

# Conducting Semi-Structured Interviews with ChatGPT and BARD on Computer Science Education



Dengel, A., Gehrlein, R., Fernes, D., Görlich, S., Maurer, J., Pham, H. H., ... & Eisermann, N. D. G. (2023, October). Qualitative Research Methods for Large Language Models: Conducting Semi-Structured Interviews with ChatGPT and BARD on Computer Science Education. In Informatics (Vol. 10, No. 4, p. 78). MDPI. Chicago <https://www.mdpi.com/2227-9709/10/4/78>

Presented by Elizabeth Gilbert

# Summary

Interview LLMs to gain insight into public opinions that were previously only obtainable by interviewing large groups of people.

Found that LLM interviews *should not* be used for research purposes.

Can help gain different perspectives on research topics.

Use to test interview guidelines before conducting real-world interviews.

# Prompting Strategies

Subject: Relevance of computer science in K-12 education.

LLMs interviewed: ChatGPT in English, ChatGPT in German, and BARD in English

Strategies:

1. **Ambiguous** responses → ask **follow-up** questions
2. **Lists** options with pros/cons → ask the AI to **choose one** of the possible options
  - a. Sometimes AI would still not choose
3. **Reduce ambiguity and non-existent source citing** by asking for its “**own opinion**”
4. **No role-playing** to not bias answers toward certain group/profession perspectives

# How well did it work?

Variation across interviews on different LLMs.

Variation across interviews on \*the same\* LLM.

Sometimes had conflicting answers across LLMs and within LLMs.

Answers were sometimes ambiguous, generic, lacking specificity.

# Did the authors trust the results?

Small “sample size” of LLMs, based on **accessibility** of the researchers

- findings may be specific to the biases, characteristics present in those models
- findings based on responses at specific time, possibly not reproducible with current version of the prompted LLMs

Authors **did not verify the responses** by asking education/computer science experts, possible that the responses were not actually reflective of those perspectives.

**Ethical concerns:** trained on datasets that could contain **biased data**.

# Lessons Learned

Unreliable and unpredictable outcomes → **should not be used for academic purposes.**

Could be used to:

- **Obtain exploratory insights and possible opinions** regarding a specific topic
- **Evaluate interview guidelines** in terms of their clarity and comprehensibility

Advice:

- **To provide validity** >> do same interview multiple times at low temperature (less surprising) to generate more coherent answers
- Use the phrase **“In your opinion. . .”** in the prompt, reduced ambiguity
- To **test interview guidelines** or to **discover multiple potential opinions**, a moderate temperature should be used (here, also different identities could be assigned)
- To adopt a new role, explicitly state **“Please take on the role of. . .”**.
  - Here, caution must be exercised to prevent the perpetuation of stereotypes.

Zhang, H., Wu, C., Xie, J., Lyu, Y., Cai, J., & Carroll, J. M.  
(2023). Redefining qualitative analysis in the AI era: Utilizing  
ChatGPT for efficient thematic analysis. arXiv preprint  
arXiv:2309.10771.

