The Long Future of the Past

Building the Public Infrastructure for Historical Knowledge in the Age of AI

The Committee of Historians for AI and Research

in concert with

AHA Ad-Hoc Committee on AI in History Research and Publications (AAAIR) and

the American Historical Association

The United States is entering a crisis of historical infrastructure. Three current battles show what is at stake. First, the Presidential Records Act—the post-Watergate law that makes presidential records public property—is now under direct legal attack; if it fails, the documentary basis for democratic accountability at the highest level of government is gravely weakened. The American Historical Association, American Oversight, and other organizations are already in court to defend it. Second, the Internet Archive—one of the few institutions preserving the born-digital record at civilization scale, including more than one trillion archived web pages—has faced existential legal and financial pressure over digital preservation.[1] Third, the Supreme Court is increasingly deciding constitutional questions through claims about history, even as historians warn that courts often misuse, flatten, or selectively interpret the historical record.[2]

Americans have long understood that preserving records is essential to liberty. Even in 1778, in the midst of the Revolutionary War, the Continental Congress appropriated scarce funds to collect and preserve state papers. As the first official historian of the United States argued, preserving records would demonstrate that Congress were both "friends of science" and "guardians of our liberties."[3] In the nineteenth century, Congress distributed hundreds of thousands of pages of public documents—including the once-secret papers of the Revolution and

---

[1] Internet Archive, "End of *Hachette v. Internet Archive*," December 4, 2024, https://blog.archive.org/2024/12/04/end-of-hachette-v-internet-archive/.

[2] See, for example, Brennan Center for Justice analysis of historians' amicus briefs before the Supreme Court; Reuters, "Historians, Watchdog Group Sue Trump Administration to Preserve White House Records," April 7, 2026, https://www.reuters.com/legal/government/historians-watchdog-group-sue-trump-preserve-white-house-records-2026-04-07/.

[3] Matthew Connelly, *The Declassification Engine*, quoting the Continental Congress debates over preserving state papers during the Revolutionary era.

the early republic—to citizens across the country as "monuments of the past, beacons to the future."[4] The United States has repeatedly treated public access to historical records not as a luxury, but as infrastructure for republican self-government.

Today, however, the institutions that preserve, interpret, and verify the past were built for a paper world, while power increasingly moves through digital records, proprietary platforms, synthetic media, and AI systems. Public life now depends on historical claims—about presidential authority, constitutional meaning, public records, national security, civil rights, climate risk, property, and institutional responsibility—yet the infrastructure for making those claims verifiable, contestable, and accountable is fragile, underfunded, and increasingly outpaced by the technologies that mediate public knowledge.

AI intensifies this crisis. Large language models already summarize court precedent, interpret regulations, answer questions about wars and elections, synthesize public records, and generate historical narratives for millions of users. While libraries, archives, governments and historians have traditionally taken responsibility for maintaining this information infrastructure, no current institution is responsibility for ensuring that a diffuse decentralized network of players keeps up with AI. At stake is the ability of civil society to reckon with accuracy, provenance, plurality, and accountability. If public institutions fail to act, proprietary systems will become the de facto memory layer of society while the records needed to challenge them disappear, become inaccessible, or are misrepresented at scale.

This proposal calls for a National Center for Historical Research, embedded in an international network of partner centers, to build the public infrastructure for trustworthy historical knowledge in the age of AI. It builds upon decades of work by archivists, librarians, digital humanists, records managers, and public memory institutions. Its purpose is not simply to support historians. It is to defend a democratic right to a verifiable past.

A Changing Context

A profound transformation is underway in how societies encounter knowledge about the past. Every day, AI systems summarize court precedents, explain wars and elections, interpret regulations, answer questions about public policy, and generate historical narratives for millions of users. Increasingly, these systems mediate the public's relationship to history.

A generation ago, historical interpretation largely remained within books, classrooms, archives, and scholarly debate. Today, millions of users encounter the past first through search engines, recommendation systems, AI assistants, and synthetic summaries. In practice, this means that AI systems are already becoming public arbiters of historical meaning.

---

[4] Connelly, *The Declassification Engine,* quoting the Continental Congress debates over preserving state papers during the Revolutionary era.

Yet no public institution currently exists with responsibility for coordinating the archives, libraries, scholarly communities, publishers, and technology systems on which trustworthy historical knowledge now depends.

The twentieth century built institutions for science, medicine, aviation, nuclear research, and computing. The twenty-first century now requires institutions for trustworthy historical knowledge. The challenge is not simply technological. It is organizational. The records, expertise, standards, and custodial institutions required for trustworthy historical knowledge already exist, but they remain fragmented across thousands of archives, libraries, universities, publishers, and public agencies with no coordinating framework equal to the scale of AI.

Americans have long acted on an implicit belief in ***a right to a verifiable past***. A right to a verifiable past does not mean access to every record. It means the ability to determine where evidence came from, what is missing, who exercised custodial authority, what restrictions apply, and how a historical claim was assembled.

Claims about public events, laws, rights, property, elections, wars, or government actions should ultimately be traceable to records, evidence, and publicly contestable interpretation. That assumption is now under strain. Decisions in law, journalism, education, national security, insurance, and public policy increasingly rely on AI-mediated representations of past events, yet the systems producing those representations often provide little visibility into evidence, provenance, uncertainty, or disagreement.

The challenge is not that archives are incomplete. No complete historical record has ever existed. Every archive is shaped by survival, destruction, secrecy, privacy, power, and chance. The goal of democratic knowledge is therefore not completeness but accountability: preserving the ability to understand where evidence comes from, what remains uncertain, what is missing, and why. In the age of AI, trustworthy knowledge depends on making gaps, silences, and competing interpretations visible rather than concealing them behind authoritative-sounding summaries.

At present, however, the public infrastructure needed to support that goal does not exist. There is no coordinated framework linking physical archives, born-digital records, libraries, universities, and AI systems.

To be clear, this proposal does not envision a National Center for Historical Research replacing existing archives, libraries, museums, publishers, documentary editing projects, or history departments. The opposite is true. The proposed Center exists because these institutions have already spent generations creating the collections, descriptive systems, scholarly editions, preservation practices, and public trust on which historical knowledge depends.

The challenge is that these institutions were not designed to confront the new problems created by AI. Archives preserve records; libraries organize and provide access to knowledge; museums

steward material culture; publishers certify and disseminate scholarship; universities train researchers and produce new interpretations. Each performs an essential function. Yet no institution is currently responsible for ensuring that historical knowledge remains transparent, inspectable, and accountable as AI systems become a primary interface between the public and the historical record.

The goal, therefore, is not substitution but coordination. A National Center would help connect and strengthen existing institutions by developing shared standards, technical infrastructure, training programs, evaluation frameworks, and public-interest partnerships that no single archive, library, publisher, or university can realistically build alone. Just as national research networks support laboratories without replacing them, and public computing infrastructure supports scientific research without replacing universities, a historical research center would amplify the capacity of existing institutions rather than compete with them.

Indeed, many of the foundational components already exist. Archivists and records managers have spent decades developing standards for description, provenance, and preservation. Libraries and digital humanities centers have built interoperable metadata systems, linked open data frameworks, APIs, IIIF protocols, web archives, and large-scale digitization initiatives. Documentary editors have created authoritative scholarly editions that model rigorous provenance and contextualization. The challenge is not the absence of innovation. It is the absence of a coordinating institution capable of bringing these efforts together and adapting them to the scale and demands of AI-mediated knowledge.

At present, there are no widely adopted standards for provenance-aware historical claims generated by AI systems, no accepted benchmarks for evaluating historical reasoning, uncertainty, disagreement, or evidentiary integrity in large language models, and no institution charged with ensuring that public knowledge about the past remains verifiable as AI increasingly mediates access to information. The Center's purpose is to fill that gap: not by centralizing authority over the past, but by helping existing institutions work together to preserve the conditions under which historical claims remain contestable, evidence-based, and publicly accountable.

The common response to this challenge is to call for more digitization. Digitization is necessary, but it is not the primary bottleneck. Historical infrastructure is fundamentally a problem of description and coordination. Records become useful only when they are identified, arranged, described, contextualized, and connected to the institutions, people, events, and activities that produced them. Likewise, no single archive, library, university, publisher, or technology firm can solve the problem alone. The challenge is to coordinate thousands of custodial institutions, scholarly communities, and technical systems into a framework that preserves provenance, enables discovery, and sustains public accountability. AI has not created this problem, but it has exposed it at unprecedented scale.

The consequences extend beyond scholarship. If public institutions fail to build alternatives, proprietary AI systems will increasingly become the default memory layer of society. Historical claims will be mediated through systems optimized for fluency, speed, and scale rather than provenance, uncertainty, and evidentiary accountability. The danger is not merely misinformation. It is the gradual replacement of contestable public records with synthetic narratives whose evidentiary foundations may be inaccessible, opaque, or lost altogether.

The challenge is also geopolitical. Historical interpretation increasingly shapes disputes over sovereignty, constitutional meaning, climate responsibility, migration, colonial legacies, and international order. States and political actors already mobilize archives, historical narratives, and AI-generated content as instruments of persuasion and legitimacy. In this environment, trustworthy historical knowledge becomes a matter not only of scholarship but of democratic resilience and national capacity.

At the same time, the opportunity is extraordinary. Historical reasoning is becoming an applied public resource. Courts rely on historical interpretation. Insurers use long-run historical data to model risk. Intelligence analysts reconstruct institutional memory and causal sequences. Governments, firms, journalists, and citizens increasingly need reliable ways to reason across decades of documents, conflicting narratives, and incomplete evidence. The challenge is no longer information scarcity but trustworthy interpretation.

This proposal argues that meeting that challenge requires a new layer of public knowledge infrastructure: a National Center for Historical Research operating within an international network of archives, libraries, universities, and public-interest technology institutions. Its mission would be to build the standards, infrastructure, expertise, and public systems necessary to preserve a right to a verifiable past in the age of AI.

The goal is not simply better AI. It is the preservation and expansion of society's capacity to reason collectively about the past.

Why This Matters

Truth: Bad History Creates Bad Outputs

The stakes are not abstract. Historical reasoning now shapes some of the most consequential decisions in public life. Courts increasingly rely on historical claims to determine constitutional meaning. Intelligence agencies reconstruct geopolitical narratives from fragmented records. Insurers price catastrophe risk through long-run historical data. Regulators, journalists,

educators, and AI systems all depend upon interpretations of the past in order to act in the present.

Yet the institutions capable of evaluating those interpretations have not kept pace with the systems now generating them.

This problem is especially visible in the judiciary. Recent Supreme Court decisions have relied heavily on selective historical interpretation, often invoking narrow readings of constitutional history while historians themselves remain deeply divided over the underlying evidence and context. Professional historians have repeatedly filed amicus briefs warning that courts are flattening complex historical debates into singular narratives that overstate certainty and understate disagreement. The issue is not that judges should avoid history. It is that democratic institutions increasingly require robust mechanisms for evaluating how historical claims are assembled, sourced, contextualized, and contested. Without such mechanisms, selective or tendentious uses of history can acquire the force of law.[5]

AI systems magnify this risk dramatically. Large language models are trained to produce coherent narratives, even when evidence is partial, contradictory, or missing. Historical archives, however, are inherently incomplete. They are shaped by power, survival, censorship, institutional bias, and accident. Most archival collections disproportionately preserve the voices of states, bureaucracies, literate elites, corporations, police, militaries, and institutions capable of recordkeeping. Marginalized communities often survive only through fragmentary traces: scattered correspondence, oral testimony, legal disputes, administrative encounters, or hostile documentation produced by others.

