From Answer to Audio: Visual Traceability for AI-Assisted Meeting Retrieval

Takuya Nakata
Center for Mathematical and Data Sciences
Kobe University
Kobe, Japan
tnakata@bear.kobe-u.ac.jp

Masahide Nakamura
Center for Mathematical and Data Sciences
Kobe University
Kobe, Japan
masa-n@cmds.kobe-u.ac.jp

Abstract

We present a visual traceability interface that enables users to verify AI-generated answers by navigating from natural-language outputs to original meeting audio. Even with textual evidence, RAG users still struggle to verify whether answers come from specific utterances or from transcription or summarization errors. Our system introduces a four-level interaction model connecting (1) AI-generated answers, (2) contributing meetings, (3) summarized meeting chunks, and (4) time-aligned utterance-level audio. The interface lets users traverse this evidence chain and directly play the corresponding audio segments. We demonstrate how audio-grounded traceability supports answer verification, error detection, and trust calibration in AI-assisted meeting analysis.

CCS Concepts

• **Human-centered computing** → **User interface management systems**; *Auditory feedback*; • **Computing methodologies** → Information extraction.

Keywords

Retrieval-Augmented Generation, Meeting Retrieval, Human–AI Interaction, AI Traceability, Audio-grounded Evidence

ACM Reference Format:
Takuya Nakata and Masahide Nakamura. 2026. From Answer to Audio: Visual Traceability for AI-Assisted Meeting Retrieval. In *Proceedings of the 2026 International Conference on Advanced Visual Interfaces (AVI '26), June 08–12, 2026, Venice, Italy*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3811427.3811496

1 Introduction

Meeting retrieval systems increasingly rely on retrieval-augmented generation (RAG) to answer questions across recorded meetings [1, 5, 9]. Recent work on open-domain question answering also highlights the importance of providing explicit evidence to support generated answers [3]. However, users struggle to verify whether AI-generated answers truly originate from specific utterances. Even when textual evidence is shown, transcription, summarization, or retrieval errors may obscure the connection between answers and the original meeting [6]. From a human–computer interaction (HCI) perspective, this creates a gap between AI capability and user trust, as users often find it difficult to calibrate their expectations of AI systems when system reasoning is not transparent [4]. Rather than treating traceability purely as an algorithmic problem in RAG pipelines [2, 10], we approach it as an interaction design challenge. We present an interface that allows users to navigate from AI-generated answers back to original meeting audio through multiple evidence levels. This demonstration contributes:

- an interactive traceability interface linking answers, meetings, chunks, and utterance-level audio,
- a navigation mechanism that exposes the evidence structure underlying RAG-based meeting retrieval, and
- an audio-grounded verification workflow that allows users to inspect AI outputs using original meeting audio.

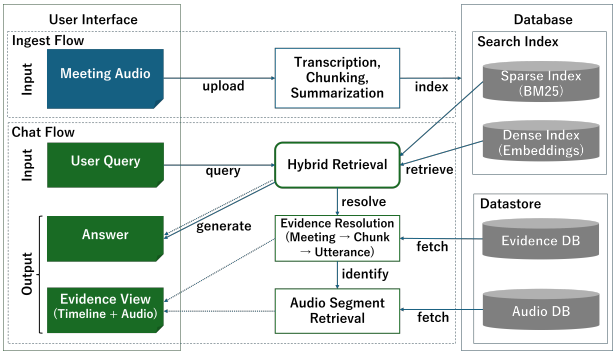

Figure 1: System architecture for meeting retrieval with audio-grounded traceability. Meeting audio is transcribed and chunked for hybrid retrieval, producing answers and evidence references linking utterances to time-aligned audio segments.

2 System Overview

2.1 Architecture

Figure 1 shows the system architecture. Meeting audio is transcribed into utterances and grouped into chunks for retrieval. Standard hybrid sparse–dense retrieval gets relevant chunks and generates answers with structured evidence. Each utterance is linked to a time-aligned audio segment for playback and verification. The main contribution is the four-level traceability interface and audio-grounded verification workflow built on this standard backend.

