

Artificial Intelligence in the Study of Modern History: Employing Digital Technologies to Analyze Political and Social Transformations and Construct Historical Knowledge

- **RA'AD AHMAD SALEM ALMLAJI, Assistant Professor, Ajloun National University, r.almlaji@anu.edu.jo.**

Abstract

Artificial intelligence (AI) now shapes how the records of modern history are transcribed, searched, and analyzed, yet assessments of its value rarely distinguish the stages of research at which it operates. This study examines how AI and related digital technologies are used to analyze political and social transformations in modern history, and how that use affects the construction of historical knowledge. An integrative literature review was conducted. Eight queries in the OpenAlex index and a purposive search for landmark studies yielded 35 peer-reviewed journal articles published between 2016 and 2026, which were verified against Crossref records and analyzed thematically. Five themes emerged: machine transcription and data quality ($n = 8$), modeling of political and social change in text ($n = 10$), analysis of visual sources ($n = 3$), archives and digital source criticism ($n = 9$), and generative AI in historical reasoning ($n = 5$). AI proved most dependable when converting handwritten and printed records into searchable text, and most productive analytically when topic models, word embeddings, and record linkage were combined with close reading and external validation. Studies of parliamentary debates, newspapers, books, and census records show that such methods can date discursive turning points, measure the spread of political innovation, and trace century-long shifts in stereotypes and class meanings. Support is weakest for generative language models, whose historical accounts vary with prompt language and with how thoroughly an event is documented. AI therefore extends the reach of

historical inquiry without replacing interpretation, and its sound use requires source criticism that covers digitization, training data, and model behavior.

Keywords: artificial intelligence, digital history, modern history, text mining, handwritten text recognition, large language models, source criticism, historical knowledge

Introduction

The modern period, understood here as the span from the late eighteenth century to the recent past, left historians a documentary record of a size that no individual can read. Parliaments published their debates, the mass press printed daily, and states counted and registered their populations in censuses and administrative files. For most of the twentieth century the historian's answer to this abundance was selection: a sample of newspapers, a run of parliamentary sessions, a single archive. The digitization of these records has altered that settlement. Putnam (2016) argued that full-text search across digitized sources has changed research practice more deeply than the profession has acknowledged, because it allows information to be found without prior knowledge of where to look and so weakens the link between discovery and place-specific expertise. Huistra and Mellink (2016) made the complementary point that digital repositories oblige scholars to reach sources through particular words, which makes the choice of search terms a methodological decision and not a clerical one.

Artificial intelligence (AI) marks a second phase of this change. In this study the term refers to machine-learning methods, whether supervised, unsupervised, or generative, that perform recognition, classification, pattern detection, or text generation on historical material. Where search retrieves documents, these methods read them: they transcribe handwriting, group texts by theme, measure how the meaning of words shifts, link individuals across records, and classify images. Romein et al. (2020) surveyed this repertoire for the historical discipline, from optical character recognition (OCR) and handwritten text recognition (HTR)

to computer vision, linked data, and quantitative text and network analysis, and observed that its effects on historical scholarship had mostly been indirect. By that time HTR had become a usable technology for archival manuscripts (Muehlberger et al., 2019; Nockels et al., 2022). More recently, large language models have begun to generate fluent historical prose, which has raised the question of whether parts of historical writing itself can be automated (Kansteiner, 2022).

Political and social transformations are a natural test for these methods. Revolutions, the growth of parliamentary government, mass migration, and changing relations of class and gender unfold over years or decades and leave their traces in language produced at scale. A method that can follow vocabulary, themes, and associations across millions of pages promises to show when a change began, how fast it spread, and who carried it. Published work has pursued exactly these aims: the transmission of new political language in the first parliament of the French Revolution (Barron et al., 2018), a century of gender and ethnic stereotypes in American text (Garg et al., 2018), and the shifting cultural markers of social class in twentieth-century books (Kozlowski et al., 2019).

Three difficulties complicate this promise. First, the digitized record is a selection shaped by institutional, financial, and political decisions about what to preserve, scan, catalogue, and make accessible (Zaagsma, 2023), and the resulting collections can misrepresent the historical landscape from which they were drawn (Beelen et al., 2023). Second, machine-produced text contains errors that travel into every later analysis (Burchardt, 2023; Hill & Hengchen, 2019). Third, generative models produce statements about the past whose relation to evidence is uncertain, which touches the standards by which the discipline warrants its claims (Hughes-Warrington, 2025; Kansteiner, 2022).

