AI-Enhanced Thematic Analysis of COVID-19 Impact: Combining Human Expertise with Generative AI

FARNAZ ASRARI, Indiana University, USA

L. JEAN CAMP, Indiana University, USA

Qualitative research provides rich insights but remains highly labor-intensive. AI tools offer opportunities to reduce researchers' labor demands. This study examines the integration of ChatGPT and Claude.ai into qualitative analysis through a case study of 1,681 open-ended responses on the impact of COVID-19 (2020–2023). While ChatGPT excelled at descriptive coding and Claude.ai at theme synthesis, both required careful human oversight. Our findings suggest that AI can enhance the efficiency of qualitative research when thoughtfully integrated with human expertise. By providing a case study, we contribute to emerging AI-assisted methodologies and advocate for continued cross-disciplinary dialogue on the potentials and risks of AI-assisted qualitative coding.

Additional Key Words and Phrases: Qualitative Methods, Generative AI, ChatGPT, Claude.ai, Human-Centered AI, Human-AI Collaboration, COVID-19

1 Introduction

Qualitative research explores human experiences through contextual inquiry [1], offering rich, nuanced insights. However, its analysis is often labor-intensive, requiring extensive iterative coding and remaining vulnerable to subjective interpretation despite multiple analysts [6, 16, 21]. As datasets grow larger, traditional qualitative methods struggle to keep pace. Emerging AI tools—particularly Large Language Models (LLMs) like ChatGPT—offer new opportunities to scale qualitative analysis by assisting in coding and classification [5, 9, 11, 13, 15, 18, 20, 23, 26]. Yet, while AI can reduce researchers' labor, it cannot replace the critical human judgment necessary for reliable interpretation.

This study examines AI-assisted qualitative coding in a COVID-19 impact study embedded within a broader privacy and risk perception survey. We analyzed 1,681 open-ended responses (2020–2023) about pandemic-related life changes. Traditional methods struggled with the dataset's scale, prompting the integration of AI-assisted thematic analysis. ChatGPT generated initial codes, which were refined by Claude.ai and further reviewed by human coders. After two independent coding phases, an expert coder finalized the codebook, balancing AI efficiency with human oversight to ensure interpretive rigor. Conducted from December 2024 to December 2025, this approach streamlined analysis while critically assessing AI's role in qualitative research.

This paper details our AI-assisted coding process, highlighting both the potential and the limitations of using AI in qualitative research, and underscores the indispensable role of human expertise.

2 Literature Review

2.1 Thematic Analysis

Thematic analysis has become a widely recognized qualitative method, largely shaped by Braun and Clarke's work [19]. Their six-phase approach—data familiarization, coding, theme identification, refinement, definition, and reporting—provides

Authors' Contact Information: Farnaz Asrari, fasrari@iu.edu, Indiana University, Bloomington, USA; L. Jean Camp, ljcamp@iu.edu, Indiana University, Bloomington, USA.

a structured yet flexible framework for analyzing qualitative data [4]. This iterative process enables researchers to identify patterns of meaning across datasets while actively constructing and interpreting themes [4, 24].

A major strength of thematic analysis is its adaptability across different epistemological approaches, allowing for both descriptive and interpretive analyses. However, it risks superficiality if not grounded in theoretical interpretation [19]. While highly effective, it remains time-intensive and laborious, particularly when handling large or complex datasets [7, 10, 25]. To manage these demands, researchers often rely on qualitative data analysis software (QDAS) such as ATLAS.ti and NVivo, which support data organization, coding, and retrieval [22]. Yet even with such tools, large-scale qualitative analysis continues to present significant logistical and interpretive challenges.

2.2 AI tools in Qualitative Research

Qualitative analysis increasingly integrates topic modeling and Large Language Models (LLMs). Baumer et al. (2016) compared computational and traditional methods, showing that topic modeling identified patterns in Facebook survey data in two days versus 2.5 months for grounded theory, though the latter produced richer themes [2]. Towler et al. (2023) similarly found that machine-assisted topic analysis (MATA) matched human-generated themes while significantly reducing analysis time for large public health datasets [20].