Archival research therefore cannot simply be "scaled" through computation alone. Historical knowledge is, by its nature, a problem of minority data. Historians have spent generations developing methods for reasoning under precisely these conditions: distinguishing silence from absence, fragment from pattern, survivorship from representativeness, and plausibility from evidence. They know that what is missing from the archive may matter as much as what survives within it.[6] Only historians possess the accumulated methodological traditions necessary to adapt AI systems to this challenge rather than allowing those systems to erase it.

This matters because AI systems tend to interpret absence as insignificance and repetition as truth. They flatten evidentiary distinctions between police reports, oral histories, propaganda, newspapers, diaries, leaked documents, and retrospective summaries unless those distinctions are

---

[5] See, for example, Brennan Center for Justice, "Historians' Friend-of-the-Court Briefs This Supreme Court Term".

[6] Michel-Rolph Trouillot, *Silencing the Past: Power and the Production of History* (Boston: Beacon Press, 1995), esp. 26–30; Saidiya Hartman, "Venus in Two Acts," *Small Axe* 12, no. 2 (June 2008): 1–14; Ann Laura Stoler, *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense* (Princeton: Princeton University Press, 2009); Marisa J. Fuentes, *Dispossessed Lives: Enslaved Women, Violence, and the Archive* (Philadelphia: University of Pennsylvania Press, 2016).

deliberately modeled. They reward coherence over uncertainty and consensus over contestation. In doing so, they risk converting the structural biases of the archive into authoritative-seeming narratives about the past.

Historians, by contrast, are trained to treat records not as transparent containers of fact, but as artifacts produced by institutions, incentives, conflicts, and historical conditions. Their expertise lies not simply in recovering information, but in reconstructing context: who produced a record, for whom, under what constraints, and why it survived.

This distinction becomes increasingly important as AI systems become a primary interface to historical knowledge.

While many organizations address pieces of this challenge, no institution is currently responsible for ensuring that AI-mediated representations of the past remain transparent, contestable, and accountable across archives, libraries, publishers, universities, and technology platforms.[7]

The lack of a coordinating institution is not merely a scholarly problem. It is a risk to democratic accountability itself. Democracies depend upon contestable records: citizens must be able to reconstruct what happened, challenge official narratives, compare evidence, and hold institutions accountable across time. When historical claims become opaque, unverifiable, or detached from accessible records, public trust weakens and conspiratorial or authoritarian narratives flourish.

The challenge, however, is also an extraordinary opportunity. Properly designed, AI systems could dramatically expand access to historical knowledge while preserving evidentiary transparency. Citizens could trace claims directly to underlying documents. Researchers could compare interpretations across languages and archives. Journalists, courts, and policymakers could inspect chains of evidence rather than relying on opaque summaries. Entire classes of neglected archives could become newly discoverable and usable for the first time.

A right to a verifiable past does not mean access to every record. It means the ability to determine where evidence came from, what is missing, who exercised custodial authority, what restrictions apply, and how a historical claim was assembled.

---

[7] Important initiatives already address portions of this challenge. These include NIST's work on AI risk management and evaluation; academic centers devoted to AI ethics, governance, and accountability; the American Historical Association's committees on AI and historical research; journalism and fact-checking organizations evaluating synthetic content and misinformation; technical efforts to measure hallucination, factuality, and model reliability; provenance and content-authenticity initiatives such as the Coalition for Content Provenance and Authenticity (C2PA); archival and library AI policies developed by individual institutions; and emerging protocols for responsible AI use in archival collections, including those developed at the University of Virginia and elsewhere. The challenge is not the absence of activity, but its fragmentation. No institution currently coordinates these efforts around the specific problem of historical knowledge at scale, linking archives, libraries, publishers, universities, and AI developers through shared standards for provenance, uncertainty, evidentiary accountability, and historical reasoning.

Governance*: Democracy requires contestable records*

Archives are not passive storehouses of old records; they are part of the machinery of democratic accountability. States generate records in order to govern, but citizens need access to those records in order to evaluate how power was exercised. Excessive secrecy, overclassification, destruction of records, and indefinite delay in release do more than inconvenience researchers—they weaken public oversight and distort collective memory.

For AI systems increasingly used in governance, law, and public information, this matters directly. If models are trained on public records shaped by decades of secrecy, selective release, or institutional omission, they inherit the epistemic footprint of the security state. AI-augmented historical methods offer a democratic corrective: discerning patterns and anomalies in state secrecy and archival destruction transparency is not merely an administrative value; it is infrastructure for self-government.

> ***Case Study: The fragility of historical infrastructure is no longer hypothetical.*** *In April 2026, the American Historical Association and American Oversight sued the federal government after the Department of Justice argued that the Presidential Records Act—the post-Watergate law governing preservation of presidential records—was unconstitutional. The dispute was not primarily about access to any single collection. It concerned whether records documenting presidential decision-making are public assets preserved for future accountability or materials ultimately subject to executive discretion. Whatever the outcome of the litigation, the episode illustrates a broader reality: the institutions, laws, and norms that preserve democratic memory are themselves contested. Historical infrastructure cannot be assumed to exist automatically. It requires ongoing institutional stewardship, legal protection, and public investment.*[8]
>
> *AI raises the stakes further. Systems increasingly trained on public records depend upon the continued existence, accessibility, and integrity of those records. If the preservation of public documentation becomes uncertain, the foundations on which future systems reason about government actions, public policy, and democratic accountability become uncertain as well.*

[8] Mike Scarcella, "Historians, Watchdog Group Sue Trump to Preserve White House Records," *Reuters*, April 7, 2026, https://www.reuters.com/legal/government/historians-watchdog-group-sue-trump-preserve-white-house-records-2026-04-07/ .

Capability: *Historical reasoning is now applied infrastructure*

Industry increasingly operates in environments saturated with text: emails, reports, contracts, tickets, customer feedback, earnings calls, policy updates, internal chats, market commentary, regulatory filings, and technical documentation. The challenge is no longer access to information but interpretation across time and across disagreement. Historical method is valuable here because firms need more than summarization—they need reliable analysis of **date-event relationships**, **dissent versus consensus**, and the changing meaning of documents within institutional context. Historians are trained to reconstruct sequences: what happened first, what changed after a policy decision, when sentiment shifted, which memo preceded a market move, whether later narratives rewrite earlier events, and how competing timelines can be reconciled. They also distinguish true consensus from temporary conformity, strategic silence, or unresolved internal conflict. In organizations, apparent agreement in formal documents may mask dissent in side channels, regional variation, or incentives not to contradict leadership. Likewise, bursts of disagreement may reflect healthy deliberation rather than dysfunction.

Additionally, industry increasingly depends on historical data for consequential decisions. For example, the world's largest reinsurers maintain dedicated geo-risk research operations that integrate decades of meteorological and loss data into their pricing of catastrophic risk.[9] This is a recognition that long-run records, however imperfect, are indispensable infrastructure for evaluating present-day exposure.

[9] Munich Re, "Climate Change and Its Consequences," Munich Re, accessed 6 May 2026, https://www.munichre.com/en/risks/climate-change.html.

Historical analysis of agreeing and conflicting sources helps distinguish durable institutional change from temporary consensus, emerging fractures from settled practice, and retrospective narratives from contemporaneous understanding. Ian Milligan's work on digital archives demonstrates that scale increases rather than eliminates the need for interpretation, while Louis Hyman's histories of firms and labor show that organizations evolve through changing incentives, power relations, and institutional adaptation. For AI-enabled enterprises, historians therefore contribute methods for interpreting change over time rather than simply summarizing documents.

Consider an acquisition in which an AI system is asked to synthesize five years of board minutes, internal memoranda, and earnings calls. The resulting narrative may accurately describe what executives ultimately agreed upon while obscuring the sequence by which that consensus emerged: abandoned strategies, recurring disagreements, shifts in leadership, or changing assumptions about markets and regulation. For a historian, those temporal patterns are often more consequential than the final consensus itself. They reveal how an institution responds to uncertainty, how decisions become durable, and where future fault lines are likely to emerge. AI can assemble the archive; historical reasoning is required to interpret how institutions actually change.

Reading institutions historically means reconstructing sequences of decision, disagreement, and adaptation rather than simply summarizing outcomes. That challenge has already driven a generation of methodological innovation within the historical discipline. The Archives Unleashed project, working with billions of web pages preserved by the Internet Archive, has shown that historians can reason at scale across heterogeneous born-digital corpora.[10] Ongoing research on the digital aftermath of September 11th, drawing on hundreds of thousands of forum posts, e-mail messages, and pager transcripts, demonstrates that techniques developed for the archive translate to the textual environments enterprises now confront daily.[11]

***Custody**: Evidence requires an unbroken chain*

The preceding arguments concern what AI systems do with the historical record. But the integrity of any output depends first on the integrity of the record's custody—and this is where the expertise of archivists and librarians becomes infrastructural rather than incidental. Every document an AI system reasons over

[10] Ruest, Nick, Jimmy Lin, Ian Milligan, and Samantha Fritz. "The Archives Unleashed Project: Technology, Process, and Community to Improve Scholarly Access to Web Archives." In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020 (JCDL '20), 157–66. New York: Association for Computing Machinery, 2020.

[11] Ian Milligan, On Every Screen: September 11th and the Early Internet (book manuscript in progress, supported by SSHRC Insight Grant "Digital Terror: The September 11th, 2001 Attacks Online," 2025–2029).

has already passed through accessioning decisions, descriptive choices, arrangement, OCR or transcription pipelines, metadata schemas, and platform migrations. Each of these is an interpretive act, and most are invisible by the time material reaches a model. A system that ingests a document as interchangeable text severs it from the finding aid, the collection-level context, the custodial history, and the provenance trail that together establish what the document is, where it came from, and how far it can be trusted.

Case Study: Founders Online and the Disappearance of Editorial Context

Founders Online offers a vivid example of how historical knowledge can become detached from the expert infrastructure that produced it. Created through a partnership among the National Archives, the University of Virginia Press, and the editors of six major documentary editions, Founders Online presents thousands of documents associated with the American founding generation. To many users—and increasingly to AI systems—the site appears to be a searchable repository of primary sources.

In reality, every document has passed through multiple layers of scholarly mediation. Documentary editors identified manuscripts, established authoritative texts, resolved ambiguous readings, selected variant versions, added contextual annotation, linked related documents, and constructed metadata that allows the material to be searched and interpreted. What appears to be a collection of primary sources is in fact the product of decades of editorial scholarship.

Over the past year, Founders Online has reportedly experienced repeated disruptions from automated AI harvesting and requests for bulk downloads intended for chatbot and model-development projects. These requests often treat the site as a corpus of raw historical text rather than as an editorially curated scholarly resource. Historians and documentary editors have similarly reported students and researchers increasingly relying on AI agents to query the collection rather than engaging directly with its editorial apparatus.

