



**HAL**  
open science

# Algorithmic Reckonings: Artificial Intelligence, Ethical Imperatives, and the Contested Future of Historical Memory

Hind Bouchareb, Frédéric Jurie

## ► To cite this version:

Hind Bouchareb, Frédéric Jurie. Algorithmic Reckonings: Artificial Intelligence, Ethical Imperatives, and the Contested Future of Historical Memory. 2026. <hal-05508518v1>

**HAL Id: hal-05508518**

**<https://hal.science/hal-05508518v1>**

Preprint submitted on 13 Feb 2026 (v1), last revised 20 Feb 2026 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

# Algorithmic Reckonings: Artificial Intelligence, Ethical Imperatives, and the Contested Future of Historical Memory

Hind Bouchareb and Frederic Jurie  
University of Caen Normandie

February 13, 2026

## Abstract

Artificial Intelligence (AI) is rapidly transforming research paradigms across all of historical inquiry, presenting both unprecedented opportunities and profound challenges. This paper investigates these complex implications through the specific and revealing lens of the History of Memory. As a subfield focused on how societies construct, contest, and make meaning of their pasts, it provides an ideal framework for analyzing AI's role not just as a research tool, but as an active agent in memory production. By examining this dynamic, the paper addresses fundamental questions of evidence, authority, and ethics that, while explored through memory studies, are vital to the broader historical profession.

The paper dissects these core ethical issues, examining specific AI use cases—such as natural language processing of testimonies and generative AI for synthetic memory objects—to illustrate their dual-edged nature. This analysis is further nuanced by considering the performative turn in AI studies, recognizing AI not merely as a representational tool but as an active agent in shaping how the past is remembered and understood, thereby shifting mnemotechnic values [106]. It further explores how AI reconfigures the politics of remembrance and the evolving role of the historian, acknowledging that machine learning technologies are already impacting how we craft and interpret the past [84].

Ultimately, this paper contends that responsible engagement with AI in Historical Memory necessitates a proactive approach. By fostering critical wisdom and ethical stewardship, historians can harness AI's capabilities to enrich understanding and promote inclusivity. The future of memory in the algorithmic age hinges on the conscious choices made to navigate this new terrain with the scholarly integrity and ethical responsibility required of all historical practice.

*Keywords:* History of Memory, Artificial Intelligence, Ethics, Digital Humanities, Collective Memory, Cultural Memory, Algorithmic Bias, Algorithmic Governance, Digital History, Performativity.

# Contents

<b>1 Introduction: Memory’s New Mirror – Reflecting Pasts in the Age of AI</b>	<b>4</b>
<b>Introduction: Memory’s New Mirror – Reflecting Pasts in the Age of AI</b>	<b>4</b>
1.1 The “Why Now?” Factor: A Confluence of Memory and Machine . . . . .	5
1.2 Defining the Terrain: History of Memory as Critical Inquiry . . . . .	6
1.3 Defining the Tool: Understanding Artificial Intelligence in a Humanities Context	6
1.4 Thesis Statement & Argument Roadmap . . . . .	7
<b>I Reconfiguring the Archives and Narratives of Memory: AI’s Potential and Perils</b>	<b>7</b>
<b>2 Expanding the Scope: AI’s Potential and Positive Applications</b>	<b>7</b>
2.1 Analyzing Vast Textual and Visual Corpora at Scale . . . . .	8
2.2 Identifying Latent Patterns and Uncovering Hidden Connections . . . . .	8
2.3 Enhancing Accessibility, Discovery, and Preservation . . . . .	9
2.4 Generating New Forms of Engagement and Interpretation . . . . .	9
<b>3 Navigating the Perils: Ethical Risks and Methodological Challenges</b>	<b>10</b>
3.1 Algorithmic Bias, Misrepresentation, and Decontextualization . . . . .	10
3.2 Erosion of Authenticity, Trivialization, and Manipulation . . . . .	11
3.3 Privacy, Consent, and Data Protection Concerns . . . . .	12
3.4 Inaccuracy, Nuance, and the Loss of Authorial Voice . . . . .	13
<b>4 The Nature of AI-Generated “Memory”: Representation, Simulation, or Artifact?</b>	<b>13</b>
<b>II The Algorithmic Politics of Remembrance: Power, Control, and the Contested Construction of Pasts</b>	<b>15</b>
<b>5 Fundamental Ethical Challenges in AI’s Interaction with Memory Archives</b>	<b>16</b>
5.1 Algorithmic Bias and Historical Silences . . . . .	16
5.2 Authenticity, Aura, and the “Black Box” of AI Interpretation . . . . .	17
5.3 Privacy, Data Subjectivity, and the Ethics of AI in Memory Work . . . . .	18
<b>6 Revisiting Power in HoM: From Human Actors to Algorithmic Governance</b>	<b>19</b>
<b>7 AI and the (Re)Centralization of Memory Production and Dissemination</b>	<b>20</b>

<b>8</b>	<b>Ethical Quagmires in Algorithmic Memory Politics</b>	<b>21</b>
8.1	Algorithmic Canonization and the “Objectivity Trap” . . . . .	21
8.2	Manipulation, Misinformation, and “Deepfake Memory” . . . . .	22
8.3	Algorithmic Discrimination and Representational Harms in Memory . . . . .	23
8.4	The Algorithmic “Right to be Forgotten” vs. the “Duty to Remember” . . . . .	23
8.5	Environmental and Societal Costs as Political Choices . . . . .	24
 <b>III The Historian’s Evolving Craft: Agency, Responsibility, and Methodological Futures in an AI-Infused World</b>		<b>25</b>
<b>9</b>	<b>The Indispensable Historian: Critical Judgment in an Age of Intelligent Machines (The Historian’s Role)</b>	<b>25</b>
<b>10</b>	<b>Fostering Critical AI Literacy and Interdisciplinarity</b>	<b>26</b>
10.1	Developing Critical AI Literacy for Historians . . . . .	26
10.2	The Imperative of Interdisciplinary Collaboration . . . . .	27
10.3	Mediation as an Ethical and Practical Imperative . . . . .	27
<b>11</b>	<b>Archiving the Algorithm: Long-Term Preservation and Methodological Transparency (Long-term preservation)</b>	<b>28</b>
<b>12</b>	<b>Navigating Professional Ethics and Responsibilities</b>	<b>29</b>
12.1	Accountability and Liability in AI-Assisted Historical Work . . . . .	29
12.2	Deskilling vs. Upskilling: The Impact of AI on Historical Labor and Training	30
12.3	Equity of Access and the Digital Divide in AI-Powered Memory Research . . .	30
 <b>IV Towards Ethical Stewardship: A Framework for Critical Practice with AI in Historical Memory</b>		<b>31</b>
<b>13</b>	<b>Recapitulation of Core Ethical Dilemmas at the Intersection of AI and HoM</b>	<b>31</b>
<b>14</b>	<b>A Proposed Framework: Points to Consider When Using AI for Historical Memory Research</b>	<b>32</b>
<b>15</b>	<b>A Call for Proactive Professional Engagement and Guideline Development</b>	<b>35</b>
<b>Conclusion: Memory’s Future, Historians’ Choices – Navigating the Algorithmic Age with Critical Wisdom</b>		<b>37</b>
<b>Acknowledgments</b>		<b>38</b>

# 1 Introduction: Memory’s New Mirror – Reflecting Pasts in the Age of AI

The study of the past is perpetually in dialogue with the present, not only in terms of the questions asked but also the tools employed. The rapid ascent of Artificial Intelligence (AI) promises to revolutionize historical research, offering unprecedented tools for analysis while simultaneously posing profound ethical challenges. However, to fully grasp the nature of this transformation, we must look beyond AI’s function as a mere analytical instrument for the historian. This paper argues that the most revealing arena for examining AI’s impact is the specific subfield of the History of Memory. By focusing on this domain, we can understand AI not just as a new tool for doing history, but as a powerful new force in the very making of history and memory that societies live by.

To clarify this distinction: historical research, in its broadest sense, is the critical discipline of reconstructing and interpreting past events based on evidence. In contrast, the History of Memory (HoM), which emerged prominently in the 1980s, shifts the focus from “what actually happened” to how pasts are perceived, constructed, represented, and mobilized in different social contexts [80, 123]. Drawing on the work of scholars like Maurice Halbwachs and Pierre Nora, HoM investigates the social frameworks of collective recall, the power dynamics of silencing and remembrance, and the creation of symbolic ‘lieux de mémoire’ that anchor a society’s identity [75, 121, 175]. This subfield is therefore uniquely equipped to analyze the rise of AI, because AI systems themselves are becoming powerful new agents of memory production. They do not simply analyze the past; they generate narratives, simulate experiences, and curate information, thereby actively shaping how the past is understood. The core ethical dilemmas of AI—bias, misrepresentation, power, and the construction of authority—are the very same themes that have animated the History of Memory for decades. This paper therefore uses the History of Memory as a critical lens to dissect AI’s impact, contending that while our focus is specific, the questions raised about evidence, interpretation, and professional ethics have crucial implications for the entire historical profession.

The scholarly community is beginning to grapple with these profound changes. In a foundational contribution, Schuh [160] theorizes AI not merely as a tool, but as a new form of “artificial memory” itself, arguing that its “analogical” logic of generation marks a fundamental departure from the “archival logic” of traditional repositories. While Schuh’s work provides a vital theoretical diagnosis of this new memory agent and its governance challenges, this paper takes a different, though complementary, approach. Our focus is less on the ontology of AI-as-memory and more on the direct, practical, and ethical consequences this shift entails for the historical profession. By using the History of Memory (HoM) as a critical lens, we aim to develop a tangible framework to guide historians, archivists, and educators as they navigate the perils and potentials of their evolving craft.

However, this article argues that we need to move beyond a purely representational understanding of AI’s role. Drawing upon the work of Matei [106], we contend that AI, particularly generative AI, is fundamentally changing the nature of collective memory by

shifting history from an object of media representation to an object of algorithmic performativity. This "performativity turn" necessitates a re-evaluation of existing ethical frameworks and the development of new guidelines that address the unique challenges posed by AI's active role in shaping historical narratives. The logic here extends the concept of "procedural rhetoric," which argues that computational systems persuade and make arguments not through text or images, but through the very rules and processes they enact [21]. In this light, an AI model is not just representing the past; its algorithms are performing an interpretation of it, embedding choices about what matters, what to connect, and what to foreground. Addressing the unique challenges posed by this active role in shaping historical narratives is therefore a central task for the profession. While AI offers unprecedented opportunities for historical research, Kansteiner [84] cautions against the uncritical belief that Large Language Models (LLMs) can fully automate the production of knowledge or replace the nuanced judgement of historians. The discipline must remain wary of issues such as historical accuracy and a variety of additional complex variables, including environmental and social factors. Moreover, the rise of powerful LLMs offers historians a new conversational approach for exploring historical data, though challenges related to their training data and potential biases must be carefully considered. As Garcia and Weilbach [61] demonstrate, LLMs augmented with specialized academic sources can provide accessible and effective research assistance, prompting a reevaluation of traditional research methods.

### 1.1 The “Why Now?” Factor: A Confluence of Memory and Machine

The current moment is particularly ripe for an inquiry into AI's role in historical memory. Several converging factors underscore this urgency. Firstly, there is an undeniable societal and academic preoccupation with memory, often termed the “memory boom” [186]. This "memory boom" is also fueled by a paradigm shift confronting memory institutions, particularly in Europe: the imminent loss of the last living witnesses to the major conflicts of the 20th century. For history and memory museums, the question of how to fill the void left by the subjective and embodied voice of survivors has become a central challenge. Faced with the political and societal imperative of the "duty of memory" (*devoir de mémoire*), which demands ever more engaging forms of transmission, Artificial Intelligence is seen as one of the few solutions capable of preserving the pedagogical power of personal testimony. It offers the promise of a mediation that simulates a human interaction, thus responding to the need to make History more personal and accessible to new generations. From debates over national monuments and historical apologies to the proliferation of museums and testimonial archives, societies globally are grappling with how to remember, represent, and reckon with their pasts. Secondly, the capabilities of AI, particularly in areas like deep learning, natural language processing (NLP), and generative models (e.g., LLMs), have advanced at an astonishing pace [69]. These technologies are no longer confined to specialized scientific domains but are increasingly accessible and applied within digital humanities research, offering new ways to process and interpret vast quantities of historical data [14, 168]. Thirdly, the digitization of historical sources (whether manuscripts, prints, images, recordings or artifacts) has created immense datasets that are often too large for traditional humanistic analysis alone,

making computational approaches, including AI, increasingly attractive and, in some cases, necessary [140, 51]. This confluence creates both unprecedented opportunities for memory studies and significant new responsibilities for scholars engaging with these powerful tools.

## 1.2 Defining the Terrain: History of Memory as Critical Inquiry

To understand the implications of AI, we must first delineate the field it seeks to engage: the History of Memory (HoM). Emerging prominently in the 1980s, HoM shifts the focus from simply reconstructing “what actually happened” to examining how pasts are perceived, constructed, represented, and mobilized in different historical and social contexts [80]. Répétition d’une phrase utilisée plus haut Drawing on foundational work by sociologists like Maurice Halbwachs, who emphasized the social frameworks that shape individual and collective recall [75], HoM scholars investigate memory not as a passive mirror of the past but as an active process of selection, interpretation, and narration. Pierre Nora’s concept of *lieux de mémoire* (sites of memory) highlighted how modern societies, supposedly having lost organic “environments of memory” (*milieux de mémoire*), actively create sites—monuments, archives, commemorations—to anchor and articulate collective memory [121]. This field is inherently critical, attuned to the power dynamics involved in memory making: what is remembered, what is forgotten (or silenced), by whom, and for what purposes [175]. Cultural memory, as explored by Jan and Aleida Assmann, further distinguishes between communicative memory (everyday, short-term) and cultural memory (institutionalized, long-term), which is mediated through texts, rituals, and monuments, forming a society’s understanding of its foundational past [6]. It is this complex, contested, and power-laden understanding of memory that AI now confronts.

## 1.3 Defining the Tool: Understanding Artificial Intelligence in a Humanities Context

Artificial Intelligence, in its contemporary application, refers to computational systems designed to perform tasks that typically require human intelligence. For the purposes of this paper, key AI subfields include: Machine Learning (ML), where systems learn from data without being explicitly programmed, encompassing supervised learning (learning from labeled data) and unsupervised learning (finding patterns in unlabeled data) [153]. Natural Language Processing (NLP) enables computers to process, understand, and generate human language, crucial for analyzing textual archives of memory. Computer Vision allows machines to interpret and understand information from images and videos, relevant for studying visual cultures of memory. More recently, Large Language Models (LLMs) and other Generative AI models have demonstrated remarkable abilities to generate novel text, images, and even code, opening new avenues for creating representations of the past, albeit with significant ethical considerations [9]. It is crucial for humanists to understand that AI systems are not sentient entities possessing understanding in a human sense; rather, they are sophisticated pattern-recognition and prediction tools, whose outputs are fundamentally shaped by the data they are trained on and the algorithms that govern them [37].

## 1.4 Thesis Statement & Argument Roadmap

This paper argues that while artificial intelligence offers transformative tools for exploring, analyzing, and even generating representations of historical memory, its integration into the field necessitates a profound ethical reckoning and a critical recalibration of historical practice. AI does not merely offer new methods; it has the potential to reshape the very nature of memory practices, the epistemological foundations of historical claims based on its outputs, and the power dynamics inherent in the construction and dissemination of past narratives. This performative dimension of AI [106] requires that, without critical engagement, AI risks amplifying historical silences, entrenching biased interpretations, creating new forms of digital misrepresentation, and undermining the historian’s role as a nuanced interpreter of the past.

To develop this argument, this paper will proceed as follows: Part I, “Reconfiguring the Archives and Narratives of Memory,” will examine AI’s potential to expand the scope of memory studies and delve into specific use cases, from textual analysis to generative AI, while simultaneously highlighting the fundamental ethical challenges related to bias, authenticity, and privacy. Part II, “The Algorithmic Politics of Remembrance,” will explore how AI intersects with power dynamics, focusing on issues of control, algorithmic canonization, manipulation, and the tensions between the “right to be forgotten” and the “duty to remember.” Part III, “The Historian’s Evolving Craft,” will consider the changing role of the historian, emphasizing critical AI literacy, interdisciplinarity, and the responsibilities associated with archival practices in an AI era. Part IV, “Towards Ethical Stewardship,” will propose a framework of points to consider for the responsible use of AI in History of Memory research, advocating for proactive professional engagement and illustrating these points with case studies. Ultimately, this paper contends that navigating the algorithmic age requires historians to be not just users of AI, but its critical interrogators and ethical guides, ensuring that technology serves to enrich, rather than erode, our understanding of the multifaceted pasts we inherit and construct.

