Gemini, generate images of historical figures with inaccurate gender or ethnic features, for example, by depicting a medieval British king as a woman. Another example comes from a UNESCO report on large language models, which found that women are often associated with children and family topics, while men are linked to business and salary (UNESCO, IRCAL, 2024).

Algorithmic bias in current technologies threatens to distort the historical record, particularly by reinforcing imperial narratives that privilege colonizer perspectives and marginalize the voices of subaltern groups (Hovy and Prabhumoye, 2021; Luthra et al., 2024). Key datasets, particularly in colonized countries, often draw from digital archives that overrepresent European experiences and overlook local cultures, knowledge systems, and environmental traditions, thereby perpetuating data colonialism (Roberts and Montoya, 2023). British colonial ethnographies in India focus on events and knowledge deemed important by the British, omitting local species and marginalized languages (Baker, 2001). As machine learning increasingly relies on these biased archives, there is a growing risk that incomplete and skewed histories will shape our understanding of the past.

Bias in NLP intensifies during language translation, particularly for indigenous and minority groups preserving history in non-Western languages. Foundational dataset biases form the basis for these issues, which surface clearly in AI translation systems. These systems often strip away cultural meaning (Anik et al., 2025) resulting in Western concepts replacing non-Western traditions. For example, translating the Hindi term “*Dharm*” (a broad ethical concept) as merely “religion” imposes a Western view. Likewise, “*Atman*” (inner essence) becomes “spirit” or “mind.”



The mistranslation of “Achar” (meaning conduct) as “pickle” provides another example. Together, these distortions underscore how algorithmic bias can distort historical narratives, silence marginalized voices, and misrepresent cultural facts.

These examples underscore the importance of expert assessment when interpreting algorithmic outputs related to historical knowledge, particularly regarding marginalized perspectives and potential bias. Ethical considerations, Experts review and transparency are crucial for obtaining reliable findings from machine learning analyses of historical data. Historians must use their expertise to evaluate results and build valid narratives. Historians’ expertise in contextual analysis, variable selection, and source evaluation finds parallels in machine learning practices. Algorithms alone cannot determine the significance or context of associations. The historian’s role is evolving to include the use of digital tools, cross-referencing diverse sources, and including oral histories for a more comprehensive, multi-perspective view. Historians’ facility for understanding the interplay of factors that shape events complements the analytical strengths of machine learning methods.

Despite its promise, the adoption of machine learning is hindered by technical requirements, skepticism about interpretive reliability, and resource constraints. These factors highlight the ongoing need for critical reflection on the impact of technology and questions about interpretation and validity. Historians benefit from machine learning while remaining mindful of its limitations.

### 3 Discussion

Each major technological advance—from manuscripts to machine learning—has fundamentally transformed historical scholarship. The text’s argument is that every phase alters production, preservation, and interpretation of historical knowledge, incorporating subjectivity and bias—whether in manuscript margins or algorithms. This reinforces the need for a thorough, critical assessment with each new tool. As digital technologies uncover hidden patterns, new interpretive challenges emerge, demanding rigorous scholarly engagement. Technologies reshape the materiality of sources, their more profound and more lasting impact lies in transforming historical methods and methodology. By changing how historians collect, process, and interpret data, technologies are not merely tools but active agents in

redefining the practice of history itself. The conclusion is that open, interdisciplinary collaboration and critical analysis are essential for historians to responsibly address the ethical and methodological questions posed by ongoing technological change.

### Author contributions

NG: Writing – original draft, Writing – review & editing.

### Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

### Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

### Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Adler, M. H. (2023). *The Writing Machine: A History of the Typewriter*. 1st ed. Routledge: London. doi: 10.4324/9781003387480
- Ahnert, R., Griffin, E., Ridge, M., and Tolfo, G. (2023). *Collaborative Historical Research in the Age of Big Data: Lessons from an Interdisciplinary Project*. 1st ed. Cambridge: Cambridge University Press. doi: 10.1017/9781009175548
- Aldirdiri, O. (2024). Navigating the digital divide: challenges and opportunities in research publishing for African scholars. *Eur. Rev.* 32, S92–107. doi: 10.1017/S1062798724000073
- Alves, D. (2014). Guest editor’s introduction: digital methods and tools for historical research. *Int. J. Humanit. Arts Comput.* 8, 1–12. doi: 10.3366/ijhac.2014.0116
- Anik, M. A., Rahman, A., Wasi, A. T., and Ahsan, M. M. (2025). “Preserving cultural identity with context-aware translation through multi-agent AI systems,” in *Proceedings of the 1st Workshop on Language Models for Underserved Communities (LM4UC 2025)* (Albuquerque, New Mexico: Association for Computational Linguistics), 51–60. doi: 10.18653/v1/2025.lm4uc-1.7
- Assael, Y., Sommerschild, T., Shillingford, B., Bordbar, M., Pavlopoulos, J., Chatzipanagiotou, M., et al. (2022). Restoring and attributing ancient texts using deep neural networks. *Nature* 603, 280–283. doi: 10.1038/s41586-022-04448-z
- Baker, M. (2001). The politics of knowledge: the case of British colonial codification of “customary” irrigation practices in Kangra, India. *Himalaya* 21, 26–35. Available online at: <https://digitalcommons.maclester.edu/himalaya/vol21/iss2/7> (Accessed September 2, 2025).
- Binkley, R. C. (1936). *Manual on Methods of Reproducing Research Materials; a Survey Made for the Joint Committee on Materials for Research of the Social Science*

Research Council and the American Council of Learned Societies. Edwards. Available online at: <https://catalog.hathitrust.org/Record/001044388> (Accessed August 26, 2025).

Bradley, J. (2012). "No job for techies: technical contributions to research in the digital humanities," in *Collaborative Research in the Digital Humanities*, eds M. Deegan, and W. McCarty (Ashgate Publishing), 11–26. Available online at: [https://www.researchgate.net/profile/John-Bradley-8/publication/264886507\\_No\\_Job\\_for\\_Techies\\_Technical\\_contributions\\_to\\_research\\_in\\_the\\_Digital\\_Humanities/links/541957ae0cf203f155add6c/No-Job-for-Techies-Technical-contributions-to-research-in-the-Digital-Humanities.pdf](https://www.researchgate.net/profile/John-Bradley-8/publication/264886507_No_Job_for_Techies_Technical_contributions_to_research_in_the_Digital_Humanities/links/541957ae0cf203f155add6c/No-Job-for-Techies-Technical-contributions-to-research-in-the-Digital-Humanities.pdf) (Accessed September 2, 2025)

Causser, T., and Terras, M. (2014). Crowdsourcing bentham: beyond the traditional boundaries of academic history. *Int. J. Humanit. Arts Comput.* 8, 46–64. doi: 10.3366/ijhac.2014.0119

Cave, E. (2000). *François Pierre Guillaume Guizot: An Intellectual Approach* (Dissertations and Theses Paper 1467). Portland State University, Department of History, 2000, Portland. doi: 10.15760/etd.1466

Chowdhary, R. (2024). Shifting paradigms of multilingual publishing and scholarship in India. *J. Electron. Publ.* 27, 121–140. doi: 10.3998/jep.5592

Chung, Y. J. (2022). Typewriters and typists in republican China, 1910s–1940s: everyday technology and race, class and gender. *Inter-Asia Cult. Stud.* 23, 345–383. doi: 10.1080/14649373.2022.2108203

Curtius, E. R. (1990). *European Literature and the Latin Middle Ages*. Princeton: Princeton University Press.

Da Silveira, L. E. (2014). Geographic information systems and historical research: an appraisal. *Int. J. Humanit. Arts Comput.* 8, 28–45. doi: 10.3366/ijhac.2014.0118

DHLU 2013 Symposium (Luxembourg, 5 December 2013)—Keynote: Tim Hitchcock-CVCE Website (2025). Available online at: [https://www.cvce.eu/en/obj/dhlu\\_2013\\_symposium\\_luxembourg\\_5\\_december\\_2013\\_keynote\\_tim\\_hitchcock-en-137c4265-35c7-46c7-a542-5899ee8bfff8.html](https://www.cvce.eu/en/obj/dhlu_2013_symposium_luxembourg_5_december_2013_keynote_tim_hitchcock-en-137c4265-35c7-46c7-a542-5899ee8bfff8.html) (Accessed August 26, 2025).