The significance extends beyond one project. Increasingly, AI systems consume the outputs of archives, libraries, and scholarly editions while obscuring the human expertise that made those resources trustworthy in the first place. Provenance is not simply a technical attribute of documents; it is also a record of the interpretive labor required to transform fragments of the past into usable historical evidence. Public infrastructure for historical AI must preserve that chain of custody and make visible the scholarly work embedded within digital collections rather than treating them as interchangeable text.

The consequences are not merely professional. They are epistemic. Once documents are detached from their editorial and archival context, users—and increasingly AI systems—lose access to the information necessary to evaluate what they are reading. A transcribed letter may contain uncertain readings, missing passages, disputed attributions, or editorial decisions that shaped how the text appears. An annotation may summarize decades of scholarship about chronology, authorship, or historical context that is invisible when only the text itself is extracted. If these

distinctions disappear, AI systems risk presenting editorial reconstructions as original documents, contested interpretations as settled fact, and fragmentary evidence as complete records.

More broadly, the disappearance of provenance transforms historical knowledge from a chain of evidence into a collection of disconnected text fragments. The result is not merely a loss of attribution but a loss of accountability. Users can no longer determine why a document should be trusted, what scholarly judgments shaped its presentation, what uncertainties remain unresolved, or where to turn when interpretations are challenged. In such a system, historical authority gradually shifts from archives, editors, and publicly accountable institutions toward opaque retrieval systems and model outputs whose evidentiary foundations are increasingly difficult to inspect. (With thanks to Patricia Searl, University of Virginia Press)

Decontextualization of records is the most consequential and least visible failure mode of historical AI. Archives are not passive repositories of raw material; they are expert infrastructure, and the chain of custody they maintain is precisely what allows a claim to be traced back to authenticated evidence. Where that chain is broken, authenticity cannot be established and authenticity is exactly what generative systems are structurally least able to supply, since a fabricated quotation and a genuine one are indistinguishable without a trail back to the source. Any public infrastructure for trustworthy historical knowledge must therefore treat archivists and librarians as principals in its design, and must require that systems expose the evidentiary chain rather than conceal it.

Preserving that chain in practice requires a distinction that is already reshaping responsible archival work: the difference between retrieval-based and training-based uses of collections. In retrieval-based use, a system searches, transcribes, or analyzes materials that remain under institutional control; the institution retains the ability to correct, withdraw, or shut the system down. In training-based use, materials are absorbed into model weights, where removal is not realistically possible—once ingested, a collection cannot be taken back. This line determines whether provenance, attribution, and accountability survive contact with the system at all. Recent institutional standards, notably the UVA Archival AI Protocol, operationalize the distinction at the collection level: retrieval-based uses are generally supportable, while training-based uses require demonstrated item-level provenance, meaningful attribution, and enforceable institutional control, including the right to stop further use.[12] Developing, testing, and disseminating shared standards of this kind—so that thousands of archives, libraries, and museums are not each left to negotiate alone with vendors—is precisely the work no single institution is positioned to do, and exactly what a coordinating center could provide.

Records in Contexts (RiC) illustrates both the promise and the challenge of provenance-rich archival systems. By linking records to creators, functions, activities, mandates, and institutional successors, RiC offers a framework for discovery that extends far beyond keyword search. Yet archivists caution that the standard's complexity may slow adoption, particularly among under-resourced institutions already struggling to describe existing holdings. The challenge facing historical AI is therefore not merely technical. It is organizational and financial. Building provenance-aware systems at scale requires

[12] Leo S. Lo, *The UVA Archival AI Protocol: A Provenance, Consent, and Control Standard for Generative AI Use of Archival Collections*, developed with the UVA Special Collections and Preservation leadership team (University of Virginia Library, 2026), https://doi.org/10.18130/5dqf-9w86.

long-term investment in archival description, metadata creation, and standards implementation across thousands of repositories.

Case Study: Why Provenance Graphs Matter More Than Search

Former National Archives official Andrew Potter recently demonstrated a problem that will become increasingly important in the age of AI. Working with records related to the federal government's Yucca Mountain nuclear waste repository program, Potter compared ordinary keyword search with a provenance-based archival graph built using the international Records in Contexts (RiC) standard.

A keyword search for "Yucca Mountain" in the National Archives Catalog returns nearly 5,000 records. Yet the search provides little understanding of the broader institutional history of the program. The repository involved the Department of Energy, Nuclear Regulatory Commission, Environmental Protection Agency, U.S. Geological Survey, Government Accountability Office, White House offices, and predecessor agencies stretching back decades. A researcher can retrieve records containing the phrase "Yucca Mountain," but the catalog does not reveal how these agencies relate to one another, what legal authorities connected them, or how responsibilities changed over time.

Potter's experiment demonstrated that a provenance graph can answer a fundamentally different question. By encoding relationships among agencies, statutes, activities, events, and institutional successors, the graph allowed researchers to discover records that no keyword search could retrieve. Most strikingly, the graph linked modern agencies such as the Department of Energy and Nuclear Regulatory Commission back to records created by their predecessor, the Atomic Energy Commission, years before the term "Yucca Mountain" existed. Those records are part of the program's documentary history but are effectively invisible to keyword search because they do not contain the modern terminology.

The significance extends beyond archives. AI systems increasingly rely on retrieval pipelines that resemble advanced forms of search. If those systems operate only on textual similarity, they inherit the same limitations: they can find words but not institutional relationships, chains of authority, or the evolution of programs across organizational change. Potter's findings suggest that the future of trustworthy historical AI may depend as much on provenance infrastructure as on model architecture. Records in Contexts offers one example of how historical and archival expertise can be encoded into machine-readable systems that preserve institutional memory rather than merely retrieving text.

As Potter concludes, the critical question is not whether records exist but whether the relationships among them are represented. In an AI-mediated information environment, those relationships may become as important as the records themselves.

Source: Andrew Potter, "What Keyword Search Can't Tell You About Nuclear Waste," Provenance Lab Series, June 2026.

What Historians Provide

Large language models are already trained on archives, books, newspapers, web records, and public documents. History is already inside AI. The question is whether historians will help govern how it is used.

Historians are experts in reasoning under conditions of incomplete evidence. They are trained to distinguish absence from uncertainty, fragment from pattern, and surviving records from representative reality. In an era when AI systems routinely generate confident claims from partial or distorted data, these are no longer academic skills—they are essential safeguards for trustworthy knowledge.

Historians are not arriving late to AI. For more than twenty years, digital historians and digital humanists have been building methods for large-scale search, text mining, web archives, digitized newspapers, government records, visual corpora, adapting network analysis and data science across different scales in which humans live (such as from micro to macro, or individual and institutional and societal) and to work with born-digital sources.

The question is not whether historians can adapt to computation; many already have. The deeper point is that historical scholarship has long developed precisely the forms of reasoning that contemporary AI systems lack or handle poorly. As societies delegate more knowledge work to machines, methods for interpreting time, incomplete evidence, disagreement, silence, provenance, and changing meaning become foundational rather than optional. Historians bring those capacities now.

Historians have spent generations developing methods for reasoning under conditions of incomplete evidence, contested interpretation, and change over time. Those methods are increasingly necessary for the design of reliable AI systems.

Capacities Historians Bring to AI

• **Chronological reasoning:** working with dates, calendars, uncertain timestamps, lag effects, simultaneity, recurrence, and distinctions between when something occurred, when it was

recorded, when it was archived, when it was digitized, and when it was later remembered, narrated, reinterpreted or politically reactivated. Historians understand that chronology is not a simple sequence of dates but a layered problem of temporal registration, competing calendars, retrospective periodization, and unequal archival survival. They distinguish event-time from record-time, publication-time from reception-time, and historical occurrence from later commemoration. This is increasingly important for AI systems, which often collapse all timestamps into a single notion of "date" and fail to model asynchronous evidence or changing temporal context. *Related work: Marnie Hughes-Warrington's History Goes to the Movies and Big History on narrative time, scale, and temporal framing; E. H. Carr's What Is History? on historical facts and chronology; Reinhart Koselleck's Futures Past on historical time, layered temporalities, and shifting horizons of expectation; Fernand Braudel's The Mediterranean on multiple temporal scales (event, conjuncture, longue durée); Jo Guldi and David Armitage's The History Manifesto on recovering long-range temporal analysis; Vanessa Ogle's work on global standardization of time and calendars; Ian Milligan on timestamped born-digital archives and web chronology. Reinhart Koselleck's conceptual history; Jo Guldi's The Dangerous Art of Text Mining; Ian Milligan's The Transformation of Historical Research in the Digital Age. Melvin Wevers and Marijn Koolen, "Digital Begriffsgeschichte"; Simon Hengchen and collaborators on changing vocabularies and word embeddings.*

This matters for AI because current systems often prefer concise, singular explanations that read coherently but misrepresent how real change occurs. Large language models tend to compress complex events into one dominant cause, overfit to the most repeated narrative in their training data, and underweight structural or less narratively salient factors. They may also confuse correlation with causation or treat retrospective commentary as equivalent to contemporaneous evidence. Systems used in policy, law, risk analysis, journalism, intelligence, or education need to represent multiple causal pathways, evidentiary uncertainty, and unresolved scholarly disagreement rather than producing falsely definitive answers. Historical methods can help design AI systems that rank explanations, compare evidence, preserve minority interpretations, and distinguish triggers from background conditions.

• **Contextual probability and historical inference:** AI systems already operate probabilistically, but probability without historical context is not the same as historical judgment. Large language models estimate likely continuations of text; historians estimate the likelihood, plausibility, and explanatory force of claims in relation to evidence, provenance, chronology, absence, power, and context. Historical research does not assume a single uncontested truth, but it does treat factuality as foundational: claims must be tested against records, rival interpretations, known silences, and the conditions under which evidence was produced. This points toward an expanded form of probabilistic reasoning for AI: not probability as textual likelihood alone, but probability disciplined by historical context.

Such work can only be led by historians in collaboration with statisticians, archivists, and computer scientists. Existing quantitative historical social science has already shown the value of modeling time, context, and parameter variation, while recent work in political methodology emphasizes inference to the best explanation rather than statistical association alone.[13]

But AI raises the stakes. A historically responsible model must learn not only what answer is probable, but what kinds of evidence make an answer reasonable, what contexts change its meaning, what absences weaken it, and what alternative explanations remain viable.[14] The aim is not to make history falsely certain. It is to build systems that can represent likelihood, uncertainty, and explanation in historically accountable ways.

• **Source Criticism:** evaluating provenance, authorship, audience, purpose, bias, transmission history, preservation history, digitization history, metadata, and evidentiary limits. Historians do not treat documents as transparent containers of fact. They ask who created a source, for whom, under what constraints, through what institutional process, and why it survived. They distinguish between a police report, diary, corporate memo, propaganda tract, oral testimony, leaked file, newspaper summary, later transcription, and machine-readable scan because each carries different evidentiary weight and different forms of distortion. They also examine how records were selected, cataloged, redacted, translated, digitized, OCR-processed, excerpted, or decontextualized over time.

