

Artificial intelligence for oral history and disaster experience: a systematic mapping study

Ahmad Azarsa

York University

Ali Asgary

asgary@yorku.ca

York University

Maleknaz Nayebi

York University

Sean Holman

University of Victoria

Systematic Review

Keywords: Oral History, Artificial Intelligence, Digital Humanities, Machine Learning, Natural Language Processing, Disaster Experience

Posted Date: September 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8751395/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Artificial intelligence is increasingly transforming oral history from a static repository of recorded testimonies into a dynamic, computationally mediated field of inquiry. This systematic mapping study classifies and analyses the state of the art in AI-driven oral history findings from 35 primary studies published between 2004 and 2025 to examine how artificial intelligence technologies are applied across the oral history data pipeline. Guided by the PRISMA-ScR 2018 framework, a structured search across Scopus, the ACM Digital Library, and Google Scholar was conducted to identify peer-reviewed research on AI-driven practices in the creation, archival processing, analysis, and visualization of oral history materials.

The field is evolving rapidly. Improvements in transformer-based automated speech recognition have greatly reduced the transcription bottleneck, especially for high-resource languages and well-preserved archival audio, though challenges remain for low-resource, accented, and noisy recordings. Additionally, natural language processing and large language models are increasingly used to organize content, analyse sentiment, and identify patterns related to memory and trauma. More recent research pushes AI into immersive and conversational systems, making it not only a tool for analysing historical data but also an active participant in public engagement and storytelling.

Despite these advances, the literature highlights persistent ethical and methodological challenges, including the risk of cultural bias, the misrepresentation of traumatic narratives, concerns over data sovereignty, and the loss of paralinguistic and embodied meaning in text-centred workflows. This review identifies key structural gaps in current research. It highlights the growing emphasis in the literature on human-in-the-loop and trauma-informed approaches, ensuring that computational efficiency is carefully balanced with the interpretive and ethical commitments central to oral history practice.

Introduction

Oral history has long been established as a versatile and successful research tool that facilitates the comprehension of diverse topics by recording human stories and subjective experiences (Addanan & Abd Aziz, 2025). By documenting personal memoirs and filling gaps in written sources, oral history bridges the divide between public and academic history, making individual stories more accessible to a wider audience (Stratford, 2019). Traditionally, this discipline has relied on the manual recording and interpretation of human stories to foster empathy and connection among communities. However, the digital transformation has fundamentally altered the landscape of this discipline, introducing what Alistair Thomson described as a paradigm shift in how oral history is recorded, archived, and disseminated (Thomson, 2007; Smyth et al., 2023).

As digital archives grow exponentially in size, the traditional methods of manual transcription and analysis have become unsustainable. For decades, the primary barrier to accessing these vast audiovisual collections has been the transcription bottleneck, due to the immense time and cost required to manually transcribe recordings to make them searchable (Lehečka et al., 2024; Picheny et al., 2019).

Historically, oral history archives have been notoriously underutilized because their audio and video contents were rarely indexed in depth, rendering them unsearchable beyond basic metadata (Frisch, 2016).

In response to these challenges, Artificial Intelligence (AI) and Machine Learning (ML) have emerged as formidable assets in the preservation and innovation of intangible cultural heritage (Fu et al., 2025). For the purposes of this review, AI is operationally defined to encompass computational systems capable of probabilistic learning, pattern recognition, or generative output. This includes Automatic Speech Recognition (ASR) to process vast multilingual collections (Byrne et al., 2004), Natural Language Processing (NLP) to extract semantic meaning (Pessanha & Salah, 2021), immersive technologies to visualize heritage in Virtual Reality (VR) environments (Franke et al., 2024) and generative and Large Language Models (LLMs) which are further shifting the field from preservation to complex computational analysis, allowing researchers to extract semantic patterns, sentiments, and even radical intangibles regarding embodied knowledge (Kenderdine et al., 2021).

Despite the growing body of research on specific technological applications, such as sentiment analysis of trauma narratives (Ashqar, 2025; Blanke et al., 2020) or the use of chatbots for disaster reporting (Tsai et al., 2024), there remains a need to systematically evaluate how these technologies intersect across the entire lifecycle of oral history data. Current literature often addresses these stages in isolation, focusing either on the technical metrics of ASR (Gref, Koehler, et al., 2018) or the pedagogical implications of digital testimonies (Kolb et al., 2025).

To structure the analysis, this review uses a four-stage representation of the oral history lifecycle: creation, archiving/transcription, computational analysis, and visualization/exhibition. This pipeline reflects how oral history materials move from recording to public engagement, and it provides a practical lens for mapping where AI interventions currently occur. This conceptual pipeline is developed in this review, informed by oral history practice and digital humanities work on archives and exhibition.

The purpose of this systematic mapping study is to explore how AI and digital tools are being integrated into the oral history domain and to map their potential implications across the full data pipeline. To achieve this, the study addresses the following research questions:

Q1. What AI technologies are being applied to oral history, and how are they distributed across the data lifecycle?

Q2. In what application domains is AI applied to oral history?

Q3. What ethical, methodological, and interpretive challenges are reported in the literature regarding the integration of AI into oral history practices?

The main contribution of this review lies in its capacity to bridge computational innovation and humanistic inquiry. By extending analysis beyond text-based transcripts to include non-verbal signals such as breathing and silence, and by incorporating embodied and immersive representations of

heritage, the literature suggests a shift toward a post-documentary approach in which digital systems function not merely as repositories but as generative spaces for historical interpretation. Understanding both the capabilities and limitations of these technologies is essential to ensure that AI enhances access and insight without compromising the ethical integrity of oral history.

Methodology

This study follows the Systematic Mapping Study approach, designed to ensure transparency, rigor, and reproducibility in identifying, screening, and mapping existing scholarship on the intersection of AI and oral history. The review process was guided by the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) framework (Tricco et al., 2018), provides structured procedures for search strategy formulation, study selection, and reporting of mapping and scoping reviews. Systematic mapping studies focus on structuring the research landscape, identifying evidence clusters, and revealing knowledge gaps, rather than aggregating effect sizes as in traditional systematic literature reviews (Petersen et al., 2015).

The main objective of this mapping study is to map and characterize how AI technologies are being applied to oral history data and to identify emerging trends, technologies, and challenges across different application domains.

Search Strategy

A comprehensive literature search was conducted in Scopus and the ACM Digital Library, selected for their coverage of interdisciplinary research spanning computer science, digital humanities, and social sciences. Google Scholar was also searched by the same terms, limited to the first 100 results sorted by relevance. To identify additional relevant studies that may not have been indexed in the selected databases. The search was conducted in Nov 2025.

The search strategy combined controlled vocabulary and free-text keywords using Boolean operators. To ensure high relevance, queries were restricted to Title, Abstract, and Keyword fields. To capture the interdisciplinary nature of AI applications in oral history, a multi-string search strategy was employed. Rather than relying on a single compound query, four complementary Boolean search strings were designed (Table 1), each targeting a different analytical dimension of the research questions.

Table 1
Boolean search strings and query logic used for literature identification.

ID	Search Focus	Boolean Search String
S1	Core Intersection (Oral History × AI)	("oral history" OR "digital oral history" OR "oral testimony" OR testimony OR "witness testimony" OR "witness archives" OR "spoken narratives" OR "cultural heritage") AND ("artificial intelligence" OR AI OR "machine learning" OR ML OR "deep learning" OR "natural language processing" OR NLP OR "large language model" OR LLM* OR "automatic speech recognition" OR ASR OR "speech recognition")
S2	Context Domain (Disaster & Crisis)	("oral history" OR "oral testimony" OR testimony OR "witness narratives") AND ("artificial intelligence" OR AI OR "machine learning" OR NLP OR ASR) AND ("climate disaster*" OR "climate change" OR disaster OR crisis OR "environmental catastrophe")
S3	Pipeline Stage (Archive, Visualization, Exhibition)	("oral history" OR "oral testimony" OR testimony) AND ("digital archive*" OR "archival system" OR "data visualization" OR "digital exhibition" OR "interactive exhibition" OR "testimony exhibition*" OR "virtual museum" OR "Virtual Reality" OR "VR" OR "Augmented Reality" OR "AR" OR "immersive exhibition") AND ("artificial intelligence" OR AI OR "machine learning" OR NLP OR ASR)
S4	Computational Methods & Tools	("oral history" OR "oral testimony" OR testimony) AND ("computational analysis" OR "algorithm" OR "machine learning" OR "natural language processing" OR NLP OR ASR OR "speech-to-text" OR "open source archive" OR "digital humanities tool")

Additional refinements were made by filtering for English-language papers published between 2004 and 2025 (21 years).

