

# **AIInterviewer: A Platform for Designing and Conducting AI-led Qualitative Interviews**

**Tobias Priesholm Gårdhus**

University of Copenhagen  
tpg@sodas.ku.dk

**Nikolas Vitsakis**

IT University of Copenhagen  
nvit@itu.dk

**Fie Lejre Frederiksen**

University of Copenhagen  
flf@sodas.ku.dk

**Anna Rogers**

IT University of Copenhagen  
arog@itu.dk

**Hjalmar Bang Carlsen**

University of Copenhagen  
hc@sodas.ku.dk

## **Abstract**

There are now multiple proposals for systems based on Large Language Models (LLMs) to conduct automated qualitative interviews, but most of the current solutions rely on proprietary LLMs, which compromises reproducibility and data security. They also rely on LLMs for all interview tasks, which limits standardisation of question wording as well as control over question order. To address these issues, we introduce the *AIInterviewer*<sup>1</sup> platform, an open-source solution based on a multi-agent pipeline that combines controlled question administration of survey software with the flexibility of LLMs. *AIInterviewer* is an interdisciplinary effort designed to implement best practices of qualitative interviewing in social science, and it can run with locally hosted models to ensure security, transparency, and reproducibility. Our platform provides a web-based GUI supporting each phase of data collection: from interview guide design and pilot testing to interview distribution and data collection monitoring.

## **1 Introduction to AI-interviewing**

The recent years saw multiple proposals for conversational AI tools to conduct interviews with humans (Chopra and Haaland, 2023; Geiecke and Jaravel, 2024; Wuttke et al., 2024; Xiao et al., 2020b; Jacobsen et al., 2025; Sturgis et al., 2025). The so-called AI-interview systems rely on Large Language Models (LLMs) for asking open-ended and follow-up questions. This approach scales up qualitative interview techniques, which have traditionally been constrained by high costs of data collection (Chopra and Haaland, 2023; Jacobsen et al., 2025). Proponents argue that AI-led interviewing offers a new methodological middle ground between qualitative depth and quantitative scale,

<sup>1</sup>Platform demo: <https://app.ainterviewer.dk>  
YouTube demo: <https://www.youtube.com/watch?v=B0WuNsFVqAU>  
Source code: <https://www.github.com/ainterviewer>

enabling researchers to collect large-N qualitative interview data in ways previously unattainable.

However, current research and the state of the art have largely relied on general-purpose commercial chatbots (Xiao et al., 2020a; Chopra and Haaland, 2023). While such setups allow researchers to conduct AI-led interviews with minimal preconfiguration, they have the following disadvantages:

- limited control over question formulation, which can undermine comparability across interviews (Gerson and Damaske, 2020);
- lack of control over the model, which undermines reliability and reproducibility (Palmer et al., 2024; Staudinger et al., 2024);
- insufficient control over data, which undermines qualitative requirements for confidentiality (Li, 2026; Shanmugarasa et al., 2025; Evertz et al., 2024).

To establish AI-interviewing as a rigorous method, researchers need tools that combine conversational AI flexibility with controlled outputs and transparent system designs.

To address this gap, we created the *AIInterviewer*, a new platform for designing, instructing, testing, and distributing AI-interviews. Interviews are conducted by a multi-agent pipeline backed by open-sourced LLMs, where each agent handles different elements of the interview process. Unlike previous work, our approach enables fine-grained control of agentic behaviour through customizable interview guides, allowing for both deterministic ordering and responsive question formulation. For the deployment stage, our platform also includes built-in tools for pre-testing the interview, large-scale distribution, and experimental randomisation across and within interview conditions. To promote transparency, reproducibility, and collaborative scientific progress, *AIInterviewer* is released under the *GNU Affero General Public License v3.0 (AGPLv3)*.

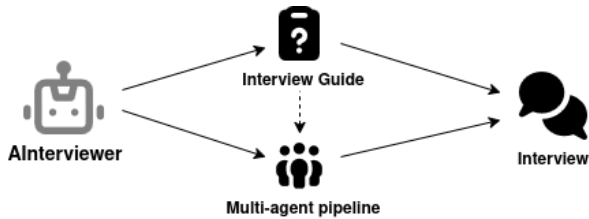


Figure 1: System design of the AInterviewer platform. The interview guide specifies the interview structure and how it may be adjusted in specific interviews based on information needs and prior conversational context.

