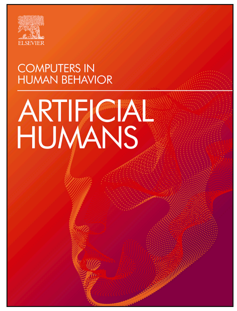
Journal Pre-proof



Harnessing the Power of AI in Qualitative Research: Exploring, Using and Redesigning ChatGPT

He Zhang, Chuhao Wu, Jingyi Xie, Yao lyu, Jie Cai, John M. Carroll

PII: S2949-8821(25)00028-3

DOI: https://doi.org/10.1016/j.chbah.2025.100144

Reference: CHBAH 100144

To appear in: *Computers in Human Behavior: Artificial Humans*

Received Date: 16 December 2024

Revised Date: 13 March 2025

Accepted Date: 19 March 2025

Harnessing the Power of AI in Qualitative Research: Exploring, Using and Redesigning ChatGPT

He Zhang[a], Chuhao Wu[a], Jingyi Xie[a], Yao lyu[a, b], Jie Cai[a, c], John M. Carroll[a]

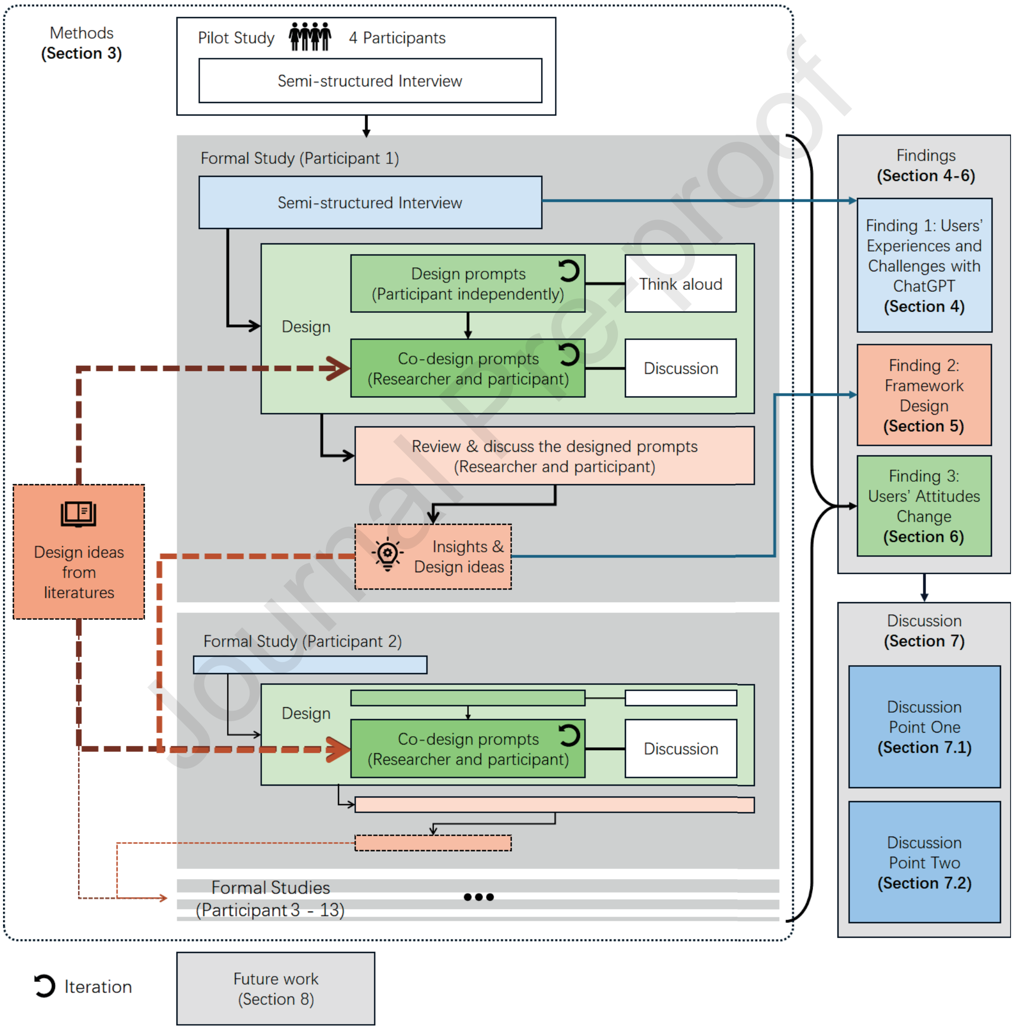
[a] College of Information Sciences and Technology, The Pennsylvania State University, Westgate Building, University Park, PA 16803 USA
[b] The School of Information, University of Michigan, MI 48109 USA
[c] The Pervasive Human Computer Interaction Laboratory, Department of Computer Science and Technology, Tsinghua University, Future Internet Technology Research Center (FIT), Haidian District, Beijing, 100084, China

Email addresses: hpz5211@psu.edu (He Zhang, Corresponding author), cjw6297@psu.edu (Chuhao Wu), jzx5099@psu.edu (Jingyi Xie), yml5549@psu.edu (Yao lyu), jie-cai@mail.tsinghua.edu.cn (Jie Cai), jmc56@psu.edu (John M. Carroll)

Graphical Abstract

Harnessing the Power of AI in Qualitative Research: Exploring, Using and Redesigning ChatGPT

Anonymous



Harnessing the Power of AI in Qualitative Research: Exploring, Using and Redesigning ChatGPT

Anonymous

Abstract

AI tools, particularly large-scale language model (LLM) based applications such as ChatGPT, have the potential to mitigate qualitative research workload. In this study, we conducted semi-structured interviews with 17 participants and held a co-design session with 13 qualitative researchers to develop a framework for designing prompts specifically crafted to support junior researchers and stakeholders interested in leveraging AI for qualitative research. Our findings indicate that improving transparency, providing guidance on prompts, and strengthening users' understanding of LLMs' capabilities significantly enhance their ability to interact with ChatGPT. By comparing researchers' attitudes toward LLM-supported qualitative analysis before and after the co-design process, we reveal that the shift from an initially negative to a positive perception is driven by increased familiarity with the LLM's capabilities and the implementation of prompt engineering techniques that enhance response transparency and, in turn, foster greater trust. This research not only highlights the importance of well-designed prompts in LLM applications but also offers reflections for qualitative researchers on the perception of AI's role. Finally, we emphasize the potential ethical risks and the impact of constructing AI ethical expectations by researchers, particularly those who are novices, on future research and AI development.

Keywords: ChatGPT, Qualitative Analysis, Prompt Design, Large Language Models, AI-assisted Research, Ethical Considerations

1. Introduction

Thematic analysis in qualitative research is a highly flexible and widely used method for identifying, analyzing, and interpreting patterns of meaning ('themes') within qualitative data [1, 2]; however, it can be time-consuming and require significant manual effort, especially when dealing with large and complex datasets [3]. As we continue to generate increasingly massive volumes of qualitative data, there's an urgent need to seek out innovative methods to streamline and enhance the process of thematic analysis [4].

The realm of artificial intelligence (AI) is swiftly evolving, poised to revolutionize the ways we approach research [5]. In particular, Machine learning algorithms have already demonstrated efficacy in processing and analyzing vast datasets across various domains [6]. Large Language Models (LLMs) have significantly expanded these capabilities, with ChatGPT emerging as a particularly noteworthy example by 2023. Its conversational interface allows users to interact through natural language, making advanced AI functionality accessible even to those without technical expertise [7]. It's distinguished by its exceptional ability to understand and generate human-like text [8, 9], making it a potential asset for qualitative researchers.

Prior research on LLMs and qualitative research has explored several key aspects: a thread of research explored the theme development, utilizing GPT to generate preliminary ideas and provide researchers with an overview of the dataset. For example, studies have compared themes generated by ChatGPT with those developed by researchers to highlight the potential of LLMs in theme development [10, 11].

However, it is important to note that the themes are highlightly descriptive, and details such as quotes may be fabricated by ChatGPT. Another thread of research examines the potential of LLM-based tools as individual coders to streamline the coding process [12] and assess reliability [11].

Current research primarily focuses on the outcomes of thematic analysis facilitated by LLMs [11, 13, 14, 15, 16] rather than on the role these models play in the analytic process, particularly regarding prompt design. We argue that prompt design is essential for fully harnessing the capabilities of LLMs. On one hand, qualitative researchers often lack the technical literacy required to design effective prompts. On the other hand, they recognize the potential of LLMs and are eager to explore ways to incorporate them into their data analysis, demonstrating openness to adopting new technologies that support their analytical work. Although junior researchers face significant challenges due to their limited experience with thematic analysis and typically require more guidance from peers and advisors, the use of LLMs might help reduce this dependency by providing additional analytical support.

While ChatGPT makes human-AI interaction unprecedentedly direct and accessible through simple textual prompts [17], the quality of results depends significantly on how effectively these prompts are designed [18, 19]. Creating optimal prompts often requires specific expertise and careful consideration of the analytical goals, which is challenging for junior researchers.

researchers' skills in designing prompts during the qualitative research process or the attitudinal dynamics they experience throughout.

This study examines the potential of ChatGPT as an instrument for thematic analysis in qualitative research. We present a systematic exploration into leveraging ChatGPT's capabilities through prompt design skill development, highlighting both its promises and limitations for practical imple- mentation in qualitative analysis. Our work addresses the following research question (RQ):

RQ. **Can the performance of ChatGPT in qualitative analysis tasks be enhanced through prompt design? If so, how?**

To address this question, we employed a two-phase methodology. First, we conducted semi-structured interviews with 17 participants to identify challenges and concerns associated with using ChatGPT for qualitative analysis. Second, we engaged 13 researchers with expertise in qualitative methods in a collaborative co-design process. Through this process, we identified specific techniques, interaction approaches, and conceptual frameworks that significantly enhance ChatGPT's efficacy in analyzing qualitative data. These insights have been synthesized into a comprehensive prompt framework (Fig. 1), specifically designed to bolster ChatGPT's capabilities in aiding the analysis of qualitative data.

Our research has also generated important discussions around three key areas: (1) the robustness and appropriate scope of ChatGPT in qualitative research applications; (2) the potential positioning of ChatGPT as either a co-researcher or a specialized tool; and (3) the evolving ethical considerations of AI in qualitative analysis, particularly how providing data-grounded evidence through effective prompt design can transform researchers' initial ethical concerns into acceptance. The framework we propose offers a generalized structure that, while developed with ChatGPT, incorporates foundational elements adaptable to future LLMs, such as task contextualization protocols and structured output formats. We envision this framework evolving alongside advancements in prompt engineering while maintaining its core utility for qualitative researchers and contributing to the broader dialogue about the intersection of AI and qualitative research [20]. The article's content structure, conceptual relationships, and methodological workflow are illustrated in Fig. 1.

2. Related Work

2.1. ChatGPT and Prompt Engineering

ChatGPT, an advanced language model developed by OpenAI, has been recognized for its versatility across various language tasks, making it a powerful tool for various applications [21, 22]. Its capabilities include generating human-like text, content creation, sentence, and paragraph completion, as well as essay and report writing [23, 24, 25, 26]. However, the quality of output from LLMs like ChatGPT is significantly influenced by the instructions or "prompts" given to them [18, 19]; the model can sometimes generate nonsensical or incorrect outputs, particularly with ambiguous prompts [27, 28, 29].

Various studies have shown that advanced prompt engineering techniques often yield more desirable results. Existing strategies include few-shot learning [17, 30], label calibration [30], providing examples [31, 18], chain-of-thought approaches [32, 33], sentence relationship explanation [34], content explanation [7], and role-playing [35]. Additionally, learning prompting structures is considered an important step for effectively using AI tools [36, 37]. Incorporating prompt design to facilitate workflows, supervision, critical evaluation, and expanding insights has also been integrated into courses, teaching, and discussion [38, 39, 36, 40].

While these related works indicate that the performance of ChatGPT can be enhanced through prompt engineering, they also highlight the challenges and limitations of current approaches [18]. Effective utilization of ChatGPT varies across domains, and domain-specific knowledge is essential for optimizing its performance [41, 42]. Experts advocate that prompt engineering should take the specific application context into account [43]. However, there is a lack of research on how to design prompts for complex, open-ended tasks such as qualitative analysis, which require a deep understanding of the domain and the ability to interpret nuanced language. Moreover, existing prompt engineering techniques often focus on improving the accuracy and coherence of the generated output but pay less attention to the model's transparency and explainability, which are crucial for building trust and facilitating collaboration between humans and AI systems [44].

In this study, we utilize a human-centric approach to prompt engineering for qualitative analysis with ChatGPT. We aim to develop a framework to empower junior qualitative researchers to effectively leverage ChatGPT's capabilities in their analysis workflows.

2.2. Current Methodologies and Features of Thematic Analysis

Thematic analysis [1] serves as a cornerstone in qualitative research, offering researchers a method to identify, analyze, interpret, and elucidate patterns or themes from their data. Its application transcends disciplines, contributing to its popularity and widespread use. Current practices encompass a range of methodologies [45, 46], each bringing a unique perspective to the process of thematic analysis.

The Six-Phase Approach [1, 47, 48] is widely recognized for its sequential process, guiding researchers through distinct phases, from data familiarization to final report preparation. This iterative methodology demands active researcher engagement in identifying themes. However, this engaged approach can be challenging for junior researchers lacking qualitative analysis experience.

Boyatzis' Codebook Approach [49] introduces a paradigm that necessitates creating a codebook before data coding. While beneficial for managing large datasets or collaborative coding [50, 51], developing a comprehensive codebook requires significant expertise, which may be daunting for junior researchers.

Thematic analysis poses several challenges, many amplified for junior researchers. Subjectivity in identifying themes [52] can lead to varying interpretations [53], resulting in multiple potential themes from the same dataset [54]. This 'researcher subjectivity ' [55] necessitates reflexivity and transparency, which may be difficult for those new to the method. Moreover, thematic analysis is often resource-intensive [56], particularly for large datasets [57, 58]. The time and effort required for coding and identifying patterns can be overwhelming for junior researchers [59].

This study aims to explore how advanced AI tools can collaborate with and support junior researchers in conducting thematic analysis. By developing a human-centric prompt design framework for ChatGPT, we investigate how junior researchers can leverage AI to address subjectivity, resources, and replicability challenges while empowering them to learn and apply thematic analysis effectively.

2.3. AI-augmented Qualitative Analysis

The increasing interest in Natural Language Processing (NLP) within the qualitative research community stems from its ability to analyze large volumes of text effectively. Sentiment analysis and topic modeling are the most frequently used NLP approaches for processing unstructured text data, such as patient feedback [60]. However, these methods have limitations, including the constrained range of detectable sentiments in sentiment analysis [61] and the difficulty in interpreting topic modeling results [62]. Guetterman et al. [63] found that while NLP approaches may lack nuance compared to human coders, they can augment human efforts and improve efficiency. This aligns with the views of Hong et al. [64] and Gebreegziabher et al. [65], who suggest that AI should assist researchers in refining and evolving their coding rather than replace human analysts.

The integration of machine learning (ML) into qualitative coding faces challenges due to fundamental differences between qualitative and quantitative methods, with effective ML support focusing on identifying coding ambiguities [66]. AI's role in qualitative analysis extends to fostering collaboration, criticality, and reflectiveness [56], while researchers exhibit complex sentiments, balancing the appeal of AI assistance with skepticism toward AI-driven analysis [67].

Recent advancements in AI, particularly LLMs, have shown potential in qualitative analysis [68, 12, 69, 10, 70, 71, 72, 73] by supporting tasks like labeling and collaborative coding [74]. These models tend to excel in deductive analysis [75] but may face limitations in inductive reasoning and maintaining code diversity [76]. While AI can enhance coding efficiency, its impact on the quality and consistency of analysis continues to spark debate, reflecting broader concerns about integrating AI into qualitative research workflows [65].

Interpretability and understandability play a key role in implementing ethical AI in practice [77]. Trust is crucial in human-AI collaboration, with perceived risk, AI capabilities, expectations, and user vulnerability shaping the degree of trust users place in AI [78]. Transparency, through explanations and communication about how AI works, is more effective in enhancing user trust than simply providing algorithmic interpretations [79].