ChatGPT's introduction by OpenAI has expanded AI-assisted qualitative research possibilities [27]. Recent studies have explored its use in sentiment analysis [11], thematic analysis [23, 26], and grounded theory [13, 18]. Lossio-Ventura et al. (2023) found ChatGPT effective in sentiment analysis of COVID-19 survey data, performing comparably to traditional tools [11].

Research on ChatGPT's qualitative applications highlights both potential and limitations. Yan et al. (2024) noted its ability to detect nuanced language patterns in thematic analysis but questioned its reliability [23]. Sinha et al. (2024) found GPT-4 useful in grounded theory by identifying overlooked data and generating analytical memos but stressed that AI should augment, not replace, human expertise [18].

2.3 Conducting AI-Assisted Qualitative Analysis

Effective use of ChatGPT in qualitative research requires well-crafted prompts that specify methodology, input format, and analytical goals [25]. Iterative refinement and task-specific frameworks improve output quality and utility [25].

2.4 Assessing Accuracy, Reliability, and Ethics

While ChatGPT aids qualitative analysis, it has limitations. Its responses can be inaccurate, imprecise, or disorganized, requiring verification, especially across multiple prompts [25, 28]. AI models excel in deductive analysis but struggle to identify emerging themes inductively [25].

Challenges include time-intensive prompt design, evaluating AI outputs, and addressing accuracy and bias concerns [25]. Trust, contextual limitations, and academic acceptance remain ongoing issues [24]. Ethical concerns, particularly regarding privacy, data security, and fairness, further complicate AI use [12]. Non-open-source models like ChatGPT raise transparency concerns due to external data processing [12]. Additionally, biases in AI training data risk reinforcing disparities in race, gender, and sexuality, posing challenges when analyzing responses from underrepresented populations [3, 8, 14].

3 Methodology

3.1 Survey Design

From 2020 to 2023, we conducted a national survey on Americans' attitudes toward IoT data sharing during COVID-19. Participants were randomly assigned to assess comfort with Internet-connected devices (smartphones, security cameras, and fitness trackers) sharing personal health and movement data. Each group evaluated data sharing with five recipients: law enforcement, healthcare providers, insurers, manufacturers, and marketers. The "general" group considered sharing for broad purposes like customization, while the "health" group assessed sharing specifically for infectious disease control.

The survey incorporated questions about participants' technical expertise, risk perceptions, understanding of COVID-19, and demographic information. To gain deeper qualitative insights into the pandemic's personal impact, we included the open-ended question: "How has COVID-19 impacted your life (e.g., daily routines, behaviors, and your food preferences)?" This question yielded rich narrative data about participants' lived experiences during the pandemic. The flow of the study is shown in the Figure 1 attached to appendix.

The survey was created using Qualtrics. It was distributed using Prolific after IRB approval. Prolific recruits a representative sample of the US population in terms of age, gender, and political affiliation [17]. The study began with an information sheet, followed by two scenarios, and then the questions.

3.2 Participant Recruitment

A 2020 pilot study with ten university students refined the survey before its Prolific launch. Students later reassessed it before deployment. Prolific participants received $3 compensation ($15/hour). Response rates increased from 262 (2020) to 485 (2023), maintaining balanced demographics. In 2022–2023, crisis resource information was added, but core questions remained unchanged. Survey details and demographics are in the appendix.

3.3 Qualitative analysis using Generative AI

We utilized ChatGPT-4 for this study, accessing it through a premium subscription. We generated initial thematic codes by importing 20–30 responses at a time due to file-reading limitations. Using the prompt:

> "I have open-ended survey responses about the COVID-19 pandemic. I will provide them in parts. Please perform qualitative coding and generate codes."

we compiled all codes into a single query for summarization. However, ChatGPT struggled to merge codes effectively, prompting us to use Claude.ai, which successfully consolidated four annual code-books into 13 themes and 80 codes. After manual refinements, the final code-book contained 11 themes and 65 codes.

Next, we instructed ChatGPT to apply the consolidated code-book to new responses. It correctly mapped codes in most cases but occasionally generated new ones for unaccounted responses. To maintain consistency, we instructed it to strictly adhere to the code-book. However, prolonged use led to performance declines, requiring session resets and code-book reuploads.