This matters for AI because most current systems ingest documents as interchangeable text tokens. In doing so, they often flatten crucial distinctions between primary and secondary sources, eyewitness and hearsay accounts, authentic and manipulated documents, contemporaneous records and retrospective summaries, or official narratives and marginalized testimony. Models may treat repetition as credibility, visibility as importance, and cleanly digitized records as more authoritative than fragmentary but significant evidence. They also inherit distortions introduced by OCR error, missing metadata, selective digitization, platform ranking systems, and archival survival bias. Related work: Andreas Fickers on digital hermeneutics and digital source criticism; Michelle Caswell and Anne Gilliland on archival imaginaries, recordkeeping power, and representation; Marisa Fuentes' Dispossessed Lives on reading violence, absence, and fragmentary archives; Matthew Connelly's The Declassification Engine on secrecy, records, and democratic accountability; Giovanna Ceserani and digital prosopographiesLara Putnam on searchable text and transformed evidentiary practice.Sternfeld: AHA 2016 on Digital Historical Context and the Challenge of Preserving Data Integrity; Karen Frost-Arnold: The Epistemic Dangers of Context Collapse Online (OUP 2021).

---

[13] For instance of a reworking of Bayesianism with historical context, see Ira Katznelson and Gregory Wawro of *Time Counts: Quantitative Analysis for Historical Social Science* (2022).

[14] For this argument, see Alexander Cirone and Arthur Spirling, "Turning History into Data: Data Collection, Measurement, and Inference in HPE," https://arthurspirling.org/documents/hpe.pdf

Without source criticism, AI systems risk producing confident answers from weak evidence, amplifying state or corporate perspectives because those records are abundant, or hallucinating certainty where the source base is partial and contested. In legal, journalistic, governmental, scientific, and educational contexts, this can produce serious downstream harms. Provenance-aware systems, by contrast, could weight evidence differently, expose chains of custody, flag uncertain attribution, preserve citation trails, distinguish document genres, and indicate when conclusions rest on incomplete or skewed archives. Historical methods therefore offer a blueprint for building AI systems that are not only fluent, but accountable and trustworthy.

Contextual Interpretation: Historians and the Next Generation of AI

Context lies at the heart of historical interpretation. Historians have debated how to situate evidence within social, political, spatial, environmental, material, and temporal contexts for as long as the discipline has existed. That expertise is becoming increasingly important as AI systems confront the limits of current approaches to knowledge generation.

Contemporary large language models rely primarily on probabilistic prediction across enormous textual corpora. These systems have achieved remarkable fluency, but their weaknesses are also becoming increasingly apparent. Hallucinations cannot simply be solved by adding more data, nor does the addition of ever-larger prompts reliably produce deeper understanding. Current architectures often struggle to distinguish foreground from background, event from structure, local meaning from systemic context, or signal from archival noise. They frequently flatten relationships between scales, compress ambiguity into coherence, and lose track of how meaning changes across time, geography, institutions, and media forms.

Historians are unusually well prepared to confront these problems because historical scholarship has long been organized around precisely such questions of contextual interpretation. Over the past half century, historians have developed sophisticated frameworks for thinking across scales and forms of evidence: the spatial turn, environmental history, material culture studies, global history, microhistory, conceptual history, network analysis, and the relationship between local experience and large structural systems. Historians routinely move between the micro and the macro, between individuals and institutions, between environmental systems and political events, between objects and ideas, and between local archives and global processes.

Importantly, these traditions have not remained isolated from computation. From cliometrics in the mid-twentieth century to contemporary digital humanities, historians have repeatedly experimented with computational methods while preserving attention to interpretation, uncertainty, and context. Projects such as the Spatial History Project under Richard White demonstrated how spatial computation could transform historical interpretation beyond narrowly quantitative approaches. Work by Dan Edelstein, Paula Findlen, and Giovanna Ceserani have

explored how large-scale relational datasets can preserve contextual richness rather than erase it. Tudor Networks of Power joined network and spatial analysis to reconstruct political and intellectual relationships within early modern England, while Myles Lavan's work on ancient slavery demonstrates how computational approaches can illuminate systems of power and social organization across fragmented historical evidence.

The larger point is not simply that historians can "use AI." It is that historians can help improve AI itself. Historical research continually pushes against the limits of current computational architectures because the archive itself resists simplification. Historical evidence is fragmentary, multilingual, multimodal, unevenly preserved, and saturated with uncertainty. It contains minority data, contradictory testimony, shifting vocabularies, incomplete metadata, damaged documents, ambiguous chronology, and changing systems of classification. Historians have developed intellectual traditions precisely for reasoning under such conditions.

This is one reason historians are increasingly important to the future development of AI architectures themselves. As multimodal systems evolve beyond text prediction toward richer forms of reasoning across images, maps, objects, handwriting, audiovisual records, environmental data, and complex relational evidence, historians provide an unusually rigorous body of expertise for modeling context, provenance, scale, and interpretation. Their role is not merely corrective or ethical. It is generative. Historians can help design systems capable of representing uncertainty transparently, preserving disagreement rather than flattening it, reasoning across heterogeneous evidence, and exposing rather than obscuring evidentiary chains.

Projects emerging from the Humanities and AI Virtual Institute (HAVI), such as the SETS project, point toward this future. These initiatives do not simply apply AI tools to humanities data; they use humanities research to challenge the assumptions, architectures, and limitations of contemporary AI systems themselves. This is a crucial shift. The humanities—and history in particular—should not be treated merely as downstream consumers of technological innovation. They are sources of methodological pressure capable of improving how AI systems reason, contextualize, and represent knowledge.

Supporting historians in this work therefore has significance far beyond the discipline itself. Historians can help build more transparent, reliable, provenance-aware, and context-sensitive AI systems precisely because historical research has always confronted the problem that AI now faces at scale: how to reason responsibly in the presence of incomplete, heterogeneous, and contested evidence.

• **The Preservation of a Usable Past:** The historical record of the late twentieth-century onwards is overwhelmingly born-digital in nature, and unlike paper records, born-digital material disappears silently. From the late 1990s onwards, human culture and activity has migrated to the web and other online platforms, creating an invaluable record of our cultures, societies, politics, businesses, and beyond, albeit at the risk of source fragility. Institutions have risen to the challenge – the Internet Archive reached the staggering one trillion archived web page milestone in late 2025 – but preserving even such numbers means that most of what is published on the web continues to disappear, and its continued preservation and access requires continued financial investment and attention to sustainability. Indeed, these archives have formed the basis of many contemporary large language models, with Common Crawl (an open-source crawl of the web) likely constituting the largest source of training information for most leading models.

As LLMs become a key source of historical knowledge today, this raises questions around how these crawls are assembled, what biases their collection methods bring, and what implications this has for historical reasoning. And as original web pages continue to disappear, what could it mean that we are building LLMS, which will serve as the knowledge infrastructure of the future, on an uncertain foundation.

Meanwhile, the archival record is unevenly distributed across regions and languages. Large portions of the Global South remain underrepresented in digitized, machine-readable form.

AI systems trained on this uneven record will reproduce and amplify these asymmetries:

- Archival asymmetry produces geopolitical bias
- Absence of Global South archives leads to systematic epistemic exclusion
- This is not only a question of fairness; it is a problem of model accuracy at global scale

Without intervention, AI systems will encode a structurally partial past as if it were universal.

The challenge is therefore not simply storage. It is the creation of durable public infrastructure for the historical record in an AI age:

- **Born-digital records:** preserving email, websites, databases, social media, messaging platforms, audiovisual files, and other records that may never exist in paper form.
- **At-risk archives:** protecting vulnerable collections threatened by conflict, climate change, institutional collapse, censorship, neglect, or proprietary shutdown.
- **Versioned knowledge:** maintaining authoritative public records over time so that synthetic summaries or AI-generated paraphrases do not displace originals such as legislative debates, court rulings, regulatory filings, or the Congressional Record.
- **Traceable information chains:** preserving provenance trails that allow society to determine when an AI model influenced a decision, spread false information, or

materially shaped an event—as illustrated by recent defamation and misinformation disputes involving AI-generated claims.

Without intervention, we risk a future in which the surviving record is thinner than the original past, and the systems interpreting that record are less transparent than the institutions they replace. *Related work: Taylor Arnold and Lauren Tilton on distant viewing; Kim Gallon on Black digital humanities as a technology of recovery; Roopika Risam on postcolonial and global digital humanities.*

Together, these capacities allow historians to do three things that are directly relevant to LLMs and AI systems: produce new knowledge from large and heterogeneous source bases; audit how existing tools distort archives, search results, and historical interpretation; and design more reliable systems for provenance-aware, temporally grounded, and contestable knowledge. The ask is not to invent this field from scratch. Historians have been doing this work for decades. We want to do it at the scale the AI moment now requires.

Historical Reasoning, Digital Sources, and the Future of Knowledge

Historians of the twenty-first century must understand not only the content of the sources they analyze, but also the changing nature of the digital systems through which those sources are produced, transformed, circulated, and made visible. The critical interrogation of sources—asking who created a record, when, for what purpose, under what institutional conditions, and with what intended audience—has always formed the foundation of historical scholarship. But the migration of human activity into digital environments has profoundly altered both the scale and character of the historical record itself.

Historians now work simultaneously with retro-digitized archives, born-digital records, algorithmically sorted search results, machine-generated metadata, platform-native communication, and increasingly synthetic forms of textual production. This transformation is not merely technical. It changes how knowledge is discovered, evaluated, interpreted, and remembered.

The impact of digital technology on historical research may ultimately prove comparable to the introduction of the microscope and telescope in the sixteenth and seventeenth centuries. Those instruments transformed scientific inquiry by changing the scale at which natural phenomena could be observed. Digital technologies similarly alter the scale at which historical evidence can be perceived and analyzed. Historians can now trace patterns across billions of words, millions of images, centuries of newspapers, or vast networks of communication. Yet scale also introduces new interpretive dangers: distortion through search ranking, decontextualization, metadata error, OCR noise, selective digitization, platform bias, and algorithmic visibility.

As a result, the central questions of historical reasoning must now expand into the digital domain:

- How has digitization changed the nature of historical research itself?
- What happens when analog sources become digital representations?
- How are digital objects created, transformed, and preserved?
- How should historians rethink the concept of the "original" in a world of copies, scans, versions, and synthetic reproduction?
- How do metadata systems shape interpretation?
- How do search engines and ranking systems alter the discovery of evidence?
- What changes when research occurs through interfaces rather than directly in archives?
- How should historians navigate across radically different scales of evidence?
- How do text, images, objects, maps, audiovisual records, databases, and sensor data require different forms of interpretation?

These questions are no longer peripheral to historical scholarship. They are foundational to the future of trustworthy knowledge itself.

Historians bring a uniquely valuable form of reasoning to this environment because historical interpretation has always operated across incomplete, heterogeneous, and contested evidence. Historical knowledge is inherently fragmentary. Archives preserve traces unevenly across time, class, language, geography, and power. Historians therefore develop methods not merely for extracting information, but for reconstructing contexts, evaluating uncertainty, recognizing silences, and reasoning across multiple scales simultaneously.

This becomes especially important in the age of AI because computational systems increasingly mediate access to historical evidence. Search engines rank visibility. Recommendation systems privilege engagement. Large language models compress vast textual corpora into fluent narrative synthesis. In each case, infrastructures shape interpretation before a user ever encounters a source. Historical reasoning therefore can no longer be separated from questions of platform architecture, interface design, metadata, ranking systems, and information retrieval.