Users prefer explanations that clarify decision-making processes and illustrate how specific actions influence outcomes [80]. XAI and Human-Centric Generative AI should cater to different users' needs, from AI experts requiring detailed visualizations to novices needing simpler explanations [81]. The effectiveness of transparency is context-dependent [82], and starting with simplified models can help build foundational knowledge before gradually introducing complexity [83]. Buschek et al. [84]'s conceptual framework categorizes user understanding support in intelligent systems into user mindsets, involvement, and knowledge outcomes, highlighting the multifaceted nature of user engagement with AI systems.

For LLMs like ChatGPT, natural language interfaces may enhance transparency by delivering effective explanations interactively [85]. Allowing users to make manual edits and visualize model decisions may further boost explainability and adaptability [86]. However, Wang and Yin [87] noted that explanations are more beneficial when users have some level of domain expertise, suggesting that researchers should maintain strong data familiarity and verify reliability through cross-referencing when using AI for qualitative analysis [88].

ChatGPT, as an advanced AI tool, shows promise in advancing qualitative analysis but may also cause trust and ethical issues. This study aims to explore these aspects from the perspective of junior researchers, investigating how to leverage ChatGPT to support qualitative research while addressing potential negative issues. By examining the explainability and integration of ChatGPT in qualitative analysis, we assess the acceptance, concerns, trust, and ethical implications associated with using such models and explore strategies to effectively leverage ChatGPT while mitigating potential negative impacts.

3. Methods

3.1. Data Collection

This study was conducted online through video conferencing software (e.g., Zoom). The research process was recorded for transcription and analysis with the informed consent of the participants.

3.1.1. Pilot Study

To gain a deeper understanding of ChatGPT's capabilities and to establish a foundation for guidelines for interviews and co-design, we initiated a pilot interview study involving four participants who have experience using qualitative methods and ChatGPT. We recruited them via social media. All of the participants who took part in the pilot interview study had been involved in one or more projects using qualitative analytic methods within the past three years, and each of them had experience using ChatGPT. Among these participants, three had received formal training or education in qualitative data analysis. The aim of this pilot study was to investigate the challenges tied to ChatGPT usage. Each pilot interview study lasted for 1-1.5 hours.

The interviews were semi-structured. We began by introducing participants to the foundational concepts of qualitative analysis and probing their experiences in this realm. Following this, we explored the challenges they encountered both in the broader scope of qualitative analysis and specifically when using ChatGPT. The thematic analysis centered on users' reflections including encompassing uses, challenges, and strategies, during their interactions with ChatGPT. Through a thorough review, analysis, and reflection on the recorded sessions and their respective coding, we found that participants indeed have some concerns and challenges in using ChatGPT, but at the same time, they also demonstrated enthusiasm for ChatGPT. This is mainly due to its rapid data processing and improved work efficiency (P1, P2, P3, P4), providing concise overviews or summaries (P2, P3, P4), generating preliminary insights (P1, P2, P3, P4), and its user-friendly question-and-answer interaction format (P4).

In addition, all participants in the pilot interview study expressed interest in understanding how to use ChatGPT better and how to design prompts more effectively.

Based on this feedback, we refined the interview guide and developed a formal study protocol to further explore ChatGPT's capabilities, uses, and strategies for addressing challenges in qualitative research. In the lens of formal study coupled with qualitative scenario analysis, we devoted special attention to these challenges with an aspiration to pinpoint potential solutions for RQ.

3.1.2. Formal Study

In this section, we paid particular attention to how participants' design of the prompts affected ChatGPT's performance, their strategies in qualitative analysis, and expected outcomes. We worked with participants through a formal study to design ChatGPT prompts that were friendly to qualitative analysis [89]. Finally, we distilled from the design solutions a framework of prompts to be applied in ChatGPT for qualitative analysis tasks.

The formal study comprised two parts: interviews and co-design. All researchers convened for two meetings to develop the protocol, make a consensus on the procedure, and conduct internal tests prior to the formal study.

The interviews were semi-structured, focusing primarily on two areas: (1) challenges, approaches, and techniques encountered in qualitative analysis research, and (2) experiences, challenges, and insights related to using ChatGPT. Additionally, for participants who had experience with qualitative analysis software (e.g., NVivo and atlas.ti), we inquired about their experiences, benefits, and limitations with these tools. This part took approximately 20 minutes.

Next, we sent participants a corpus of qualitative data (a transcript of a focus group) and asked participants to use ChatGPT[1] to perform qualitative analysis coding on the content in the corpus while sharing screen. All participants interacted with ChatGPT in English. First, we allowed participants to design their own prompts (up to 5 times). During this process, researchers asked participants to think aloud and provide a detailed explanation of the intentions, requirements, and strategies in each designed prompt, and to evaluate the generated results. During the iteration (updating prompts) process, we also asked participants to explain their intentions and provide reasons for modifications. In addition, we asked participants to compare and evaluate the generated results from previous interactions and provide reasons for their respective attitudes.

Second, the researchers actively participated in the prompt design process by offering targeted suggestions to improve the prompts initially designed independently by participants. These suggestions were informed by strategies recommended in prior literature, such as few-shot learning [30], chain-of-thought approaches [32], role-playing [35], adding explanatory tags [18], categorizing prompts [33], understanding sentence structure and relationships [34], considering instruction usage and content [7], and integrating natural language and code in mixed inputs [31]. These insights helped shape the redesign of prompts, allowing for more nuanced and effective interactions with ChatGPT.

After each round of prompt refinement, researchers and participants collaboratively reviewed the generated outputs to evaluate their quality, align-

[1]We used ChatGPT based on the GPT-3.5 model instead of GPT-4.0 or later version due to accessibility concerns. At the time of the study, GPT-4.0 had usage limits (maximum access every four hours) and required a subscription, which we chose not to mandate. Additionally, prior research and our own testing showed no significant performance improvement in GPT-4.0 for the tasks in this study [90].

ment with expectations, and areas for further improvement. Researchers also collected feedback on participants' satisfaction with the updated results and incorporated their optimization suggestions into subsequent iterations. Notably, the iterative refinement process did not end with individual participants. Effective design suggestions provided by one participant were shared with subsequent participants, enabling a cumulative improvement in the overall prompt design process. In short, after the design process, researchers and participants collectively reviewed the prompts developed throughout the study. They evaluated the effectiveness of these prompts in enhancing ChatGPT's ability to process qualitative data and synthesize key insights and design recommendations for future prompt development. This final reflection underscored the iterative and collaborative nature of the process, ensuring that the outcomes were both practical and grounded in shared experiences. This approach facilitated knowledge transfer across sessions and ensured that successful strategies were disseminated and adapted for broader application.

The trial-and-error process was highly iterative, with researchers and participants co-designing prompts in successive cycles until the generated results met the participants' expectations. Each cycle emphasized refining the outputs and examining the underlying rationale of the prompt designs. By discussing the rationale behind prompt elements and their impact on outcomes, researchers and participants gained deeper insights into the interplay between prompt structure and AI performance, further enriching the co-design process. This collaborative approach underscored the importance of combining individual creativity with systematic, evidence-based design principles to achieve optimal results.

3.2. Participants Recruitment

In this study, we distributed recruitment questionnaires via social media and the authors' academic networks to recruit participants who met the following minimum inclusion criteria: being at least 18 years old, having qualitative analysis experience, and having used ChatGPT or similar tools. Participants who did not meet the inclusion criteria were informed during the questionnaire process, and no relevant information would be retained, while those who did meet the criteria could voluntarily provide their contact information to facilitate the arrangement of subsequent research processes. Any identifiable data will be permanently deleted after the study concludes. Our recruitment process was based on a first-come, first-served strategy, and finally, we recruited 17 participants (12 females and five males). The participants ranged from 20 to 32 years old (median = 27, SD ≈ 3.70). Regarding the termination of recruitment, we continuously conducted preliminary analyses of the collected data during the data collection process and found that theoretical saturation had been reached, meaning that the newly collected data did not offer any new insights into the research questions. Therefore, we concluded that the existing data were sufficient to support the study's conclusions and decided to cease further recruitment. P1-P4 were recruited to take part in the pilot study (Section 3.1.1), while the remaining 13 participated in the formal study (Section 3.1.2). Detailed demographic information is presented in Table 1. Except for one participant, all others in the pilot study had undergone formal training or courses in qualitative analysis. For formal study, we reached out to potential participants via our academic network, most of whom hailed from universities and were actively involved in qualitative research, while being fluent in English (if they are not native speakers). All 13 participants were researchers skilled in qualitative methods and had prior experience with ChatGPT. Among these participants, 11 were doctoral students, one was a master's student, and one worked in the industry but held a doctoral degree. All participants had formal training or education in qualitative data analysis. The pilot and formal studies were processed under the university's Institutional Review Board (IRB) approval. Informed consent was obtained from all participants. All participants received a $10 gift card (or equivalent) for their time and effort.

3.3. Data Analysis

The researchers conducted reflexive thematic analysis (RTA) on the collected data [47, 91]. The analysis followed a six-step procedure: dataset familiarization, data coding, initial theme generation, theme development and review, theme refinement and definition, and report composition. Specifically, after each interview and co-design session, the research team briefly discussed the outcomes. All recordings were transcribed and coded by the first author and at least one other author. Transcripts were independently coded by the first author and at least one other team member, with at least one of them having directly hosted the interviews and co-design sessions to ensure sufficient familiarity with the data. Using an inductive coding approach, codes were generated directly from participants' responses to capture the nuanced meanings inherent in the data. Rather than applying predetermined categories, we allowed codes to emerge naturally, ensuring that our

Index	Age	Gender	Major	Occupation	Region	Training in Qualitative Analysis	Knowledge of Qualitative Analysis
				Pilot study			
1	25-30	Male	Healthcare	Industry	East Asia	✓	Passing Knowledge
2	31-35	Male	IT Consulting Services	Industry	East Asia	✓	Knowledgeable
3	18-24	Female	Enological Engineering	Undergraduate Student - Junior	East Asia	×	Passing Knowledge
4	25-30	Female	Electronic Engineering	Graduate Student - PhD	East Asia	✓	Passing Knowledge
				Formal Study			
5	25-30	Female	Human-Computer Interaction	Graduate Student - PhD	North America	✓	Knowledgeable
6	25-30	Female	Communication	Graduate Student - PhD	North America	✓	Knowledgeable*
7	18-24	Female	Design	Graduate Student - Master	East Asia	✓	Passing Knowledge
8	25-30	Female	Human-Computer Interaction	Graduate Student - PhD	East Asia	✓	Knowledgeable
9	25-30	Male	Information Science	Graduate Student - PhD	North America	✓	Knowledgeable
10	25-30	Female	Learning, Design and Technology	Graduate Student - PhD	North America	✓	Knowledgeable
11	18-24	Male	Computer Science	Graduate Student - PhD	East Asia	✓	Knowledgeable
12	25-30	Female	Human-Computer Interaction	Graduate Student - PhD	North America	✓	Knowledgeable
13	31-35	Female	Human-Computer Interaction	Graduate Student - PhD	North America	✓	Knowledgeable
14	25-30	Female	Information Science	Industry - PhD	North America	✓	Expert
15	25-30	Male	Human-Computer Interaction	Graduate Student - PhD	North America	✓	Knowledgeable
16	25-30	Female	Information Science	Graduate Student - PhD	North America	✓	Knowledgeable
17	25-30	Female	History	Graduate Student - PhD	North America	✓	Knowledgeable

* Researchers who specialize in the prompt design.
Passing Knowledge: Having participated as a researcher in at least one qualitative research project.
Knowledgeable: Published one or more peer-reviewed academic papers based on qualitative methods as the primary author (first author).
Expert: Primarily conducting research using qualitative methods and holding a doctoral degree.

Table 1: Demographics

analysis remained grounded in the data. Our coding approach was inherently inductive, with codes emerging directly from the data rather than being imposed by pre-existing theoretical constructs. Subsequently, we explored the relationships between the codes and aggregated them into initial themes, an iterative process that evolved as the study progressed. Throughout this iterative process, the researchers met at least once a week to discuss progress, the outcomes of the interviews and co-design sessions, as well as any findings and issues, and to continuously refine the themes and process. Through these discussions, the researchers reached consensus to resolve discrepancies in coding and theme interpretation, which were then synthesized into our findings.

4. Users' Experiences and Challenges with ChatGPT

Participants expressed significant concerns about ChatGPT's transparency (interpretability and verifiability), performance (consistency and accuracy), the difficulty of designing prompts, and the cost of reviewing results. Table 2 summarizes the key challenges reported by the formal study participants (P5–P17), along with the proportion of participants who mentioned each challenge. These findings further guided our subsequent design efforts.

Challenges	Participants	Count
Lack of Transparency	All participants	13/13
Difficulty of Designing Prompts	P7, P8, P9, P10, P11, P12, P13, P14, P16, P17	10/13
Insufficient Understanding of ChatGPT's Capabilities	P5, P7, P9, P10, P12, P13, P15, P16, P17	9/13
Demand for Customized Solutions	P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, P15, P16	12/13

Table 2: Challenges in Using ChatGPT for Qualitative Analysis Tasks Reported by Formal Study Participants (13 Participants, P5–P17)

4.1. *Lack of Transparency*

Participants' initial skepticism towards using ChatGPT for qualitative analysis was largely rooted in concerns about the lack of transparency in its outputs. As P13 mentioned:

> "*It's so convenient, both convenient and scary. Although it's convenient and might be smarter than me, I still don't know who it is, because I can't identify it just using keywords.*" (P13)

P13 highlights the unease researchers felt about relying on an AI tool without clearly understanding how it generated its responses.

4.2. *Difficulty of Designing Prompts*

Participants reported that the current cost of designing prompts and learning how to design prompts is high. The internet is flooded with many prompt design schemes and tutorials, often confusing users about where to start. As P8 mentioned, "*There are too many junk prompts [tutorials/examples] online right now.*"

Additionally, when users design prompts, it frequently necessitates multiple attempts, the outcomes of which appear to be random. Also, the efficiency and accuracy with which ChatGPT operates is inextricably linked to the caliber of the prompts it receives. On the one hand, ChatGPT offers the allure of automated and efficient outputs, while on the other, the effort to finely tune and optimize these prompts often verges on the boundary of practicality. There is a need to stick a balance between automated ease and manual customization. Several participants encapsulated this sentiment:

> "*I need to test it for a while. For instance, using ChatGPT, various prompts might lead to divergent outcomes, and I might have to repeatedly test to determine which type of prompt elicits the desired result.*" (P14)

> "*The process (of designing prompts) is also a bit time-consuming, and it might be easier for me to do some coding work manually.*" (P8)

The high overhead of designing effective prompts may cause potential users to default to their manual processes, thus underscoring the pressing need for a more streamlined and intuitive prompt-design framework.

4.3. Insufficient Understanding of ChatGPT's Capabilities

Participants' lack of knowledge about ChatGPT's capabilities or incorrect usage can lead to a decline in performance and discourage them from using the tool. As previously mentioned, the learning cost associated with understanding ChatGPT's capabilities is substantial in the context of information overload. Participants in our study were neither expert users of ChatGPT nor researchers in related applications, and their understanding of its capabilities was limited.

> "*I didn't know it could generate tables before.*" (P17)

> "*When I used it before, I didn't know what I wanted... I didn't know what ChatGPT could do, so I couldn't maximize its performance. Using ChatGPT still feels like using a Swiss Army knife for work, not very efficient.*" (P15).

> "*I'm quite shocked. I felt I didn't expect it to have these capabilities, because I hadn't tried asking in so many ways before.*" (P12)

> "*At first, I didn't expect ChatGPT to generate such a result for me, but I find it quite aligned with the ideal result.*" (P7)

The feedback from participants directly reflects the awkward situation of limited-use scenarios due to a lack of understanding of ChatGPT's capabilities. P15 used a current example to analogize this awkward situation, indicating that even when using the "most powerful" AI systems to date, they can only be effectively applied to tasks if the users understand how to utilize them.