## Part I

# Reconfiguring the Archives and Narratives of Memory: AI’s Potential and Perils

## 2 Expanding the Scope: AI’s Potential and Positive Applications

The advent of sophisticated Artificial Intelligence tools offers the field of History of Memory unprecedented opportunities to expand its analytical scope and deepen its inquiries. For a discipline that grapples with vast and often diffuse traces of the past—from extensive archival collections to the ephemeral digital footprints of contemporary commemoration—AI

presents capabilities that can augment, and in some cases transform, the historian’s traditional toolkit. This section will explore several key areas where AI’s capacity to process, analyze, and represent memory at scale can make significant positive contributions, illustrated by real-world applications.

## 2.1 Analyzing Vast Textual and Visual Corpora at Scale

One of the most immediate benefits of AI is its ability to process and find patterns within datasets of a magnitude that would be prohibitive for individual researchers or even large teams [104]. This enhanced scale and efficiency in analysis allows for the processing of millions of words or images far more rapidly than human researchers. For instance, tasks previously unimaginable, such as analyzing the entire digitized archive of the USC Shoah Foundation’s testimonies or the vast digitized newspaper archives described by Terras [172] and Robertson and Zaragoza [146] for recurring themes, become feasible [152]. Transfer learning methods are improving the ability of Named Entity Recognition (NER) models to decipher historical texts, even with scarce annotated data and language variations [88]. Similarly, computer vision techniques enable the large-scale analysis of iconographic trends by training models to identify specific symbols across thousands of images, allowing quantitative tracking of shifts in the visual language of public memory [105].

A prime example of this is the *Living with Machines* project, a collaboration between historians and data scientists that used AI to analyze historical newspapers and other textual sources from the 19th century [1, 171]. By employing Natural Language Processing (NLP), the project uncovered new patterns in industrialization, labor, and social change, providing fresh insights into the social and economic impacts of the Industrial Revolution in Britain [171]. Similar efforts to create large-scale structured databases of historical news, like the Newswire dataset, aim to facilitate diverse research but also highlight the need for careful consideration of inherent biases [163].

## 2.2 Identifying Latent Patterns and Uncovering Hidden Connections

Beyond simple keyword searching, AI can facilitate more nuanced analyses. Sentiment analysis tools, while requiring careful calibration, can offer insights into the emotional tenor of texts related to specific memories [98, 114]. Topic modeling algorithms like LDA can uncover latent thematic structures or narrative frames within document sets that might be missed by close reading alone, allowing historians to map the conceptual landscape of memory discourses [19, 83]. In the visual domain, AI can aid in the detection of stylistic evolution [117], analyzing features in art or photography to identify stylistic clusters and trace evolving conventions in commemorative art [48]. This allows historians to use NLP to track the evolution of public discourse surrounding specific historical events, revealing shifts in commemorative language over time [72]. In the same vein, LLMs are now being used to assist research directly, with frameworks like HistoLens [189] combining NER and knowledge graph construction, and COSMOS [190] extracting life trajectories.

In the visual domain, AI can aid in detecting stylistic evolution, analyzing features in art to identify stylistic clusters [48], or using Content-Based Image Retrieval (CBIR) to uncover visual connections between disparate records [40]. Foundation models are now also being used to create image datasets directly from historical documents [74]. Furthermore, network analysis, powered by AI algorithms, can map and visualize relationships between individuals, organizations, texts, and concepts within a memory field [148]. This can illuminate hidden power structures, the flow of ideas, or connections between disparate actors in a social movement [148, 22]. Combining AI with GIS also enables geospatial analysis of memory sites, examining their spatial distribution to understand the formation or contestation of memory in specific regions [71].

### 2.3 Enhancing Accessibility, Discovery, and Preservation

AI holds significant promise for making memory archives more accessible, discoverable, and durable. Advanced semantic search functionalities can help researchers and the public navigate vast digital collections more effectively [165], while AI-driven tools can automate the creation of detailed metadata, tag content, and suggest connections between disparate archival items, potentially uncovering forgotten voices [33, 143]. AI can also connect disparate archival holdings through linked data initiatives, helping researchers trace individuals across institutions [78]. For audiovisual archives, AI-driven automated transcription and indexing can dramatically increase searchability [93]. AI-powered translation tools, while imperfect, can begin to break down linguistic barriers, facilitating comparative studies of memory across different cultural contexts [27, 89, 54].

The **Transkribus** platform is a widely adopted successful example, using AI-powered Handwritten Text Recognition (HTR) to transcribe and make accessible vast collections of historical documents. By allowing users to train custom models, it tailors the transcription process to specific historical contexts, and has been instrumental in digitizing sources like the *Archives of the Spanish Inquisition* [138, 64]. In the realm of preservation and reconstruction, the **Rekrei** project uses crowdsourced images and AI to digitally recreate cultural heritage sites and artifacts destroyed by conflict or natural disasters. It provides a powerful tool for preserving cultural memory and fostering global awareness of heritage loss by offering virtual access to lost treasures from the Mosul Museum or the Notre-Dame Cathedral [144].

### 2.4 Generating New Forms of Engagement and Interpretation

The rise of generative AI opens a new frontier for actively creating novel representations of and interactions with the past. This technology enables new forms of engagement and potentially historical empathy through tools like interactive chatbots "embodying" historical figures or AI-generated first-person narratives based on historical data [86]. These tools can foster different kinds of intuitive grasp of past perspectives, assuming careful design.

A landmark example of this new form of engagement is the USC Shoah Foundation's "Dimensions in Testimony" project. This initiative uses AI to create interactive biographies, allowing museum visitors and students to ask questions and receive real-time responses from

a video projection of a Holocaust survivor. By drawing from a vast database of pre-recorded interviews, the system creates a compelling illusion of a live conversation, fostering a powerful sense of personal connection and historical empathy. This innovative approach promises to preserve the pedagogical power of survivor testimony for future generations, long after the survivors themselves are gone. While ethically complex—a point we will return to—it represents a profound shift in how we might engage with personal memory and difficult pasts.

This approach is not limited to the memory of genocide. A similar and innovative project has been implemented at the Lewarde Historical Mining Centre in Northern France. The "À la rencontre des anciens mineurs" (Meeting the Former Miners)<sup>1</sup> installation uses AI to allow visitors to ask questions, with the system responding by playing specific excerpts from oral testimonies gathered by the museum. This example, rooted in social and industrial history, demonstrates that these AI-assisted testimony technologies are becoming increasingly common mediation tools for a wide range of memories, seeking to compensate for the loss of direct witnesses and to keep a disappearing culture and experience alive.

AI also allows for visualizing lost or inaccessible pasts by creating plausible reconstructions of historical sites or scenes lacking visual evidence, which can aid interpretation and public understanding [43, 120]. Furthermore, carefully designed AI simulations can serve as educational tools for exploring "What If" scenarios, fostering a deeper understanding of historical causality and contingency. While the potential for misuse is significant, these capabilities, when harnessed responsibly, mark a profound shift in how societies can creatively explore, learn from, and engage with their histories.

### 3 Navigating the Perils: Ethical Risks and Methodological Challenges

The transformative potential of Artificial Intelligence detailed in the previous section is inextricably linked to profound ethical and methodological challenges. The deployment of AI in the sensitive domain of historical memory is fraught with risks that demand critical attention, as they can significantly distort how the past is understood, perpetuate historical injustices, and create new forms of epistemic corruption. This section dissects these core perils, using examples of controversial projects to illustrate the real-world consequences of incautious implementation.

#### 3.1 Algorithmic Bias, Misrepresentation, and Decontextualization

A primary danger of using AI in historical memory is the risk of inheriting, amplifying, and "laundering" biases present in the archival record [119, 37]. Models trained on historically biased texts or images can perpetuate stereotypes and underrepresent marginalized voices [9]. Since archives often over-represent dominant groups [175], applying AI tools without

---

<sup>1</sup><https://www.chm-lewarde.com/fr/nouveau-a-la-rencontre-des-anciens-mineurs-le-dispositif-numerique/>

critical intervention can make marginalized memories even less visible. Beyond the content biases within archives, AI is confronted with an even more fundamental methodological bias: that of the medium itself. Archives, in their modern Western conception, are predominantly written documents. This primacy of the written word systematically excludes societies whose modes of transmitting history and memory are based on orality, as is the case in many cultures. An AI trained on digitized textual corpora is therefore structurally incapable of grasping this 'immaterial memory.' In the post-colonial context, for example, the history of local populations remains largely inaudible because it did not leave conventional written traces. A major challenge, therefore, is to consider how to integrate these non-textual forms of memory into databases that can be used by AI without denaturing them. This can manifest as systematic misrepresentation or erasure. A stark example is the *Faces of the Past* project, which used facial recognition on historical photographs. The project faced severe criticism for its poor performance on non-European faces, reflecting the well-documented biases in facial recognition technology and underscoring the risk of using tools that algorithmically erase or misidentify certain communities [136, 150, 28, 5].

Similarly, AI-driven narrative generation can produce biased accounts. The *Deep History* project, which aimed to generate historical narratives from large text datasets, was criticized for producing stories that reinforced existing biases and stereotypes, lacking the nuance and context essential for responsible historical interpretation [116]. This highlights the broader risks of oversimplification and decontextualization, where reducing complex human narratives to quantifiable data points strips away crucial context and the ineffable aspects of lived experience [134, 139]. The emphasis on aggregate patterns can lead to the loss of the individual voice and subjectivity, inadvertently obscuring uniqueness, agency, and culturally specific forms of expression vital to memory.

### 3.2 Erosion of Authenticity, Trivialization, and Manipulation

Generative AI's capacity to create highly plausible synthetic content poses a severe threat to the integrity of the historical record. The potential for "Deepfake History"—the creation of entirely fabricated but realistic-looking narratives, images, or videos—risks a profound erosion of public trust and blurs the lines between evidence and fabrication [50, 183, 177, 67]. Reducing complex historical figures to simplified AI outputs risks trivialization and misrepresentation, as generative models suffer from bias amplification and a lack of contextual understanding regarding sensitive historical aspects [9, 182]. This is compounded by the "black box" nature of many AI models, where the inability to scrutinize *how* a conclusion is reached hinders critical assessment and validation [151, 112]. Interventions like AI-driven colorization or "enhancement" of historical media create ambiguous digital artifacts, altering their historical aura and potentially misleading viewers by creating a false sense of immediacy [99]. This debate moved from the theoretical to the mainstream with Peter Jackson's 2018 film, "They Shall Not Grow Old." The project employed sophisticated AI to colorize, stabilize, and even estimate the correct frame rate of World War I footage, creating a startlingly immediate viewing experience. The real-world implications of this erosion of authenticity are already visible on major platforms. Schuh [160] provides a detailed case

study of an AI-retouched photograph of France Gall and Serge Gainsbourg on Wikimedia Commons. The image was not only colorized but also had its background completely erased and its resolution "enhanced," creating an artifact that was factually inaccurate yet visually plausible. Crucially, Schuh highlights the subsequent debate among Wikipedia editors, which pitted archival principles of documentary traceability against a "realist" desire for a higher-quality image, regardless of its artificial origins. This example perfectly illustrates how AI tools create ambiguous digital objects and force communities to negotiate new, and often contradictory, standards of historical truthfulness in the public sphere. Furthermore, creating "digital revenants" of historical figures for interactive experiences raises serious ethical issues regarding consent and respect for the dead and their descendants [109]. Interventions like AI-driven "restorations" of historical media create ambiguous digital artifacts, altering their historical aura and potentially misleading viewers [11, 99].

Furthermore, interactive AI experiences dealing with trauma could lead to emotional manipulation or unintended psychological impact if not designed with ethical safeguards [157, 127]. Creating "digital revenants" raises serious ethical issues regarding consent and respect for the dead [109], while the concentration of AI development raises concerns about corporate or state shaping of historical storytelling [192, 37, 13].

### **3.3 Privacy, Consent, and Data Protection Concerns**

Many memory archives contain sensitive personal information, and applying AI to them introduces significant privacy risks. Even ostensibly anonymized data can sometimes be re-identified through AI-powered analysis, challenging the adequacy of past consent agreements made before such technologies were conceivable [122, 191]. This is particularly acute in network analysis of sensitive activities, such as mapping connections between activists or dissenters, which could pose surveillance risks or generate spurious correlations [180]. Analyzing personal testimonies also raises major data protection concerns, demanding strict adherence to regulations like GDPR and ethical protocols for anonymization. The fundamental "Garbage In, Garbage Out" principle applies: the quality of any AI analysis, and its ethical standing, hinges entirely on the quality and ethical sourcing of its input data [155].

The fundamental "Garbage In, Garbage Out" principle therefore applies with particular acuity. The quality of any AI analysis, and its ethical standing, hinges entirely on the quality and ethical sourcing of its input data [155]. This principle, however, must be expanded to include not only erroneous information but also the societal silences and 'lies' inscribed within the archives. For instance, it is only very recently, nearly 80 years after the events, that French women have begun to testify about rapes committed by Allied soldiers during the Liberation in 1944-45. An AI trained on the archives and press of that era would inevitably inherit the societal taboo and produce a biased vision of the 'liberators,' ignoring violence that only delayed testimony could reveal. The archive is not a fixed truth, and AI risks petrifying silences that are, in fact, temporary social constructs.

### 3.4 Inaccuracy, Nuance, and the Loss of Authorial Voice

The factual accuracy of generative AI is a primary methodological concern. Several studies have shown that the answers provided by various generative AIs to factual history questions can be imprecise, if not completely wrong [29, 77]. A combined lack of data and embedded biases can lead to these "hallucinations" [103]. If "factual accuracy alone does not guarantee good history" [159, p.27], it is nevertheless an essential prerequisite. This problem is magnified by the strong authoritative dimension of AI-generated statements, which can be accepted uncritically due to automation bias. Indeed, whereas a neural network "sees only in shades of grey" [158], the outputs of generative AI systems often conceal the probabilistic nature of their calculations behind an enunciation in the indicative mood, rather than the conditional mood.

This risk of inaccuracy extends to tools like AI-powered translation. NMT systems still struggle with complex syntax, cultural idioms, irony, and the emotional tones crucial in memory-laden texts, potentially flattening a source's specificity and leading to misinterpretation [16]. Bias in translation models is also a major issue, as they can replicate gender, racial, or cultural biases present in their training corpora [156, 137]. The ease of use of these tools can create an "illusion of unmediated access" and "translation goggles," leading to the uncritical acceptance of machine-generated output. This loss of nuance and potential for error represents a failure to respect the integrity of the source material and the voice of the original author.

### Lessons from the Field

The case studies explored offer valuable lessons. Ensuring transparency and accountability is paramount; projects must meticulously document data sources, methodologies, and assumptions, as exemplified by *Transkribus* and *Living with Machines* [62]. As El-Deeb et al. [42] note, the "black box" nature of some algorithms necessitates clear explanations. Addressing bias and representation requires active measures, including diverse datasets, community involvement, and techniques like intersectional bias detection [28]. Ultimately, these perils highlight that balancing automation with human oversight is crucial. While AI can automate research tasks, human historians must remain central to interpretation and contextualization, ensuring that technology serves, rather than subverts, the goals of responsible historical inquiry, a discussion that continues to evolve in forums on the topic [108].

## 4 The Nature of AI-Generated “Memory”: Representation, Simulation, or Artifact?

Beyond analyzing existing traces of memory, the recent surge in generative Artificial Intelligence capabilities—particularly Large Language Models (LLMs) and image/video synthesis tools—introduces a novel and complex dimension to the study of historical memory: the

AI-generated “memory” object itself. When an AI system produces a narrative of a historical event it was not witness to, a “restored” and colorized photograph of a past scene, or an interactive chatbot embodying a historical figure, what is the epistemological and ontological status of this output? Is it a new form of historical representation, a sophisticated simulation, a digital artifact with its own unique properties, or something else entirely? This section critically examines these questions, exploring how AI-generated content challenges and potentially reconfigures our understanding of memory, authenticity, and the historian’s relationship to the past, particularly through the lens of AI’s performative capacity [106].

A primary concern revolves around the concept of representation. Traditionally, historical representations (texts, images, monuments) are understood as human-mediated interpretations of the past, shaped by evidence, authorial intent, cultural context, and the conventions of the medium [184]. While these representations are never direct transcriptions of reality, they maintain a traceable, albeit complex, link to human agency and source material. AI-generated content, however, complicates this model. An LLM generating a historical narrative does so by learning statistical patterns from vast datasets. Its internal architecture does not function like a database of facts, but more like a complex system of layered “key-value memories” [63]. This means the model “recalls” information by activating and combining learned patterns, rather than retrieving a discrete, preserved source. The resulting output may be coherent and plausible, but its connection to specific evidence is probabilistic and diffuse, posing a fundamental challenge to historical methods of verification and citation. AI-generated content, however, complicates this model. An LLM generating a historical narrative does so by learning statistical patterns from vast datasets of text, not by engaging in critical source evaluation or possessing intentionality in the human sense [9]. The “knowledge” it produces is a probabilistic amalgamation of its training data, potentially leading to outputs that are plausible and coherent yet factually inaccurate, anachronistic, or imbued with the biases inherent in that data [182]. This raises the question: can such an output be considered a historical representation in the same vein as a historian’s monograph, or is it more akin to a sophisticated form of automated pastiche?

