Journal on Interactive Systems, 2025, 16:1, doi: 10.5753/jis.2025.5416

Can (A)I help you? Comparing human and GenAI analysis of HCI qualitative research results

Mariana Gomes Borges [**PUCRS** | ***mariana.borges@edu.pucrs.br***]
Claiton Marques Correa [**IFFar** | ***claiton.correa@iffar.edu.br***]
Diego Moreira da Rosa [**IFRS** | ***diego.rosa@restinga.ifrs.edu.br***]
Andrea Gnecco [**PUCRS** | ***andrea.gnecco@edu.pucrs.br***]
Milene Selbach Silveira [**PUCRS** | ***milene.silveira@pucrs.br***]

Polytechnic School, Pontifical Catholic University of Rio Grande do Sul (PUCRS), Av. Ipiranga 6681, Porto Alegre, CEP 90619-900, Brazil.

Received: 10 January 2025 • **Accepted:** 14 October 2025 • **Published:** 02 November 2025

Abstract: Generative AI (GenAI) is experiencing rapid growth, particularly in its application as a tool for qualitative text analysis—a key element of Human-Computer Interaction (HCI) research. This study examines the potential of GenAI, specifically ChatGPT, to assist in the analysis of qualitative research data. Four qualitative HCI studies, previously conducted and analyzed by our research group, were selected for this investigation. ChatGPT was employed to perform AI-assisted analyses on the raw data from these studies, and the AI-generated insights were then compared with the human-led analyses already completed. The results reveal significant alignment between the human and AI-assisted analyses, indicating that GenAI can serve as an effective support tool in qualitative research. However, while GenAI offers considerable advantages in enhancing research efficiency, human oversight remains crucial to ensuring accurate interpretation and contextual alignment. This study also provides practical recommendations for researchers interested in incorporating GenAI into their qualitative analysis processes.

Keywords: Generative AI, Qualitative Research, User Research, Text Analysis, ChatGPT, Human-AI Collaboration

1 Introduction

The integration of Artificial Intelligence (AI) into qualitative research has been a concept that was introduced earlier. Prior to the development of current Large Language Models (LLM), the most prevalent use of AI in qualitative research data analysis involved the application of Natural Language Processing (NLP), with sentiment analysis and semantic coding being common approaches [Morgan, 2023].

Recently, the introduction and evolution of Generative Artificial Intelligence (GenAI) tools like ChatGPT and Google Gemini have ushered in an era of new possibilities for collaboration in qualitative research analysis because of their capacity to process and generate human-like texts.

In this wake, many researchers are investigating using GenAI in qualitative analysis. For example, Yan *et al.* [2024] investigated the use of ChatGPT in thematic analysis and provided design recommendations for future seamless human-AI collaboration in qualitative research. Morgan [2023] compared its results using GenAI to those obtained through traditional manual coding methods. He aimed to assess the suitability of ChatGPT for this purpose and discuss its implications for the future of qualitative research. Feuston and Brubaker [2021] explored how AI could be integrated into various stages of qualitative analysis, mainly when dealing with large datasets. The authors aimed to understand the nuanced perspectives of qualitative researchers regarding AI adoption and provide recommendations on how AI systems can be designed to better support and augment their analytical practices without compromising the integrity and depth of qualitative research.

Despite these early explorations, the scientific problem remains: while GenAI tools offer efficiency and scalability, their reliability, interpretability, and alignment with qualitative research values, such as contextual sensitivity and reflexivity, are still unclear. This is particularly relevant in Human-Computer Interaction (HCI), where a rich and in-depth analysis of user experiences often relies on a nuanced interpretation of text-based data, such as interviews, open-ended survey responses, and workshop transcripts [Wang *et al.*, 2020; Yang *et al.*, 2020]. As GenAI tools become more integrated into research workflows, there is an urgent need to understand their real potential and limitations in supporting, not replacing, human-led interpretation in qualitative analysis.

Furthermore, the challenge extends beyond mere application to the methods of interaction with these models. The effectiveness of GenAI is highly dependent on the quality of user prompts. Regarding originality and relevance, this object is quite current and could also include advanced techniques such as Instructional and Iterative prompts, Retrieval-Augmented Generation (RAG), Chain-of-Thought (CoT), and Retrieval-Augmented Thoughts (RAT) as prompt improvement strategies. Although a detailed analysis of these advanced prompting methods is beyond the scope of this paper, their existence highlights the complexity of achieving reliable results and reinforces the need for foundational studies like ours.

This study addresses this gap by investigating the role of GenAI, specifically ChatGPT, in assisting the analysis of qualitative data in HCI. It asks:

- How does GenAI, specifically ChatGPT, perform when

applied to different types of qualitative data in HCI research?
- To what extent do GenAI-generated insights align with those produced by human researchers?
- What are the limitations, risks, and opportunities of integrating GenAI into qualitative analysis workflows?

In this scenario, we sought to delve into the dynamics of human-AI collaboration in qualitative analysis by examining the experiences of performing analysis using ChatGPT. We selected four of our HCI research projects, each employing different qualitative data collection methods, and subjected the row data and the study's protocol to an AI-powered textual analysis. The first study consisted of an **online survey** to uncover the behavior and difficulties of technology users in different multilingual contexts. The second study consisted of a **focus group** to identify how data types are used with different structures of narrative data visualizations. The third study is a set of **workshop sessions** conducted to observe how end-users of data visualization use customizable features to explore a given narrative developed using our proposed narrative data visualization approach to support visualization design. The last research consisted of **interviews** to understand how design systems' users, typically user experience designers, were applying interaction design patterns and how they were relating these patterns to design systems.

The main contribution of this work is to explore the potential of GenAI, specifically ChatGPT, in assisting with the analysis of qualitative research data in the HCI field. In addition, we also provide practical recommendations for researchers interested in incorporating GenAI into their qualitative analysis processes.

The remainder of the paper is organized as follows. Section 2 addresses the research on how GenAI can aid qualitative analysis. It also positions our contribution in the context of existing research. Section 3 details the five main steps undertaken in the research, which include selecting studies, defining a prompt template, generating AI-assisted analyses, comparing these with human-led analyses, and consolidating the findings. Section 4 then presents the outcomes of the AI-assisted and human-led analyses for each of the four selected studies, highlighting key similarities and differences. Section 5 contextualizes these results within the broader field of AI in qualitative research, comparing and contrasting the findings with those of other studies. Finally, Section 6 concludes by summarizing the key takeaways, acknowledging limitations, and suggesting potential avenues for future research.

2 AI and qualitative research

Qualitative analysis is an important methodological approach in HCI. It is essential for allowing researchers to understand users' experiences, social dynamics, and other contextual matters of technology use. Recent advances in GenAI, particularly with models like GPT-4o, have prompted researchers to investigate its potential in qualitative data analysis. This has led to discussions about efficiency, trustworthiness, and the importance of human agency in the research process. This section reviews studies examining how AI, especially ChatGPT, can support, challenge, and potentially transform qualitative research.

Jiang *et al*. [2021] explored how human-AI collaboration can enhance qualitative research. The authors argue that while AI can be helpful in managing and coding large qualitative datasets, it should not diminish the researcher's ability to interpret data flexibly or make creative insights. Based on interviews with qualitative researchers, the study highlights the importance of maintaining human agency and acknowledging the inherent uncertainty in qualitative research. It emphasizes that AI assistance should avoid imposing rigid or overly automated processes.

Wachinger *et al*. [2024] investigated ChatGPT's ability to analyze qualitative interview transcripts compared to a human researcher. The researchers identified a significant thematic overlap between analyses conducted by humans and those generated by AI, especially when using prompts specifically designed for thematic analysis. While ChatGPT performed well in suggesting descriptive codes and connecting them to theoretical frameworks—even in situations with a loose theoretical fit—it still needed human oversight to guarantee contextual accuracy and maintain methodological rigor. This study highlights both the promise and the limitations of using LLMs for qualitative research, especially regarding integrating AI into existing research workflows and theoretical reasoning.

dos Anjos *et al*. [2024] evaluated the effectiveness of models like ChatGPT 4.0 and Claude 2.0 in conducting qualitative research in the field of science education. Using Cognitive Network Mediation Theory (CMT) to analyze student interviews, the study found that Claude 2.0 demonstrated a superior ability to identify cognitive mediations and distinguish between pre- and post-test conditions. Additionally, it provided more detailed references to specific excerpts from the interviews. The research concludes that AI, particularly Claude 2.0, is valuable for qualitative research. However, it emphasizes the critical role of human oversight in identifying potential limitations of the models.

Liu *et al*. [2025] examined how AI can support real-time qualitative analysis during semi-structured interviews. They emphasized that while AI could assist with tasks such as note-taking, question management, and real-time sentiment detection, researchers consistently insisted on maintaining control over critical interpretive decisions. The study suggests that AI's most effective role may be to reduce cognitive load and enhance human sensemaking, rather than replacing the authority and judgment of the human researcher.

Our study adds to the research by using ChatGPT to analyze raw qualitative data from four previous HCI studies and comparing its outputs to those of human analyses. We discovered significant overlap in the identification of themes, which suggests that GenAI can effectively support qualitative research. However, in line with concerns raised by other researchers, we highlight the importance of human oversight to ensure accurate interpretation and maintain the contextual richness of the data. This study contributes to the emerging field by providing practical recommendations for integrating AI into qualitative research workflows, advocating for a collaborative approach rather than a substitutive role for AI in the analysis process.

3 Methodology

To answer the research question *Can generative AI help researchers with the analysis of qualitative research results?*, we structured the research into five main steps as depicted in Fig. 1. Steps 1, 2, and 5 were performed by all authors collaboratively, while Steps 3 and 4 (inside the central rectangle) were repeated for each of the four studies and executed individually by one author who was responsible for the specific study[1].

Step 1 consisted of selecting four studies to be part of this research. The studies were selected by convenience, all being developed within our research group and consisting of previous research by the authors of this work. This selection strategy ensured that the studies met the following criteria:

- **HCI-based:** All studies were in the field of Human-Computer Interaction.
- **Qualitative research:** The studies were based on qualitative research, i.e., they included qualitative data that had been analyzed by humans without the help of GenAI.
- **Variability of methods:** The methods used for data collection varied between studies, including online survey, focus group, workshop, and interviews. This allowed for comparison and a more comprehensive analysis. More details about each study, including the exact methods used for data collection and text analysis, can be found in Section 4.
- **Raw data access:** By selecting studies from our own research group, we guaranteed access to all raw data collected during the studies. In addition to the unrestricted access, we also had previous knowledge of the data structure, which made manipulation easier.
- **Control over the human-made analysis:** The authors themselves are the researchers who carried out the initial analysis; therefore they have more authority for the discussion.
- **Ethical conduct from start to finish:** All studies had been previously approved by the Ethics Committee, ensuring that the data had been collected and analyzed in compliance with ethical criteria.

Due to the factors listed above, the selection of studies from our group was important for the operationalization of this research work. It should be noted, however, that this selection strategy gives the study an exploratory nature, increases the risk of bias, and reduces the possibilities of generalization. Despite these liabilities, the advantages of such an approach outweigh the disadvantages, and we believe that the results presented below provide relevant recommendations and insights for HCI researchers who wish to use AI in their qualitative studies.

Step 2 was conducted in parallel with the first step and consisted of free explorations with the artificial intelligence chatbot to define the best prompt template. The template aimed to strike a balance between generality and specificity to ensure comparability across the four studies.

[1]All prior user studies were approved by the University Ethics Committee following ethical standards.

The ChatGPT[2] interface by OpenAI[3] was chosen for being a widely popular tool among researchers and the general public. During the execution of this study, the interface was backed by the GPT-4o model. The construction of the prompt template was not predefined but emerged through iterative experimentation.

The adoption of human-AI collaboration in research has become an even more relevant object of study after the popularization of large language models in the 2000s. In this work, we do not intend to exhaust all possibilities of collaboration nor evaluate all models available on the market (which are already numerous). Therefore, we chose the ChatGPT interface supported by the GPT-4o model as a reference for our experiment. Among the reasons for choosing the OpenAI environment, we can mention its pioneering nature and great accessibility, state-of-the-art technology, and the active interest of the academic community in this tool. The OpenAI model is widely available through its natural language interface (ChatGPT) and through APIs, being a pioneer in popularizing access to LLMs and generative AI. The model also stands out in traditional benchmarks for the evaluation of natural language processing[4]. Finally, due to these characteristics, the ChatGPT interface (and its backing models) has been the target of several research works, providing a good reference framework [Barambones *et al*., 2024; Morgan, 2023; Saare, 2024; Sinha *et al*., 2024; Wachinger *et al*., 2024; Yan *et al*., 2024].

During the construction of the prompt template, the researchers tried to summarize all relevant information needed for the model to produce the best analysis. This template would then be adapted and used for each of the selected studies. It is important to note that this process of prompt crafting directly influenced the outcomes of each study.

Fig. 2 shows the prompt template defined during Step 2. The prompt starts **(1)** with an imperative sentence that introduces what the researcher wants from the tool. The next block of text **(2)** should describe the method to be used for the analysis. In our case, this description was as close as possible to the method used by human researchers during their previous analyses of the results. This ensured a fair comparison of human-led and AI-assisted analyses. Paragraph **(3)** provides additional context on the study. Context information here can include details about the study objectives, locations, target population, and data collection method. The next two clauses **(4)** and **(5)** are optional and related to language information. If the prompt, results, and analysis are all written in the same language, this part is probably unnecessary. In other cases, the results may be written in multiple languages, and the analysis is expected in a specific language. Because GenAI models are known for being sensitive to language choice, returning different results for different languages, we recommend inserting the language indications for the tool. In our case, all prompts and analyses were made in English, but the raw re-

[2]`https://chatgpt.com`

[3]`https://openai.com`

[4]OpenAI models, including GPT-4o, have already scored highly in benchmarks such as HellaSwag (`https://rowanzellers.com/hellaswag/`) and MMLU (`https://paperswithcode.com/task/multi-task-language-understanding`), which compare the efficiency of LLMs in NLP tasks.

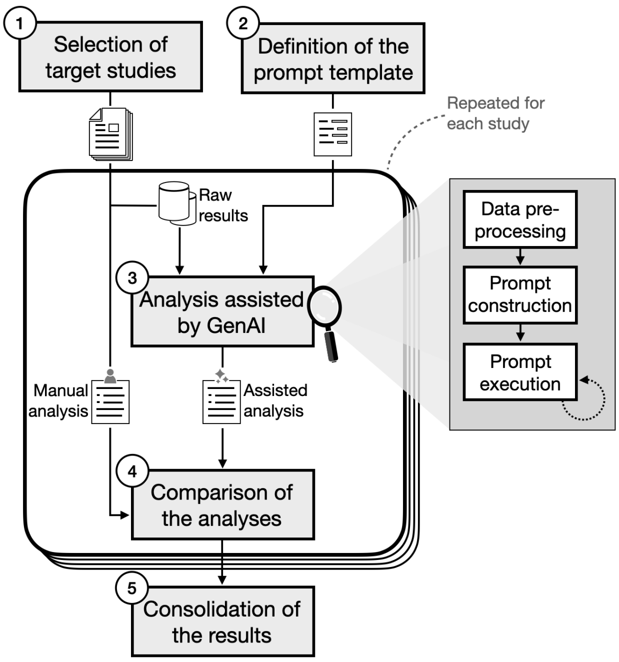

Figure 1. Overview of the methodology used for the study.

sults included a mix of English, Portuguese, and Italian texts (see language details for each study in Section 4). Finally, in part **(6)**, the researcher should provide the raw data to be analyzed by the AI tool. The current version of the ChatGPT interface allows users to attach files to the prompt, which may facilitate this task. In the current study, all data was submitted in plain text format. Any details further describing the data format can also be included here.

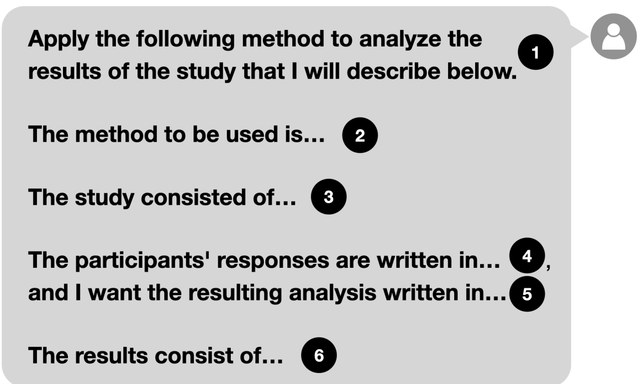

Figure 2. The prompt template defined for the study.

The steps displayed inside the central rectangle in Fig. 1 (Steps 3 and 4) were repeated for each of the four studies. Basically, these steps consisted of generating an AI-assisted analysis using ChatGPT and comparing it with human-made analysis done previously. The two inputs for this stage were the original study and the prompt template defined in one of the previous steps. The original studies material contained the raw results data and the analysis made by researchers for each study.

Step 3 takes the raw results and the prompt template to generate an analysis assisted by AI. This step is better described by three sub-steps that can be used as an inspiration for other researchers that want/need to adopt GenAI in their analysis procedures (see box on the right of Fig. 1). The first sub-step consists of performing any sort of data pre-processing deemed necessary. This sub-step should be used to format the raw results into something easier for the tool to process and understand. This is also the moment to check the tool's privacy policy and exclude from the results any sensitive data or data that cannot be publicized under the risk of violating previously assumed ethical commitments.

The second sub-step of Step 3 is prompt construction. At this moment, the gaps in the prompt template are filled with information about the specific study being analyzed. The raw results may be included as plain text or as attached files. The final sub-step of Step 3 is prompt execution. In our approach, we tried to produce a more comprehensive prompt from the start, but other approaches might prefer a more interactive conversation with the AI chatbot. If the response from the AI tool is considered incomplete or unsatisfactory, the researcher can opt to adjust the prompt and re-execute it or try to refine the response with follow-up prompts.

In **Step 4**, each researcher compared the previous human-made analysis with the analysis produced with AI assistance. We sought to find differences and similarities between the executions. In addition, possible advantages and disadvantages of each of the approaches were listed. The main criteria used to compare the two analyses were:

- **Number of categories:** The number of categories used in the thematic analysis.
- **Similarity between categories**: How similar the categories created in the two analyses were.
- **Granularity of categories**: Whether the categories were more fine- or coarse-grained.
- **General presentation and summary**: General aspects regarding the presentation of the analyses.

Step 5, the final step, consisted of consolidating the results of the comparisons for each of the four studies. During such consolidation, the team searched for general impressions and recommendations regarding the application of GenAI in the analysis of qualitative research results.

3.1 Ethical Issues

As previously stated, the four qualitative studies, Study 1 from da Rosa [2025]; da Rosa and Silveira [2025], Study 2 from Correa [2023]; Correa and Silveira [2024], Study 3 from Borges [2022], and Study 4 from Gnecco [2022], have all been approved by the Ethics Committee of the Pontifical Catholic University of Rio Grande do Sul[5], institution which they were conducted.

4 Results

This section presents the results obtained from the analysis of the four studies. A consolidated discussion of these results is presented in Section 5.

4.1 Study 1 - Online survey

This study consisted of a survey conducted through an online questionnaire. The study was part of a broader research project investigating interaction design approaches focused on multilingualism. More specifically, the survey aimed to uncover the behavior and difficulties of technology users in different multilingual contexts. To allow comparison between different environments, the questionnaire was applied to participants from two universities: one in Europe and the other in South America. Table 1 shows a summary of Study 1.

Table 1. Summary of Study 1: Online survey.

Research context:	Investigation of technology users' behavior in multilingual contexts.
Collection method:	Survey / online questionnaire.
Analysis method:	Text qualitative analysis.
Data format:	Responses to two open-ended questions in text format.
Scope of the study:	60 non-empty responses to the two open questions (out of 196 completed questionnaires).

As is the case in many surveys of this type, the questionnaire was composed of closed and open questions. In this analysis, we focused only on two open-ended questions. **Question 1** was: *Do you engage in any other activity using different languages? Which one(s)?* And **Question 2** was: *Are there any other experiences (positive or negative) related to multilingual interaction that you would like to share? Which one(s)?*. A total of 196 participants submitted valid responses for the questionnaire, from which 44 filled out Question 1 and 16 filled out Question 2, adding up to the 60 responses used for the analysis.

Because the open questions were not the main focus of the original research, no particular analysis method was applied to the qualitative results.

[5]Study 1 CAAE Number: 65338122.9.0000.5336, Study 2 and Study 3 CAAE Number: 54348321.1.0000.5336, and Study 4 CAAE Number: 68631523.1.0000.5336

The participants' responses to the two open questions were treated as complementary information, elucidating responses to the closed-ended questions. The responses were qualitatively analyzed, searching for similarities and noteworthy issues.

we classified the method that was applied to analyze the text as a *non-structured qualitative analysis*. In the original manual analysis, responses were first grouped by location (either Europe or South America) and then grouped in fine-grained categories by similarity in a process similar to thematic analysis. Responses from European participants were grouped in 11 categories, meanwhile the responses from South Americans resulted in seven categories. It was also observed in the original analysis that Europeans reported more daily activities, while South Americans reported more activities linked to the work environment.

Regarding Question 2, almost all responses were from European respondents; therefore, no separation by country was applied. The responses were initially classified as positive experiences and negative experiences (problems), with reports of problems being much more numerous. Reports of negative experiences were then grouped into four main categories, which included: *difficulty of understanding websites/apps lacking translation*; *wrong language settings*; *problems with code-switching and -mixing*; and *difficulties with configuration, typing, and searching, in less common languages*. Concluding remarks considering both questions were also included in the original analysis. For example, responses served to illustrate a list of six challenges and opportunities related to the intersection of HCI and multilingualism.

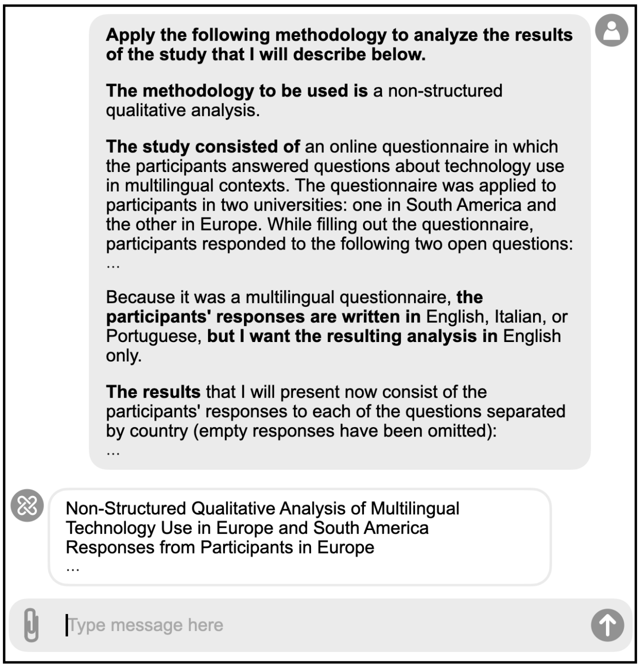

Figure 3. The prompt submitted to ChatGPT for the analysis of results from Study 1 - Online survey.

The prompt for Study 1 was built following the chosen template and considering the aforementioned study description (see Fig. 3). Since the online questionnaire was multilingual, special care was taken in regard to notifying the language of the responses, which were written either in English, Italian, or Portuguese. All the 60 analyzed responses were formatted together in a single text with indications of the location (Europe or South America), question (Question 1 or Question 2), and respondent (participants were identified by random identifiers to ensure anonymization).

ChatGPT's response to the prompt included a summarizing title, an analysis of the results grouped by country and question, and a final conclusion. With respect to the analysis of Question 1, the AI tool grouped responses into generic categories. For the responses from Europe, four categories were identified *scientific and professional activities*, *social interactions*, *sports and hobbies*, and *miscellaneous*. Responses from South America for the same question were grouped into three categories: *professional and educational activities*, *leisure activities*, and *travel and miscellaneous*. For each location, ChatGPT highlighted the main characteristics of the responses: *daily integration of multiple languages in both professional and personal domains* in Europe; and *professional and educational contexts* in South America. Regarding Question 2, ChatGPT also recognized that responses were more numerous in Europe. The assisted analysis started by classifying responses into positive and negative experiences. Negative experiences reported by participants in Europe were associated with three categories: *digital frustrations*, *cultural challenges*, and *administrative hurdles*. The conclusion generated by ChatGPT consisted of a short paragraph that attempted to summarize the analyzed dataset but was more focused on the answers to Question 2. In its final remark, the AI tool sought to extrapolate the analysis of the results to a broader context: "The insights from both groups highlight the importance of improving digital tools and interfaces to better support multilingual users in diverse contexts".

Despite the difficulty to describe the analysis methodology adopted in the original study, the obtained assisted analysis presents many similarities when compared to the manual human-made analysis. The general grouping and categorization of the responses presented together with concluding remarks, typical aspects of a qualitative analysis, can be considered similarities. Similar key findings can also be observed in both analysis, such as the characterization of multilingual activities reported by Europeans and South Americans.

On the other hand, remarkable differences can also be observed. The categorization of responses to the first Question (multilingual activities) was more fine-grained in the manual analysis, while the AI tool opted for a more coarse-grained classification. ChatGPT ended up with only four categories for the European sample and three categories for the South American sample, compared to 11 and 7, respectively in the manual analysis. This difference is understandable, given that the manual analysis consisted of merely grouping activities that were similar in its description, meanwhile the assisted analysis opted for a categorization closer to that of thematic analysis. This difference in defining themes was not so pronounced in the case of Question 2, when the number of responses was much lower, and both analysis arrived at a similar number of categories.

The categorization performed by ChatGPT over the dataset was not free of caveats though. Some of the chosen themes can be considered too generic such as the *miscellaneous* and *travel and miscellaneous* categories in the analysis of responses to Question 1. In the analysis of Question 2, some categories can be considered ill-defined or overlapping: for instance, *digital frustrations* and *cultural challenges*. These problems could be alleviated by choosing a better-defined analysis method during prompt construction, such as thematic analysis. An alternative would be to try refining the AI tool's response by issuing follow-up prompts. These actions, however, would be out of the scope of this experiment. Table 2 presents the comparison between the manual and AI-based analyses for Study 1.

4.2 Study 2 - Focus group

This study involved a remote focus group conducted using Zoom, with six selected participants (identified as P1 to P6) and two facilitators. The research aims to assist designers in developing data narrative visualizations that enhance understanding and engagement from the end user's perspective. The focus group sought to identify how data types are used with different genres, narrative structures, visual narratives, and other elements composing a narrative visualization.

Participants were encouraged to express their opinions interactively during the session, which we recorded for further analysis. The activity involved planning a narrative data visualization based on content presented by the facilitators. After the session, we transcribed the discussions to enable a detailed examination of the participants' approaches and preferences in creating narrative visualizations.

All text transcriptions in the proposed prompt replicate the same analysis using ChatGPT (see Fig 4). The method used for both approaches was thematic analysis. Table 3 summarizes this study.

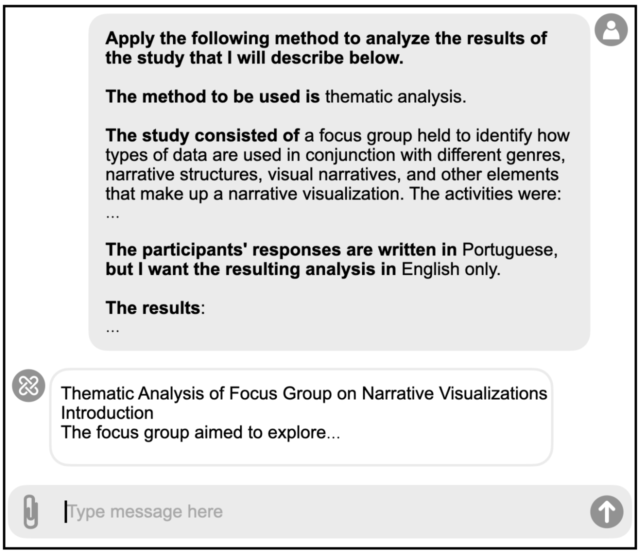

Figure 4. The prompt submitted to ChatGPT for the analysis of results from Study 2 - Focus group.

The results from the manual analysis identified three themes: narrative genres, visualization techniques, and visual structures. Regarding the *genre* theme, the participants expressed common ideas. P4 suggested combining the Annotated Chart with the Partitioned Poster genre, explaining that when users explore a poster, there may be moments when they encounter information without understanding the designer's

Table 2. Comparison between manual and AI analysis – Study 1

Subject	Manual analysis	AI analysis
Grouping and theme categorization	General grouping and categorization of the responses presented together with concluding remarks.	From both Question 1 and Question 2, the broader context classification.
Number of categories	Question 2 has led to a more similar number of categories. Question 1: 11 and 7 categories found in the European and South American sets.	Question 1: 4 and 3 categories found. In Question 2, the negative experiences were categorized into a similar number of categories.

Table 3. Summary of Study 2: Focus Group.

Research context:	Investigation of the relation between data types and narrative elements.
Collection method:	Focus group.
Analysis method:	Thematic analysis.
Data format:	Audio transcribed to plain text.
Scope of the study	Video call session conducted with 6 participants discussing and interacting.

intended message. P4 emphasized that adding explanatory points could enhance interaction. P1 said the Annotated Chart could also help show historical moments or significant legal changes. The participants discussed different *visualization techniques*. Some participants opted for filters based on country and year, while P4 suggested using a correlation technique, such as a scatter plot, to observe data dispersion. P5 added that using cards to display values, total enrollments, total investments, and a matrix or table could help clarify. Regarding *visual structures*, the group leaned toward drill-down stories and the Martini glass structure, with P1 noting that the drill-down approach would work well since the visualization covered multiple continents, allowing for an expansion based on user preferences.

The results of the ChatGPT-assisted analysis identified seven categories, each linked to a theme, a description, and illustrative quotes. The first category, familiarity with narrative visualization, is related to the theme *varied levels of familiarity with narrative visualization*, in which the description indicated that some participants already knew about narrative data visualizations and others did not. The quote examples showed that P5, P2, and P1 were already familiar with narrative visualization, while P3 and P4 knew storytelling concepts but not the specific genres presented. Another category, decision-making in narrative planning, was linked to the theme of *collaborative and iterative decision-making process*. The description highlighted the flexibility and adaptation required during the creative process, emphasizing how narratives evolve as different elements come together. The category challenges in abstracting and planning, associated with *difficulty in abstract planning without concrete data*, described how participants, especially P4 and P6, struggled to plan narrative visualizations without direct access to datasets.

The preferences for visualization genres included the theme *preference for versatile and interactive genres*, showing that participants favored genres that allowed for greater interactivity and flexibility, such as the partitioned poster and annotated chart. P1 described the partitioned poster as a versatile tool that unifies data in one place with interactivity. In the category perception of genre applicability, linked to the theme of *genre selection based on audience and narrative goals*, participants emphasized that the choice of genre should be guided by the audience's needs and narrative goals rather than just the data. P5 suggested that any dataset could fit into any genre, depending on how the narrative is constructed. In the category interactivity and user engagement, associated with the theme *importance of interactivity in enhancing user engagement*, participants agreed that interactivity is crucial for making narrative visualizations more engaging and effective. The ability to filter and explore data was identified as a significant factor in choosing a genre, with P5 mentioning that filtering data by country or year could enhance user engagement. Lastly, the category learning and adaptation was linked to the theme of *learning new concepts and adapting to them*. For instance, P1 remarked on the importance of learning about genre categorization for effective communication. ChatGPT then generated a brief conclusion summarizing these results.

After comparing these two approaches, manual and ChatGPT-assisted, we observed that ChatGPT generated seven categories and themes, and three themes were identified throughout the manual analysis. The manual analysis grouped the themes into broader topics, while ChatGPT generated more categories and themes. However, this fragmentation can be risky because not all participants expressed opinions about all the themes. On the other hand, the tool brought complete descriptions and examples in all cases. Two approaches identified this study's main topics and brought illustrative examples. The combination of manual and ChatGPT-assisted analyses offers a good framework for future qualitative research, balancing the depth of human analysis with the efficiency and comprehensiveness of GenAI tools. Table 4 presents the comparison between the manual and AI-based analyses for Study 2.

4.3 Study 3 - Workshop

This study is part of a research project that blends End-User Development and narrative data visualizations to investigate how end-users utilize customization features provided within a narrative visualization to explore the data. In this sense, we aimed to understand how these features enable users to explore the narrative and tailor it to their preferences and understanding.

We held workshop sessions with end-users to observe how they use the customizable features to explore a given narrative developed using our proposed narrative data visualization approach to support visualization design. The narrative

Table 4. Comparison between manual and AI analysis – Study 2

Subject	**Manual analysis**	**AI analysis**
Grouping and theme categorization	Two approaches identified similar main topics and brought illustrative examples. Themes are grouped into broader topics.	More categories and themes identified.
Number of categories	4 categories	7 categories

data visualization was based on the partitioned poster genre. This choice relies on the partitioned genre being considered optimal due to its characteristic of presenting multiview narratives.

The sessions were attended by 45 participants divided into nine groups. The workshops were facilitated using the Figma tool, a platform that allowed for the creation of interactive prototypes of narrative visualizations and the simulation of a customization design space. Participants were encouraged to explore the narrative and utilize the available customization features to tailor the visualization to their preferences. The researchers observed the participants' interactions and collected feedback through discussions and questionnaires with a mix of open and closed questions. The data gathered during these workshops were then analyzed using thematic analysis to identify patterns and themes in how the participants utilized the customization features and their overall experience with the customizable narrative visualization. Table 5 summarizes the study design. Fig 4 presents an excerpt of the prompt submitted to the ChatGPT.

Table 5. Summary of Study 3: Workshop sessions.

Research context:	Investigation of how end-users utilize customization features to tailor a narrative data visualization.
Collection method:	Questionnaire.
Analysis method:	Thematic Analysis.
Data format:	Responses to mix of open and closed questions in text format.
Scope of the study:	9 groups (45 attendees) non-empty responses.

We posed seven questions. Four of these questions were dedicated to mapping the attendees' profiles. The remaining questions were related to participants' perceptions of the study. We asked them about their prior knowledge of narrative data visualizations, perceptions about the tailoring techniques provided during the session, and how they support the narrative exploration. Finally, we asked them for suggestions regarding the design space and topic.

To ease the understanding, we will present our results organized by the kind of analysis performed. First, we present the analysis conducted by a human being employing the method chosen. Then, we present the same task, which was performed with the assistance of chatGPT.

Regarding the analysis conducted by a human, after analyzing the results, we organized them into three categories: *access to alternative viewpoints*, *customization of narratives*, and *general contributions to customization*.

The category of *access to alternative viewpoints* addresses attendees' contribution regarding their perception of how the

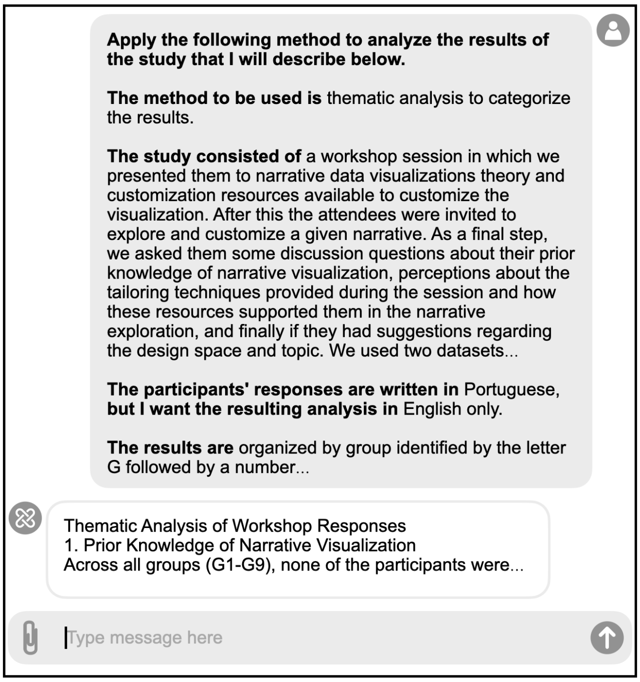

Figure 5. The prompt submitted to ChatGPT for the analysis of results from Study 3 - Workshop.

customizable resources supported them in exploring the narrative—the category of *customization of narratives* addresses. The third category, *general contributions to customization*, addresses participants' suggestions for further customization possibilities. These suggestions included the ability to customize component colors without compromising image proportions.

In addition to this classification, we reported general results based on the comments and observations made during the workshop sessions' execution. These results covered their engagement, the use of the design space, and the use of the available customization resources. We also analyzed the differences in how each group approached the narrative according to their background.

Regarding the analysis performed by chatGPT, we observed that the IA identified four themes after analyzing the participants' responses. Moreover, within each theme, the IA provided a set of categories. The themes identified were *prior knowledge of narrative visualization*, *customization flexibility*, *suggestions for improvement*, and *user-driven enhancement requests*.

For the first theme, prior knowledge of narrative visualization, the IA proposed the category of *lack of prior knowledge* to describe participants' lack of prior knowledge of the subject.

Following, the IA organized the categories of second and third themes according to their classification as positive or negative aspects of narrative visualization. The positive aspects were *flexibility and exploration*, *enhanced data emphasis*, and *user engagement and experience*. The negative aspects, in turn, relate to *limited effectiveness* and *the need for additional features*.

Finally, for the last theme, user-driven enhancement requests, the categories suggested by the chatGPT's analysis were *color customization*, *component resizing and proportions*, and *additional data and presentation features*.

By comparing the results of the human and chatGPT analysis, we can recognize different approaches over the same content. The themes identified in the human analysis are broad and related to the research project's context. Themes like access to alternative viewpoints, customization of narratives, and general contributions to customizations represent an attempt to capture the tool's functionality and essential aspects of the research context, such as user experience impact and narrative comprehension. The themes proposed by chatGPT were focused on tools' functionality and users' perceptions of them.

Another aspect we noted is that the human analysis tries to capture the meaning of participants' responses regarding actions and motivations. Evidence of this is the discussion of how each participant profile approached the narrative according to their background (e.g., data science students versus other IT students). In contrast, the chatGPT analysis is more descriptive, summarizing the main topics and participants' suggestions. The IA analysis did not dive into the implications of the results for the tool's design or users' comprehension of narratives.Indeed, a close examination revealed no substantial parallels between the aforementioned methodologies. Notwithstanding, the principal discrepancies that have been identified are illustrated in the Table 6.

To sum up, the human analysis demonstrates a deeper comprehension of the research context, and it is more prosperous than chatGPT analysis, which is helpful but more descriptive than interpretative.

4.4 Study 4 - Interviews

This study represents a portion of an investigation focusing at understanding the utilization of interaction design patterns in design systems. The objective of the interviews was to understand how design systems' users, who are typically user experience designers, were applying interaction design patterns and how they were relating these patterns to design systems.

We conducted a series of ten semi-structured interviews with user experience designers from a variety of backgrounds and career stages. The sole requisite for participants is that they must have previously been employed by an organization with a design system in place. In the present study, we posit that the selection of a single participant from the original ten as the subject of interest was sufficient to obtain the requisite material.

The interviews were conducted via the online video meeting platforms Zoom and Google Meet, according to the preferences of the respective participants. Researchers documented the participants' responses and engaged in observation of their interactions. Subsequently, the transcripts of the interviews were subjected to thematic analysis in accordance with the principles of grounded theory methodology, using an appropriate software for this purpose called MAXQDA (version 2022). Table 7 provides this study structure.

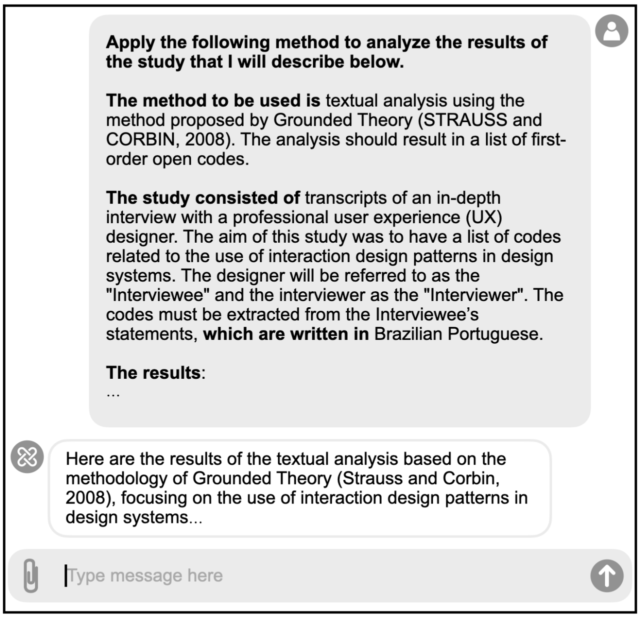

Figure 6. The prompt submitted to ChatGPT for the analysis of results from Study 4 - Interviews.

The interview process entailed the administration of all nine questions previously outlined in the interview guide. The initial four questions addressed the candidate's prior familiarity with and utilization of interaction design patterns. The subsequent three questions addressed the practice of collaboration and the incorporation of references when applying interaction patterns in professional endeavors. The final two questions addressed the topic of design systems and their relationship to the aforementioned patterns. Fig. 6 shows the prompt used for the AI-assisted analysis.

The manual analysis of the interview data identified nine open codes. *The interaction patterns have already been tested in the context of work*. This first code is related to the designer's view of the interaction pattern as a tool that validates the user and other professionals' experiences. The second code identified by the manual analysis pertains to the *utilization of online references*, which refers to the application of design patterns in the course of one's daily work. The following code pertains to the notion held by the participant that a *novel pattern may be generated during a collaborative interaction* with a colleague. Furthermore, the fourth code is also relevant to the subject of the collaborative process, which pertains to the *challenges encountered when attempting to collaborate with other professionals in the field of design* with the aim of exchanging knowledge. The fifth and sixth codes identified pertain to the documentation of interaction design patterns. The former code addresses the issue of *how a design pattern's documentation is not readily accessible from its practical application in the field*. The latter code pertains to the rhythm of daily work and highlights the fact that the *stressful routine impedes the documentation of a pattern*. This may result in the potential loss of an excellent pattern, which may have been tested and validated with users, due to the lack of time to document it and disseminate it to other

Table 6. Comparison between manual and AI analysis – Study 3

Subject	**Manual analysis**	**AI analysis**
Grouping and theme categorization	Broader themes. Captured the meaning regarding motivations and actions from the participants.	More descriptive analysis.

Table 7. Summary of Study 4: Interviews.

Research context:	Investigation of user experience designers utilize interaction design patterns in design systems
Collection method:	Interviews.
Analysis method:	Open codification phase from the Grounded Theory textual analysis methodology.
Data format:	Responses to the open questions in text format.
Scope of the study	One participant of a total of ten was included in a nine-question open interview process.

professionals.

The preceding two codes are highly pertinent to the seventh code, which demonstrates the interviewee's perception of the user experience field. The seventh code pertains to the numerous transformations that the user experience design profession has undergone in recent years. Furthermore, it can be seen as an indication of the general perception that the *role of the user experience designer is one that is in a state of constant evolution*. This, in turn, gives rise to a corresponding evolution in technology and interaction patterns.

Ultimately, the two remaining codes illustrate a participant's perception of the *utility of interaction design patterns in a design system's documentation* and how it could facilitate the *adaptation of patterns to the diversity of contexts present in the real field*. This approach is more aligned with the design systems theme.

With regard to the analysis of the interview used in the study, which was conducted by chatGPT, fifteen codes were identified. The first code to be commented is the *lack of documentation of interaction design patterns*. In the analyzed interview, the lack of formal documentation for design patterns was frequently mentioned as a challenge for consistency. The following three codes refer to the *creation of components in the Figma tool*, the *use of these components* and the *difficulty of reusing these components due to the lack of consistency among designers regarding interaction design patterns*. The fifth code identified expressed *fear of becoming outdated in the profession*, mentioning efforts to stay updated with new tools and techniques. These first five codes express the participant's concern about the consistency in interaction design solutions and the importance given to patterns in this context.

There are four other codes related to adaptation and changing interaction patterns according to context. The first code cites the *frequent need to adapt patterns even without mentioning context*, the second code cites *adaptation depending on context*, and the last one mention *the use of online references to be updated about the last patterns tendencies*. These aforementioned codes can be interpreted as an important signal of how often there is a need for an adaptation of a pattern to real scenarios in the practice field. However, on the topic of customization, this leads the participant to concerns about the *excessive customization done by designers*, which can result in a lack of uniformity in projects.

The subsequent group of codes pertains to the domains of communication, feedback, and documentation within the sphere of interaction design. The themes are interrelated in that the initial point of discussion is the *lack of effective communication among designers*, which can lead to the importance being placed on the collaboration process among these designers, and *how constant feedback is a critical part of the design process, despite the emotional challenges it may bring*. The argument put forth by the interviewee posits that the adaptation or *creation of a pattern during the course of an ongoing project may act as a barrier to evolution due to the fear of altering what has already been done and approved*. Furthermore, the participant put forth the suggestion that *professionals should be granted access to comprehensive, objective, and readily available documentation*, citing the potential benefits of a good interaction pattern documentation living in Figma.

Finally, there are the latest two codes that relate to the perception of design systems and the importance of standardization of design patterns. The discussion given from the interviewee's perspective is that in the first instance, even if there is a search for external references, which was aforementioned in another code, it is very *important to maintain consistency of interface patterns in a context using an existing design system*. In the same context, the participant has shown what ended up being another code, the belief in the *importance of a centralization of patterns within the project or company*. A resumed comparison between those two methods can be seem in Table 8.

The manual approach was driven to a smaller number of codes, and the AI approach was able to identify a greater variety of codes. The human-generated codes seemed to be more interpreted, and the AI codes seemed to be more objective and descriptive. Nevertheless, when it comes to these two types of analysis, manual analysis and AI-driven analysis, we could see that there was a similarity in a certain types of codes, and it is clear that both proved to be effective in identifying potentially relevant quality material for further analysis.

5 Discussion

The use of generative AI, particularly ChatGPT, in qualitative research is an emerging field. Our findings align with previous studies, such as those by Saare [2024], Skjuve *et al.* [2023], and Barambones *et al.* [2024], which confirm the potential of AI to accelerate the initial analysis of data, but

Table 8. Comparison between manual and AI analysis – Study 4

Subject	Manual analysis	AI analysis
Grouping and theme categorization	Similarities in codification: both analyses named a code for "lack of documentation" in interaction design patterns in the context studied. Less variability of codes. The human-generated codes seemed to be more interpreted.	More variability of codes. The AI codes seemed to be more objective and descriptive.
Number of categories	9 codes.	15 codes.

also reinforce the indispensableness of human oversight to ensure the depth of the interpretation. However, by replicating a multimethod research design (questionnaires, focus groups, workshops, and interviews), our study advances this understanding by revealing how the effectiveness and limitations of AI vary significantly depending on the qualitative methodology employed.

5.1 Analysis of AI's depth, validity, and biases

Hamilton *et al*. [2023], Sinha *et al*. [2024], and Lopez-Fierro and Nguyen [2024] explored the role of AI, specifically ChatGPT and GPT-4, in assisting qualitative research. Hamilton *et al*. [2023] compared AI-generated themes with those identified by human coders in interviews, finding overlap but emphasizing that AI cannot replace the interpretive depth human researchers provide. Sinha *et al*. [2024] examined GPT-4's role in open coding within a grounded theory framework, noting that while AI can expedite coding by identifying uncoded segments and offering alternative interpretations, human refinement is still crucial. They also highlighted technical limitations, such as token constraints for large datasets. Lopez-Fierro and Nguyen [2024] focused on human-AI collaboration in qualitative coding, suggesting that AI can enhance transparency in the process and prompt new insights. However, human oversight is essential to ensure proper contextualization and alignment with research goals. All three studies stress that AI can complement but not replace human researchers' nuanced and interpretive work.

We agree with the need for human oversight of AI-generated results. In this research, ChatGPT was very helpful in generating an initial descriptive analysis of the raw data, according to the prompts provided. Based on the input, it generated detailed themes or codes, reflecting its ability to respond accurately to structured prompts. However, as observed in previous studies, human oversight was essential for evaluating and refining AI-generated results. For instance, in the online survey (Study 1), ChatGPT generated themes similar to those identified by human analysis, but the latter was more detailed, uncovering more categories. In the focus group (Study 2), the AI identified broader categories, sometimes misclassifying responses when not all participants had engaged with specific topics. ChatGPT generated more codes in the interview (Study 4). In contrast, its analysis in the workshop (Study 3) was overly descriptive and lacked the necessary depth to explore critical aspects of the research thoroughly.

The direct comparison between human-led and AI-assisted analysis across our four studies revealed a consistent pattern: while AI excelled at identifying descriptive and broad themes, human analysis consistently provided greater detail and contextual depth. Our work demonstrates that human validation is not just a "checking" step, but a dialectical process of refinement, where the researcher questions, deepens, and contextualizes the AI's raw outputs.

5.2 Recommendations for using generative AI in qualitative analysis

Table 9 summarizes the key recommendations derived from our findings. These recommendations aim to guide researchers in the responsible and effective integration of GenAI tools into qualitative workflows.

Based on the findings, we highlight a recommendation - **human validation** - indicating that it's necessary to ensure a human validates the result of AI analysis. Once AI analysis does not replace human interpretation, researchers must evaluate and interpret the results, ensuring they align with the research context and goals.

While the analyzed studies demonstrate the potential and limitations of integrating generative AI, particularly ChatGPT, into qualitative research workflows, this study offers a distinct approach by replicating a multi-method research design using ChatGPT. Unlike studies such as those by Yan *et al*. [2024] and Hamilton *et al*. [2023], which primarily focus on thematic analysis and AI's contributions to coding efficiency, the research is based on four distinct qualitative methodologies: questionnaires, focus groups, workshops, and interviews. Applying ChatGPT prompts to replicate the original research findings aims to explore how AI-generated data compares across these varied methods.

These contrasts reveal that while ChatGPT can effectively assist in initial analyses, the methods used and researchers' perceptions significantly influence the final results. Perkins and Roe [2024] explored the use of GenAI in qualitative and quantitative analysis. They provided recommendations to researchers using these tools and identified that the AI-generated result's relevance depends on the ability to create great prompts. Considering this, another recommendation is a **prompt design** that can effectively mitigate it. It's essential that the prompt includes information about the research context, language, data collection methods, and desired analysis approach. Additionally, the iterative process of refining prompts and re-engaging with the AI could further improve outcomes, though this was outside the scope of this study. Nevertheless, the prompt design allows future researchers to engage in more in-depth prompt-AI interaction.

Still, about the prompt design, a recommendation about

Table 9. Key recommendations provided.

Prompt design	Craft clear, detailed prompts to guide the AI's analysis.
Data preprocessing	Format raw data for AI comprehension and ensure anonymization.
Human validation	Ensure human validation of AI outputs to align with research goals.
Documentation	Keep records of prompts, outputs, and how AI was used in the analysis.

data processing and anonymization emphasizes that the data format must be easily and understandable by the AI. In addition, all sensitive information must be removed to comply with privacy regulations and ethical considerations.

It is noteworthy the practice of **documentation**. Document the prompts used, the AI results generated, and how they were incorporated into the final analysis to ensure the study's replicability. Table 9 summarizes the recommendations provided in this study.

These recommendations are not exhaustive, but are based on empirical evidence from our case studies. Future research could expand upon these insights by testing refined prompt strategies, combining multiple GenAI tools, or exploring human-AI co-creation scenarios.

Across all studies, ChatGPT identified the main topics in the contexts, demonstrating that our prompt design is handy across different qualitative methods. However, as previously pointed out, human evaluation remains indispensable to ensure the analysis is aligned with the research goals and contextual nuances.

6 Final Remarks

This study aimed to explore the potential of human-AI collaboration in qualitative analysis by examining the use of ChatGPT to support researchers in interpreting textual data. We used four different qualitative methods from HCI studies: questionnaires, focus groups, workshops, and interviews. By applying the same prompt template across varied contexts, we assessed how GenAI performs in diverse research scenarios. Our findings show that AI like GenAI can be helpful in qualitative research, providing efficiency and new perspectives. However, we also emphasize the crucial role of human supervision in ensuring the accuracy and relevance of AI-generated analysis.

We noted that GenAI has the potential to facilitate the analysis of qualitative data, particularly in HCI research, where textual analysis is commonly used. AI's capability to generate themes, codes, and summaries can significantly expedite the research process, allowing researchers to concentrate on interpretation and synthesis. However, we also found that AI-generated analyses can sometimes be too broad or lack the nuanced understanding of human researchers. The differences observed between manual and AI-assisted analyses underscore the importance of human validation and interpretation, i.e., it means that while AI can speed up the analysis process, it is essential for human researchers to review and interpret the results to ensure their accuracy and contextual relevance.

The study's implications extend beyond HCI research, offering valuable insights for any field that relies on data analysis. Integrating GenAI into research workflows can be a milestone in approaching qualitative analysis, making it more efficient and accessible. In addition, our methodology and prompt are broad enough to be applied by other researchers.

To summarize, we contributed to the growing literature on human-AI collaboration in qualitative research. The findings suggest that GenAI can be a powerful tool for enhancing research efficiency and generating unique insights. Yet, the study serves as a reminder that AI is a tool to augment, not replace, human expertise. As GenAI continues to evolve, we must explore its potential and limitations, striving for a balanced and synergistic partnership between humans and AI in pursuing knowledge.

Following, we present the study's limitations and future work we envision to the research of Human-AI collaboration.

6.1 Limitations

A limitation of this study lies in the approach used regarding ChatGPT. We chose not to provide additional information to refine the AI-generated results, which may have limited the depth and specificity of the analyses. We decided to assess the tool's capabilities in its default configuration, but we acknowledge that including more detailed guidance could lead us to different results. In addition, we can explore prompt engineering in the future by trying configurations that were out of our scope.

Additionally, it's important to note that the findings in this study are not comprehensive but rather represent a perspective based on our experience with ChatGPT. Comparing the AI-generated results with our manual analyses introduces a potential bias, as the same researchers were involved in both stages. This overlap in roles may have influenced, perhaps unconsciously, the interpretation of the data and the assessment of the quality of the analyses generated by ChatGPT. To overcome this, we performed rounds of discussion among the authors to evaluate the results.

6.2 Future Work

Human-AI collaboration is an increasing topic in all communities' research. Due to the rapid advancement of AI tools, assisted research has gained considerable attention. In this scenario, we envision future work that we can address in the short, medium, and long term.

In this work, we focused on the ChatGPT tool because of its popularity. In the short term, we envision conducting qualitative research to investigate the capabilities and limitations of other emerging GenAI tools, such as Google Gemini.

In the medium term, we envision investigating AI's impact on diverse qualitative methods. In this work, we explored four distinct qualitative methods. However, we expect to widen the range of qualitative methods to cover other methods.

In the long term, we plan to investigate the long-term implications of human-AI collaboration by conducting longitudinal studies to track its evolution over time and examine its impact on research practices, roles, and the nature of knowledge production.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior – Brasil (CAPES) – Finance Code 001 and supported by the Ministry of Science, Technology, and Innovations, with resources from Law No. 8.248, dated October 23, 1991, within the scope of PPI-SOFTEX, coordinated by Softex, and published in the Residência em TIC 02 - Aditivo, Official Gazette 01245.012095/2020-56. Da Rosa thanks the Federal Institute of Rio Grande do Sul (IFRS) for the grant of a paid leave during the study period. Da Rosa also acknowledges CAPES for the grant of a scholarship during a six-month period of studies in Switzerland (*Projeto Institucional de Internacionalização CAPES-PrInt*). Correa thanks the Federal Institute Farroupilha (IFFar) for the grant of a paid leave during the study period. Correa also acknowledges CAPES for the grant of a scholarship during a six-month period of studies in Canada (*Projeto Institucional de Internacionalização CAPES-PrInt*).

Authors' Contributions

Mariana Gomes Borges - Conceptualization, Writing, Review - original draft, Writing - review & editing
Claiton Marques Correa - Conceptualization, Writing, Review -original draft
Diego Moreira da Rosa - Conceptualization, Writing, Review - original draft
Andrea Gnecco - Conceptualization, Writing, Review -original draft
Milene Selbach Silveira - Project administration

Competing interests

There is no competing financial and/or non-financial interests in relation to the work described.

Availability of data and materials

All data generated or analyzed during this study are included in this published article. There are no supplementary materials associated with this article.

References

Barambones, J., Moral, C., de Antonio, A., Imbert, R., Martínez, L., and Villalba-Mora, E. (2024). Chatgpt for learning hci techniques: A case study on interviews for personas. *IEEE Transactions on Learning Technologies*. DOI: https://doi.org/10.1109/TLT.2024.3386095.

Borges, M. G. (2022). Apoio à construção de visualizações narrativas de dados com ênfase em gêneros narrativos. Master's thesis, Pontifical Catholic University of Rio Grande do Sul.

Correa, C. M. (2023). *An approach for designing and exploring customizable narrative data visualizations*. PhD thesis, Pontifical Catholic University of Rio Grande do Sul.

Correa, C. M. and Silveira, M. S. (2024). An approach for designing and exploring customizable narrative data visualizations. In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, pages 1–13. ACM. DOI: https://doi.org/10.1145/3702038.3702080.

da Rosa, D. and Silveira, M. (2025). Human-computer interaction patterns of design for multilingualism. In *Anais do XXIV Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, pages 862–880, Porto Alegre, RS, Brasil. SBC. DOI: https://doi.org/10.5753/ihc.2025.12228.

da Rosa, D. M. (2025). *Human-Computer Interaction Patterns of Design for Multilingualism*. PhD thesis, Pontifical Catholic University of Rio Grande do Sul.

dos Anjos, J. R., de Souza, M. G., de Andrade Neto, A. S., and de Souza, B. C. (2024). An analysis of the generative ai use as analyst in qualitative research in science education. *Revista Pesquisa Qualitativa*, 12:1–29. DOI: https://doi.org/10.33361/RPQ.2024.v.12.n.30.724.

Feuston, J. L. and Brubaker, J. R. (2021). Putting tools in their place: The role of time and perspective in human-ai collaboration for qualitative analysis. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–25. DOI: https://doi.org/10.1145/3479856.

Gnecco, A. (2022). O uso dos padrões de design de interação em design systems. Master's thesis, Pontifical Catholic University of Rio Grande do Sul.

Hamilton, L., Elliott, D., Quick, A., Smith, S., and Choplin, V. (2023). Exploring the use of ai in qualitative analysis: A comparative study of guaranteed income data. *International Journal of Qualitative Methods*, 22:1–13. DOI: https://doi.org/10.1177/16094069231201504.

Jiang, J. A., Wade, K., Fiesler, C., and Brubaker, J. R. (2021). Supporting serendipity: Opportunities and challenges for human-ai collaboration in qualitative analysis. *Proceedings of the ACM on Human-Computer Interaction*, 5:1–23. DOI: https://doi.org/10.1145/3449168.

Liu, Z., Dai, J., Conati, C., and McGrenere, J. (2025). Envisioning AI support during semi-structured interviews across the expertise spectrum. *Proceedings of the ACM on Human-Computer Interaction*, 9:1–29. DOI: https://doi.org/10.1145/3710909.

Lopez-Fierro, S. and Nguyen, H. (2024). Making Human-AI contributions transparent in qualitative coding. In *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024, pp. 3-10*. International Society of the Learning Sciences. DOI: https://doi.org/10.22318/cscl2024.352932.

Morgan, D. L. (2023). Exploring the use of artificial intelligence for qualitative data analysis: The case of chatgpt. *International journal of qualitative methods*, 22:1–10. DOI: https://doi.org/10.1177/1609406923121124.

Perkins, M. and Roe, J. (2024). Generative AI tools in academic research: Applications and implications for qualitative and quantitative research methodologies. *arXiv preprint arXiv:2408.06872*. DOI: https://doi.org/10.48550/arXiv.2408.06872.

Saare, K. (2024). Natural language based AI tools in interaction design research: Using chatgpt for qualitative user research insight analysis.

Sinha, R., Solola, I., Nguyen, H., Swanson, H., and Lawrence, L. (2024). The role of generative AI in qualitative research: Gpt-4's contributions to a grounded theory analysis. In *Proceedings of the Symposium on Learning, Design and Technology*, pages 17–25. DOI: https://doi.org/10.1145/3663433.3663456.

Skjuve, M., Følstad, A., and Brandtzaeg, P. B. (2023). The user experience of ChatGPT: findings from a questionnaire study of early users. In *Proceedings of the 5th international conference on conversational user interfaces*, pages 1–10. DOI: https://doi.org/10.1145/3571884.3597144.

Wachinger, J., Bärnighausen, K., Schäfer, L. N., Scott, K., and McMahon, S. A. (2024). Prompts, pearls, imperfections: Comparing chatgpt and a human researcher in qualitative data analysis. *Qualitative Health Research*, 0:1–16. DOI: https://doi.org/10.1177/10497323241244669.

Wang, D., Churchill, E., Maes, P., Fan, X., Shneiderman, B., Shi, Y., and Wang, Q. (2020). From Human-Human collaboration to Human-AI collaboration: Designing AI systems that can work together with people. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, page 1–6, New York, NY. Association for Computing Machinery. DOI: https://doi.org/10.1145/3334480.3381069.

Yan, L., Echeverria, V., Fernandez-Nieto, G. M., Jin, Y., Swiecki, Z., Zhao, L., Gašević, D., and Martinez-Maldonado, R. (2024). Human-AI collaboration in thematic analysis using ChatGPT: A user study and design recommendations. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–7. DOI: https://doi.org/10.1145/3613905.3650732.

Yang, Q., Steinfeld, A., Rosé, C., and Zimmerman, J. (2020). Re-examining whether, why, and how human-ai interaction is uniquely difficult to design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–13, New York, NY. Association for Computing Machinery. DOI: https://doi.org/10.1145/3313831.3376301.