Eligibility Criteria

To ensure the review focused on high-quality, relevant scholarship, strict inclusion and exclusion criteria were applied (Table 2):

Table 2
Inclusion and exclusion criteria applied for the selection of primary studies.

Inclusion Criteria:	Exclusion Criteria:
• Peer-reviewed journal articles, full conference papers, and select high-relevance workshop papers or preprints that provide unique technical insights or specific context domain not yet available in peer-reviewed journals. • Studies explicitly applying or discussing AI, ML, NLP, ASR, or immersive visualization in relation to oral history, testimony, or digital heritage. • Papers published between 2004 and 2025 in English. • Research contributing either empirical results, system design, or conceptual frameworks relevant to oral history.	• Non-peer-reviewed reports, editorials, or short abstracts. • Studies focusing solely on traditional digital archives without AI or computational elements. • Purely theoretical AI papers unrelated to cultural or oral data. • Articles that use “oral history” metaphorically (e.g., business or sports “histories”). • Studies utilizing Immersive Technologies, Virtual Reality (VR), Augmented Reality (AR), or 3D visualization were included only if they served as the interface for oral history computational data or AI-driven interaction; purely descriptive digitization projects were excluded.

Classification Scheme

To systematically categorize the heterogeneous body of literature, we developed a classification scheme based on the oral history data pipeline. This framework synthesizes established models of the digital curation lifecycle (Higgins, 2008) but is adapted specifically for the linear workflow of oral history processing.

We selected this pipeline as our primary categorization lens because, unlike generic “big data” models, it allows for the precise mapping of AI interventions to the specific archival functions they augment. Studies were coded according to their primary contribution to one or more of four distinct stages:

- **Creation:** Technologies used for elicitation, interviewing, and embodied capture.
- **Archive:** Technologies for transcription, segmentation, and infrastructure.
- **Analysis:** Technologies for semantic processing, sentiment analysis, and pattern recognition.
- **Visualization and Exhibition:** Technologies for exhibition, VR interaction, and public engagement.

The conceptual structure of this analytical framework is illustrated in Fig. 1. Following classification, data was extracted using a structured protocol to capture bibliographic details, context domains, and specific technological interventions.

Study Selection and Screening Process

The search strategy was designed to ensure comprehensive coverage of high-impact venues in both Computer Science and Digital Humanities. The primary search was conducted in Scopus and the ACM Digital Library (n = 248). These databases were selected as the primary indexing services because they encompass the full proceedings of key conferences and high-impact journals relevant to the research domain.

To mitigate publication bias and identify recent preprints or emerging scholarship not yet indexed in primary databases, a supplementary search was conducted in Google Scholar. In accordance with systematic review guidelines regarding the diminishing relevance of search engine results (Haddaway et al., 2015), screening in Google Scholar was restricted to the first 100 results sorted by relevance. This threshold was selected to balance recall with feasibility, as retrieval precision for relevant academic content significantly decreases beyond this point. All identified records were processed for screening. Bibliographic data was exported to Zotero (Zotero, 2024) for management and subsequently imported into Rayyan (Ouzzani et al., 2016), a web-based tool used to facilitate independent, blinded screening.

The selection process consisted of three distinct phases applied against the eligibility criteria defined in Table 2. In the first phase, titles and abstracts were reviewed to exclude clearly irrelevant items. This phase excluded 241 records.

The remaining 107 records were sought for retrieval. After removing 7 technical duplicates identified during full-text review, 100 unique studies were assessed in full. A total of 65 records were excluded for failing to meet specific criteria, primarily due to the absence of an oral history focus or lack of an explicit AI methodology.

To ensure comprehensive coverage of relevant venues not fully indexed in primary databases (Wohlin, 2014), the reference lists of the included studies were manually screened. No additional unique studies meeting the inclusion criteria were identified in this phase. Consequently, 35 studies were included in the final mapping.

Consequently, 35 studies met all eligibility criteria and were included in the final review. Figure 2, the PRISMA-ScR flow diagram, summarizes this selection process.

Data Extraction

A structured data extraction protocol was developed to capture bibliographic metadata and intervention characteristics from the included studies. The extraction focused on the intersection of technological capability and oral history practice. The data extraction fields are presented in Table 3.

Table 3
Data Extraction Fields and Categories

Category	Description
Bibliographic Details	Author(s), Year, Title, Source, DOI.
Context Domain	The specific application area (e.g., Indigenous communities, Holocaust testimonies, Cultural Heritage, Disaster Management).
Technological Intervention	The specific computational tools utilized, classified into ASR/Transcription models (e.g., Wav2Vec, Whisper), NLP/LLMs, or Immersive Technologies (VR/AR).
Pipeline Stage	Classification of the study's primary contribution into one of four stages: Creation, Archive, Computational Analysis, or Visualization & Exhibition.
Outcomes & Challenges	Reported benefits (e.g., efficiency, accessibility) and limitations (e.g., algorithmic bias, hallucination, ethical concerns).

Following data extraction, synthesis was conducted using a thematic analysis approach. To support systematic comparison across studies, an analytical framework based on the oral history data pipeline was employed as an organizing lens for synthesis. This framework was not treated as a methodological intervention, but as a categorization scheme to map how different AI technologies intervene at distinct stages of the oral history lifecycle.

Specifically, studies were coded according to their primary contribution to one or more of four pipeline stages: Creation, Archive, Computational Analysis, and Visualization and Exhibition. This pipeline-based framework enabled consistent classification of heterogeneous research contributions spanning technical, human–computer interaction, and digital humanities perspectives.

Results

Descriptive Findings

This section details the descriptive characteristics of included publications, followed by a thematic analysis organized around the lifecycle of oral history data. A complete inventory of the included studies, detailing their specific methodologies and application domains, is provided in supplementary Table SI 1.

Descriptive Characteristics of the Field

Publication Trends and Temporal Evolution

The data indicate that the application of computational methods to oral history is currently undergoing a period of rapid expansion. As shown in Fig. 3, research output has grown in recent years, with a distinct surge in activity corresponding to the widespread adoption of advanced generative AI models.

Specifically, the combined output from 2024 and 2025 accounts for 17 articles, which represents nearly half (48.6%) of the entire review corpus. This concentration of recent scholarship suggests that the field is moving from an experimental phase into a period of active implementation.

Qualitatively, this timeline reflects two distinct technological eras. The earlier period, anchored by foundational work such as Byrne et al. (2004), focused largely on the technical challenges of transcribing spontaneous speech using older statistical models. In contrast, the current spike in publication is driven by the capabilities of Transformer-based architectures and Large Language Models, which have opened new possibilities for semantic analysis and synthetic media (Lehečka et al., 2024; Cherukuri et al., 2025).

Geographical and Archival Concentration

The geospatial distribution of the included literature is not uniform. Instead, research activity coalesces around specific centres of gravity, typically major digital archives or specialized computational linguistics groups. As detailed in Table 4, the field is primarily dominated by two distinct research clusters: the United States (n = 10) and Central Europe (n = 10), which together account for the majority of the published research.

Table 4
Geographical distribution of primary studies and associated archival collections.

Region	Count	Primary Focus	Key Archives / Collections
United States	10	Social & Clinical Application: Trauma analysis, PTSD prediction, Indigenous sovereignty, community storytelling.	USC Shoah Foundation (VHA), Den sho Digital Repository (JAIOH), World Trade Centre Health Registry, StoryCorps.
Central Europe (Germany, Czech Rep.)	10	Infrastructure & ASR: Transcription portals, acoustic modelling, multilingual ASR, error correction workflows.	MALACH Corpus, CLARIN-D, Regional Oral History Archives (e.g., Prague Spring).
Asia (China, Japan)	3	Intangible Heritage: Dialect preservation, rural traditions, disaster memory.	Local Intangible Cultural Heritage (ICH) Databases, Great East Japan Earthquake Archive.
Middle East (Israel, Palestine)	2	Conflict & Memory: Connective memory, sentiment analysis of contested narratives.	Holocaust Testimonies, Nakba Oral History Archive.
Global / Other (UK, Canada, Switzerland, Portugal)	10	Embodied & Virtual Heritage: 3D visualization, motion capture, mining history, biocultural memory.	Hong Kong Martial Arts Archive, Industrial/Mining Archives, Biodiversity Archives.

The North American Application Hub

The United States is the largest single contributor to the corpus, producing research focused on high-level applications and social impact. This cluster is heavily influenced by the presence of massive, established repositories such as the USC Shoah Foundation’s Visual History Archive (VHA) and the Densho Digital Repository of Japanese American incarceration narratives.