The notion of simulation also comes to the fore. AI can create highly immersive and interactive “experiences” of the past, such as virtual reality reconstructions of historical sites populated by AI-driven characters or chatbots designed to mimic historical personalities [86, 85]. These simulations can offer engaging pedagogical tools or novel ways to explore historical environments, potentially enhancing presence and learning [187, 139]. However, they also risk blurring the lines between historical reconstruction and fictional enactment. The perceived realism of an AI simulation might lead users, particularly those without deep historical literacy, to accept the simulated experience as an unproblematic reflection of past reality, obscuring the interpretative choices, simplifications, and potential inaccuracies embedded within the simulation’s design [56]. As Ian Bogost argues in the context of procedural rhetoric, simulations make arguments through their underlying rules and processes [21]; understanding the “arguments” made by AI memory simulations becomes a critical task for historians.

Alternatively, AI-generated content can be viewed as a new category of digital artifact,

possessing its own distinct characteristics and “aura,” or lack thereof, in a Benjaminian sense [11]. An AI-colored photograph is not the original photograph, nor is it a human artist’s interpretation of color; it is a new object created through algorithmic processes. Its historical value may lie less in its claim to accurately represent the past’s chromatic reality and more in what it reveals about contemporary desires for historical immediacy, the capabilities of AI technology at a given point, and the cultural assumptions embedded in the colorization algorithms [99]. These AI-generated artifacts become, in themselves, objects worthy of historical study, reflecting the ways a society uses technology to engage with, and reconstruct, its memory. They could even be considered new forms of Pierre Nora’s *lieux de mémoire*—sites where collective memory crystallizes, albeit in a digitally native and algorithmically mediated form [121]. If so, their analysis demands attention not just to their content, but to their mode of production, their algorithmic underpinnings, and their reception. Generative models like GANs [70] are increasingly used for such reconstructions, for example, in architecture [43, 120], pushing the boundaries of what constitutes a historical artifact.

Engaging with AI-generated “memory” requires historians to develop new critical literacies. It necessitates a shift from primarily analyzing human-authored representations of the past to also interrogating the processes, biases, and epistemological claims of non-human, algorithmic “authors.” The allure of AI’s productive power must be tempered by a rigorous examination of what these generated outputs truly represent, how they function as simulations, and what their status as novel digital artifacts means for the practice of history and the social life of memory. Ignoring these questions risks a future where algorithmically constructed pasts, detached from critical historical inquiry, gain unmerited authority and shape collective understanding in ways that are opaque and potentially misleading.

Ultimately, this distinction between representation, simulation, and artifact highlights a fundamental shift in the media logic of memory. Schuh [160] powerfully articulates this as a move from an “archival logic” to a “logic of analogical generation.” Traditional historical artifacts, even when digitized, are rooted in an archival logic of preserving, retrieving, and decoding stable information. In contrast, AI-generated content follows an analogical logic, where a probabilistic model reconfigures vast training data into a novel output that mimics the past without being a direct trace of it. Understanding this distinction is crucial for historians, as it reframes our task from interpreting evidence to critiquing the very processes of its algorithmic generation.

## Part II

# The Algorithmic Politics of Remembrance: Power, Control, and the Contested Construction of Pasts

## 5 Fundamental Ethical Challenges in AI's Interaction with Memory Archives

The preceding exploration of AI use cases within historical memory studies has highlighted a recurring tension: the transformative potential of these technologies is inextricably linked to profound ethical challenges. While some risks are specific to particular applications, several fundamental ethical issues permeate AI's interaction with memory archives more broadly. These challenges demand critical attention, as they can significantly impact how the past is understood, represented, and preserved, potentially perpetuating historical injustices or creating new forms of epistemic distortion. This section will synthesize and delve deeper into three core areas of ethical concern: algorithmic bias and the perpetuation of historical silences; issues of authenticity, aura, and the "black box" of AI interpretation; and the complexities of privacy and data subjectivity in AI-driven memory work.

### 5.1 Algorithmic Bias and Historical Silences

Archives are not neutral repositories; they are products of power, selection, and interpretation, often marginalizing or silencing certain groups [175, 167]. Historians have long developed critical methodologies for "reading against the archival grain" to uncover the lives of those systematically erased by the historical record, such as the enslaved women whose fragmented stories [59] painstakingly reconstructs from violent and dehumanizing colonial documents. The danger of AI is that it is computationally incapable of this critical, counter-hegemonic reading. Instead, if applied uncritically, AI systems trained on these biased archives risk both replicating and amplifying these historical imbalances on an unprecedented scale. A core problem is the *inheritance and amplification of bias*. Machine learning models learn from their input data; if trained on archives reflecting colonial perspectives, for instance, they are likely to reproduce those perspectives, potentially marginalizing other experiences and effectively "laundering" historical biases under a veneer of algorithmic objectivity [119, 37]. This can lead to direct *representational harms*, where biased AI systems produce skewed depictions of historical groups or events. For example, facial recognition systems trained on limited demographics may misidentify individuals from underrepresented groups in historical photos [28], or NLP models might perpetuate stereotypes found in training data. Furthermore, the development and application of AI tools can *exacerbate the digital divide in memory* [25, 17]. Resource-intensive AI development, concentrated in dominant institutions, risks further marginalizing the memories of less-resourced communities

or the Global South, as choices about digitization and AI application become acts of power. While efforts exist, the *challenge of debiasing* AI remains technically complex and often only partially successful [58, 162], complicated by the fraught interpretive act of defining "fairness" in diverse historical contexts [57, 8]. Intersectional approaches are crucial, recognizing compounded biases across race, gender, and class [38]. The difficulty lies in the ability to translate the concept of intersectionality into quantitative methods, particularly when it is seen as a strategy for optimizing an algorithm rather than as an "analytical sensibility" [124]. In order to guard against the risk of ethics washing, or more specifically fairness washing, it is necessary to question the conceptual framework and technical measures implemented to limit bias. As Youjin Kong has clearly shown, three major problems coexist: the reduction of intersectionality to the consideration of several categories or social attributes, without taking into account overlaps and reinforcement of oppressions; the difficulty of determining the combinations of attributes to "protect", at the risk of defining too many sub-groups or selecting only the variables deemed statistically important (gender and race, for example); the reduction of fairness to distributive justice. Kong advocates for a "strong fairness in AI", which would not only mitigate biases, but also seek to counter the mechanisms of oppression present in society [90]. In the field of history, this means using AI to promote the history of minority groups, for example by increasing the visibility and discoverability of their archives in digital repositories compared with the average, and not just by guaranteeing them the same visibility as data from majority groups.

To counter this risk of petrifying historical silences, a proactive approach is necessary. This first involves broadening the very definition of 'archives' usable by AI beyond just public, predominantly administrative, records. It becomes imperative to integrate a greater diversity of sources—private archives, corporate records, association papers, collections of oral testimonies such as those at France's INA—in order to multiply perspectives. Furthermore, a mitigation strategy could consist of enriching AI models by training (or fine-tuning) them on a corpus of recognized historical scholarship validated by the scientific community. This approach would allow the AI to benefit from already contextualized analyses. It is not, however, without risk: it could lead to the algorithmic canonization of certain historiographical schools or the reproduction of outdated interpretations if the corpus is not carefully selected and updated.

## 5.2 Authenticity, Aura, and the "Black Box" of AI Interpretation

AI's capacity to generate or alter content, combined with the opacity of its decision-making, raises critical questions about authenticity and trustworthiness. There is a significant risk of the *erosion of historical authenticity and the "Liar's Dividend."* AI-generated deepfakes can blur lines between genuine artifacts and fabrications [50], not only enabling deliberate misinformation but also fostering generalized skepticism towards all historical evidence, making factual grounding difficult [31]. Blockchain verification systems are explored as a countermeasure but are not foolproof [30]. AI interventions like colorization also create objects with an *altered aura*. Following Benjamin [11], these AI-mediated artifacts have a complex, potentially misleading relationship to the original, raising questions about whether

they offer closeness to the past or interpose an ahistorical algorithmic layer [99]. The "*black box*" *problem and interpretive opacity* compound these issues [32]. The difficulty in understanding how complex AI models arrive at conclusions [151, 97] hinders critical assessment, validation, and integration into nuanced historical arguments, challenging methodological norms of transparency. Explainable AI (XAI) techniques aim to address this but are not universally applicable [115]. This opacity, coupled with perceived sophistication, can lead to *algorithmic authority and the risk of de-skilling*, where historians might uncritically accept AI findings or become overly reliant on tools without fully grasping their limitations. Moreover, historical proof relies on citing and critically analysing primary sources, which generative AIs generally do not allow. However, RAGs could offer historians a suitable solution, enabling them to benefit from the advantages offered by LLMs [132], while maintaining a focus on the primary source.

### 5.3 Privacy, Data Subjectivity, and the Ethics of AI in Memory Work

Memory archives frequently contain sensitive personal information, demanding careful ethical consideration when AI is applied. *Re-identification and anonymity challenges* are significant; even ostensibly anonymized data can sometimes be re-identified through AI linking disparate datasets or inferring identities from subtle patterns, particularly in rich personal archives [122]. This raises questions about *informed consent in the age of AI*. Historical subjects who contributed to archives previously could not have foreseen current AI capabilities, challenging the adequacy of past consent and highlighting researchers' ethical responsibilities [191]. AI's capacity for *algorithmic profiling and inference of sensitive attributes* (like political affiliation or health status) from historical data, even if not explicitly stated, poses potential harms if such inferred data is used or publicized. The tension between the "*Right to be Forgotten*" and AI's "*Total Recall*" capacity is exacerbated [52], as AI makes perpetual access and recirculation of potentially painful past information easier, conflicting with needs for reconciliation or privacy [107]. Regulations like GDPR [181] and techniques like differential privacy [47] aim to protect individuals. GDPR, as well as multiple national privacy laws, only concern living individuals. Similarly, IRBs and RECs are only concerned with the protection of living research participants. When the rights of the dead are considered, it is generally about how the disclosure of certain information could harm their living relatives. Several historians, notably Antoon De Baets, have asked which rights the dead might have, and which obligations historians should respect regarding to them. In the name of the posthumous dignity of the dead, De Baets asserts that historians have a "duty to weigh the privacy and reputation of the dead against the public interest" when deciding which information to disclose [41, p.123]. This also means that it is sometimes necessary to pseudonymize the data acquired on the dead, although no legislation requires it. As is often the case, the use of AI does not change the nature of the problem, but amplifies it, both because the volume of digital traces left by an individual, sometimes without being aware of it, is increasing, and because the technological capacity to retrieve and collate data linked to a person is constantly improving. AI makes it all the more necessary to reflect on the dignity of the dead, and on the limits to be set to the historian's work in exploiting

data acquired without the authorization of the people concerned, particularly if they were not public figures. Finally, applying AI to archives documenting trauma requires *respect for vulnerable populations and traumatic memories*, guided by an ethics of care to avoid re-traumatization, disrespect, or exploitation, potentially necessitating community consultation and trauma-informed design [176]. Addressing these fundamental challenges requires ongoing critical reflection, interdisciplinary dialogue, and the development of robust ethical frameworks by the historical profession.

## 6 Revisiting Power in HoM: From Human Actors to Algorithmic Governance

A central tenet of the History of Memory (HoM) is its critical attention to the operations of power in shaping what is remembered, how it is remembered, by whom, and for what purposes [123, 55]. Historians of memory have long investigated how states, social groups, institutions, and individuals vie to establish dominant narratives of the past, often marginalizing or actively suppressing alternative or counter-memories [175]. This "politics of regret," of commemoration, and of forgetting is a fundamental aspect of how societies construct their identities and legitimize present conditions [34]. The introduction of Artificial Intelligence into this already contested terrain does not merely add a new tool; it potentially introduces new actors, new forms of control, and new mechanisms of "algorithmic governance" over historical memory [188], compelling a revisitation of how power operates in the digital age.

Traditionally, the power to shape collective memory was often associated with control over key institutions and media: archives, educational curricula, state-sponsored commemorations, museums, and mass media [3]. Those who controlled these "means of memory production" could exert significant influence over public understanding of the past. While these traditional loci of power remain vital, AI introduces new dimensions. Firstly, *power becomes embedded in code and data*; it is not just wielded by human actors using technology, but resides within the algorithms and datasets themselves [119, 37]. Design choices about data inclusion, parameters, and optimization objectives are inherently political, systematically privileging certain memories. Secondly, we witness the *rise of new "Memory Keepers."* Large tech corporations controlling dominant AI platforms (search engines, social media, generative models) act as powerful, often unaccountable, arbiters of historical information [66], their algorithmic curation subtly shaping public interpretation. This shift of power to platform corporations has been a long-term development. As Vaidhyanathan [178] influentially argued, the "Googlization of everything" represents a profound restructuring of how we access knowledge, where public information is filtered through a system optimized for commercial engagement rather than scholarly or civic principles. Applied to historical memory, this means that the "relevance" of the past is increasingly determined by opaque, corporate-owned algorithms, turning public memory itself into a curated and monetizable product. Thirdly, AI enables *automated decision-making in memory dissemination*. Algorithmic gatekeeping prioritizes certain narratives as "relevant" or "authoritative," shifting from explicit human editorial control to a more diffuse, data-driven, and potentially less

contestable form of power. Finally, the *scale and speed of algorithmic influence* are unprecedented. AI can amplify narratives—accurate or misleading—far faster and more broadly than traditional media, potentially overwhelming nuanced accounts and introducing new challenges like viral deepfakes.

This shift necessitates an expansion of HoM’s analytical framework. While continuing to scrutinize the actions of human agents (states, activists, communities), historians of memory must now also critically examine the "agency" of algorithms, the political economy of AI development, and the ways in which algorithmic systems reconfigure the power dynamics within the field of memory. The subsequent sections in this part will explore how these new algorithmic power dynamics manifest in issues such as the (re)centralization of memory production, the risks of algorithmic canonization and manipulation, and the ongoing struggle over the "right to be forgotten" versus the "duty to remember" in an AI-suffused world. Understanding these algorithmic dimensions of power is crucial for grasping the contemporary politics of historical remembrance.

## 7 AI and the (Re)Centralization of Memory Production and Dissemination

The historical study of memory has often tracked the interplay between centralized forces (e.g., state narratives, national museums) and decentralized memory practices (e.g., local communities, family histories, grassroots movements). The advent of digital technologies, particularly the early internet and social media, was initially heralded by some as a democratizing force, potentially allowing for a wider proliferation of diverse and counter-memories [82]. However, the development and deployment of sophisticated Artificial Intelligence systems, especially those requiring vast computational resources and massive datasets, raise concerns about a potential (re)centralization of power over the means of memory production, curation, and dissemination, albeit in new forms.

This (re)centralization trend stems from several factors. The *concentration of AI development and infrastructure* places cutting-edge research, particularly for large-scale models, predominantly within a few large tech corporations and state-backed initiatives [37, 36]. These entities control the necessary computational power, proprietary datasets, and engineering talent, meaning the tools shaping historical memory are owned by a select few whose interests may not align with broader societal needs. This trend can be understood as a form of "digital extractivism" applied to collective memory [160]. In this model, vast public and historical datasets—including centuries of digitized cultural heritage—are treated as a raw resource, a form of "memory capital," to be extracted and refined by a few powerful private entities. These corporations then control not only the proprietary models but also the "actionable memory" these models produce, effectively commercializing access to our collective past and shaping it according to their own logics. This is reinforced by *proprietary algorithms and "walled gardens" of data*. Many AI platforms operate as "black boxes," their algorithms and training data opaque [129]. When used to surface historical information (e.g., search results), the criteria are often hidden, potentially biasing access

towards platform-mediated interpretations. Furthermore, the *power of default settings and algorithmic paternalism* subtly guides users towards certain content or interpretations [188]. These defaults, operating at scale, can have a widespread centralizing effect on public historical consciousness, reflecting platform priorities rather than diverse scholarship. We also see increasing *state adoption of AI for "official" memory projects*, such as national digital archives with AI interpretation layers or tools to combat "historical revisionism" (as defined by the state). This represents a powerful new form of centralized memory control, risking the codification of state-sanctioned narratives [46, 96]. Finally, significant *economic barriers to independent AI-powered memory work* exist. The computational resources needed for large models or datasets can be prohibitive for smaller institutions, archives, or community projects, particularly in the Global South, leading to a de facto centralization where only well-funded entities can fully leverage AI's potential.

This (re)centralization dynamic does not necessarily negate the potential for AI to be used by smaller actors or for counter-hegemonic memory work. Open-source models, community data initiatives, and critical AI literacy can offer pathways for resistance and alternative uses. This (re)centralization dynamic does not necessarily negate the potential for AI to be used by smaller actors or for counter-hegemonic memory work. Open-source models, community data initiatives, and critical AI literacy can offer pathways for resistance and alternative uses. Projects like the BigScience Workshop's BLOOM model [94] demonstrate that large-scale, multilingual AI development can be achieved through open, collaborative, and non-commercial efforts, providing a crucial alternative to the proprietary systems of large tech corporations. However, the dominant trend towards concentration of AI power poses a significant challenge to the diversity and democratic contestation of historical memory. It underscores the need for critical scrutiny of who owns and controls the algorithmic infrastructure shaping our engagement with the past, and for advocating for more equitable access and transparent governance of these powerful new "means of memory production."

