Conversing with the Past: Immersive Access to Cultural Digital Archives with LLMs and Spatial Interfaces

Michalis Vlachos
University of Lausanne
Lausanne, Switzerland
michalis0@gmail.com

Stergios Konstantinidis
University of Lausanne
Lausanne, Switzerland
stergios@unil.ch

Arman Dzhrahatspanian
Immerse it
Kiev, Ukraine
immerseit1@gmail.com

Min-Yen Kan
School of Computing
National University of Singapore
Singapore, Singapore
kanmy@comp.nus.edu.sg

Abstract

Historical digitized archives, such as newspapers, preserve cultural memory but remain difficult to explore due to noisy OCR and rigid keyword-based search interfaces. We present a research demo that combines retrieval augmented generation with an immersive mixed reality interface on Apple Vision Pro, where interaction is driven entirely by voice and hand gestures. The system integrates LLM-based OCR refinement with a RAG pipeline over historical periodicals, enabling natural language queries, spatial browsing of results, and inspection of supporting evidence through timeline and geospatial visualizations. An empirical evaluation analyzes retrieval quality and system latency, illustrating how evidence-grounded information retrieval supports sensemaking in large archival collections.
A video of the demo is provided here:
https://youtu.be/WeGhnDTTJLY

CCS Concepts

• **Information systems → Users and interactive retrieval**.

Keywords

Immersive interfaces, digital archives, RAG, speech to text, Apple Vision Pro

ACM Reference Format:
Michalis Vlachos, Stergios Konstantinidis, Arman Dzhrahatspanian, and Min-Yen Kan. 2026. Conversing with the Past: Immersive Access to Cultural Digital Archives with LLMs and Spatial Interfaces. In *Companion Proceedings of the 31st International Conference on Intelligent User Interfaces (IUI Companion '26), March 23–26, 2026, Paphos, Cyprus.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3742414.3794770

1 Introduction

There exist many digitized archives, including historical newspapers and magazines, whose access is invaluable to researchers and the public. However, discovery in existing repositories often relies on exact keywords, which are further degraded by optical character recognition (OCR) errors. Large language models enable two complementary improvements for archival access. First, LLM-based correction can sanitize OCR output, improving text quality for downstream tasks. While this has not been systematically studied for historical archives, our experiments suggest modest improvements. Second, via retrieval augmented generation (RAG), one can support conversational search that retrieves by meaning, cites primary passages, and synthesizes information across documents, enabling questions that significantly extend traditional keyword search. Beyond improving retrieval, we introduce a new mode of access by situating these capabilities within an immersive interface on Apple Vision Pro. Mixed reality provides continuous space for browsing results, arranging evidence, and posing follow-up questions through speech. For both professionals and casual visitors, an interface that listens, retrieves, shows, and explains transforms archival exploration from passive viewing into active inquiry.

What are the advantages of this search interface. Scholars and visitors often start with vague intents, not exact strings. They need to move between overview and detail, compare passages, and see what supports a generated claim. Immersive space allows parallel placement of query history, retrieved snippets, full articles and detailed document view. Voice-first input reduces friction for iterative refinement. Together this supports rapid sensemaking, especially when sources are long or when screen real estate is a bottleneck on mobile or desktop.

Contributions. This paper contributes: (i) a prototype system that couples speech recognition with a retrieval augmented generation service and a spatial user interface, enabling evidence-grounded conversational access to historical newspapers and magazines. Such an immersive interaction allows users to move fluidly between overview and detail, compare passages, inspect supporting evidence, and refine queries through voice. By leveraging spatial layout for parallel presentation of query history, retrieved snippets, and full articles scans, the system reduces interaction friction. (ii) We also contribute an empirical evaluation demonstrating the retrieval quality of modern LLMs in the specific archival setting, together with a detailed latency profile of the end-to-end system.