Research emerging from this region often bypasses foundational questions about infrastructure to address complex sociological and clinical inquiries. For instance, Son et al. (2023) utilized the oral histories of 9/11 responders not merely for archival preservation, but as clinical data to predict PTSD symptom trajectories. Similarly, work surrounding the Densho collection has utilized advanced NLP to map intergenerational trauma and semantic patterns within survivor testimonies (Chen et al., 2024a; Cherukuri et al., 2025).

The European Infrastructure Hub

Parallel to the American focus on application, the Central European cluster, specifically Germany and the Czech Republic, is defined by a rigorous focus on technical infrastructure and linguistic optimization. This region acts as the engine room for oral history technology.

Researchers here have primarily concentrated on solving the transcription bottleneck for Germanic and Slavic languages. Key initiatives include the development of the CLARIN Transcription Portal (Draxler et al., 2025) and the optimization of ASR models specifically for the acoustic challenges of historical recordings (Gref, Koehler, et al., 2018; Lehečka et al., 2024). Unlike U.S. studies, which often use off-the-shelf commercial tools to analyse content, European studies usually contribute to the fundamental engineering of these tools.

Global Perspectives on Heritage and Conflict

Beyond these two primary hubs, the review identified a dispersed but critical body of work addressing Intangible Cultural Heritage and conflict narratives in the Middle East and Asia. These studies often arise from urgent preservation mandates. For example, research from China (n = 2) focuses on utilizing machine learning to document endangered dialects and rural oral traditions before they disappear (Fu et al., 2025). Meanwhile, scholarship from the Middle East (Israel and Palestine) engages directly with the role of AI in interpreting contested historical narratives, testing how Large Language Models handle highly sensitive, polarized political testimonies (Ashqar, 2025; Bar-Gil, 2025).

RQ1: The AI-Oral History Data Pipeline

To understand where AI intervenes in the oral history lifecycle, this review mapped the included studies into a four-stage pipeline: Creation, Archive, Analysis, and Visualization. As detailed in the mapping matrix (Table 5), the literature is currently heavily skewed toward the Archive and Analysis stages, reflecting a field that prioritizes processing existing backlogs over the development of new collection methods.

Table 5
Mapping of included studies across the Oral History Data Pipeline stages.

ID	Author (Year)	Short Title	Creation	Archive	Analysis	Visualization
1	Lehečka et al. (2024)	<i>Comparative Analysis of Wav2Vec Models</i>		✓		
2	Pessanha & Salah (2021)	<i>Computational Look at OH Archives</i>		✓	✓	
3	Franke et al. (2024)	<i>Polyvocal Approach to Virtual Heritage</i>				✓
4	Wan & Luo (2025)	<i>Study of Anti-war Memorial Hall</i>				✓
5	Matsubayashi et al. (2022)	<i>Sustainable Community Memory Archive</i>	✓	✓		✓
6	Fu et al. (2025)	<i>AI in Intangible Cultural Heritage</i>		✓	✓	✓
7	S̃vec et al. (2023)	<i>Asking Questions Framework</i>			✓	✓
8	Zhuang & Ma (2025)	<i>LLM-Powered Simulations</i>	✓			✓
9	Meanwell et al. (2024)	<i>Automated Transcription Service</i>		✓		
10	Byrne et al. (2004)	<i>Automatic Recognition of Spontaneous Speech</i>		✓	✓	
11	Picheny et al. (2019)	<i>Challenging Speech Recognition (MALACH)</i>		✓		
12	Tsai et al. (2024)	<i>Co-Design with Indigenous Communities</i>	✓			
13	Chen et al. (2024)	<i>Demystifying OH with NLP (Densho)</i>			✓	
14	Kolb et al. (2025)	<i>Visual Modality in Digital Testimonies</i>				✓
15	Bulín et al. (2025)	<i>Exploring Archives Using AI</i>			✓	✓
16	Smyth et al. (2023)	<i>Multimodal Approaches (Fourth Paradigm)</i>			✓	

ID	Author (Year)	Short Title	Creation	Archive	Analysis	Visualization
17	Luzietti et al. (2021)	<i>FONTI 4.0: Speech-to-Text Evaluation</i>		✓		
18	Bar-Gil (2025)	<i>Holocaust Remembrance in Digital Age</i>		✓	✓	
19	Gref, Koehler, et al. (2018)	<i>Improved Transcription for Digital Humanities</i>		✓		
20	Gref, Schmidt, et al. (2018)	<i>Robust SR for German Interviews</i>		✓		
21	Rollo (2025)	<i>Memory & OH as Biocultural Heritage</i>		✓		
22	Cherukuri et al. (2025)	<i>LLMs for OH Understanding</i>			✓	
23	Pessanha (2022)	<i>Non-verbal Signals in OH Archives</i>			✓	
24	Salesky et al. (2016)	<i>Keyword Search Assessment</i>		✓	✓	✓
25	Halperin et al. (2023)	<i>Conversational Storytelling Agent</i>	✓			
26	Kenderdine et al. (2021)	<i>Radical Intangibles (Embodied)</i>	✓			✓
27	Ashqar (2025)	<i>Sentiment Analysis of Nakba Histories</i>			✓	
28	Draxler et al. (2024)	<i>Speech Technology Services</i>		✓		
29	Brown & Shackel (2023)	<i>Text Mining in Historical Archaeology</i>			✓	
30	Karanikolas et al. (2015)	<i>Processing Oral Dialectal Data</i>		✓		✓
31	Draxler et al. (2025)	<i>Transcribing with Transcription Portal</i>		✓		
32	Lehečka et al. (2023)	<i>Transformer-based SR Models</i>		✓		
33	Blanke et al. (2020)	<i>Neural Networks for Holocaust Memories</i>			✓	
34	Dueck (2024)	<i>AI for Indigenous Oral Histories</i>		✓		

ID	Author (Year)	Short Title	Creation	Archive	Analysis	Visualization
35	Son et al. (2023)	<i>Predicting PTSD with AI (WTC)</i>			✓	

The resulting distribution shows a pronounced concentration of research activity in the Archive and Analysis stages. The Archive stage is the most prominent, with 18 studies (51.4%) applying AI to transcription, segmentation, or archival infrastructure. The Analysis stage follows closely, represented by 15 studies (42.9%) focusing on computational interpretation of oral history data. In contrast, AI applications in the Creation stage remain limited, appearing in only 5 studies (14.3%), while 11 studies (31.4%) address Visualization and Exhibition (Fig. 4). Percentages reflect the proportion of studies associated with each pipeline stage; because several studies span multiple stages, percentages do not sum to 100%.

AI in Creation: Active Elicitation and Embodied Capture

Research addressing the Creation stage remains limited (14.3%), accounting for a small portion of the corpus, yet it introduces a significant conceptual shift. Rather than treating AI as a passive post-processing tool, these studies position it as an active agent in the generation of oral history materials. Conversational Storytelling Agents (CSA) have been deployed to support large-scale narrative collection in contexts such as housing insecurity and disaster management, demonstrating that AI-mediated interviewing can reduce social pressure for some narrators and facilitate more open disclosure (Halperin et al., 2023; Tsai et al., 2024).

Beyond facilitation, a small subset of studies explores AI-driven simulation and embodied capture as boundary cases for oral history creation. Zhuang and Ma (2025) for example, employed large language models to generate synthetic historical dialogue, positioning simulation as a speculative tool for examining under-documented or counterfactual perspectives rather than as a substitute for human testimony. Similarly, Kenderdine et al. (2021) expanded the notion of historical record through the concept of “radical intangibles,” integrating motion capture and AI to preserve embodied cultural knowledge such as martial arts traditions. While these approaches depart from conventional testimonial practice, they engage with longstanding oral history concerns regarding voice, presence, and the limits of archival representation. Taken together, they illustrate how AI-enabled creation functions less as a replacement for oral testimony and more as a site where the boundaries of oral history are actively negotiated.

The Archive Stage: Solving Scale and Multilingualism

The Archive stage represents the most developed area of AI application, encompassing 18 studies (51.4%). Research in this category focuses primarily on automated transcription, segmentation, and scalable archival services. Multiple studies evaluate transformer-based speech recognition models on spontaneous and accented oral history speech, consistently reporting improvements over earlier acoustic modelling approaches (Byrne et al., 2004; Lehečka et al., 2024).

In addition to model evaluation, several studies describe the implementation of large-scale transcription infrastructures. Systems such as the Automated Transcription Service (Meanwell et al., 2024) and the Transcription Portal (Draxler et al., 2024, 2025) are reported to support batch processing and human-in-the-loop correction workflows, enabling the transcription of previously inaccessible archival collections. Studies focusing on multilingual and Indigenous datasets report variation in transcription accuracy across languages and dialects, particularly in contexts involving code-switching or low-resource speech (Dueck, 2024).

