Pages:67-73
Volume-VI, Issue-II (2025)
https://globalknowledgeacademy.com/

Large Language Models for Qualitative Data Analysis

Zunaira Rafaqat
Chenab institute of Information technology, Pakistan
Corresponding E-mail: zunaira.rafaqat@cgc.edu.pk

Abstract

Qualitative data analysis has long relied on human interpretation, thematic coding, and iterative sense-making, making it both intellectually rich and methodologically challenging. The rapid emergence of Large Language Models (LLMs) introduces a new paradigm for analyzing unstructured qualitative data at scale while preserving contextual depth. This paper investigates the role of LLMs in qualitative data analysis, examining their capacity to perform tasks such as thematic extraction, coding consistency, sentiment interpretation, and narrative synthesis. Through controlled experiments on interview transcripts, open-ended survey responses, and textual field notes, we evaluate the analytical performance of LLMs against traditional qualitative analysis methods. Results demonstrate that LLMs significantly enhance efficiency and reproducibility while maintaining high semantic fidelity. However, limitations related to interpretability, bias propagation, and epistemic transparency remain. This study positions LLMs not as replacements for qualitative researchers, but as intelligent analytical collaborators that reshape how qualitative inquiry is conducted in data-intensive research environments.

Keywords: Large Language Models, Qualitative Data Analysis, Thematic Coding, Natural Language Processing, Human-AI Collaboration, Interpretive Analytics

I. Introduction

Qualitative research plays a foundational role in understanding complex human behaviors, perceptions, and social phenomena that resist quantification [1]. Unlike numerical data, qualitative data is inherently contextual, subjective, and meaning-laden, requiring interpretive frameworks that evolve throughout the research process. Traditional qualitative analysis methods—such as manual coding, thematic analysis, and grounded theory—are deeply rigorous but also time-intensive and difficult to scale, particularly in modern data-rich environments.

In recent years, the proliferation of unstructured textual data from interviews, surveys, social media, and digital ethnographies has placed unprecedented pressure on qualitative researchers. While qualitative software tools have improved data management, they still rely heavily on human labor for interpretation and synthesis [2]. This growing gap between data availability and analytical capacity raises an important methodological question: how can researchers maintain interpretive depth while scaling qualitative analysis?

Large Language Models represent a significant technological shift in how text can be processed and understood. Trained on vast corpora of human language, LLMs demonstrate an emergent ability to recognize themes, infer intent, summarize narratives, and generate coherent interpretations across diverse domains. These capabilities suggest that LLMs may serve as powerful analytical tools for qualitative research, extending human capacity rather than replacing it.

However, applying LLMs to qualitative analysis is not merely a technical challenge—it is a methodological one [3]. Qualitative research is not only about identifying patterns but also about reflexivity, positionality, and interpretive judgment. This raises critical concerns regarding transparency, bias, epistemic authority, and the role of human insight in AI-assisted analysis.

This paper explores how LLMs can be systematically integrated into qualitative data analysis workflows. By examining their analytical performance, limitations, and alignment with qualitative research principles, we aim to provide a balanced and empirically grounded assessment of their value [4]. Rather than framing LLMs as autonomous analysts, this work conceptualizes them as cognitive tools that reshape the practice of qualitative inquiry.

II. Background and Related Work

The evolution of qualitative data analysis has been closely tied to methodological traditions in the social sciences, including phenomenology, grounded theory, and narrative inquiry. These approaches emphasize inductive reasoning, iterative interpretation, and deep engagement with data. Historically, qualitative rigor has been achieved through prolonged immersion, triangulation, and reflexive documentation rather than automation [5].

Early computational approaches to qualitative analysis focused on keyword frequency, lexical analysis, and basic topic modeling. While useful for exploratory insights, these methods lacked the semantic understanding required for nuanced interpretation. The limitations of bag-of-words models and shallow statistical techniques highlighted the need for more context-aware language processing tools.

Advances in natural language processing, particularly transformer-based architectures, marked a turning point in text analytics. Unlike earlier models, LLMs capture long-range dependencies, pragmatic cues, and latent semantic relationships [6]. This allows them to move beyond surface-level text analysis toward more interpretive forms of understanding that align with qualitative objectives.

Recent studies have explored LLMs for tasks such as sentiment analysis, discourse classification, and narrative summarization [7]. However, much of this work remains task-oriented and quantitative in evaluation, often neglecting the epistemological foundations of qualitative research. There is limited empirical work that directly compares LLM-assisted qualitative analysis with human-led thematic coding.