## 2 Interview System Design

Figure 1 consists of the interview system, and the system that supports the interview designers in testing and distributing the interviews. We begin with the former. As shown in Figure 1, it comprises the interview guide which structures a given interview (§2.1), and the multi-agent LLM pipeline that generates outputs conditioned on the interview guide (§2.2). The features that support the testing and distribution of interviews will be covered in §3.

### 2.1 Interview guide

Each interview is defined by an *interview guide* that provides structure in accordance with guidelines for effective interview design, established in social science research (Gerson and Damaske, 2020; Kallio et al., 2016; Kelly et al., 2010; Naz et al., 2022). The guide organises the interview questions into thematic sections. Each section consists of a set of *main questions* and *probes* (follow-up questions), each informed by *context*. The interview designer’s task is to specify and populate an interview guide, for which we provide a customizable template. For example, designers can specify the overall interview focus, the focus of each section, precise question formulations, and question ordering—all designed to support high-quality data collection.

#### 2.1.1 Context Fields

The interview guide allows designers to provide contextual information that models can use to adjust question generation. For example, the *framing* field allows the inclusion of information relevant to the entire interview (e.g., “*This interview is about how people use generative AI*”), which persists through the thematic shifts. It is also possible to specify context that should only be used within a given thematic section (e.g., “*This section*

*is about the use of generative AI to solve work-related tasks*”).

#### 2.1.2 Main questions

Main questions are pre-defined by the interview designer, and must be asked in each interview to ensure cross-comparability between interviews. To ensure comparability between interviews, the interview designer can set<sup>2</sup> them to be asked verbatim as written in the interview guide. Conversely, interview designers can also instruct the system to reformulate pre-defined questions to resolve misunderstandings or to re-introduce previous conversational context, thereby improving data quality (Conrad and Schober, 2000; Herlihy et al., 2024; Sarkar et al., 2025).

#### 2.1.3 Probes

We define probes as potential follow-up questions that may be asked depending on the respondent’s answer to a main question. Probes are a core component of semi-structured and unstructured interview methodology (Small and Calarco, 2022; Robinson, 2023). Their responsive generation allows the interviewer to support the interviewee’s own narratives and perspectives in relation to the original question, thereby producing in-depth data essential for interpretative inference (Gerson and Damaske, 2020; Lareau, 2021; Weiss, 1995; Small and Calarco, 2022). Probing should support interviewees in developing their own answers, and probes should therefore always be relevant to the original response.

Since probes are context-dependent, the interview guide only provides examples of such questions, and the system retains flexibility to reformulate their wording or generate a more contextually relevant follow-up question based on the unfolding conversation. Alternatively, if no probes are pre-defined, the model will use both its instructions and the context of the interview to generate probes (for more information on probing agents, see §2.2.1).

### 2.2 Multi-agent Pipeline

Interviews are conducted via a multi-agent pipeline, in which each agent is an instantiation of a LLM with its own task. These tasks are specified within

<sup>2</sup>All decisions made about the flow of the interview are based on the structure of the interview guide and explicit classification steps. Hence, if it is needed to ensure that main questions are asked verbatim as set in the interview guide, we can simply skip the step where it is reformulated. See §2.2 for more details.