4.4. Demand for Customized Solutions

Although our findings emphasize the importance of being informed about ChatGPT's capabilities, this does not mean that users need to understand all of its functions completely. In other words, what users more urgently need is knowledge of how to use ChatGPT to better accomplish the tasks they are truly concerned with. This was manifested in our study as a demand for customized solutions. After introducing some of ChatGPT's capabilities and suggestions on how to use it for processing qualitative data to the participants, they provided positive feedback. This does not necessarily indicate their recognition of ChatGPT's abilities in other domains, but it did increase their understanding of its capabilities in handling qualitative data. Some illustrative quotes are as follows:

> "*I think especially for these prompts. You know, each [task's] topic might be different, like the prompts for programming and the prompts for qualitative research, they might all be different. I think at least for my future qualitative research, I will definitely follow the steps you suggested for writing prompts. They are indeed more specific than what I used to write.*" (P10)

Furthermore, P7 highlighted the need for a standardized yet flexible framework, offering clear, concise guidance on input formats and expected outcomes. This approach would streamline prompt design, ensuring efficiency in interactions with AI tools like ChatGPT and potentially higher quality in outcomes.

> "*I think there can be a more standardized [process] because everyone has a procedure when doing qualitative analysis... Maybe the software tutorial will tell me what to input or in what format to achieve a certain result. Just present these processes to me. There's no need to be overly detailed...*" (P7)

The findings discussed above highlight the challenges participants faced when using ChatGPT for qualitative research analysis and emphasize the need for a standardized yet flexible framework that provides suggestions and guidance on effectively utilizing ChatGPT. Participants' demands for customized solutions and clear guidance on input formats and expected outcomes underscore the importance of developing a framework that addresses these concerns. To bridge this gap and create a framework that meets the needs of qualitative researchers, we further collaborated with participants and engaged in an iterative analysis and design process. This process, detailed in the following section, aimed to develop a framework that provides structured guidance while maintaining flexibility, ultimately enabling researchers to effectively harness ChatGPT's power for their qualitative analysis tasks.

5. Analyses of the Design Process

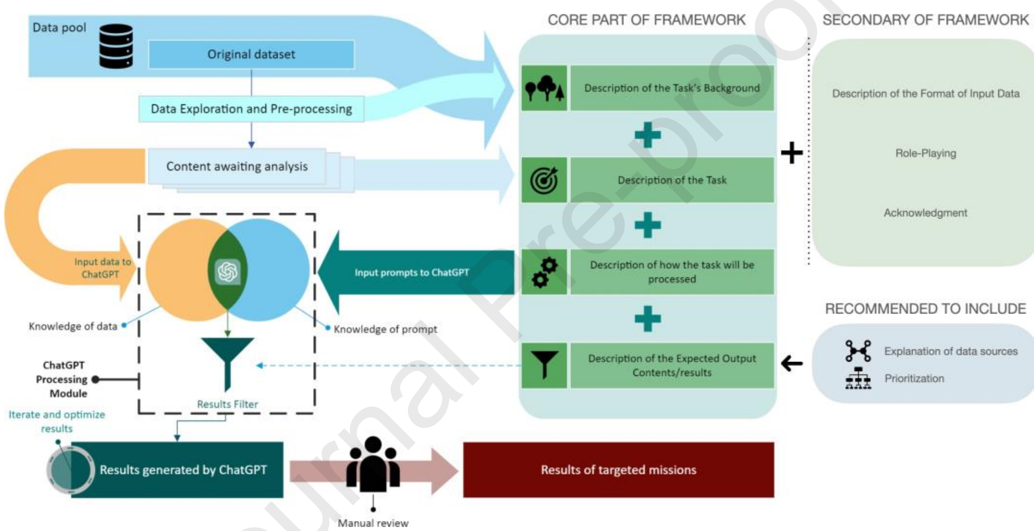


Figure 1: A suggested workflow for applying ChatGPT to handle qualitative analysis tasks. The core of the (prompt design) framework includes descriptions of tasks (including methods), task backgrounds, and output format, enabling ChatGPT to analyze input data with strong robustness. The secondary part of the framework includes descriptions of data structure, role-playing, and friendly wording, which can further enhance the robustness of ChatGPT in task processing.

Inputting structured prompts is crucial for maintaining high performance in ChatGPT [92, 93]. Although the phrasing of the prompts can vary, they should contain fundamental information that forms a core element of the pattern [33]. Hence, we summarize the key elements, features, rules, and strategies possessed by the prompts used in achieving the participants' desired (satisfactory) outcomes, as shown in Table 3. This includes strategies in the independent coding process of participants, as well as various prompt design methods and testing strategies proposed by researchers based on the goals expected by the participants, which are considered effective. The specific process to determine these features was as follows: (1) We analyzed the results generated by ChatGPT for each participant during the formal study. (2) We selected the output results that participants deemed satisfactory, which they believed could be used directly as results, or had completed some important qualitative analysis tasks, such as thematic analysis. (3) We reviewed the prompts that led to the corresponding output results and conducted a comprehensive analysis with the interview content to determine the specific feature categories.

Participant No. (Index)	Background / Conceptual Understanding	Focus on Methodology (Goal of Task)	Focus on Analytical Process (context)	Data Format (Inputs)	Data Format (Outputs)	Role-Playing	Prioritization	Transparency & Traceability	Acknowledgment of Expertise
5	•	•	•	0	0	0	•	•	0
6	•	•	0	0	0	•	•	•	0
7	•	•	•	0	0	•	•	•	0
8	•	•	0	Ot	Ot	•	0	•	•
9	•	•	0	Ot	Ot	0	•	•	0
10	•	•	•	0	0	0	•	•	0
11	•	•	•	0	tO	0	•	•	0
12	•	•	•	Ot	Ot	0	0	•	0
13	•	•	•	0	0	•	•	•	0
14	•	•	•	Ot	Ot	0	•	•	0
15	•	•	•	Ot	Ot	0	•	•	0
16	•	•	•	0	0	0	0	•	0
17	•	•	0	0	tO	•	0	•	•

0: a strategy not employed by the participant in crafting the prompt.
• : a strategy utilized by the participant in crafting the prompt.
Ot: a strategy adopted by the participant following the researcher's suggestion.

Table 3: Summary of Participant Design Prompts

Please note that the strategies presented in Tables 3 and A.4 are the results of summaries, and the prompts used by participants were not the same. Furthermore, these strategies are interconnected in subtle ways, which we will explore in greater detail in the following sections. At the core of these strategies is the iterative nature of prompt design, a foundational concept for interacting with LLMs like ChatGPT. Participants recognized the conversational and flexible nature of ChatGPT, as P7 noted:

> *"I feel the advantage of using ChatGPT lies in this conversational aspect, where depending on your input, you might get slightly different outcomes every time."*

This iterative approach enables users to refine prompts to suit various needs, offering explanations, elaborating queries, or facilitating knowledge transfer. It allows them to critically evaluate and adjust outputs, incorporating their creative thinking into the interaction process [94]. For example, while detailed procedures were helpful in better utilization, participants preferred not to confine their interactions to rigid commands, making the framework more versatile and adaptable to their unique tasks.

The iterative process ensures that ChatGPT's outputs align with research objectives. However, as P8 pointed out:

> *"[With multiple iterations,] it might lose focus on the original task or context."*

The framework emphasizes that the LLM's objectives and context must be repeatedly reinforced to maintain consistency during iterative prompt refinement. This process is further enriched by acknowledging ChatGPT's expertise, creating a more stable interaction model while enhancing user engagement and trust in the system. As such, iteration is not a standalone element but an integral part of the overall framework, emphasizing that each refinement cycle introduces new opportunities for critical assessment and improved output alignment.

By framing prompt design as a cyclic process, we underscore how iterative refinement and acknowledgment of expertise work together to improve the effectiveness and reliability of ChatGPT in qualitative analysis. Each iteration helps address nuances and ensures that strategies within the framework remain interconnected, supporting the ultimate goal of producing meaningful and consistent outputs.

5.1. Explanation of Prompts Design Ideas

In this section, we will elaborate on the characteristics and reasons that should be considered in a prompt design, where the prompts used in the formal study will be shown in Table 3 and A.4.

5.1.1. A Good Prompt Design Should Have: Background or Conceptual Understanding

In qualitative research, researchers start by familiarizing themselves with the data's context to gain a nuanced understanding of the subject matter. Similarly, when using ChatGPT for thematic analysis, providing a clear and descriptive background in the prompts serves as a foundation for accurate analysis. By offering ChatGPT a comprehensive task description—including the purpose, desired outcomes, and specific nuances—this prompt strategy mirrors the familiarization stage in qualitative studies, where context is critical for meaningful interpretation.

> *"(Example of prompts) "Here is a transcript from a focus group interview about 'Transitioning to Remote Work'. Each paragraph is from one participant. Please read it first." (P9)*

This approach guides the ChatGPT in generating targeted responses, enhancing the quality and relevance of its outputs. When ChatGPT understands the overarching task goals and background, it is better equipped to perform nuanced analysis, thereby increasing the likelihood of successful data interpretation and reducing the need for extensive post-processing.

5.1.2. A Good Prompt Design Should Have: a Description of the Methodology (Goal of Task)

Defining the task methodology within prompts is akin to setting clear research objectives in qualitative studies, where researchers specify their analysis goals and approach. In this study, prompts need to detail the exact task that ChatGPT was to perform, namely, analyzing qualitative data related to the theme of "remote work." This prompt structure reflects the goal-setting phase in qualitative research, where clarity about objectives helps focus the analysis.

> *"(Example of prompts) "can you do a thematic analysis of their responses? " (P6)*

For example, participants in this study were instructed to specify that ChatGPT should identify patterns and themes within the data. Testing showed that detailed task descriptions improved ChatGPT's effectiveness, as noted by participant P6, who mentioned that their familiarity with prompt engineering led them to articulate specific requests from the start.

> "*Because I am familiar with things like ChatGPT's prompt engineering, I would write my request more specifically from the beginning. (P6)*"

Clearly stating the research goal within prompts ensures that ChatGPT stays aligned with the intended analytical purpose, much as goal-setting does for researchers.

5.1.3. A Good Prompt Design Should Have: a Description of the Analytical Process

Instructing ChatGPT on the desired analytical process parallels the coding and thematic development stage in qualitative analysis, where specific frameworks or methods direct how data is interpreted. Adding guidance in the prompts, such as P16 specifying that ChatGPT should use the Job Demands-Resources Model for analysis, enables the AI to apply a structured methodology, thus ensuring alignment with research frameworks.

> *"(Example of prompts) Is there any shortcomings or benefits can be mapped to personal demands or personal resources in the JD-R mode." (P16)*

P9's prompts could direct ChatGPT to review the entire dataset, identify recurring themes, cluster these into broader categories, and present each with representative data excerpts and a summary. This process guidance ensures that outputs are consistent with qualitative research principles, producing robust, methodologically sound results.

> *"(Example of prompts) Please do a thematic analysis and summarize no more than 10 themes from this transcript." (P9)*

5.1.4. A Good Prompt Design Should Have: a Definition of the Input Data Format

In qualitative analysis, data preparation is essential for effective interpretation. ChatGPT performs better when analyzing formatted data than disorganized data [93], indicating that pre-cleaning data is necessary. Cleaning the data beforehand ensures that ChatGPT is processing only the most pertinent and reliable information, thereby maximizing the value of its analytical capabilities. However, thanks to ChatGPT's capabilities in understanding context and its overall robustness [95], the preparation of formatted datasets can be less stringent compared to traditional data cleaning methods [96, 97]. The most critical aspect in preparing the dataset for ChatGPT is differentiating the information ownership, i.e., who said what.

> *"(Example of prompts) The format of the transcript looks like this: Participant <participant name>: <transcript of this participant's comments regarding transitioning to remote work>." (P15)*

In addition to using a structured corpus, describing the nature and structure of the input data within the prompts is equally important. The prompt should clearly express the type of data ChatGPT will be analyzing, and we should elucidate the features of the data, such as its conversational structure, data structures entered into ChatGPT, and the data's complexity, like the length and roles.

In our study, although not all participants used this prompt rule, it can effectively avoid potential errors such as discontinuity issues arising from multiple segmented inputs and misunderstandings of the corpus.

5.1.5. A Good Prompt Design Should Have: a Definition of the Output Format

Apart from defining the input data format, specifying the output content format is equally important. Setting a standardized output format is a common strategy. Such standardized formats enhance the consistency of ChatGPT's task output. Also, this practice reflects the reporting and results presentation stage in qualitative studies, where structured results support efficient analysis and comprehension. Notably, while not all participants explicitly defined the output format in their prompts, the results from the formal study and a user perspective show that satisfactory outcomes have a more readable format, which is deemed beneficial for quick reading and subsequent analysis.

> *(Example of prompts)*
> *"Please analysis data again and make outputs follow format as below:*
> *(New line) Topic: challenges, strategies, benefits, downsides, productivity*
> *(New line) Argument points: reasons, suggestions, perceptions*
> *(New line) Supports: raw data, participant number" (P8)*

In the formal study, some participants adopted the suggestions and requested that each topic be presented alongside relevant excerpts from the input data in a specific manner (as an example shown in Fig. A.3, part of the output results from P15.) and summarized the topics and their significance, which further increased the transparency of the results.

Moreover, specifying the output format, such as a table, enables users to easily transfer it to spreadsheet software like Excel, which can be achieved by selecting and copying the table result from ChatGPT, and then pasting it into an Excel sheet, facilitating further analysis, as demonstrated in the example shown in Fig. A.4.

5.1.6. *A Good Prompt Design Should Have: Role-Playing and Acknowledgment of Expertise*

Role-playing, where ChatGPT assumes an "expert" perspective in qualitative analysis, seems to be a common and effective optimization strategy, and this strategy is somewhat unique to LLM interactions. However, it can be compared to perspective-taking and reflexivity in qualitative research, where researchers adopt different lenses to analyze data comprehensively. Assigning ChatGPT a role, such as a "qualitative analysis expert," helps it focus on performing complex andnuanced tasks with greater accuracy.

> *(Example of prompts) "You are now a research expert in qualitative analysis..." (P7)*

Many previous literature mentioned the use of role-playing to enhance or surpass the performance of ChatGPT [98, 99, 100, 101, 102]. By placing ChatGPT in a specific role, participants could direct its attention to tasks like analyzing corpus data, enhancing both task focus and precision. While other strategies, such as providing detailed instructions, can sometimes surpass the effects of role-playing, this method remains valuable for engaging the model in an "expert" mindset, much like researchers apply their expertise to qualitative analysis.

Complementing role-playing and acknowledgment of expertise involves reinforcing the model's perceived expertise and maintaining a positive interaction environment. This strategy, while not part of the core prompt design, can influence the AI's performance in subsequent interactions by improving user experience and fostering a more stable output model [103]. Participants noted that acknowledging the model's expertise—through positive reinforcement or polite feedback—helped maintain consistency and improved its responsiveness, such as the results obtained by P8 (see Fig. A.6). For instance, P8 emphasized the need to "praise" ChatGPT to encourage smarter responses, believing that a positive environment could mitigate the limitations of the model's memory:

> "*(On the evaluation of the output) So smart! It still needs to be praised. (Explanation) If you compliment it, it might be a bit smarter. Because ChatGPT's memory is limited, you should praise it a bit. Otherwise, it might get 'dumber' the more you use it, so you need to give it a compliment.* (P8)"

However, the practical impact of acknowledgment alone was limited, as seen in P17's practice, where adding more specific strategies, such as standardized output formats, proved more effective. Nonetheless, polite and encouraging words were perceived as contributing to a better "learning environment" for AI, as P17 explained:

> "*Because I think whatever I say to the AI, that's what the AI will eventually become. I want a polite and kind AI. I hope to provide some good data references for training the AI.*"

While acknowledgment of expertise might not directly enhance output quality, its role in creating a positive interaction environment supports long-term benefits, such as improved AI responsiveness, better user experience, and trust in the model's capabilities. By integrating role-playing and acknowledgment of expertise into prompt design, researchers can strategically align input, interaction, and output quality for more effective LLM use.