## 8 Ethical Quagmires in Algorithmic Memory Politics

The increasing entanglement of Artificial Intelligence with the production, dissemination, and governance of historical memory gives rise to a host of complex ethical quagmires specifically situated at the intersection of algorithms and memory politics. These are not merely technical issues but deeply political ones, as AI systems become implicated in shaping public understanding of the past, legitimizing certain narratives, and potentially marginalizing others. This section examines several key ethical challenges that emerge when AI is deployed within the contested arena of historical memory.

### 8.1 Algorithmic Canonization and the "Objectivity Trap"

AI systems, particularly those powering search engines, recommendation algorithms, or even generative models creating historical summaries, can inadvertently or intentionally lead to an "algorithmic canonization" of certain historical interpretations or memories. This arises

partly from the *illusion of neutrality*, where algorithms are often perceived by the public as objective arbiters of information [65, 4]. When AI consistently prioritizes certain historical sources or narratives, these can gain an unwarranted aura of authority, becoming de facto "official" versions, even if they reflect biases. This process often involves *reinforcing dominant narratives*, as AI models trained predominantly on mainstream historical accounts or archives tend to reproduce established societal memories, making it harder for alternative or marginalized memories to gain visibility [119]. Furthermore, the *"filter bubble" effect*, driven by personalization algorithms, can extend to historical understanding, predominantly exposing users to interpretations aligning with their existing beliefs and limiting engagement with diverse perspectives, potentially polarizing historical understanding [126, 169]. Ultimately, algorithmic canonization poses significant *challenges to pluralism in memory*. By flattening the inherent complexity and contestation of historical memory, it risks promoting singular or oversimplified understandings at the expense of scholarly debate and diverse lived experiences.

## 8.2 Manipulation, Misinformation, and “Deepfake Memory”

The capacity of AI to generate highly realistic synthetic content (deepfakes) and to disseminate information (and misinformation) at scale poses a severe threat to the integrity of historical memory and its political uses. Malicious actors can engage in the *weaponization of AI for historical revisionism*, creating and propagating fabricated "evidence" like fake photos or synthetic accounts to promote political agendas, deny atrocities, or rewrite national histories [50, 183, 67]. The proliferation of such AI-generated misinformation contributes to an *erosion of trust in historical sources* more broadly, making it difficult to establish shared factual bases for discussing the past and benefiting those who wish to sow confusion (the "liar's dividend") [31, 149]. AI can also power *automated propaganda and influence operations* related to historical events, tailoring messages to specific audiences and overwhelming online discussions with biased or false narratives to manipulate public opinion. AI can therefore address the three levels of abuse and irresponsible use of history identified by Antoon de Baets: heuristic (irresponsible destruction, collection and use of sources; fabrication of sources); epistemological (irresponsible description and analysis of data); and pragmatic (lies about oneself or works; censorship of works; providers of information, assignments, contracts and funding; editors and publishers; peer reviewers; and audiences) [41, p.17-22, 182].

Compounding these threats is the *difficulty of detection and attribution*; identifying sophisticated AI-generated misinformation is technically challenging, and attributing it to specific actors is often even harder, making accountability elusive. Although assessing intent was important for distinguishing between recklessness, negligence and deception, it will become increasingly difficult to characterise the misuse of history from a moral and legal perspective. While irresponsible or malicious uses of history are not new, AI introduces many new challenges by operating on a vast scale and blurring the lines of responsibility.

However, it is crucial to note that AI acts here as a powerful amplifier of a pre-existing phenomenon rather than the creator of the problem. The manipulation of history for political or commercial ends is an old practice. A contemporary case in point in France is the Puy

du Fou theme park, which disseminates a biased, counter-revolutionary, and ultra-Catholic vision of French history to a very wide audience under the guise of entertainment. Millions of visitors, often lacking the tools of historical criticism, absorb this narrative as factual truth. AI, in this context, can be understood as an uncritical visitor: it is incapable of distinguishing a rigorous reconstruction from effective propaganda. If trained on corpora that include such popular and biased productions, it risks reproducing and massively disseminating this misinformation, lending it a new form of algorithmic authority [15].

### 8.3 Algorithmic Discrimination and Representational Harms in Memory

The biases embedded in AI systems, as discussed earlier, can translate into direct forms of algorithmic discrimination and representational harm when applied to historical memory, particularly for marginalized communities. This can manifest as *systematic misrepresentation or erasure*. If AI models are trained on data that underrepresents or stereotypically portrays certain groups (e.g., ethnic minorities, women, LGBTQ+ individuals), the resulting AI-driven narratives, search results, or visual representations may perpetuate these harms, effectively distorting or erasing their memories [10, 76, 170]. Algorithmic systems might also create *differential access to memory resources*, inadvertently making materials related to dominant groups more discoverable than those pertaining to marginalized communities, thus fostering an unequal digital landscape. Applying biased AI tools within *sensitive historical contexts*, such as archives related to genocide or colonial violence, carries a high risk of causing profound offense and re-traumatization if outputs are inaccurate, disrespectful, or perpetuate harmful stereotypes.

### 8.4 The Algorithmic “Right to be Forgotten” vs. the “Duty to Remember”

The tension between societal needs for forgetting (for reconciliation, privacy, moving on) and the ethical or political “duty to remember” certain pasts (for prevention, justice, identity) is significantly complicated by AI’s capabilities [100, 53]. AI’s *capacity for “total recall”*—storing, processing, and retrieving vast amounts of historical data efficiently—can make it difficult for individuals or societies to move beyond past events, as previously obscured information becomes perpetually accessible [107]. This poses *challenges to the individual “Right to be Forgotten,”* as AI-driven search and aggregation make it harder for individuals to de-index or reduce the accessibility of certain aspects of their past, even if historically recorded. Moreover, AI can become a *tool for enforcing or resisting state-mandated forgetting/remembering*, used by states to selectively erase or suppress memories, or conversely, to ensure constant remembrance of sanctioned narratives, turning AI into a battleground for memory control. This highlights the complex *ethics of algorithmic forgetting*: if AI systems are designed to “forget,” the criteria used are ethically and politically charged, raising concerns about censorship and the erasure of inconvenient truths or marginalized histories. The *right to be forgotten* poses an ethical challenge for historians, who must perform a “balancing test to determine whether omitting sensitive facts and opinions about the

dead, the benefit gained in terms of privacy and reputation protection outweighs the harm inflicted on freedom of expression and historical truth" [41, p.130]. This raises the question of whether such ethical judgments could or should be entrusted to AI. While research in this field is progressing, AI moral decision-making currently lacks a consensus on the principles and criteria needed, whereas human moral decision-making often depends more heavily on contextual factors [141]. Navigating these ethical quagmires requires not only technical solutions but also robust public debate, democratic oversight, and a renewed commitment from historians to critically engage with the political and ethical dimensions of AI's growing influence on historical memory.

## 8.5 Environmental and Societal Costs as Political Choices

A focus on the ethics of AI that is limited to data and algorithms is incomplete. It is paradoxical to question the content of AI outputs while ignoring the profound environmental and societal consequences of the systems that produce them. The decision to invest in and deploy AI is not merely a methodological choice but a political act with significant, often hidden, material costs [37]. These considerations are especially critical for memory institutions like archives and libraries, whose core mission often includes long-term preservation and equitable access.

First, the environmental footprint of large-scale AI is immense and globally unequal. The production of AI hardware relies on the extraction of minerals, often under exploitative labor conditions. The training and operation of complex models, particularly LLMs, consume vast amounts of energy and water for data center cooling [9]. Finally, the rapid cycle of technological obsolescence contributes to a growing e-waste crisis, the burdens of which are disproportionately borne by communities in the Global South. As Crawford [37] argues in *Atlas of AI*, this asymmetry in environmental harm mirrors the asymmetries of power seen in data production, creating new forms of resource colonialism under the guise of technological progress. An ethical assessment of an AI project in historical memory must therefore account for its position within this global chain of extraction and disposal.

Second, the financial investment required for AI creates significant opportunity costs within cultural heritage institutions. The high cost of computational infrastructure (e.g., GPU clusters), specialized software, and expert staff (whether recruiting expert staff or upskilling current one) represents a major reallocation of resources. For an archive or university, funding a large AI initiative may mean defunding other vital areas: the conservation of fragile physical artifacts, the digitization of less "AI-friendly" collections, or the hiring of more archivists or public-facing staff. These budgetary decisions are de facto statements of institutional priority, potentially valuing computational analysis over other forms of historical labor and preservation.

Conversely, while AI is often framed as a strategic necessity in an era of shrinking cultural and humanities budgets, the economic narrative of technological efficiency is more complex. Like previous digital transformation efforts, AI-powered platforms such as Transkribus are rarely one-time investments and may not deliver the promised cost savings. Initial enthusiasm about AI-driven metadata generation or transcription must be tempered by careful

economic analysis. The apparent efficiency gains can mask substantial hidden costs: ongoing software licenses, computational resources, technical staff training, and long-term dependencies on proprietary AI companies. Further research is needed to comprehensively assess the real economic impacts, examining not just immediate operational costs but also the broader institutional implications of technological dependence [24, 68]. From this perspective, AI emerges not as a straightforward solution to budget constraints, but as a nuanced technological intervention requiring critical and sustained economic scrutiny.

Ultimately, navigating these realities requires transparency and a conscious weighing of trade-offs. The allure of innovation cannot be a sufficient justification on its own. Institutions must critically assess whether the potential for new discoveries and efficiencies afforded by AI outweighs the environmental impact and the opportunity costs for other essential forms of historical and archival work. These are not technical questions, but fundamental political and ethical questions about what aspects of the past we choose to invest in, and how.

### Part III

## The Historian's Evolving Craft: Agency, Responsibility, and Methodological Futures in an AI-Infused World

### 9 The Indispensable Historian: Critical Judgment in an Age of Intelligent Machines (The Historian's Role)

The integration of Artificial Intelligence into the study of historical memory, with its attendant promises and perils, necessitates a profound reflection on the role and responsibilities of the historian. Far from rendering the historian obsolete, the age of intelligent machines underscores the indispensable nature of human critical judgment, contextual understanding, ethical awareness, and interpretive skill. AI can be a powerful tool, but it remains just that—a tool. Its effective and ethical deployment in historical inquiry depends entirely on the historian's capacity to guide its application, scrutinize its outputs, and situate its "findings" within broader humanistic frameworks. This section argues that the historian's role evolves to become not only a user of AI but, more crucially, its critical interrogator, ethical navigator, and thoughtful curator.

Central to this evolving role is the continued application of core historical skills, albeit extended to new domains. *Source criticism in the algorithmic age* remains fundamental [20, 174]. Historians must now evaluate not only traditional evidence but also the "algorithmic sources"—the training data, model architecture, parameters, and potential biases of AI tools—requiring a new layer of methodological rigor. Equally vital is *contextualization as*

*a core humanistic contribution.* While AI excels at pattern recognition, it lacks the deep understanding of historical periods, cultural nuances, and human motivations that informs historical interpretation. The historian's ability to contextualize AI-derived patterns and integrate them into richer narratives remains a uniquely human task [72]. Furthermore, historians serve as the *interpreter, not merely a data processor.* Constructing coherent, meaningful, evidence-based arguments involves synthesis, narrative construction, and theoretical engagement far beyond AI's current capabilities [140]. Historians must now evaluate not only traditional evidence but also the "algorithmic sources"—the training data, model architecture, parameters, and potential biases of AI tools. This is not merely a technical task but a critical, interpretive one. As Ramsay [142] has argued for the digital humanities, the act of building and deconstructing computational systems is itself a form of scholarly argument. To ethically use AI, therefore, historians must be able to critically assess the "interpretive choices" embedded within the algorithm itself.

The rise of AI also introduces new responsibilities. Historians must engage in *ethical navigation and responsible stewardship*, ensuring their use of AI respects privacy, avoids harm, and promotes honest engagement with the past, including transparency about tool usage and limitations [2]. When AI generates content, the historian becomes a critical *curator and critique of AI-generated "memories,"* evaluating their accuracy, plausibility, and underlying assumptions, and clearly distinguishing them from human scholarship. Finally, historians have a role in *advocating for transparency and accountability in AI for history*, contributing to public understanding of AI's limitations and calling for accountability when it distorts the past. In essence, the AI era demands an "augmented historian"—digitally literate yet grounded in critical inquiry, ethical reflection, and nuanced interpretation.

## 10 Fostering Critical AI Literacy and Interdisciplinarity

The responsible and effective integration of Artificial Intelligence into the study of historical memory necessitates more than just individual historians honing their critical judgment; it demands a broader cultivation of "critical AI literacy" within the historical profession and a fundamental commitment to interdisciplinary collaboration. Navigating the complexities of AI—its technical underpinnings, its embedded biases, its societal implications, and its potential for historical inquiry—cannot be achieved in disciplinary isolation. Historians must engage with experts from other fields, and the profession as a whole must invest in developing new pedagogical approaches and research methodologies.

### 10.1 Developing Critical AI Literacy for Historians

Critical AI literacy for historians extends beyond merely knowing how to operate specific AI software or tools. It involves a deeper, more conceptual understanding. Historians need a foundational grasp of *core AI concepts and limitations*, including what AI can and cannot do, how models are trained, the risk of confusion between correlation and causation, and the probabilistic nature of many systems, recognizing that AI processes patterns, not

"understands" history [26]. Crucially, this includes *recognizing algorithmic bias and its historical roots*, identifying potential biases stemming from training data or design choices, and understanding how these distort analysis [119, 10]. Historians must be equipped for *evaluating the epistemological claims of AI outputs*, critically assessing their methodological underpinnings rather than accepting them at face value. This literacy also encompasses *ethical reasoning and AI governance awareness*, understanding broader debates around AI ethics (privacy, fairness, etc.) and relevant governance frameworks. Finally, *integrating AI into historical pedagogy* is essential, preparing future historians to engage thoughtfully with these technologies by incorporating critical AI literacy into curricula [125]. However, initial training is not enough, especially when the pace of technological innovation is accelerating. Initiatives to share open-access learning resources, such as Programming Historian<sup>2</sup>, should be encouraged.

## 10.2 The Imperative of Interdisciplinary Collaboration

The multifaceted challenges and opportunities presented by AI in historical memory studies necessitate robust collaboration across disciplinary boundaries. Historians cannot and should not attempt to become AI experts in isolation, nor can computer scientists fully grasp the nuances of historical inquiry or memory theory without humanistic input. Essential partnerships include those between *historians and computer scientists/data scientists* to develop useful and ethically sound tools, combining domain expertise with technical skill [14, 147, 1]. *Engagement with ethicists and legal scholars* is vital for navigating the profound ethical and legal questions raised by AI concerning bias, privacy, and misinformation. *Collaboration with archivists and librarians* is key, as these professionals curate, preserve, and provide access to the data AI relies on and grapple with AI's implications for information management [33]. For projects dealing with sensitive memories, *involving affected communities through participatory design* is ethically imperative to ensure respect, accuracy, and alignment with community interests [166, 35, 18]. Fostering such collaboration requires *building interdisciplinary research centers and training programs*, with institutional support for joint projects and hybrid training [7, 130, 128]. Embracing critical AI literacy and genuine interdisciplinarity is essential for conducting rigorous, ethical, and relevant historical memory research in the 21st century.

## 10.3 Mediation as an Ethical and Practical Imperative

Finally, beyond the training of historians and interdisciplinary collaboration, the responsible integration of AI, particularly in public-facing memory projects (museums, online platforms, educational applications), demands a third pillar: mediation. Mediation is the active process by which an institution explains and contextualizes not only the historical content but also the tools used to present it. In the case of AI, this means not simply offering a 'magical' experience, but making the tool's functioning transparent: explaining what data it was

---

<sup>2</sup><https://programminghistorian.org/>

trained on, what its known biases are, and what its limitations are. This approach transforms the user, whether a visitor or a student, from a passive consumer into a critical and conscious actor, capable of understanding that the tool is not an oracle but the result of a series of human, technical, and ethical choices.

## 11 Archiving the Algorithm: Long-Term Preservation and Methodological Transparency (Long-term preservation)

The increasing use of Artificial Intelligence in historical memory research introduces new and complex challenges for long-term preservation, methodological transparency, and the future citability and critique of AI-assisted scholarship. Traditional historical scholarship relies on the citability of sources and the transparency of methods, allowing future historians to re-examine evidence and reinterpret arguments. If AI-driven historical "findings" are to become a legitimate and enduring part of the scholarly record, then the "algorithmic black box" must, to the greatest extent possible, be made archivable, understandable, and re-analyzable by future generations of researchers. This necessitates a concerted effort to develop best practices for "archiving the algorithm" itself, alongside the data it processes and the interpretations it generates.

