🌻 Step 1 – Conducting the chat interviews

In the world of machine learning, a clear distinction can be made between supervised and unsupervised approaches (Ziulu et al., 2024). Using genAI to conduct interviews and code texts blurs this boundary. In our case, we developed our semi-generic instructions for interviewing, giving the AI instructions on how to behave, and how to make follow-up questions based on the interview objectives. Once the data collection is done, we create a separate genAI prompt to code causal links as a trial-and-error process, monitoring the quality of the coding post-hoc. We did not have an explicitly stated ground truth about exactly how the interview should look or which causal claims were “really” present within each text passage or how their causes and effects should be labelled, as we believe neither of these questions have a definitive answer; rather, we monitored AI’s responses coding post-hoc, iterating the prompt over many cycles to improve its performance. "Prompt engineering" (Ferretti, 2023) like this can be considered a kind of supervision because it steers the AI’s responses in a desired way.

Once the prompt was finalised, the interview AI was left to conduct interviews without further supervision. This prompt can remain broadly the same across different studies. However, the response of the AI can be highly sensitive to small differences in the "prompt" and other settings (Jang & Lukasiewicz, 2023). Small adjustments made for specific studies, such as adjusting the instructions to focus better on research objectives, remain a vital point of human intervention.

This paper presents results from a proof-of-concept analogue study. We employed online workers as respondents, recruited via Amazon’s MTurk platform (Shank, 2016). We decided to investigate respondents’ ideas about problems facing the USA, as this generic theme was likely to elicit opinions from randomly chosen participants. This unsophisticated way of recruiting respondents means that the results cannot be generalised to a wider population in this case.

We had no specific evaluative questions in mind; We aimed to demonstrate a method which can be easily adapted to a specific research question.

A short semi-structured interview guideline was designed on the theme of "What are the important current problems facing the USA and what are the (immediate and underlying) reasons for those problems?". We aimed to construct an overall collective “ToC” around problems in the USA. As it does not encompass a specific intervention this theory is not an example of a program theory.

© Causal Map Ltd 2025 · CC BY-NC 4.0

This interview guideline was implemented via an online interview "AI interviewer" called "Qualia", which uses the OpenAI Application Programming Interface (API) to control the AI's behaviour. Qualia is designed to elicit stories from multiple individual respondents, in an AI-driven chat format. Individual respondents are sent a link to an interview on a specific topic and, after consenting, are greeted by the interviewer. Rather than following a set list of questions, the interviewer is instructed to adapt its responses and follow-up questions depending on the respondents' answers, circling back to link responses and asking for more information as appropriate, focusing on the interview's objective mentioned above. These behaviours are based on the instructions written by the authors.

The respondents, who had the level of “Master” on Amazon's MTurk service, each completed an interview. The Amazon workers were given up to 19 minutes to complete the interview.

We repeated this interview at three different timepoints in September, October and November 2023, inviting approximately N=50 respondents each time. The data from the three timepoints was pooled.

References

Ferretti (2023). *Hacking by the Prompt: Innovative Ways to Utilize ChatGPT for Evaluators*. https://doi.org/10.1002/ev.20557.

Jang, & Lukasiewicz (2023). *Consistency Analysis of ChatGPT*. https://doi.org/10.48550/arXiv.2303.06273.

© Causal Map Ltd 2025 · CC BY-NC 4.0