5.1.7. A Good Prompt Design Should Have: Output Prioritization

Participants had certain priority requirements for the results, mainly due to (1) considerations for higher readability and (2) the elimination of some secondary information, focusing on the analysis of the main themes in the corpus. Directing ChatGPT to prioritize specific aspects of the data mirrors the data reduction stage in qualitative analysis, where researchers focus on key themes and discard peripheral information. This strategy ensures that ChatGPT concentrates on the most relevant data, enhancing the clarity and relevance of its output.

Several participants gave positive feedback on asking ChatGPT to prioritize the results. Taking P10 as an example, she mentioned:

> "*I currently feel that it's a bit too much, and I might not use so many codes.*"

Therefore, the researcher suggested that participants add requirements for the number of codes:

> "*You can tell it the number of codes you want, for example, 'I want a codebook with less than 10 codes.'*"

Fig. A.5 shows the results of ChatGPT's output after P6 used this strategy. She believes that this output result can better pinpoint key codes. At this stage, the practical effect was positive and revealed some users' habits and personalized needs when handling tasks. We hope the proposed framework serves as a reminder for novice users interacting with ChatGPT, helping them recall the traditional analysis process and apply ChatGPT's capabilities to enhance their analytical tasks.

5.1.8. A Good Prompt Design Should Have: Rules to Make the Output Transparent and Traceable

Transparency is crucial in both qualitative research and AI-assisted analysis. In qualitative studies, researchers document analytic decisions to ensure traceability and transparency, allowing others to follow the logic behind the findings.

The lack of transparency is a significant challenge for generative AI [29]. Due to the black box, transparency problems have long been criticized by researchers and users [104]. The interviews in the formal study confirmed that the lack of transparency is one of the main reasons for users to be cautious about using ChatGPT for qualitative data analysis. All participants had concerns about the transparency of the content generated by ChatGPT to varying degrees. Although the black box is not entirely untrustworthy [105], the results of the formal study show that all participants demanded further clarification from ChatGPT to enhance the transparency of the results. Adding explanatory information, traceable sources, and standardized output formats can significantly increase users' trust in ChatGPT results [106]. We've previously mentioned standardized output formats, and another effective method to increase transparency is to provide traceable information sources for output results [107].

When using ChatGPT for qualitative corpus analysis, demanding better interpretability and higher transparency may require just a single sentence, as done by P13. However, combining this prompt with other strategies can achieve better results, achieving better readability as shown in Fig. A.3 (P15) or Fig. A.5 (P6).

In this formal study, we found that asking ChatGPT to analyze each line of data independently, rather than conducting an overall analysis of the input data, is a more effective strategy. Although overall analysis can still produce usable or insightful insights, through comparison, independently analyzing each response plays a more significant role in subsequent in-depth studies and may lead to more discoveries. This doesn't mean that analysis should be conducted without considering context. A comprehensive analysis can be achieved by combining other prompt strategies, such as considering priority.

6. User's Attitude on ChatGPT's Qualitative Analysis Assistance: From No to Yes

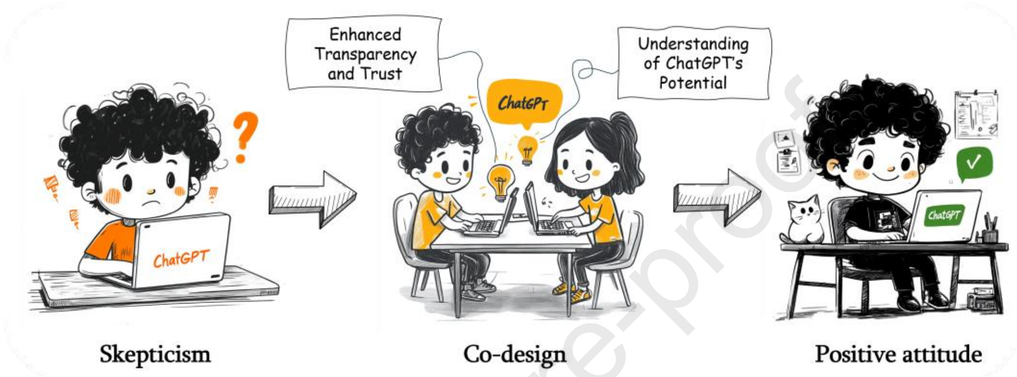


Figure 2: From Skepticism to Positive Attitude: A Schematic Illustration of the User's Attitude Evolution Toward ChatGPT. The figure features three sequential cartoon scenes: the left scene depicts an initial state of doubt or uncertainty, the middle highlights the co-design process, and the right scene portrays a positive attitude after co-design practice.

In the initial phase of our study, participants expressed considerable skepticism about ChatGPT's utility for qualitative analysis. Their concerns primarily centered on issues such as the lack of transparency in output (i.e., difficulties in interpreting and verifying the underlying reasoning), inconsistent performance, the challenge of designing effective prompts, and the substantial effort required to review and refine generated outputs. These concerns not only reflected practical challenges but also raised ethical issues, including potential over-reliance on opaque AI outputs and the risk of inadvertently propagating biases.

However, after engaging in an iterative co-design process, all participants from the formal study demonstrated a consistent trend of shift toward a more positive and accepting attitude. This transformation was driven by two key factors: enhanced transparency (and the resulting trust) and an expanded understanding of ChatGPT's potential. The process of this attitude shift is illustrated in Figure 2.

6.1. Enhanced Transparency and Trust

The systematic application of prompt engineering techniques substantially improved the transparency and traceability of ChatGPT's responses. By clearly linking outputs to specific data sources and clarifying the underlying reasoning, these techniques not only increased participants' confidence but also helped address ethical concerns regarding accountability and bias. For example, one participant remarked:

> "*After designing the prompts step by step, I think both the format and content have been improved, such as the source of the data and the ability to integrate it into a table format... If I were to use ChatGPT for batch coding, I would use such a template [framework].*" (P15)

This quote illustrates how a structured prompt design framework can enhance output transparency and usability, thereby increasing trust in ChatGPT's support for qualitative analysis tasks. Further reinforcing this point, another participant noted:

> "*In terms of preliminary screening and categorization, it has already saved a lot of time, which I think is a very okay process. Another great thing is that I know its data source... I can also trace back to see who made this point and what it originally meant.*" (P8)

Despite these improvements, participants also emphasized the need for human oversight. As one participant stated:

> "*I think the results from ChatGPT are very good. And when presenting these [analysis results], it even connects them with the user, which I find very intuitive. This [format] will increase my trust in ChatGPT. Still, I won't directly state my research results based solely on its findings. I won't do that; I would definitely double-check. [Overall,] I think ChatGPT is exceptionally useful.*" (P5)

This quote illustrates that while the framework significantly improved ChatGPT's transparency and credibility, researchers still valued the role of human judgment in validating the AI's outputs.

6.2. Expanded Understanding of ChatGPT's Potential

The co-design process also enabled participants to gain a deeper understanding of ChatGPT's capabilities, which broadened their awareness of its potential applications in qualitative research. This enhanced familiarity played a crucial role in shifting perceptions. For instance, one participant observed:

> "*From my experience today, I feel its [ChatGPT's] summarization ability is quite strong. Also, I hadn't thought before that qualitative data could be analyzed with it.*" (P7)

Through iterative prompt refinement, participants not only addressed their initial concerns but also discovered that well-designed prompts could unlock a wider range of functionalities from ChatGPT. This experiential learning ultimately contributed to a unanimous shift in attitudes—from skepticism to a more positive, supportive view of ChatGPT as an effective tool to support qualitative analysis. Based on these findings, it is evident that the prompt design framework we developed (presented in Section 5) is beneficial for leveraging LLMs to assist qualitative analysis work.

7. Discussion

Our research results show that LLMs-based applications (such as ChatGPT) have the potential to conduct qualitative analysis on corpora through well-designed prompts, addressing concerns of human analysts. In our study, we integrated established prompt design methodologies [90] to enhance ChatGPT's efficacy in processing structured data. We addressed the challenge of command failures, as highlighted by Zamfirescu-Pereira et al. [90], not by simply appending multiple prompts, but by strategically integrating them and simplifying the interaction process. In the following sections, we will expand on the discussions by Jiang et al. [67] and Feuston and Brubaker [56] about the collaboration between humans and AI in qualitative research, especially for junior researchers. Our inspiration comes from the views of the participants and is drawn from the processes and methods of qualitative research [108].

In addition, we will discuss the ethical considerations of using ChatGPT from the user's perspective and its impact on the conceptual transition process (from rejection to acceptance). We draw inspiration from Idhe's book

"Instrumental Realism" [109], concepts from Husserl [110], and ideas from Foucault [111]. Combining practice, tools, and phenomenology, we discuss how new AI tools like ChatGPT can influence user attitudes by shifting paradigms.

7.1. Overcoming the Challenges in Prompt Design

Our findings highlight the difficulties researchers face when designing prompts for ChatGPT to support qualitative analysis tasks, particularly for junior researchers who lack expertise and experience in prompt design [18, 19]. To address these challenges, we propose a framework that incorporates strategies identified in our findings and aligns these strategies with existing research on prompt engineering techniques [17, 30, 32, 35].

While our framework was initially developed with ChatGPT in mind, we argue that its foundational elements are broadly applicable across different LLMs. These elements provide a scaffold that can adapt to varying models and prompt engineering advancements, making the framework relevant for current and future applications. The framework remains flexible and adaptable by focusing on essential prompt structures rather than model-specific optimizations, thus retaining its long-term value.

The framework-based approach also considers the specific challenges associated with qualitative analysis [57, 58, 52, 53, 55], and aligns closely with traditional qualitative research methodologies, which adds rigor to its design and implementation. By leveraging ChatGPT's capabilities through effective prompt design, researchers can potentially mitigate these challenges and enhance the efficiency of the qualitative analysis process. However, it is crucial to recognize the importance of human oversight and validation in the qualitative analysis process [56, 67, 88, 64, 65]. Our framework provides users with a referenceable guide to understand the role of each component within the prompts, enhancing transparency in the process. This disclosure reinforces the effectiveness of voluntarily sharing information about GenAI without any negative consequences, further contributing to a more transparent workflow [112].

AI developers should consider designing tools that facilitate context-aware prompting, enabling structured definitions of research contexts, analytical objectives, and workflows, and emphasizing model outputs that offer enhanced transparency and clear explanations of reasoning processes. Developers are encouraged to integrate customizable role-playing capabilities, allowing qualitative researchers to define specific analytical roles for ChatGPT, enhancing task-specific relevance. Moreover, systems should accommodate flexible qualitative data input formats to ensure seamless integration into diverse analytical contexts.

Qualitative researchers can focus on clearly structured prompt designs that explicitly articulate research objectives, methodologies, data formats, and desired analytical outputs, thereby enabling targeted and effective AI assistance. Researchers must emphasize the importance of human oversight, recognizing AI-generated insights as preliminary and necessitating further human validation, refinement, and contextualization. Comprehensive documentation of AI interactions, including prompts and model-generated outputs, is crucial to maintaining methodological transparency and analytical rigor. Researchers are advised to adopt iterative prompt refinement strategies, continuously enhancing prompts based on feedback from AI-generated outputs and evolving research needs.

To effectively integrate ChatGPT into qualitative research workflows, researchers should begin by thoroughly defining the research context, including research questions, data characteristics, and analytical methodologies, within prompts. Methodological alignment, explicitly specifying analytical methods such as inductive coding, thematic analysis, or constant comparative analysis, is crucial for aligning AI outputs with qualitative standards. Iterative cycles of prompt testing, response evaluation, and prompt refinement should be standard practice, ensuring progressive improvement of AI-generated outputs. Clearly instructing ChatGPT on its analytical role (e.g., assistant coder, thematic synthesizer) enhances output precision and relevance. Maintaining systematic documentation throughout interactions with ChatGPT supports transparency and ensures replicability. Lastly, implementing a rigorous human-AI validation loop, wherein researchers critically validate, refine, and contextualize AI-generated outputs before finalizing findings, reinforces methodological integrity and the quality of analytical insights.

7.2. Enhancing Junior Researchers' Understanding of ChatGPT's Capabilities and Proficiency in Qualitative Analysis Processes

7.2.1. Theoretical Contributions to AI-Assisted Qualitative Research

The framework aims to make the prompt design process more structured and transparent, enabling researchers to elicit more interpretable and verifiable responses from ChatGPT. This approach addresses concerns regarding the lack of transparency in ChatGPT's outputs [9] and aligns with research on

XAI [81]. Moreover, the framework improves junior researchers' understanding of ChatGPT's capabilities in qualitative analysis by providing guidance on effectively communicating context, methodology, and data formats to the AI [67, 74, 75].

Our framework offers structured guidance that helps junior researchers understand ChatGPT's capabilities in qualitative analysis and gain proficiency in the fundamental processes of qualitative research. By incorporating elements such as context, methodology, data organization, and transparency into prompt design, the framework breaks down complex analysis processes into manageable, structured steps, making qualitative analysis and prompt engineering more accessible for those with limited experience [18]. This approach is especially valuable for early-career researchers who may lack formal training in qualitative methods, offering them practical insights into best practices for data interpretation, theme identification, and methodological rigor.

Furthermore, the framework's emphasis on transparency and traceability encourages junior researchers to engage in a more critical and reflective approach to qualitative analysis and beyond. By making ChatGPT's outputs interpretable and verifiable, the framework cultivates habits of accountability and methodological scrutiny [106, 78], which are essential for high-quality research. Through prompt strategies that elicit clear, structured, and traceable results, junior researchers are empowered to interrogate potential biases in AI outputs and refine their own analytical processes, building confidence [113] in their ability to conduct robust research.

7.2.2. Practical Implications for Research Workflows

Beyond theoretical benefits, our framework offers practical advantages for qualitative research workflows, especially for those new to the field. For junior researchers lacking formal training, the framework serves as a step-by-step guide through complex analytic tasks. It breaks down processes like thematic analysis into manageable steps, from data familiarization to coding and theme development, with ChatGPT assisting at each stage. This guided approach accelerates the learning curve – junior analysts gain hands-on experience by working alongside the AI on real data, which helps solidify their understanding of qualitative techniques. The efficiency gains afforded by ChatGPT (e.g., rapidly generating initial codes or summarizing large text corpora) can significantly reduce the time and labor required for early coding cycles or transcript reviews. These gains lower the barrier to entry for novices, allowing them to focus on higher-level interpretation and reflection sooner.

Beyond merely providing a structural template, our framework fundamentally transforms access to sophisticated research methodologies for junior researchers. By systematically bridging the expertise gap between novice and experienced qualitative researchers, we address a longstanding barrier in the field that has traditionally limited methodological innovation to those with extensive training. This democratization represents a significant paradigm shift in how qualitative research competencies are acquired and applied across disciplines [113].

Moreover, the structured workflow ensures consistency and reduces the risk of novice mistakes by following the prompt framework. Junior researchers are less likely to skip reflexive steps or misapply thematic analysis procedures. The immediate feedback and suggestions from ChatGPT also provide a form of mentorship, reinforcing proper practices (e.g., reminding the user to consider alternative interpretations or ground themes in the data). In sum, the framework improves qualitative workflows by combining the scalability and speed of AI with the reflective and iterative nature of human analysis, resulting in a more efficient yet methodologically sound process.

7.2.3. Practical Applications of the Framework for Interdisciplinary Research Management and Collaboration

While the framework was designed specifically to support qualitative analysis, ChatGPT's integration within this structure demonstrates a versatility that extends beyond this specific domain. By adapting the principles of structured prompt design to specific contexts, users can leverage ChatGPT's capabilities to enhance workflows, reduce cognitive load, and foster innovation across diverse disciplines. For researchers, this broad potential underscores the importance of developing transferable skills in AI-assisted methodologies, enabling them to navigate the increasingly interdisciplinary nature of modern research. The strategies outlined, such as iterative refinement, structured input, and transparent methodologies, have potential applications in other areas of research and practice. For instance, ChatGPT has shown promise in facilitating quantitative analysis, where it can assist with data cleaning, interpretation of results, and automating statistical explanations [114]. Similarly, its use in programming has enabled efficient debugging, code generation, and documentation [31, 19], while in creative writing, ChatGPT supports tasks such as generating ideas, composing drafts, and refining language for specific audiences [115, 112, 36, 39, 25]. Additionally, while much of the focus in AI research is on STEM fields, there is a growing recognition of the need to extend AI applications, like ChatGPT, to the humanities. Researchers from non-STEM disciplines often find their perspectives overlooked in technological methodologies and applications [116]. Incorporating AI into the humanities would not only broaden the scope of interdisciplinary research but also ensure inclusivity and resonance across diverse academic communities.