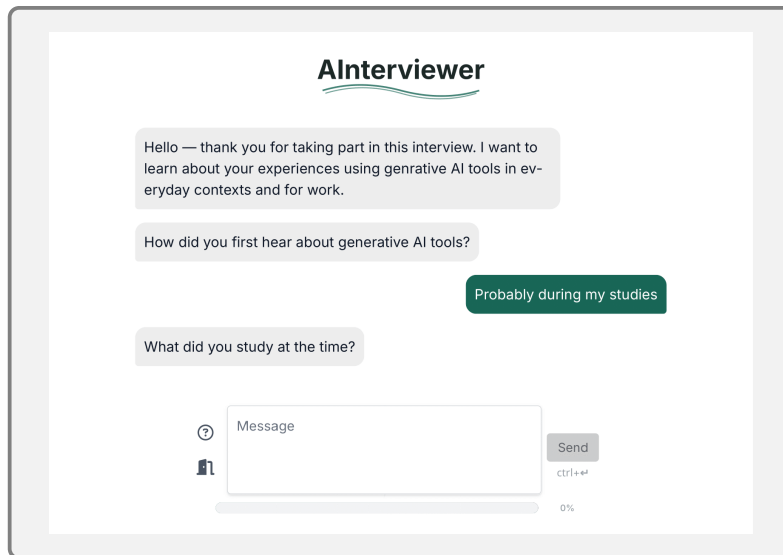


Figure 3: The chat interface that interviewees interact with the AInterviewer.

reformulate a question to account for information previously provided by the respondent. The purpose of the conversational transition is to ensure that the question fits the flow of the conversation without altering its substantive content. For example, if a respondent has already discussed their use of AI at work prior to a hypothetical question about AI and work, the reformulation agent may generate a transition (e.g. ‘*Let’s talk more about your use of AI at work*’), thereby adding responsive passages without altering the core content of the question. In contrast, reformulating the content of the question involves updating the question itself to incorporate information already provided by the respondent, typically resulting in a more specific question (e.g., ‘*You said that you mainly use AI to help with writing tasks; can you tell me more about the specific writing tasks you use it for?*’) This agent is needed because of our reliance on an otherwise stringent structure imposed by the interview guide.

### 2.2.3 Classification agents

Classification agents are responsible for altering the flow of the interview including triggering other agents such as the reformulation agent. They do so by evaluating whether specific conditions are met, based on a structured prompt template and relevant context which can include descriptions from the interview guide and transcript of the ongoing interview. The output of the classification agent is a binary value, which triggers the relevant action. We currently implement three classification agents at different points of the interview, to check for the

following conditions:

- Whether the question has already been answered in a previous part of the interview, which may trigger the reformulation agent. This is based on the transcript of the current interview, along with the main question description, and is evaluated before each main question is asked.
- Whether the respondent refuses to answer a question, which likewise results in a transition to a new question, and is evaluated before asking each probe.
- Whether a sufficient number of follow-up questions have been asked for the current main question, resulting in a transition to a new question. This is based on the transcript of the current main question, along with the section and main question description, and is evaluated after each probe.

### 2.3 Interviewee user experience

Interviewees participate in the interview through a web-based application, as shown in [Figure 3](#). The upper portion of the screen displayed the chat history, while the lower portion features an expanding text box in which interviewees type their responses. The design of the interface has been inspired by messaging applications such as Messenger ([Meta Platforms, Inc., 2011](#)), with a strong emphasis on minimal UI and high-contrast visual elements ([Lister et al., 2020](#); [Dong, 2019](#)). The conversation is

currently text-based, with its flow mirroring real-life conversations, with dialogue structured around a turn-taking format. Voice-based interviewing interface is in development.

### 3 Workflow for Interview Designers

*AInterviewer* can support social science researchers across all stages of interview-based studies, enabling them to easily design, prototype, distribute, and monitor interviews without extensive prior experience in NLP. The full workflow is shown in Figure 4, and we describe the main steps below.

**Designing an interview guide** As discussed in §2.1, interview designers need to define the thematic sections of the interview, formulate the main questions and probes within each section, and determine the overall question order. In practice, interview guides often undergo multiple iterations before deployment. Accordingly, interview designers may proceed to pilot testing (see below) with a preliminary interview guide, rather than aim to produce a fully developed guide at once.

**Agent configuration** This phase involves selecting the LLM models to use for each agent (either via locally hosted models or through third-party API providers) and setting their hyperparameters. In addition, the default agent prompts can be customised to tailor the agents to specific interview objectives and/or specific model.

**Pilot testing** This functionality simulates interviews in a controlled environment, enabling the researchers to test their guide and agent configuration before deploying it to real subjects. Interviewee responses can be either pre-defined by researchers to test specific conditions, or automatically generated by customizable synthetic agents. The synthetic respondents can be created using pre-defined personas or randomised background characteristics, as shown in Figure 4. During the pilot phase, users can test multiple aspects of the AI interview, including (1) the content of the questions and their formulation, (2) the order of the interview, (3) the performance of the different LLM agents, and (4) the system’s ability to handle high-risk situations.

**Distributing the interview** Once the interview design is finalised, the researchers can distribute interviews to participants by sharing automatically generated unique URLs or QR codes. The platform also supports the automatic allocation of respon-

dents across different interview versions. For example, controlled interview versions may differ in the wording of specific questions, interview length, or supplementary materials, allowing researchers to test different treatment effects (e.g., differences in participant disclosure based on question ordering).

**Data management and sharing** Researchers can access interview data directly through the Control Panel in its native HTML format or export it in their preferred file format (e.g., CSV). The resulting data files can then be imported into qualitative analysis software such as NVivo (Lewis, 2004) or into software environments that support quantitative text analysis, such as R. In addition, *AInterviewer* supports sharing the exact system configuration of a given interview study, enabling straightforward replication of the full setup using the accompanying open-source system, from the interview guide, to the precise agent prompts.

### 4 System evaluation: pilot study

*AInterviewer* was evaluated in a pilot study with 40 on-site participants. This study is described in detail in separate concurrent work that focuses on the comparative analysis of the quality of data obtained by *AInterviewer* vs trained human interviewers (Guven et al., 2025). For completeness, we include a summary of these results with the present technical description of *AInterviewer* system. This study was conducted with OpenHermes-2.5-Mistral-7B<sup>4</sup> as the underlying LLM for the multi-agent pipeline.

**Study setup.** The participants<sup>5</sup> were randomly assigned to be interviewed by either a trained human interviewer or *AInterviewer* via a chat interface. Informed consent was obtained, and participants were aware of the experimental condition (e.g., whether they were participating in a human-led or AI-led interview). In total, 946 responses were collected and evaluated for three metrics widely used in qualitative research: exposure (operationalized as length of replies obtained from

<sup>4</sup>An open-source model finetuned on Mistral 7B (Jiang et al., 2023): <https://huggingface.co/teknium/OpenHermes-2.5-Mistral-7B>

<sup>5</sup>The participants were university students, recruited through face-to-face outreach on campus and via flyers and posters. The topic of the interviews was university students’ help-seeking practices and experiences in the academic setting. The compensation for participation was a voucher for a free lunch at the university canteen. All the interviews were conducted in English. The study was approved by the Institutional Review Board at the University of Copenhagen.