The research that addresses these matters is dispersed across history, archival and information science, computational social science, computer science, and philosophy.

Existing syntheses cover digital history in general (Romein et al., 2020), archives and AI (Colavizza et al., 2022), or a single technology such as HTR or named entity recognition (Ehrmann et al., 2024; Nockels et al., 2022). None of these brings the technical, empirical, and epistemological strands together around the specific problem of studying political and social change in the modern period. A historian who wishes to know what AI has demonstrably contributed to that problem, and under what conditions its results can be trusted, has no single account to consult.

This study addresses that gap. Its purpose is to synthesize peer-reviewed research on the use of AI and related digital technologies in the study of modern history, with attention to the analysis of political and social transformations and to the construction of historical knowledge. The latter phrase denotes the process through which sources are selected, read, interpreted, and turned into warranted claims about the past. Three questions guide the review:

1. Which AI techniques are applied to which kinds of sources in research on modern history?
2. What have these applications shown about political and social transformations, and how are their findings validated?
3. Which methodological and epistemological conditions govern the construction of historical knowledge with AI?

The study contributes in two ways. It offers historians a structured map of a literature that is otherwise scattered across disciplines with different vocabularies. It also advances an argument about where the contribution of AI lies. The reviewed evidence suggests that the value of AI is uneven along the chain of historical research: well established where sources are prepared for reading, substantial where patterns in language are measured and then interpreted, and contested where machines are asked to produce historical narrative.

Recognizing this unevenness permits a more exact account of what the technology changes in the historian's work and what it leaves in place.

Methodology

Research Design

The study used an integrative review design. Whitemore and Knafl (2005) describe the integrative review as the review method that permits the combination of diverse methodologies, including empirical and theoretical work, and they structure it in five stages: problem identification, literature search, data evaluation, data analysis, and presentation. The design suited the present topic because the relevant literature consists of technical benchmarks, empirical historical studies, methodological proposals, surveys, and theoretical essays, which a review restricted to one study type would have excluded. The problem identification stage is reported in the Introduction; the remaining stages are described below.

Search Strategy

The literature search was carried out on October 7, 2026, in OpenAlex, an open index of scholarly publications. Eight queries were written to cover the technical, empirical, archival, and theoretical sides of the topic (Table 1). Each query was run with a lower limit on the year of publication, results were ranked by relevance, and the highest-ranked records were retained for screening: between 14 and 20 per query and 137 in total. Five records were returned by more than one query, which left 132 unique records.

Keyword queries built on historical vocabulary did not surface several widely cited studies that are published in general science, sociology, economics, and computing journals and that describe their work in the terms of those fields. A purposive search was therefore added. Thirteen articles of this kind were located by author and title and assessed against the same eligibility criteria as the other records.

Table 1*Search Queries, Records Screened, and Articles Included*

Query	Search string	Earliest year	Records screened	Articles included
1	artificial intelligence machine learning historical research historians methods	2015	20	1
2	digital history text mining topic modeling newspapers political discourse	2012	19	3
3	handwritten text recognition historical documents archives Transkribus	2015	20	5
4	large language models historians generative AI historiography	2022	19	4
5	digital history state of the field historians digitized sources text-searchable transnational	2010	15	4
6	parliamentary debates text mining history revolution word embeddings historical change stereotypes	2015	15	2
7	OCR errors digitized newspapers bias representativeness historical research	2015	14	3
8	archives and artificial intelligence overview debates digital archives machine learning access	2018	15	0
Purposive search	Author and title of landmark studies	Not applicable	13	13
Total			150	35

Note. An article retrieved by more than one query is counted under the first query that returned it. The 137 records from the eight queries include five duplicates across queries.

Eligibility Criteria

Records were screened by title and abstract. An article was included if it met all of the following criteria: (a) it was a peer-reviewed journal article with a registered digital object identifier; (b) it was published in English between 2016 and 2026; and (c) it applied a machine-learning or computational method to historical sources from the modern period, evaluated a text-recognition or image-recognition technology on which such research depends, or examined the methodological, archival, or epistemological implications of AI for historical research. Records were excluded if they were preprints, conference papers, book chapters, books, theses, or reports; if they were earlier or duplicate versions of an included

article; or if they concerned AI in fields unrelated to history, such as medicine, economics, or education, which the broader queries also returned. Screening retained 22 of the 132 unique records. Together with the 13 purposively added articles, the final corpus comprised 35 articles.

