

ISPOR Europe 2025 • Glasgow, Scotland • 9-12 November 2025

Evaluating AI Chatbot Empathy and Probing Skills in Automated Patient In-Trial Interviews: A Proof-of-Principle Experiment

Bill Byrom, PhD¹ & Tina Byrom, PhD²
1: Signant Health, Nottingham, UK 2: Loughborough University, Loughborough, UK

Background

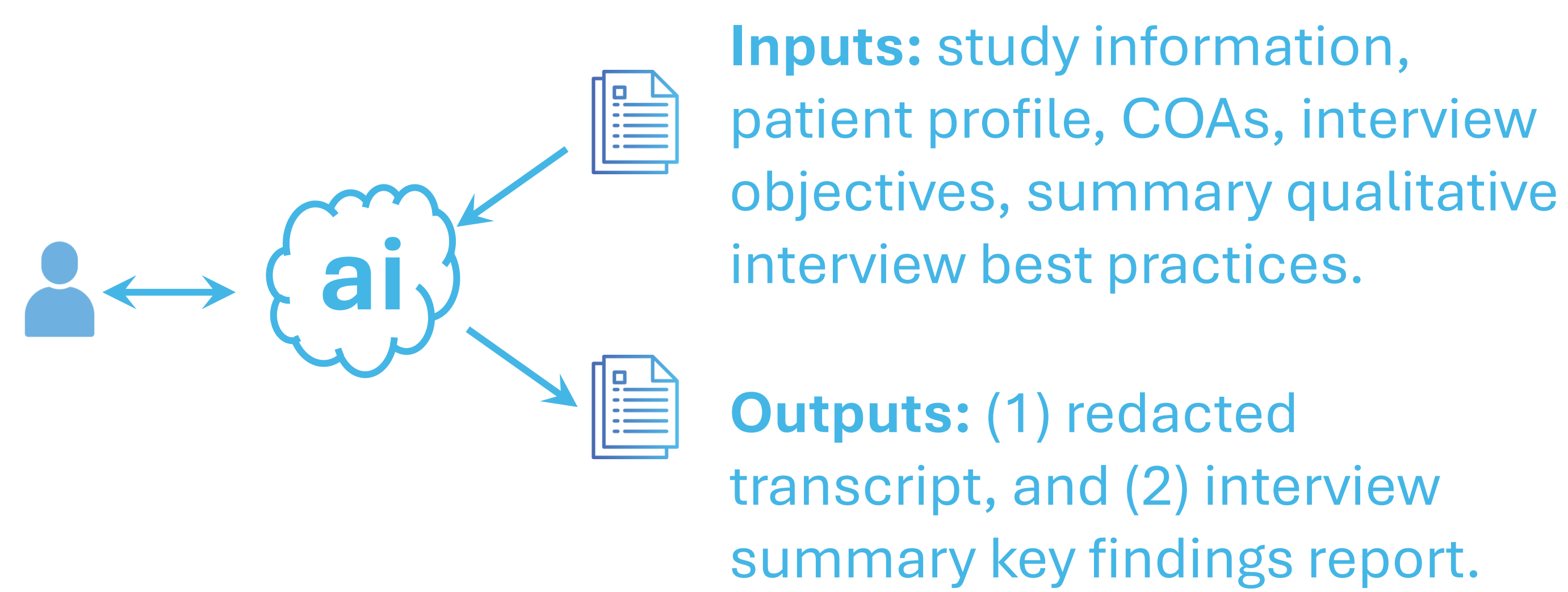
In-trial interviews are emerging as a valuable tool in clinical trials to capture patient experiences and perspectives throughout drug development [1,2]. They are typically conducted as 60-minute one-on-one telephone interviews by trained, independent qualitative interviewers. For practical purposes, in-trial interviews are normally conducted on a sub-sample of study participants. Resulting interview transcripts are redacted and synthesized using qualitative data analysis (e.g., thematic analysis) [3].

Aims

To evaluate the feasibility of using an AI chatbot to conduct qualitative in-trial interviews with clinical trial participants, according to qualitative interview scientific best practices.

Methods

A proof-of-principle experiment was conducted using Claude 3.5 Sonnet (Anthropic, October 2024) to perform an exit interview exploring study participation experience and COA measures by role play with a mock participant from a hypothetical NSCLC trial. The figure below illustrates the inputs provided to the AI model, and the requested output.



Results

We rated interview conduct with reference to qualitative interview best practices using a ‘green-amber-red’ indicator scale (green: acceptable quality; amber: minor limitations; red: major failings). In general, the chatbot delivered a relatively successful qualitative interview, with all aspects assessed rated either green or amber (See Table).

References

[1] Basch E. Toward patient-centered drug development in oncology. N Engl J Med. 2013; 369: 397–400. [2] FDA . Patient-focused drug development: methods to identify what is important to patients. 2022. [3] Braun V., Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006; 3: 77–101.

Interview element	Evaluation	Overall readiness
Providing opening context and obtaining consent	- Followed the interview guide to provide an appropriate introduction and explain purpose, use of data, anonymity, recording of conversation. - Requested consent to proceed.	✓
Establishing rapport	Friendly tone.	✓
Active listening	- Paraphrasing failed to test understanding using a follow up question (e.g., <i>"Am I understanding that correctly?"</i>). - Provided context-related empathetic remarks. - Used follow-up questioning on occasion.	●
Empathy	- Provided context-related empathetic remarks. - Some unqualified assumptions about patient feelings in empathetic responses (e.g., <i>"That sounds frustrating..."</i>).	●
Probing appropriately	- Used follow up probing questions to understand detail. - Sometimes combined multiple probes into a single question.	●
Understanding colloquialisms	Understood colloquialisms (e.g., <i>"eat like a horse"</i>).	✓
Understanding study context	- Identified PRO items stated as missing from one measure were in fact included in a different measure used in the study. - Understood the content of the PROMs used.	✓
Closing the interview	- Asked participant whether they had anything else to share, but failed to check if there were further items before closing. - Thanked participant appropriately.	●
Following the interview guide	- The chatbot covered all the topics in the interview guide. - Some questions contained multiple questions which should have been asked separately. - More depth could have been explored in some topic areas.	●

✓ Acceptable quality

● Shows promise but improvements through further training / refinement needed

● Major failing / limitation

The chatbot was able to provide a friendly welcome, use probing questions to explore symptom impact (e.g., *"Could you tell me more about how the shortness of breath affected you in your daily life?"*), and show empathy during conversation (e.g., *"That sounds like quite a journey for each visit..."*). Greater consistency/improvement areas included: (a) timing probing questions to not over-direct conversation, (b) making unqualified assumptions about patient feelings, and (c) asking consecutive questions / probes together.

Conclusion

AI chatbot technology has the potential to efficiently administer qualitative in-trial interviews, at a scale not practical with human interviewers. Further work is needed refining the approach, evaluating in multilingual settings, and extending to voice and associated transcription.