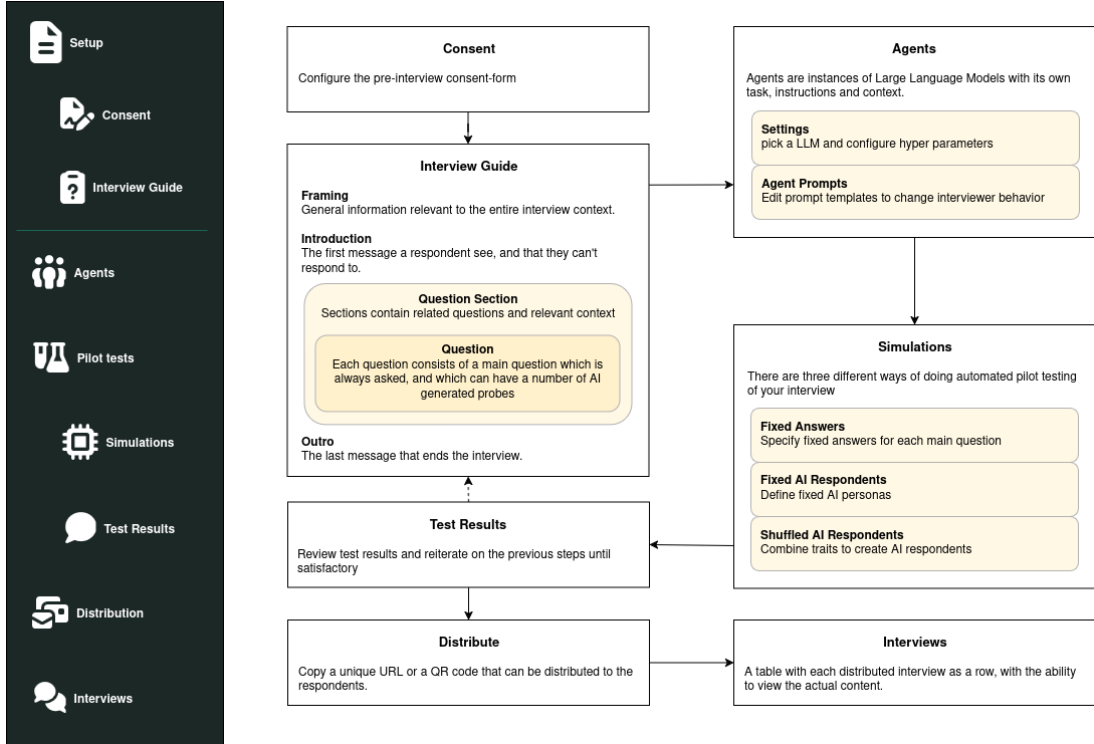


Figure 4: The full workflow for Interview Designers. The setup section (Consent, Interview guide) contains information that must be pre-determined prior to the conduct of the study. The Agents section allows the customising of the multi-agent pipeline, which can be used to pilot prototypes in simulations. Once a study is finalised, it can be easily distributed to participants to partake in AI-led interviews.

	U-val	p-val
Response Length	142.0	0.470
Interview Length	329.0	0.012
Questions per Interview	392.0	0.000
Mean Specificity	185.5	1.000
Mean Relevance	161.0	1.000

Table 1: Results of the Mann–Whitney U analyses comparing the two independent groups (interviews led by trained human interviewers vs *AIInterviewer*) across all outcome variables. For each measure, the table reports the U statistic and the Bonferroni-adjusted p-value.

the participants and number of question answered), relevance (i.e. addressing the topic of the interview) and specificity (providing concrete details of the participants’ lived experience). Specificity and relevance were manually coded by the authors of the study on a 5-point scale.

**Results.** We compare the two independent groups through a Mann-Whitney U (MWU) test (Mann and Whitney, 1947), followed by a Bonferroni correction (Bender and Lange, 2001) to control for multiple comparisons, as shown in Table 1.

When assessing response lengths on an individual question basis, we find that human interview-

ers elicited longer responses (38% increase), although this effect was not statistically significant ( $p_{MWU} = 0.0470$ ,  $U = 142.0$ ). However, when comparing total response length over whole interviews, the trend reverses, with AI-led interviews eliciting substantially larger corpora ( $p_{MWU} = 0.012$ ,  $U = 329.0$ ). This effect is explained by the time required to formulate questions between the human and AI interviewers, with AI interviewers outperforming humans by a wide margin.

Regarding response quality, our evaluative metrics indicate no significant differences in specificity and relevance: relevances score were the same ( $M=4.9$ ). For specificity, human interviews scored higher, but this effect was not statistically significant ( $M=3.8$  vs  $M=3.5$ ). We interpret these findings as overall promising for AI-led interviewing, although by no means indicating that it can replace expert human interviewers.

## 5 Conclusion and Future Work

We present *AIInterviewer*, an open-source platform for conducting large-scale AI-led qualitative interviews that is highly customisable and fully transparent in its design. Compared to existing solutions

that mostly rely on commercial LLMs accessed via APIs, *AInterviewer* offers a greater control over both the underlying model and the interview process, which is designed to strictly follow an interview guide. Our solution is grounded in best practices of qualitative interview research, and supports the researchers across the entire workflow of developing, testing, and deploying the interviews.

**Future work.** Currently *AInterviewer* supports text-based interviews, and future work will focus on incorporating speech functionality to enable more natural interactions between participants and the interview system. We also plan to equip the platform with tools for qualitative and quantitative analysis of the collected interviews.

**Limitations.** While the results of the pilot study summarized in §4 are positive, they are preliminary and should not be interpreted as the final verdict on the quality of AI-led interviewing. It is generally a new method for social science research, and its evaluation methodology remains to be established. This includes not only the criteria for comparison to expert human interviewers, but also AI-specific criteria including agent safety, adherence to instructions, linguistic quality of generated questions, etc. Currently, there is no consensus on which constructs and operationalisations should be included in evaluation (Wuttke et al., 2024; Geiecke and Jaravel, 2024; Chopra and Haaland, 2023).