Digital humanities scholars have increasingly argued that these transformations require not simply new tools, but a new "style of reasoning" appropriate to a networked, platform-dominated, and deeply unequal knowledge environment. As Annie van den Oever argues in dialogue with Andreas Fickers, digital humanities has generated a new constellation of epistemic virtues extending beyond older ideals of objectivity, autonomy, and precision.[15]

These newer virtues include:

[15] Annie van den Oever, "Six Memos for the New Millennium: A Dialogue with Andreas Fickers on Epistemic Virtues in the Digital Humanities," in Nicholas Baer and Annie van den Oever, eds., *Technics: Media in the Digital Age*(Amsterdam: Amsterdam University Press, 2024), 313–336, DOI 10.1515/9789048564569-019.

- **sharing**, in which knowledge infrastructures become collective rather than proprietary;
- **collaboration**, where historians, archivists, engineers, librarians, designers, and data scientists work within interdisciplinary "trading zones";
- **participation**, through broader public involvement in preservation and interpretation;
- **transparency**, exposing how data, interfaces, and models shape conclusions;
- **traceability**, preserving chains of evidence and transformation;
- **distributed accountability**, recognizing that responsibility now extends across infrastructures and systems rather than individual scholars alone;
- **bibliodiversity**, preserving multilingual and regionally grounded forms of knowledge production;
- and forms of **epistemic and infrastructural justice** attentive to how computational systems render some communities visible while excluding others.

These are not merely ethical aspirations. They are increasingly practical requirements for trustworthy AI systems.

The transition is profound. Older epistemic models often imagined knowledge as the product of autonomous experts producing objective and universal conclusions. The emerging digital environment instead reveals knowledge to be collaborative, infrastructural, probabilistic, unequal, and politically entangled. Historians are particularly well equipped to navigate this shift because the discipline has long confronted the instability of evidence, the politics of archives, and the contextual nature of interpretation.

This is why historians must become central participants in the future development of AI systems rather than passive observers of technological change. Historians do not simply contribute "content" to AI. They contribute frameworks for understanding how knowledge itself is constructed, contextualized, ranked, preserved, and contested.

Indeed, many of the core problems now confronting AI are fundamentally historical problems:

- how to distinguish evidence from repetition;
- how to represent uncertainty without collapsing into incoherence;
- how to preserve disagreement rather than flattening it into consensus;
- how to reason across incomplete archives;
- how to navigate between local context and large-scale patterns;
- how to preserve provenance across transformations of data;
- and how to ensure that infrastructures themselves remain publicly accountable.

Historical reasoning therefore should not be understood as supplementary to AI development. It is increasingly foundational to building systems capable of trustworthy interpretation in democratic societies. These same capacities also define a boundary. Recognizing what historical

and archival expertise makes possible is inseparable from recognizing what should not be delegated to automated systems at all.

What Should Not Be Outsourced

Trustworthy historical AI depends on recognizing where human judgment and institutional accountability cannot be delegated. Certain decisions remain the responsibility of historians, archivists, and librarians, regardless of how capable systems become:

- **Authentication and chain of custody**. Determining whether a document is genuine, altered, or fabricated, and establishing the unbroken trail back to its origin, requires accountable human judgment. This is precisely what generative systems cannot supply, and it grows more critical as AI-generated material enters the documentary record.
- **What counts as evidence**. Distinguishing a primary source from a derivative, an eyewitness account from hearsay, or an authentic record from a later reconstruction is an interpretive act with consequences no automated classifier should foreclose.
- **Accessioning and deaccessioning**. Deciding what enters or leaves the archive—what is preserved, at what level, and what is removed—shapes the historical record itself and cannot be reduced to a retrieval problem.
- **Governance of sensitive collections**. Materials subject to donor restrictions, community protocols, privacy obligations, or legal seal—including records of Indigenous communities, enslaved people, and living individuals—require human stewardship attentive to harm, consent, and context.
- **Authority over training use**. The decision whether a collection may be absorbed into a model that cannot later release it must remain with the custodial institution, exercised deliberately rather than by default.

These are not limitations to be engineered away but indeed are the points at which the discipline's accountability to the public, to descendant communities, and to the historical record is exercised.

Building the Next Generation of Historical Expertise

One of the central challenges facing the relationship between history and AI is institutional rather than intellectual. Today, those working inside AI laboratories and computer science departments often have little training in historical reasoning, archival interpretation, or source criticism, while historians with deep expertise in archives and contextual interpretation frequently lack meaningful pathways into the technical environments where AI systems are being designed. The result is a widening structural divide precisely at the moment when historical reasoning is becoming increasingly important to public knowledge systems.

This divide did not emerge accidentally. There are strong institutional pressures that separate technical and historical expertise early in professional training. Students with quantitative or computational interests are often directed toward engineering, computer science, statistics, or the natural sciences, both because those disciplines promise economic mobility and because universities and governments have spent decades attempting to expand technical capacity and workforce diversity within STEM fields. As a result, many students who might once have pursued hybrid intellectual trajectories are routed away from historical training before such combinations become institutionally viable.

At the same time, historians face structural barriers that complicate sustained engagement with computational research. Historical scholarship in the United States remains unusually concentrated: a relatively small number of institutions disproportionately produce the next generation of academic historians. Professional advancement at those institutions continues to depend heavily upon the production of archive-based, single-authored monographs. These norms reward forms of individualized scholarship very different from the collaborative, infrastructure-intensive, and team-based modes of research common within scientific and computational fields.

There is also a second dimension that is easy to overlook: the gradual separation of historical scholarship from archival practice itself. For much of the twentieth century, the two were continuous. History PhDs were trained in and staffed the archives, and that shared formation built and sustained the records pipelines on which public accountability depends: the systematic appraisal, description, and preservation of government records being one enduring example. Over recent decades, however, archival education and historical training have diverged into largely separate professional tracks, with distinct credentials, distinct institutions, and few pathways between them. The result is not only that historians are underprepared for computational work, but that the discipline has lost the dual-trained scholar—at home in both historical interpretation and custodial practice—precisely when AI makes that combination essential. Systems that reason over archives at scale require people who understand both how meaning is made from evidence and how evidence is appraised, arranged, described, and preserved.

This divergence has deepened over the past several decades. Since the late twentieth century, many of the most influential currents within the historical profession have emerged from social and cultural history, fields devoted to recovering the experiences of communities and actors previously excluded from dominant narratives. This work transformed the discipline intellectually and politically, producing richer and more democratic forms of historical scholarship. Yet the archival and interpretive demands of such research also reinforced methods centered on contextualized close reading, microhistorical interpretation, and intensive engagement with fragmentary evidence. At the very moment computation was becoming increasingly central elsewhere in the university, historians were often moving toward forms of scholarship that understandably prioritized interpretation over quantification. As Ian Milligan notes, historians today operate within a condition of "digital abundance," yet the discipline's interpretive traditions remain rooted in close engagement with contextualized evidence rather than automated synthesis or scale alone.[16]

As a consequence, historians today may be further from mainstream computational training than at almost any point in the modern history of the discipline. Ironically, this was not always the case. Historians in the early twentieth century—and again during the quantitative and cliometric turn of the 1960s and 1970s—were often expected to possess significant facility with statistics, demography, census analysis, economic modeling, and quantitative reasoning. Quantitative methods expanded dramatically within historical scholarship during the 1960s and 1970s before declining sharply after the mid-1980s.[17] Since the 1980s, many of these approaches have migrated institutionally into economics, sociology, geography, political science, and demography programs, frequently outside history departments altogether.[18]

The result is not simply a skills gap. It is the absence of an institutional pipeline capable of producing scholars fluent simultaneously in archives, historical reasoning, digital infrastructure, and AI systems. Existing scholarship on digital history repeatedly identifies the same structural problem: historians increasingly work with computationally mediated archives and large-scale digital corpora, yet the institutional systems for training, infrastructure, collaboration, and long-duration technical maintenance remain fragmented and underfunded.[19]

At precisely the moment when AI systems increasingly depend upon historical data and historical interpretation, the United States lacks durable institutions capable of training experts who can move confidently between archives, computational methods, public-interest technology,

---

[16] Ian Milligan, The Transformation of Historical Research in the Digital Age (Cambridge: Cambridge University Press, 2022).

[17] Steven Ruggles, "The Revival of Quantification: Reflections on Old New Histories," *Social Science History* 45 (2021): 1–25.

[18] Claudia Goldin, "Cliometrics and the Nobel," Journal of Economic Perspectives 9, no. 2 (1995): 191–208.

[19] Andreas Fickers, "Digital Hermeneutics: The Reflexive Turn in Digital Public History?" in *Handbook of Digital Public History*, 139–148, DOI 10.1515/9783110430295-012; Matthew K. Gold, ed., Debates in the Digital Humanities (Minneapolis: University of Minnesota Press, 2012).

and large-scale knowledge infrastructures. Existing structures are poorly suited to the task. Most history departments were not designed to sustain long-term technical infrastructures, permanent engineering staff, or laboratory-style collaborative research environments.[20] Most computer science departments do not train students in provenance, archival reasoning, or historiography. Libraries and archives possess critical expertise but are rarely funded to operate as large-scale research laboratories. Meanwhile, postdoctoral structures common in the sciences remain comparatively rare, fragile, and expensive in history and digital humanities, limiting the development of long-term collaborative research ecosystems.

[20] Matthew K. Gold and Lauren F. Klein, eds., Debates in the Digital Humanities 2016 (Minneapolis: University of Minnesota Press, 2016).

Addressing this problem therefore requires more than curricular reform. It requires new institutions. What is needed is a network of laboratories, centers, and long-duration collaborations in which historians, archivists, librarians, data scientists, computer scientists, engineers, and public-interest technologists can work together on the fundamental problems of historical knowledge in the age of AI. Such institutions would create environments where historical expertise is not peripheral to technological development, but constitutive of it.

The emerging literature on digital humanities suggests that this transformation also requires a new "style of reasoning." As Annie van den Oever argues in dialogue with Andreas Fickers, the digital humanities are increasingly organized around epistemic virtues such as collaboration, traceability, transparency, participation, distributed accountability, and infrastructural justice rather than older ideals of solitary scholarly autonomy alone.[21] These frameworks matter directly for AI because they foreground precisely the problems contemporary systems struggle with most: provenance, interpretive uncertainty, multilinguality, infrastructural inequality, and the politics of visibility inside networked systems.

These centers would therefore provide long-term technical staff and infrastructure rather than temporary grant-based experimentation; stable collaborative environments capable of sustaining ambitious archival and computational projects over decades; graduate and postdoctoral pathways rewarding interdisciplinary fluency rather than penalizing it; and direct institutional bridges connecting universities, archives, public agencies, and AI developers. Students entering doctoral programs would see viable career pathways combining historical scholarship with computational expertise. Early-career scholars would no longer need to choose between archival rigor and technical engagement. Historians could participate directly in shaping AI architectures, provenance systems, retrieval pipelines, and public-interest infrastructures rather than encountering them only after deployment.

The larger goal is not simply to modernize historical training. It is to cultivate a new generation of scholars capable of stewarding trustworthy knowledge systems in democratic societies. The challenge posed by AI is ultimately not only technological. It is institutional, epistemic, and civic. Meeting it will require institutions capable of producing experts who can reason simultaneously about archives, infrastructures, interpretation, computation, and public accountability across long historical timescales.

A Big Move

This proposal is not a grant program for a single discipline. It is a nation-building knowledge project: the construction of public infrastructure for trustworthy historical knowledge in the age of AI.