2 Related Work

Prior work shows that immersive media for cultural heritage improves engagement, interpretation, and learning, while raising

ACM ISBN 979-8-4007-1985-1/26/03
https://doi.org/10.1145/3742414.3794770

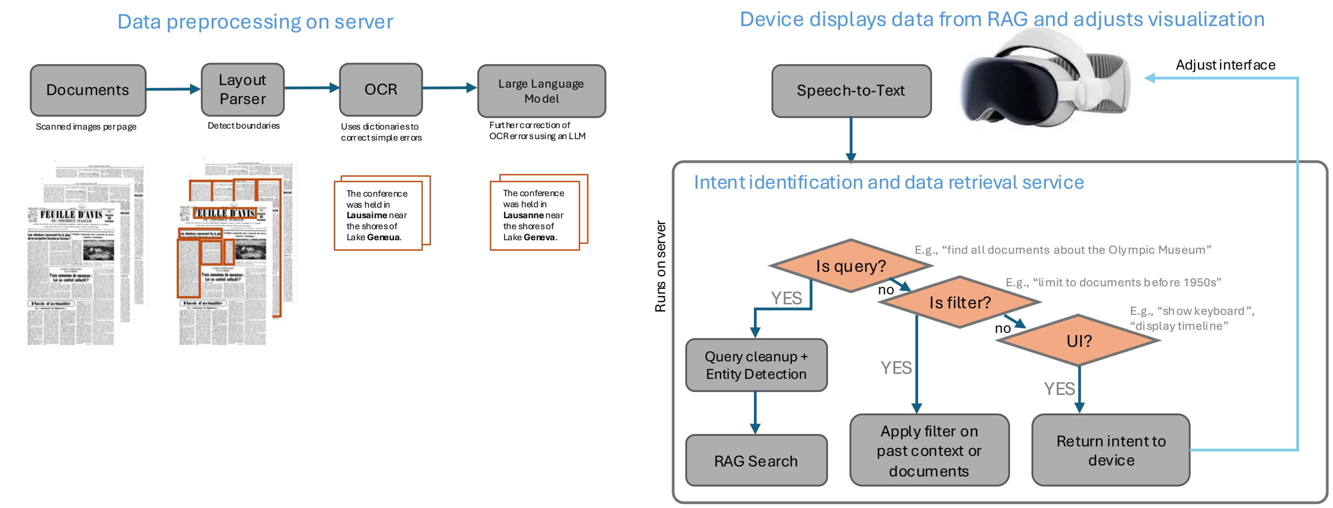

Figure 1: Overview of our pipeline. Left-side (pre-processing) Pages of the documents are scanned to images, boundaries are detected, OCR'd, and OCR text is corrected using LLMs. Right-side (information retrieval). On-device, we classify if the command is intended for the UI or the RAG service to further expedite the system responsiveness.

tradeoffs around immersion, usability, and curatorial control [2, 12]. Museum deployments report increased presence and engagement alongside cost, maintenance, and sustainability concerns [15, 19], and comparative studies of AR, augmented reality, and VR versions of the same exhibit demonstrate effects on experience quality, task performance, and cognitive load [22]. System-oriented work documents interaction patterns for immersive storytelling, intangible heritage, and VR museum systems supporting interpretation and meaning making [21, 23].

Related research in digital libraries and archives examines interaction techniques for exploring historical manuscripts and books [10], highlighting spatial and temporal encodings such as maps and timelines as central to sense making in complex corpora. Recent work further explores AI, conversational agents, genAI, and advanced interaction techniques for libraries, archives, and museums [6], identifying retrieval augmented generation as a promising approach for dialog based access to large collections. Several systems combine retrieval augmented generation with immersive cultural heritage experiences, including spatially aware RAG for virtual museums [16], multimodal reasoning and semantic interaction frameworks [11], and multimodal RAG for archaeological provenance analysis [24]. These systems primarily address exhibit centered or expert use cases, whereas our work targets everyday archival interaction, emphasizing conversational access, grounding in retrieved documents, and transparent evidence for reading and verification.