Trained coders reviewed ChatGPT's coding, deeming 63.66% of assignments fully accurate which were mostly for short and clear responses. This review led to code-book refinements, adding new categories to better capture the pandemic's impact across economic, political, and lifestyle dimensions. To validate results, independent human coders applied the final code-book without AI assistance, ensuring unbiased labeling.

The expert coder then integrated human and AI-generated codes, using the refined prompt:

"I will import a code-book followed by participant responses. Please map suitable codes to each for thematic analysis."

Even at this stage, manual revisions were needed as broader dataset analysis revealed emerging patterns. ChatGPT's inconsistency in extended sessions required frequent resets. The final revision consolidated themes while expanding subcategories, especially in shopping behaviors and mental health impacts.

ChatGPT demonstrated notable capabilities, producing highly descriptive codes and inferring implicit themes. However, it occasionally deviated from the code-book, struggled with "other" category assignments, and failed to distinguish minimal changes due to pre-existing conditions. Despite these challenges, it effectively identified patterns, such as interpreting work-from-home adjustments based on indirect cues.

ChatGPT also exhibited ethical sensitivity, offering condolences in response to loss-related statements before resuming coding tasks. Finally, we assessed AI-human agreement using Cohen's Kappa, achieving a substantial 0.827 score with 64.25% total agreement. This high reliability reflects the nature of our short, descriptive responses, a format where AI tools excel.

4 Finding and Discussion

Our integration of AI tools in qualitative analysis highlighted both strengths and limitations. ChatGPT demonstrated strong descriptive coding, effectively capturing nuanced patterns, aligning with prior findings [18, 23]. Claude.ai excelled at synthesizing code-books. However, these tools may have performed well due to the short, descriptive nature of our data. As Terry et al. (2017) emphasize, qualitative research must go beyond description to achieve theoretical depth [19]. Deep engagement with data remains essential for uncovering insights beyond statistical modeling.

Despite its strengths, ChatGPT struggled with consistent adherence to the code-book, often generating overly specific or non-compliant codes. It also exhibited performance degradation over extended use, requiring iterative refinements, careful prompt design, and manual oversight—challenges noted in prior research [18, 23, 25]. Whether AI reduces labor in qualitative analysis remains uncertain, as human evaluation is still essential.

Expert coders agreed AI cannot replace human judgment or eliminate bias, as it is itself inherently biased. ChatGPT's coding was fully acceptable only 63.66% and 64.25% of the time in two evaluation phases. However, as a secondary coder, it achieved a Cohen's Kappa score of 0.827 with the expert coder, indicating substantial agreement. Yet, this came with considerable manual effort, including careful prompt design, segmenting data, and validating outputs.

Beyond methodology, ethical concerns around privacy, security, and bias remain critical [11, 14]. Additionally, AI paywalls risk privileging well-funded researchers, shifting research priorities from societal impact to profitability.

5 Conclusion

Our study highlights AI's transformative yet complex role in qualitative research. Integrating ChatGPT and Claude.ai in thematic analysis demonstrated AI's potential to enhance efficiency, especially with large datasets. ChatGPT excelled in descriptive coding, while Claude.ai effectively synthesized themes, illustrating their complementarity. However, AI remains an assistive tool, not a replacement for human expertise, requiring careful oversight, iterative refinement, and ethical consideration.

References

[1] Patrik Aspers and Ugo Corte. 2019. What is qualitative in qualitative research. *Qualitative sociology* 42 (2019), 139–160.
[2] Eric P. S. Baumer, David Mimno, Shion Guha, Emily Quan, and Geri K. Gay. 2017. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology* 68, 6 (June 2017), 1397–1410. doi:10.1002/asi.23786