---

[21] Annie van den Oever, "Six Memos for the New Millennium: A Dialogue with Andreas Fickers on Epistemic Virtues in the Digital Humanities," in Technics: Media in the Digital Age.

Comparable precedents include the Human Genome Project, CERN, the NSF supercomputing initiatives, and the public investments that transformed astronomy, climate science, and medicine. Those efforts succeeded because governments and philanthropies recognized that certain forms of infrastructure become foundational to the future capacity of societies themselves. Historical infrastructure has now reached that threshold.

The opportunity before us is not merely defensive. It is not simply about slowing misinformation, preserving archives, or mitigating the risks of AI, although all of those tasks are urgent. The larger opportunity is transformative: to create entirely new capacities for understanding human societies across time, scale, geography, language, and systems of knowledge.

Recent work already suggests what becomes possible when historical methods, large-scale archives, and AI systems are brought together. Ian Milligan's work with web archives and the Internet Archive has demonstrated that born-digital records can allow scholars to reconstruct cultural and political transformations at scales unimaginable a generation ago.[22] Research using large digital corpora drawn from the aftermath of September 11th—including web forums, emails, message boards, and public discussion—has shown that historians can begin reconstructing how societies collectively adapted to crisis in real time rather than only retrospectively.[23] Similar methods could transform how future historians study the COVID-19 pandemic, climate disasters, migration crises, financial instability, or the evolution of democratic norms.

Jo Guldi's *Dissentometer* project points toward another frontier. Rather than flattening public discourse into singular narratives, it explores whether AI systems can preserve patterns of disagreement, contested interpretation, minority viewpoints, and shifting consensus across millions of documents and multiple languages. Such systems could help scholars and citizens understand not only what societies believed, but where they disagreed, how coalitions formed, when dissent disappeared, and how institutions managed conflict across time. In a world increasingly mediated by AI-generated summaries, preserving disagreement may become as important as preserving information itself.[24]

At even larger scales, comparative historical work by scholars such as Walter Scheidel has shown how long-run historical datasets can illuminate recurring patterns in inequality, labor

---

[22] Ian Milligan, *The Transformation of Historical Research in the Digital Age* (Cambridge: Cambridge University Press, 2022), https://doi.org/10.1017/9781009026055.

[23] See, for example, Ian Milligan and Jimmy Lin, "The Web as History," in *The SAGE Handbook of Web History*, ed. Niels Brügger and Ian Milligan (London: SAGE, 2018), 47–63; and the Archives Unleashed project, which has demonstrated large-scale historical analysis using born-digital web archives preserved by the Internet Archive. See also Ian Milligan, "Illusionary Order: Online Databases, Optical Character Recognition, and Canadian History, 1997–2010," *Canadian Historical Review* 94, no. 4 (2013): 540–69, https://doi.org/10.3138/chr.1065.

[24] Steph Buongiorno et al., "Multi-Scale Computational Analysis of Wikipedia's Telling of History Across 357 Languages," *Research Square* preprint (2025), https://doi.org/10.21203/rs.3.rs-8042139/v1.

systems, state formation, demographic change, and social transformation across civilizations and centuries.[25] Researchers such as Peter Turchin have attempted to model structural instability across empires and political systems using historical databases spanning centuries.[26] Thinkers such as Nassim Nicholas Taleb have emphasized the importance of fragility, uncertainty, tail risk, and systemic shocks in understanding how societies respond to unpredictable events.[27] Whatever the limitations of such approaches—and historians are often among their most important critics—they demonstrate the extraordinary possibilities opened by connecting historical reasoning to planetary-scale datasets.

Yet these projects also reveal the limits of history conducted outside historical scholarship itself. Large-scale computational analysis can identify patterns, correlations, clusters, and anomalies, but historical interpretation requires more than pattern recognition alone. Historians contribute methods for reasoning about context, causation, chronology, archival silence, contested evidence, institutional power, and changing meaning across time. They distinguish between what survives in the archive and what was representative of lived experience. They recognize that records are produced artifacts shaped by systems of governance, exclusion, censorship, preservation, and memory. Historical reasoning therefore remains essential precisely because AI systems increasingly encounter archives at scales where decontextualization becomes easier and more dangerous.[28]

The potential here is analogous to what CERN achieved for particle physics. Before large-scale shared infrastructure, certain questions simply could not be asked because no individual laboratory possessed sufficient instrumentation, collaboration, or scale. Historical scholarship now faces a similar threshold. Vast multilingual archives, born-digital corpora, environmental datasets, satellite imagery, newspapers, oral histories, government records, audiovisual collections, and cultural repositories exist, but remain fragmented across institutions, languages, formats, and incompatible systems. AI has created the possibility of connecting and reasoning across these archives at unprecedented scale—but only if trustworthy public infrastructure exists to support that work.

Just as CERN enabled discoveries impossible within isolated laboratories, a global historical infrastructure would allow scholars and citizens to ask questions previously beyond the scale of human inquiry. Researchers might reconstruct planetary patterns of migration and disease across

---

[25] Walter Scheidel, *The Great Leveler: Violence and the History of Inequality from the Stone Age to the Twenty-First Century* (Princeton, NJ: Princeton University Press, 2017), https://doi.org/10.1515/9781400884605.
[26] Peter Turchin, *Ages of Discord: A Structural-Demographic Analysis of American History* (Chaplin, CT: Beresta Books, 2016); Peter Turchin et al., "Quantitative Historical Analysis Uncovers a Single Dimension of Complexity That Structures Global Variation in Human Social Organization," *Proceedings of the National Academy of Sciences*115, no. 2 (2018): E144–51, https://doi.org/10.1073/pnas.1708800115.
[27] Nassim Nicholas Taleb, *The Black Swan: The Impact of the Highly Improbable*, 2nd ed. (New York: Random House, 2010); Nassim Nicholas Taleb, *Antifragile: Things That Gain from Disorder* (New York: Random House, 2012).
[28] Jo Guldi, *The Dangerous Art of Text Mining: A Methodology for Digital History* (Cambridge: Cambridge University Press, 2023), https://doi.org/10.1017/9781009270458.

centuries; compare how societies adapted to pandemics, financial crises, or environmental collapse; trace the global circulation of political concepts and social movements across languages and media systems; identify long-term shifts in public trust and democratic legitimacy; or recover marginalized archives and forms of historical experience long excluded from dominant narratives. Astonishing discoveries remain possible—not despite AI, but because responsible AI may finally allow humanity to reason across the full scale of its surviving historical record.

This transformation also requires new epistemic virtues suited to a networked and computational knowledge environment. Digital historians and media scholars such as Andreas Fickers have argued that digital knowledge infrastructures require values different from older models of solitary expertise and closed scholarly production. Emerging virtues include transparency, traceability, openness, collaboration, participation, bibliodiversity, infrastructural justice, and distributed accountability.[29] These are not abstract ideals. They are practical requirements for democratic knowledge systems in a world increasingly mediated by AI.

The challenge is therefore larger than the future of historical scholarship alone. AI systems are rapidly becoming the interface through which citizens encounter law, medicine, journalism, science, public administration, and collective memory. As T*he Declassification Engine* project demonstrates, archives are inseparable from democratic accountability itself: states generate records in order to govern, but citizens require access to those records in order to evaluate how power was exercised.[30] AI systems trained on incomplete, classified, selectively preserved, or distorted archives inherit those asymmetries directly. The future of democratic accountability will increasingly depend on whether societies can preserve inspectable chains of evidence linking AI-generated claims back to verifiable records.

This is why the proposal's central goal is not merely to "include historians in AI." It is more ambitious: to help define the epistemic architecture of democratic knowledge in the AI century.

The Solution: A 100-Year Plan for Institutionalizing Responsibility for Information in an AI World

Vellum documents from the middle ages are still available to us today. We need to be planning on these time scales, because if we lose material, it is lost for good. In a time of ever condensing time spans, Historians have the responsibility of being stewards for the long duree of historical memory.

---

[29] Annie van den Oever, "Six Memos for the New Millennium: A Dialogue with Andreas Fickers on Epistemic Virtues in the Digital Humanities," in *Technics: Media in the Digital Age*, ed. Nicholas Baer and Annie van den Oever (Amsterdam: Amsterdam University Press, 2024), 313–36, https://doi.org/10.1515/9789048564569-019.
[30] Matthew Connelly, *The Declassification Engine: What History Reveals About America's Top Secrets* (New York: Pantheon Books, 2023).

We propose a new institutional ecosystem for historical expertise in the age of AI: an international network of centers with complementary areas of emphasis, anchored by a National Center for Historical Research. These centers would support faculty appointments, graduate training, visiting fellowships for researchers and students, and archive-based fellowships that align preservation priorities with active research needs. They would provide sustained access to data science and engineering expertise through long-term collaboration rather than one-off consulting. Most importantly, they would cultivate a new generation of scholars able to reason across complex networks of documents, datasets, and model-generated representations, while establishing standards for provenance, preservation, and accountability in the public interest.

Seen in this light, the challenge facing historical knowledge is not primarily one of preservation, and still less one of scanning. It is a problem of coordination. Archives preserve records. Libraries steward collections. Publishers curate scholarship. Historians interpret evidence. Technologists build systems for retrieval and analysis. Yet no institution currently exists with responsibility for connecting these functions into a coherent public infrastructure. The National Center for Historical Research would fill that gap, serving as the coordinating institution through which description, provenance, access, standards, and historical expertise can operate at the scale the AI era now requires.

Any large-scale digitization initiative must also acknowledge that no digital archive can ever be fully complete. Privacy restrictions, classified records, business confidentiality, cultural protocols, donor agreements, and legal constraints will continue to limit what can be publicly released. Some institutions already create digital surrogates that omit restricted materials entirely. Future historical AI systems must therefore be designed not only to work with incomplete archives but also to make those absences visible. The challenge is not simply preserving records. It is preserving knowledge about what is missing, why it is missing, and how those gaps shape interpretation.

• Global Network of Centers of Excellence

A coordinated international network of centers organized around thematic strengths and regional needs, sharing common infrastructure, standards, fellowships, and preservation responsibilities. Rather than duplicating one model everywhere, centers could specialize in areas such as the history of medicine, history of technology, government documents and public records, climate and environmental history, migration and diaspora archives, legal history, intellectual history, pre-modern global, microhistory for data, Indigenous knowledge systems, or multilingual textual corpora. An initial configuration might include one major center in Europe, one in the United States (perhaps with future satellite branches across the states), and additional centers in Africa, India, Australia, Canada, and other strategically important regions.

The network would operate as a federated system: shared pipelines for digitization and metadata, common provenance standards, interoperable datasets, joint fellowships, rotating scholar

residencies, and collaborative training programs. Centers would back up one another's collections and systems, reducing the risk of data loss through conflict, climate events, censorship, institutional failure, or technical decay. This model combines local expertise and regional stewardship with global resilience and scale.

Such a network would serve three core functions.