AVI '26, Venice, Italy

ACM ISBN 979-8-4007-2342-1/26/06
https://doi.org/10.1145/3811427.3811496

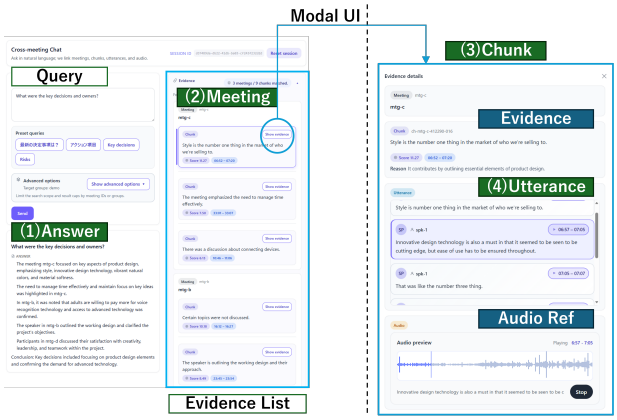

Figure 2: Interactive interface for tracing AI-generated answers back to meeting audio. Users navigate from (0) query to (1) answer, (2) meetings, (3) chunks, and (4) utterance-level audio evidence.

2.2 Four-Level Traceability Model

Figure 2 shows the core interaction model of the interface, which consists of four levels:

(1) Answer: A natural-language response generated by the system.
(2) Meeting: A list of meetings that contributed to the answer, each with a brief relevance explanation.
(3) Chunk: Summarized segments within each meeting used to construct the answer.
(4) Utterance and Audio: Individual utterances with timestamps that can be played directly.

These levels move from abstract to granular evidence, letting users inspect provenance from summaries to raw audio as needed. This design addresses challenges associated with long-context reasoning and the loss of intermediate evidence in retrieval-augmented systems [12].

3 Interaction Design and Demonstration

This work is presented as an interactive demonstration. Instead of quantitatively evaluating retrieval accuracy or answer quality, the demo lets attendees experience how answers can be inspected and verified through audio-grounded evidence.

3.1 Interface

The interface uses a fixed two-column layout: queries and answers on the left, and hierarchical evidence cards on the right for side-by-side comparison. Meeting cards show titles and relevance explanations, chunk cards provide summaries and expandable utterances, and utterances include timestamps and playback controls. This card hierarchy and progressive disclosure guide users from answer summaries to primary audio while preserving query context. Visualization approaches have been widely explored to expose the internal reasoning or attention structures of neural models and improve interpretability of AI systems [11]. The demonstration highlights several interaction capabilities:

- Verifying answers through direct audio evidence
- Detecting transcription or summarization errors
- Navigating evidence across multiple abstraction levels

Rather than presenting a single "correct" answer, the interface supports interactive exploration of meeting content. These capabilities support trust calibration by allowing users to decide when to rely on AI-generated answers and when to consult primary audio evidence. This perspective aligns with HCI research on inspectable memory structures and transparent interaction for conversational agents [8].

3.2 Demonstration

In a typical scenario, a user asks a question such as "What risks were discussed regarding option A?" The system retrieves relevant meetings and generates an answer. The evidence panel shows contributing meetings and chunks. Users can expand chunks and play utterance-level audio to verify the answer. Attendees can try different queries and observe how the evidence trail changes. Precomputed indices ensure responsive interaction during the demo.

4 Limitations and Discussion

Traceability improves transparency but does not guarantee correctness: users must still interpret how utterances support claims, and multi-level navigation may increase cognitive load for complex queries. The current card-based visual design is also relatively conventional, and future work should explore richer evidence overviews, visual encodings, and adaptive interfaces while preserving access to primary audio.

5 Conclusion

This demonstration presented an interactive traceability interface that enables users to move from AI-generated answers to original meeting audio. By framing traceability as an interaction design problem rather than a purely algorithmic one, the system allows users to verify and explore AI-assisted meeting retrieval. This contribution complements recent advances in conversational QA and RAG models [7] by focusing on user-centered inspection and verification rather than model accuracy. We believe that enabling users to move from answer to audio offers a practical way to engage with AI outputs and supports more transparent and trustworthy intelligent interfaces for meeting analysis.