A primary obstacle is the *ephemerality of AI models and software environments*. Complex AI architectures often depend on specific software libraries and hardware configurations that rapidly become obsolete, threatening long-term reproducibility [79]. Essential for future understanding is *preserving training data and its provenance*. Since model outputs are fundamentally shaped by training data, future historians need access to the specific dataset version used, along with comprehensive metadata on its origin, scope, limitations, and pre-processing steps [185, 135]. Methodological transparency also requires *documenting model architecture, parameters, and code*, including the rationale behind design choices, to enable future researchers to understand how results were generated [133]. Frameworks like "model cards" and "datasheets for datasets" offer useful approaches [111, 62]. However, the *challenge of archiving large-scale models and datasets* presents significant technical and financial hurdles for traditional institutions [110]. These preservation efforts must also navigate *ethical considerations in archiving AI and data*, particularly regarding privacy in sensitive datasets and the handling of proprietary models. The ultimate goal is *ensuring future re-interpretability and critical engagement*, allowing historians not just to replicate but to critically re-engage with past AI-assisted research in light of new knowledge. Achieving this requires *developing new archival practices and infrastructures*, potentially involving techniques like emulation or containerization, through collaboration between historians, archivists, librarians, and computer scientists [45, 44, 49]. Failure to address these issues risks creating an opaque and ephemeral body of scholarship.

## 12 Navigating Professional Ethics and Responsibilities

The integration of Artificial Intelligence into historical memory research not only reshapes methodologies but also places new demands on the professional ethics and responsibilities of historians.

While historians and applied AI ethicists come from vastly different traditions, it may be worthwhile to compare the ethical frameworks that have emerged in each field in recent years. International recommendations on trustworthy AI have proliferated, and these guidelines have increasingly been applied to the humanities as well. However, it is important to note that there is no universal consensus on ethical codification within the historical profession. Internationally, few historians' associations or societies have chosen to adopt formal codes of ethics. Antoon De Baets's observation in 2009 regarding the paucity of ethical reflection in the actual practice of historical research remains valid [41, p.173].

One commonality between historians' ethics and AI ethics is a tendency to prioritize processual ethics over procedural ethics as practiced by research ethics committees or institutional review boards. While some have argued for the complementarity of these two approaches, it is true that both historical research and AI development frequently fall outside the purview of formal ethical oversight bodies, unless they involve living human subjects or personal data.

Existing codes of conduct, such as the American Historical Association's Statement on Standards of Professional Conduct, emphasize principles of integrity, honesty in research, respect for evidence, and responsible engagement with the public [2]. These foundational principles remain paramount and must be thoughtfully applied and extended to the novel challenges posed by AI. This section examines key areas where historians' professional ethics are particularly tested by the use of AI in memory work, focusing on accountability, the implications for scholarly labor (deskilling vs. upskilling), and issues of equity and access [101].

### 12.1 Accountability and Liability in AI-Assisted Historical Work

The use of AI, especially complex or "black box" models, raises critical questions about accountability when historical interpretations or representations generated with AI assistance prove to be flawed, biased, or harmful. Historians hold *responsibility for algorithmic outputs*; they cannot abdicate accountability for their claims or evidence simply because AI was involved [12]. This necessitates a deep understanding of the tools and a critical stance. *Transparency in method and limitations* is crucial for professional integrity, requiring clear disclosure of AI methodologies, datasets, parameters, and known biases to allow for peer review. If AI-assisted work causes harm, historians have an ethical duty toward *addressing algorithmic harms* by acknowledging, correcting, and mitigating issues, potentially involving revisions or engagement with affected communities. In collaborative projects, *navigating shared responsibility* is essential; clear agreements on roles, responsibilities, and ethical oversight are needed from the outset when working with computer scientists and others [164]. Ethical issues are all the more acute in public history projects, where sources of funding and

their links with historical supervision need to be made explicit [87, 154].

## 12.2 Deskillling vs. Upskilling: The Impact of AI on Historical Labor and Training

The increasing sophistication of AI tools raises questions about their impact on the nature of historical labor and the skills required of future historians. There is *potential for deskilling certain tasks*, as AI might automate routine activities like basic searching or transcription. While this could free up time, over-reliance risks diminishing core competencies if not managed carefully. More significantly, AI necessitates *the need for new skills and "upskilling"* among historians, including critical AI literacy, data management, understanding computational methods, and effective interdisciplinary collaboration [14]. This requires adaptation in graduate training and professional development. Consequently, there may be a need for *redefining scholarly contribution and value* to recognize the intellectual labor involved in designing AI projects, curating data, and interpreting complex outputs [173]. Amidst this technological shift, it remains vital to *maintain a humanistic focus*, ensuring that training prioritizes critical thinking, nuanced interpretation, ethical reasoning, and deep contextual knowledge, with technology serving, not supplanting, these core skills.

Since the debate initiated by Franco Moretti around the notion of distant reading, more and more voices from digital humanities have called for "a space for quantitatively minded digital scholarship that goes beyond the trend of big data, allowing us to craft digital hermeneutic strategies for a handful of texts or even for just one" [155]. Whether this approach is referred to as "mixed methods" or "scalable reading" [92], it always aims to see digital analysis tools not as an end, but as a means among others.

Unlike the algorithmic and visualization tools that have characterized the digital humanities, the development of AI in historical research will not replay the debate between quantitativists and qualitativists. As we have seen, its applications lie not only in the processing of massive corpora, but also in the renewed questioning and analysis of single texts, insinuating themselves even into the writing practices of historians.

## 12.3 Equity of Access and the Digital Divide in AI-Powered Memory Research

The benefits of AI in historical memory research risk being unevenly distributed, potentially exacerbating existing inequalities. *Unequal access to computational resources and expertise* creates barriers for scholars at smaller institutions, independent researchers, or those in the Global South, potentially establishing new hierarchies in knowledge production [113]. Even supposedly *bias in "open" resources* can be problematic, as open-source models or public datasets may still carry embedded biases or require significant technical know-how, limiting true accessibility. Furthermore, the question of *whose memories are AI-ready* arises, as digitization efforts often focus on dominant cultures or well-funded institutions, leaving other memories in analog obscurity and further marginalizing them. Addressing these inequities requires *promoting open science and equitable collaboration*. Professional organi-

zations and funding agencies have a responsibility to support initiatives that foster more equitable access to AI tools, data, and training, encourage inclusive collaborations, and prioritize projects addressing historical silences or empowering marginalized communities. Ultimately, navigating these professional responsibilities requires proactive, reflective, and collective engagement from historians.

## Part IV

# Towards Ethical Stewardship: A Framework for Critical Practice with AI in Historical Memory

## 13 Recapitulation of Core Ethical Dilemmas at the Intersection of AI and HoM

The preceding analysis has navigated the complex and rapidly evolving landscape where Artificial Intelligence intersects with the History of Memory. We have explored AI's transformative potential to expand the scale and scope of memory research—from processing vast textual and visual archives to generating novel interactive experiences with the past. However, woven throughout this exploration has been a consistent thread of profound ethical dilemmas that demand careful consideration and proactive engagement from the historical profession. Before proposing a framework for critical practice, it is pertinent to briefly recapitulate these core ethical challenges.

These challenges coalesce around several key themes. *Bias, representation, and historical justice* remain primary concerns, given the risk of AI inheriting, amplifying, and "laundering" historical biases, leading to misrepresentation or erasure of marginalized memories and perpetuating injustices. Issues of *authenticity, interpretation, and epistemological integrity* arise from AI's capacity for synthetic content generation and its opaque interpretive processes, threatening historical authenticity, eroding public trust, and challenging scholarly norms. The concentration of AI development raises questions about *power, control, and algorithmic governance of memory*, including risks of (re)centralization, algorithmic canonization, and manipulation. The application of AI to archives containing personal data necessitates careful navigation of *privacy, consent, and data subjectivity*, complicated by AI's "total recall" capacity. Finally, these factors impose new demands regarding *professional responsibility, accountability, and equity* on historians, requiring critical AI literacy, transparency, and attention to potential deskilling or exacerbation of inequalities. These interwoven dilemmas underscore that AI is a powerful socio-technical force requiring ethical stewardship guided by rigor, fairness, transparency, and a commitment to historical truth.

## 14 A Proposed Framework: Points to Consider When Using AI for Historical Memory Research

Navigating the complex ethical and methodological landscape at the intersection of Artificial Intelligence and the History of Memory requires a proactive and principled approach from researchers, archivists, educators, and institutions. Building upon the critical analysis presented in this paper, this section proposes a framework of key points to consider when conceptualizing, developing, deploying, or evaluating AI applications within historical memory studies. These points are intended not as rigid regulations, but as a heuristic guide to foster responsible innovation, critical reflection, and ethical stewardship. They can be adapted and expanded based on the specific context of the research or project.

### I. Data Governance, Provenance, and Bias Mitigation

The foundation of any AI system is its data. Ensuring ethical and rigorous data practices is paramount.

- A.1 Conduct Thorough Provenance Audits for Training Data:* Meticulously document the origin, scope, collection methods, and known limitations or biases of all datasets used to train or inform AI models. The history of large open datasets like "The Pile" [60], which faced challenges due to the inclusion of copyrighted and problematic content scraped from the web, underscores the critical importance of this step. Understand the historical context of the archive's creation [167]. Prioritize documentation quality over dataset size, "only creating datasets as large as can be sufficiently documented" [9].
- A.2 Implement Bias Detection and Mitigation Strategies:* Actively seek to identify potential demographic, ideological, or representational biases within datasets. Where feasible and appropriate, employ technical or methodological strategies to mitigate these biases, while acknowledging that perfect "debiasing" is often elusive [57].
- A.3 Prioritize Representativeness and Inclusivity in Data Collection and Curation:* When creating new datasets or digitizing archives for AI use, strive for representativeness and actively seek out sources that can counter historical silences and include marginalized voices [39]. Document any gaps.
- A.4 Adhere to Data Privacy and Protection Regulations:* Ensure full compliance with relevant data protection laws (e.g., GDPR) and ethical guidelines for handling personal or sensitive information within historical memory archives, especially when using AI for analysis or dissemination [191, 118].
- A.5 Utilize Datasheets for Datasets:* Adopt practices like "Datasheets for Datasets" to systematically document the motivation, composition, collection process, recommended uses, and potential risks of datasets used in AI-driven memory research [62].

*A.6 Assess the financial and environmental costs:* Consider the impact of institutional financial choices and benchmark the energy consumption of the project. Take these dimensions into account in the ethical assessment of project risks and benefits [9]

## II. Algorithmic Transparency, Explainability, and Validation

The "black box" nature of some AI systems challenges scholarly norms of transparency and verifiability.

*B.1 Strive for Algorithmic Transparency and Explainability (XAI):* Where possible, favor AI models whose decision-making processes are more interpretable. When using complex models, explore XAI techniques to gain insights into how outputs are generated, and clearly communicate the level of opacity involved [151, 145].

*B.2 Document Model Design, Parameters, and Limitations:* Provide detailed documentation of the AI model's architecture, algorithms, chosen parameters, and the specific version of software/libraries used. Clearly state the model's known limitations and the scope of its applicability. Practices like "Model Cards" can be beneficial [111].

*B.3 Validate AI Outputs Rigorously Against Diverse Historical Evidence:* Do not accept AI-generated "findings" or representations at face value. Triangulate AI outputs with traditional historical sources, qualitative analysis, and expert domain knowledge. Subject AI-assisted historical claims to rigorous peer review [23].

*B.4 Distinguish Between Correlation and Causation:* Be cautious about inferring historical causation from statistical correlations identified by AI. Algorithmic patterns require careful historical interpretation to establish their meaning and significance [131].

## III. Human Oversight, Agency, and Critical Interpretation

AI should augment, not replace, human historical judgment and ethical reasoning.

*C.1 Maintain the Historian's Critical Role in the Loop:* Ensure that human historians with relevant domain expertise and ethical awareness are actively involved in all stages of an AI project—from design and data selection to interpretation of results and dissemination.

*C.2 Establish Mechanisms for Correction and Contestation of AI Outputs:* Develop processes for identifying, correcting, and publicly acknowledging errors or biases in AI-generated historical content or analysis. Allow for contestation and alternative interpretations [81].

*C.3 Clearly Label AI-Generated or AI-Assisted Content:* Transparently identify historical content (narratives, images, analyses) that has been wholly or substantially generated or modified by AI to avoid misleading audiences about its provenance and nature [102].

*C.4 Prioritize Interpretive Depth over Computational Scale:* While AI allows for analysis at scale, ensure that this does not come at the expense of nuanced historical interpretation, contextual understanding, and engagement with the complexity of human experience in the past.

## **IV. Stakeholder Engagement, Community Consultation, and Social Responsibility**

AI-driven memory work has societal implications that extend beyond academia.

*D.1 Engage with Affected Communities, Especially for Sensitive Memories:* For projects dealing with traumatic, contested, or culturally sensitive memories, consult and collaborate with relevant stakeholder communities (e.g., victim groups, descendant communities, indigenous peoples) to ensure respectful, ethical, and meaningful engagement [166].

*D.2 Consider Co-Design and Participatory Approaches:* Where appropriate, involve community members in the design and development of AI tools or memory projects to ensure they align with community needs, values, and perspectives [35].

*D.3 Assess Potential Societal Impacts and Dual-Use Concerns:* Reflect on the broader societal implications of the AI research, including potential misuse of the technology or its findings (e.g., for political manipulation, perpetuation of stereotypes) [161].

*D.4 Promote Public Understanding and Critical AI Literacy:* Historians have a role in educating the public about how AI is being used to interpret and represent the past, fostering critical engagement rather than passive consumption of AI-generated historical content.

*D.5 Implement Active Mediation Strategies for the Public:* For any project using AI that is intended for a non-expert audience, develop and implement clear mediation strategies. These should explain in an accessible manner how the AI tool works, the data it is based on, its limitations, and its potential biases. The goal is to foster critical literacy among users and to ensure that engagement with the past remains a reflective experience rather than a passive consumption of algorithmic content.

## **V. Professional Development, Collaboration, and Preservation**

The historical profession must adapt to responsibly integrate AI.

*E.1 Invest in Critical AI Literacy Training for Historians:* Support the development of pedagogical resources and training programs to equip historians with the necessary skills to understand, use, and critique AI.

*E.2 Foster Equitable Interdisciplinary Collaboration:* Encourage and support genuine collaborations between historians, computer scientists, ethicists, archivists, and other relevant experts, based on mutual respect and shared goals.

*E.3 Develop Standards for Archiving AI Models, Data, and Code:* Promote and adopt practices for the long-term preservation and accessibility of AI-related research materials to ensure reproducibility, future critique, and methodological transparency [133, 79].

*E.4 Advocate for Ethical AI Governance and Public Interest Technology:* Encourage historical associations and individual scholars to contribute to broader discussions about ethical AI governance and the development of AI in the public interest.

This framework is intended to stimulate ongoing discussion and adaptation. As AI technologies continue to evolve, so too must our ethical and methodological considerations for their application in the sensitive and vital work of historical memory.

## 15 A Call for Proactive Professional Engagement and Guideline Development

The framework of considerations outlined above provides a starting point for individual researchers and project teams. However, addressing these profound challenges requires a collective, proactive engagement from the historical profession. This call for critical self-reflection is part of a larger tradition within digital scholarship. More than a decade ago, **Liu2012culturalcriticism** challenged the digital humanities to move beyond the mere instrumental use of tools and to critically engage with the political, social, and cultural infrastructures that sustain them. In the age of AI, this imperative is more urgent than ever; simply reacting to technological developments as they arise is insufficient.

The framework of considerations outlined above provides a starting point for individual researchers and project teams. However, addressing the profound ethical and methodological challenges posed by Artificial Intelligence at its intersection with historical memory requires more than individual vigilance; it necessitates a collective, proactive, and sustained engagement from the historical profession as a whole. Simply reacting to technological developments as they arise is insufficient. Instead, historical associations, academic departments, journals, and funding agencies must take a leading role in fostering critical dialogue, developing ethical guidelines, and shaping the responsible integration of AI into historical practice.

Several key areas warrant focused professional attention:

- *Development of Ethical Guidelines and Best Practices:* Professional historical organizations (such as the American Historical Association, the Royal Historical Society, and their international counterparts) should consider developing specific addenda or supplementary guidelines to their existing codes of conduct that directly address the ethical use of AI in historical research, teaching, and public engagement [2, 73]. These guidelines could draw upon frameworks similar to the one proposed in this paper, tailored to the specific needs and values of the historical discipline. This might include

recommendations on data provenance, algorithmic transparency, bias mitigation, consent, and the responsible representation of AI-generated historical content. In doing so, they would not be starting from scratch, but could learn from the ongoing, real-world negotiations within other knowledge communities. The Wikimedia Foundation, for instance, has already established detailed internal guidelines concerning the use of AI for image retouching and text generation, mandating human review, clear labeling, and justification for any alterations. These community-led efforts highlight the pressing need for such protocols and can serve as a valuable model for the historical profession.