> **Audit.** As AI systems become central intermediaries for knowledge, society needs institutions capable of evaluating how information is produced, interpreted, and circulated. Today, many systems generate persuasive historical claims without meaningful evidence trails, clear provenance, or mechanisms for contestation. This creates a governance gap: consequential decisions increasingly rely on synthetic knowledge, while no durable public institution is charged with ensuring that those claims remain verifiable, accountable, and open to challenge. A network of centers could establish benchmarks for historical reasoning, audit model outputs, test for archival bias, and develop standards for trustworthy use in law, government, education, and industry.
>
> **Build.** The same network could create public-facing tools that demonstrate what trustworthy historical AI should look like. Rather than opaque black boxes, future systems should allow users to inspect sources, compare interpretations, trace claims to documentary evidence, and understand where uncertainty remains. Existing prototypes suggest what is possible: Louis Hyman's historically grounded research workflows, experimental scholarly systems such as ChatPhD, Jo Guldi's DemocracyViewer, and Ian Milligan's work using web archives and born-digital collections for large-scale historical research. These efforts show that AI can be designed not only to answer questions, but to expose reasoning, disagreement, evidence, and historical context.
>
> **Train.** No institution in North America currently provides a robust pipeline combining historical scholarship, archival practice, and AI research. Yet the need is clear. Matthew Connelly's work on secrecy, state records, and declassification demonstrates how large documentary corpora can illuminate democratic accountability. Jo Guldi's work with text mining and civic-scale analysis shows how computational methods can recover long-range political and social patterns. CESTA (in particular Spatial History project, MRoL and Classics) 's work with pre-modern digital history across times and space. This coordinated network could translate such precedents into sustained doctoral training, postdoctoral fellowships, cross-sector residencies, and professional development for historians, archivists, and technologists alike.

The larger principle is that citizens should enjoy a **right to a verifiable historical explanation** when AI systems make consequential claims about the past. It is essential to the American democratic experience - yet unenumerated.

If an AI system informs a court filing, insurance dispute, national security assessment, regulatory decision, hiring process, or public controversy, users should be able to know what records were used, how claims were derived, what evidence was excluded, and where uncertainty persists. Historical method provides the intellectual foundation for such standards because historians are trained to weigh conflicting evidence, evaluate provenance, and make interpretive judgments explicit.

A verifiable past does **not** mean access to every document. It means access to:

- provenance,
- description,
- chains of custody,
- knowledge of restrictions,
- and knowledge of absence.

Updated systems are needed. Cryptographic provenance ledgers can preserve chains of custody across digitization, transcription, annotation, and AI-mediated reuse. “Reverse firewall” architectures can ensure that archives retain authority over how collections are accessed and reused rather than surrendering control once materials enter training pipelines. Records in Contexts (RiC) standards offer a framework for expressing relationships among records, creators, institutions, mandates, activities, and successor agencies, enabling discovery through provenance rather than keywords alone. Together these approaches point toward a new generation of historical infrastructure: systems designed not merely to retrieve information, but to preserve evidence, context, accountability, and the visibility of what is absent as well as what survives.

No existing institution is organized at the scale required to carry out this mission. Universities reward individual scholarship; archives preserve records but are rarely funded to build advanced analytic systems; technology firms optimize proprietary products rather than public accountability. A network of Centers for Historical Research, linked to an international network of partner centers, could coordinate these functions around a shared mandate: trustworthy knowledge about the past in the age of AI.

Without being led by historians, tech solutions will abound. But historical scholarship is needed to adjudicate what is useful, lasting, applied, sustains the right to a verifiable past. The center is what enables and contains that.

Perspective: The Preponderant Significance of Supporting U.S. History in 2026

> *[We must not be distracted from] the urgent imperative of the USA assuming a central role in this space internationally. … I think the proposal should be clear that it is advocating for the creation of a singular USA-based center with a national scope (perhaps with satellite facilities*

in the future), one that will work in tandem with similar institutions internationally - Peter Olsen Harbich

The National Center for Historical Research will respond to the urgent imperative present in the international community of historians for national-scale responses to the introduction of artificial intelligence. The Center will coordinate its efforts in tandem with a network of similar institutions around the globe and work to advance all areas of historical inquiry. However, it will also have a particular mandate to pursue advanced research of archival collections housed in the United States and to coordinate the application of frontier technology to state repositories such as the Library of Congress and the National Archives and Records Administration. This mandate will position the Center as the principal independent research facility in the United States.

[Louis Hyman, do you care to offer a perspective on the importance of American higher ed in 2026 here?]

Perspective: The International Opportunities Represented in 2026

The international network is a beautiful idea - which comes at a time when the USA is both destroying research infrastructures and international collaborations. Is there any way to address this [...]? - Mats Malm, Secretary of the Swedish Academy

[Marnie Hughes-Warrington, would you be willing to contribute a statement arguing for the importance of international research centers here?]

[Andreas Fickers, would you be willing to contribute a statement arguing for the importance of international research centers here?]

Four Foundational Programs

The first five years would establish four interconnected forms of public infrastructure.

1. Truth Infrastructure

The project would build the evaluative and standards architecture for historically trustworthy AI: historical reasoning benchmarks, temporal reasoning datasets, provenance standards, disagreement maps, archive audit tools, and annual public audits of major AI systems. These systems would function as the equivalent of a public standards institute for historical reasoning: establishing common benchmarks for chronology, causation, provenance, uncertainty, dissent, and interpretive plurality. The goal is not merely to measure fluency, but to create public accountability for how AI systems represent the past.

Recent experiments using provenance graphs built on Records in Contexts standards have demonstrated that relationship-aware archival systems can recover institutional histories invisible to keyword search alone (Potter 2026). A National Center could accelerate development of such provenance-aware retrieval infrastructure across archives, libraries, and AI systems.

2. Access Governance

Provenance standards alone are insufficient if AI systems lack legitimate pathways to underlying historical evidence. Increasingly, archives, publishers, libraries, and content providers are deploying technical and contractual barriers designed to limit large-scale automated access to their collections. These restrictions often arise for good reasons: copyright obligations, privacy concerns, server costs, collection stewardship, and the need to prevent uncontrolled extraction of scholarly or archival resources. Yet the result is a growing fragmentation of the historical record across thousands of institutions with differing policies, technologies, and legal frameworks.

A National Center for Historical Research could play a critical coordinating role by developing shared access protocols that allow compliant research systems to interact with historical evidence while respecting the rights and responsibilities of custodial institutions. Rather than attempting to centralize all historical data in a single repository, the Center would work with archives, libraries, publishers, and scholarly projects to establish common frameworks for retrieval, attribution, provenance, licensing, and auditability. In effect, the Center would help create the public-interest infrastructure through which trustworthy AI systems can gain access to evidence while preserving institutional control, scholarly credit, and democratic accountability.

3. Memory Infrastructure

The project would establish durable systems for preserving and democratizing the historical record itself. This includes born-digital preservation pipelines, distributed preservation networks, multilingual archival partnerships, and scholar-driven archival activation programs. Rather than funding isolated digitization projects, Archive Transformation Fellowships would allow scholarly communities to identify neglected archives whose preservation and digitization could unlock decades of downstream teaching, research, and public use. This work would directly address the risks already visible in disappearing web archives, platform collapse, geopolitical archival imbalance, and the concentration of cultural memory inside fragile commercial systems.

A major obstacle to large-scale historical infrastructure is not scanning technology but archival description. International surveys of archives suggest that a substantial proportion of archival holdings remain only partially described, often at the collection or series level rather than at the folder or item level. In many repositories, researchers can identify broad collections but cannot yet discover individual documents through machine-assisted search. Digitization alone does not solve this problem. Records must first be identified, arranged, described, and contextualized before they become discoverable to either researchers or AI systems. Investments in historical infrastructure therefore require sustained support not only for scanning and storage, but also for the archival labor of description, metadata creation, and intellectual control. Without this work, large portions of the historical record remain effectively invisible even when they physically survive.

The popular image of historical infrastructure focuses on digitization: scanners, storage, and online access. In practice, the more difficult challenge is intellectual control. Vast portions of the historical record remain only partially described, often at the collection or series level rather than the folder or item level.

The bottleneck is rarely the scanner. It is the finding aid. A million scanned pages are of limited value if no one knows what they contain, how they relate to one another, what institution created them, what restrictions apply, or why they matter. AI systems excel at summarizing and generalizing across large corpora, but historical understanding often depends on exceptional documents: the overlooked memorandum, the missing correspondence, the anomalous report, the revealing draft. Such materials rarely become visible through scale alone. They become visible because archivists and historians have described them, contextualized them, and connected them to broader chains of evidence. Description is what turns documents into evidence.

Description can be partially automated, but it cannot simply be automated. Machine learning systems can assist with transcription, entity extraction, metadata generation, relationship mapping, and preliminary classification at scales no human workforce could achieve alone. Yet these systems must themselves be designed, trained, evaluated, and curated. Decisions about what entities matter, how records are grouped, which relationships are represented, what levels of uncertainty are exposed, and how provenance is preserved remain fundamentally human choices. The challenge is therefore not whether AI can help describe archives. It is how to build systems in which automation amplifies archival expertise rather than replacing it. The most successful historical infrastructure of the coming decades will likely combine machine-scale processing with human stewardship: automated pipelines that accelerate discovery while remaining accountable to archivists, historians, and the communities whose records they preserve.

Without sustained investment in description, digitization risks producing large quantities of searchable images without meaningful pathways for interpretation.

4. Public-Interest Historical AI

The project would demonstrate what historically grounded AI systems can look like in practice. These systems would not function as opaque black boxes. They would expose provenance trails, preserve disagreement, allow users to inspect evidence, surface uncertainty, distinguish between conflicting interpretations, and show how conclusions were reached. The goal is AI that shows its work: systems designed not merely for speed and fluency, but for democratic accountability and historical transparency.

5. Training and Institutional Transformation

Finally, the project would build the first durable pipeline of historians, archivists, technologists, and public-interest AI researchers trained together around democratic knowledge infrastructure. Fellowships would be nationally and internationally competitive rather than restricted to elite doctoral pipelines, allowing students from less-resourced institutions to enter advanced collaborative training networks. The aim is not merely to "teach historians to code," but to create long-duration interdisciplinary communities capable of stewarding archives, infrastructures, and public knowledge systems across generations.[31] Historical reasoning and fact-checking laboratories in schools, libraries, and public institutions would further extend these capacities into civic life, helping citizens evaluate provenance, inspect evidence trails, understand uncertainty, and critically navigate AI-generated claims.

Universities, libraries, archives, and philanthropic institutions are uniquely positioned to build trustworthy alternatives to purely commercial memory systems. The question is no longer whether AI will shape humanity's understanding of the past. It already does. The question is whether democratic societies will build institutions capable of ensuring that this future remains grounded in evidence, plurality, transparency, and public accountability.

The First Five Years: A $100 Million Program of Work

An initial five-year investment would establish durable public infrastructure, produce visible public benefits, and create a new pipeline of expertise for democratic knowledge in the age of AI. The goal is not merely to fund individual research projects. It is to build the foundational infrastructure necessary for trustworthy historical reasoning at scale: standards for evidence and provenance, systems for preserving collective memory, public-interest AI tools, and institutional

[31] Jo Guldi and David Armitage, *The History Manifesto* (Cambridge: Cambridge University Press, 2014), https://doi.org/10.1017/CBO9781139923880.

pathways capable of training scholars, archivists, engineers, and citizens together around the stewardship of historical knowledge.

The proposed program therefore centers on four interconnected forms of infrastructure: **Truth Infrastructure, Memory Infrastructure, Public-Interest Historical AI, and Training and Institutional Transformation.** Together, these programs would establish the institutional foundations for historically trustworthy AI systems while opening entirely new possibilities for discovery, public access, democratic accountability, and large-scale historical research.

