Bridging the Expertise Gap: An AI-Assisted Workflow for Qualitative Analysis of Developer Meeting Transcripts in NSF ATE Program Evaluation

SHALEE HODGSON[1*], KATE ROTINDO[1], DAVID C. ANDERSON[2]

[1]*Research and Evaluation, Impact Allies, Vero Beach, FL 32960, USA*

[2] *KonnectXR (KXR), Impact Allies, Vero Beach, FL 32960, USA*

*shalee.hodgson@impactallies.com

Abstract: External evaluators of NSF Advanced Technological Education (ATE) projects are often asked to document technical progress in fields where they are not the experts. This article describes a replicable, artificial intelligence (AI)-assisted workflow that we developed during the Year 2 evaluation of KonnectXR (KXR): Extended Reality in Technician Education (NSF Award Nos. 2435348-2435353) to address that problem. The workflow applies a structured, large language model (LLM)-assisted reading of developer meeting transcripts, a source that records technical milestones and decisions as they happen. Using it, we documented the development team's progress more accurately and relied less on our own technical expertise. The resulting evidence fed directly into NSF goal-level reporting. We close with what the approach means for evaluators of technically complex ATE projects.

Keywords: program evaluation, NSF ATE, AI-assisted analysis, large language models, qualitative methods, meeting transcripts

Materials and Methods

NSF Advanced Technological Education (ATE) evaluations are expected to document technical progress with rigor and accuracy [1], [2],[3]. External evaluators, however, are usually generalists. They bring real expertise in evaluation methodology, but not necessarily in the technical domains their projects work in. This particular case includes topics such as authentication, learning management system (LMS) integration, and extended reality (XR) development. While the evaluation team could interview the principal investigator (PI) and review documents, these rely on curated, after-the-fact accounts, while the development team's meeting transcripts capture the work as it happens. This article describes a structured, large language model (LLM)-assisted workflow for turning those long, technical transcripts into evidence that an evaluation can use.

We developed the workflow during the Year 2 evaluation of KonnectXR (KXR): Extended Reality in Technician Education, a three-year NSF ATE collaborative project (Award Nos. 2435348-2435353) that Impact Allies (IA) leads with six partner institutions. KXR is building an XR platform for technician education, with technical work in authentication, interoperability, networking, and instructional design. Extended reality has shown growing value in technician training, even as real barriers to implementation remain [4],[5].

KXR is organized around four primary goals: integrate single sign-on (SSO) with Family Educational Rights and Privacy Act (FERPA)-compliant encryption and Learning Tools Interoperability (LTI) 1.3 across major LMS platforms (Goal 1); develop a multi-user networking framework that supports at least

50 simultaneous users (Goal 2); create six XR lab experiences grounded in Universal Design for Learning (UDL) principles, with tutorials and instructor resources (Goal 3); and conduct user interface and user experience (UI/UX) research with 30 instructors and 100 students (Goal 4). Since the technical development work is concentrated in the first three goals, those are the ones this workflow helped us evaluate.

The development team meets every other week and records its meetings with Read AI, an AI-powered assistant that produces timestamped, speaker-labeled transcripts and short summaries. Across Year 2, this produced more than 20 transcripts spanning August 2025 through June 2026. Our team works in program evaluation and education research rather than software engineering, and we had come to see these transcripts as a high-value but underused record of progress against the four goals.

To enhance our work, we built the workflow iteratively over the Year 2 evaluation. It has seven steps that run across three phases: collection and preparation, AI-assisted analysis, and evaluation integration (Figure 1). The seven steps read left to right: Collection and Preparation (Steps 1-3), AI-Assisted Analysis (Steps 4-5), and Evaluation Integration (Steps 6-7). The middle phase uses a large language model; the first and third are manual steps by the evaluation team. SME = subject matter expert; LLM = large language model.