- *Fostering Interdisciplinary Dialogue and Training Initiatives:* The profession should actively promote and fund workshops, conferences, and training initiatives that bring together historians, computer scientists, ethicists, archivists, and community stakeholders to discuss AI's impact on history and memory. Graduate programs in history need to consider how to integrate critical AI literacy and digital methodologies into their curricula [14, 147].
- *Supporting Open Science and Equitable Access:* Historical associations and funding bodies can play a crucial role in advocating for and supporting open-source AI tools, open datasets (where ethically appropriate), and infrastructures that promote equitable access to computational resources for historical research, particularly for scholars from under-resourced institutions or the Global South [113].
- *Establishing Peer Review Standards for AI-Assisted Scholarship:* History journals and academic presses will need to develop clear standards for the peer review of scholarship that employs AI methodologies. This includes expectations for methodological transparency, documentation of data and code, and critical assessment of algorithmic choices and outputs. Reviewers with relevant digital and AI literacy will be increasingly necessary.
- *Public Engagement and Education on AI and History:* Historians have a vital role to play in educating the public about the capabilities and limitations of AI in interpreting the past, the risks of AI-driven historical misinformation, and the importance of critical thinking when encountering AI-generated historical content [179]. Professional organizations can support public history initiatives that address these issues.
- *Advocacy for Ethical AI Governance Beyond Academia:* The historical profession should consider lending its voice and expertise to broader societal discussions about the ethical governance of AI, particularly concerning issues of historical truth, cultural heritage, and the potential for AI to be used for political manipulation or the distortion of memory. Historians can offer valuable long-term perspectives on the societal consequences of new technologies.
- *Ongoing Monitoring and Adaptation:* Given the rapid pace of AI development, any guidelines or best practices established will require ongoing review and adaptation.

The historical profession needs to create mechanisms for continuously monitoring AI's evolving impact on historical memory studies and updating its ethical and methodological responses accordingly.

Proactive engagement is not merely about mitigating risks; it is also about responsibly harnessing AI's potential to enrich historical understanding and make the study of memory more inclusive and impactful. By taking collective ownership of these challenges and opportunities, the historical profession can help ensure that the integration of AI into its practices is guided by scholarly integrity, ethical responsibility, and a commitment to serving the public good. The alternative is to cede the shaping of our algorithmic engagement with the past to purely technological or commercial imperatives, a prospect that carries significant risks for the future of historical knowledge and memory.

## **Conclusion: Memory's Future, Historians' Choices – Navigating the Algorithmic Age with Critical Wisdom**

The encroachment of Artificial Intelligence into nearly every facet of human endeavor presents both exhilarating prospects and profound anxieties. For the History of Memory, a field dedicated to understanding how societies construct, contest, and make meaning of their pasts, AI offers a powerful, yet ethically fraught, new frontier. This paper has navigated this evolving terrain, charting AI's potential to revolutionize research methodologies—from analyzing vast archives at unprecedented scale to generating novel forms of engagement with historical narratives—while simultaneously dissecting the complex ethical dilemmas that accompany these advancements. We have seen how AI can risk amplifying historical biases, eroding notions of authenticity, reconfiguring power dynamics in memory production, and challenging the very nature of historical inquiry and professional responsibility. The performative turn, as articulated by Matei [106], further compels us to recognize AI not merely as a representational tool but as an active agent in shaping how the past is remembered and understood.

At its core, the intersection of AI and historical memory compels us to confront fundamental questions about knowledge, power, and representation. The allure of algorithmic efficiency and the promise of uncovering hidden patterns in the past must be tempered by a critical awareness of AI's limitations, its embedded biases, and its potential for misuse. AI systems are not neutral conduits to the past; they are socio-technical artifacts shaped by human choices, commercial interests, and existing societal inequalities. Their outputs, therefore, require the same, if not more, rigorous scrutiny that historians apply to any other source or interpretation [95, 91].

The path forward is not one of Luddite rejection nor of uncritical adoption. Instead, it demands a nuanced and proactive engagement from historians, characterized by critical AI literacy, a commitment to interdisciplinary collaboration, and an unwavering adherence to ethical principles. The historian's role, far from being diminished, becomes even more crucial: as a critical interrogator of algorithmic claims, an ethical navigator of complex data

landscapes, a thoughtful curator of AI-generated content, and a vital public intellectual helping society understand how its memories are being shaped by these new technologies. The "Points to Consider" framework proposed herein offers a practical starting point for this engagement, but it must be complemented by a collective professional commitment to developing robust guidelines, fostering equitable access, and adapting pedagogical and research practices.

Ultimately, the future of historical memory in an algorithmic age will be determined by the choices historians make today. Will we cede the interpretation and representation of the past to opaque algorithms and powerful tech platforms, or will we actively shape the development and deployment of AI in ways that enhance historical understanding, promote justice, and preserve the integrity of our discipline? The challenge is to harness AI's capabilities responsibly, ensuring that it serves as a tool for deeper insight and broader inclusivity, rather than a force that flattens complexity, silences dissent, or distorts our collective grasp of the past. Navigating this new era requires not just technical skill, but critical wisdom—a wisdom rooted in the enduring humanistic values of evidence-based inquiry, ethical reflection, and a profound respect for the multifaceted human experiences that constitute history itself. The future of memory is not predetermined by technology; it is, and must remain, a domain of critical human choice and scholarly responsibility.

## Author Contributions

According to the CRediT taxonomy: **Frederic Jurie:** Writing – original draft. **Hind Bouchareb:** Writing – review & editing.

## Acknowledgments

The authors wish to express their gratitude to Corentin Rousman, Curator at the Pole Museal of the City of Mons Memory and Military History, for his insightful and helpful comments. We also gratefully acknowledge the financial support of the 'AI Remembers' European project, which made this research possible.

## References

- [1] Ruth Ahnert et al. *Collaborative Historical Research in the Age of Big Data: Lessons from an Interdisciplinary Project*. Elements in Historical Theory and Practice. Cambridge, UK: Cambridge University Press, 2023. DOI: 10.1017/9781009175548.
- [2] American Historical Association. *Statement on Standards of Professional Conduct*. 2019. URL: <https://www.historians.org/jobs-and-professional-development/statements-standards-and-guidelines-of-the-discipline/statement-on-standards-of-professional-conduct> (visited on 02/10/2024).

- [3] Benedict Anderson. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso, 1983.
- [4] C. W. Anderson. *Apostles of Certainty: Data Journalism and the Politics of Doubt*. Oxford, UK: Oxford University Press, 2018.
- [5] Jerone Andrews et al. “Ethical considerations for responsible data curation”. In: *Advances in Neural Information Processing Systems 36*. Neural Information Processing Systems Foundation, 2023, pp. 55320–55360. URL: [https://papers.neurips.cc/paper/2023/file/ad3ebc951f43d1e9ed20187a7b5bc4ee-Paper-Datasets\\_and\\_Benchmarks.pdf](https://papers.neurips.cc/paper/2023/file/ad3ebc951f43d1e9ed20187a7b5bc4ee-Paper-Datasets_and_Benchmarks.pdf).
- [6] Jan Assmann. *Cultural Memory and Early Civilization: Writing, Remembrance, and Political Imagination*. Cambridge University Press, 2012. ISBN: 978-0-521-19161-5. DOI: 10.1017/CB09780511996306. URL: <https://www.cambridge.org/core/books/cultural-memory-and-early-civilization/>.
- [7] Anne Balsamo and Carl Mitcham. “Interdisciplinarity in Ethics and the Ethics of Interdisciplinarity”. In: *The Oxford Handbook of Interdisciplinarity*. Ed. by Robert Frodeman, Julie Thompson Klein, and Carl Mitcham. Oxford University Press, 2010, pp. 259–272.
- [8] Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press, 2019.
- [9] Emily M. Bender et al. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’21)*. ACM, 2021, pp. 610–623. DOI: 10.1145/3442188.3445922. URL: <https://dl.acm.org/doi/10.1145/3442188.3445922>.
- [10] Ruha Benjamin. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press, 2019.
- [11] Walter Benjamin. “The Work of Art in the Age of Mechanical Reproduction”. In: *Illuminations*. Ed. by Hannah Arendt. Trans. by Harry Zohn. New York, NY, USA: Schocken Books, 1968 [1955], pp. 217–251. URL: <https://books.google.com/books?id=yeE0AQAAIAAJ>.
- [12] Sarah V. Bentley et al. “The Digital Divide in Action: How Experiences of Digital Technology Shape Future Relationships with Artificial Intelligence”. In: *AI & Ethics* 4.4 (2024), pp. 901–915. DOI: 10.1007/s43681-024-00452-3. URL: <https://doi.org/10.1007/s43681-024-00452-3>.
- [13] Naja BENTZEN. *Online Information Manipulation and Information Integrity*. Tech. rep. EPRS briefing report on online manipulation and integrity issues in information environments. European Parliamentary Research Service, 2024.
- [14] David M. Berry. *Understanding Digital Humanities*. London, UK: Palgrave Macmillan, 2012. ISBN: 9780230292659. URL: <https://www.palgrave.com/gp/book/9780230292659>.

- [15] Florian Besson et al. *Le Puy du Faux Enquête sur un parc qui déforme l'Histoire*. Les Arènes, 2022.
- [16] Łucja Biel. “Quality in institutional EU translation: parameters, policies and practices”. In: *Quality Aspects in Institutional Translation*. Language Science Press, Nov. 2017, pp. 31–57. DOI: 10.5281/zenodo.1048183. URL: <https://doi.org/10.5281/zenodo.1048183>.
- [17] Abeba Birhane. “Algorithmic Injustice: A Relational Ethics Approach”. In: *Patterns* 2.2 (2021), p. 100205. ISSN: 2666-3899. DOI: 10.1016/j.patter.2021.100205.
- [18] Abeba Birhane et al. “The Values Encoded in Machine Learning Research”. In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’22)*. ACM, 2022, pp. 173–184. DOI: 10.1145/3531146.3533082. URL: <https://doi.org/10.1145/3531146.3533082>.
- [19] David M. Blei. “Probabilistic Topic Models”. In: *Communications of the ACM* 55.4 (2012), pp. 77–84. DOI: 10.1145/2133806.2133826. URL: <https://dl.acm.org/doi/10.1145/2133806.2133826>.
- [20] Marc Bloch. *The Historian’s Craft*. Trans. by Peter Putnam. Originally published 1949. New York: Vintage Books, 1953.
- [21] Ian Bogost. *Persuasive Games: The Expressive Power of Videogames*. The MIT Press, 2007. ISBN: 9780262268912. DOI: 10.7551/mitpress/5334.001.0001. URL: <https://doi.org/10.7551/mitpress/5334.001.0001>.
- [22] Stephen P. Borgatti, Martin G. Everett, and Jeffrey C. Johnson. *Analyzing Social Networks*. 2nd. Thousand Oaks, CA, USA: SAGE Publications Ltd, 2018. ISBN: 9781506389997. URL: <https://us.sagepub.com/en-us/nam/analyzing-social-networks/book246451>.
- [23] Christine L. Borgman. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press, 2015.
- [24] Elena Borin and Fabio Donato. “Financial Sustainability of Digitizing Cultural Heritage: The International Platform Europeana”. In: *Journal of Risk and Financial Management* 16.10 (2023), p. 421. ISSN: 1911-8074. DOI: 10.3390/jrfm16100421. (Visited on 07/16/2025).
- [25] Danah Boyd and Kate Crawford. “Critical Questions for Big Data: Provocations for a Cultural, Technological and Scholarly Phenomenon”. In: *Information, Communication & Society* 15.5 (2012), pp. 662–679.
- [26] Meredith Broussard. *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge, MA: MIT Press, 2018.
- [27] Siobhan Brownlie. “Translation and the (de)construction of memory in a network of great historical documents”. In: *Translation Studies* 6.1 (2013), pp. 48–63. DOI: 10.1080/14781700.2012.748387. URL: <https://doi.org/10.1080/14781700.2012.748387>.

- [28] Joy Buolamwini and Timnit Gebru. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification”. In: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. Ed. by Sorelle A. Friedler and Christo Wilson. Vol. 81. Proceedings of Machine Learning Research. PMLR, Feb. 2018, pp. 77–91. URL: <https://proceedings.mlr.press/v81/buolamwini18a.html>.
- [29] Mathieu Chartier et al. “HiBenchLLM: Historical Inquiry Benchmarking for Large Language Models”. In: *Data & Knowledge Engineering* 156 (2025), p. 102383. ISSN: 0169-023X. DOI: 10.1016/j.datak.2024.102383. URL: <https://www.sciencedirect.com/science/article/pii/S0169023X24301677>.
- [30] Yourong Chen et al. “A Survey on Blockchain Systems: Attacks, Defenses, and Privacy Preservation”. In: *High-Confidence Computing* 2.2 (2022), p. 100048. ISSN: 2667-2952. DOI: 10.1016/j.hcc.2021.100048.
- [31] Robert Chesney and Danielle Keats Citron. “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security”. In: *California Law Review* 107.6 (2019), pp. 1753–1820. URL: <https://scholarship.law.berkeley.edu/californialawreview/vol107/iss6/3>.
- [32] Cary Coglianese and David Lehr. “Transparency and Algorithmic Governance”. In: *Administrative Law Review* 71.1 (2019), pp. 1–56. ISSN: 0001-8368.
- [33] Giovanni Colavizza et al. “Archives and AI: An overview of current debates and future perspectives”. In: *ACM Journal on Computing and Cultural Heritage (JOCCH)* 15.1 (2021), pp. 1–15. DOI: 10.1145/3437970. URL: <https://dl.acm.org/doi/10.1145/3437970>.
- [34] Alon Confino. “Collective Memory and Cultural History: Problems of Method”. In: *The American Historical Review* 102.5 (1997), pp. 1386–1403. DOI: 10.2307/2171069. URL: <https://doi.org/10.2307/2171069>.
- [35] Sasha Costanza-Chock. *Design Justice: Community-Led Practices to Build the Worlds We Need*. Cambridge, MA: MIT Press, 2020.
- [36] Nick Couldry and Ulises A. Mejias. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford, CA: Stanford University Press, 2019.
- [37] Kate Crawford. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT, USA: Yale University Press, 2021. ISBN: 9780300264630. URL: <https://yalebooks.yale.edu/book/9780300264630/atlas-of-ai/>.
- [38] Kimberle Crenshaw. “Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics”. In: *University of Chicago Legal Forum* 1989.1 (1989), pp. 139–167.
- [39] Catherine D’Ignazio and Lauren F. Klein. *Data Feminism*. Cambridge, MA: MIT Press, 2020.

- [40] Ritendra Datta et al. “Image retrieval: Ideas, influences, and trends of the new age”. In: *ACM Computing Surveys (CSUR)* 40.2 (2008), pp. 1–60. DOI: 10.1145/1348246.1348248. URL: <https://doi.org/10.1145/1348246.1348248>.
- [41] Antoon De Baets. *Responsible History*. New York, NY: Berghahn Books, 2009. ISBN: 9781845455415.
- [42] Sara El-Deeb et al. “Towards Responsible AI: Exploring AI Frameworks, Ethical Dimensions and Regulations”. In: *Market Grooming*. Ed. by Sumesh Singh Dadwal, Hamid Jahankhani, and Kenneth Revett. Emerald Publishing Limited, 2024, pp. 139–157. ISBN: 978-1-83549-002-0. DOI: 10.1108/978-1-83549-001-320241007.
- [43] Emilio Delgado-Martos et al. “Automatic Virtual Reconstruction of Historic Buildings through Deep Learning. A Critical Analysis of a Paradigm Shift”. In: *Beyond Digital Representation: Advanced Experiences in AR and AI for Cultural Heritage and Innovative Design*. Ed. by Andrea Giordano, Michele Russo, and Roberta Spallone. Springer Nature Switzerland, 2024, pp. 415–426. ISBN: 978-3-031-36155-5.
- [44] Janet Delve and David Anderson, eds. *Preserving Complex Digital Objects*. London: Facet Publishing, 2014.
- [45] Digital Preservation Coalition. *Digital Preservation Coalition*. International organization promoting digital preservation best practices. URL: <https://www.dpconline.org>.
- [46] José van Dijck, Thomas Poell, and Martijn de Waal. *The Platform Society: Public Values in a Connective World*. Oxford: Oxford University Press, 2018.
- [47] Cynthia Dwork. “Differential Privacy”. In: *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006*. Springer, 2006, pp. 1–12.
- [48] Ahmed Elgammal et al. “CAN: Creative Adversarial Networks, Generating “Art” by Learning About Styles and Deviating from Style Norms”. In: *Proceedings of the Eighth International Conference on Computational Creativity (ICCC 2017)*. 2017, pp. 83–91. URL: <https://arxiv.org/abs/1706.07068>.
- [49] Tom Evens and Laurence Hauttekeete. “Challenges of digital preservation for cultural heritage institutions”. In: *Journal of Librarianship and Information Science* 43.3 (2011), pp. 157–165. DOI: 10.1177/0961000611400452.
- [50] Hany Farid. “Image Forensics”. In: *Annual Review of Vision Science* 5 (2019). Discusses deepfakes within image forensics context., pp. 549–573. DOI: 10.1146/annurev-vision-091718-014827. URL: <https://www.annualreviews.org/doi/10.1146/annurev-vision-091718-014827>.
- [51] Domenico Fiormonte, Teresa Numerico, and Francesca Tomasi. *The Digital Humanist: A Critical Inquiry*. Punctum Books, 2015. ISBN: 9780692580448.
- [52] Luciano Floridi. *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford University Press, 2014.