Computational Analysis: Semantic and Affective Interpretation

The Analysis stage is addressed by 15 studies (42.9%) and reflects a shift from retrieval-based approaches toward computational interpretation of meaning, affect, and psychological states. Several studies apply machine learning models to identify trauma-related linguistic patterns in oral history narratives. For example, (Son et al., 2023) report that predictive models trained on oral histories of World Trade Centre responders can estimate PTSD symptom trajectories.

Other work applies natural language processing and large language models to semantically organize large oral history collections. Cherukuri et al. (2025) demonstrate that LLM-based classification can group incarceration narratives into thematic categories at scale, while Chen et al. (2024) apply topic modelling techniques to improve navigation in Holocaust testimony archives. Beyond text, some studies incorporate paralinguistic features into analysis. Pessanha (2022) reports that acoustic signals such as breathing and silence can be computationally extracted and used to estimate psychological distress.

Visualization & Exhibition: Interactive and Immersive Interfaces

AI-supported Visualization and Exhibition appear in 11 studies (31.4%) and focus on public-facing engagement with oral history collections. These studies primarily examine interactive digital testimonies, virtual reality environments, and immersive exhibition designs that integrate AI-driven retrieval or generative systems. Kolb et al. (2025) report increased user engagement in three-dimensional interactive testimony environments compared to two-dimensional interfaces. Other studies describe virtual exhibitions designed to present multiple historical perspectives within a shared digital space. Franke et al. (2024) and Wan & Luo (2025) report the use of spatialized virtual environments to enable users to navigate divergent or conflicting narratives associated with cultural heritage and memorial contexts. Across this category, AI systems function primarily as computational backends that support retrieval, generation, or conversational interaction within immersive presentation frameworks.

The reviewed literature exhibits a balanced but differentiated methodological landscape, reflecting parallel advances in infrastructural processing, computational interpretation, and experiential mediation. As summarized in Table 6, research activity is distributed across three non-exclusive methodological

categories, indicating that many studies integrate multiple technical approaches rather than adhering to a single paradigm.

Table 6

Distribution of primary studies by methodological focus (N = 35), categories are non-exclusive, individual studies may contribute to multiple methodological areas.

Methodological Category	Number of Studies	Percentage (N = 35)	Key Technologies	Representative Studies
Infrastructure and Signal Processing	17	48.6%	Automatic Speech Recognition (ASR), acoustic modelling, multilingual transcription, segmentation pipelines	(Byrne et al., 2004; Draxler et al., 2025; Gref, Schmidt, et al., 2018; Lehečka et al., 2024; Meanwell et al., 2024)
Computational Interpretation and Meaning-Making	16	45.7%	Natural Language Processing (NLP), Large Language Models (LLMs), topic modelling, sentiment and trauma analysis, semantic clustering	(Ashqar, 2025; Blanke et al., 2020; Chen et al., 2024a; Cherukuri et al., 2025; Son et al., 2023)
Experiential, Embodied, and Interactive Systems	17	48.6%	Virtual and augmented reality (VR/AR), interactive digital testimonies (IDT), conversational agents, affective computing, embodied capture	(Franke et al., 2024; Halperin et al., 2023; Kenderdine et al., 2021; Kolb et al., 2025a; Pessanha, 2022)

Infrastructure and Signal Processing represent one of the most established areas, accounting for nearly half of the reviewed studies. Work in this category focuses on addressing the long-standing transcription bottleneck in oral history through advances in ASR, acoustic modelling, and scalable archival services. Early studies highlighted the difficulty of processing spontaneous, accented, and emotionally charged speech, while more recent contributions demonstrate substantial performance gains through Transformer-based architectures and the deployment of production-level transcription platforms (Byrne et al., 2004; Lehečka et al., 2023; Meanwell et al., 2024). The presence of multiple large scale service-oriented systems further indicates a shift from experimental prototypes toward operational infrastructure that supports human-in-the-loop correction and archival workflows (Draxler et al., 2024, 2025).

Computational Interpretation and Meaning-Making constitutes a second major methodological strand, emphasizing the extraction of semantic, affective, and thematic structure from oral history materials. Studies in this category move beyond retrieval-oriented indexing to model meaning at scale, applying natural language processing and large language models to cluster narratives, detect thematic patterns, and analyse trauma-related language (Chen et al., 2024a; Cherukuri et al., 2025; Son et al., 2023). Several

works extend this approach by incorporating affective and non-verbal signals, challenging text-centric assumptions and demonstrating that emotional and physiological cues embedded in audio recordings carry significant interpretive value (Pessanha, 2022). Collectively, these studies position AI not only as a tool for access but also as a mechanism for large-scale interpretive analysis.

Experiential, Embodied, and Interactive Systems represent an equally prominent methodological orientation, highlighting growing interest in how AI-mediated oral history is encountered by users. Research in this area emphasizes immersive environments, conversational agents, and interactive digital testimonies designed to facilitate engagement, dialogue, and embodied understanding. Virtual reality and interactive testimony systems have been shown to enhance user engagement and perceived social presence, while also introducing new design and ethical considerations related to authenticity and representation (Franke et al., 2024; Kolb et al., 2025). In parallel, conversational storytelling agents are increasingly used not only for exhibition but also for data creation, enabling the collection of contemporary narratives in contexts where human interviewing resources are limited (Halperin et al., 2023; Tsai et al., 2024). Work on embodied heritage further expands the scope of oral history by capturing physical performance and movement alongside spoken narratives, underscoring a broader redefinition of what constitutes historical evidence (Kenderdine et al., 2021).

Across these methodological categories, the frequent overlap between classifications highlights the inherently interdisciplinary nature of this research area. Rather than advancing through a linear sequence from infrastructure to analysis and then to experience, many studies combine multiple approaches within a single research design. This pattern reflects a field that is simultaneously strengthening its technical foundations while actively exploring new forms of historical interpretation and mediation.

RQ2: Distribution by Context Domain

To examine the primary contexts in which AI is applied to oral history, the included studies were categorized into four context domains. As shown in Table 7, research in this area is predominantly concentrated on historically and socially sensitive materials rather than on technical archival management alone.

Table 7
Distribution of primary studies by research context domain (N = 35).

Context Domain	Count	Percentage	Key Focus Areas
Historical Testimonies, Trauma, & Memory	17	48.6%	Computational memory studies, narrative fidelity, Holocaust & conflict archives, trauma prediction.
Cultural Heritage & Digital Archives	8	22.9%	Archival infrastructure, searchability, multimodal digital humanities theory, museology.
Disaster, Crisis, & Resilience	5	14.3%	PTSD prediction, affective computing, social crisis (e.g., housing insecurity), disaster management.
Endangered Languages & Intangible Heritage	5	14.3%	Preservation of dialects, Traditional Ecological Knowledge (TEK), embodied heritage.

The largest category, Historical Testimonies, Trauma, and Memory, accounts for nearly half of the reviewed corpus (n = 17, 48.6%). Studies in this domain primarily engage with collections documenting conflict, genocide, and human rights violations, where AI is applied to support large-scale analysis of emotionally complex narratives (Bar-Gil, 2025). These works frequently focus on identifying patterns related to trauma, memory, and narrative structure within extensive testimonial archives such as the Visual History Archive.

The second largest group, Cultural Heritage, and Digital Archives (n = 8, 22.9%), addresses challenges related to the organization, accessibility, and discoverability of large digital collections. Research in this domain emphasizes archival infrastructure, search functionality, and multimodal representation within digital humanities and museum contexts.

The remaining domains, Disaster, Crisis, and Resilience (n = 5, 14.3%) and Endangered Languages and Intangible Heritage (n = 5, 14.3%), represent more specialized but emerging areas of application. In these studies, AI is used to document time-sensitive knowledge, support crisis response, or preserve endangered linguistic and cultural practices, often in contemporary or near real-time settings.

RQ3: Ethical and Methodological Challenges

The integration of AI into oral history is not purely additive; it introduces significant friction between computational logic and the ethical imperatives of the discipline. As summarized in Table 8, the review identified four primary categories of challenge: interpretive loss, algorithmic bias, infrastructural opacity, and ethical sovereignty.

Table 8
Reported ethical and methodological challenges in AI-driven oral history.