## 6 Ethical Considerations

**Safety risks for the participants.** Like in other proposed implementations of AI-interview systems, the generation process itself remains stochastic. This entails the usual safety problems for all applications of large language models, including unintended outputs, steering the conversation towards unintended topics, or hallucinations when prior conversational content is referenced in long-form conversations (Huang et al., 2025; Narayanan Venkit et al., 2024). There is also the possibility that the interviewees could deliberately undermine or ‘jailbreak’ the system during interviews (Jin et al., 2024; Shneiderman, 2020). These issues apply to all AI-led interview systems, and addressing them will require a field-wide effort.

**Safe interviewing of vulnerable participants.** Interview systems handling responses from potentially vulnerable participants need to ensure the highest level of safety for them in the interview

process. For example, participants may mention something like self-harm, which should be flagged for professional assistance. While the explicit addition of agents responsible for flagging potential safety-critical conversational turns and interviews is planned, the current version of our system is not equipped to handle sensitive information from vulnerable population samples, which makes it appropriate for use only in highly controlled settings. Explicit safeguarding of participants during AI-led interviews has also not been implemented in most prior work (Wuttke et al., 2024; Chopra and Haaland, 2023; Geiecke and Jaravel, 2024; Xiao et al., 2020a,b)

## 7 Acknowledgements

This work was supported by a research grant ([VIL69261]) from Villum Fonden.

## References

- Ralf Bender and Stefan Lange. 2001. Adjusting for multiple testing—when and how? *Journal of clinical epidemiology*, 54(4):343–349.
- Felix Chopra and Ingar Haaland. 2023. Conducting Qualitative Interviews with AI.
- Frederick G. Conrad and Michael F. Schober. 2000. Clarifying Question Meaning in a Household Telephone Survey\*. *Public Opinion Quarterly*, 64(1):1–28.
- Rui Dong. 2019. Minimalist style of ui interface design in the age of self-media. In *Proc. Of International Conference on Information and Social Science*, pages 217–221.
- Jonathan Evertz, Merlin Chlosta, Lea Schönherr, and Thorsten Eisenhofer. 2024. Whispers in the machine: Confidentiality in llm-integrated systems. *arXiv preprint arXiv:2402.06922*.
- Friedrich Geiecke and Xavier Jaravel. 2024. *Conversations at Scale: Robust AI-led Interviews with a Simple Open-Source Platform*. Issue: 4974382 Place: Rochester, NY.
- Kathleen Gerson and Sarah Damaske. 2020. *The Science and Art of Interviewing*. Oxford University Press.
- Semra Yuksel Guven, Tobias Gårdhus, Andreas Bjerre-Nielsen, and Hjalmar Bang Carlsen. 2025. *Comparing AI-led to human-led chat-based interviews: Motivations, initial results and challenges*. Working paper.