This cross-domain applicability highlights the broader implications of the framework for researchers and practitioners, which is the most significant contribution lies in its systematic approach to bridging disciplinary divides through methodological translation. By articulating core principles of human-AI collaboration that remain consistent across domains, we provide a common language for researchers from diverse backgrounds to leverage AI capabilities without sacrificing methodological rigor.

In broader terms, by lowering skill barriers, our approach encourages more researchers in fields like design, education, healthcare, and beyond to integrate qualitative insights into their projects, enriching studies that otherwise might rely solely on quantitative data.

7.2.4. Evolving Roles of ChatGPT in Qualitative Research and Future Directions: Tool vs. Co-Researcher

Our work also inspired reflection on the evolving relationship between human researchers and AI tools. An interesting layer of understanding involves the potential roles ChatGPT can assume in qualitative research, either as a tool or as a co-researcher. As a tool, ChatGPT can streamline coding processes, uncovering patterns within data and generating initial insights that researchers can further interpret and validate. This role aligns with traditional uses of AI, emphasizing its utility in automating repetitive tasks and augmenting human analysis [117]. However, treating ChatGPT solely as a tool may lead to over-reliance, where researchers undervalue critical engagement with the AI's outputs [113].

Conversely, considering ChatGPT as a co-researcher invites exploration into its more active involvement in thematic analysis, where it could contribute interpretative suggestions, identify themes, or even challenge researcher assumptions. While this perspective underscores the collaborative potential of human-AI interaction, it raises critical questions about whether AI can truly "understand" qualitative data in the way a human researcher does. Ethical considerations, such as attribution of intellectual contributions, transparency in processes, and the risk of bias, become paramount in this scenario [118]. Future studies should investigate how adopting these roles impacts the quality and depth of qualitative analysis and develop collaboration strategies that align with ethical guidelines.

7.3. Evolving Trust and Ethical Shfit in ChatGPT for Qualitative Research

Our study revealed a significant transformation in participants' attitudes toward ChatGPT's use in qualitative research, evolving from skepticism to cautious acceptance. Initially due to its "black-box" nature (concerns over transparency, consistency, and control), the role of ChatGPT in qualitative analysis was met with skepticism. Participants were worried that the AI's coding and theme generation processes were opaque, conflicting with the qualitative research norms that require transparency and audit trails linking data to findings [119]. The inability to observe or verify ChatGPT's intermediate reasoning made researchers feel a loss of control over their analysis. This led them to question whether AI could handle the nuanced interpretative work of qualitative coding and whether it would adhere to the research principles they valued, such as integrity [120]. By leveraging our structured prompt design framework, users enhanced the clarity and traceability of ChatGPT's responses, gradually adopting a more positive attitude. This transformation reflects the dynamic nature of trust formation in human-AI collaboration.

7.3.1. Trust Formation and Transparency

Based on Mayer et al.'s organizational trust model [121], we can explain this evolution through three key dimensions: ability (competence to perform qualitative analysis), integrity (consistency and transparency of analytical processes), and benevolence (alignment with researcher values). Trust is cognitively driven [122], and our co-design process facilitated trust calibration by making ChatGPT's analytical processes more transparent and traceable. This enabled participants to develop what Muir [123] referred to as "appropriate reliance," rather than blind trust or complete distrust.

Participants' initial ethical concerns focused on "cognitive opacity" [124] - the inability to fully understand how AI-generated responses are produced and how conclusions are derived. This opacity raised legitimate concerns about responsible research practices, particularly in qualitative analysis, where interpretive transparency is crucial [125, 126, 127].

Our framework emphasizes transparency and traceability, addressing these ethical concerns through "meaningful transparency" [128], where transparency extends beyond mere visibility, enabling understanding and accountability for LLM-generated content [129].

Participants experienced what Kizilcec mentioned "*Transparency may promote or erode users' trust in a system by changing beliefs about its trustworthiness*" [130]. This represents a shift from viewing AI ethics through consequentialist frameworks (focusing solely on outcomes) [131, 132] toward procedural ethics (focusing on the fairness and transparency of processes) [133]. The critical factor in this transition was not merely improving ChatGPT's analytical capabilities, but rather enhancing process transparency [134], making visible how conclusions are derived.

7.3.2. Symbiotic Integration and Phenomenological Perspectives

The transformation we observed represents what Orlikowski calls "sociomaterial practices" [135], the recursive intertwining of social and material elements [136, 137] in research. Participants did not simply adopt ChatGPT as a replacement for traditional analysis but integrated it through the combination of traditional and novel analytical approaches, a process akin to the concept of methodological bricolage [138, 139, 140].

Describing the relationship between the human analyst and ChatGPT, one can envisage a comprehensive, symbiotic interaction [141]. In this context, ChatGPT is not just an auxiliary tool. It becomes an extension of the human cognitive process [142], enhancing analytical capabilities while simultaneously reshaping one's epistemological outlook. Trust is cognitively driven [122], and cautious humanization [143] while maintaining control is essential [144, 145].

Examining this relationship through Husserl's phenomenological lens deepens our understanding [110]. Interaction with technology, in this view, is not just transactional but a profound, lived experience. Participants used ChatGPT and engaged with it, evolving their understanding and establishing trust in its analytical capabilities. Foucault's notion of "technologies of the self" further enriches this discussion [111]. Seen through this lens, ChatGPT serves not only as an analytical tool but also as a catalyst for personal and professional transformation. It encourages scholars to reconsider their methodologies and the very limits of knowledge [146]. Learning knowledge from the interaction process and reflecting on it is of great significance, especially for early-career individuals and student groups.

7.3.3. Ethical Dimensions and Future Implications

The future of AI in qualitative research can be further conceptualized through Horvitz's notion of "mixed-initiative" interaction [147], a collaborative process where humans and AI systems contribute complementary strengths [148]. This approach moves beyond the traditional levels of automation paradigm [149], which frames AI as either replacing or assisting humans toward a model of true collaboration. It also calls for advancements in interpretable machine learning design [150], ensuring that AI-generated outcomes not only enhance human capabilities but also remain comprehensible and aligned with human reasoning in a deterministic and human-centered manner.

Informed by the shift in researchers' attitudes as revealed in our findings, it becomes evident that the ethical dimensions of this integration are multifaceted. Nissenbaum's emphasis on context in shaping our ethical expectations from technology is highly pertinent here [151]. The shift from skepticism to acceptance among researchers in our study underscores a growing reliance on tools like ChatGPT. While this reliance speaks to the efficiency and potential of AI, it simultaneously raises concerns about the unintentional reinforcement of biases and the potential narrowing of the scope of qualitative inquiry. The construction and contemplation of ethical expectations for AI is not only timely for novice researchers but also conducive to stimulating more in-depth reflections, particularly in contexts where novices are more susceptible to the immediate influence of ChatGPT's interactions [152].

The ethical implications of this scenario are twofold. First, there is a risk that AI systems, including ChatGPT, might inadvertently perpetuate existing biases in their training data [153]. This highlights the crucial importance of training data, especially in qualitative research, where a nuanced understanding and interpretation of data are key. Biased or toxic training data could lead to the development of an AI that is "unfriendly" and "unethical". As the influence of AI in everyday life increases and human-AI interactions become more frequent, such a "toxic" AI could harm human interests in various ways, including the quality of interaction experiences, emotional responses, and moral perspectives. Second, the growing trust in AI's capabilities may lead to a diminished emphasis on the critical, reflective role traditionally played by human analysts in qualitative research. This shift might result in less depth in analysis, as AI tools may not fully replicate the complex cognitive processes inherent in human analysis [154].

Therefore, our findings call for a balanced approach that ensures AI tools like ChatGPT augment and assist in research without overshadowing the indispensable human elements of intuition, skepticism, and ethical judgment that are crucial in qualitative analysis. Learning and reflecting through this interaction process is of great significance, especially for early-career individuals and student groups.

8. Limitations and More Future Work

This study primarily explores ChatGPT's effectiveness in analyzing text-based qualitative data, focusing on tasks like coding. However, several limitations should be acknowledged.

First, due to accessibility constraints, we utilized GPT-3.5 instead of the more advanced GPT-4. While this may impact our findings, as GPT-4 or newer versions offer enhanced reasoning capabilities, larger context windows, and improved accuracy for complex analytical tasks that could potentially provide greater depth and precision [155], it is important to note that a key contribution of our work, the framework for prompts design, has enduring value for tools like ChatGPT that rely on prompt engineering as a core interaction method. This framework remains applicable regardless of model version and provides a possible guidance for effective qualitative analysis with LLMs.

Second, qualitative research often encompasses a broader range of data types and methods, such as observational studies and visual analysis, which were beyond our study's scope. Future research is needed to investigate how LLMs like ChatGPT can be adapted or extended to support these other qualitative approaches, providing a more comprehensive understanding of their potential across diverse qualitative methods.

Third, our participant sample primarily consisted of individuals already experienced in qualitative analysis and ChatGPT, potentially limiting the generalizability of our findings [10, 156]. While this selection ensured participants could provide informed insights on integrating AI tools into qualitative research, it may also introduce biases regarding the framework's perceived utility. Although we identified how varying levels of prior ChatGPT exposure influenced attitudes toward LLM-assisted analysis—a key finding—our sampling approach may not adequately represent novice users who lack qualitative analysis experience or ChatGPT familiarity. These users might require additional foundational knowledge to effectively implement our prompts framework despite its step-by-step guidance. Furthermore, despite including participants from multiple regions, certain backgrouind (culture, gender, age, education) remained underrepresented, which could introducing potential biases.

For a broader future work, in addition to the points already discussed in Section 7, we recognize one of the aspirations expressed by qualitative analysts (P7, P9, P15, P16) is the creation of an integrated LLMs toolkit tailored specifically for qualitative analysis. We have taken this recommendation into account for our future work trajectory. Furthermore, insights from one of our study participants (P4) regarding expectations for future iterations of ChatGPT have been enlightening. While our results show that ChatGPT can effectively analyze qualitative data, it's crucial to acknowledge the vast diversity in user backgrounds. Coupled with the complexities tied to each cultural context, it underscores the importance of AI technologies being attuned to regional and cultural nuances. Such cultural cognizance not only ensures the pertinence of the produced content but also augments user trust and engagement. Moving forward, there is potential to develop a version of ChatGPT that amalgamates personalized knowledge bases, thereby catering to the distinct needs of varied global communities.

9. Conclusion

This study first identified the risks and challenges of ChatGPT in qualitative analysis through a pilot study. It then explored the attitudes of a group of qualitative analyst towards the application of ChatGPT in qualitative research through interviews and co-design, and in collaboration with this study group, developed a well-received framework for prompts design. Our research explored the powerful capabilities of AI in qualitative analysis using ChatGPT, which could potentially significantly reduce the labor-intensive tasks and coding costs in qualitative analysis in the future. Our research findings indicate that enhancing transparency, providing guidance on prompts, and strengthening users' understanding of LLM capabilities can significantly improve user interaction with ChatGPT and reverse negative attitudes towards using such applications in research. In the discussion, we focus on the challenges, potential, and impact on novice researchers associated with the application of ChatGPT, based on our findings. Furthermore, we delve into the ethical considerations brought about by advanced AI, especially in the context of qualitative analysis, starting from the shift in user attitudes observed during the research process. In the broader context of applying LLM workflows, the proposed framework can serve as an essential reference, supporting researchers in continuously developing more personalized application solutions. We proposed several pressing future works to further expand, delve into, and understand LLMs, providing insights Moreover, this framework facilitates a deeper understanding among users regarding the underlying mechanisms of tasks performed by LLMs, encouraging users to actively engage with and critically assess LLM-generated outputs rather than simply relying on AI-produced results. We hope that this study can help users better apply new technologies to enhance efficiency, and we believe this task-oriented framework approach has potential generalizability beyond qualitative analysis, and could be adapted and applied across various scenarios involving the use of LLMs.

Acknowledgments

We would like to take this opportunity to express our gratitude to the participants of this study for their valuable insights. We would also like to clarify that we used ChatGPT 4o and Claude 3.5 Sonnet solely for spell-checking and grammar revisions to enhance content clarity. Our research team assumes full responsibility for the content of this publication. We did not use generative AI for data collection, analysis, or image generation.

CRediT authorship contribution statement

Anonymous

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

References

[1] V. Braun, V. Clarke, Thematic analysis., American Psychological Association, 2012. doi:10.1037/13620-004.

[2] V. Braun, V. Clarke, Using thematic analysis in psychology, Qualitative research in psychology 3 (2) (2006) 77–101. doi:10.1191/1478088706qp063oa.

[3] L. S. Nowell, J. M. Norris, D. E. White, N. J. Moules, Thematic analysis: Striving to meet the trustworthiness criteria, International journal of qualitative methods 16 (1) (2017) 1609406917733847. doi:10.1177/1609406917733847.

[4] P. Bazeley, Qualitative Data Analysis: Practical Strategies, SAGE Publications, 2013.
URL https://books.google.com/books?id=33BEAgAAQBAJ

[5] T. H. Davenport, R. Ronanki, et al., Artificial intelligence for the real world, Harvard business review 96 (1) (2018) 108–116.
URL https://www.bizjournals.com/boston/news/2018/01/09/hbr-artificial-intelligence-for-the-real-world.html

[6] Y. K. Dwivedi, L. Hughes, E. Ismagilova, G. Aarts, C. Coombs, T. Crick, Y. Duan, R. Dwivedi, J. Edwards, A. Eirug, V. Galanos, P. V. Ilavarasan, M. Janssen, P. Jones, A. K. Kar, H. Kizgin, B. Kronemann, B. Lal, B. Lucini, R. Medaglia, K. Le Meunier-FitzHugh, L. C. Le Meunier-FitzHugh, S. Misra, E. Mogaji, S. K. Sharma, J. B. Singh, V. Raghavan, R. Raman, N. P. Rana, S. Samothrakis, J. Spencer, K. Tamilmani, A. Tubadji, P. Walton, M. D. Williams, Artificial intelligence (ai): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy, International Journal of Information Management 57 (2021) 101994. doi:https://doi.org/10.1016/j.ijinfomgt.2019.08.002.
URL https://www.sciencedirect.com/science/article/pii/S026840121930917X

[7] H. Dang, S. Goller, F. Lehmann, D. Buschek, Choice over control: How users write with large language models using diegetic and non-diegetic

prompting, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, 2023. doi:10.1145/3544548.3580969.
URL https://doi.org/10.1145/3544548.3580969

[8] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., Language models are unsupervised multitask learners, OpenAI blog 1 (8) (2019) 9.
URL https://api.semanticscholar.org/CorpusID:160025533

[9] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang, W. Ye, Y. Zhang, Y. Chang, P. S. Yu, Q. Yang, X. Xie, A survey on evaluation of large language models, ACM Trans. Intell. Syst. Technol. 15 (3) (mar 2024). doi:10.1145/3641289.
URL https://doi.org/10.1145/3641289

[10] J. Wachinger, K. Bärnighausen, L. N. Schäfer, K. Scott, S. A. McMahon, Prompts, pearls, imperfections: comparing chatgpt and a human researcher in qualitative data analysis, Qualitative Health Research (2024) 10497323241244669doi:https://doi.org/10.1177/10497323241244669.

[11] Z. Xiao, X. Yuan, Q. V. Liao, R. Abdelghani, P.-Y. Oudeyer, Supporting qualitative analysis with large language models: Combining codebook with gpt-3 for deductive coding, in: Companion Proceedings of the 28th International Conference on Intelligent User Interfaces, IUI '23 Companion, Association for Computing Machinery, New York, NY, USA, 2023, p. 75–78. doi:10.1145/3581754.3584136.
URL https://doi.org/10.1145/3581754.3584136

[12] R. Bijker, S. S. Merkouris, N. A. Dowling, S. N. Rodda, Chatgpt for automated qualitative research: content analysis, Journal of medical Internet research 26 (2024) e59050. doi:https://doi.org/10.2196/59050.