[3] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*. 610–623.
[4] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.
[5] António Pedro Costa. 2023. Qualitative Research Methods: do digital tools open promising trends? *Revista Lusófona de Educação* 59, 59 (2023).
[6] Jan De Vos. 2009. Now that you know, how do you feel? The Milgram experiment and psychologization. *Annual Review of Critical Psychology* 7 (2009), 223–246.
[7] Therese Ferguson and Tenesha Gordon. 2019. "Miss, this is a lot of work": Exploring Part-Time Students Experiences of Qualitative Research. *The Qualitative Report* (Nov. 2019). doi:10.46743/2160-3715/2019.3986
[8] Stefan Harrer. 2023. Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *EBioMedicine* 90 (2023).
[9] Jialun Aaron Jiang, Kandrea Wade, Casey Fiesler, and Jed R. Brubaker. 2021. Supporting Serendipity: Opportunities and Challenges for Human-AI Collaboration in Qualitative Analysis. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–23. doi:10.1145/3449168
[10] Jasy Suet Yan Liew, Nancy McCracken, Shichun Zhou, and Kevin Crowston. 2014. Optimizing Features in Active Machine Learning for Complex Qualitative Content Analysis. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*. Association for Computational Linguistics, Baltimore, MD, USA, 44–48. doi:10.3115/v1/W14-2513
[11] Juan Antonio Lossio-Ventura, Rachel Weger, Angela Y Lee, Emily P Guinee, Joyce Chung, Lauren Atlas, Eleni Linos, and Francisco Pereira. 2023. A Comparison of ChatGPT and Fine-Tuned Open Pre-Trained Transformers (OPT) Against Widely Used Sentiment Analysis Tools: Sentiment Analysis of COVID-19 Survey Data (Preprint). doi:10.2196/preprints.50150
[12] Juan Antonio Lossio-Ventura, Rachel Weger, Angela Y Lee, Emily P Guinee, Joyce Chung, Lauren Atlas, Eleni Linos, and Francisco Pereira. 2023. A Comparison of ChatGPT and Fine-Tuned Open Pre-Trained Transformers (OPT) Against Widely Used Sentiment Analysis Tools: Sentiment Analysis of COVID-19 Survey Data (Preprint). doi:10.2196/preprints.50150
[13] Blaž Mesec. 2023. The language model of artificial inteligence chatGPT - a tool of qualitative analysis of texts. doi:10.22541/au.168182047.70243364/v1
[14] David L. Morgan. 2023. Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT. *International Journal of Qualitative Methods* 22 (Jan. 2023), 16094069231211248. doi:10.1177/16094069231211248
[15] SJRK Padminivalli V, MVP Chandra Sekhara Rao, and Naga Sai Ram Narne. 2024. Sentiment based emotion classification in unstructured textual data using dual stage deep model. *Multimedia Tools and Applications* 83, 8 (2024), 22875–22907.
[16] Catherine Pope and Nicholas Mays. 2013. Qualitative research in health care. (2013).
[17] Prolific Researcher Help. 2025. How Do I Screen Participants Using Pre-Screening Questions? https://researcher-help.prolific.com/en/article/95c345 Accessed: 2025-02-06.
[18] Ravi Sinha, Idris Solola, Ha Nguyen, Hillary Swanson, and LuEttaMae Lawrence. 2024. The Role of Generative AI in Qualitative Research: GPT-4's Contributions to a Grounded Theory Analysis. In *Proceedings of the Symposium on Learning, Design and Technology*. ACM, Delft Netherlands, 17–25. doi:10.1145/3663433.3663456
[19] Gareth Terry, Nikki Hayfield, Victoria Clarke, Virginia Braun, et al. 2017. Thematic analysis. *The SAGE handbook of qualitative research in psychology* 2, 17-37 (2017), 25.
[20] Lauren Towler, Paulina Bondaronek, Trisevgeni Papakonstantinou, Richard Amlôt, Tim Chadborn, Ben Ainsworth, and Lucy Yardley. 2023. Applying machine-learning to rapidly analyze large qualitative text datasets to inform the COVID-19 pandemic response: comparing human and machine-assisted topic analysis techniques. *Frontiers in Public Health* 11 (Oct. 2023), 1268223. doi:10.3389/fpubh.2023.1268223
[21] Brett Wolff, Frank J. Mahoney, Anna Leena Lohiniva, and Melissa Corkum. 2019. Collecting and Analyzing Qualitative Data. *The CDC Field Epidemiology Manual* (2019). https://api.semanticscholar.org/CorpusID:201895083
[22] Megan Woods, Trena Paulus, David P Atkins, and Rob Macklin. 2016. Advancing qualitative research using qualitative data analysis software (QDAS)? Reviewing potential versus practice in published studies using ATLAS. ti and NVivo, 1994–2013. *Social science computer review* 34, 5 (2016), 597–617.
[23] Lixiang Yan, Vanessa Echeverria, Gloria Milena Fernandez-Nieto, Yueqiao Jin, Zachari Swiecki, Linxuan Zhao, Dragan Gašević, and Roberto Martinez-Maldonado. 2024. Human-AI Collaboration in Thematic Analysis using ChatGPT: A User Study and Design Recommendations. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–7. doi:10.1145/3613905.3650732
[24] Lixiang Yan, Vanessa Echeverria, Gloria Milena Fernandez-Nieto, Yueqiao Jin, Zachari Swiecki, Linxuan Zhao, Dragan Gašević, and Roberto Martinez-Maldonado. 2024. Human-AI Collaboration in Thematic Analysis using ChatGPT: A User Study and Design Recommendations. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–7. doi:10.1145/3613905.3650732
[25] He Zhang, Chuhao Wu, Jingyi Xie, Yao Lyu, Jie Cai, and John M Carroll. 2023. Redefining qualitative analysis in the AI era: Utilizing ChatGPT for efficient thematic analysis. *arXiv preprint arXiv:2309.10771* (2023).
[26] He Zhang, Chuhao Wu, Jingyi Xie, Yao Lyu, Jie Cai, and John M. Carroll. 2024. Redefining Qualitative Analysis in the AI Era: Utilizing ChatGPT for Efficient Thematic Analysis. http://arxiv.org/abs/2309.10771 arXiv:2309.10771 [cs].
[27] Mert Şen, Şevval Nur Şen, and Tuğrul Gökmen Şahin. 2023. A New Era for Data Analysis in Qualitative Research: ChatGPT! *Shanlax International Journal of Education* 11, S1-Oct (Oct. 2023), 1–15. doi:10.34293/education.v11iS1-Oct.6683
[28] Mert Şen, Şevval Nur Şen, and Tuğrul Gökmen Şahin. 2023. A New Era for Data Analysis in Qualitative Research: ChatGPT! *Shanlax International Journal of Education* 11, S1-Oct (Oct. 2023), 1–15. doi:10.34293/education.v11iS1-Oct.6683