<https://arxiv.org/pdf/2309.10771>

Presented by Hao He

# High-Level Summary of the Paper

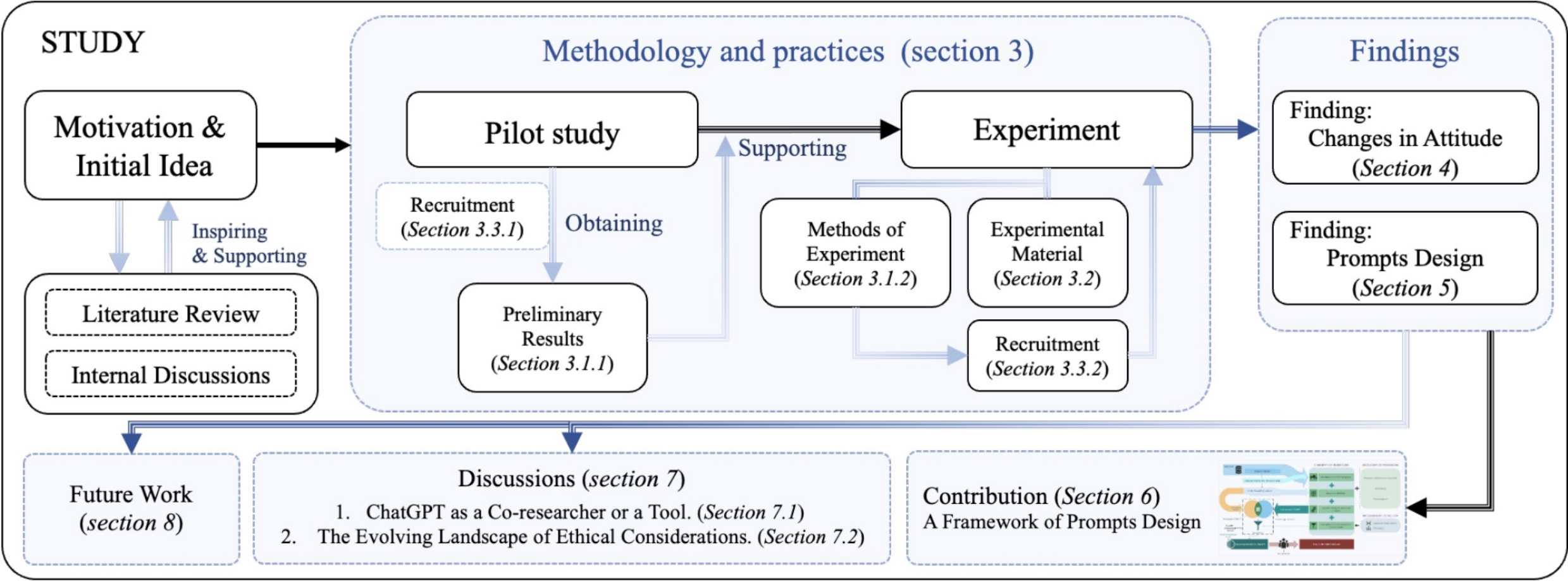
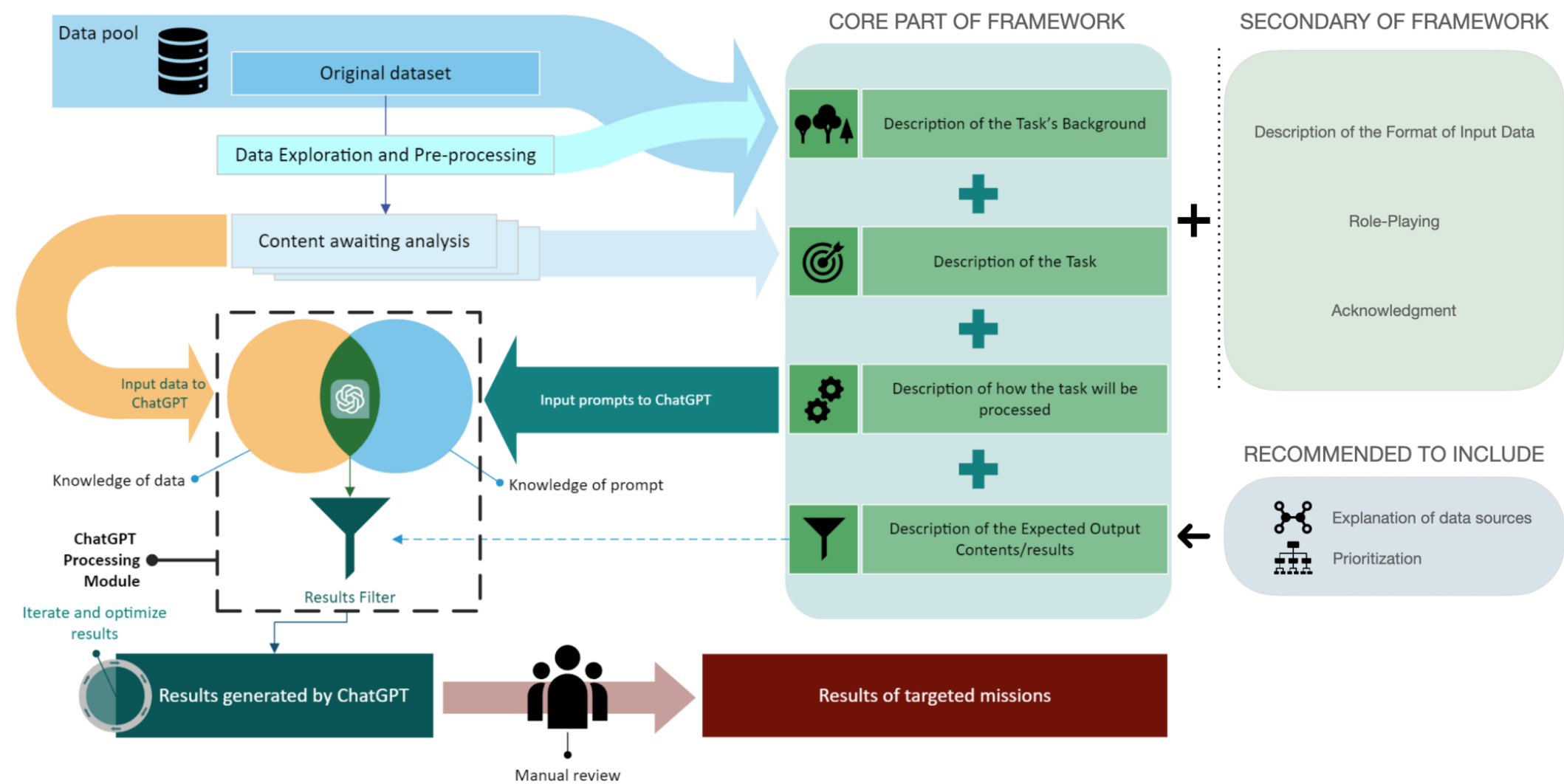


