Designing an LLM-Supported Qualitative Analysis Tool for Iterative Inquiry

Anna Zhelizniak
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
azhelizn@uwaterloo.ca

James R Wallace
School of Public Health Sciences
University of Waterloo
Waterloo, Ontario, Canada
james.wallace@uwaterloo.ca

Abstract

We present a demonstration of an LLM-assisted qualitative analysis system that supports iterative, data-grounded interpretation. The system integrates model feedback directly into a document-centered workflow, allowing researchers to write and refine interpretations while receiving structured evaluations grounded in reference datasets. Rather than generating analytical outputs, the system evaluates user-written text, highlighting both supported and potentially misleading interpretations. This design is intended to encourage critical engagement and position the model as a reflective assistant within the analytical process.

Keywords

qualitative analysis, large language models, human-AI collaboration, interaction design, thematic analysis, iterative thematic inquiry

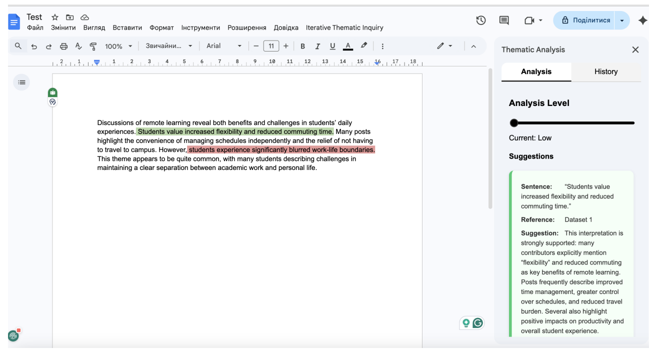


Figure 1: System interface integrated into Google Docs. Interpretations are written directly in the document, with supported (green) and potentially misleading (red) statements highlighted inline. Detailed feedback is provided in the side panel.

1 Background

Qualitative analysis helps researchers understand complex human experiences by working closely with text and building interpretations over time. Methods such as thematic analysis and grounded theory involve reading data repeatedly, writing down ideas, and gradually refining them [1, 2]. This process is inherently iterative: researchers move back and forth between data and interpretation, adjusting their understanding as they go. While this is what makes qualitative work meaningful, it also makes it slow and difficult to scale, especially when dealing with large datasets.

Recent work has explored how large language models (LLMs) can support this process, for example by generating codes, summaries, or themes [3, 5]. These approaches can save time and help researchers get started more quickly. At the same time, they introduce a new challenge: model outputs can be easy to accept at face value. Because they are often fluent and confident, they may discourage researchers from questioning how well the results are actually grounded in the data.

In this work, we take a different approach. Instead of asking the model to generate interpretations, we use it to evaluate interpretations written by the researcher. The system checks how well a statement is supported by the data and highlights both strong and weaker cases. This shifts the role of the model from producing answers to supporting reflection. This design is inspired by Iterative Thematic Inquiry (ITI), where researchers externalize interpretations early and continuously revisit them as their understanding develops [4]. Rather than treating interpretation as an outcome, ITI frames analysis as an ongoing loop of writing, reflection, and revision. Our system uses this idea by embedding evaluation directly into the writing process itself.

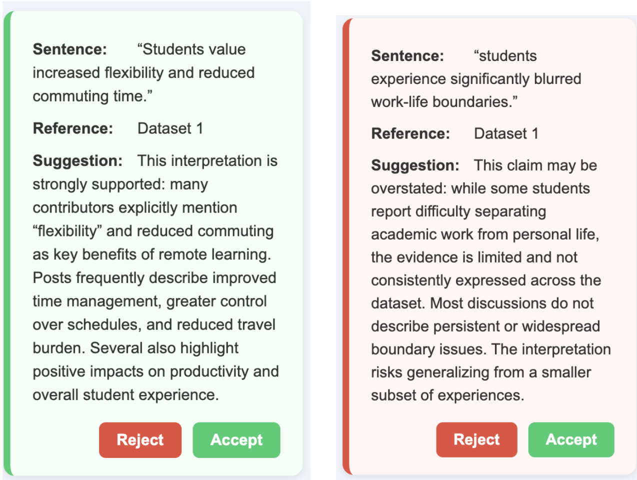
2 System Overview

The system is designed as a lightweight layer on top of a familiar writing environment, implemented as a Google Docs add-on. Instead of introducing a separate analysis tool, it embeds support directly into the document where interpretations are written. This allows researchers to stay within their natural workflow, writing and revising interpretations while receiving feedback in context.