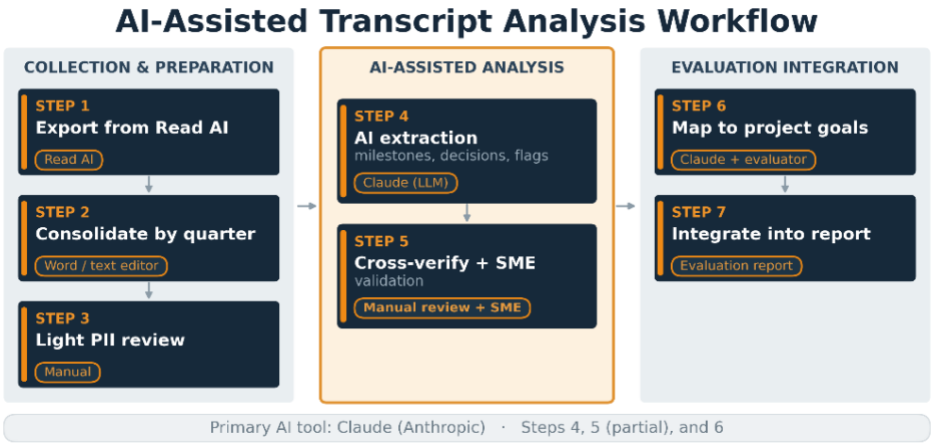

Figure 1. AI-Assisted Transcript Analysis Workflow

Step 1–3: Transcript Collection and Preparation

The evaluator exported transcripts from Read AI as .txt files and grouped them into quarterly bundles in chronological order. Before analysis, we reviewed each one to remove personally identifiable information (PII). In this case, participant names became role descriptors, for example "lead developer" or "PI." Full de-identification is completed before any transcript content is included in a published deliverable.

Step 4: AI-Assisted Extraction

We uploaded the bundles to Claude (Anthropic), a large language model with strong document analysis capabilities. Later in the process, we used Claude Cowork to set up a file that could be extracted, which improved the workflow. Once Claude had the files, we used a structured prompt that asked it to extract technical milestones, challenges or blockers, decisions and their rationale, team roles, partnership developments, and evaluation-relevant data, such as pending approvals. We also asked the model to mark anything it was not confident about and to keep confirmed progress apart from open items. The reality check for us as evaluators lived in the KXR dashboard and project documentation. This information held the goals and deliverables to ensure we were checking against what was real [6],[7].

This step gives us an account of what the developers said. It does not give us independent technical claims, and we treated that line as a hard one. Our job was not to judge whether the networking architecture was sound; it was to check that the extraction matched the transcript, and evaluators already know how to do that [6],[8].

Step 5: Cross-Verification
As we developed the process and the output, we worked to ensure every AI summary was verified against the source transcripts. We flagged anything that overstated certainty, dropped context, or put a statement in the wrong person's mouth. This cross-checking is something we do on any evaluation, but the technical judgment is not. This is why sending the summaries to the subject matter experts (SMEs) was a critical step. The split was practical: we confirmed the summary matched the transcript, and the SMEs confirmed the technical meaning. Skipping this step was never on the table; LLM hallucination rates are documented well enough [9],[10],[11].

Step 6–7: Goal Mapping and Report Integration
The last two steps put the verified findings to work. We mapped them to the project's goals and objectives, with the KXR dashboard and prior reporting as another source of verification. We then wrote the synthesis into the Year 2 evaluation report, as narrative evidence and as a cited methodology source [12].

Results

Once the workflow was developed and tested, we used it to analyze the full Year 2 transcript set (August 2025–June 2026). The output was a technically accurate analysis of progress across the planned KXR objectives and activities, helping us address a key evaluation question about grant deliverables and progress. Some of what the transcripts held, though, was not available to us anywhere else.

The clearest example was the progress on the 2D desktop client and its role in the project. The 2D work started as a way to support accessibility. However, somewhere in Year 2 it became a core instructional tool. It now shares an application programming interface (API) layer with the virtual reality (VR) client, allowing an instructor to run a session without a headset. Nobody framed it that way for us in an interview. We read it in the developers' own back-and-forth, and it changed how we wrote about Goals 2 and 3. The transcripts also confirmed several project risks.

- Working LTI connection with Canvas in February 2026 and D2L Brightspace approval delays.
- Personnel changes that had not been formally documented.
- Early FERPA mapping and Voluntary Product Accessibility Template (VPAT) work.
- A pedagogical question the team was starting to wrestle with about VR versus 2D simulation.

Each of these touched indicators under Goals 1 through 3.

Cross-verification paid for itself on the D2L Brightspace integration. The AI's first summary treated the MnState approval as routine administrative work, nearly done, and we read it the same way at first. The transcripts told us something different. The reality was that the team was describing a structured security review at MnState that was expert-supported, outside the project's control, and directly tied to Goal 1's interoperability work. With that new context, we were able to revise the indicator, going from a box about to be checked to an active external dependency needing specialized coordination. Neither the dashboard nor a PI interview would have caught this, and as evaluators we were working outside our technical depth. Without the verification step, the error would have been ours. The PI's read on the method was simple: the documentation matched what the technical team had actually done. This lessened the back-and-forth and the challenges of documenting it correctly in the evaluation report.

For this work, Claude (Anthropic) handled extraction, synthesis, and goal mapping; we used Perplexity separately for citation research and kept it apart from the transcript analysis [13]. The workflow is a targeted method, not a universal one. The cross-verification step cannot be skipped, and it is worth being clear that transcript analysis documents what developers said, not an independent technical validation of it.