[13] S.-C. Dai, A. Xiong, L.-W. Ku, Llm-in-the-loop: Leveraging large language model for thematic analysis (2023). arXiv:2310.15100.
URL https://arxiv.org/abs/2310.15100

[14] C. Ziems, W. Held, O. Shaikh, J. Chen, Z. Zhang, D. Yang, Can large language models transform computational social science?, Computational Linguistics 50 (1) (2024) 237–291.

[15] W. S. Mathis, S. Zhao, N. Pratt, J. Weleff, S. De Paoli, Inductive thematic analysis of healthcare qualitative interviews using open-source large language models: How does it compare to traditional methods?, Computer Methods and Programs in Biomedicine 255 (2024) 108356. doi:https://doi.org/10.1016/j.cmpb.2024.108356.

[16] R. H. Tai, L. R. Bentley, X. Xia, J. M. Sitt, S. C. Fankhauser, A. M. Chicas-Mosier, B. G. Monteith, An examination of the use of large language models to aid analysis of textual data, International Journal of Qualitative Methods 23 (2024) 16094069241231168. doi:https://doi.org/10.1177/16094069241231168.

[17] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, in: H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, H. Lin (Eds.), Advances in Neural Information Processing Systems, Vol. 33, Curran Associates, Inc., 2020, pp. 1877–1901.
URL https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[18] J. Zamfirescu-Pereira, R. Y. Wong, B. Hartmann, Q. Yang, Why johnny can't prompt: How non-ai experts try (and fail) to design llm prompts, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, 2023. doi:10.1145/3544548.3581388.
URL https://doi.org/10.1145/3544548.3581388

[19] A. J. Fiannaca, C. Kulkarni, C. J. Cai, M. Terry, Programming without a Programming Language: Challenges and Opportunities for Designing Developer Tools for Prompt Programming, in: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, ACM, pp. 1–7. doi:10.1145/3544549.3585737.

[20] A. Bryman, Social Research Methods, Oxford University Press, 2016. URL https://books.google.com/books?id=N2zQCgAAQBAJ

[21] A. Bahrini, M. Khamoshifar, H. Abbasimehr, R. J. Riggs, M. Esmaeili, R. M. Majdabadkohne, M. Pasehvar, Chatgpt: Applications, opportunities, and threats, in: 2023 Systems and Information Engineering Design Symposium (SIEDS), IEEE, Charlottesville, VA, USA, 2023, pp. 274–279. doi:10.1109/SIEDS58326.2023.10137850.

[22] T. Teubner, C. M. Flath, C. Weinhardt, W. van der Aalst, O. Hinz, Welcome to the era of chatgpt et al. the prospects of large language models, Business & Information Systems Engineering 65 (2) (2023) 95–101. doi:10.1007/s12599-023-00795-x.

[23] M. Liebrenz, R. Schleifer, A. Buadze, D. Bhugra, A. Smith, Generating scholarly content with chatgpt: ethical challenges for medical publishing, The Lancet Digital Health 5 (3) (2023) e105–e106. doi:10.1016/S2589-7500(23)00019-5.

[24] H. Alkaissi, S. I. McFarlane, Artificial hallucinations in chatgpt: implications in scientific writing, Cureus 15 (2) (2023) 1–4. doi:10.7759/cureus.35179.

[25] L. Bishop, A computer wrote this paper: What chatgpt means for education, research, and writing, Research, and Writing (January 26, 2023) (2023). doi:10.2139/ssrn.4338981.

[26] C. Macdonald, D. Adeloye, A. Sheikh, I. Rudan, Can chatgpt draft a research article? an example of population-level vaccine effectiveness analysis, Journal of global health 13 (2023). doi:10.7189/jogh.13.01003.

[27] Y. Shen, L. Heacock, J. Elias, K. D. Hentel, B. Reig, G. Shih, L. Moy, Chatgpt and other large language models are double-edged swords (2023). doi:10.1148/radiol.230163.

[28] H. Hassani, E. S. Silva, The role of chatgpt in data science: how ai-assisted conversational interfaces are revolutionizing the field, Big data and cognitive computing 7 (2) (2023) 62. doi:10.3390/bdcc7020062.

[29] Y. K. Dwivedi, N. Kshetri, L. Hughes, E. L. Slade, A. Jeyaraj, A. K. Kar, A. M. Baabdullah, A. Koohang, V. Raghavan, M. Ahuja, H. Albanna, M. A. Albashrawi, A. S. Al-Busaidi, J. Balakrishnan, Y. Barlette, S. Basu, I. Bose, L. Brooks, D. Buhalis, L. Carter, S. Chowdhury, T. Crick, S. W. Cunningham, G. H. Davies, R. M. Davison, R. Dé, D. Dennehy, Y. Duan, R. Dubey, R. Dwivedi, J. S. Edwards, C. Flavián, R. Gauld, V. Grover, M.-C. Hu, M. Janssen, P. Jones, I. Junglas, S. Khorana, S. Kraus, K. R. Larsen, P. Latreille, S. Laumer, F. T. Malik, A. Mardani, M. Mariani, S. Mithas, E. Mogaji, J. H. Nord, S. O'Connor, F. Okumus, M. Pagani, N. Pandey, S. Papagiannidis, I. O. Pappas, N. Pathak, J. Pries-Heje, R. Raman, N. P. Rana, S.-V. Rehm, S. Ribeiro-Navarrete, A. Richter, F. Rowe, S. Sarker, B. C. Stahl, M. K. Tiwari, W. van der Aalst, V. Venkatesh, G. Viglia, M. Wade, P. Walton, J. Wirtz, R. Wright, Opinion paper: "so what if chatgpt wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy, International Journal of Information Management 71 (2023) 102642. doi:https://doi.org/10.1016/j.ijinfomgt.2023.102642.
URL https://www.sciencedirect.com/science/article/pii/S0268401223000233

[30] Z. Zhao, E. Wallace, S. Feng, D. Klein, S. Singh, Calibrate Before Use: Improving Few-shot Performance of Language Models, in: Proceedings of the 38th International Conference on Machine Learning, PMLR, pp. 12697–12706.
URL https://proceedings.mlr.press/v139/zhao21c.html

[31] E. Jiang, E. Toh, A. Molina, K. Olson, C. Kayacik, A. Donsbach, C. J. Cai, M. Terry, Discovering the syntax and strategies of natural language programming with generative language models, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22, Association for Computing Machinery, New York, NY, USA, 2022. doi:10.1145/3491102.3501870.
URL https://doi.org/10.1145/3491102.3501870

[32] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. V. Le, D. Zhou, Chain-of-Thought Prompting Elicits Reasoning in Large Language Models 35 24824–24837.

URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html

[33] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, D. C. Schmidt, A prompt pattern catalog to enhance prompt engineering with chatgpt (2023). arXiv:2302.11382.

[34] S. Linzbach, T. Tressel, L. Kallmeyer, S. Dietze, H. Jabeen, Decoding prompt syntax: Analysing its impact on knowledge retrieval in large language models, in: Companion Proceedings of the ACM Web Conference 2023, WWW '23 Companion, Association for Computing Machinery, New York, NY, USA, 2023, p. 1145–1149. doi:10.1145/3543873.3587655.
URL https://doi.org/10.1145/3543873.3587655

[35] A. Gao, Prompt Engineering for Large Language Models. doi:10.2139/ssrn.4504303.

[36] L. E. Nacke, How to write better chi papers (with ai), in: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '24, Association for Computing Machinery, New York, NY, USA, 2024. doi:10.1145/3613905.3636272.
URL https://doi.org/10.1145/3613905.3636272

[37] L. Wang, X. Chen, X. Deng, H. Wen, M. You, W. Liu, Q. Li, J. Li, Prompt engineering in consistency and reliability with the evidence-based guideline for llms, npj Digital Medicine 7 (1) (2024) 41. doi:https://doi.org/10.1038/s41746-024-01029-4.

[38] E. Mollick, L. Mollick, Assigning ai: Seven approaches for students, with prompts (2023). arXiv:2306.10052.
URL https://arxiv.org/abs/2306.10052

[39] J. Tu, H. Hadan, D. M. Wang, S. A. Sgandurra, R. H. Mogavi, L. E. Nacke, Augmenting the author: Exploring the potential of ai collaboration in academic writing (2024). arXiv:2404.16071.
URL https://arxiv.org/abs/2404.16071

[40] M. Aubin Le Quéré, H. Schroeder, C. Randazzo, J. Gao, Z. Epstein, S. T. Perrault, D. Mimno, L. Barkhuus, H. Li, Llms as research tools:

Applications and evaluations in hci data work, in: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, CHI EA '24, Association for Computing Machinery, New York, NY, USA, 2024. doi:10.1145/3613905.3636301.
URL https://doi.org/10.1145/3613905.3636301

[41] S. Tian, Q. Jin, L. Yeganova, P.-T. Lai, Q. Zhu, X. Chen, Y. Yang, Q. Chen, W. Kim, D. C. Comeau, R. Islamaj, A. Kapoor, X. Gao, Z. Lu, Opportunities and Challenges for ChatGPT and Large Language Models in Biomedicine and Health. arXiv:2306.10070, doi:10.48550/arXiv.2306.10070.

[42] S. Wang, P. Jin, A Brief Summary of Prompting in Using GPT Modelsdoi:10.32388/IMZI2Q.

[43] T. F. Heston, C. Khun, Prompt Engineering in Medical Education 2 (3) 198–205. doi:10.3390/ime2030019.

[44] F. M. Calisto, J. a. Fernandes, M. Morais, C. Santiago, J. a. M. Abrantes, N. Nunes, J. C. Nascimento, Assertiveness-based agent communication for a personalized medicine on medical imaging diagnosis, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, 2023. doi:10.1145/3544548.3580682.
URL https://doi.org/10.1145/3544548.3580682

[45] M. Williams, T. Moser, The art of coding and thematic exploration in qualitative research, International Management Review 15 (1) (2019) 45–55.
URL http://www.imrjournal.org/uploads/1/4/2/8/14286482/imr-v15n1art4.pdf

[46] U. D. Parameswaran, J. L. Ozawa-Kirk, G. Latendresse, To live (code) or to not: A new method for coding in qualitative research, Qualitative social work 19 (4) (2020) 630–644. doi:10.1177/1473325019840394.

[47] V. Clarke, V. Braun, N. Hayfield, Thematic analysis, Qualitative psychology: A practical guide to research methods 3 (2015) 222–248. doi:http://digital.casalini.it/9781473933415.

[48] M. Maguire, B. Delahunt, Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars., All Ireland Journal of Higher Education 9 (3) (2017).
URL https://ojs.aishe.org/index.php/aishe-j/article/view/335/553

[49] R. Boyatzis, Transforming Qualitative Information: Thematic Analysis and Code Development, Transforming Qualitative Information: Thematic Analysis and Code Development, SAGE Publications, 1998.
URL https://books.google.com/books?id=_rfClWRhIKAC

[50] K. A. R. Richards, M. A. Hemphill, A practical guide to collaborative qualitative data analysis, Journal of Teaching in Physical education 37 (2) (2018) 225–231. doi:10.1123/jtpe.2017-0084.

[51] G. Guest, K. MacQueen, E. Namey, Applied Thematic Analysis, SAGE Publications, 2011.
URL https://books.google.com/books?id=Hr11DwAAQBAJ

[52] H. Morgan, Understanding thematic analysis and the debates involving its use, The Qualitative Report 27 (10) (2022) 2079–2090. doi:10.46743/2160-3715/2022.5912.

[53] A.-M. Ortloff, M. Fassl, A. Ponticello, F. Martius, A. Mertens, K. Krombholz, M. Smith, Different researchers, different results? analyzing the influence of researcher experience and data type during qualitative analysis of an interview and survey study on security advice, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, 2023. doi:10.1145/3544548.3580766.
URL https://doi.org/10.1145/3544548.3580766

[54] E. Kuang, X. Jin, M. Fan, "merging results is no easy task": An international survey study of collaborative data analysis practices among ux practitioners, in: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22, Association for Computing Machinery, New York, NY, USA, 2022. doi:10.1145/3491102.3517647.
URL https://doi.org/10.1145/3491102.3517647

[55] V. Braun, V. Clarke, Reflecting on reflexive thematic analysis, Qualitative research in sport, exercise and health 11 (4) (2019) 589–597. doi:10.1080/2159676X.2019.1628806.

[56] J. L. Feuston, J. R. Brubaker, Putting tools in their place: The role of time and perspective in human-ai collaboration for qualitative analysis, Proc. ACM Hum.-Comput. Interact. 5 (CSCW2) (oct 2021). doi:10.1145/3479856.
URL https://doi.org/10.1145/3479856

[57] A. Castleberry, A. Nolen, Thematic analysis of qualitative research data: Is it as easy as it sounds?, Currents in Pharmacy Teaching and Learning 10 (6) (2018) 807–815. doi:https://doi.org/10.1016/j.cptl.2018.03.019.
URL https://www.sciencedirect.com/science/article/pii/S1877129717300606

[58] G. Guest, E. Namey, M. Mitchell, Collecting Qualitative Data: A Field Manual for Applied Research, SAGE Publications, 2013.
URL https://books.google.com/books?id=--3rmWYKtloC

[59] G. Terry, N. Hayfield, V. Clarke, V. Braun, Thematic analysis, The SAGE handbook of qualitative research in psychology 2 (2017) 17–37. doi:10.4135/9781526405555.n2.

[60] M. Khanbhai, P. Anyadi, J. Symons, K. Flott, A. Darzi, E. Mayer, Applying natural language processing and machine learning techniques to patient experience feedback: A systematic review 28 (1) e100262. arXiv:33653690, doi:10.1136/bmjhci-2020-100262.

[61] L. Yue, W. Chen, X. Li, W. Zuo, M. Yin, A survey of sentiment analysis in social media 60 (2) 617–663. doi:10.1007/s10115-018-1236-4.

[62] R. Churchill, L. Singh, The Evolution of Topic Modeling 54 1–35. doi:10.1145/3507900.

[63] T. C. Guetterman, T. Chang, M. DeJonckheere, T. Basu, E. Scruggs, V. G. V. Vydiswaran, Augmenting Qualitative Text Analysis with Natural Language Processing: Methodological Study 20 (6) e9702. doi:10.2196/jmir.9702.

[64] M.-H. Hong, L. A. Marsh, J. L. Feuston, J. Ruppert, J. R. Brubaker, D. A. Szafir, Scholastic: Graphical human-ai collaboration for inductive and interpretive text analysis, in: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology, UIST '22, Association for Computing Machinery, New York, NY, USA, 2022. doi:10.1145/3526113.3545681.
URL https://doi.org/10.1145/3526113.3545681

[65] S. A. Gebreegziabher, Z. Zhang, X. Tang, Y. Meng, E. L. Glassman, T. J.-J. Li, Patat: Human-ai collaborative qualitative coding with explainable interactive rule synthesis, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23, Association for Computing Machinery, New York, NY, USA, 2023. doi:10.1145/3544548.3581352.
URL https://doi.org/10.1145/3544548.3581352

[66] N.-C. Chen, M. Drouhard, R. Kocielnik, J. Suh, C. R. Aragon, Using machine learning to support qualitative coding in social science: Shifting the focus to ambiguity, ACM Transactions on Interactive Intelligent Systems (TiiS) 8 (2) (2018) 1–20. doi:10.1145/3185515.

[67] J. A. Jiang, K. Wade, C. Fiesler, J. R. Brubaker, Supporting serendipity: Opportunities and challenges for human-ai collaboration in qualitative analysis, Proc. ACM Hum.-Comput. Interact. 5 (CSCW1) (apr 2021). doi:10.1145/3449168.
URL https://doi.org/10.1145/3449168

[68] Y. Gamieldien, J. M. Case, A. Katz, Advancing qualitative analysis: An exploration of the potential of generative ai and nlp in thematic coding, Available at SSRN 4487768 (2023). doi:http://dx.doi.org/10.2139/ssrn.4487768.