Finally, device specific analyses characterize Apple Vision Pro as a mixed reality platform with video see-through, high-resolution displays, and hand- and eye-based input [5, 7]. Although not focused on archival retrieval, they motivate design choices relevant to our system, including speech driven queries, spatial information organization, and balancing focused reading with peripheral awareness. Our emphasis on grounding generated responses aligns with broader efforts to promote transparent AI-mediated interaction, with answers coupled with the related evidence which can help mitigate misinformation [1, 18].

3 Retrieval Augmented Generation on Digital Archives

Information Retrieval. Our system operates on the digital archives of the cantonal library of Lausanne, Switzerland, comprising newspapers and magazines spanning more than 300 years. An existing system (https://www.scriptorium.ch) provides keyword-based access but does not support semantic or conversational search. We adopt a retrieval augmented generation pipeline that enables free text queries. Documents are embedded in a semantic, multi-lingual, space using a large language model, and user queries are mapped into the same space. A k-nearest-neighbor search retrieves candidate documents, which are reranked to remove less relevant results. The remaining documents are passed to the LLM to generate an answer grounded in retrieved content. We further apply LLMs in the data processing phase to sanitize the raw OCR text, and our experiments show that this improves downstream retrieval performance (see Tables 1, 2, 3).

User Interaction. The interface leverages the gesture and augmented-reality capabilities of Apple Vision Pro. Interaction is primarily speech driven, with optional keyboard input, using on-device speech-to-text, eye gaze for selection, and hand gestures for clicking, sliding, and zooming. Lightweight on-device intent detection implemented with Apple's CoreML resolves on-device the commands intended for the interface (e.g., "filter results for year 1955" or "show map"), without invoking the remote RAG service.

The full pipeline of the system is illustrated in Figure 1 and can be broken down in the following stages.

(1) Speech capture and dictation, streaming partials to keep latency low.

OCR	k=1	k=5	k=10
Raw OCR	65.09	72.55	80.64
OCR with LLM	66.43	74.44	82.20

Table 1: Answer Faithfulness (%)

OCR	k=1	k=5	k=10
Raw OCR	36.06	43.63	48.75
OCR with LLM	38.92	45.03	48.92

Table 2: Answer Correctness (%)

OCR	k=1	k=5	k=10
Raw OCR	44.01	61.84	56.68
OCR with LLM	49.76	61.91	59.73

Table 3: Answer Relevancy (%)

(2) Intent identification of the query; if intended for the UI it is resolved directly, otherwise it is sent to the RAG service.
(3) Receiving the text query, it is mapped into an embedding space.
(4) kNN results are retrieved, reranked, and filtered
(5) Answer generation from the kNN results
(6) Packaging of evidence snippets, thumbnails, and metadata for spatial layout.

Combining RAG with spatial interfaces enables new interaction practices, such as treating retrieved evidence as manipulable objects that persist across conversational turns, rather than transient search results. Users can spatially cluster, reorder, and revisit evidence while refining queries through speech, supporting sensemaking as an embodied and cumulative process rather than a sequence of isolated searches.

Interface and Visualization. The interface provides a persistent search bar and supports English and French speech-to-text. Generated answers are presented alongside two visualization paradigms. A timeline-based evidence board arranges retrieved snippets with source metadata and supports expansion to full articles. In augmented reality mode, the map can be pinned to a physical surface, and geospatial references extracted from retrieved articles are used to place documents in space. Evidence items are represented as cuboids, whose height reflects the multitude of articles referencing a location (see Fig. 2).