Data Evaluation

Each included article was checked against its Crossref record to confirm the publication type, journal, year, volume, issue, and pagination, and to determine whether a retraction notice was attached to the record; none was. Because the corpus combines study types that cannot be assessed with a single appraisal instrument, no numerical quality score was assigned. Publication in a peer-reviewed journal served as the minimum quality threshold, and the weight given to each source in the synthesis reflects how directly it bears on the research questions.

Data Extraction and Analysis

For every article, the following items were recorded from the abstract and bibliographic record: the technique used or discussed, the type of source, the period and region covered, the stated aim, the principal finding, and any limitation the authors reported. The extracted material was analyzed thematically following the phases set out by Braun and Clarke (2006): familiarization with the data, generation of initial codes, search for themes, review of themes, definition and naming of themes, and reporting. Coding was inductive. Codes that described the same stage of historical research were grouped, which produced five themes ordered from the preparation of sources to their analysis and interpretation.

Limitations

Several limitations qualify the results. The search used a single index and relevance-ranked samples of each result set, so the corpus is a structured selection and not a census of the field. Only English-language journal articles were included, which excludes monographs,

edited volumes, and scholarship in other languages, among them Arabic. Extraction relied on abstracts and bibliographic records, so details reported only in the full texts are not reflected. Screening and coding were performed by a single reviewer without an independent second coder. The purposive additions involved judgment about which studies count as landmarks. Finally, the sources and case studies in the corpus are predominantly European and North American, a point taken up in the Results.

Results

Characteristics of the Reviewed Literature

The corpus consists of 35 articles published between 2016 and 2026. Twenty of them (57%) appeared in 2022 or later, which reflects the recent growth of work on handwritten text recognition and on generative models. The articles are spread across the disciplines that share the topic. Historical journals (*The American Historical Review*, *History*, *History and Theory*, *Media History*, and *Historical Methods*) supply the reflective and methodological essays. *Digital Scholarship in the Humanities* is the most frequent outlet, with seven articles. Archival and information science is represented by the *Journal of Documentation*, *Archival Science*, the *Journal of the Association for Information Science and Technology*, and the *Journal on Computing and Cultural Heritage*. Several of the most cited empirical studies were published outside the humanities, in the *Proceedings of the National Academy of Sciences*, the *American Sociological Review*, and the *Journal of Economic Literature*. Table 2 summarizes the five themes and the articles assigned to each.

Table 2*Themes Identified in the Corpus (N = 35)*

Theme	<i>n</i>	Principal techniques	Typical sources	Articles
1. Machine transcription and data quality	8	HTR, OCR, multimodal language models	Manuscripts, printed books, newspapers	Burchardt (2023); Crosilla et al. (2025); Hill and Hengchen (2019); Muehlberger et al. (2019); Najam and Faizullah (2023); Nockels et al. (2022, 2024); Romein et al. (2025)
2. Modeling political and social change in text	10	Topic models, word embeddings, information-theoretic measures, named entity recognition, record linkage	Parliamentary debates, newspapers, books, census records	Abramitzky et al. (2021); Barron et al. (2018); Ehrmann et al. (2024); Garg et al. (2018); Guldi (2018, 2019); Hurtado Bodell et al. (2026); Kozłowski et al. (2019); Ros and Wevers (2025); Viola and Verheul (2020)
3. Analysis of visual sources	3	Convolutional neural networks, face detection, contrastive image and text models	Newspaper images, documentary photographs, television	Arnold and Tilton (2019); Smits and Wevers (2023); Wevers and Smits (2020)
4. Archives and digital source criticism	9	Full-text search, machine learning for archival access, sensitivity review	Digitized and born-digital archives	Beelen et al. (2023); Colavizza et al. (2022); Huistra and Mellink (2016); Jaillant and Caputo (2022); Jensen (2021); Oberbichler et al. (2022); Putnam (2016); Romein et al. (2020); Zaagsma (2023)
5. Generative AI and historical reasoning	5	Large language models, generative reconstruction	Model outputs on historical questions	Hughes-Warrington (2025); Kansteiner (2022); Khosrowi and Finn (2025); Papaioannou (2026); Ulloa et al. (2026)

Note. HTR = handwritten text recognition; OCR = optical character recognition.

Theme 1: Machine Transcription and the Quality of Historical Data

The first theme concerns the step that precedes all analysis: turning images of documents into text. Muehlberger et al. (2019) reported, on the basis of the Transkribus platform, that HTR had become a usable technology that can be combined with mass digitization to produce accurate transcripts of archival manuscripts. They proposed a cooperative model to sustain the infrastructure and noted unresolved problems of funding and resources. The uptake that followed is documented by Nockels et al. (2022), whose systematic review collected 381 publications from 2015 to 2020 that mention the platform.