Acknowledgments. This work was supported by the National Science Foundation Advanced Technological Education program (Award Nos. 2435348–2435353). The authors thank the KXR project

team and members of partner institutions whose discussions informed this work. The views expressed are those of the authors and do not necessarily reflect those of NSF.

Disclosures. AI tools: Claude (Anthropic) was used as a research instrument in the evaluation workflow described in this article, as detailed in the Materials and Methods section. Claude was also used to assist with manuscript drafting and structural organization. All content was reviewed, verified, and revised by the human authors. No AI tool is listed as an author. This disclosure is consistent with emerging standards for the use of AI in scholarly publications [14]. Conflicts of interest: Reference [5] includes David C. Anderson as a co-author; this citation is included based on scholarly merit and its relevance to the project context described herein. The authors declare no other conflicts of interest.

References

[1] National Science Foundation, “Advanced Technological Education (ATE),” *NSF - U.S. National Science Foundation*, 2024. https://www.nsf.gov/funding/opportunities/ate-advanced-technological-education/nsf24-584/solicitation

[2] D. M. Hata, "Value-Creation Evaluation Framework for Evaluating NSF Advanced Technological Education Projects," *J. Adv. Technol. Educ*., vol. 3, no. 2, pp. 158–171, 2024. DOI: 10.5281/zenodo.13367680.

[3] M. López, L. Wilson Becho, V. Marshall, and B. Hooks Singletary, “The State of Evaluation in the ATE Program 2025 - EvaluATE,” *EvaluATE*, 2025. https://evalu-ate.org/miscellaneous/2025-state-ate-eval/ (accessed Mar. 22, 2026).

[4] B. Boland, "Extended Reality Vocational Training’s Ability to Improve Soft Skills Development and Increase Equity in the Workforce," *AI, Comput. Sci. Robot. Technol*., vol. 2023, no. 2, Art. no. 22, 2023. DOI: 10.5772/acrt.22.

[5] B. Upadhyay, K. C. Madathil, S. Hegde, D. Anderson, E. Wooldridge, D. Presley, L. Perez, and B. Reid, "Barriers Toward the Implementation of Extended Reality (XR) Technologies to Support Education and Training in Workforce Development Programs," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 68, no. 1, pp. 265–269, Aug. 2024. DOI: 10.1177/10711813241275080.

[6] R. H. Tai, L. R. Bentley, X. Xia, J. M. Sitt, S. C. Fankhauser, A. M. Chicas-Mosier, and B. G. Monteith, "An Examination of the Use of Large Language Models to Aid Analysis of Textual Data," *Int. J. Qualitative Methods*, vol. 23, 2024. DOI: 10.1177/16094069241231168.

[7] A. Vijayan, "A Prompt Engineering Approach for Structured Data Extraction from Unstructured Text Using Conversational LLMs," in *Proc. 2023 6th Int. Conf. on Algorithms, Computing and Artificial Intelligence* (ACAI), 2024. DOI: 10.1145/3639631.3639663.

[8] S. Qiao, X. Fang, J. Wang, R. Zhang, X. Li, and Y. Kang, "Generative AI for thematic analysis in a maternal health study: coding semistructured interviews using large language models," *Appl. Psychol. Health Well-Being*, vol. 17, no. 3, Art. no. e70038, 2025. DOI: 10.1111/aphw.70038.

[9] R. T. Williams, "Paradigm shifts: exploring AI’s influence on qualitative inquiry and analysis," *Frontiers in Research Metrics and Analytics*, vol. 9, Art. no. 1331589, 2024. DOI: 10.3389/frma.2024.1331589.

[10] N. Shafran et al., "Hallucination Rates and Reference Accuracy of ChatGPT and Bard in Systematic Reviews," *J. Med. Internet Res*., vol. 26, Art. no. e53164, 2024. DOI: 10.2196/53164.

[11] J. Kantor, "Best practices for implementing ChatGPT, large language models, and artificial intelligence in qualitative and survey-based research," *JAAD Int*., vol. 14, pp. 22–23, Mar. 2024. DOI: 10.1016/j.jdin.2023.10.001.

[12] J. W. Creswell and C. N. Poth, *Qualitative Inquiry and Research Design: Choosing Among Five Approaches,* 4th ed. Thousand Oaks, CA: SAGE, 2018, ch. 7–8.

[13] A. J. Nashwan and H. Abukhadijah, "Harnessing Artificial Intelligence for Qualitative and Mixed Methods in Nursing Research," *Cureus*, vol. 15, no. 11, Art. no. e48570, 2023. DOI: 10.7759/cureus.48570.

[14] COPE Council, “Authorship and AI tools,” *COPE: Committee on Publication Ethics*, Dec. 18, 2024. https://doi.org/10.24318/cCVRZBms