Donovan, M. (2023). *How AI Is Helping Historians Better Understand Our Past*. Available online at: <https://www.technologyreview.com/2023/04/11/1071104/ai-helping-historians-analyze-past/> (Accessed April 11, 2023).

Drucker, J. (2013). "Performative materiality and theoretical approaches to interface," in *DHQ: Digital Humanities Quarterly* 7. Available online at: <https://dhq-static.digitalhumanities.org/pdf/000143.pdf>

Eberle, O., Büttner, J., El-Hajj, H., Montavon, G., Müller, K.-R., and Valleriani, M. (2024). Insightful analysis of historical sources at scales beyond human capabilities using unsupervised machine learning and XAI. *Sci. Adv.* 10:eadj1719. doi: 10.1126/sciadv.adj1719

Eisenstein, E. L. (1980). *The Printing Press as an Agent of Change*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107049963

England, K., and Boyer, K. (2009). Women's work: the feminization and shifting meanings of clerical work. *J. Soc. Hist.* 43, 307–340. doi: 10.1353/jsh.0.0284

Fantoli, M. (2023). *Steven E. Jones, Roberto Busa, S. J., and the Emergence of Humanities Computing: The Priest and the Punched Cards*, Routledge, London – New York 2016. *Mediterranea. International Journal on the Transfer of Knowledge*, 673–681. doi: 10.21071/mijtk.v8i-0.15563

Fine, L. M. (1993). Beyond the typewriter: gender, class, and the origins of modern american office work, 1900–1930. By Sharon Hartman strom urbana: university of illinois press 1992. Xvii + 427 Pp. Illustrations, tables, notes, and index. \$42.50. ISBN 0-252-01806-0. *Bus. Hist. Rev.* 67, 476–79. doi: 10.2307/3117374

Foster, C. (1985). Microfilming activities of the historical records survey, 1935–42. *Am. Archiv.* 48, 45–55. doi: 10.17723/aarc.48.1.605415455010j71q

Gibbon, E., Ley, J., Strahan, W., and Cadell, T. (1776). *The History of the Decline and Fall of the Roman Empire* by Edward Gibbon, Esq.; Volume the First[Sixth]. With Boston Public Library. London: Printed for W. Strahan and T. Cadell. Available online at: [http://archive.org/details/historyofdecline01gibb\\_0](http://archive.org/details/historyofdecline01gibb_0) (Accessed August 26, 2025).

Gitelman, L. (2000). *Scripts, Grooves, and Writing Machines: Representing Technology in the Edison Era*. Stanford: Stanford University Press. doi: 10.1515/9781503617353

Gomes, D., and Costa, M. (2014). The importance of web archives for humanities. *Int. J. Humanit. Arts Comput.* 8, 106–123. doi: 10.3366/ijhac.2014.0122

Grafton, A. (1993). *Joseph Scaliger: A Study in the History of Classical Scholarship*. Oxford University Press: Oxford. doi: 10.1093/oso/9780199206018.001.0001

Hendrickx, I., and Marquilhaes, R. (2011). From old texts to modern spellings: an experiment in automatic normalisation. *J. Lang. Technol. Comput. Linguist.* 26, 65–76. doi: 10.21248/jlcl.26.2011.147

Hovy, D., and Prabhumoye, S. (2021). Five sources of bias in natural language processing. *Lang. Linguist. Compass* 15, 1–19. doi: 10.1111/lnc3.12432

Kansteiner, W. (2022). Digital doping for historians: can history, memory, and historical theory be rendered artificially intelligent? *Hist. Theory* 61, 119–133. doi: 10.1111/hith.12282

Kaplan, F. (2020). "Big data of the past, from Venice to Europe," in *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems* (New York, NY, USA), ASPLOS'20, March 13, 1. doi: 10.1145/3373376.3380611

Kelley, D. R. (1998). *Faces of History: Historical Inquiry from Herodotus to Herder*. Yale: Yale University Press. Available online at: <https://www.jstor.org/stable/j.ctt32bs9h> (Accessed June 12, 2025).

Luther, F. (1959). *Microfilm: A History, 1839-1900*. National Microfilm Association. Available online at: <https://catalog.hathitrust.org/Record/001758639> (Accessed June 11, 2025).