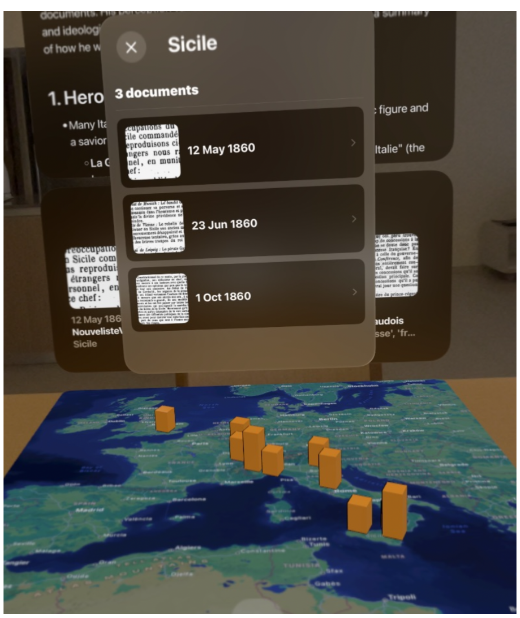

Figure 2: Interacting with an AR map anchored on the table

4 Experiments

We used a corpus of 145,000 articles from 16 newspapers and periodicals in the archival collection of the Lausanne Library, dating back to the eighteenth century. Embeddings were stored using a Chroma index. Although we experimented with a range of large language models, including OpenAI [17], Google Gemini [9], Qwen [4], and Mistral [14], as well as multiple OCR systems, namely Tesseract [20], EasyOCR [13], and Apple Vision OCR [3], space limitations prevent a full comparison. In this work, we therefore report results obtained using OpenAI GPT 4.0 as the backbone language model and Apple Vision OCR, provided by macOS, which achieved the best empirical performance. All experiments and the RAG service were executed on a Synology 923+ data server equipped with a 2 core 3.10 GHz CPU and 20 GB of RAM.

Answer Quality. We evaluate retrieval quality using an offline evaluation with RAGAS [8], which applies an LLM-as-a-judge to evaluate three information retrieval metrics. Faithfulness ($= \frac{|\text{Claims supported by context}|}{|\text{Total claims in the answer}|}$) measures factual consistency between the generated answer and retrieved context. Therefore, it checks if there are claims in the answer not supported by the documents retrieved, suggesting the the LLM used its own internal knowledge to answer parts of the question. Correctness combines semantic similarity and factual accuracy by comparing the generated answer to ground truth (Correctness $= w_1 \cdot$ (Semantic Similarity) + $w_2 \cdot$ (Factual Similarity)). Relevancy ($= \frac{1}{n}\sum_{i=1}^{n} \frac{q \cdot q_i}{\|q\|\|q_i\|}$) captures query alignment by generating n synthetic questions q_i from the answer and averaging their cosine similarity with the original query q.

We evaluate the proposed pipeline on 130 curated queries, with respect to LLM-refined OCR and retrieval depth $k \in \{1, 5, 10\}$. As shown in Tables 1–3, the LLM model benefits from increased retrieval depth and LLM-refined OCR, achieving peak correctness and faithfulness when retrieving more evidence. LLM-based OCR preprocessing consistently improves relevancy, highlighting the importance of semantic refinement for complex queries.

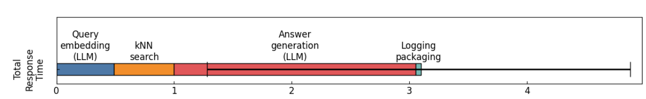

Figure 3: Mean Total Response Time (with std) and time spent across the subcomponents of the retrieval service

The system's temporal performance was evaluated across 130 queries to establish a baseline for real-time interaction. As illustrated in Figure 3, the end-to-end latency is primarily driven by the answer generation phase, followed by the kNN search, and Query embedding.

Figure 4: A query that requires an exact numeric answer

Figure 5: Using the AR-driven geospatial visualization

5 Demonstration

We review the three cases that we show in the demo video.

Query 1: Price of a Mazda 626 in year 2000. This numerical query retrieves a relevant advertisement through the RAG system and produces a precise price of 28,950 Swiss francs (Figure 4). The system allows users to drill down to the exact periodical issue that supports this result, thereby providing direct evidence for the provenance of the answer. For comparison, posing the same question to ChatGPT 5.1 yields an estimated price range of 18,000 to 25,000 francs based on web search results.