Humanities applications accounted for 67% of these publications, technological papers for 25%, user studies for 5%, and tutorials for 3%. The proportions show that within a few years the technology had passed from computer science into regular use by archivists and historians.

Accuracy, however, depends on script, language, and training. Romein et al. (2025) compared several recognition engines on datasets in different scripts and found that transformer-based engines performed best without further training on documents in Latin script, whereas other engines did better on particular non-Latin scripts once they had been trained or fine-tuned for them. Crosilla et al. (2025) benchmarked multimodal large language models on the same task. They found strong performance on English texts, weaker performance on other languages, no significant capacity for self-correction, and results on historical documents that were not consistently better or worse than those of a specialized platform. For Arabic handwriting, Najam and Faizullah (2023) reviewed deep-learning approaches published between 2020 and 2023 and identified a shortage of high-quality, domain-specific datasets as a principal obstacle. Taken together, these studies indicate that the ability to work computationally on a body of sources is unevenly distributed across languages and scripts. Historians of regions whose records are written in under-resourced scripts face a higher threshold before any analysis can begin.

Printed sources raise the same issue through OCR. Hill and Hengchen (2019) compared an OCR version of Eighteenth Century Collections Online with a manually keyed version of the same texts. They tested how topic modeling, authorship attribution, collocation analysis, and vector space modeling respond to recognition errors, with the further aim of estimating OCR quality where no clean reference exists. Burchardt (2023) tested searches in leading digitized newspaper databases and found an average error rate of 18% for single words in body text and far higher rates in advertisements. He concluded that database owners

should supply detailed metadata so that users can practice source criticism on search results. Nockels et al. (2024), drawing on 120 published texts, added that HTR depends on prior digitization pipelines, can omit the archival history of documents, and can entrench existing biases. They listed copyright, credit for transcribers, and environmental cost among the issues that remain open.

The finding for this theme is that transcription is the most mature application of AI in historical research and the one with the clearest evidence of adoption. Its outputs are nonetheless data with a measurable error structure. That structure differs by script, language, layout, and text type, and it has to be known before transcribed text is used as evidence.

Theme 2: Modeling Political and Social Change in Text

The largest group of articles applies machine learning to text in order to describe change over time. The techniques are topic models, which group words that tend to occur together into themes; word embeddings, which represent the meaning of words by their contexts of use; information-theoretic measures of difference between texts; named entity recognition; and automated record linkage.

Three studies address parliamentary politics. Barron et al. (2018) analyzed about 40,000 speeches from the first parliament of the French Revolution and measured how far each speech departed from earlier patterns of word use and how far its patterns persisted afterward. Speakers on the left introduced new patterns at higher rates, speakers on the right tended to preserve existing ones, and holders of the newly created offices of assembly president and committee chair had a marked influence on the course of debate. The study also dated a transition, midway through the parliament, at which committees gained greater power over legislation. The authors presented these results as consistent with established interpretations while bringing to light dynamics that had gone unnoticed. Guldi (2019) applied word counts, topic modeling, and divergence measures to nineteenth-century British

parliamentary debates. Her results supported social-historical scholarship on the growing place of food and water in national policy, drew attention to understudied debates on fisheries, and showed that Liberal and Conservative speakers related differently to the past and the future. Ros and Wevers (2025) examined 8.2 million sentences spoken in the Dutch House of Representatives between 1945 and 1995. By modeling 250 topics and the networks that connect them, they found that the modularity of debate almost doubled after 1960 while its density fell. They interpreted this as a narrowing of political discussion through the specialization of members of parliament, a process they termed epistemic capture.

Social transformations have been studied mainly through the press and through books. Viola and Verheul (2020) combined topic modeling with the discourse-historical approach to analyze Italian-language newspapers published in the United States between 1898 and 1920, and traced how the immigrant community negotiated its ties to the host country and the homeland as the experience of migration changed. Hurtado Bodell et al. (2026) used seeded topic models, in which researchers supply words that anchor topics to known concepts, on millions of Swedish newspaper articles from 1945 to 2019. They derived longitudinal measures of how immigration was framed and identified turning points that divide the discourse into distinct eras. Garg et al. (2018) trained word embeddings on 100 years of American text and showed that changes in the associations between words for gender and ethnic groups and words for occupations and attributes track demographic and occupational shifts recorded in census data, including those linked to the women's movement of the 1960s and to Asian immigration. Kozlowski et al. (2019) analyzed millions of books published over 100 years. They found that the markers of social class shifted continuously during the economic transformations of the twentieth century while the basic dimensions of class meaning remained stable, with the exception of education, which became closely tied to affluence.