Challenge Category	Description	Representative Studies
Loss of Paralinguistic Fidelity	Automated transcription pipelines systematically discard non-verbal signals (breathing, silence, hesitation), resulting in the loss of emotional meaning and flattening the oral record into text.	(Gref, Koehler, et al., 2018; Pessanha, 2022; Pessanha & Salah, 2021)
Algorithmic Bias & Cultural Misinterpretation	Models trained in general, high-resource datasets fail to capture the nuance of trauma narratives or specific cultural contexts, often misclassifying sentiment in politically contested histories.	(Ashqar, 2025; Bar-Gil, 2025; Fu et al., 2025)
Infrastructural Opacity ("Black Box")	The widening gap between historians and engineers, combined with reliance on commercial cloud services, renders archival retrieval logic opaque and difficult to interrogate.	(Brown & Shackel, 2023; Meanwell et al., 2024)
Data Sovereignty & Sustainability	Conflicts between open AI training and Indigenous data protocols, the high environmental cost (carbon footprint) of training large multilingual models.	(Dueck, 2024; Tsai et al., 2024; Lehečka et al., 2024)

Discussion

Technical Applications and Contextual Domains

This systematic review examined how artificial intelligence is reshaping the practice and infrastructure of oral history. Taken together, the findings point to a substantive paradigm shift: oral history is increasingly treated not as a static repository of recorded voices but as a dynamic, computational pipeline. While this transformation offers clear benefits in terms of scale, accessibility, and analytical reach, it also introduces methodological, epistemic, and ethical tensions that challenge long-standing principles of the discipline.

From Preservation to Active Elicitation

One of the most significant theoretical shifts identified in this review is the movement of technology from a passive instrument of preservation toward an active agent in the creation of oral history. Historically, recording technologies were designed to capture human testimony without shaping it. In contrast, a small but growing body of work positions AI as an interlocutor, facilitator, or even simulator of historical narratives.

Although research at the Creation stage remains limited, accounting for 14.3% of the corpus (n = 5), its conceptual implications are substantial. Notably, studies in this domain tend to appear in Human Computer Interaction venues rather than history journals (e.g., Halperin et al., 2023; Zhuang & Ma, 2025).

This distribution suggests a disciplinary lag in which technical experimentation is advancing faster than the development of humanities-based methodological and ethical frameworks. As a result, AI-mediated elicitation currently evolves at the margins of oral history scholarship, raising questions about authorship, agency, and the role of shared human vulnerability in testimonial practice.

The Tension Between Scale and Interpretive Fidelity

Across the Archive and Analysis stages, the literature consistently reveals a central tension between computational scale and interpretive fidelity. Advances in transformer-based speech recognition have resolved mainly the volume problem, enabling the transcription and indexing of previously inaccessible massive oral history collections (Lehečka et al., 2023). However, this efficiency comes with trade-offs. As Pessanha (2022) argues oral history meaning often resides in the breath, the pause, and the silence features that text-centric ASR pipelines routinely discard. Consequently, the datafication of oral history risks prioritizing what can be easily transcribed over what is felt. This reductionism extends to the semantic layer, where the use of Large Language Models for sentiment analysis introduces the risk of cultural flattening (Fu et al., 2025). Studies caution that models trained predominantly on Western, high-resource datasets may misinterpret or oversimplify complex affective registers, particularly in survivor testimonies and politically contested narratives (Ashqar, 2025; Bar-Gil, 2025). These findings underscore that scalability alone does not guarantee epistemic accuracy in oral history research.

Organizational and Infrastructural Challenges: The "Black Box" Archive

The integration of AI into oral history archives also reshapes institutional labour and expertise. Several studies highlight a widening gap between humanities scholars and the engineers who design and maintain computational systems (Brown & Shackel, 2023). Without transparent, reproducible workflows, historians risk becoming dependent on tools they cannot thoroughly interrogate, transforming archives into black boxes where decisions about classification, relevance, and retrieval are algorithmically mediated.

This challenge is compounded by infrastructural dependency. Reliance on commercial cloud platforms introduces financial and sustainability concerns, particularly for small institutions and community-based archives (Meanwell et al., 2024). For Indigenous communities and speakers of low-resource languages, the costs associated with fine-tuning or maintaining custom models may be prohibitive, reinforcing existing inequalities and increasing the risk of algorithmic exclusion (Dueck, 2024).

Ethical Frontiers: Sovereignty and Sustainability

Beyond technical and organizational concerns, the review identifies pressing ethical questions related to data governance and environmental sustainability. Principles of Indigenous data sovereignty conflict with extractive AI paradigms that treat oral histories as raw material for large-scale model training. As Tsai et al. (2024) emphasize, openness does not equate to ethical availability. Communities must retain authority over how their narratives are processed, represented, and repurposed.

In addition, the environmental costs of large-scale computation are increasingly difficult to ignore. The energy demands of training and deploying multilingual models introduce carbon footprint considerations into what has traditionally been framed as a preservation-oriented field (Lehečka et al., 2024). These concerns suggest that ethical evaluation of AI in oral history must extend beyond representational harm to include ecological impact.

Limitation and Threats to Validity

While this review offers a comprehensive overview of current research, several limitations should be acknowledged. First, the field of artificial intelligence evolves at a pace that often exceeds traditional academic publication cycles. Many recent advances in generative AI are disseminated through preprints and workshop venues (Cherukuri et al., 2025; Ashqar, 2025), meaning that one preprint study discussed here has not yet undergone extended peer review. As a result, specific findings should be interpreted with caution, particularly those involving rapidly emerging models and experimental systems.

The reviewed corpus also reveals a clear domain bias, with a strong emphasis on Holocaust-related research drawing on large, well-established digital archives. While these collections provide valuable and methodologically rich case studies, this concentration limits the generalizability of the findings to oral history collections that have already been digitized. A substantial portion of oral history materials remain in analogue formats due to financial, technical, or institutional constraints. This imbalance reflects a broader digital divide within the field, where communities with limited digital infrastructure, including many tribal nations and local historical societies, face persistent barriers to participating in computational oral history research.

In addition, although studies on languages such as Czech and German are included, most advanced semantic analysis tools discussed in this review are optimized for English. The performance and cultural validity of sentiment analysis, topic modelling, and large language models in low-resource and non-English languages remain underexamined within the current literature. This linguistic imbalance restricts the applicability of many findings to global oral history contexts and underscores the need for more inclusive computational approaches.

Finally, the literature reviewed is primarily characterized by short-term or project-based evaluations. Few studies examine the long-term effects of AI-mediated oral history practices on narrators, communities, or archives' interpretive practices over time. Consequently, the enduring social, ethical, and epistemic implications of integrating AI into oral history infrastructures remain insufficiently understood and warrant sustained longitudinal investigation.

Conclusion and Future Research Directions

This review demonstrates that artificial intelligence is fundamentally reconfiguring the oral history data pipeline. By mapping existing research across the stages of Creation, Archive, Analysis, and Visualization, the review reveals an uneven but rapidly accelerating integration of computational methods into oral history practice.

The Archive stage has achieved a high level of technological maturity, significantly reducing the transcription bottleneck. The adoption of end-to-end Deep Learning models like Wav2Vec 2.0 and Whisper has created a strong foundation for large-scale digitization, making vast collections that were once inaccessible searchable. However, this technological progress shifts the challenge to the Analysis stage, where interpretive risks increase. As AI systems transition from merely recognizing content to inferring meaning, the risk of epistemic distortion rises, especially when dealing with traumatic or politically sensitive narratives.

At the cutting edge of Creation and Visualization, AI offers new modes of engagement that challenge traditional ideas about presence, empathy, and mediation. Interactive digital testimonies and immersive exhibitions illustrate how sustained engagement can persist in a post-witness era, although the use of AI agents for active elicitation remains debated. The rise of Conversational Storytelling Agents raises questions about whether human interviewers are still necessary, and it raises ethical concerns about shared vulnerability in documenting sensitive histories. Ultimately, this review argues that AI is not just a tool for efficiency but an active mediator that transforms how memory interactions unfold. Success in integrating these technologies depends on a Human-in-the-Loop approach that balances computational power with curatorial ethics and trauma-informed practices.

Based on the gaps identified in this review, future research should focus on creating trauma-informed AI architectures. Standard language models often overlook the seriousness of trauma, sometimes hallucinating facts or softening negative sentiments in summaries. Future studies need to fine-tune models specifically on oral history collections to accurately recognize linguistic indicators of trauma, silence, and distress without misrepresenting them. Additionally, it is crucial to reconsider the dynamic between human interviewers and AI systems. Instead of replacing humans, future approaches should aim for augmented interviewing methods in which AI supports human inquiry by suggesting prompts or identifying emerging themes in real time.

Ultimately, the field needs to go beyond traditional computer science metrics and create standardized evaluation frameworks for humanities AI. Existing metrics like Word Error Rate (WER) provide only a limited understanding of whether an oral history has retained its cultural and interpretive richness. It will be crucial to develop interdisciplinary metrics that evaluate emotional accuracy, ensuring that computational advancements align with the fundamental principles of oral history scholarship.

Declarations

Ethical statements

This article does not contain any studies with human participants performed by any of the authors.

Competing interests

The authors declare no competing interests.