6 GenAI Usage Disclosure

LLMs were used to assist with drafting and editing the manuscript and developing the demonstration system. All design decisions and interpretations were determined by the authors.

Acknowledgments

This research was partially supported by JSPS KAKENHI Grant Numbers JP25H01167, JP25K02946, JP25K24389, JP24K02765, JP24K02774, JP23K17006, JP23K28091, JP23K28383.

References

[1] Sartaj Bhuvaji, Prachitee Chouhan, Madhuroopa Irukulla, Jay Singhvi, Wan D. Bae, and Shayma Alkobaisi. 2025. A Retrieval-Augmented Framework For

Meeting Insight Extraction. In *Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing* (Catania International Airport, Catania, Italy) *(SAC '25)*. Association for Computing Machinery, New York, NY, USA, 899–906. doi:10.1145/3672608.3707915

[2] Juan Chen, Baolong Bi, Wei Zhang, Jingyan Sui, Xiaofei Zhu, Yuanzhuo Wang, Lingrui Mei, and Shenghua Liu. 2025. Rethinking All Evidence: Enhancing Trustworthy Retrieval-Augmented Generation via Conflict-Driven Summarization. arXiv:2507.01281 [cs.CL] https://arxiv.org/abs/2507.01281

[3] Parastoo Jafarzadeh and Faezeh Ensan. 2025. An evidence-based approach for open-domain question answering. *Knowledge and Information Systems* 67, 2 (01 Feb 2025), 1969–1991. doi:10.1007/s10115-024-02269-2

[4] Rafal Kocielnik, Saleema Amershi, and Paul N. Bennett. 2019. Will You Accept an Imperfect AI? Exploring Designs for Adjusting End-user Expectations of AI Systems. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3290605.3300641

[5] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Vancouver, BC, Canada) *(NIPS '20)*. Curran Associates Inc., Red Hook, NY, USA, Article 793, 16 pages.

[6] Xingxian Liu, Bin Duan, Bo Xiao, and Yajing Xu. 2023. Query-Utterance Attention With Joint Modeuing For Query-Focused Meeting Summarization. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1–5. doi:10.1109/ICASSP49357.2023.10096486

[7] Zihan Liu, Wei Ping, Rajarshi Roy, Peng Xu, Chankyu Lee, M. Shoeybi, and Bryan Catanzaro. 2024. ChatQA: Surpassing GPT-4 on Conversational QA and RAG. *Advances in Neural Information Processing Systems 37* (2024). doi:10.52202/079017-0493

[8] Zhuoshi Pan, Qianhui Wu, Huiqiang Jiang, Xufang Luo, Hao Cheng, Dongsheng Li, Yuqing Yang, Chin-Yew Lin, H. Vicky Zhao, Lili Qiu, and Jianfeng Gao. 2025. On Memory Construction and Retrieval for Personalized Conversational Agents. arXiv:2502.05589 [cs.CL] https://arxiv.org/abs/2502.05589

[9] Virgile Rennard, Guokan Shang, Julie Hunter, and Michalis Vazirgiannis. 2023. Abstractive Meeting Summarization: A Survey. *Transactions of the Association for Computational Linguistics* 11 (07 2023), 861–884. doi:10.1162/tacl_a_00578

[10] Nirmal Roy, Leonardo F. R. Ribeiro, Rexhina Blloshmi, and Kevin Small. 2024. Learning When to Retrieve, What to Rewrite, and How to Respond in Conversational QA. arXiv:2409.15515 [cs.CL] https://arxiv.org/abs/2409.15515

[11] Jesse Vig. 2019. A Multiscale Visualization of Attention in the Transformer Model. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Marta R. Costa-jussà and Enrique Alfonseca (Eds.). Association for Computational Linguistics, Florence, Italy, 37–42. doi:10.18653/v1/P19-3007

[12] Gongbo Zhang, Zihan Xu, Qiao Jin, Fangyi Chen, Yilu Fang, Yi Liu, Justin F. Rousseau, Ziyang Xu, Zhiyong Lu, Chunhua Weng, and Yifan Peng. 2025. Leveraging long context in retrieval augmented language models for medical question answering. *npj Digital Medicine* 8, 1 (02 May 2025), 239. doi:10.1038/s41746-025-01651-w