Two further techniques supply the entities on which social and political history depends. Automated record linkage connects the same individual across censuses and other registers. Abramitzky et al. (2021) evaluated several automated methods against hand linking and reported false positive rates below 5%, a measurable trade-off between false positives and match rates, and closely similar estimates of the quantities of interest across methods. Named entity recognition identifies persons, places, and organizations in text. In their survey, Ehrmann et al. (2024) discussed the particular difficulties that historical documents pose for this task, among them noisy input from OCR, historical variation in language, and limited training resources.

A common feature of these studies is the way in which results are secured. The empirical studies do not treat the output of a model as self-evident. Garg et al. (2018) compared embedding measures with census data, and Kozłowski et al. (2019) validated their dimensions against surveys. Barron et al. (2018) set their findings beside the existing historiography of the Revolution. Viola and Verheul (2020) paired a statistical model with an established qualitative method, and Hurtado Bodell et al. (2026) built prior knowledge into the model itself. Guldi (2018) gave this practice a general form in the procedure she called critical search, in which the researcher moves repeatedly between the choice of search terms, algorithmic summaries of the corpus, and the guided reading of selected passages. The finding for this theme is that computational text analysis has produced dated, measurable, and in several cases new observations about political and social change, and that the credibility of those observations rests on triangulation with independent evidence and with reading.

Theme 3: Analysis of Visual Sources

Research on images is smaller in volume. Wevers and Smits (2020) showed that convolutional neural networks can be used on digitized Dutch newspapers to separate photographs from illustrations, to cluster visually similar advertisements, and to classify

images according to categories defined by domain experts. They argued that such methods open the visual side of digitized archives, which text-based research had left aside. Arnold and Tilton (2019) proposed the framework of distant viewing for large visual corpora. Its central claim is that extracting metadata from images is an interpretive act that should be made explicit. One of their two demonstrations used the documentary photographs of the Farm Security Administration and Office of War Information and compared the styles of the photographers who built that collection. Smits and Wevers (2023) turned to multimodal models trained on paired images and text, which can classify categories they were not trained on, and argued that these models allow image and text to be analyzed together. They also warned that the models carry biases from their training data, a concern of particular weight for historical images.

The finding for this theme is that image analysis at scale is technically feasible and has begun to extend the evidentiary base of modern history beyond text. The categories through which machines see historical images remain a matter of interpretation.

Theme 4: Archives, Representativeness, and Digital Source Criticism

The fourth theme concerns the collections on which every method depends. Zaagsma (2023) argued that decisions about what to digitize, how to catalogue and classify it, which metadata to add, and how to provide access are political decisions with political consequences, and he placed them in the context of unequal relations between and within the Global North and the Global South. Beelen et al. (2023) offered a way to measure the result. Their environmental scan compares a digitized collection of nineteenth-century British provincial newspapers with Victorian press directories, a contemporary reference source, and so exposes selection decisions made at the microfilming and digitization stages. The approach treats representativeness as an empirical property of a corpus that can be tested.

A second group of articles deals with the skills that follow from this situation. Putnam (2016) cautioned that digital search removes the contextual learning that place-based archival research once imposed, with intellectual and political costs. Jensen (2021) proposed a digital archival literacy for all historians, based on treating the move from analogue to digital archives as a change of medium with its own logic of production, content, distribution, and use. Huistra and Mellink (2016) recommended search technologies that depend less on exact phrasing. Oberbichler et al. (2022) described integrated workflows in which humanities scholars, computer scientists, and librarians work on historical newspapers together, and argued that each group must understand the practices and expectations of the others. Romein et al. (2020) similarly stressed the need for hermeneutics and awareness of data in digital historical work.

AI is also changing archives themselves. Colavizza et al. (2022) organized the literature on archives and AI around theoretical and professional considerations, the automation of recordkeeping, the organization of and access to archives, and new forms of digital archives. Jaillant and Caputo (2022) observed that most born-digital records held by cultural institutions are inaccessible for reasons of privacy, copyright, commercial interest, and technology. They proposed machine-assisted sensitivity review as a practical route to opening such records, named error, opacity, and bias as risks, and insisted on fairness, accountability, and transparency. The matter bears directly on the history of the late twentieth and early twenty-first centuries, whose records are largely born digital.