Funding

This research was supported by the Social Sciences and Humanities Research Council (SSHRC) Award No. 895-2025-1013.

Author Contribution

A.As. conceived and supervised the project. A.Az. conducted the systematic search, data extraction, and analysis. A.As. and A.Az. wrote the main manuscript text and prepared the figures. M.N. and S.H. provided critical review and editing. All authors reviewed and approved the final manuscript. S.H., A.As, and M.N. received the funding for the research project.

Acknowledgement

This research was supported by the Social Sciences and Humanities Research Council (SSHRC) Award No. 895-2025-1013.

Data Availability

All data generated or analysed during this study are included in this published article and its supplementary information files. The PRISMA checklist, full search strategies, and the list of included studies with extraction details are provided in the Supplementary Information.

References

1. Addanan FA, Abd Aziz A (2025) A Systematic Literature Review of the Oral History Used in Education Settings. *AJELP: Asian J Engl Lang Pedagogy* 13:61–74.
<https://doi.org/10.37134/ajelp.vol13.1.4.2025>
2. Ashqar HI (2025) Sentiment Analysis of Nakba Oral Histories: A Critical Study of Large Language Models. *Proceedings of the 1st International Workshop on Nakba Narratives as Language Resources*, 30–36
3. Bar-Gil O (2025) Holocaust remembrance in the digital age: The transformative influence of technology, digital archives, and connective memory. *Memory Stud* 18(6):1279–1301.
<https://doi.org/10.1177/17506980241312341>
4. Blanke T, Bryant M, Hedges M (2020) Understanding memories of the Holocaust—A new approach to neural networks in the digital humanities. *Digit Scholarsh Humanit* 35(1):17–33.

<https://doi.org/10.1093/llc/fqy082>

5. Brown M, Shackel P (2023) Text Mining Oral Histories in Historical Archaeology. *Int J Hist Archaeol* 27(3):865–881. <https://doi.org/10.1007/s10761-022-00680-5>
6. Bulín M, Švec J, Ircing P, Frémund A, Polák F (2025) Exploring Oral History Archives Using State-of-the-Art Artificial Intelligence Methods. *Acta Informatica Pragensia* 14(2):207–214. <https://doi.org/10.18267/j.aip.268>
7. Byrne W, Doermann D, Franz M, Gustman S, Hajic J, Oard D, Picheny M, Psutka J, Ramabhadran B, Soergel D, Ward T, Zhu W-J (2004) Automatic Recognition of Spontaneous Speech for Access to Multilingual Oral History Archives. *IEEE Trans Speech Audio Process* 12(4):420–435. <https://doi.org/10.1109/TSA.2004.828702>
8. Chen H, Kim J (Annie), Chen J, Sakata A (eds) (2024) Demystifying oral history with natural language processing and data analytics: A case study of the Densho digital collection. *The Electronic Library*, 42(4), 643–663. <https://doi.org/10.1108/EL-12-2023-0303>
9. Cherukuri KS, Moses PA, Sakata A, Chen J, Chen H (2025) Large Language Models for Oral History Understanding with Text Classification and Sentiment Analysis (No. arXiv:2508.06729). arXiv. <https://doi.org/10.48550/arXiv.2508.06729>
10. Corporation for Digital Scholarship (2024) Zotero. <https://www.zotero.org>
11. Draxler C, Heuvel H, van van den, Hessen A, Ircing P, Lehečka J (2024) Speech Technology Services for Oral History Research (No. arXiv:2405.02333). arXiv. <https://doi.org/10.48550/arXiv.2405.02333>
12. Draxler C, Ircing P, Heuvel J, van den H, Ardolino F, van Hessen A (2025) August 17). Transcribing Oral History Recordings Using the Transcription Portal. <https://doi.org/10.1515/9783110767377>
13. Dueck GW (2024) Using AI to help Preserve Indigenous Oral Histories. 2024 IEEE International Humanitarian Technologies Conference (IHTC), 1–5. <https://doi.org/10.1109/IHTC61819.2024.10855026>
14. Franke IF, Conte SD, Libbi CA, De Boer V, Hartmann T (2024) A Polyvocal Approach to Virtual Heritage: An Immersive Case Study. *J Comput Cult Herit* 17(4):1–23. <https://doi.org/10.1145/3678177>
15. Frisch M (2016) Oral History in the Digital Age: Beyond the Raw and the Cooked. *Australian Hist Stud* 47(1):92–107. <https://doi.org/10.1080/1031461X.2015.1122073>
16. Fu Y, Shi K, Xi L (2025) Artificial intelligence and machine learning in the preservation and innovation of intangible cultural heritage: Ethical considerations and design frameworks. *Digit Scholarsh Humanit* 40(2):487–508. <https://doi.org/10.1093/llc/fqaf034>
17. Gref M, Köhler J, Leh A (2018) Improved transcription and indexing of oral history interviews for digital humanities research. *LREC - Int Conf Lang Resour Evaluation*, 3124–3131
18. Gref M, Schmidt C, Köhler J (2018) Improving Robust Speech Recognition for German Oral History Interviews Using Multi-Condition Training, 13th ITG Symposium on Speech Communication

19. Haddaway NR, Collins AM, Coughlin D, Kirk S (2015) The Role of Google Scholar in Evidence Reviews and Its Applicability to Grey Literature Searching. *PLoS ONE* 10(9):e0138237. <https://doi.org/10.1371/journal.pone.0138237>
20. Halperin BA, Hsieh G, McElroy E, Pierce J, Rosner DK (2023) Probing a Community-Based Conversational Storytelling Agent to Document Digital Stories of Housing Insecurity. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–18. <https://doi.org/10.1145/3544548.3581109>
21. Higgins S (2008) The DCC Curation Lifecycle Model. 3(1)
22. Karanikolas NN, Galiotou E, Papazachariou D, Athanasakos K, Koronakis G, Ralli A (2015) Towards a computational processing of oral dialectal data. *Proceedings of the 19th Panhellenic Conference on Informatics*, 337–341. <https://doi.org/10.1145/2801948.2801966>
23. Kenderdine S, Hibberd L, Shaw J (2021) Radical Intangibles: Materializing the Ephemeral. *Museum Soc* 19(2):252–272. <https://doi.org/10.29311/mas.v19i2.3638>
24. Kolb D, Maiolo S, Maier P, Genz F, Müller S, Kranzlmüller DA (2025a) Effects of Visual Modality on Conversations with Interactive Digital Testimonies: Preparing for the Post-Witness Era. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–25. <https://doi.org/10.1145/3706598.3713111>
25. Lehečka J, Psutka JV, Smidl L, Ircing P, Psutka J (2024) A Comparative Analysis of Bilingual and Trilingual Wav2Vec Models for Automatic Speech Recognition in Multilingual Oral History Archives. *Interspeech 2024*, 1285–1289. <https://doi.org/10.21437/Interspeech.2024-472>
26. Lehečka J, Švec J, Psutka JV, Ircing P (2023) Transformer-based Speech Recognition Models for Oral History Archives in English, German, and Czech. *INTERSPEECH 2023*, 201–205. <https://doi.org/10.21437/Interspeech.2023-872>
27. Luzietti RB, Pretto N, Kaplan F, Dufaux A, Canazza S (2021) FONTI 4.0: Evaluating Speech-to-Text Automatic Transcription of Digitized Historical Oral Sources. In E. Fersini, M. Passarotti, & V. Patti (Eds.), *Proceedings of the Eighth Italian Conference on Computational Linguistics CliC-it 2021* (pp. 387–393). Accademia University Press. <https://doi.org/10.4000/books.aaccademia.10947>
28. Matsubayashi M, Uda N, Joho H, Mizoue C, Toshimori A, Oba Y (2022) A Prototype System of Sustainable Community Memory Archive for Public Libraries. *Libr Inform Sci Res E-J* 32(1). <https://doi.org/10.32655/LIBRES.2022.1.3>
29. Meanwell E, Walsh A, Berg P, Tuna E (2024) Automated Transcription Service: Enabling Transcription at Scale for Social Science Research. *Practice and Experience in Advanced Research Computing 2024: Human Powered Computing*, 1–7. <https://doi.org/10.1145/3626203.3670545>
30. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A (2016) Rayyan—A web and mobile app for systematic reviews. *Syst Reviews* 5(1):210. <https://doi.org/10.1186/s13643-016-0384-4>
31. Pessanha F (2022) Non-verbal Signals in Oral History Archives. *Proceedings of the 2022 International Conference on Multimodal Interaction*, 668–672. <https://doi.org/10.1145/3536221.3557036>