Figure 1 shows the interface. The main document area contains the researcher's writing, where interpretations are expressed in natural language. The system analyzes this text continuously, highlighting segments that require attention.

The interaction follows a simple but structured loop. The researcher writes an interpretation based on the data, and the system evaluates how well that interpretation is supported by the reference dataset. Feedback is then presented directly in the document, prompting the researcher to review, accept, or revise the statement. This creates a continuous cycle of writing, evaluation, and refinement.

The system provides two main types of feedback: supported interpretations and potentially misleading interpretations. Supported interpretations are marked in green and indicate that the claim is consistently reflected in the data. Potentially misleading interpretations are marked in red and indicate that the claim may rely on



(a) Supported interpretation. **(b) Potentially misleading interpretation.**

Figure 2: Examples of suggestion cards shown in the side panel. The system provides both supportive and critical feedback, encouraging the researcher to validate, refine, or reconsider interpretations instead of passively accepting model output.

limited evidence or overgeneralize from a smaller set of observations.

Figure 2 shows examples of suggestion cards in the side panel. Each card includes the sentence under review, the reference dataset, and a short explanation of the system's reasoning. These explanations are designed to be readable and grounded in the data, helping the researcher understand not just what the system suggests, but why.

Each suggestion requires an explicit decision. The researcher can accept or reject the feedback, which introduces a deliberate moment of reflection rather than allowing suggestions to be passively incorporated into the text. This interaction design is central to the system: instead of automating analysis, it structures how the researcher engages with their own interpretations.

The system supports two modes of interaction. In the automated mode, suggestions are generated continuously as the researcher writes, appearing as hints in the document. In the manual mode, the researcher can select a specific passage and request evaluation explicitly. These modes allow flexibility in how much guidance the researcher wants at different stages of analysis.

To illustrate the workflow, consider a researcher analyzing student discussions. They write two interpretations: "Students value increased flexibility and reduced commuting time" and "Students experience significantly blurred work-life boundaries." The system confirms the first as supported and flags the second as potentially overstated. After reviewing the feedback, the researcher accepts both suggestions, recognizing the strength of the first claim and the limitations of the second. This does not immediately change the text, but informs the next step, where the researcher may revise the second interpretation to better reflect the data.

Overall, the system is intentionally simple. It focuses on a small set of interactions (writing, evaluating, and revising) but structures them in a way that keeps the researcher engaged with the data. By embedding evaluation directly into the writing process, the system is designed to support iterative interpretation without taking control of it.

3 Design Goals

The system was designed around several interaction goals motivated by challenges in LLM-supported qualitative analysis.

Researcher control. The system positions the user as the primary author of analytical statements. Model outputs are presented as suggestions that can be reviewed, accepted, or ignored.

Critical engagement with suggestions. The interaction design encourages users to actively examine model feedback instead of directly incorporating it into the document. Suggestions require explicit review through accept/reject actions.

Grounding in data. System feedback is linked to reference datasets and accompanied by short explanations, allowing interpretations to remain connected to the underlying material.

Iterative refinement. Inspired by Iterative Thematic Inquiry (ITI), the workflow supports repeated revisiting and revision of interpretations as the analysis develops.

Lightweight interaction. The system is embedded directly into Google Docs and focuses on a small set of interactions in order to minimize disruption to existing writing workflows.

4 Conclusion and Future Work

We presented an LLM-assisted qualitative analysis system that integrates evaluation directly into the writing process. Future work will include user studies exploring how researchers engage with different forms of model feedback, particularly how interaction design shapes reflection, revision, and reliance on system suggestions during qualitative analysis.

References

[1] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. doi:10.1191/1478088706qp063oa

[2] Kathy Charmaz. 2014. *Constructing Grounded Theory* (2 ed.). SAGE Publications, London, UK.

[3] Tim Fischer and Chris Biemann. 2024. Exploring large language models for qualitative data analysis. In *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*. Association for Computational Linguistics, Miami, USA, 423–437. doi:10.18653/v1/2024.nlp4dh-1.41

[4] David L. Morgan and Andreea Nica. 2020. Iterative thematic inquiry: A new method for analyzing qualitative data. *International Journal of Qualitative Methods* 19 (2020), 1–11. doi:10.1177/1609406920955118

[5] Ansh Sharma and James R Wallace. 2025. DeTAILS: Deep Thematic Analysis with Iterative LLM Support. In *Proceedings of the 7th ACM Conference on Conversational User Interfaces (CUI '25)*. Association for Computing Machinery, New York, NY, USA, Article 28, 7 pages. doi:10.1145/3719160.3735657