The finding for this theme is that the archive read by AI is a constructed object. Source criticism must now cover selection, recognition quality, interfaces, and algorithms in addition to the documents.

Theme 5: Generative AI and Historical Reasoning

The most recent articles examine large language models. Kansteiner (2022) expected such models to supply historians with intermediate products such as translations, summaries, and chronologies. He argued that they nonetheless fail at tasks the discipline values: they cannot guarantee the truth of what they state, they cannot follow a piece of information through layers of text, and they lack ethical self-reflexivity. He concluded that academic historical writing would continue to require human agency, and suggested that the models could serve theorists as instruments for experiments on historical writing. Hughes-Warrington (2025) approached the ethics of such systems through the logic of historical writing and argued for moving beyond lists of principles toward attention to the forms of reasoning that histories display. Khosrowi and Finn (2025) asked whether generative systems used to reconstruct destroyed artifacts can yield evidence, and argued that their outputs can count at least as higher-order evidence.

Two empirical audits give these arguments a factual basis. Ulloa et al. (2026) put questions built on common denialist claims to five language models about four mass atrocities, in English and in a second language relevant to each case. Responses were generally accurate for the Holocaust, which is widely documented, and showed inconsistencies and susceptibility to denialist framings for less represented cases such as the Cambodian genocide. The authors attributed the difference to the availability of training data and to the probabilistic character of the models. Papaioannou (2026) compared accounts of the events of 1922 in Asia Minor generated by two models and found that terminology, the attribution of responsibility, and historiographical framing varied with the language of the prompt.

The finding for this theme is that generative models are of use for auxiliary tasks under verification. As sources of historical statements they reproduce the unevenness and the framings of the material on which they were trained.

Synthesis of the Findings

Read together, the five themes answer the research questions. On the first question, the techniques in use correspond to stages of research: recognition models for manuscripts and print; topic models, embeddings, entity recognition, and record linkage for text and registers; neural image classifiers for visual material; and generative models for text production. The sources are overwhelmingly those that modern states and publics produced in series, namely parliamentary debates, newspapers, books, censuses, and photographs.

On the second question, the applications have yielded specific results on political change (the diffusion of new political language in a revolutionary assembly, the differing temporal orientations of parties, the specialization of postwar parliamentary debate) and on social change (successive framings of immigration, the movement of gender and ethnic stereotypes, the stability and change of class meanings). In each case the result was secured by comparison with evidence external to the model.

On the third question, the conditions for sound knowledge construction are a known error structure in the data, a tested relation between the digitized corpus and the historical record, explicit interpretive choices, and human verification of generated statements. Table 3 sets out these conditions by stage of research.

Table 3

Contribution of AI, Principal Risk, and Reported Safeguard at Each Stage of Historical Research

Stage of research	Contribution of AI	Principal risk	Safeguard reported in the literature
Source preparation	Converts manuscripts and print into searchable text at scale	Error rates that vary by script, language, and text type	Training or fine-tuning on the collection; metadata on recognition quality (Burchardt, 2023; Romein et al., 2025)
Corpus building and search	Retrieval across collections without prior knowledge of location	Unrepresentative digitized samples; dependence on exact wording	Comparison with contemporary reference sources; iterative search (Beelen et al., 2023; Guldi, 2018)
Pattern analysis	Measures change in themes, meanings, and links between records over decades	Statistical patterns taken for historical explanation	Triangulation with external data, qualitative method, and existing historiography (Garg et al., 2018; Viola & Verheul, 2020)
Visual analysis	Classifies and clusters images	Categories and biases inherited from training data	Expert-defined categories; explicit interpretive coding (Arnold & Tilton, 2019; Wevers & Smits, 2020)
Narrative generation	Drafts translations, summaries, and chronologies	Unreliable statements; variation by prompt language and by coverage of the event	Human verification; audits across languages and cases (Kansteiner, 2022; Ulloa et al., 2026)

Two further observations follow from the corpus as a whole. The first is geographic and linguistic. The empirical studies of political and social change in Themes 2 and 3 concern France, Britain, the Netherlands, Sweden, and the United States, and the article devoted to Arabic handwriting identifies a lack of training data as a principal difficulty. Knowledge built with AI is therefore, at present, largely knowledge of the regions whose records were digitized early and in well-resourced languages. This leaves a clear agenda for work on Arabic and other under-resourced archives. The second observation concerns the historian's role. Across the themes, AI moves effort away from locating and transcribing documents and toward evaluating data, models, and outputs. The interpretive responsibility that defines the discipline is not reduced by this shift. It is extended to new objects, and source criticism is the practice through which the discipline can absorb the technology on its own terms.