Query 2: *Fête des Vignerons* events in the 1900s. The Fête des Vignerons is a multi-day celebration honoring Swiss wine makers, held approximately every 25 years. The results for this query illustrate the usefulness of the timeline view, which highlights publications from 1905, 1927, 1955, 1977, and 1999, corresponding to the years in which the celebration took place.

Query 3: What do you know about Garibaldi before 1880? This question refers to Giuseppe Garibaldi's 1860 military campaign in Sicily, a pivotal episode in the unification of Italy. Newspapers from the 1860s present the events in contrasting tones, reflecting the political and ideological divisions of the period. The interface emphasizes these documents through a timeline displaying sources from May and June 1860, illustrating how perceptions evolved as Garibaldi advanced through southern Italy (Fig. 5). This is complemented by the geographic map view, which shows the Italian regions mentioned in the articles and helps relate journalistic coverage to the specific locations involved in the campaign.

6 Conclusion

Our demo combines modern information retrieval and RAG-based search with the interactive flexibility of immersive spatial computing on Apple Vision Pro, enabling the discovery of digital archives and the synthesis of answers. Beyond highlighting a novel user experience through new interface paradigms, the application also demonstrates the full workflow, from query to grounded answer. We hope that tools like this contribute to fact checking and help combat misinformation by making grounded historical evidence more accessible.

Acknowledgments

This work was supported by the Swiss National Science Foundation (SNSF), Switzerland, under Grant No. 237991.

References

[1] Esma Aïmeur, Sabrine Amri, and Gilles Brassard. 2023. Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining* 13, 1 (Feb 2023), 30. doi:10.1007/s13278-023-01028-5

[2] Muhammad Shahid Anwar, Jie Yang, Jaroslav Frnda, Ahyoung Choi, Nilufar Baghaei, and Miram Ali. 2025. Metaverse and XR for cultural heritage education: applications, standards, architecture, and technological insights for enhanced immersive experience. *Virtual Reality* 29, 2 (March 2025), 51. doi:10.1007/s10055-025-01126-z

[3] Apple Inc. 2023. Vision Framework: Text Recognition. Developer documentation. https://developer.apple.com/documentation/vision/recognizing_text_in_images Accessed 2025-12-21.

[4] Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wen Ge, Yu Han, Fei Huang, et al. 2023. Qwen Technical Report. arXiv:2309.16609 [cs.CL] https://arxiv.org/abs/2309.16609

[5] Verena Biener, Florian Jack Winston, Dieter Schmalstieg, and Alexander Plopski. 2025. Long-Term Experiences From Working with Extended Reality in the Wild. doi:10.48550/arXiv.2509.05067 arXiv:2509.05067.

[6] Giovanna Di Marzo Serugendo and Lamia Friha. 2025. AI-Powered User Experience and Interactions with Galleries, Libraries, Archives and Museum (GLAM). *Revue électronique suisse de science de l'information (RESSI)* , Spécial CIHN24 (April 2025), . doi:10.55790/journals/ressi.2025.e1809

[7] Jan Egger et al. 2024. Is the Apple Vision Pro the Ultimate Display? A First Perspective and Survey on Entering the Wonderland of Precision Medicine. *Journal of Imaging* 10, 8 (2024), 147. doi:10.2196/52785

[8] Shahul Es, Jithin James, Luis Espinosa-Anke, and Steven Schockaert. 2025. Ragas: Automated Evaluation of Retrieval Augmented Generation. arXiv:2309.15217 [cs.CL] https://arxiv.org/abs/2309.15217

[9] Gemini Team Google, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, et al. 2023. Gemini: A Family of Highly Capable Multimodal Models. arXiv:2312.11805 [cs.CL] https://arxiv.org/abs/2312.11805