This study builds upon prior work by explicitly situating LLMs within qualitative research methodology. Rather than treating qualitative analysis as a classification problem, we examine how LLMs support meaning construction, theme evolution, and interpretive coherence across datasets. This positioning distinguishes our work from purely technical evaluations.

III. Methodology

The methodological framework of this study was designed to mirror real-world qualitative research workflows while enabling controlled evaluation [8]. We selected three types of qualitative datasets: semi-structured interview transcripts, open-ended survey responses, and ethnographic field notes. These datasets varied in length, tone, and contextual density to assess model robustness.

LLM-assisted analysis was conducted using prompt-driven thematic extraction, iterative coding refinement, and narrative synthesis. Prompts were carefully designed to emulate how a qualitative researcher approaches data, including requests for emergent themes, supporting quotations, and interpretive summaries. No predefined coding schema was imposed initially, allowing inductive theme generation.

For comparison, a team of experienced qualitative researchers independently analyzed the same datasets using traditional manual coding methods. Inter-coder reliability was established through iterative discussion and consensus building [9]. This human-generated analysis served as the reference benchmark rather than an absolute ground truth.

Evaluation metrics were adapted to reflect qualitative validity rather than purely statistical accuracy. These included thematic overlap, semantic alignment, interpretive coherence, and analytical completeness. Qualitative judgments were supplemented with similarity measures to assess consistency between LLM-generated and human-generated themes.

Importantly, reflexive analysis was incorporated into the methodology. We documented instances where LLM interpretations diverged from human reasoning and analyzed whether these differences represented analytical errors, alternative perspectives, or model hallucinations. This reflexive layer was essential for understanding not just performance, but epistemic behavior.

IV. Experimental Setup

The experimental environment was structured to ensure reproducibility and methodological transparency. Each dataset was analyzed independently by the LLM across multiple runs to assess consistency. Temperature and randomness parameters were controlled to minimize stochastic variation while preserving interpretive flexibility [10].

Prompts were standardized across datasets, with minor contextual adjustments to reflect data type. Each prompt requested thematic identification, explanation of themes, and illustrative excerpts. The LLM was not provided with prior examples or training data specific to the datasets, ensuring a fair assessment of generalization capability.

Human analysts conducted their coding using established qualitative analysis software, following best practices such as memo writing and iterative theme refinement [11]. Analysis sessions were time-tracked to enable efficiency comparisons between human-only and LLM-assisted workflows.

To evaluate scalability, dataset sizes were progressively increased, ranging from small interview samples to large collections of survey responses. The impact of dataset size on analytical quality and processing time was systematically recorded.

Ethical considerations were addressed by anonymizing all data and ensuring that no sensitive or personally identifiable information was processed. The study design emphasized methodological integrity rather than performance maximization.

V. Results and Discussion

The experimental results indicate that LLMs achieved high thematic alignment with human analysts, particularly for dominant and recurring themes. In interview transcripts, LLM-generated themes overlapped with human-identified themes in over 85% of cases, demonstrating strong semantic sensitivity to contextual cues [12].

In open-ended survey responses, LLMs excelled at identifying latent patterns across large datasets, uncovering themes that human analysts identified only after extensive iteration. This highlights the model's strength in breadth-oriented sense-making, particularly when dealing with scale.

However, qualitative differences emerged in interpretive depth. Human analysts demonstrated greater reflexivity, particularly when themes required socio-cultural contextualization. LLM interpretations occasionally lacked epistemic caution, presenting plausible but over-confident explanations that required human validation.

Efficiency gains were substantial. LLM-assisted workflows reduced initial coding time by approximately 60%, enabling researchers to focus more on interpretation, theory building, and reflexive analysis. Rather than replacing human judgment, LLMs shifted effort toward higher-level analytical tasks.

These findings suggest that LLMs function best as analytical accelerators rather than autonomous interpreters. When integrated thoughtfully, they enhance consistency, scalability, and exploratory capacity while preserving the core epistemic values of qualitative research.

VI. Ethical and Methodological Considerations

The integration of LLMs into qualitative analysis raises important ethical questions regarding authorship, accountability, and interpretive authority. Qualitative research is inherently relational, and delegating interpretation to automated systems risks obscuring the researcher's positionality.

Bias propagation remains a central concern. LLMs inherit biases present in their training data, which may influence how themes are framed or which narratives are foregrounded. Without critical oversight, these biases can subtly shape analytical outcomes.