[69] G. Bryda, D. Sadowski, From words to themes: Ai-powered qualitative data coding and analysis, in: World Conference on Qualitative Research, Springer, 2024, pp. 309–345. doi:https://doi.org/10.1007/978-3-031-65735-1_19.

[70] M. S. Jalali, A. Akhavan, Integrating ai language models in qualitative research: Replicating interview data analysis with chatgpt, System

Dynamics Review 40 (3) (2024) e1772. doi:https://doi.org/10.1002/sdr.1772.

[71] C. Overney, B. Sald´ıas, D. Dimitrakopoulou, D. Roy, Sensemate: An accessible and beginner-friendly human-ai platform for qualitative data analysis, in: Proceedings of the 29th International Conference on Intelligent User Interfaces, IUI '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 922–939. doi:10.1145/3640543.3645194.
URL https://doi.org/10.1145/3640543.3645194

[72] K. A. Fuller, K. A. Morbitzer, J. M. Zeeman, A. M. Persky, A. C. Savage, J. E. McLaughlin, Exploring the use of chatgpt to analyze student course evaluation comments, BMC Medical Education 24 (1) (2024) 423. doi:https://doi.org/10.1186/s12909-024-05316-2.

[73] A. Barany, N. Nasiar, C. Porter, A. F. Zambrano, A. L. Andres, D. Bright, M. Shah, X. Liu, S. Gao, J. Zhang, et al., Chatgpt for education research: exploring the potential of large language models for qualitative codebook development, in: International conference on artificial intelligence in education, Springer, 2024, pp. 134–149. doi:https://doi.org/10.1007/978-3-031-64299-9_10.

[74] A. Katz, S. Wei, G. Nanda, C. Brinton, M. Ohland, Exploring the Efficacy of ChatGPT in Analyzing Student Teamwork Feedback with an Existing Taxonomy. arXiv:2305.11882, doi:10.48550/arXiv.2305.11882.

[75] L. A. Siiman, M. Rannastu-Avalos, J. Pöysä-Tarhonen, P. Häkkinen, M. Pedaste, Opportunities and Challenges for AI-Assisted Qualitative Data Analysis: An Example from Collaborative Problem-Solving Discourse Data, in: Y.-M. Huang, T. Rocha (Eds.), Innovative Technologies and Learning, Lecture Notes in Computer Science, Springer Nature Switzerland, pp. 87–96. doi:10.1007/978-3-031-40113-8_9.

[76] J. Gao, K. T. W. Choo, J. Cao, R. K. W. Lee, S. Perrault, Coaicoder: Examining the effectiveness of ai-assisted human-to-human collaboration in qualitative analysis (2023). arXiv:2304.05560, doi:10.48550/arXiv.2304.05560.

[77] H. Vainio-Pekka, M. O.-O. Agbese, M. Jantunen, V. Vakkuri, T. Mikkonen, R. Rousi, P. Abrahamsson, The role of explainable ai in the research field of ai ethics (2023). doi:10.1145/3599974.
URL https://doi.org/10.1145/3599974

[78] A. Jacovi, A. Marasović, T. Miller, Y. Goldberg, Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in ai, in: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21, Association for Computing Machinery, New York, NY, USA, 2021, p. 624–635. doi:10.1145/3442188.3445923.
URL https://doi.org/10.1145/3442188.3445923

[79] J. Zerilli, U. Bhatt, A. Weller, How transparency modulates trust in artificial intelligence, Patterns (2022). doi:https://doi.org/10.1016/j.patter.2022.100455.
URL https://www.cell.com/patterns/pdf/S2666-3899(22)00028-9.pdf

[80] R. Singh, T. Miller, H. Lyons, L. Sonenberg, E. Velloso, F. Vetere, P. Howe, P. Dourish, Directive explanations for actionable explainability in machine learning applications (2023). doi:10.1145/3579363.
URL https://doi.org/10.1145/3579363

[81] S. Mohseni, N. Zarei, E. D. Ragan, A multidisciplinary survey and framework for design and evaluation of explainable ai systems (2021). doi:10.1145/3387166.
URL https://doi.org/10.1145/3387166

[82] T. Schrills, T. Franke, How do users experience traceability of ai systems? examining subjective information processing awareness in automated insulin delivery (aid) systems 13 (4) (2023). doi:10.1145/3588594.
URL https://doi.org/10.1145/3588594

[83] A. Springer, S. Whittaker, Progressive disclosure: When, why, and how do users want algorithmic transparency information? (2020). doi:10.1145/3374218.
URL https://doi.org/10.1145/3374218

[84] D. Buschek, M. Eiband, H. Hussmann, How to support users in understanding intelligent systems? an analysis and conceptual framework of user questions considering user mindsets, involvement, and knowledge outcomes (2022). doi:10.1145/3519264.
URL https://doi.org/10.1145/3519264

[85] D. C. Hernandez-Bocanegra, J. Ziegler, Explaining recommendations through conversations: Dialog model and the effects of interface type and degree of interactivity (2023). doi:10.1145/3579541.
URL https://doi.org/10.1145/3579541

[86] T. Spinner, R. Kehlbeck, R. Sevastjanova, T. Stähle, D. A. Keim, O. Deussen, M. El-Assady, generaitor: Tree-in-the-loop text generation for language model explainability and adaptation (2024). doi:10.1145/3652028.
URL https://doi.org/10.1145/3652028

[87] X. Wang, M. Yin, Effects of explanations in ai-assisted decision making: Principles and comparisons (2022). doi:10.1145/3519266.
URL https://doi.org/10.1145/3519266

[88] P. Christou, How to use artificial intelligence (ai) as a resource, methodological and analysis tool in qualitative research? 28 (7) 1968–1980. doi:10.46743/2160-3715/2023.6406.

[89] L. Levac, S. Ronis, Y. Cowper-Smith, O. Vaccarino, A scoping review: The utility of participatory research approaches in psychology, Journal of Community Psychology 47 (8) (2019) 1865–1892. doi:https://doi.org/10.1002/jcop.22231.

[90] J. Zamfirescu-Pereira, H. Wei, A. Xiao, K. Gu, G. Jung, M. G. Lee, B. Hartmann, Q. Yang, Herding ai cats: Lessons from designing a chatbot by prompting gpt-3, in: Proceedings of the 2023 ACM Designing Interactive Systems Conference, DIS '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 2206–2220. doi:10.1145/3563657.3596138.
URL https://doi.org/10.1145/3563657.3596138

[91] N. Cooper, T. Horne, G. R. Hayes, C. Heldreth, M. Lahav, J. Holbrook, L. Wilcox, A systematic review and thematic analysis of community-collaborative approaches to computing research, in: Proceedings of the

2022 CHI Conference on Human Factors in Computing Systems, CHI '22, Association for Computing Machinery, New York, NY, USA, 2022. doi:10.1145/3491102.3517716.
URL https://doi.org/10.1145/3491102.3517716

[92] J. Jiang, K. Zhou, Z. Dong, K. Ye, W. X. Zhao, J.-R. Wen, Structgpt: A general framework for large language model to reason over structured data (2023). arXiv:2305.09645.

[93] F. M. Megahed, Y.-J. Chen, J. A. Ferris, S. Knoth, L. A. Jones-Farmer, How generative ai models such as chatgpt can be (mis)used in spc practice, education, and research? an exploratory study, Quality Engineering 0 (0) (2023) 1–29. arXiv:https://doi.org/10.1080/08982112.2023.2206479, doi:10.1080/08982112.2023.2206479.
URL https://doi.org/10.1080/08982112.2023.2206479

[94] R. Paul, L. Elder, Critical and creative thinking, Dillon Beach, CA: The Foundation for Critical Thinking (2004).
URL https://www.thefont.co.za/wp-content/uploads/2020/03/007-Paul-Elder-2008-A-FULL-Guide-to-Crit-Thinking-Concepts_Tool.pdf

[95] W. Jiao, W. Wang, J. tse Huang, X. Wang, Z. Tu, Is chatgpt a good translator? yes with gpt-4 as the engine (2023). arXiv:2301.08745.

[96] X. Chu, I. F. Ilyas, S. Krishnan, J. Wang, Data cleaning: Overview and emerging challenges, in: Proceedings of the 2016 International Conference on Management of Data, SIGMOD '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 2201–2206. doi:10.1145/2882903.2912574.
URL https://doi.org/10.1145/2882903.2912574

[97] D. E. White, N. D. Oelke, S. Friesen, Management of a large qualitative data set: Establishing trustworthiness of the data, International journal of qualitative methods 11 (3) (2012) 244–258. doi:10.1177/160940691201100305.

[98] A. Kong, S. Zhao, H. Chen, Q. Li, Y. Qin, R. Sun, X. Zhou, Better zero-shot reasoning with role-play prompting (2023). arXiv:2308.07702.

[99] B. Eager, R. Brunton, Prompting higher education towards ai-augmented teaching and learning practice, Journal of University Teaching & Learning Practice 20 (5) (2023) 02. doi:10.53761/1.20.5.02.

[100] Y. Liu, G. Deng, Z. Xu, Y. Li, Y. Zheng, Y. Zhang, L. Zhao, T. Zhang, Y. Liu, Jailbreaking chatgpt via prompt engineering: An empirical study (2023). arXiv:2305.13860.

[101] Y. Dong, X. Jiang, Z. Jin, G. Li, Self-collaboration code generation via chatgpt (2023). arXiv:2304.07590.

[102] Z. Yuan, Y. Lou, M. Liu, S. Ding, K. Wang, Y. Chen, X. Peng, No more manual tests? evaluating and improving chatgpt for unit test generation (2023). arXiv:2305.04207.

[103] Z. Kilhoffer, A. Nlkolich, M. R. Sanfilippo, Z. Zhou, Ai accountability policy, NTIA-2023-0005-0810 (2023).
URL https://www.ideals.illinois.edu/items/127130

[104] A. Adadi, M. Berrada, Peeking inside the black-box: A survey on explainable artificial intelligence (xai), IEEE Access 6 (2018) 52138–52160. doi:10.1109/ACCESS.2018.2870052.

[105] D. Castelvecchi, Can we open the black box of ai?, Nature News 538 (7623) (2016) 20. doi:10.1038/538020a.

[106] A. Barredo Arrieta, N. D´ıaz-Rodr´ıguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Information Fusion 58 (2020) 82–115. doi:https://doi.org/10.1016/j.inffus.2019.12.012.
URL https://www.sciencedirect.com/science/article/pii/S1566253519308103

[107] V. Franzoni, From black box to glass box: advancing transparency in artificial intelligence systems for ethical and trustworthy ai, in: O. Gervasi, B. Murgante, A. M. A. C. Rocha, C. Garau, F. Scorza, Y. Karaca, C. M. Torre (Eds.), Computational Science and Its Applications – ICCSA 2023 Workshops, Springer Nature Switzerland, Cham, 2023, pp. 118–130. doi:10.1007/978-3-031-37114-1_9.

[108] K. B. Jensen, The qualitative research process, 1st Edition, Routledge, London, 2002, Ch. 14, pp. 247–265. doi:10.4324/9780203465103.

[109] D. Ihde, Instrumental Realism: The Interface Between Philosophy of Science and Philosophy of Technology, A Midland book, Indiana University Press, 1991.
URL https://books.google.com/books?id=vYl0tagdh1QC

[110] E. Husserl, D. Moran, Ideas: General introduction to pure phenomenology, Routledge, 2012. doi:10.4324/9780203120330.

[111] M. Foucault, et al., Technologies of the self, in: Technologies of the self: A seminar with Michel Foucault, Vol. 18, Amherst, 1988.
URL https://works.raqsmediacollective.net/wp-content/uploads/2023/04/sarai_reader_03_shaping_technologies.pdf#page=105

[112] H. Hadan, D. M. Wang, R. H. Mogavi, J. Tu, L. Zhang-Kennedy, L. E. Nacke, The great ai witch hunt: Reviewers' perception and (mis)conception of generative ai in research writing, Computers in Human Behavior: Artificial Humans 2 (2) (2024) 100095. doi:https://doi.org/10.1016/j.chbah.2024.100095.
URL https://www.sciencedirect.com/science/article/pii/S2949882124000550

[113] R. Hadi Mogavi, C. Deng, J. Juho Kim, P. Zhou, Y. D. Kwon, A. Hosny Saleh Metwally, A. Tlili, S. Bassanelli, A. Bucchiarone, S. Gujar, L. E. Nacke, P. Hui, Chatgpt in education: A blessing or a curse? a qualitative study exploring early adopters' utilization and perceptions, Computers in Human Behavior: Artificial Humans 2 (1) (2024) 100027. doi:https://doi.org/10.1016/j.chbah.2023.100027.
URL https://www.sciencedirect.com/science/article/pii/S2949882123000270

[114] L. Ma, B. Sun, Machine learning and ai in marketing – connecting computing power to human insights, International Journal of Research in Marketing 37 (3) (2020) 481–504. doi:https://doi.org/10.1016/j.ijresmar.2020.04.005.
URL https://www.sciencedirect.com/science/article/pii/S0167811620300410

[115] T. N. Fitria, Artificial intelligence (ai) technology in openai chatgpt application: A review of chatgpt in writing english essay, in: ELT Forum: Journal of English Language Teaching, Vol. 12, 2023, pp. 44–58. doi:10.15294/elt.v12i1.64069.

[116] A. Marcoci, A. C. Thresher, N. C. Martens, P. Galison, S. S. Doeleman, M. D. Johnson, Big stem collaborations should include humanities and social science, Nature Human Behaviour (2023) 1–2doi:10.1038/s41562-023-01674-x.

[117] N. Pangakis, S. Wolken, N. Fasching, Automated annotation with generative ai requires validation (2023). arXiv:2306.00176.
URL https://arxiv.org/abs/2306.00176

[118] H. Schroeder, M. A. L. Quéré, C. Randazzo, D. Mimno, S. Schoenebeck, Large language models in qualitative research: Can we do the data justice? (2024). arXiv:2410.07362.
URL https://arxiv.org/abs/2410.07362

[119] T. Koch, Establishing rigour in qualitative research: the decision trail, Journal of advanced nursing 19 (5) (1994) 976–986.

[120] J. W. Drisko, Strengthening qualitative studies and reports: Standards to promote academic integrity, Journal of social work education 33 (1) (1997) 185–197. doi:https://doi.org/10.1080/10437797.1997.10778862.

[121] R. C. Mayer, J. H. Davis, F. D. Schoorman, An integrative model of organizational trust, Academy of management review 20 (3) (1995) 709–734. doi:https://doi.org/10.5465/amr.1995.9508080335.

[122] L. Seitz, S. Bekmeier-Feuerhahn, K. Gohil, Can we trust a chatbot like a physician? a qualitative study on understanding the emergence of trust toward diagnostic chatbots, International Journal of Human-Computer Studies 165 (2022) 102848. doi:https://doi.org/10.1016/j.ijhcs.2022.102848.
URL https://www.sciencedirect.com/science/article/pii/S1071581922000751

[123] B. M. Muir, Trust between humans and machines, and the design of decision aids, International journal of man-machine studies 27 (5-6) (1987) 527–539. doi:https://doi.org/10.1016/S0020-7373(87)80013-5.

[124] M. Carabantes, Black-box artificial intelligence: an epistemological and critical analysis, AI & society 35 (2) (2020) 309–317. doi:https://doi.org/10.1007/s00146-019-00888-w.

[125] R. Tuval-Mashiach, Raising the curtain: The importance of transparency in qualitative research., Qualitative Psychology 4 (2) (2017) 126. doi:https://doi.org/10.1037/qup0000062.

[126] S. J. Tracy, Qualitative quality: Eight "big-tent" criteria for excellent qualitative research, Qualitative inquiry 16 (10) (2010) 837–851. doi:https://doi.org/10.1177/1077800410383121.

[127] B. C. O'Brien, I. B. Harris, T. J. Beckman, D. A. Reed, D. A. Cook, Standards for reporting qualitative research: a synthesis of recommendations, Academic medicine 89 (9) (2014) 1245–1251. doi:10.1097/ACM.0000000000000388.