32. Pessanha F, Salah AA (2021) A Computational Look at Oral History Archives. *J Comput Cult Herit* 15(1):1–16. <https://doi.org/10.1145/3477605>
33. Petersen K, Vakkalanka S, Kuzniarz L (2015) Guidelines for conducting systematic mapping studies in software engineering: An update. *Inf Softw Technol* 64:1–18. <https://doi.org/10.1016/j.infsof.2015.03.007>
34. Picheny M, Tüske Z, Kingsbury B, Audhkhasi K, Cui X, Saon G (2019) Challenging the Boundaries of Speech Recognition: The MALACH Corpus. *Interspeech 2019*, 326–330. <https://doi.org/10.21437/Interspeech.2019-1907>
35. Rollo MF (2025) Interconnected Nature and People: Biosphere Reserves and the Power of Memory and Oral Histories as Biocultural Heritage for a Sustainable Future. *Sustainability* 17(9):4030. <https://doi.org/10.3390/su17094030>
36. Švec J, Bulin M, Fremund A, Polak F (2023) August 20). Asking Questions: An Innovative Way to Interact with Oral History Archives. *INTERSPEECH*, Dublin, Ireland
37. Salesky E, Ray J, Shen W (2016) Operational Assessment of Keyword Search on Oral History. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 317–321. <https://aclanthology.org/L16-1049/>
38. Smyth HK, Nyhan J, Flinn A (2023) Exploring the possibilities of Thomson's fourth paradigm transformation—The case for a multimodal approach to digital oral history? *Digit Scholarsh Humanit* 38(2):720–736. <https://doi.org/10.1093/llc/fqac094>
39. Son Y, Clouston SAP, Kotov R, Eichstaedt JC, Bromet EJ, Luft BJ, Schwartz HA (2023) World Trade Center responders in their own words: Predicting PTSD symptom trajectories with AI-based language analyses of interviews. *Psychol Med* 53(3):918–926. <https://doi.org/10.1017/S0033291721002294>
40. Stratford E (2019) Oral History and Narrative. In D. Richardson, N. Castree, M. F. Goodchild, A. Kobayashi, W. Liu, & R. A. Marston (Eds.), *International Encyclopedia of Geography* (1st ed., pp. 1–7). Wiley. <https://doi.org/10.1002/9781118786352.wbieg1038.pub2>
41. Thomson A (2007) Four Paradigm Transformations in Oral History. *Oral History Rev* 34(1):49–70. <https://doi.org/10.1525/ohr.2007.34.1.49>
42. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MDJ, Horsley T, Weeks L, Hempel S, Akl EA, Chang C, McGowan J, Stewart L, Hartling L, Aldcroft A, Wilson MG, Garritty C, Straus SE (2018) PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med* 169(7):467–473. <https://doi.org/10.7326/M18-0850>
43. Tsai C-H, Huang C, Chen Y-C, Zendejas E, Krafka S, Zendejas J (2024) Co-Design Smart Disaster Management Systems with Indigenous Communities. *Digit Government: Res Pract* 5(3):1–21. <https://doi.org/10.1145/3660643>
44. Wan X, Luo Y (2025) A Study of Anti-war Memorial Hall of Leshan City based on Virtual Museum Technology. In *Proceedings of the 2025 2nd International Conference on Informatics Education and Computer Technology Applications (IECA '25)*, (pp. 493–497)

45. Wohlin C (2014) Guidelines for snowballing in systematic literature studies and a replication in software engineering. *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*, 1–10. <https://doi.org/10.1145/2601248.2601268>

46. Zhuang Z, Ma R (2025) Augmenting Engagement with Oral History Archives through LLM -Powered Simulations. *Proceedings of the Association for Information Science and Technology*, 62(1), 1753–1755. <https://doi.org/10.1002/pra2.1531>

Figures

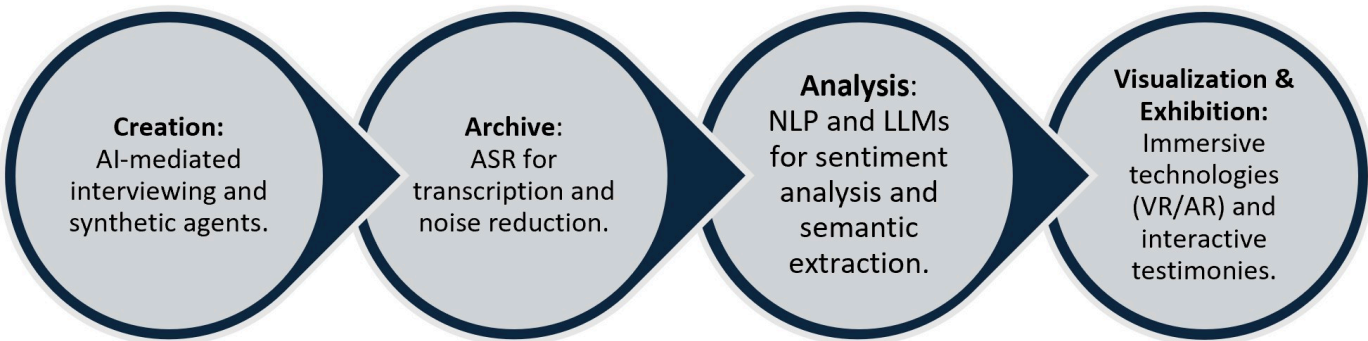


Figure 1

Oral History Data Pipeline and Associated Computational Interventions

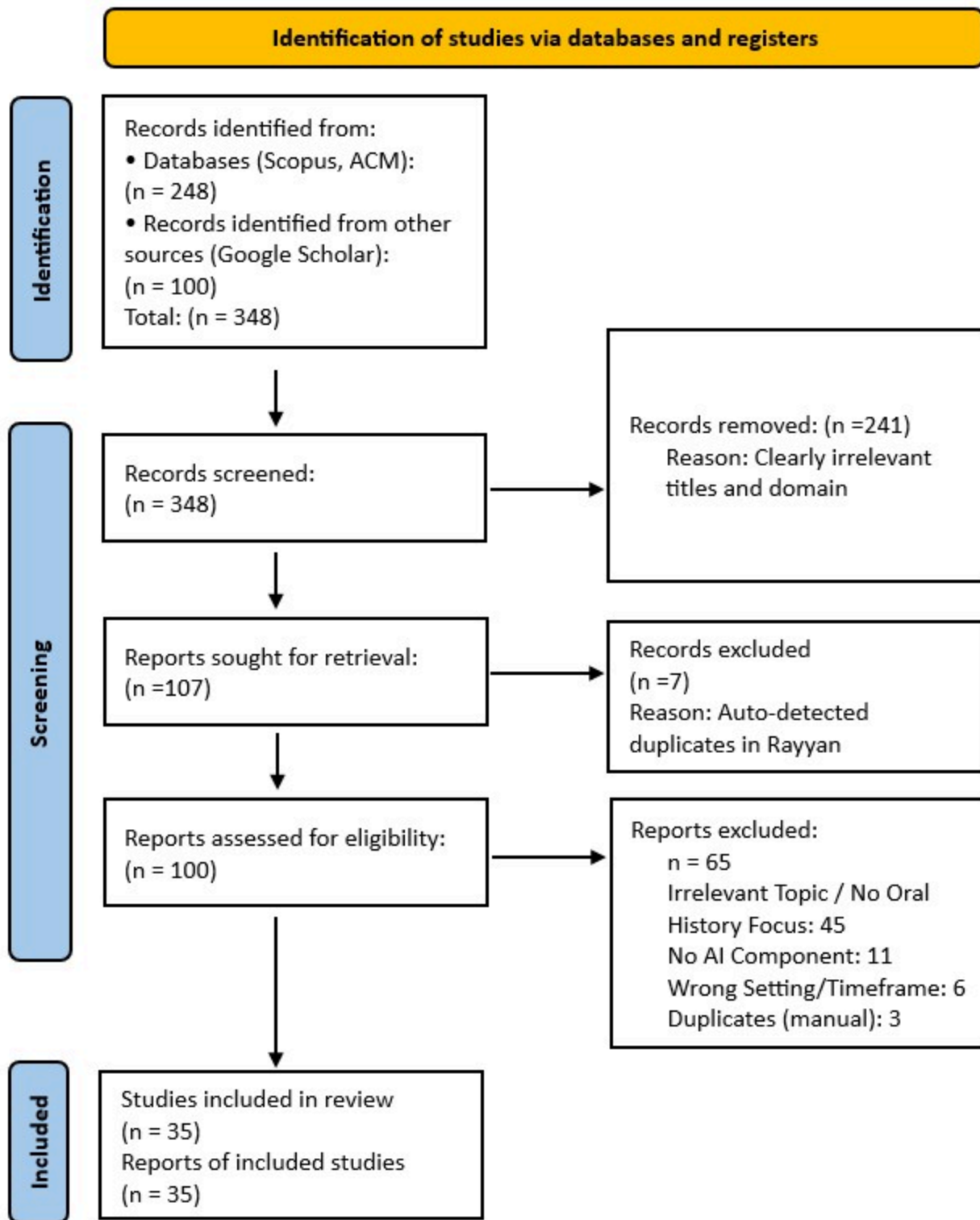


Figure 2

PRISMA-ScR 2018 flow diagram illustrating the identification, screening, eligibility assessment, and inclusion of studies examining Artificial Intelligence applications in oral history research.

