SAI What?! Ping the Bots Before You Probe the People:
Testing Large Language Models for Pre-Cognitive Interviewing in Survey Research

Keywords: Cognitive Interviewing, Large Language Models, Persona Prompting, Questionnaire Design, Survey Evaluation and Development

With the rapid evolution and availability of large language models (LLMs) such as GPT-4 (OpenAI, 2023), Claude, and Gemini, there is growing interest in integrating these tools into the survey research process. There is emerging and current work focusing on tasks such as developing questionnaires on specific topics (Padgett et al., 2024), assessing readability of survey items (Olson and Buskirk, 2025), revising existing scales (Yun et al., 2023) and open-ended response classification or summarization (Gilardi, Gessler, & Kubli, 2023; Lerner at al., 2024). This work expands these applications by investigating applications of LLMs to the earliest stage of survey instrument development: using LLMs to simulate cognitive interviews prior to engaging human respondents. In essence the LLM models would be used to conduct what we are calling a "pre-cognitive interview," where AI is used to evaluate how questions might be interpreted before human field testing. This early-stage AI-assisted cognitive interview could offer early insights into comprehension, recall, judgment, and response challenges that may arise among human participants (Willis, 2005) and could then be incorporated by researchers in drafting more applicable (human) cognitive interviewing scripts and protocols.

This study systematically investigates the potential of LLMs to surface issues typically identified through human cognitive interviewing before involving human participants. Survey questions are drawn from rigorously developed instruments and LLMs are prompted using standard probes and enhanced prompting methods including chain-of-thought prompting (Wei et al., 2022) and metacognitive scaffolding (Zhang et al., 2022). We will also explore role prompting methods embedded into a session prompt where we ask the LLM to conduct the pre-cognitive interview as though they were an expert in cognitive interviewing using specific names of historical, noted luminaries in the field like Gordon Willis or Roger Tourangeau or an expert who has mastered information in such texts such as Tourangeau et al. (2000). The pre-cognitive interviewing will be embedded within a series of personas that are simulated (Argyle et al., 2023) to reflect key demographic and cognitive characteristics found in actual cognitive interview participants (Miller et al., 2014), allowing us to

approximate how different population segments might interpret survey items and explore features of persona definition that might impact the quality of the pre-cognitive interviews relative to human populations.

Multiple LLMs are evaluated under varied prompt conditions within basic and enhanced persona definitions and across replications. The outputs from the pre-cognitive interviews are thematically coded and compared to findings from comparable human cognitive interviews. The comparisons will focus specifically on assessing overlap in identified issues, thematic breadth, and diagnostic value. We also examine whether human coders—and automated classifiers—can distinguish between AI- and human-generated interviews, and explore patterns in language and problem detection that may be unique to each source. Importantly, the goal is not to replace human cognitive interviewing, but to examine how pre-cognitive interviews with LLMs may be optimized for use to inform and enhance protocol development.

Our work contributes to a growing methodological literature on AI-assisted research, particularly in the context of simulating human judgment and opinion (Gilardi et al., 2023; Argyle et al., 2023). It also raises questions about the fidelity of LLM simulations to actual human cognition—a concern especially relevant to fields dependent on nuanced language understanding. Nonetheless, the potential value of LLMs as low-cost, rapid-turnaround diagnostic tools is substantial, particularly for iterative instrument development in resource-constrained environments. By "pinging the bots before probing the people," we explore how artificial intelligence might complement human expertise in cognitive interviewing process and within survey research writ large.

References

- Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). *Out of one, many: Using language models to simulate human samples*. Political Analysis, 31(3), 337–351.

- Gilardi, F., Gessler, T., & Kubli, M. (2023). *ChatGPT outperforms humans in simulating political opinions*. *Proceedings of the National Academy of Sciences*, 120(36), e2305016120.

- Lerner, J. et al. (2024) Optimizing Open-Ended Questions for Natural Language Processing and Enhancing Survey Research Quality with LlMs. Paper presented at the 79th annual conference of the American Association of Public Opinion Research; Atlanta, GA, May 15-17, 2024. Paper accessed on June 18, 2024 from: https://aapor.confex.com/aapor/2024/meetingapp.cgi/Paper/3202

- Miller, K., Willson, S., Chepp, V., & Padilla, J. (2014). *Cognitive Interviewing Methodology*. Wiley.

- Olson, K., & Buskirk, T. D. (2025). *Chatbot is a two-syllable word… or is it? Using generative AI for survey question readability assessments*. Paper presented at the 80th Annual Conference of the American Association for Public Opinion Research (AAPOR), St. Louis, MO, May 14–16, 2025.

- OpenAI. (2023). *GPT-4 Technical Report*. https://openai.com/research/gpt-4

- Padgett, Z. et al. (2024). Evaluating the Quality of Questionnaires Created with Survey monkey's Build with AI. Paper presented at the 79th annual conference of the American Association of Public Opinion Research; Atlanta, GA, May 15-17, 2024. Paper accessed on June 20, 2024 from: https://aapor.confex.com/aapor/2024/meetingapp.cgi/Paper/3198

- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. "The psychology of survey response." (2000)

- Wei, J., Wang, X., Schuurmans, D., et al. (2022). *Chain-of-thought prompting elicits reasoning in large language models*. arXiv:2201.11903.

- Willis, G. B. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Sage Publications.

- Zhang, X., Bai, Y., Lin, Z., et al. (2022). *Automatic chain of thought prompting in large language models*. arXiv:2210.03493.