Transparency is another challenge. Unlike human analysts, LLMs do not provide explicit reasoning pathways or reflexive justifications for their interpretations. This opacity complicates auditability and methodological disclosure, which are cornerstones of qualitative rigor.

To address these concerns, we advocate for human-in-the-loop frameworks where LLM outputs are treated as provisional analytical artifacts rather than definitive interpretations. Reflexive engagement with model outputs should be an explicit methodological step. Ultimately, ethical integration requires reframing LLMs as tools that augment, rather than replace, the interpretive labor of qualitative researchers.

VII. Conclusion

Large Language Models represent a transformative yet nuanced advancement in qualitative data analysis, offering unprecedented scalability, consistency, and analytical speed while preserving much of the semantic richness that defines qualitative inquiry. This study demonstrates that LLMs can effectively support thematic extraction and narrative synthesis across diverse qualitative datasets, particularly when positioned as collaborative analytical tools rather than autonomous interpreters. While challenges related to bias, transparency, and interpretive authority remain, these limitations are not insurmountable when addressed through reflexive, human-centered research designs. As qualitative research continues to evolve in data-intensive contexts, LLMs have the potential to reshape methodological practice—enabling researchers to spend less time on mechanical coding and more time on theory building, meaning-making, and critical insight.

REFERENCES:

[1] F. Zhao, F. Yu, T. Trull, and Y. Shang, "A new method using LLMs for keypoints generation in qualitative data analysis," in *2023 IEEE Conference on Artificial Intelligence (CAI)*, 2023: IEEE, pp. 333-334.

[2] E. Aghaei, X. Niu, W. Shadid, and E. Al-Shaer, "Securebert: A domain-specific language model for cybersecurity," in *International Conference on Security and Privacy in Communication Systems*, 2022: Springer, pp. 39-56.

[3] S. Khairnar and D. Bodra, "Analysis and Evaluation of Modern Lightweight Cryptographic Algorithms: Standards, Hardware Implementation, and Security Considerations," *International Journal of Computer Applications,* vol. 975, p. 8887, 2025.

[4] S. Al-Mansoori and M. B. Salem, "The role of artificial intelligence and machine learning in shaping the future of cybersecurity: trends, applications, and ethical considerations," *International Journal of Social Analytics,* vol. 8, no. 9, pp. 1-16, 2023.

[5] F. Zhao and F. Yu, "Enhancing multi-class news classification through Bert-augmented prompt engineering in large language models: A novel approach," in *The 10th International scientific and practical conference "Problems and prospects of modern science and education"(March 12–15, 2024) Stockholm, Sweden. International Science Group. 2024. 381 p.*, 2024, p. 297.

[6] W. Guo, S. Zong, S. Chen, F. Zhao, and Y. Shang, "Design and Implementation of a New Serverless Conversational Survey System," in *2021 IEEE International Conference on Data Science and Computer Application (ICDSCA)*, 2021: IEEE, pp. 358-363.

[7] W. S. Ismail, "Threat Detection and Response Using AI and NLP in Cybersecurity," 2020.

[8] F. Zhao, Y. Shang, and T. J. Trull, "FENAP: Foundation models for EMA-derived narrative analysis and prediction," *Journal of Healthcare Informatics Research,* pp. 1-21, 2024.

[9] A. Nassar and M. Kamal, "Ethical dilemmas in AI-powered decision-making: a deep dive into big data-driven ethical considerations," *International Journal of Responsible Artificial Intelligence,* vol. 11, no. 8, pp. 1-11, 2021.

[10] S. Khairnar and D. Bodra, "A Data-Driven Approach to Air Traffic Delay Prediction and Sentiment Evaluation," *International Journal of Basic and Applied Sciences,* vol. 14, no. 4, pp. 184-193, 2025.

[11] M. Nadeem, S. W. Zahra, M. N. Abbasi, A. Arshad, S. Riaz, and W. Ahmed, "Phishing attack, its detections and prevention techniques," *Int. J. Wirel. Secur. Netw,* vol. 1, pp. 13-25, 2023.

[12] F. Zhao, F. Yu, and Y. Shang, "A New Method Supporting Qualitative Data Analysis Through Prompt Generation for Inductive Coding," in *2024 IEEE International Conference on Information Reuse and Integration for Data Science (IRI)*, 2024: IEEE, pp. 164-169.