- Christine Herlihy, Jennifer Neville, Tobias Schnabel, and Adith Swaminathan. 2024. On overcoming miscalibrated conversational priors in llm-based chatbots. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, UAI '24. JMLR.org.
- Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and 1 others. 2025. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2):1–55.
- Rune Moberg Jacobsen, Samuel Rhys Cox, Carla F Griggio, and Niels Van Berkel. 2025. Chatbots for Data Collection in Surveys: A Comparison of Four Theory-Based Interview Probes. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, pages 1–21.
- Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, L elio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timoth ee Lacroix, and William El Sayed. 2023. *Mistral 7b*. Preprint, arXiv:2310.06825.
- Haibo Jin, Andy Zhou, Joe D. Menke, and Haohan Wang. 2024. [Jailbreaking large language models against moderation guardrails via cipher characters](#). In *Advances in Neural Information Processing Systems*, volume 37, pages 59408–59435. Curran Associates, Inc.
- Hanna Kallio, Anna-Maija Pietil a, Martin Johnson, and Mari Kangasniemi. 2016. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of advanced nursing*, 72(12):2954–2965.
- Susan E Kelly, Ivy Bourgeault, Robert Dingwall, and 1 others. 2010. Qualitative interviewing techniques and styles. *The SAGE handbook of qualitative methods in health research*, 19:307–326.
- Annette Lareau. 2021. *Listening to people: A practical guide to interviewing, participant observation, data analysis, and writing it all up*. University of Chicago Press.
- R Barry Lewis. 2004. Nvivo 2.0 and atlas. ti 5.0: A comparative review of two popular qualitative data-analysis programs. *Field methods*, 16(4):439–464.
- Tianshi Li. 2026. Agentic llms as powerful deanonymizers: Re-identification of participants in the anthropic interviewer dataset. *arXiv preprint arXiv:2601.05918*.
- Kate Lister, Tim Coughlan, Francisco Iniesto, Nick Freear, and Peter Devine. 2020. Accessible conversational user interfaces: considerations for design. In *Proceedings of the 17th international web for all conference*, pages 1–11.
- Henry B Mann and Donald R Whitney. 1947. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60.
- Meta Platforms, Inc. 2011. Messenger.
- Pranav Narayanan Venkit, Tatiana Chakravorti, Vipul Gupta, Heidi Biggs, Mukund Srinath, Koustava Goswami, Sarah Rajtmajer, and Shomir Wilson. 2024. [An audit on the perspectives and challenges of hallucinations in NLP](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 6528–6548, Miami, Florida, USA. Association for Computational Linguistics.
- Nuzhat Naz, Fozia Gulab, and Aslam Mahnaz. 2022. [Development of qualitative semi-structured interview guide for case study research](#). *Competitive Social Science Research Journal*, 3(2):42–52.
- Alexis Palmer, Noah A. Smith, and Arthur Spirling. 2024. [Using Proprietary Language Models in Academic Research Requires Explicit Justification](#). *Nature Computational Science*, 4(1):2–3.
- Oliver C Robinson. 2023. Probing in qualitative research interviews: Theory and practice. *Qualitative Research in Psychology*, 20(3):382–397.
- Rupak Sarkar, Bahareh Sarrafzadeh, Nirupama Chandrasekaran, Nagu Rangan, Philip Resnik, Longqi Yang, and Sujay Kumar Jauhar. 2025. Conversational user-ai intervention: A study on prompt rewriting for improved llm response generation. *arXiv preprint arXiv:2503.16789*.
- Yashothara Shanmugarasa, Shidong Pan, Ming Ding, Dehai Zhao, and Thierry Rakotoarivelo. 2025. Privacy meets explainability: Managing confidential data and transparency policies in llm-empowered science. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–8.
- Ben Shneiderman. 2020. Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered ai systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 10(4):1–31.
- Mario Luis Small and Jessica McCrory Calarco. 2022. *Qualitative literacy: A guide to evaluating ethnographic and interview research*. Univ of California Press.
- Moritz Staudinger, Wojciech Kusa, Florina Piroi, Aldo Lipani, and Allan Hanbury. 2024. [A reproducibility and generalizability study of large language models for query generation](#). In *Proceedings of the 2024 Annual International ACM SIGIR Conference on Research and Development in Information Retrieval in the Asia Pacific Region, SIGIR-AP 2024*, page 186–196, New York, NY, USA. Association for Computing Machinery.

Patrick Sturgis, Thomas Robinson, Laura Fung, and Caroline Roberts. 2025. [SOCbot: Using Large Language Models to measure and classify occupations in surveys.](#)

Robert S Weiss. 1995. *Learning from strangers: The art and method of qualitative interview studies.* Simon and Schuster.

Alexander Wuttke, Matthias Aßenmacher, Christopher Klamm, Max M. Lang, Quirin Würschinger, and Frauke Kreuter. 2024. [AI Conversational Interviewing: Transforming Surveys with LLMs as Adaptive Interviewers.](#) *arXiv preprint*. Issue: arXiv:2410.01824 arXiv: 2410.01824 [cs].

Ziang Xiao, Michelle X Zhou, Wenxi Chen, Huahai Yang, and Changyan Chi. 2020a. If I hear you correctly: Building and evaluating interview chatbots with active listening skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14.

Ziang Xiao, Michelle X. Zhou, Q. Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020b. [Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-ended Questions.](#) *ACM Transactions on Computer-Human Interaction*, 27(3):1–37.