Fig. 1. Flowchart of this study. The diagram displays the main content distribution, mind map, and workflow of this study.



# High-Level Summary of the Paper



# Prompting Strategies

- Background or Conceptual Understanding
- Focus on Methodology (Goal of Task)
- Focus on Analytical Process
- Define the Format of Input
- Define the Format of Output
- Role-Playing
- Prioritization
- Classification, Transparency and Tracibility
- Acknowledgement of Expertise

# How Do Authors Trust the Results? How Well Does it Work?

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

- : a strategy not employed by the participant in crafting the prompt.
- : a strategy utilized by the participant in crafting the prompt.
- ◐: a strategy adopted by the participant following the researcher’s suggestion.

# Lessons Learned

- ChatGPT can be used as a tool / co-researcher for thematic analysis
- ChatGPT (LLM, AI) is actually gaining human trust from qualitative researchers, contrasting earlier non-LLM studies



# Integrating AI Language Models in Qualitative Research: Replicating Interview Data Analysis with ChatGPT

7 Feb 2024

---

Feb 20<sup>th</sup>, 2024

Jeffrey Chen(weigenc)



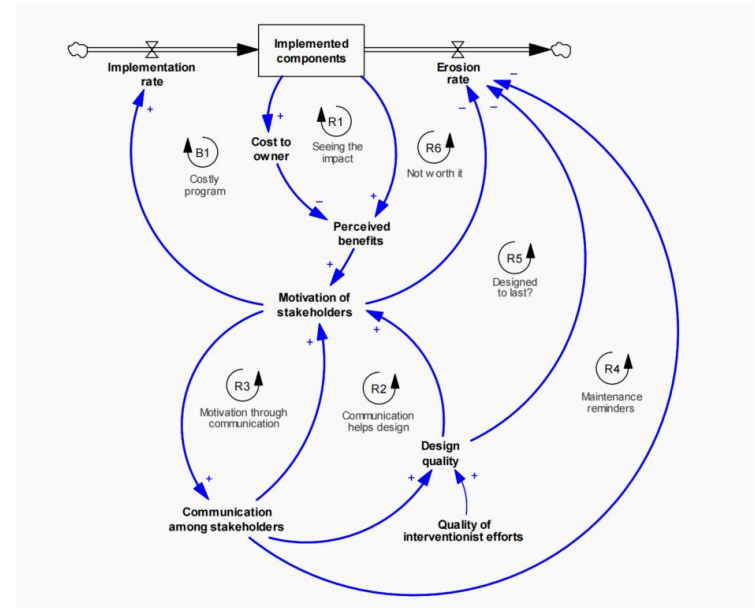
# Agenda

---

- Original Analysis
- LLM +
- Discussion

# Original Content

- Researchers use the interview data to do qualitative analysis and draw the causal loop diagram