References

- Abramitzky, R., Boustan, L., Eriksson, K., Feigenbaum, J., & Pérez, S. (2021). Automated linking of historical data. *Journal of Economic Literature*, 59(3), 865–918.
<https://doi.org/10.1257/jel.20201599>
- Arnold, T., & Tilton, L. (2019). Distant viewing: Analyzing large visual corpora. *Digital Scholarship in the Humanities*, 34(Suppl. 1), i3–i16.
<https://doi.org/10.1093/llc/fqz013>
- Barron, A. T. J., Huang, J., Spang, R. L., & DeDeo, S. (2018). Individuals, institutions, and innovation in the debates of the French Revolution. *Proceedings of the National Academy of Sciences*, 115(18), 4607–4612. <https://doi.org/10.1073/pnas.1717729115>
- Beelen, K., Lawrence, J., Wilson, D. C. S., & Beavan, D. (2023). Bias and representativeness in digitized newspaper collections: Introducing the environmental scan. *Digital Scholarship in the Humanities*, 38(1), 1–22. <https://doi.org/10.1093/llc/fqac037>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Burchardt, J. (2023). Are searches in OCR-generated archives trustworthy? *Jahrbuch für Wirtschaftsgeschichte / Economic History Yearbook*, 64(1), 31–54.
<https://doi.org/10.1515/jbwg-2023-0003>
- Colavizza, G., Blanke, T., Jeurgens, C., & Noordegraaf, J. (2022). Archives and AI: An overview of current debates and future perspectives. *Journal on Computing and Cultural Heritage*, 15(1), 1–15. <https://doi.org/10.1145/3479010>
- Crosilla, G., Klic, L., & Colavizza, G. (2025). Benchmarking large language models for handwritten text recognition. *Journal of Documentation*, 81(7), 334–354.
<https://doi.org/10.1108/JD-03-2025-0082>

- Ehrmann, M., Hamdi, A., Linhares Pontes, E., Romanello, M., & Doucet, A. (2024). Named entity recognition and classification in historical documents: A survey. *ACM Computing Surveys*, 56(2), 1–47. <https://doi.org/10.1145/3604931>
- Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16), E3635–E3644. <https://doi.org/10.1073/pnas.1720347115>
- Guldi, J. (2018). Critical search: A procedure for guided reading in large-scale textual corpora. *Journal of Cultural Analytics*, 3(1). <https://doi.org/10.22148/16.030>
- Guldi, J. (2019). The measures of modernity: The new quantitative metrics of historical change over time and their critical interpretation. *International Journal for History, Culture and Modernity*, 7(1), 899–939. <https://doi.org/10.18352/hcm.589>
- Hill, M. J., & Hengchen, S. (2019). Quantifying the impact of dirty OCR on historical text analysis: Eighteenth Century Collections Online as a case study. *Digital Scholarship in the Humanities*, 34(4), 825–843. <https://doi.org/10.1093/llc/fqz024>
- Hughes-Warrington, M. (2025). Ethics for artificial historians. *History and Theory*, 64(2), 159–177. <https://doi.org/10.1111/hith.12377>
- Huistra, H., & Mellink, B. (2016). Phrasing history: Selecting sources in digital repositories. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 49(4), 220–229. <https://doi.org/10.1080/01615440.2016.1205964>
- Hurtado Bodell, M., Magnusson, M., & Keuschnigg, M. (2026). Seeded topic models in digital archives: Analyzing interpretations of immigration in Swedish newspapers, 1945–2019. *Sociological Methods & Research*, 55(1), 120–156. <https://doi.org/10.1177/00491241241268453>