- [53] Luciano Floridi et al. “AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations”. In: *Minds and Machines* 28.4 (2018), pp. 689–707. DOI: 10.1007/s11023-018-9482-5. URL: <https://doi.org/10.1007/s11023-018-9482-5>.
- [54] Mikel L. Forcada. “Making sense of neural machine translation”. In: *Translation Spaces* 6.2 (2017), pp. 291–309. DOI: 10.1075/ts.6.2.06for. URL: <https://doi.org/10.1075/ts.6.2.06for>.
- [55] Michel Foucault. *Discipline and Punish: The Birth of the Prison*. Trans. by Alan Sheridan. New York, NY: Pantheon Books, 1977.
- [56] Gonzalo Frasca. “Simulation versus Narrative: Introduction to Ludology”. In: *The Video Game Theory Reader*. Ed. by Mark J.P. Wolf and Bernard Perron. New York, NY, USA: Routledge, 2003, pp. 221–235. URL: <https://www.routledge.com/The-Video-Game-Theory-Reader/Wolf-Perron/p/book/9780415934940>.
- [57] Sorelle A. Friedler et al. “A comparative study of fairness-enhancing interventions in machine learning”. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT\* ’19)*. ACM, 2019, pp. 329–338. DOI: 10.1145/3287560.3287589. URL: <https://doi.org/10.1145/3287560.3287589>.
- [58] Batya Friedman and Helen Nissenbaum. “Bias in Computer Systems”. In: *ACM Transactions on Information Systems* 14.3 (1996), pp. 330–347. DOI: 10.1145/230538.230561.
- [59] Marisa J. Fuentes. *Dispossessed Lives: Enslaved Women, Violence, and the Archive*. University of Pennsylvania Press, 2016.
- [60] Leo Gao et al. “The Pile: An 800GB Dataset of Diverse Text for Language Modeling”. In: *CoRR* abs/2101.00027 (2021). DOI: 10.48550/arXiv.2101.00027. arXiv: 2101.00027. URL: <https://arxiv.org/abs/2101.00027>.
- [61] Giselle Gonzalez Garcia and Christian Weilbach. *If the Sources Could Talk: Evaluating Large Language Models for Research Assistance in History*. Ed. by Artjoms Sela, Fotis Jannidis, and Iza Romanowska. 2023. URL: <https://ceur-ws.org/Vol-3558/paper6960.pdf>.
- [62] Timnit Gebru et al. “Datasheets for Datasets”. In: *Communications of the ACM* 64.12 (2021), pp. 86–92. DOI: 10.1145/3458723. URL: <https://dl.acm.org/doi/10.1145/3458723>.
- [63] Mor Geva et al. “Transformer Feed-Forward Layers Are Key-Value Memories”. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Ed. by Marie-Francine Moens et al. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 5484–5495. DOI: 10.18653/v1/2021.emnlp-main.446. URL: <https://aclanthology.org/2021.emnlp-main.446/>.

- [64] Trishita Ghosh et al. “Advances in Online Handwritten Recognition in the Last Decades”. In: *Computer Science Review* 46 (2022), p. 100515. DOI: 10.1016/j.cosrev.2022.100515.
- [65] Tarleton Gillespie. “The Relevance of Algorithms”. In: *Media Technologies: Essays on Communication, Materiality, and Society*. Ed. by Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot. Cambridge, MA: MIT Press, 2014, pp. 167–193.
- [66] Tarleton Gillespie. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press, 2018.
- [67] Liang Yu Gong and Xue Jun Li. “A Contemporary Survey on Deepfake Detection: Datasets, Algorithms, and Challenges”. In: *Electronics* 13.3 (2024), p. 585. ISSN: 2079-9292. DOI: 10.3390/electronics13030585. URL: <https://www.mdpi.com/2079-9292/13/3/585>.
- [68] Megan Gooch, Rebecca Kahn, and Eburne Kugeler. “Human Perspectives and Social Infrastructures: Prioritising People in GLAM Digitisation”. In: *Journal of Open Humanities Data* 11.1 (2025). ISSN: 2059-481X. DOI: 10.5334/johd.274. (Visited on 07/16/2025).
- [69] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. Cambridge, MA: MIT Press, 2016. ISBN: 9780262035613. URL: <https://www.deeplearningbook.org/>.
- [70] Ian J. Goodfellow et al. “Generative Adversarial Nets”. In: *Advances in Neural Information Processing Systems* 27 (2014).
- [71] Ian N. Gregory and Richard G. Healey. “Historical GIS: structuring, mapping and analysing geographies of the past”. In: *Progress in Human Geography* 31.5 (2007), pp. 638–653. DOI: 10.1177/0309132507081495. URL: <https://doi.org/10.1177/0309132507081495>.
- [72] Jo Guldi and David Armitage. *The History Manifesto*. Cambridge, UK: Cambridge University Press, 2014. ISBN: 9781107076341. URL: <https://www.cambridge.org/core/books/history-manifesto/720AD79879A2B48852E46F5A7F3F33D4>.
- [73] Thilo Hagenorff. “The Ethics of AI Ethics: An Evaluation of Guidelines”. In: *Minds and Machines* 30.1 (2020), pp. 99–120. ISSN: 1572-8641. DOI: 10.1007/s11023-020-09517-8.
- [74] Hassan El-Hajj and Matteo Valleriani. “Prompt me a Dataset: An Investigation of Text-Image Prompting for Historical Image Dataset Creation Using Foundation Models”. In: *International Conference on Image Analysis and Processing*. Springer. 2023, pp. 247–257.

- [75] Maurice Halbwachs. *On Collective Memory*. Ed. and trans. by Lewis A. Coser. Edited and translated from French with introduction by Lewis A. Coser. Chicago: University of Chicago Press, 1992. ISBN: 978-0-226-11596-2. URL: <https://press.uchicago.edu/ucp/books/book/chicago/0/bo3619875.html>.
- [76] Alex Hanna et al. “Towards a Critical Race Methodology for Algorithmic Fairness”. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT\* ’20)*. ACM, 2020, pp. 501–512. DOI: 10.1145/3351095.3372826. URL: <https://doi.org/10.1145/3351095.3372826>.
- [77] Jakob Hauser et al. “Large Language Models’ Expert-level Global History Knowledge Benchmark (HiST-LLM)”. In: *Advances in Neural Information Processing Systems*. Vol. 37. Curran Associates, Inc., 2024, pp. 32336–32369. URL: [https://proceedings.neurips.cc/paper/2024/file/38cc5cba8e513547b96bc326e25610dc-Paper-Datasets\\_and\\_Benchmarks\\_Track.pdf](https://proceedings.neurips.cc/paper/2024/file/38cc5cba8e513547b96bc326e25610dc-Paper-Datasets_and_Benchmarks_Track.pdf).
- [78] Tom Heath and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers, 2011. DOI: 10.2200/S00334ED1V01Y201102WBE001. URL: <https://doi.org/10.2200/S00334ED1V01Y201102WBE001>.
- [79] Ben Hutchinson et al. “Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure”. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’21)*. ACM, 2021, pp. 560–575. DOI: 10.1145/3442188.3445910. URL: <https://doi.org/10.1145/3442188.3445910>.
- [80] Patrick H. Hutton. *History as an Art of Memory*. Burlington, VT: University of Vermont Press, 1993. ISBN: 0874516374. URL: <https://www.upress.virginia.edu/title/9780874515488/>.
- [81] Sheila Jasanoff. *The Ethics of Invention: Technology and the Human Future*. New York: W. W. Norton & Company, 2016.
- [82] Henry Jenkins. *Convergence Culture: Where Old and New Media Collide*. New York: New York University Press, 2006.
- [83] Matthew L. Jockers. “Macroanalysis: Digital Methods and Literary History”. In: (2013).
- [84] Wulf Kansteiner. “Digital Doping for Historians: Can History, Memory, and Historical Theory Be Rendered Artificially InTELLigent?” In: *History and Theory* 61.4 (2022), pp. 119–133. ISSN: 1468-2303. DOI: 10.1111/hith.12282.
- [85] Matthew Wilhelm Kapell and Andrew B. R. Elliott, eds. *Playing with the Past: Digital Games and the Simulation of History*. Bloomsbury, 2013.

- [86] Andreas Kaplan and Michael Haenlein. “Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence”. In: *Business Horizons* 62.1 (2019). Broader article, touches on AI interactions., pp. 15–25. DOI: 10.1016/j.bushor.2018.08.004. URL: <https://doi.org/10.1016/j.bushor.2018.08.004>.
- [87] Theodore J. Karamanski. “Ethics and Public History: An Introduction”. In: *The Public Historian* 8.1 (1986), pp. 5–12. ISSN: 0272-3433. DOI: 10.2307/3377069. (Visited on 07/14/2025).
- [88] Sojung Lucia Kim et al. *Transfer Learning Across Several Centuries: Machine and Historian Integrated Method to Decipher Royal Secretary’s Diary*. 2023. arXiv: 2306.14592.
- [89] Philipp Koehn. *Neural Machine Translation*. Cambridge, UK: Cambridge University Press, 2020. URL: <https://www.cambridge.org/core/books/neural-machine-translation/2A4E19DFB0C5D1C8C5ACD2F5E0E69C74>.
- [90] Youjin Kong. “Are “Intersectionally Fair” AI Algorithms Really Fair to Women of Color? A Philosophical Analysis”. In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. Association for Computing Machinery, 2022, pp. 485–494. DOI: 10.1145/3531146.3533114.
- [91] Sergey P. Kostrikov, Alexandr A. Iliyuhov, and Elena V. Kryukova. “Historical Science and the Digitalization”. In: *Socio-Economic Systems: Paradigms for the Future*. Ed. by Elena G. Popkova, Victoria N. Ostrovskaya, and Aleksei V. Bogoviz. Springer International Publishing, 2021, pp. 181–190. ISBN: 978-3-030-56433-9. DOI: 10.1007/978-3-030-56433-9\_20.
- [92] Benjamin Krautter. “The Scales of (Computational) Literary Studies: Martin Mueller’s Concept of Scalable Reading in Theory and Practice”. In: 2023, pp. 261–286. ISBN: 9783111317779. DOI: 10.1515/9783111317779-011.
- [93] Douglas Lambert. “Oral History Indexing”. In: *The Oral History Review* 50.2 (2023), pp. 169–192. URL: <https://academic.oup.com/ohr/article/50/2/169/7190625>.
- [94] Teven Le Scao et al. “BLOOM: A 176B-Parameter Open-Access Multilingual Language Model”. In: *CoRR* abs/2211.05100 (2022). DOI: 10.48550/arXiv.2211.05100. arXiv: 2211.05100. URL: <https://doi.org/10.48550/arXiv.2211.05100>.
- [95] Georges Lee et al. “Artificial Intelligence, History and Knowledge Representation”. In: *Computers and the Humanities* 16.1 (1982), pp. 25–34. ISSN: 00104817. URL: <http://www.jstor.org/stable/30204722> (visited on 02/11/2026).
- [96] Lawrence Lessig. *Code and Other Laws of Cyberspace*. New York: Basic Books, 1999.
- [97] Zachary C. Lipton. “The Mythos of Model Interpretability”. In: *ACM Queue* 16.3 (2018), p. 30. DOI: 10.1145/3236386.3241340. URL: <https://doi.org/10.1145/3236386.3241340>.

- [98] Bing Liu. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. 2nd. Cambridge, UK: Cambridge University Press, 2020. ISBN: 9781108735841. URL: <https://www.cambridge.org/core/books/sentiment-analysis/67F83688DF3C0E03290A8DD9C2F5DA83>.
- [99] Tom Livingstone. “Colourisation and the Archive: Repurposing World War One”. In: *Frames Cinema Journal* 15 (2015). URL: <https://framescinemajournal.com/article/colourisation-and-the-archive-repurposing-world-war-one/>.
- [100] Jesus L. Lobo, Sergio Gil-Lopez, and Javier Del Ser. “The Right to Be Forgotten in Artificial Intelligence: Issues, Approaches, Limitations and Challenges”. In: *2023 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 2023, pp. 179–180. DOI: 10.1109/CAI54212.2023.00036. URL: <https://ieeexplore.ieee.org/document/10258152>.
- [101] Massimo Lollini. “Humanism in the Digital Age”. In: *Humanist Studies & the Digital Age* 1.1 (2011), pp. 1–7.
- [102] Shayne Longpre et al. “Position: Data Authenticity, Consent, and Provenance for AI Are All Broken: What Will It Take to Fix Them?”. In: *Proceedings of the 41st International Conference on Machine Learning (ICML 2024)*. Position paper. OpenReview, 2024. URL: <https://openreview.net/forum?id=3hSTecKy1b>.
- [103] Mykola Makhortykh et al. “Shall Androids Dream of Genocides? How Generative AI Can Change the Future of Memorialization of Mass Atrocities”. In: *Discover Artificial Intelligence* 3.1 (2023), pp. 1–17. ISSN: 2731-0809. DOI: 10.1007/s44163-023-00072-6.
- [104] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press, 1999. ISBN: 9780262133608, 0262133601.
- [105] Lev Manovich. *Cultural Analytics*. Cambridge, MA: The MIT Press, 2020. ISBN: 9780262037105.
- [106] Stefania Matei. “Generative Artificial Intelligence and Collective Remembering. The Technological Mediation of Mnemotechnic Values”. In: *Journal of Human-Technology Relations* 2 (2024). ISSN: 2773-2266. DOI: 10.59490/jhtr.2024.2.7405.
- [107] Viktor Mayer-Schönberger. *Delete: The Virtue of Forgetting in the Digital Age*. Princeton, NJ, USA: Princeton University Press, 2009. ISBN: 0691138613. URL: <https://press.princeton.edu/books/paperback/9780691138619/delete>.
- [108] R. Darrell Meadows and Joshua Sternfeld. “Artificial Intelligence and the Practice of History: A Forum”. In: *The American Historical Review* 128.3 (2023), pp. 1345–1349. DOI: 10.1093/ahr/rhad362.

- [109] Matthias Meitzler et al. “Digital Afterlife and the Future of Collective Memory”. In: *Memory Studies Review* 1.2 (2024), pp. 274–291. DOI: 10.1163/29498902-202400013. URL: [https://brill.com/view/journals/mesr/1/2/article-p274\\_005.xml](https://brill.com/view/journals/mesr/1/2/article-p274_005.xml).
- [110] María Antonia Diaz Mendoza, Emiro De La Hoz Franco, and Jorge Eliecer Gómez Gómez. “Technologies for the Preservation of Cultural Heritage—A Systematic Review of the Literature”. In: *Sustainability* 15.2 (2023), p. 1059. ISSN: 2071-1050. DOI: 10.3390/su15021059. URL: <https://www.mdpi.com/2071-1050/15/2/1059>.
- [111] Margaret Mitchell et al. “Model Cards for Model Reporting”. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT\* ’19)*. ACM, 2019, pp. 220–229. DOI: 10.1145/3287560.3287596. URL: <https://doi.org/10.1145/3287560.3287596>.
- [112] Brent Daniel Mittelstadt et al. “The Ethics of Algorithms: Mapping the Debate”. In: *Big Data & Society* 3.2 (2016), p. 2053951716679679.
- [113] Shakir Mohamed, Marie-Therese Png, and William Isaac. “Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence”. In: *Philosophy & Technology* 33.4 (2020), pp. 659–684. DOI: 10.1007/s13347-020-00405-1.
- [114] Saif M. Mohammad. “Sentiment Analysis: Automatically Detecting Valence, Emotions, and Other Affectual States from Text”. In: *Emotion Measurement*. Ed. by Herbert L. Meiselman. Comprehensive chapter on sentiment analysis techniques. Woodhead Publishing / Elsevier, 2017, pp. 323–379.
- [115] Christoph Molnar. *Interpretable machine learning*. Online book on interpretable machine learning techniques. 2020. URL: <https://christophm.github.io/interpretable-ml-book/>.
- [116] Maureen Ajiambo Muleka. “The Future of Historical Writing: Will Big Data Rewrite the Past? Exploring the Nuances”. In: *Journal of Research and Academic Writing* 1.1 (2024), pp. 22–30. ISSN: 3007-5343. DOI: 10.58721/jraw.v1i1.566.
- [117] Tempestt Neal et al. “Surveying Stylometry Techniques and Applications”. In: *ACM Comput. Surv.* 50.6 (Nov. 2018). ISSN: 0360-0300. DOI: 10.1145/3132039. URL: <https://doi.org/10.1145/3132039>.
- [118] Helen Nissenbaum. “Privacy as Contextual Integrity”. In: *Washington Law Review* 79.1 (2004), pp. 119–158.
- [119] Safiya Umoja Noble. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, 2018.
- [120] Alberto Nogales et al. “ARQGAN: An Evaluation of Generative Adversarial Network Approaches for Automatic Virtual Inpainting Restoration of Greek Temples”. In: *Expert Systems with Applications* 180 (2021), p. 115092. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2021.115092.