[10] Srushti Goud, Valerio M. Salerno, Arianna Magnani, and Salvatore Sorce. 2025. Interaction and User Interface Features for Exploring Ancient Books and Manuscripts. In *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces (IUI '25 Companion)*. Association for Computing Machinery, New York, NY, USA, 70–73. doi:10.1145/3708557.3716349

[11] Yifan Hu. 2025. MuseRAG++: A Deep Retrieval-Augmented Generation Framework for Semantic Interaction and Multi-Modal Reasoning in Virtual Museums. Preprint. doi:10.21203/rs.3.rs-7281889/v1

[12] C. Innocente et al. 2023. A framework study on the use of immersive XR technologies in the cultural heritage domain. *Journal of Cultural Heritage* 63 (2023), 166–180. doi:10.1016/j.culher.2023.06.001

[13] JaidedAI. 2024. EasyOCR, project page. Project website. https://www.jaided.ai/easyocr/ Accessed 2025-12-21.

[14] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7B. arXiv:2310.06825 [cs.CL] https://arxiv.org/abs/2310.06825

[15] Diana Marques and Robert Costello. 2018. Concerns and Challenges Developing Mobile Augmented Reality Experiences for Museum Exhibitions. *Curator: The Museum Journal* 61, 4 (2018), 541–558. doi:10.1111/cura.12279

[16] Elisa Ayumi Masasi de Oliveira, Rafael Teixeira Sousa, Andressa Araújo Bastos, Luiza Martins de Freitas Cintra, and Arlindo Rodrigues Galvão Filho. 2025. Immersive Virtual Museums with Spatially-Aware Retrieval-Augmented Generation. In *Proceedings of the 2025 ACM International Conference on Interactive Media Experiences (IMX '25)*. Association for Computing Machinery, New York, NY, USA, 437–440. doi:10.1145/3706370.3731643

[17] OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, et al. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL] https://arxiv.org/abs/2303.08774

[18] Gordon Pennycook and David G. Rand. 2021. The psychology of fake news. *Trends in Cognitive Sciences* 25, 5 (2021), 388–402. doi:10.1016/j.tics.2021.02.007

[19] Maria Shehade and Theopisti Stylianou-Lambert. 2020. Virtual Reality in Museums: Exploring the Experiences of Museum Professionals. *Applied Sciences* 10, 11 (2020), 4031. doi:10.3390/app10114031

[20] Ray Smith. 2007. An Overview of the Tesseract OCR Engine. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*. IEEE Computer Society, Los Alamitos, CA, USA, 629–633. doi:10.1109/ICDAR.2007.4376991

[21] Christina Tsita, Maya Satratzemi, Alexandros Pedefoudas, Charalabos Georgiadis, Maria Zampeti, Evi Papavergou, Syrago Tsiara, Eleni Sismanidou, Petros Kyriakidis, Dionysios Kehagias, and Dimitrios Tzovaras. 2023. A Virtual Reality Museum to Reinforce the Interpretation of Cultural Heritage. *Heritage* 6, 5 (2023), 218. doi:10.3390/heritage6050218

[22] Isabelle Verhulst, Andy Woods, Laryssa Whittaker, James Bennett, and Polly Dalton. 2021. Do VR and AR versions of an immersive cultural experience engender different user experiences? A comparison of user experiences in Virtual Veronese. *Computers in Human Behavior* 125 (2021), 106943. doi:10.1016/j.chb.2021.106951

[23] Jian Yu, Zhan Wang, Yifan Cao, Hao Cui, and Wei Zeng. 2025. Centennial Drama Reimagined: An Immersive Experience of Intangible Cultural Heritage through Contextual Storytelling in Virtual Reality. *Journal on Computing and Cultural Heritage* 18, 1, Article 11 (2025), 22 pages. doi:10.1145/3705613

[24] Tuo Zhang, Yuechun Sun, and Ruiliang Liu. 2025. Provenance Analysis of Archaeological Artifacts via Multimodal RAG Systems. arXiv:2509.20769 [cs.IR] https://arxiv.org/abs/2509.20769