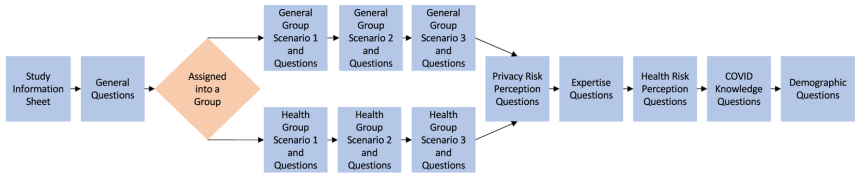

Fig. 1. Survey flow showing participant pathways through scenario-based questions and assessments.

A Appendix

Table 1. Participant Recruitment over 4 Years

Year	Date	Total	Rejected	Withdraw	Timed-Out
2020	1-10 May	256	17	22	4
2021	24 March	458	10	24	4
2022	21-23 April	482	20	17	4
2023	8 May	485	0	19	4

Table 2. Participant Demographics (2020–2024)

Category	**2020**	**2021**	**2022**	**2023**	**2024**
Group Assignment (N)					
General	129	227	241	244	270
Health	127	231	241	241	263
Total	256	458	499	485	533
Gender (%)					
Male	48	48	46	48	48
Female	49	51	48	50	50
Non-binary	2	2	1	<1	1
Education (%)					
Secondary or Less	29	24	32	29	29
Post-Secondary	68	74	65	69	71
Income (%)					
<30K	29	19	22	16	19
>30K	71	78	71	79	77
Technical Skill (%)					
Level 1-3	56	55	57	57	60
Level 4-7	32	28	23	31	26
Security Skill (%)					
Level 1-3	25	22	23	28	27
None	42	77	75	72	65