[128] H. Felzmann, E. Fosch-Villaronga, C. Lutz, A. Tamò-Larrieux, Towards transparency by design for artificial intelligence, Science and engineering ethics 26 (6) (2020) 3333–3361. doi:https://doi.org/10.1007/s11948-020-00276-4.

[129] M. Ananny, K. Crawford, Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability, new media & society 20 (3) (2018) 973–989. doi:https://doi.org/10.1177/1461444816676645.

[130] R. F. Kizilcec, How much information? effects of transparency on trust in an algorithmic interface, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, Association for Computing Machinery, New York, NY, USA, 2016, p. 2390–2395. doi:10.1145/2858036.2858402.
URL https://doi.org/10.1145/2858036.2858402

[131] J. Laine, M. Minkkinen, M. Mäntymäki, Ethics-based ai auditing: A systematic literature review on conceptualizations of ethical principles and knowledge contributions to stakeholders, Information & Management (2024) 103969doi:https://doi.org/10.1016/j.im.2024.103969.

[132] T. Hagendorff, The ethics of ai ethics: An evaluation of guidelines, Minds and machines 30 (1) (2020) 99–120. doi:https://doi.org/10.1007/s11023-020-09517-8.

[133] M. K. Lee, A. Jain, H. J. Cha, S. Ojha, D. Kusbit, Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation, Proc. ACM Hum.-Comput. Interact. 3 (CSCW) (Nov. 2019). doi:10.1145/3359284.
URL https://doi.org/10.1145/3359284

[134] M. D. Molina, S. S. Sundar, When ai moderates online content: effects of human collaboration and interactive transparency on user trust, Journal of Computer-Mediated Communication 27 (4) (2022) zmac010.

[135] W. J. Orlikowski, Sociomaterial practices: Exploring technology at work, Organization studies 28 (9) (2007) 1435–1448. doi:https://doi.org/10.1177/0170840607081138.

[136] A. Pickering, The mangle of practice: Time, agency, and science, University of Chicago Press, 2010.

[137] B. Latour, Reassembling the social: An introduction to actor-network-theory, Oxford university press, 2005.

[138] A. Yardley, et al., Piecing together—a methodological bricolage, in: Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, Vol. 9, 2008.

[139] J. L. Kincheloe, Describing the bricolage: Conceptualizing a new rigor in qualitative research, Qualitative inquiry 7 (6) (2001) 679–692.

[140] M. Rogers, Contextualizing theories and practices of bricolage research., Qualitative Report 17 (2012) 7.

[141] G. Jacucci, A. Spagnolli, J. Freeman, L. Gamberini, Symbiotic interaction: A critical definition and comparison to other human-computer paradigms, in: G. Jacucci, L. Gamberini, J. Freeman, A. Spagnolli (Eds.), Symbiotic Interaction, Springer International Publishing, Cham, 2014, pp. 3–20. doi:10.1007/978-3-319-13500-7_1.

[142] M. H. Jarrahi, Artificial intelligence and the future of work: Human-ai symbiosis in organizational decision making, Business Horizons 61 (4) (2018) 577–586. doi:https://doi.org/10.1016/j.bushor.2018.03.007.
URL https://www.sciencedirect.com/science/article/pii/S0007681318300387

[143] E. Go, S. S. Sundar, Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions, Computers in Human Behavior 97 (2019) 304–316. doi:https://doi.org/10.1016/j.chb.2019.01.020.
URL https://www.sciencedirect.com/science/article/pii/S0747563219300329

[144] D. A. Moore, P. J. Healy, The trouble with overconfidence., Psychological review 115 (2) (2008) 502. doi:https://doi.org/10.1037/0033-295X.115.2.502.

[145] D. J. McAllister, Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations, Academy of management journal 38 (1) (1995) 24–59. doi:https://doi.org/10.5465/256727.

[146] T. Wu, S. He, J. Liu, S. Sun, K. Liu, Q.-L. Han, Y. Tang, A brief overview of chatgpt: The history, status quo and potential future development, IEEE/CAA Journal of Automatica Sinica 10 (5) (2023) 1122–1136. doi:10.1109/JAS.2023.123618.

[147] E. Horvitz, Principles of mixed-initiative user interfaces, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '99, Association for Computing Machinery, New York, NY, USA, 1999, p. 159–166. doi:10.1145/302979.303030.
URL https://doi.org/10.1145/302979.303030

[148] G. Tecuci, M. Boicu, M. T. Cox, Seven aspects of mixed-initiative reasoning: An introduction to this special issue on mixed-initiative assistants, AI Magazine 28 (2) (2007) 11–11. doi:https://doi.org/10.1609/aimag.v28i2.2035.

[149] P. Hemmer, M. Schemmer, N. Kühl, M. Vössing, G. Satzger, Complementarity in human-ai collaboration: Concept, sources, and evidence (2024). arXiv:2404.00029.
URL https://arxiv.org/abs/2404.00029

[150] F. Doshi-Velez, B. Kim, Towards a rigorous science of interpretable machine learning (2017). arXiv:1702.08608.
URL https://arxiv.org/abs/1702.08608

[151] H. Nissenbaum, Privacy as contextual integrity symposium: Technology, values, and the justice system, Washington Law Review 79 (2004) 119. doi:https://heinonline.org/HOL/Page?handle=hein.journals/washlr79&id=129.

[152] I. Gessinger, K. Seaborn, M. Steeds, B. R. Cowan, Chatgpt and me: First-time and experienced users' perceptions of chatgpt's communicative ability as a dialogue partner, International Journal of Human-Computer Studies 194 (2025) 103400. doi:https://doi.org/10.1016/j.ijhcs.2024.103400.
URL https://www.sciencedirect.com/science/article/pii/S1071581924001836

[153] F. Motoki, V. Pinho Neto, V. Rodrigues, More human than human: Measuring chatgpt political bias, Available at SSRN 4372349 (2023). doi:https://dx.doi.org/10.2139/ssrn.4372349.

[154] M. Koivisto, S. Grassini, Best humans still outperform artificial intelligence in a creative divergent thinking task, Scientific reports 13 (1) (2023) 13601. doi:https://doi.org/10.1038/s41598-023-40858-3.

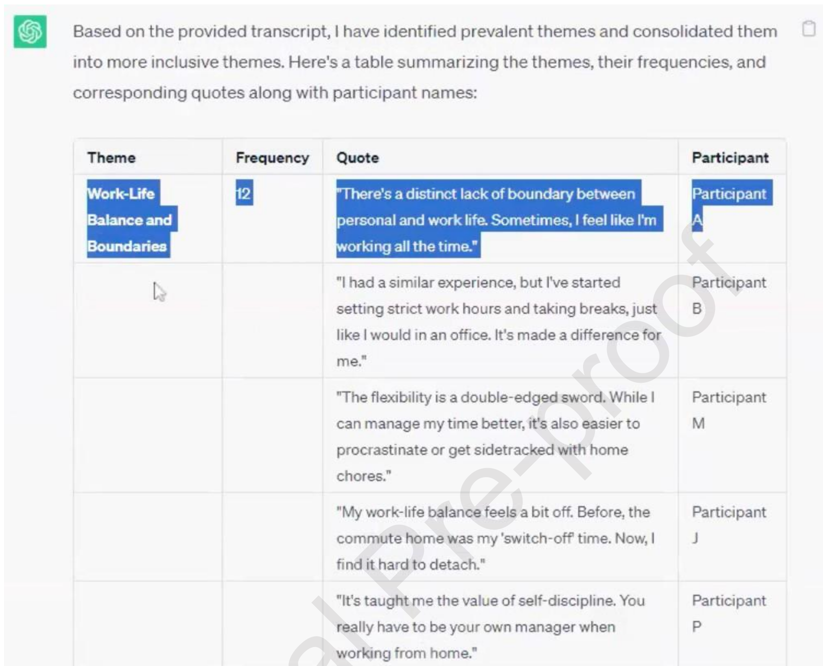
[155] Y. Liu, D. Iter, Y. Xu, S. Wang, R. Xu, C. Zhu, G-eval: Nlg evaluation using gpt-4 with better human alignment (2023). arXiv:2303.16634.
URL https://arxiv.org/abs/2303.16634

[156] F. M. Calisto, N. Nunes, J. C. Nascimento, Modeling adoption of intelligent agents in medical imaging, International Journal of Human-Computer Studies 168 (2022) 102922. doi:https://doi.org/10.1016/j.ijhcs.2022.102922.

Appendix A. Appendix

Strategies	Examples of Prompts
Background / Conceptual Understanding	"*Here is a transcript from a focus group interview about 'Transitioning to Remote Work'. Each paragraph is from one participant. Please read it first.*" (P9)
Description of the Methodology (Goal of Task)	"*can you do a thematic analysis of their responses?*" (P6)
Description of the Analytical Process (context)	"*Please do a thematic analysis and summarize no more than 10 themes from this transcript.*" (P9)
Definition of Data Format (Inputs)	"*The format of the transcript looks like this: Participant <participant name>: <transcript of this participant's comments regarding transitioning to remote work>.*" (P15)
Definition of Data Format (Outputs)	"*Please analysis data again and make outputs follow format as below:* *(New line) Topic: {challenges}, {strategies}, {benefits}, {downsides}, {productivity}* *(New line) Argument points: {reasons}, {suggestions}, {perceptions}* *(New line) Supports: {raw data}, {participant number}*" (P8)
Role-Playing	"*You are now a research expert in qualitative analysis...*" (P7)
Prioritization	"*I want a codebook with less than 10 codes. I hope these codes are main themes from the transcript ...*" (P10)
Transparency &Traceability	"*...and tell me what's your rationale for your analysis.*" (P13)
Acknowledgment of Expertise	"*Good job – I cannot finish my work without you, you are so sweet. Follow the requirements as below, and continue analyzing data.*" (P5)

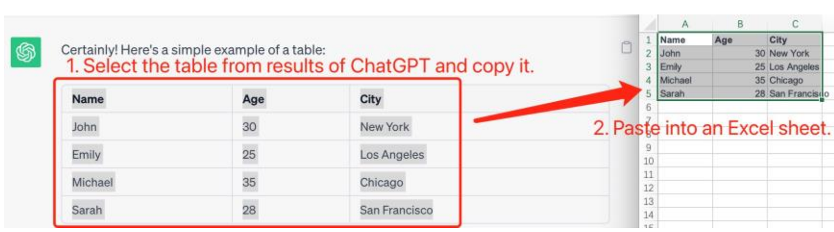
Table A.4: Strategies and Examples of Prompts.



Based on the provided transcript, I have identified prevalent themes and consolidated them into more inclusive themes. Here's a table summarizing the themes, their frequencies, and corresponding quotes along with participant names:

Theme	Frequency	Quote	Participant
Work-Life Balance and Boundaries	12	"There's a distinct lack of boundary between personal and work life. Sometimes, I feel like I'm working all the time."	Participant A
		"I had a similar experience, but I've started setting strict work hours and taking breaks, just like I would in an office. It's made a difference for me."	Participant B
		"The flexibility is a double-edged sword. While I can manage my time better, it's also easier to procrastinate or get sidetracked with home chores."	Participant M
		"My work-life balance feels a bit off. Before, the commute home was my 'switch-off' time. Now, I find it hard to detach."	Participant J
		"It's taught me the value of self-discipline. You really have to be your own manager when working from home."	Participant P

Figure A.3: An example of some results: Participant (P15) added requirements regarding the format of the output results in the prompts. The original prompt text reads: "*For the output, put the result in a table. The first column is the name of the theme, the second column is its frequency, the third column includes the quotes that belong to this theme and the name of the participant who made this comment, one row per quote.*".



Name	Age	City
John	30	New York
Emily	25	Los Angeles
Michael	35	Chicago
Sarah	28	San Francisco

Figure A.4: An example of how to transfer the table from ChatGPT to Excel.

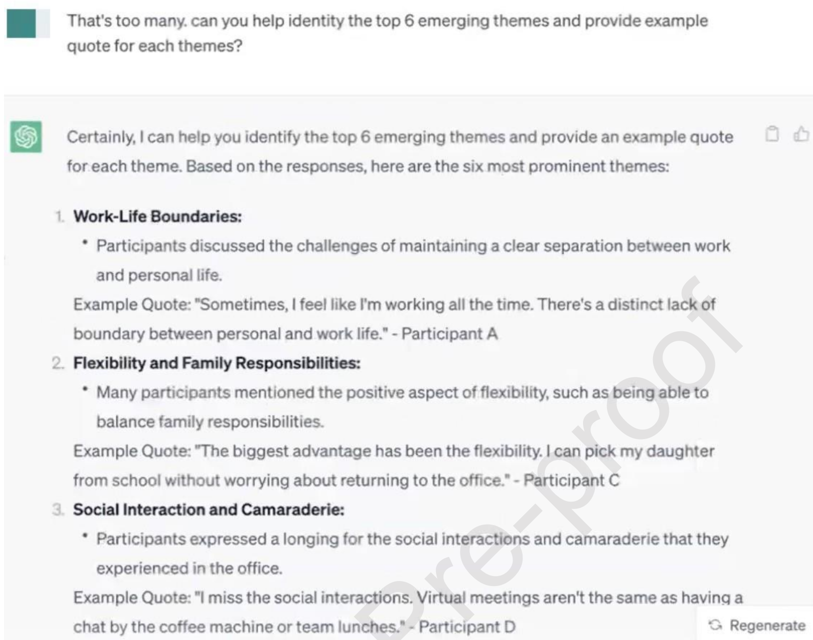


Figure A.5: Examples of ChatGPT's output after adding priority requirements.

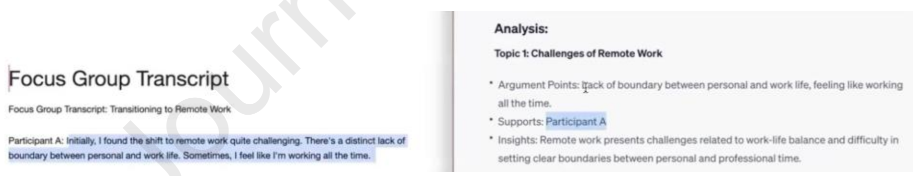


Figure A.6: Examples of ChatGPT's output from P8's prompts.

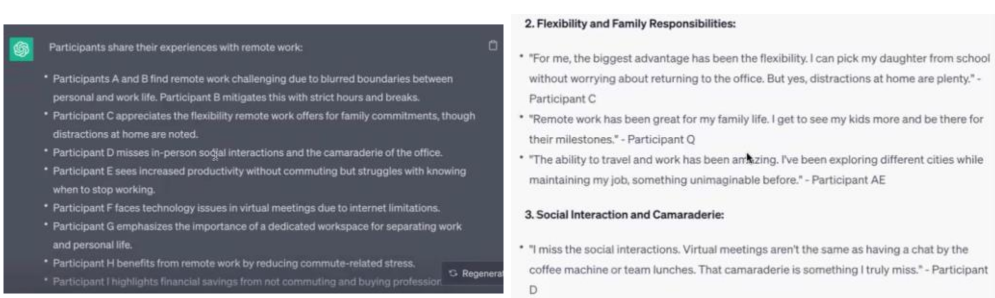


Figure A.7: Some of the results obtained by participants using ChatGPT. On the left is the result for P5, and on the right is the result for P6.

Highlights

Harnessing the Power of AI in Qualitative Research: Exploring, Using and Redesigning ChatGPT

He Zhang, Chuhao Wu, Jingyi Xia, Yao Lyu, Jie Cai, John M. Carroll

- This study identifies the key challenges of applying ChatGPT in qualitative analysis, including prompt design and AI interpretability, and proposes a structured framework to address these issues.
- The prompt design framework, developed with researcher feedback and aligned with traditional qualitative methods, enhances ChatGPT's effectiveness in qualitative analysis by improving context definition, methodological guidance, and data structuring.
- The adaptability of the framework is discussed, highlighting its potential to support evolving AI models and serve as a reusable resource across future LLM-based tools.
- Ethical considerations, such as transparency and accountability in AI-assisted analysis, are examined to promote reliable and responsible use of ChatGPT in qualitative research.
- Practical implications for junior researchers are emphasized, as the framework provides a foundational tool to improve prompt design skills and proficiency in AI-supported qualitative analysis.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: