Flexibility & Iteration: Exploring the Potential of Large Language Models in Developing and Refining Interview Protocols

Jessica L. Parker
Massachusetts College of Pharmacy and Allied Health Sciences, jessica.parker@mcphs.edu

Veronica M. Richard
Dissertation by Design, veronica@dissertationbydesign.com

Kimberly Becker
Iowa State University, kimberly@academicinsightlab.org

Flexibility & Iteration: Exploring the Potential of Large Language Models in Developing and Refining Interview Protocols

Abstract

This article investigates the potential of Large Language Model (LLM) tools like ChatGPT in aiding researchers in the development and refinement of interview protocols. We found that ChatGPT could generate appropriate interview questions, craft key questions, provide feedback on protocols, and simulate interviews, indicating its potential to reduce time and effort, particularly when human resources are limited. This article builds upon previous authors' insights and suggestions regarding developing and refining interview protocols to maximize the chances of achieving research aims, especially for novice researchers. Additionally, the researchers highlight the flexibility of these tools in adapting to different research contexts and cultural considerations. Ethical considerations and human oversight are emphasized as critical components in the responsible implementation of these tools. The research also paves the way for further exploration into the integration of LLMs into other aspects of research processes and offers suggestions for the use of LLMs in interview protocol development and refinement. The findings encourage a broader discussion on the evolving role of technology in academic research and present an exciting avenue for future studies in hybrid human-AI engagements in scholarly pursuits.

Keywords

large language model, ChatGPT, interview protocol, qualitative methods

Creative Commons License

Acknowledgements

Flexibility & Iteration: Exploring the Potential of Large Language Models in Developing and Refining Interview Protocols

Jessica L. Parker[1], Veronica Richard[2], and Kimberly Becker[3]
[1]School of Healthcare Business, Massachusetts College of Pharmacy and Health Sciences, Boston, Massachusetts, USA
[2]Dissertation by Design, Raleigh, North Carolina, USA
[3]Academic Insight Lab, Raleigh, North Carolina, USA

This article investigates the potential of Large Language Model (LLM) tools like ChatGPT in aiding researchers in the development and refinement of interview protocols. We found that ChatGPT could generate appropriate interview questions, craft key questions, provide feedback on protocols, and simulate interviews, indicating its potential to reduce time and effort, particularly when human resources are limited. This article builds upon previous authors' insights and suggestions regarding developing and refining interview protocols to maximize the chances of achieving research aims, especially for novice researchers. Additionally, the researchers highlight the flexibility of these tools in adapting to different research contexts and cultural considerations. Ethical considerations and human oversight are emphasized as critical components in the responsible implementation of these tools. The research also paves the way for further exploration into the integration of LLMs into other aspects of research processes and offers suggestions for the use of LLMs in interview protocol development and refinement. The findings encourage a broader discussion on the evolving role of technology in academic research and present an exciting avenue for future studies in hybrid human-AI engagements in scholarly pursuits.

Keywords: large language model, ChatGPT, interview protocol, qualitative methods

Introduction

Qualitative researchers are valued for their interests in people's experiences and meaning making of their experiences. In these interests, qualitative researchers develop the abilities to understand subtle details and recognize deeper meanings or patterns that might not be immediately apparent. These processes can include interpreting emotions, motivations, beliefs, and the cultural contexts in which phenomena occur. This human-centric approach has defined the field of qualitative research for generations (Denzin & Lincoln, 1994). However, the recent release of large language models (LLMs) like ChatGPT (OpenAI, 2023) presents novel opportunities for qualitative researchers due to the ability of LLMs to understand human input and generate human-like text that is contextually appropriate.

One promising area in academia for applying the robust power of LLMs is in qualitative data collection tools. We see specific opportunities for developing and refining interview protocols with LLMs to provide an innovative solution to age-old challenges in qualitative research – the extensive time commitment involved in protocol development and refinement and resource drain involved in developing robust interview protocols (McGrath et al., 2019;

Rabionet, 2011; Rubin & Rubin, 2012). As Brinkmann (2018) contended, efficiency in developing and refining interview protocol questions and prompts refers to implementing the "best and least wasteful route" in meeting protocol requirements (p. 589). In these two areas, ChatGPT emerges as a tool for leveraging technology to reduce time- and human-related resources.

The interview piloting stages can be particularly taxing in scenarios where participant availability is limited, leading to a risk of squandering the interview population (Castillo-Montoya, 2016; Jacob & Furgerson, 2012). Traditionally, the term 'squandering' refers to the risk of exhausting or overusing a pool of potential interview participants during the piloting and refinement process, reducing the available population for the actual study (Jacob & Furgerson, 2012). By employing LLMs like ChatGPT for the preliminary stages, researchers can systematically refine their protocols without overusing their participant pool, thus ensuring the availability of participants for the crucial data collection stage. Overall, the AI approach has potential to enhance the efficiency and effectiveness of the interview process in qualitative research in a way that is accessible for all researchers.

ChatGPT's adeptness at understanding and generating context-appropriate responses suggests a strong potential role in helping researchers develop and refine interview protocols. Therefore, the objective of this paper is to explore how ChatGPT can aid researchers in the interview protocol development and refinement process.
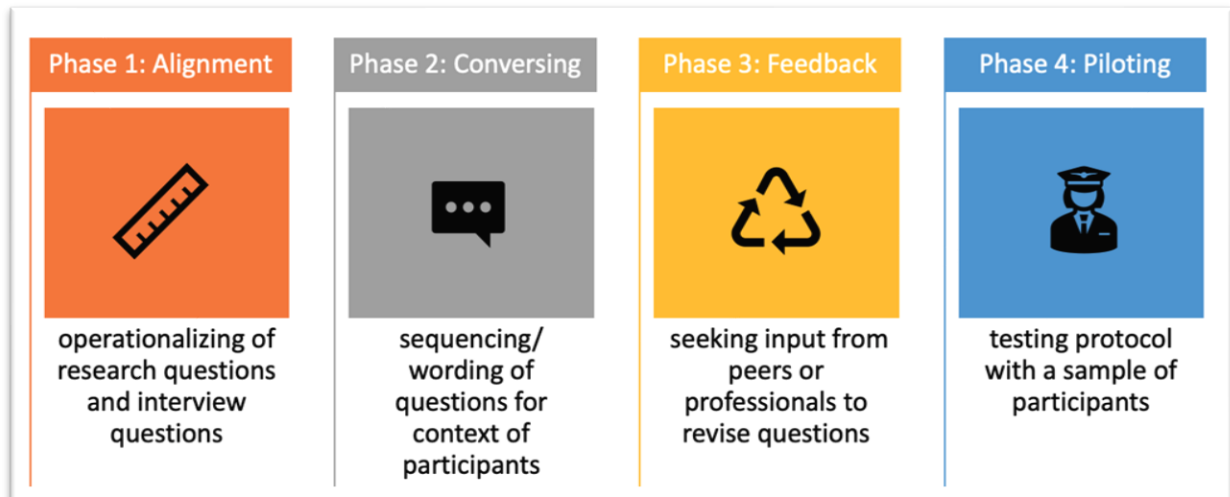
Background

Interviews have become a powerful and justified method in qualitative research due to researchers' ability to capture the stories, complexities, and nuances of the human experience through this method (Brinkmann, 2018). However, developing interview protocols and conducting interviews is an art and a science, requiring intensive human involvement, expertise, and time (Brinkmann & Kvale, 2015; Creswell & Poth, 2018; Roulston, 2012; Rubin & Rubin, 2012). Researchers have argued that novice qualitative researchers often need support and structured processes in developing and refining interview protocols (Jacob & Furgerson, 2012; Roberts, 2020; Turner III, 2010). For novice researchers, the nuances of phrasing, alignment, ethical considerations, and crafting questions that elicit meaningful, in-depth information can present a steep learning curve (Roberts, 2020). Researchers new to qualitative research also require support in developing reflective and reflexive skills needed regarding data collection, such as developing and refining interview protocols (Chenail, 1997; Roulston, 2012). Castillo-Montoya's (2016) interview protocol refinement (IPR) framework offers a structured process for strengthening interview protocols and improving the quality of data derived from interviews. The IPR framework includes four phases, as shown in Figure 1.

The first phase involves aligning interview protocol items and research questions, where the goal is to remove unnecessary interview questions/prompts that do not serve the purpose of the study. During the second phase, researchers develop what Castillo-Montoya (2016) refers to as an "inquiry-based conversation," in which researchers rely on their understanding of the usual behaviors, societal standards, and routines of potential participants to word questions/prompts while considering the sequencing of questions/prompts to align with the chosen methodological approach. In this stage, it is also necessary for the researcher to consider the sequencing of interview questions/prompts to begin the interview, to address the key questions/prompts, and necessary prompting to transition during the interview and concluding the interview (Rubin & Rubin, 2012). In the third phase, researchers seek feedback on the protocol (Castillo-Montoya, 2016). This feedback can include language and wording issues that might limit participant understanding of the question/prompts (e.g., jargon, academic language) and identify areas that are vague or confusing. The fourth and final phase

consists of piloting the interview to determine if each question/prompt elicits the information desired (Castillo-Montoya, 2016). In this phase, the goal is to simulate as much as possible the interview process including via building rapport, explaining processes and consent, and trying out the logistics of space, recording, and timing (Creswell & Poth, 2018) in order to make necessary revisions prior to entering the field.

Figure 1
Visual Adaption of the IPR Framework

Phase 1: Alignment	Phase 2: Conversing	Phase 3: Feedback	Phase 4: Piloting
operationalizing of research questions and interview questions	sequencing/ wording of questions for context of participants	seeking input from peers or professionals to revise questions	testing protocol with a sample of participants

Throughout each phase, researchers continually modify the protocol. This iterative refinement process is often the most time-consuming aspect of preparing to conduct interviews (Rabionett, 2011). As such, inherent challenges are associated in working through each phase. For example, researchers are commonly faced with limited time and resources. One especially detrimental resource shortage that exists is a small participant pool, which may limit the final phase of piloting the interview protocol (Castillo-Montoya, 2016) and ultimately limit the researchers' ability to elicit responses to accomplish their research aims.

In this article, we advocate that such limitations could be mitigated using LLMs like ChatGPT, which present new avenues for developing and refining interview protocols. Prior research has shown that LLMs can mimic human conversation (y Arcas, 2022) and the most recent version of ChatGPT – ChatGPT 4 – is capable of advanced reasoning (Sejnowski, 2023). Until now, conventional resources for training qualitative researchers (i.e., textbooks and other paper-based sources) could not offer interactive, real-time assistance. However, with the advent of LLMs, researchers now have a viable tool for prototyping not only the development and refinement of interview protocols but also piloting the protocol to enhance trustworthiness and quality of the actual interview process. To date, using LLMs for this purpose in qualitative research, particularly in developing and refining interview protocols through to piloting processes, remains unexplored.

Interview Protocol Development and Refinement

Interviews are integral to qualitative research, serving as an invaluable tool for obtaining rich, in-depth data and facilitating a profound understanding of individuals' experiences, perspectives, and subjective interpretations (Brinkmann, 2018; McGrath et al., 2019; Patton, 2015). Unlike more rigid methods like structured surveys, qualitative interviews allow for an interactive and flexible exploration of unique experiences, as well as insight into various phenomena (Brinkmann, 2018; DiCicco-Bloom & Crabtree, 2006; McGrath et al.,

2019; Rubin & Rubin, 2012; Smith et al., 2022). Researchers use interviews to amplify marginalized voices and to explore complex and sensitive topics, offering a nuanced and detailed view of their experiences (Brinkmann, 2018; McGrath et al., 2019). Additionally, through the open-ended nature of interviews, qualitative researchers aim to understand people's experiences and meaning making to gain deeper insight into phenomena (Creswell & Poth, 2018; Denzin & Lincoln, 1994; Patton, 2015). Overall, qualitative researchers use interviews to enhance their understanding of subjective experiences and to make inclusive and comprehensive contributions to their fields.

In qualitative research, developing and refining interview protocols involves complexities, particularly for novice researchers, who may make common mistakes such as asking leading questions, using technical jargon, and framing questions too narrowly (Roberts (2020). Researchers have identified other challenges such as aligning interview questions with research aims (Brinkmann & Kvale, 2015; Ojeda et al., 2011; Roulston, 2012), crafting suitable open-ended questions to elicit rich responses (Cyr & Anderson, 2013; Rubin & Rubin, 2012), and incorporating cultural sensitivity to ensure relevance to the target population (McGrath et al., 2019; Ojeda et al., 2011). Researchers also advocate for piloting the interview protocol to identify language-based challenges, question/prompt understanding, and to verify the effectiveness and plausible aspects of the protocol (Castillo-Montoya, 2016; Ojeda et al., 2011). Overall, these elements enhance the quality and pertinence of the collected data.

ChatGPT as a Potential Solution

ChatGPT is a deep learning model designed to understand and generate natural language text. Chatbots like ChatGPT have recently garnered significant attention within the academic community due to their wide range of applications, from enhancing academic writing (Dergaa et al., 2023, Parker et al., 2023) to creating personalized learning tools (Khan et al., 2023). In the face of such attention, there is a need for additional research on best practices for using LLMs to plan and conduct research, as well as research to explore the outcomes related to the use of LLMs in qualitative research (Christou, 2023). Researchers have specifically noted the potential of generative AI in various aspects of research, such as analyzing text (Rahman et al., 2023) and for annotation and coding (Lennon et al., 2021; Xiao et al., 2023). This recognition points to the opportunity for increased utilization of qualitative research, which has sometimes been underused due to the time and cost associated with annotating (coding) qualitative data.

Castillo-Montoya's (2016) Framework

A seminal article about qualitative research interviews is Castillo-Montoya's (2016) IPR framework, a four-phase systematic process to develop and fine-tune interview protocols for qualitative research. This framework emphasizes aligning interview questions with research objectives and focuses on creating an inquiry-based conversation, receiving feedback on protocols, and pilot testing. By using the framework, researchers can enhance the reliability and quality of data obtained from research interviews and can be applied to structured, semi-structured, or non-structured interviews. In ensuring alignment, sensitivity, and context-appropriateness of questions/prompts, and by carefully constructing the inquiry process, the IPR framework helps researchers obtain robust and detailed data that is congruent with the research's aims. The framework also offers a shared language to illustrate rigorous steps taken in developing interview protocols and can be a valuable tool for researchers seeking to capture participants' experiences and meaning making.

Method

To explore ChatGPT's suitability to aid researchers in developing and refining interview protocols, we used ChatGPT to progress through each phase of Castillo-Montoya's (2016) IPR framework. In each phase, we prompted ChatGPT (Table 1) and reviewed its output before advancing to the next phase.

We began by prompting ChatGPT to generate a list of interview questions for a developing study. A prompt consists of "a set of instructions provided to an LLM that programs the LLM by customizing it and/or enhancing or refining its capabilities" (White et al., 2023, p. 1). For this study, we used the "persona" pattern, which is useful in instances when users desire the LLM output to consistently adopt a specific viewpoint or role (White et al., 2023). We provided details regarding the future study's purpose, research questions, and theoretical framework. In Phase 1, we prompted ChatGPT to create a matrix and map the interview questions onto the research questions. In Phase 2, we prompted ChatGPT to construct an inquiry-based conversation which follows social rules of ordinary conversation by including a variety of questions (i.e., introductory, transitional, key, and closing; Castillo-Montoya, 2016). The conversation is scripted with likely follow-up and prompts (Rubin & Rubin, 2012). In Phase 3, we prompted ChatGPT to play the role of a research assistant and provide feedback on aspects of the interview protocol using Castillo-Montoya's (2016) "Activity Checklist for Close Reading of an Interview Protocol" (p. 825). This role-play process addresses Castillo-Montoya's (2016) recommendation for receiving feedback on protocols in circumstances where limited participants are available for piloting. In Phase 4, we prompted ChatGPT to simulate an interview by providing expected responses.

Table 1
Input Prompt Sequence and Evaluation Criteria

IPR Framework Phase (Castillo-Montoya, 2016)	Input Prompt(s)
Phase 1: Ensuring interview questions align with research questions	Initial Prompt: I am a professor conducting a study on student acceptance of technological tools for automated writing evaluation (AWE). The primary question driving the original study is: How do students describe their acceptance of technological tools for automated writing evaluation? Sub-questions include: (1) How do students use technology for scholarly writing development? (2) How do students perceive the individual benefits of technology for scholarly writing development? (3) How do students perceive the potential drawbacks or challenges of technology for scholarly writing development? I am using a modified version of the Technology Acceptance Model which includes these constructs: perceived usefulness, perceived ease of use, behavioral intent to use the technology, actual use, AI anxiety, relative advantage, subjective norm, and trust. Help me develop a list of interview questions. Follow-up prompt: Create a matrix that shows how the previously developed interview questions map onto the research questions.

Phase 2: Constructing an inquiry-based conversation	Create an interview protocol that follows social rules of ordinary conversation, contains a variety of questions (introductory, key, transition, and closing), and a script with likely follow-up and prompt questions.
Phase 3: Receiving feedback on interview protocols	Pretend you are a research assistant providing feedback on the interview protocol. Conduct a close read of the protocol and comment on each of the following criteria: (1) Beginning questions are factual in nature, (2) Key questions are majority of the questions and are placed between beginning and ending questions, (3) Questions at the end of interview protocol are reflective and provide participant an opportunity to share closing comments , (4) A brief script throughout interview protocol provides smooth transitions between topic areas , (5) Interviewer closes with expressed gratitude and any intents to stay connected or follow up, (6) Overall, interview is organized to promote conversational flow, (7) Questions/statements are free from spelling error(s), (8) Only one question is asked at a time Most questions ask participants to describe experiences and feelings, (9) Questions are mostly open ended , (10) Questions are written in a non-judgmental manner, (11) All questions are needed, (12) Questions/statements are concise, (13) Questions/statements are devoid of academic language, and (14) Questions/statements are easy to understand.
Phase 4: Piloting the interview protocol	Imagine you are a doctoral student who recently used a technological tool for automated writing evaluation. Answer the interview questions in as much detail as possible so I can pilot the interview protocol.

Results

ChatGPT's output in Phase 1 included a list of interview questions and a matrix mapping the interview questions onto the research questions. Most of the questions and prompts included in the output were what Rubin and Rubin (2012) call "main questions" or questions that are designed to answer the research questions. We then asked ChatGPT to create a matrix for mapping the interview questions onto research questions as shown in Figure 1.

Figure 1
ChatGPT's Output for Creating Interview Questions and Mapping Them onto Research Questions

Certainly! Here's a set of interview questions aligned with your research focus:

1. Perceived Usefulness:
* How do you see the usefulness of automated writing evaluation tools in your writing process?
* Can you describe an instance where AWE tools have helped or hindered your scholarly writing?

2. Perceived Ease of Use:
* How would you describe your experience in learning to use AWE tools?
* What challenges, if any, did you face when first using these tools?

3. Behavioral Intent to Use the Technology:
* What influences your decision to use or not use AWE tools in your writing?
* How likely are you to continue using these tools in the future?

4. Actual Use:
* How often do you use AWE tools in your writing?
* In what ways have these tools been integrated into your writing routine?

5. AI Anxiety:
* Have you ever felt apprehensive about using AI-based writing evaluation tools? If so, why?
* What can be done to alleviate any anxiety or concerns you have about using AWE tools?

6. Relative Advantage:
* How do AWE tools compare to traditional methods of writing evaluation?
* Can you identify specific advantages or disadvantages?

7. Subjective Norm:
* How do peers or mentors influence your use of AWE tools?
* Is there social pressure or encouragement within your academic community regarding the use of these tools?

8. Trust:
* How much trust do you place in the accuracy and reliability of AWE tools?
* Have you ever encountered issues that made you question the trustworthiness of these tools?
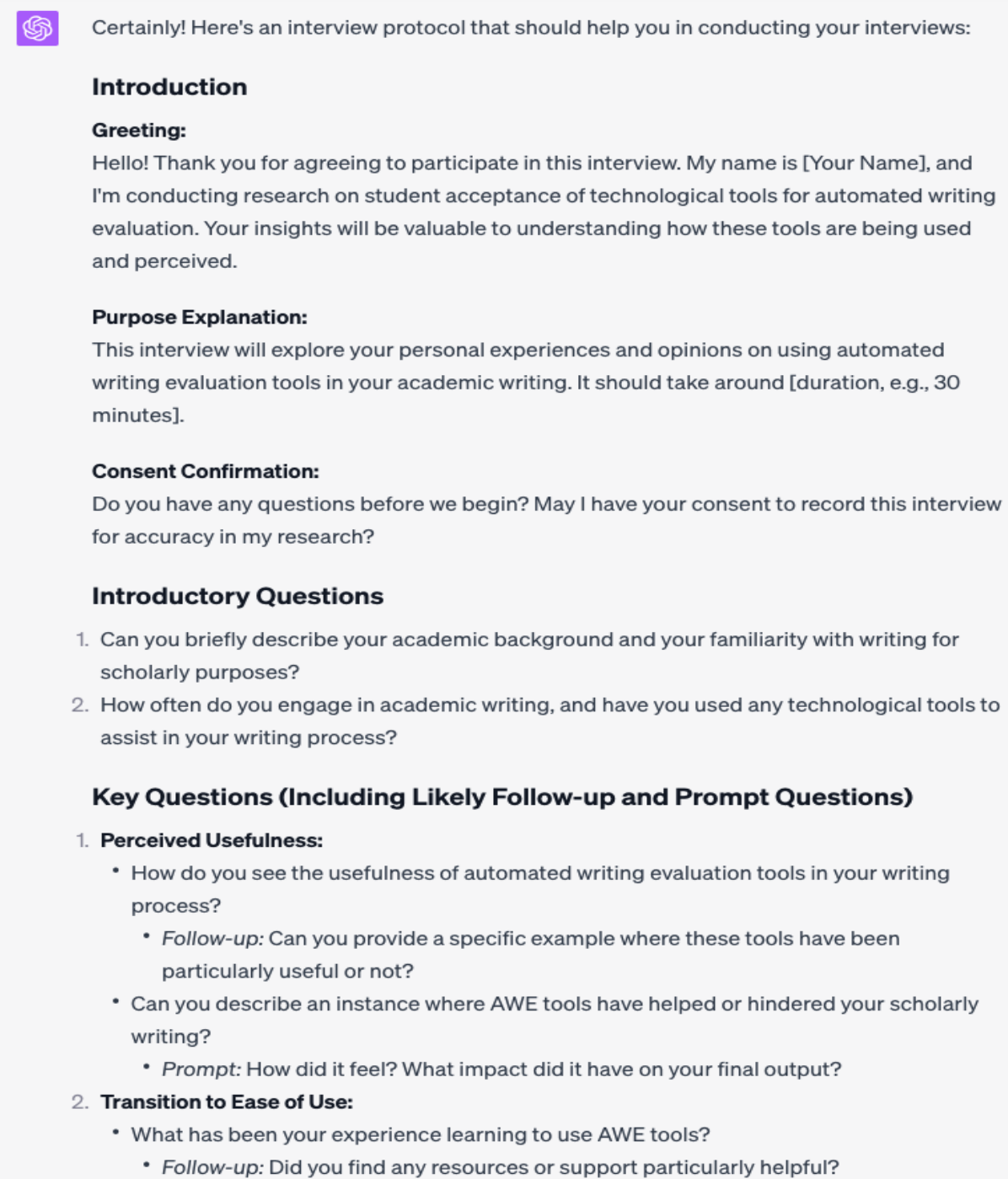
Matrix Mapping:

Interview Questions	Research Questions
1. Perceived Usefulness	Main question, Sub-question (2)
2. Perceived Ease of Use	Sub-question (3)
3. Behavioral Intent to Use the Tech	Sub-question (1), (3)
4. Actual Use	Sub-question (1)
5. AI Anxiety	Sub-question (3)
6. Relative Advantage	Sub-question (2), (3)
7. Subjective Norm	Main question, Sub-question (1)
8. Trust	Sub-question (3)

In Phase 2, we asked ChatGPT to construct an inquiry-based conversation. An inquiry-based conversation follows social rules of ordinary conversation, includes a variety of questions (i.e., introductory, transitional, key, and closing; Castillo-Montoya, 2016), and is scripted with likely follow-up and prompts (Rubin & Rubin, 2012). ChatGPT's output is displayed in Figures 2 and 3.

Figure 2
The First Half of ChatGPT's Output When Prompted to Construct an Inquiry-Based Conversation

Certainly! Here's an interview protocol that should help you in conducting your interviews:

Introduction

Greeting:
Hello! Thank you for agreeing to participate in this interview. My name is [Your Name], and I'm conducting research on student acceptance of technological tools for automated writing evaluation. Your insights will be valuable to understanding how these tools are being used and perceived.

Purpose Explanation:
This interview will explore your personal experiences and opinions on using automated writing evaluation tools in your academic writing. It should take around [duration, e.g., 30 minutes].

Consent Confirmation:
Do you have any questions before we begin? May I have your consent to record this interview for accuracy in my research?

Introductory Questions
1. Can you briefly describe your academic background and your familiarity with writing for scholarly purposes?
2. How often do you engage in academic writing, and have you used any technological tools to assist in your writing process?

Key Questions (Including Likely Follow-up and Prompt Questions)
1. **Perceived Usefulness:**
   * How do you see the usefulness of automated writing evaluation tools in your writing process?
     * *Follow-up:* Can you provide a specific example where these tools have been particularly useful or not?
   * Can you describe an instance where AWE tools have helped or hindered your scholarly writing?
     * *Prompt:* How did it feel? What impact did it have on your final output?
2. **Transition to Ease of Use:**
   * What has been your experience learning to use AWE tools?
     * *Follow-up:* Did you find any resources or support particularly helpful?

Figure 3
The Second Half of ChatGPT's Output When Prompted to Construct an Inquiry-Based Conversation (Continued from Figure 2)

3. **AI Anxiety & Trust:**
   - Have you ever felt apprehensive about using AI-based writing evaluation tools? If so, why?
     - *Prompt:* What could be done to alleviate your concerns?
   - How much trust do you place in the accuracy and reliability of these tools?
     - *Follow-up:* Have you ever encountered issues that made you question the trustworthiness of these tools?

4. **Transition to Behavioral Intent and Actual Use:**
   - How often do you use AWE tools, and how have they been integrated into your writing routine?
     - *Prompt:* What influences your decision to use or not use these tools?

Closing Questions

1. Looking forward, how do you see yourself using or not using these tools in your future academic writing?
2. Do you have any suggestions for how these tools could be improved to better serve students like yourself?
3. Is there anything else you'd like to share on this topic that we haven't covered?
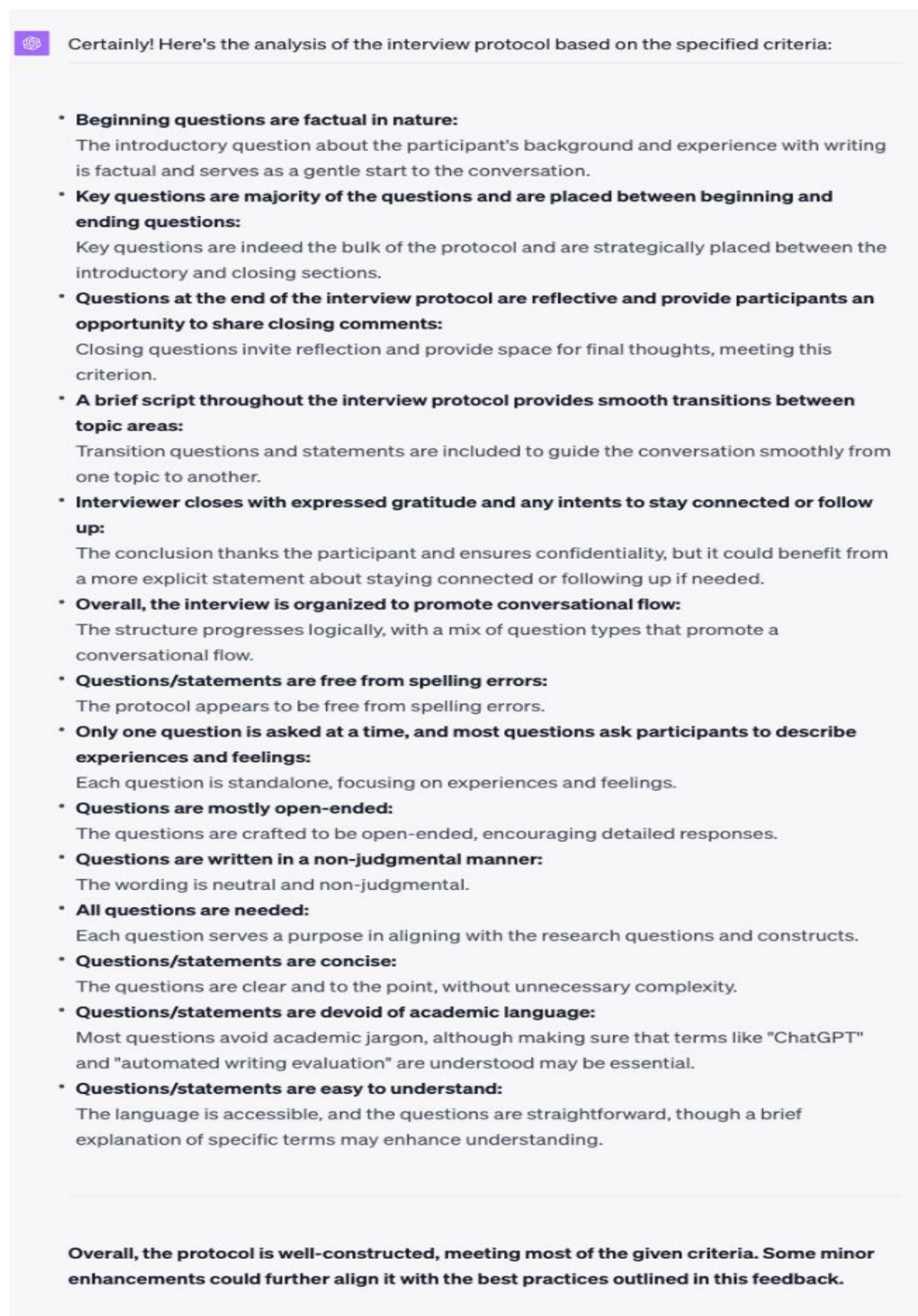
Thank You and Closure:

Thank you so much for your time and insights today. Your perspectives are incredibly valuable to this research. If you have any further questions or thoughts, please feel free to reach out to me at [contact information].

This protocol covers a variety of question types and includes follow-up and prompt questions that allow for a natural and engaging conversation. It maintains a respectful and inquisitive tone, ensuring that the interviewee feels comfortable and valued throughout the process.

In Phase 3, we prompted ChatGPT to role-play as a research assistant and provide feedback on aspects of the interview protocol using Castillo-Montoya's (2016) "Activity Checklist for Close Reading of an Interview Protocol" (p. 825), displayed in Figure 4.

Figure 4

ChatGPT's Output When Prompted to Pilot the Interview Protocol

Certainly! Here's the analysis of the interview protocol based on the specified criteria:

- **Beginning questions are factual in nature:**
  The introductory question about the participant's background and experience with writing is factual and serves as a gentle start to the conversation.
- **Key questions are majority of the questions and are placed between beginning and ending questions:**
  Key questions are indeed the bulk of the protocol and are strategically placed between the introductory and closing sections.
- **Questions at the end of the interview protocol are reflective and provide participants an opportunity to share closing comments:**
  Closing questions invite reflection and provide space for final thoughts, meeting this criterion.
- **A brief script throughout the interview protocol provides smooth transitions between topic areas:**
  Transition questions and statements are included to guide the conversation smoothly from one topic to another.
- **Interviewer closes with expressed gratitude and any intents to stay connected or follow up:**
  The conclusion thanks the participant and ensures confidentiality, but it could benefit from a more explicit statement about staying connected or following up if needed.
- **Overall, the interview is organized to promote conversational flow:**
  The structure progresses logically, with a mix of question types that promote a conversational flow.
- **Questions/statements are free from spelling errors:**
  The protocol appears to be free from spelling errors.
- **Only one question is asked at a time, and most questions ask participants to describe experiences and feelings:**
  Each question is standalone, focusing on experiences and feelings.
- **Questions are mostly open-ended:**
  The questions are crafted to be open-ended, encouraging detailed responses.
- **Questions are written in a non-judgmental manner:**
  The wording is neutral and non-judgmental.
- **All questions are needed:**
  Each question serves a purpose in aligning with the research questions and constructs.
- **Questions/statements are concise:**
  The questions are clear and to the point, without unnecessary complexity.
- **Questions/statements are devoid of academic language:**
  Most questions avoid academic jargon, although making sure that terms like "ChatGPT" and "automated writing evaluation" are understood may be essential.
- **Questions/statements are easy to understand:**
  The language is accessible, and the questions are straightforward, though a brief explanation of specific terms may enhance understanding.

Overall, the protocol is well-constructed, meeting most of the given criteria. Some minor enhancements could further align it with the best practices outlined in this feedback.

In Phase 4, we evaluated ChatGPT's ability to simulate an interview by providing responses expected of the study sample. ChatGPT's output is displayed in Figures 5 and 6.

Figure 5
ChatGPT's Output of a Simulated Interview

Certainly! Here are the answers to the interview questions, imagining myself as a doctoral student who has recently used an automated writing evaluation (AWE) tool.

Introductory Questions

1. **Academic Background & Familiarity with Writing:**
   - I'm currently a Ph.D. candidate in Computer Science, focusing on machine learning. I frequently engage in academic writing for journals, conferences, and my dissertation.
2. **Use of Technological Tools:**
   - I have used various tools for academic writing, including citation management software, grammar checkers, and an AWE tool to assess the structure and coherence of my work.

Key Questions

1. **Perceived Usefulness:**
   - I find AWE tools highly useful. They provide instant feedback, which helps in revising and enhancing the quality of my writing. For instance, the tool I used suggested improvements in my argumentation in a recent paper, making it more persuasive.
2. **Ease of Use:**
   - Learning to use the AWE tool was straightforward. Initially, I had to experiment to understand its features, but there were online tutorials that helped guide me.
3. **AI Anxiety & Trust:**
   - At first, I was a bit skeptical about relying on AI for evaluating complex writing, but my apprehensions were reduced as I saw the tool's accuracy in analyzing various elements. I trust the tool for basic evaluations but still prefer human feedback for more nuanced assessments.
4. **Behavioral Intent and Actual Use:**
   - I use AWE tools for initial drafts and revisions, especially when I need quick feedback. They've become an integral part of my writing routine, and I intend to continue using them.

Figure 6
A Continuation of the Simulated Interview

Closing Questions

1. **Future Use:**
   - I definitely see myself continuing to use these tools, particularly as they evolve and provide even more insightful feedback. They save time and give me an additional layer of confidence in my writing.

2. **Suggestions for Improvement:**
   - I'd love to see more context-sensitive feedback and perhaps integration with other writing tools, like reference managers. A more personalized experience could enhance the utility of AWE tools.

3. **Additional Comments:**
   - Overall, I think AWE tools are an exciting development in the academic writing landscape. They're not a replacement for human evaluation but a complement that allows for more iterative refinement of writing.

I hope this detailed response helps in piloting your interview protocol! Feel free to ask for any clarifications or further details.

Discussion

Using ChatGPT to progress through each phase of the IPR framework (Castillo-Montoya, 2016) demonstrates its potential utility for developing and refining interview protocols. Because ChatGPT is a sophisticated natural language processing tool and has been shown capable of engaging in conversation and handling multi-turn dialogue as an interviewee (see Eysenbach, 2023; Lund & Wang, 2023; Mann, 2023), its capabilities are not surprising. However, we identified several limitations in each phase of refining the interview protocol, warranting the importance of human oversight.

Our exploration in the first phase, which focused on ensuring interview questions align with research questions, revealed several notable strengths and weaknesses in ChatGPT's capabilities. A strength was ChatGPT's ability to instantly produce a matrix aligning the research questions with the interview questions. This strategy could help researchers readily identify gaps and ensure interview questions are purposefully constructed to operationalize the main areas of inquiry. The automatic creation of a matrix is a time-saving technique for quickly validating the overall methodological alignment. Despite this capability, researchers must be aware that chatbots frequently need follow-up prompting to produce optimal outputs. In this case, on the first prompting, ChatGPT provided two suboptimal outputs: (1) no questions aligned with the primary research question and (2) the inclusion of yes/no questions. Much like working with a human partner, chatbots require iterative meaning negotiation. These iterations can serve as valuable learning opportunities, as they require researchers to consider the output carefully and critically and then to renegotiate if necessary.

In the second phase, we prompted ChatGPT to construct an inquiry-based conversation. An inquiry-based conversation is characterized by a variety of questions with follow-up questions and prompts that follow the social rules of ordinary conversation (Castillo-Montoya, 2016). We found that ChatGPT demonstrated a largely positive role, particularly in creating key and follow-up questions. Nevertheless, ChatGPT failed to produce satisfactory introductory, transitional, and closing questions. For instance, the "ice-breaker" question in the introductory section was irrelevant to the research purpose or study questions, and only one transitional phrase (i.e., "Before we delve into this topic...") was included. Despite these shortcomings, we believe its ability to generate key questions adds value because key questions solicit the most important information from the interviewee (Rubin & Rubin, 2012). Furthermore, in aligning with qualitative research aims, follow-up questions are often derived during the interview based on the interviewee's stories and explanations (Creswell & Poth, 2018; Roulston, 2012; Smith et al., 2022)

The third phase consisted of receiving feedback on the interview protocol and highlighting the novelty of ChatGPT's role-playing ability as a research assistant. To conduct this phase, we prompted ChatGPT with Castillo-Montoya's (2016) "Activity Checklist for Close Reading of an Interview Protocol." We were particularly interested in determining ChatGPT's capabilities in this task because there is often a scarcity of available participants or others outside of the sample but who are familiar with the research topic and participant population to provide relevant input and suggestions. Although ChatGPT was able to provide feedback, its feedback was not critical enough to be useful for refinement. In response to each question, ChatGPT simply stated that the criteria were met and did not offer suggestions for improvement. In a follow-up prompt, we asked ChatGPT to make some suggestions, and it provided some mediocre feedback for improving the protocol, as shown in Figure 7. In this round of suggestions, ChatGPT acknowledged the need to revise the introductory questions, breakdown complex questions, check for conciseness, and simplify or define terms. ChatGPT also suggested adding reflection questions to the conclusion and conducting a pilot study. However, its ability to provide specific suggestions for rephrasing was absent. The final phase of piloting the interview protocol highlighted ChatGPT's ability to suggest responses expected of a human. Utilizing LLM models like ChatGPT for piloting qualitative research protocols signifies a substantial advancement in research methodology. It addresses a critical concern in qualitative research – the dilemma of wasting valuable human resources while testing and refining interview questions. This potential of ChatGPT has particular relevance in the training of novice researchers, who must learn through practice without exhausting limited resources. In many research contexts, the target population might be limited, unique, or hard to reach. Wasting these precious samples on piloting questions may result in lost opportunities to gather substantive data. Moreover, novice researchers need extensive practice to hone their interviewing skills and develop questions that are clear, unbiased, and effective. This practice phase might traditionally require actual participants, but LLM models like ChatGPT can simulate human responses, providing a more sustainable training ground prior to in-person piloting. Finally, AI-driven piloting allows researchers to iterate and modify their questions rapidly, without the need for scheduling and coordinating with human participants. These benefits maximize a more efficient refinement process while simultaneously mitigating the ethical dilemma of engaging human participants in trials that might not directly contribute to the research outcomes. Adding this process to methodological plans also serves to preserve the integrity of the participant's contribution to the actual study.

Figure 7
Output on Improving the Protocol and Questions

Certainly! Here are the suggestions to improve the interview questions based on the analysis:

1. **Beginning Questions:**
   * Consider including clear introductory questions about the participant's specific experiences with AWE tools.
2. **Key Questions:**
   * Break down complex questions with follow-ups into separate queries to maintain focus.
   * Include more specific questions related to each construct in your model, such as examples of perceived usefulness or specific anxieties related to AI.
3. **Closing Questions:**
   * Provide a clear opportunity for the participant to add anything else they feel is relevant.
4. **Transitions Between Topic Areas:**
   * Add brief explanatory statements to facilitate smoother transitions between topics.
5. **Conciseness:**
   * Review key questions to ensure they are as concise as possible without losing essential details.
6. **Academic Language:**
   * Simplify or define terms like "AWE tools" during the interview to ensure understanding.
7. **Open-Ended Questions:**
   * Ensure that all key questions are open-ended to elicit comprehensive responses.
8. **Overall Flow:**
   * The overall flow seems well-structured. However, a pilot interview could reveal areas for further improvement.
9. **Inclusion of Reflective Questions:**
   * Consider adding more reflective questions, especially towards the end, that encourage the participant to consider their overall experience with AWE tools.

These suggestions aim to enhance the clarity, conciseness, and comprehensiveness of the interview questions, aligning them more closely with the research objectives.

Effectiveness of LLMs in Simulating Human Responses

The use of LLMs in piloting research protocols opens new possibilities but also raises questions about how accurately models like ChatGPT can simulate human responses. While the LLM chatbot can generate responses based on vast amounts of data, the authenticity and nuance of human understandings, emotions, and contexts may not be fully captured. Research into the effectiveness of LLMs in this role is a vital ongoing consideration; however, we have

developed a set of preliminary guidelines (Table 2) for researchers to consider as we move toward a hybrid human-AI research world.

Accessibility and Cost

Implementing AI-driven piloting can be more cost-effective in the long run but might require initial investments in technology and training. Ensuring that these tools are accessible to all researchers, regardless of institutional backing, is a crucial consideration.

Ethical Use of AI

While LLM models alleviate some ethical concerns related to human participants, they introduce new ones, such as data privacy and the responsible use of AI. As the use of generative AI is such a new endeavor in academia, guidelines and standards are in a nascent stage of development and must be further refined to ensure ethical implementation.

Integration with Traditional Methods

The use of LLMs does not replace the need for human intuition and expertise in question development. It should be integrated with traditional methods, such as expert review, to ensure that the questions are both scientifically robust and ethically sound. This is a developing transdisciplinary research area (Eaton, 2021) that explores the idea of a post-plagiarism world. Eaton argues that hybrid human-AI writing will become so prevalent that discerning the boundary between human contribution and artificial intelligence involvement will be a fruitless endeavor.

Conclusion

Incorporating ChatGPT for piloting and refining qualitative research protocols marks a promising advancement, especially in training novice researchers. By mitigating the waste of human resources such as one's participant sample and offering a flexible and ethical platform for piloting, LLMs can revolutionize how researchers prepare for interviews. Careful consideration of effectiveness, accessibility, ethical use, and integration with traditional methods will be key to maximizing this potential.

The development of this study was influenced by our desire to integrate ChatGPT as a resource into a research methods course for doctoral students. Many courses necessitate students to develop and pilot an interview protocol, which is known to be challenging for novice researchers (Roberts, 2020; Roulston, 2012) and time-consuming for even the most seasoned researchers (Brinkmann, 2018; Pope et al., 2000). Unlike traditional static resources (i.e., textbooks and articles), LLMs can engage in an interactive dialogue with the user, assisting in the ongoing refinement of interview questions. Further, LLMs are always available, accessible to all, and can provide real-time assistance, making them an ideal resource for students and researchers who work at various hours and balance multiple commitments.

ChatGPT, or other LLMs, are not the ultimate answer. As demonstrated, ChatGPT often repeated its answers and responded sub-optimally by including close-ended responses and using double-barreled questions. This highlights the need for researchers to accept that making ongoing adjustments in understanding and communication with LLMs is par for the course of hybrid human-AI interactions, just as it is with human-to-human interactions. As humans acclimate to engaging with LLMs for daily tasks, there is and will continue to be debates about what is appropriate. Already we noticed that although some scholars embrace the use LLMs

for tasks like data analysis and writing (Hosseini & Horbach, 2023), others raised concerns about their ethical implications, biases, and potential loss of human intuition and insight (De Angelis et al., 2023).

Considering the early and varied nature of the guidance to date in the literature, we offer a list of considerations for researchers interested in hybridizing their operationalization of interview protocol design (Table 2).

Table 2

Considerations for Researchers Hybridizing Their Operationalization of Interview Protocol Design

Consideration	Explanation	Example
Cultural Sensitivity	ChatGPT's exposure to diverse texts helps test questions across cultural contexts, ensuring respect and consideration.	Testing a survey in multiple countries
Quality of Output	Evaluation of ChatGPT's quality output, assessing reliability and validity; comparison with traditional methods.	Comparing AI-generated and human-made protocols
LLMs as a Tool, Not Replacement	LLMs complement researchers, not replace them; human understanding of context remains vital.	Using AI to draft but human to finalize a paper
Ethical Considerations	The use of LLM tools like ChatGPT raises new ethical issues like data privacy and bias.	Ensuring no personal data is stored by the AI
LLM Limitations	Acknowledging AI limitations, like the inability to make intuitive or creative leaps like a human.	Using human oversight for creative writing
Further Research Directions	Further research needed to identify best practices using LLMs in research planning and conduct.	Exploring how AI can be used in qualitative data analysis

Our study demonstrates that Language Model tools (LLMs) such as ChatGPT can significantly enhance the development and refinement of interview protocols by performing various essential tasks including generating contextually appropriate questions and simulating real-life interviews. The multifaceted capabilities of LLMs can streamline the process, reduce the required time and effort, and adapt to various research contexts and cultural considerations. However, it is imperative to balance this innovation with ethical considerations, quality control, and human oversight to ensure that the use of these tools complements human expertise without supplanting it, maintaining a balance between innovation and integrity. Finally, our study opens new doors for future research, especially in exploring how LLMs can be further integrated into other aspects of research processes, such as data collection, analysis, and interpretation. It encourages a broader conversation about the evolving role of technology in academic research and the potential for hybrid human-AI engagements to enrich and enhance the rigor, creativity, and efficiency of scholarly pursuits. The horizon appears promising, yet it invites careful navigation to fully harness the potential of these innovative tools in a responsible and effective manner.

References

Brinkmann, S. (2018). The interview. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 576-599). SAGE.

Brinkmann, S., & Kvale, S. (2015). *Interviews: Learning the craft of qualitative research interviewing* (3rd ed.). SAGE.

Castillo-Montoya, M. (2016). Preparing for interview research: The interview protocol refinement framework. *The Qualitative Report*, *21*(5), 811-831. https://doi.org/10.46743/2160-3715/2016.2337

Chenail, R. J. (1997). Interviewing exercises: Lessons from family therapy. *The Qualitative Report, 3*(2), 1-8. https://doi.org/10.46743/2160-3715/1997.2025

Christou, P. (2023). How to use artificial intelligence (AI) as a resource, methodological and analysis tool in qualitative research? *The Qualitative Report, 28*(7), 1968-1980. https://doi.org/10.46743/ 2160-3715/2023.6406

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry & research design: Choosing among five approaches*, (4th ed.). SAGE.

Cyr, A. A, & Anderson, N. D. (2013). Updating misconceptions: Effects of age and confidence. *Psychonomic Bulletin & Review, 20* 574-580. https://doi.org/10.3758/s13423-012-0357-0

De Angelis, L., Baglivo, F., Arzilli, G., Privitera, G., Ferragina, P., Tozzi, A., & Rizzo, C. (2023). ChatGPT and the rise of large language models: The new AI-driven infodemic threat in public health. *Frontiers in Public Health*, *11,* 1166120. https://doi.org/10.3389/fpubh.2023.1166120

Denzin, N. K., & Lincoln, Y. S. (1994). Introduction: Entering the field of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 1-17). SAGE.

Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, *40*(2), 615-622. https://doi.org/10.5114/biolsport.2023.125623

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. *Medical Education*, *40*(4), 314-321. https://doi.org/10.1111/j.1365-2929.2006.02418.x

Eaton. (2021). *Plagiarism in higher education: Tackling tough topics in academic integrity*. Libraries Unlimited.

Eysenbach, G. (2023). The role of ChatGPT, generative language models, and artificial intelligence in medical education: A conversation with ChatGPT and a call for papers. *JMIR Medical Education*, *9*, e46885. https://doi.org/10.2196/46885

Hosseini, M., & Horbach, S. P. (2023). Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Research Integrity and Peer Review, 8*(1), 4. https://doi.org/10.21203/rs.3.rs-2587766/v1

Khan, R., Jawaid, M., Khan, A., & Sajjad, M. (2023). ChatGPT - reshaping medical education and clinical management. *Pakistan Journal of Medical Sciences*, *39*(2), 605-607. https://doi.org/10.12669/pjms.39.2.7653

Jacob, S. A., & Furgerson, S. P. (2012). Writing interview protocols and conducting interviews: tips for students new to the field of qualitative research. *The Qualitative Report, 17*, 1-10. https://doi.org/10.46743/2160-3715/2012.1718

Lennon, R. P., Fraleigh, R., Van Scoy, L. J., Keshaviah, A., Hu, X. C., Snyder, B. L., Miller, E. L., Calo, W. A., Zgierska, A. E., & Griffin, C. (2021). Developing and testing an automated qualitative assistant (AQUA) to support qualitative analysis. *Family*

Medicine and Community Health, *9*(Suppl 1), e001287. https://doi.org/10.1136/fmch-2021-001287

Lund, B. D., & Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News*, *40*(3), 26-29. https://doi.org/10.1108/lhtn-01-2023-0009

Mann, D. L. (2023). Artificial intelligence discusses the role of artificial intelligence in translational medicine. *JACC: Basic to Translational Science, 8*(2), 221-223. https://doi.org/10.1016/j.jacbts.2023.01.001

McGrath, C., Palmgren, P. J., & Liljedahl, M. (2019). Twelve tips for conducting qualitative research interviews. *Medical Teacher*, *41*(9), 1002-1006. https://doi.org/10.1080/0142159x.2018.1497149

OpenAI. (2023). *ChatGPT* (July 15 version) [Large language model]. https://chat.openai.com/chat

Ojeda, L., Flores, L. Y., Meza, R., & Morales, A. (2011). Culturally competent qualitative research with Latino immigrants. *Hispanic Journal of Behavioral Sciences*, *33*(2), 184-203. https://doi.org/10.1177/0739986311402626

Parker, J., Becker, K., & Carroca, C. (forthcoming). ChatGPT for automated writing evaluation in scholarly writing instruction. *Journal of Nursing Education.*

Patton, M. Q. (2015). *Qualitative research and evaluation methods: Integrating theory and practice* (4th ed.). SAGE.

Pope, C. (2000). Qualitative research in health care: Analysing qualitative data. *BMJ*, *320*(7227), 114-116. https://doi.org/10.1136/bmj.320.7227.114

Rabionet, S. E. (2011). How I learned to design and conduct semi-structured interviews: An ongoing and continuous journey. *The Qualitative Report*, *16*(2), 563-566. https://doi.org/10.46743/2160-3715/2011.1070

Rahman, M., Terano, H. R., Rahman, N., Salamzadeh, A., & Rahaman, S. (2023). ChatGPT and academic research: A review and recommendations based on practical examples. *Journal of Education, Management and Development Studies*, *3*(1), 1-12. https://doi.org/10.52631/jemds.v3i1.175

Roberts, R. (2020). Qualitative interview questions: Guidance for novice researchers. *The Qualitative Report, 25*(9), 3185-3203. https://doi.org/10.46743/2160-3715/2020.4640

Roulston, K. (2012). The pedagogy of interviewing. In J. F. Gubrium, J. A. Holstein, A. B. Marvasti, & K. D. McKinney (Eds). *The SAGE handbook of interview research: The complexity of the craft* (2nd ed., pp. 61-74), SAGE.

Rubin, H. J., & Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data* (3rd ed.). SAGE.

Sejnowski, T. J. (2023). Large language models and the reverse Turing test. *Neural Computation*, *35*(3), 309-342. https://doi.org/10.1162/neco_a_01563

Smith, J. A., Flowers, P., & Larkin, M. (2022). *Interpretive phenomenological analysis: Theory, method and research* (2nd ed.). SAGE.

Turner, D. W., III (2010). Qualitative interview design: A practical guide for novice investigators. *The Qualitative Report*, *15*(3), 754-760. https://doi.org/10.46743/2160-3715/2010.1178

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., and Schmidt, D.C. (2023), A prompt pattern catalog to enhance prompt engineering with ChatGPT. *EdArXiv preprint*. https://doi.org/10.48550/arXiv.2302.11382

Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P. Y. (2023, March). 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* (pp. 75-78). https://doi.org/10.1145/3581754.3584136

y Arcas, B. (2022). Do large language models understand us? *Daedalus*, *151*(2), 183-197. https://doi.org/10.1162/daed_a_01909

Author Note

Dr. Jessica Parker is a researcher and educator who is passionate about demystifying the research and writing process for scholars. She owns Dissertation by Design and Academic Insight Lab, and her research interests are at the intersection of technology and education. She is particularly intrigued by the potential of generative AI for academic purposes, exploring how this technology can revolutionize the way we conduct research, teach, and learn. Jessica has worked with a diverse range of researchers and scholars and continues to teach Doctor of Health Sciences students at Massachusetts College of Pharmacy and Health Sciences (MCPHS) University. Please direct correspondence to jessica.parker@mcphs.edu.

Dr. Veronica Richard is a qualitative methodologist at Dissertation by Design. She has held various academic positions, including adjunct, assistant, and associate professor roles at the University of Northern Colorado, Indiana University Northwest, and Concordia University Chicago. Her experience spans working with undergraduate, master's, and doctoral students, primarily in the fields of literacy and research methods. Throughout her career, Veronica has been an integral part of nine research teams, many of which were grant-funded and dedicated to supporting at-risk youth. Veronica earned her Ph.D. from the University of Northern Colorado in 2010, specializing in Applied Statistical Research and Research Methods with a cognate in Reading.

Dr. Kimberly Becker is an applied linguist who specializes in disciplinary academic writing and English for research publication purposes. She has a Ph.D. in applied linguistics and technology (Iowa State University, 2022) and an M.A. in teaching English as a second language (Northern Arizona University, 2004). Kimberly's research and teaching experience as a professor and communication consultant has equipped her to support native and non-native English speakers in written, oral, visual, and electronic communication. Her most recent publications are related to the use of ethical AI for automated writing evaluation and a co-authored e-book titled *Preparing to Publish*, about composing academic research manuscripts.

Article Citation

Parker, J. L., Richard, V., & Becker, K. (2023). Flexibility & iteration: Exploring the potential of large language models in developing and refining interview protocols. *The Qualitative Report, 28*(9), 2772-2791. https://doi.org/10.46743/2160-3715/2023.6695