1. Truth Infrastructure

The first priority is the creation of public standards and evaluative systems for historically trustworthy AI. Today, large language models are primarily evaluated for fluency, efficiency, and benchmark performance, not for historical reasoning, evidentiary integrity, or preservation of uncertainty. Yet AI systems increasingly summarize legal precedents, interpret political histories, answer questions about public controversies, and synthesize archival material for governments, courts, journalists, corporations, and citizens alike.

This initiative would establish the equivalent of a public standards infrastructure for historical reasoning: a "NIST for historical AI."[32] Deliverables would include:

- **Historical Reasoning Benchmarks for AI Systems.** National evaluation suites testing whether AI systems can reason about chronology, causation, provenance, event boundaries, archival uncertainty, and contested interpretation rather than merely generating fluent summaries.
- **Temporal Reasoning Datasets.** Large-scale multilingual datasets designed to teach and evaluate distinctions between event-time, record-time, publication-time, memory-time, and retrospective interpretation. These systems would model lag effects, recurrence, simultaneity, competing chronologies, and evolving historical narratives.[33]
- **Public Provenance Standards for AI-Generated Claims.** Open technical and policy standards requiring historically consequential claims generated by AI systems to expose source trails, confidence indicators, citation layers, and documentation of missing or conflicting evidence.
- **Access Regimes for LLM retrieval.** Protocols negotiated between the National Center for Historical Research, archives, and scholarly publishers facilitating data access for LLMs designed or authorized by the Center for historical research.
- **Annual Historical AI Audit.** A recurring public audit assessing major AI systems for historical transparency, preservation of dissent, treatment of uncertainty, archival bias,

[32] Compare the role of the National Institute of Standards and Technology in establishing public technical benchmarks and standards infrastructures.
[33] Jo Guldi and David Armitage, *The History Manifesto* (Cambridge: Cambridge University Press, 2014), https://doi.org/10.1017/CBO9781139923880.

multilingual representation, and patterns of historical invention. The audit would evaluate how systems represent contested legal precedents, constitutional history, national security events, marginalized communities, and rapidly changing public controversies. Published annually, it would function as a "State of Historical AI" index tracking improvements, failures, risks, and emerging best practices over time.

- **Historical Disagreement Maps.** Public-facing interfaces allowing users to inspect where historians, archives, governments, or source traditions disagree about major events, legal precedents, civic controversies, or historical interpretations. Rather than flattening disagreement into singular narratives, these systems would preserve contestability as a democratic virtue.
- **Archive Audit Tools.** Open-source diagnostic systems measuring OCR error, metadata gaps, language imbalance, geographic skew, and selective survival across digitized collections used in AI training and retrieval systems.[34]

Together, these systems would establish historical accountability as a measurable and inspectable property of AI systems rather than an afterthought or marketing claim.

2. Memory Infrastructure

A second priority is ensuring the long-term survival, accessibility, and pluralism of the historical record itself.

The modern historical archive is increasingly born-digital: emails, websites, databases, social media, audiovisual files, cloud documents, messaging systems, PDFs, software repositories, and digital government records. Unlike paper archives, born-digital materials can disappear silently and instantaneously. The collapse of platforms such as MySpace permanently destroyed millions of cultural artifacts, while web decay and proprietary platform control continue to erase large portions of public memory each year.[35]

At the same time, archival visibility remains deeply unequal across languages, regions, and geopolitical systems. Much of the Global South remains underrepresented in machine-readable historical corpora, creating structural distortions in the historical record available to AI systems.[36]

This initiative would establish durable and distributed memory infrastructure through:

---

[34] Andreas Fickers, "Digital Hermeneutics: The Reflexive Turn in Digital Public History?" in *Handbook of Digital Public History*, ed. Serge Noiret, Mark Tebeau, and Gerben Zaagsma (Berlin: De Gruyter, 2022), https://doi.org/10.1515/9783110430295-012.
[35] Ian Milligan, *The Transformation of Historical Research in the Digital Age* (Cambridge: Cambridge University Press, 2022), https://doi.org/10.1017/9781009026055.
[36] Roopika Risam, *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*(Evanston, IL: Northwestern University Press, 2018).

- **A Global History Digitization Initiative.** A large-scale program to digitize and connect historical materials around major themes of global history: migration, labor, empire, climate, commodities, oceans, borders, religion, finance, and circulation.
- **Global South Archival Partnerships.** Long-term collaborations with universities, archives, libraries, and community institutions across Africa, Latin America, South Asia, Southeast Asia, and the Caribbean to expand multilingual, machine-readable corpora and reduce geopolitical imbalance in historical datasets.
- **Archive Transformation Fellowships.** Scholar-driven archival activation programs enabling scholarly communities—not only elite institutions—to identify neglected archives whose preservation and digitization could unlock decades of downstream research, teaching, public access, and interdisciplinary scholarship. Rather than funding isolated archival visits, these fellowships would transform high-value collections into shared public infrastructure.
- **A Scholarly Demand Signal Network.** A recurring international process through which scholarly communities identify endangered, neglected, or underutilized archives for preservation investment.
- **Long-Term Preservation and Distributed Memory Systems.** Research and infrastructure for durable digital storage, redundancy protocols, authenticity verification, and distributed web archiving systems modeled partly on the work of the Internet Archive and related initiatives. The goal is to prevent concentration of historical memory inside fragile or proprietary systems and instead create resilient public infrastructures capable of surviving over decades and centuries.

This work is not merely about preservation. It is about ensuring that the historical record remains plural, inspectable, multilingual, and democratically accessible in the AI century.

3. Public-Interest Historical AI

A third priority is demonstrating what historically grounded AI systems can look like in practice.

Current AI systems are optimized primarily for fluency, engagement, and efficiency. Historical systems designed in the public interest would instead prioritize provenance, transparency, inspectability, uncertainty, dissent preservation, and democratic accountability. They would expose evidence trails rather than conceal them.

Deliverables would include:

- **Trustworthy Retrieval Systems for Courts and Government.** Prototype systems for judges, clerks, regulators, legislative staff, journalists, and public agencies that retrieve historical materials with provenance indicators, citation trails, archival context, and uncertainty flags.

- **Open-Source Archival Research Assistants.** Historian-designed AI systems capable of helping users search, compare, transcribe, summarize, and translate archival materials while preserving source distinctions and evidentiary context.
- **Provenance-Aware Interfaces.** Systems allowing users to inspect how claims were generated, which records were included or excluded, and where disagreements or uncertainties remain.
- **AI Systems That Preserve Disagreement.** Research systems modeled partly on projects such as Dissentometer, designed to surface patterns of dissent, minority interpretation, and contested evidence rather than collapsing all historical discourse into singular narratives.

The objective is not simply to make AI "better." It is to demonstrate a fundamentally different model of AI: systems designed to expose reasoning rather than obscure it, and to strengthen democratic accountability rather than weaken it.

4. Training and Institutional Transformation

Finally, the initiative would establish a durable international pipeline of historians, archivists, librarians, technologists, and public-interest AI researchers trained together around democratic knowledge infrastructure.

The challenge is not merely technical. It is institutional. Existing doctoral pipelines remain highly concentrated, while few programs currently provide sustained training across archives, AI systems, digital preservation, historical reasoning, and public infrastructure.

This initiative would therefore support:

- **International Historical AI Fellowships.** Competitive fellowships open across institutions and national systems rather than restricted to elite doctoral pipelines alone. These fellowships would widen participation, internationalize expertise, and create pathways for scholars from less-resourced institutions to enter advanced collaborative research networks.
- **Scholars-in-Residence Programs.** Long-duration residencies pairing historians with engineers, archivists, librarians, and data scientists to work on public-interest AI problems such as hallucination detection, democratic transparency, provenance systems, collective memory, or causal reasoning.
- **Public Data Innovation Fellowships.** Fellowships bringing civic technologists, librarians, investigative researchers, records managers, and open-government practitioners into collaborative research environments focused on public information systems.

- **Permanent Technical Staff for Historical Infrastructure.** Full-time engineers, data curators, librarians, and research software developers responsible for maintaining long-duration public infrastructure rather than relying solely on temporary grant labor.
- **Archival AI Residencies.** Fellowships allowing practicing archivists and librarians to step out of daily operations for six to twelve months to develop applied AI expertise and carry it back to their home institutions. Modeled on the Council on Library and Information Resources postdoctoral program, these residencies would carry a return commitment, ensuring capacity stays in the field rather than draining into industry.
- **Institutional Collection Fellowships.** Awards made to archives and libraries—not to individuals—to responsibly transform a single significant collection, funding the staff time, technical labor, and compute that grant cycles rarely support. As a condition, the resulting pipelines and protocols would be published openly, so that one institution's investment becomes shared infrastructure others can adopt.
- **Joint Graduate Residencies in Archives.** A requirement that doctoral training supported by the center include a working residency inside a custodial institution, not only a digital humanities laboratory—rebuilding the dual-trained scholar, fluent in both historical reasoning and archival practice, that the discipline has largely ceased to produce.
- **Historical Reasoning and Fact-Checking Labs.** Pilot programs in schools, libraries, and public institutions teaching students and citizens how to inspect provenance, evaluate AI-generated claims, understand uncertainty, distinguish primary from secondary evidence, and critically navigate digital archives.

The goal is not merely to teach historians to code. It is to build durable interdisciplinary communities capable of stewarding democratic knowledge infrastructures over generations.

Why Now

The window for action is immediate. The next generation of AI systems is being trained now. The archives they ingest—and the reasoning standards they encode—will become increasingly difficult to reverse later.

At the same time, archival loss is accelerating. Born-digital records disappear silently. Public records are increasingly controlled by private platforms. Vulnerable archives remain underfunded. Governments continue to classify, redact, or destroy records. And AI systems increasingly mediate how citizens encounter law, science, journalism, medicine, and public memory.

The next decade will determine whether historical knowledge remains plural, contestable, inspectable, and publicly grounded—or becomes concentrated inside opaque systems optimized primarily for engagement and scale.

What Is to Be Gained

The stakes are not only defensive. Getting AI and history right could radically expand public access to knowledge and transform humanity's ability to understand itself.

Citizens could explore the history of public institutions, environmental policy, migration, labor, medicine, or civil rights through systems grounded directly in documentary evidence. Families could reconstruct histories of movement, work, and community across generations. Scholars could compare societies across centuries at scales previously impossible. New forms of multilingual historical synthesis could illuminate patterns of crisis adaptation, democratic resilience, public health response, environmental change, or social transformation across civilizations and eras.

Astonishing discoveries remain possible. Just as CERN created the conditions for discoveries impossible within isolated laboratories, large-scale historical infrastructure could allow researchers to ask questions previously beyond the scale of human inquiry. AI-assisted historical systems may help reconstruct how societies adapted to events such as September 11th or the COVID-19 pandemic in real time across millions of communications, archives, and public records. They may reveal hidden connections across languages, institutions, and archives. They may uncover lost histories, recover marginalized voices, and create new forms of democratic accountability by making the historical record more inspectable than ever before.

The predictive power of the past has never been more relevant. But prediction alone is not the goal. The larger promise is democratic: building public systems capable of preserving humanity's memory while expanding humanity's capacity for historical reasoning, collective self-understanding, and accountable governance.