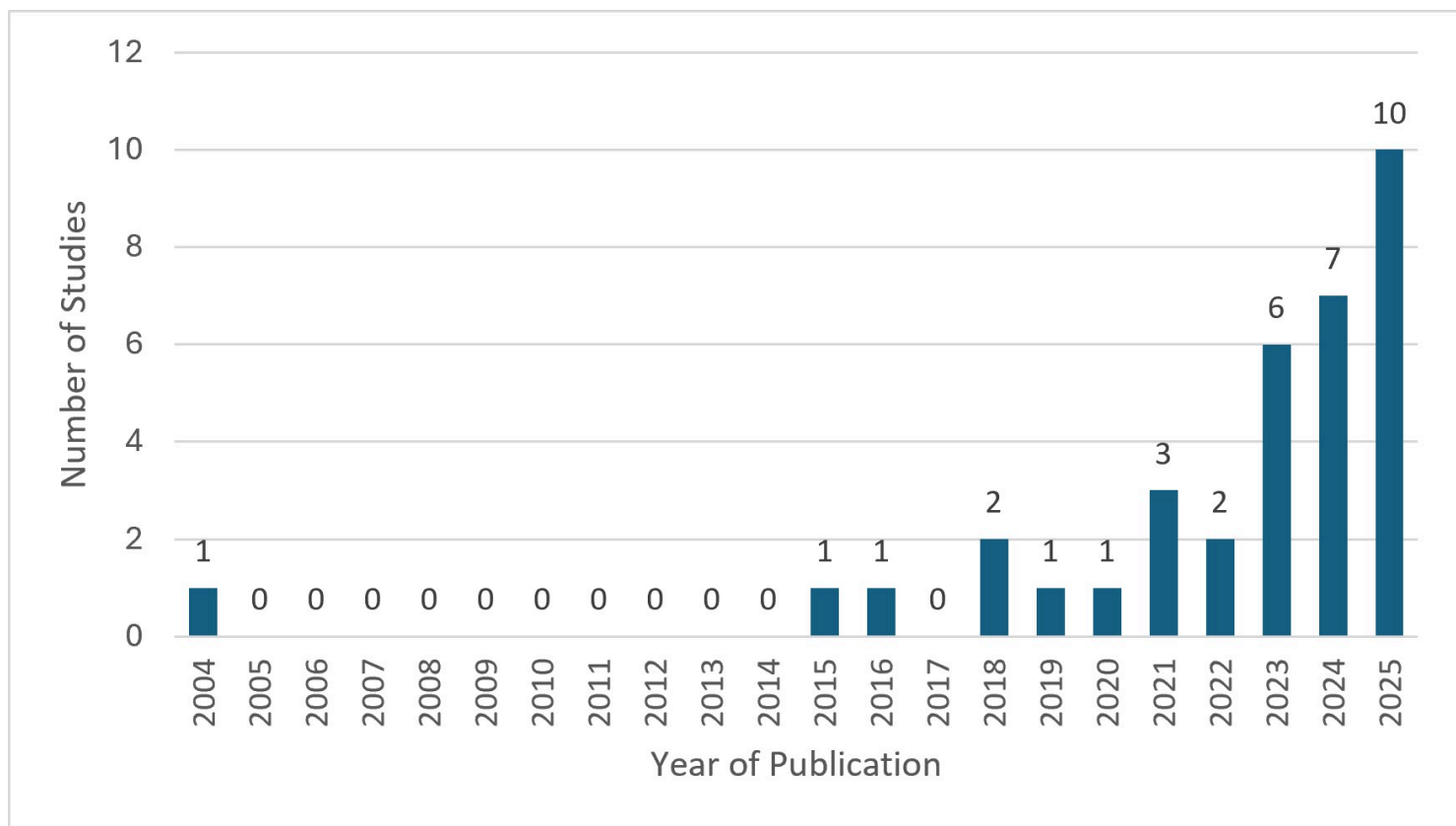


Figure 3

Annual distribution of included primary studies (n=35). The timeline reveals a significant hiatus following foundational work in 2004, with a resurgence beginning in 2015 and an exponential increase in publication volume observed in the 2023–2025 period.

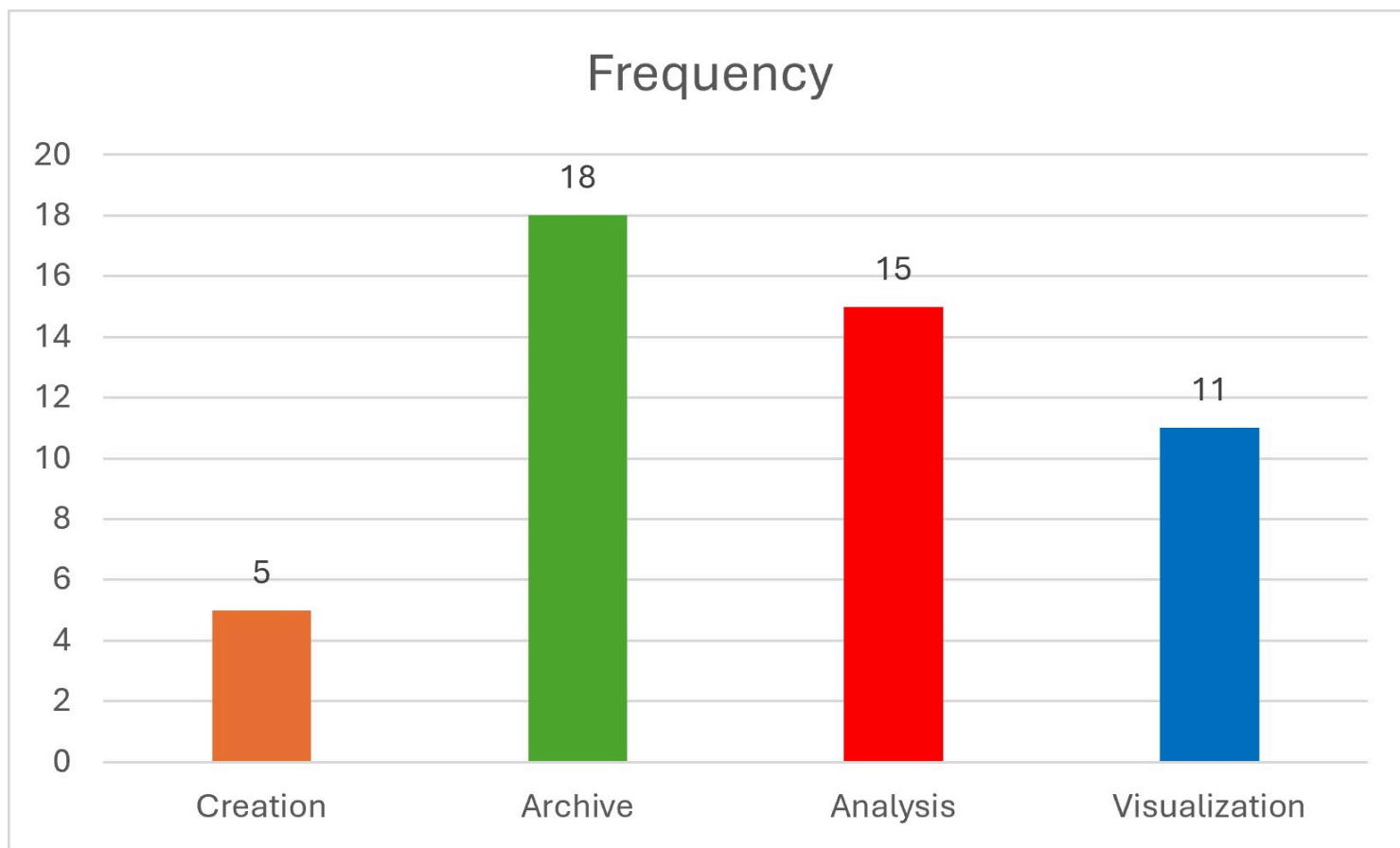


Figure 4

Frequency of papers with focus on different sections of Oral History pipeline

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFileScreeningLogBlind.csv](#)
- [PRISMA ScR Checklist.docx](#)