- [121] Pierre Nora. “Between Memory and History: Les Lieux de Mémoire”. In: *Representations* 26.Spring (1989), pp. 7–24. ISSN: 0734-6018. DOI: 10.2307/2928520. URL: <https://www.jstor.org/stable/2928520>.
- [122] Paul Ohm. “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization”. In: *UCLA Law Review* 57 (2010), pp. 1701–1777. URL: <https://scholarship.law.georgetown.edu/facpub/204/>.
- [123] Jeffrey K. Olick. *The Politics of Regret: On Collective Memory and Historical Responsibility*. New York, NY, USA: Routledge, 2007. ISBN: 9780415956826. URL: <https://www.routledge.com/The-Politics-of-Regret-On-Collective-Memory-and-Historical-Responsibility/Olick/p/book/9780415956826>.
- [124] Anaelia Ovalle et al. “Factoring the Matrix of Domination: A Critical Review and Reimagining of Intersectionality in AI Fairness”. In: *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. Montréal, QC, Canada: Association for Computing Machinery, Aug. 2023, pp. 496–511. DOI: 10.1145/3600211.3604705. (Visited on 07/14/2025).
- [125] Yasin Özkan and Mustafa Samur. “The Use of Artificial Intelligence in History Education”. In: *Virtual Technology Innovations in Education*. IGI Global Scientific Publishing, 2025, pp. 329–352. DOI: 10.4018/979-8-3693-6030-9.ch012. URL: <https://www.igi-global.com/chapter/the-use-of-artificial-intelligence-in-history-education/362313>.
- [126] Eli Pariser. *The Filter Bubble: What the Internet Is Hiding from You*. New York: Penguin Press, 2011.
- [127] Thomas D. Parsons. “Ethical Challenges of Using Virtual Environments in the Assessment and Treatment of Psychopathological Disorders”. In: *Journal of Clinical Medicine* 10.3 (2021), p. 378. ISSN: 2077-0383. DOI: 10.3390/jcm10030378.
- [128] Katalin Parti and Akos Szigeti. “The Future of Interdisciplinary Research in the Digital Era: Obstacles and Perspectives of Collaboration in Social and Data Sciences - An Empirical Study”. In: *Cogent Social Sciences* 7.1 (2021). Ed. by Sandro Serpa, p. 1970880. ISSN: 2331-1886. DOI: 10.1080/23311886.2021.1970880.
- [129] Frank Pasquale. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press, 2015.
- [130] Prantosh Paul and PS Aithal. “Digital Society: Its Foundation and Towards an Interdisciplinary Field”. In: *Proceedings of National Conference on Advances in Information Technology, Management, Social Sciences and Education*. 2018, pp. 1–6.
- [131] Judea Pearl and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books, 2018.

- [132] Aurélien Pellet, Julien Perez, and Marie Puren. “Generative Artificial Intelligence and Historical Research: Challenges, Potentials, and Limitations. Application of RAG to French Parliamentary Debates of the Third Republic (1881-1940)”. In: *A Conversation between AI and the Humanities*. Lyon, France, Nov. 2024. URL: <https://hal.science/hal-04832663>.
- [133] Roger D. Peng. “Reproducible Research in Computational Science”. In: *Science* 334.6060 (2011), pp. 1226–1227. DOI: 10.1126/science.1213847. URL: <https://www.science.org/doi/10.1126/science.1213847>.
- [134] Andrew Piper. *Enumerations: Data and Literary Study*. Chicago, IL, USA: University of Chicago Press, 2018. URL: <https://press.uchicago.edu/ucp/books/book/chicago/E/bo28200653.html>.
- [135] Christine Plumejeaud et al. “Data Curation to Reveal True Historical Uncertainty: From Navigocorpus to Portic, an Interdisciplinary Story of Dirty Data Cleaning”. In: *Histoire & mesure* 2 (2023), pp. 39–71. URL: <https://journals.openedition.org/histoiremesure/>.
- [136] Anil Poudel. “Face Recognition on Historical Photographs”. Master’s Thesis. Uppsala University, Department of Information Technology, 2021.
- [137] Marcelo O. R. Prates, Pedro H. C. Avelar, and Luís C. Lamb. “Assessing gender bias in machine translation: a case study with Google Translate”. In: *Neural Computing and Applications* 32.10 (2020), pp. 6363–6381. DOI: 10.1007/s00521-019-04144-6. URL: <https://doi.org/10.1007/s00521-019-04144-6>.
- [138] Gila Prebor. “From Digitization and Images to Text and Content: Transkribus as a Case Study”. In: *Proceedings of the Association for Information Science and Technology* 60.1 (2023), pp. 1102–1103.
- [139] Todd Presner. *Ethics of the Algorithm: Digital Humanities and Holocaust Memory*. Princeton University Press, 2024. ISBN: 9780691258980.
- [140] Lara Putnam. “The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast”. In: *The American Historical Review* 121.2 (2016), pp. 377–402. DOI: 10.1093/ahr/121.2.377. URL: <https://academic.oup.com/ahr/article/121/2/377/2586815>.
- [141] Jiwat Ram. “Moral Decision-Making in AI: A Comprehensive Review and Recommendations”. In: *Technological Forecasting and Social Change* 217 (Aug. 2025), p. 124150. ISSN: 0040-1625. DOI: 10.1016/j.techfore.2025.124150. (Visited on 06/10/2025).
- [142] Stephen Ramsay. *Reading Machines: Toward an Algorithmic Criticism*. University of Illinois Press, 2011. ISBN: 9780252036415.
- [143] Ridho Reinanda, Edgar Meij, and Maarten de Rijke. “Knowledge Graphs: An Information Retrieval Perspective”. In: *Foundations and Trends in Information Retrieval* 14.4 (2020), pp. 289–444. DOI: 10.1561/15000000063.

- [144] Rekrei Team. *Crowdsourcing the Reconstruction of Lost Heritage*. Article on the Rekrei crowdsourced heritage reconstruction project. Accessed: 2023-10-01. 2021. URL: <https://www.gim-international.com/content/article/crowdsourcing-the-reconstruction-of-lost-heritage?output=pdf>.
- [145] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. ““Why should I trust you?”: Explaining the predictions of any classifier”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 2016, pp. 1135–1144. DOI: 10.1145/2939672.2939778. URL: <https://doi.org/10.1145/2939672.2939778>.
- [146] Stephen E. Robertson and Hugo Zaragoza. “The Probabilistic Relevance Framework: BM25 and Beyond”. In: *Foundations and Trends in Information Retrieval* 3.4 (2009), pp. 333–389. DOI: 10.1561/15000000019.
- [147] Geoffrey Rockwell and Stéfan Sinclair. *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. Cambridge, MA: MIT Press, 2016.
- [148] C. Annemieke Romein et al. “State of the Field: Digital History”. In: *History* 105.365 (2020), pp. 291–312. DOI: 10.1111/1468-229X.12969. URL: <https://onlinelibrary.wiley.com/journal/1468229x>.
- [149] Roy Rosenzweig. *Clio Wired: The Future of the Past in the Digital Age*. New York: Columbia University Press, 2011.
- [150] Kirsten E. Roy. “Defrosting the chill: How facial recognition technology threatens free speech”. In: *Roger Williams University Law Review* 27.1 (2022), 185–? URL: [https://docs.rwu.edu/rwu\\_LR/vol27/iss1/9](https://docs.rwu.edu/rwu_LR/vol27/iss1/9).
- [151] Cynthia Rudin. “Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead”. In: *Nature Machine Intelligence* 1.5 (2019), pp. 206–215. ISSN: 2522-5839. DOI: 10.1038/s42256-019-0048-x.
- [152] Joanne W. Rudolf. *The Fortunoff Video Archive for Holocaust Testimonies*. 1996. URL: <https://www.tandfonline.com/doi/abs/10.1080/01439689600260101>.
- [153] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 4th. Upper Saddle River, NJ, USA: Pearson, 2020. ISBN: 9780134610993. URL: <https://www.pearson.com/en-us/subject-catalog/p/artificial-intelligence-a-modern-approach/P200000003500/>.
- [154] Carl Ryant. “The Public Historian and Business History: A Question of Ethics”. In: *The Public Historian* 8.1 (1986), pp. 31–38. ISSN: 0272-3433. DOI: 10.2307/3377072. (Visited on 07/14/2025).
- [155] Moacir P. de Sá Pereira. “Mixed Methodological Digital Humanities”. In: *Debates in the Digital Humanities 2019*. Ed. by Matthew K. Gold and Lauren F. Klein. Debates in the Digital Humanities. Minneapolis, MN: University of Minnesota Press, 2019, pp. 405–413. ISBN: 978-1-5179-0693-1.

- [156] Beatrice Savoldi et al. “Gender Bias in Machine Translation”. In: *Transactions of the Association for Computational Linguistics* 9 (2021), pp. 845–874. DOI: 10.1162/tacl\_a\_00401. URL: [https://doi.org/10.1162/tacl\\_a\\_00401](https://doi.org/10.1162/tacl_a_00401).
- [157] Raphael Schlembach and Nicola Clewer. “‘Forced Empathy’: Manipulation, Trauma and Affect in Virtual Reality Film”. In: *International Journal of Cultural Studies* 24.5 (2021), pp. 827–843. ISSN: 1367-8779. DOI: 10.1177/13678779211007863.
- [158] Benjamin Schmidt. “Representation Learning”. In: *The American Historical Review* 128.3 (2023), pp. 1350–1353. ISSN: 0002-8762. DOI: 10.1093/ahr/rhad363. URL: <https://academic.oup.com/ahr/article/128/3/1350/7093827>.
- [159] Zachary M. Schrag. *The Princeton Guide to Historical Research*. Skills for Scholars. Princeton, NJ: Princeton University Press, 2021. ISBN: 9780691210964.
- [160] Julien Schuh. “AI As Artificial Memory: A Global Reconfiguration of Our Collective Memory Practices?” In: *Memory Studies Review* 1.2 (2024), pp. 231–255. DOI: 10.1163/29498902-202400012. URL: [https://brill.com/view/journals/mesr/1/2/article-p231\\_003.xml](https://brill.com/view/journals/mesr/1/2/article-p231_003.xml).
- [161] Andrew D. Selbst et al. “Fairness and Abstraction in Sociotechnical Systems”. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT\* ’19)*. Association for Computing Machinery, 2019, pp. 59–68. DOI: 10.1145/3287560.3287598.
- [162] Donghee Shin. *Debiasing AI: Rethinking the Intersection of Innovation and Sustainability*. New York: Routledge, Apr. 2025. ISBN: 978-1-00-353024-4. DOI: 10.1201/9781003530244.
- [163] Emily Silcock et al. *Newsire: A Large-Scale Structured Database of a Century of Historical News*. 2024. arXiv: 2406.09490.
- [164] Jatinder Singh et al. *Responsibility & Machine Learning: Part of a Process*. 2016. DOI: 10.2139/ssrn.2860048. SSRN: 2860048.
- [165] Amit Singhal. “Modern Information Retrieval: A Brief Overview”. In: *IEEE Data Eng. Bull.* 24.4 (2001), pp. 35–43. URL: <http://sites.computer.org/debull/A01DEC-CD.pdf>.
- [166] Mona Sloane et al. “Participation Is Not a Design Fix for Machine Learning”. In: *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO 2022)*. ACM, 2022, 1:1–1:6. DOI: 10.1145/3551624.3555285. URL: <https://doi.org/10.1145/3551624.3555285>.
- [167] Ann Laura Stoler. “Colonial Archives and the Arts of Governance”. In: *Archival Science* 2.1–2 (2002), pp. 87–109. DOI: 10.1007/BF02435632. URL: <https://doi.org/10.1007/BF02435632>.

- [168] Chris Alen Sula and Heather V. Hill. “The Early History of Digital Humanities: An Analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004)”. In: *Digital Scholarship in the Humanities* 34.Supplement\_1 (2019), pp. i190–i206. DOI: 10.1093/llc/fqz072.
- [169] Cass R. Sunstein. *Republic: Divided Democracy in the Age of Social Media*. Princeton, NJ: Princeton University Press, 2017.
- [170] John Tagg. *The Burden of Representation: Essays on Photographies and Histories*. Amherst, MA: University of Massachusetts Press, 1988.
- [171] Living With Machines Team. *Living With Machines: Computational Histories of the Age of Industry*. Ed. by Ruth Ahnert, Emma Griffin, and Jon Lawrence. Open access collaborative history book; project edition. Accessed: 2025-02-07. 2024. URL: <https://read.uolpress.co.uk/projects/living-with-machines>.
- [172] Melissa M. Terras. *Digital Images for the Information Professional*. Ashgate, 2008. ISBN: 9780754648604.
- [173] Sue Timmis et al. “Rethinking Assessment in a Digital Age: Opportunities, Challenges and Risks”. In: *British Educational Research Journal* 42.3 (2016), pp. 454–476. ISSN: 1469-3518. DOI: 10.1002/berj.3215.
- [174] John Tosh. *The Pursuit of History: Aims, Methods and New Directions in the Study of Modern History*. 6th. London: Routledge, 2015.
- [175] Michel-Rolph Trouillot. *Silencing the Past: Power and the Production of History*. Boston, MA: Beacon Press, 1995. ISBN: 9780807059197. URL: <https://www.beacon.org/Silencing-the-Past-P919.aspx>.
- [176] Emily Tseng, Thomas Ristenpart, and Nicola Dell. “Mitigating Trauma in Qualitative Research Infrastructure: Roles for Machine Assistance and Trauma-Informed Design”. In: *Proceedings of the ACM on Human-Computer Interaction* 9.2 (2025), pp. 1–37.
- [177] Cristian Vaccari and Andrew Chadwick. “Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News”. In: *Social Media+ Society* 6.1 (2020), p. 2056305120903408.
- [178] Siva Vaidhyanathan. *The Googlization of Everything: (And Why We Should Worry)*. 1st ed. University of California Press, 2011. ISBN: 9780520258822. URL: <http://www.jstor.org/stable/10.1525/j.ctt1pn9z8> (visited on 06/30/2025).
- [179] Teemu Valtonen et al. “Media Literacy Education in the Age of Machine Learning”. In: *Journal of Media Literacy Education* 11.2 (2019), pp. 20–36. ISSN: 2167-8715. DOI: 10.23860/JMLE-2019-11-2-2.
- [180] Tommaso Venturini, Mathieu Jacomy, and Pablo Jensen. “What do we see when we look at networks: Visual network analysis, relational structuralism, and unsupervised learning”. In: *Big Data & Society* 8.1 (2021). DOI: 10.1177/20539517211018488. URL: <https://doi.org/10.1177/20539517211018488>.

- [181] Paul Voigt and Axel Von dem Bussche. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer, 2017.
- [182] Laura Weidinger et al. “Ethical and social risks of harm from Language Models”. In: *arXiv preprint arXiv:2112.04359* (2021). URL: <https://arxiv.org/abs/2112.04359>.
- [183] M. Westerlund. “The Emergence of Deepfake Technology: A Review”. In: *Technology Innovation Management Review* 9.11 (2019). Broader review relevant to deepfake history., pp. 39–52. DOI: 10.22215/timreview/1282. URL: <https://timreview.ca/article/1282>.
- [184] Hayden White. *Metahistory: The Historical Imagination in Nineteenth-Century Europe*. Baltimore, MD, USA: Johns Hopkins University Press, 1973. ISBN: 0801814693. URL: <https://jhupbooks.press.jhu.edu/title/metahistory>.
- [185] Mark D. Wilkinson et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific Data* 3 (2016), p. 160018. DOI: 10.1038/sdata.2016.18.
- [186] Jay Winter. *Remembering War: The Great War between Memory and History in the Twentieth Century*. New Haven, CT: Yale University Press, 2006. ISBN: 9780300101911. URL: <https://yalebooks.yale.edu/book/9780300101911/remembering-war/>.
- [187] Bob G. Witmer and Michael J. Singer. “Measuring Presence in Virtual Environments: A Presence Questionnaire”. In: *Presence: Teleoperators and Virtual Environments* 7.3 (1998), pp. 225–240. DOI: 10.1162/105474698565686.
- [188] Karen Yeung. ““Hypernudge”: Big Data as a mode of regulation by design”. In: *Information, Communication & Society* 20.1 (2017), pp. 118–136. DOI: 10.1080/1369118X.2016.1186713.
- [189] Yifan Zeng. *HistoLens: An LLM-Powered Framework for Multi-Layered Analysis of Historical Texts – A Case Application of Yantie Lun*. 2024. arXiv: 2411.09978.
- [190] Ying Zhang et al. *Paths of a Million People: Extracting Life Trajectories from Wikipedia*. 2024. arXiv: 2406.00032.
- [191] Michael Zimmer. ““But the data is already public”: on the ethics of research in Facebook”. In: *Ethics and Information Technology* 12.4 (2010), pp. 313–325. DOI: 10.1007/s10676-010-9227-5. URL: <https://doi.org/10.1007/s10676-010-9227-5>.
- [192] Shoshana Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, 2019.