Luther, F. (1996). René dagron, inventor of microfilm. *Hist. Photogr.* 20, 345–361. doi: 10.1080/03087298.1996.10443695

Luthra, M., Todorov, K., Jeurgens, C., and Colavizza, G. (2024). Unsilencing colonial archives via automated entity recognition. *J. Doc.* 80, 1080–1105. doi: 10.1108/JD-02-2022-0038

McKeown, M. (2021). *Erasmus: An Author Profile*. *CLT Journal*. Available online at: <https://blog.cltext.com/erasmus-author-profile/> (Accessed May 10, 2021).

Milligan, I. (2019). *History in the Age of Abundance?: How the Web Is Transforming Historical Research*. Montreal, QC: McGill-Queen's University Press. doi: 10.1515/9780773558212

Milligan, I. (2022). *The Transformation of Historical Research in the Digital Age*. 1st ed. Cambridge: Cambridge University Press. doi: 10.1017/9781009026055

Musembe, C. N., Kwanya, T., and Chweya, N. (2025). Enhancing digital access to Kenya's archival materials through virtual and augmented reality technologies. *Digit. Transform. Soc.* 4, 235–250. doi: 10.1108/DTS-08-2024-0148

Ndegwa, H., Bosire, E., and Odero, D. (2022). The status of the digital preservation policies and plans of the institutional repositories of selected public universities in Kenya. *Insights* 35, 1–13. doi: 10.1629/uksg.590

Nyhan, J., Rockwell, G., Sinclair, S., and Ortolja-Baird, A. eds. (2023). *On Making in the Digital Humanities: The Scholarship of Digital Humanities Development in Honour of John Bradley*. London: UCL Press. doi: 10.2307/j.ctv2wk727j

Old Bailey Proceedings Online (2025). *The Proceedings of the Old Bailey, 1674–1913*. Google Search. Available online : <https://www.oldbaileyonline.org/history/> (Accessed August 31, 2025).

Owsley, H. C. (1990). *Frank Lawrence Owsley: Historian of the Old South: a Memoir*. Vanderbilt University Press.

Roberts, J. S., and Montoya, L. N. (2023). "In consideration of indigenous data sovereignty: data mining as a colonial practice," in *Proceedings of the Future Technologies Conference (FTC) 2023. Volume 2*, ed. K. Arai (Lecture Notes in Networks and Systems; Springer Nature Switzerland). doi: 10.1007/978-3-031-47451-4\_13

Rockwell, G., and Passarotti, M. (2019). *The Index Thomisticus as a Big Data Project*. *Umanistica Digitale*.

Roy, S. C. (2021). Peer review process - its history and evolution. *Sci. Cult.* 87, 36–44. doi: 10.36094/sc.v87.2021.Peer\_Review\_Process.Roy.36

Serjeantson, R. W. (2006). "Proof and persuasion," in *The Cambridge History of Science*, 1st ed., eds. K. Park, and L. Daston. Cambridge: Cambridge University Press. doi: 10.1017/CHOL9780521572446.006

Sula, C. A., and Hill, H. V. (2019). The early history of digital humanities: an analysis of computers and the humanities (1966–2004) and literary and linguistic computing (1986–2004). *Digit. Scholarsh. Human.* 5:fqz072. doi: 10.1093/lilc/fqz072

Thomas, W. G., and Ayers, E. L. (2003). An overview: the differences slavery made: a close analysis of two american communities. *Am. Hist. Rev.* 108, 1299–1307. doi: 10.1086/ahr/108.5.1299

UNESCO, IRCAI. (2024). *Challenging Systematic Prejudices: An Investigation into Gender Bias in Large Language Models*.

Valleriani, M., Kräutli, F., Lockhorst, D., and Shlomi, N. (2023). "Vision on vision: defining similarities among early modern illustrations on cosmology," in *Scientific Visual Representations in History*, eds. M. Valleriani, G. Giannini, and E. Giannetto (Springer International Publishing: New York). doi: 10.1007/978-3-031-11317-8\_4

Zamani, M., El-Hajj, H., Vogl, M., Kantz, H., and Valleriani, M. (2023). A mathematical model for the process of accumulation of scientific knowledge in the early modern period. *Hum. Soc. Sci. Commun.* 10:533. doi: 10.1057/s41599-023-01947-w