- Jaillant, L., & Caputo, A. (2022). Unlocking digital archives: Cross-disciplinary perspectives on AI and born-digital data. *AI & Society*, 37(3), 823–835.
<https://doi.org/10.1007/s00146-021-01367-x>
- Jensen, H. S. (2021). Digital archival literacy for (all) historians. *Media History*, 27(2), 251–265. <https://doi.org/10.1080/13688804.2020.1779047>
- Kansteiner, W. (2022). Digital doping for historians: Can history, memory, and historical theory be rendered artificially intelligent? *History and Theory*, 61(4), 119–133.
<https://doi.org/10.1111/hith.12282>
- Khosrowi, D., & Finn, F. (2025). Can generative AI produce novel evidence? *Philosophy of Science*, 92(5), 1405–1416. <https://doi.org/10.1017/psa.2025.10123>
- Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84(5), 905–949. <https://doi.org/10.1177/0003122419877135>
- Muehlberger, G., Seaward, L., Terras, M., Ares Oliveira, S., Bosch, V., Bryan, M., Colutto, S., Déjean, H., Diem, M., Fiel, S., Gatos, B., Greinöcker, A., Grüning, T., Hackl, G., Haukkovaara, V., Heyer, G., Hirvonen, L., Hodel, T., Jokinen, 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.
<https://doi.org/10.1108/JD-07-2018-0114>
- Najam, R., & Faizullah, S. (2023). Analysis of recent deep learning techniques for Arabic handwritten-text OCR and post-OCR correction. *Applied Sciences*, 13(13), Article 7568. <https://doi.org/10.3390/app13137568>
- Nockels, J., Gooding, P., Ames, S., & Terras, M. (2022). Understanding the application of handwritten text recognition technology in heritage contexts: A systematic review of

Transkribus in published research. *Archival Science*, 22(3), 367–392.

<https://doi.org/10.1007/s10502-022-09397-0>

Nockels, J., Gooding, P., & Terras, M. (2024). The implications of handwritten text recognition for accessing the past at scale. *Journal of Documentation*, 80(7), 148–167.

<https://doi.org/10.1108/JD-09-2023-0183>

Oberbichler, S., Boroş, E., Doucet, A., Marjanen, J., Pfanzelter, E., Rautiainen, J., Toivonen, H., & Tolonen, M. (2022). Integrated interdisciplinary workflows for research on historical newspapers: Perspectives from humanities scholars, computer scientists, and librarians. *Journal of the Association for Information Science and Technology*, 73(2), 225–239. <https://doi.org/10.1002/asi.24565>

Papaioannou, C. (2026). Language and the framing of historical narrative in large language models: The case of Asia Minor (1922). *AI and Ethics*, 6(3), Article 262.

<https://doi.org/10.1007/s43681-026-01138-8>

Putnam, L. (2016). The transnational and the text-searchable: Digitized sources and the shadows they cast. *The American Historical Review*, 121(2), 377–402.

<https://doi.org/10.1093/ahr/121.2.377>

Romein, C. A., Kemman, M., Birkholz, J. M., Baker, J., De Gruijter, M., Meroño-Peñuela, A., Ries, T., Ros, R., & Scagliola, S. (2020). State of the field: Digital history.

History, 105(365), 291–312. <https://doi.org/10.1111/1468-229X.12969>

Romein, C. A., Rabus, A., Leifert, G., & Ströbel, P. B. (2025). Assessing advanced handwritten text recognition engines for digitizing historical documents. *International*

Journal of Digital Humanities, 7(1), 115–134. <https://doi.org/10.1007/s42803-025-00100-0>

- Ros, R., & Wevers, M. (2025). Epistemic capture through specialization in post-World War II parliamentary debate. *Computational Humanities Research, 1*, Article e6.
<https://doi.org/10.1017/chr.2025.10008>
- Smits, T., & Wevers, M. (2023). A multimodal turn in digital humanities: Using contrastive machine learning models to explore, enrich, and analyze digital visual historical collections. *Digital Scholarship in the Humanities, 38*(3), 1267–1280.
<https://doi.org/10.1093/llc/fqad008>
- Ulloa, R., Zucker, E. M., Bultmann, D., Simon, D. J., & Makhortykh, M. (2026). From prosthetic memory to prosthetic denial: Auditing whether large language models are prone to mass atrocity denialism. *AI & Society, 41*(4), 3889–3903.
<https://doi.org/10.1007/s00146-025-02719-7>
- Viola, L., & Verheul, J. (2020). Mining ethnicity: Discourse-driven topic modelling of immigrant discourses in the USA, 1898–1920. *Digital Scholarship in the Humanities, 35*(4), 921–943. <https://doi.org/10.1093/llc/fqz068>
- Wevers, M., & Smits, T. (2020). The visual digital turn: Using neural networks to study historical images. *Digital Scholarship in the Humanities, 35*(1), 194–207.
<https://doi.org/10.1093/llc/fqy085>
- Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing, 52*(5), 546–553. <https://doi.org/10.1111/j.1365-2648.2005.03621.x>
- Zaagsma, G. (2023). Digital history and the politics of digitization. *Digital Scholarship in the Humanities, 38*(2), 830–851. <https://doi.org/10.1093/llc/fqac050>