## LLM - high level

---

1. Input interview *text*
2. Ask GPT to identify the variable
3. Ask GPT to create causal link
4. Gpt returns *text* description
5. Human draw according to the *text*
6. Comparing to original result





## LLM - Prompt

---

"Specifically, we used the interview transcripts from the original study and asked ChatGPT the following prompt:"

I want to create a *causal loop (CLD) diagram* from the text I provide. The text is the transcript of multiple interviews. *In the first step*, I just want to identify the key variables that can be used in the CLD. Could you please go through these interview data and *extract the key variables of interest? Read the text thoroughly*



## LLM - Prompt

---

“ASKED ChatGPT to identify the causal links between variables and propose as many feedback loops as possible using the following prompt”

This is good. Thank you. Now, in the next step, I want to *identify the causal links between variables*. Note that you can consider all the variables you identified above, and it is okay if the relationship between the two variables comes from different parts of the interviews. Also, note that you *can always go back* and read the interview data and find more additional variables if you think something is missing or you need more context and variables to establish causal relationships. Then, please identify and list the causal links between variables. After identifying the causal links, *identify and list as many feedback loops as possible*



## LLM - Strategy

---

“We repeated these prompts for all the interview data **within the same chat session.**”

“We repeated the whole process **three times** (in three separate chat sessions) and **rotated the order** of the interview documents that ChatGPT received.”

# LLM - Results

ChatGPT identified feedback loops that *were not reported* in the original study

Feedback loops identified by ChatGPT	CLDs drawn by authors based on ChatGPT's outputs
"Increased awareness and education about health benefits lead to higher motivation and personal health goals, which in turn encourage further participation and engagement in the wellness program, leading to more awareness."	<pre>graph TD; A[Awareness and education about health benefits] --&gt; B[Motivation and personal health goals]; B --&gt; C[Participation and engagement in the wellness program]; C --&gt; A; R((R))</pre>
"As employee health improves due to lifestyle changes and program participation, there might be a decrease in the perceived need for intense participation, balancing out the engagement levels."	<pre>graph TD; A[Program participation] --&gt; B[Lifestyle change]; B --&gt; C[Employee health]; C --&gt; D[Perceived need for participation]; D --&gt; A; B((B))</pre>
"Time constraints and workload impact program participation negatively, but as health improves through participation, there might be an increase in overall energy and efficiency, potentially easing time constraints." *	<pre>graph TD; A[Time constraints and workload] --&gt; B[Program participation]; B --&gt; C[Employee health]; C --&gt; D[Overall energy and efficiency]; D --&gt; A; R((R))</pre>



# Discussion

---

- Advantage: Combining LLM's direct, *unbiased* analysis with the researcher's in-depth analytical perspective can provide a more comprehensive and balanced understanding.
- Disadvantage: Lack the capacity for nuanced understanding and integration of data
- Does not identify the erosion of implemented components
- It *does not trace the causal chains in the same direction* -> wrong analysis on some casual links



# Discussion

---

- What exact tasks is the LLM used for?

Qualitative analysis

- What are the prompting strategies?

1. Iterative Refinement: Refine and adjust queries incrementally based on LLMs' preliminary responses to enhance relevance and accuracy of the results

2. Change Query Sequence: change the input sequence try to get the response.

- How well does it work?

Gpt find something new

- Lessons learned.

It can provide relatively unbiased results

Because people interpret the data from an organizational perspective

- How do the authors trust the results?

The authors think GPT can become an assistance but can not replace the human