

Talk to the bot: A scoping review on using AI-powered transcription tools in qualitative research

Ying Jun Lim^{1,2} MD, Ngiam Chuan Tan^{2,3} FAMS

Dear Editor,

Untangling complex social phenomena for health-related problems relies on qualitative research methods like one-to-one in-depth interviews and focus group discussions.¹ However, transcription of these interviews and discussions is costly, labour-intensive² and prone to errors.¹ Furthermore, manual transcription is translational. The transcriber interprets the interview during transcription, inevitably incorporating their own biases during the act of translation,³ affecting subsequent analyses.

Artificial intelligence (AI)-powered transcription tools are set to replace manual transcription. However, there are doubts about the reliability and validity of the transcription, especially in healthcare research. Thus, to understand the usability of AI transcription tools in real-world research settings, researchers need to assess whether these tools can meet their specific needs. This study conducted a scoping review to summarise the studies that have evaluated AI transcription tools for qualitative interviews, and created a guide to help healthcare researchers choose a suitable tool.

A literature search conducted on 2 May 2025 on PubMed, JSTOR and IEEE, generated 6, 83 and 790 hits, respectively. The following search terms were used: ("text analytics" or "text analysis" or "speech analysis" or "speech to text" or "automated speech recognition" or "speech processing" or "visual analytics" or "video analysis") AND ("research interview" or "qualitative interview" or "qualitative research" or "qualitative data") in all fields. Ten additional articles identified via a Google Scholar search were also included. A total of 886 articles were identified after removing duplicates. After excluding articles for reasons such as the study not involving transcription, qualitative research interviews, or being descriptive technical papers, 9 comparative analysis studies were identified.

These 9 articles used 11 different automatic speech recognition tools. Seven studies were published from 2024 to 2025, 1 in 2017 and 1 in 2018. Three studies

transcribed interviews in multiple languages,^{1,4,5} 3 involved interviews conducted in English,^{2,3,6} and the rest were conducted in German,⁷ Indonesian⁸ and Swedish.⁹ Majority involved one-on-one interviews, with 1 study involving both one-on-one and group interviews.²

AI transcription tools were relatively accurate in transcribing interviews, with 8 out of 9 studies supporting its use. The accuracy varied with language, with English interviews unsurprisingly performing best.^{1,5} Audio conditions (e.g. minimal background noise, no crosstalk)^{1,2} and speech (e.g. not heavily accented, slow tempo)³ also improved accuracy. Some researchers found that these errors did not have a significant impact on meaning, and were easily rectified.^{2,6}

Conducting qualitative research in the healthcare setting presents an additional challenge. Medical jargon, colloquial language and non-verbal or tonal cues limit the accuracy of these tools.^{1,2,7,9} More dangerously, biases inherent to the training methods of AI tools reinforce stereotypes. Bircan and Ceylan observed errors (e.g. substitution, insertion errors) that resulted in biases, which reinforced Orientalist and Islamophobic stereotypes.⁴ For instance, "staying in Arab countries is difficult with no visa" was erroneously transcribed to "staying in Arab countries is difficult with all these conflicts."

AI transcription tools can lead to considerable cost and time savings. Eftekhari³ saved GBP3772 from allocated transcribing costs by using Microsoft Teams (Microsoft, Redmond, WA, US), a free tool. KALDI (<http://github.com/kaldi-asr/kaldi>) transcribed at 1.1 times the recording duration¹ Microsoft Teams almost instantaneously³ and Google's Speech-to-Text about 7 times faster than humans.⁷

However, these time savings were, in turn, spent correcting the transcripts manually. Tolle et al. reported correction time ranging from 1.7-fold for Portuguese to 16-fold for Tamil,¹ while Eftekhari required 1.5–3.5 hours correcting interviews averaging 65 minutes.³

AI transcription tools were generally easy to use, with positive feedback. Participants, despite most

The Annals is an open access journal, allowing non-commercial use under CC BY-NC-SA 4.0.

¹ SingHealth Family Medicine Residency Programme, Singapore

² Department of Research, SingHealth Polyclinics, Singapore

³ Duke NUS Medical School, Singapore

Correspondence: Dr Ying Jun Lim, Department of Research, SingHealth Polyclinics, 167 Jalan Bukit Merah, Connection One, Tower 5, #15-10 Singapore 150167. Email address: yingjun.lim@mohh.com.sg

Accepted: 4 February 2026

Published Online First: 23 February 2026

having no or limited experience with these tools, could independently install the software and use the tools.^{1,8} Tolle et al.'s participants evaluated the tool highly, scoring it 4.42 (standard deviation [SD] 0.90) on perceived usefulness and 4.50 (SD 0.80) on likeliness of future use via 5-point Likert scales.¹

Healthcare research contains especially sensitive data. Researchers can maintain privacy by ensuring that data are kept locally.¹ Cloud- and website-based tools, such as Microsoft Teams and Google's Speech-to-Text, require data upload to external servers, introducing security risks.^{2,5} Researchers should also evaluate vendors' security protocols and compliance with privacy regulations such as the Health Insurance Portability and Accountability Act, and choose vendors that have the necessary approvals and passed relevant security reviews.⁵

Researchers need to prioritise the features of different AI transcription tools to align with the research team's needs. For instance, in dealing with sensitive healthcare data, an AI transcription tool with local data storage offers better security, but requires additional hardware and programming expertise. Paid tools generally offer a more seamless experience and better integration with downstream data analysis software,

but can be costly, cloud-based and therefore susceptible to security breaches. Cloud-based tools should not be considered when working with highly sensitive data, as they require uploading of recordings and may monetise data, compromising participant anonymity.^{10,11} Clinical researchers can consider free or low-cost options, such as aTrain,¹² NoScribe (<https://github.com/kaixxx/noScribe?tab=readme-ov-file>) and MacWhisper (<https://goodsnooze.gumroad.com/l/macwhisper>).

Transcription of interviews that feature technical or colloquial language, strong accents or controversial topics tend to be prone to inaccuracies, leading to higher word error rates (WER). Correcting a transcript with WER >20% can be time-consuming, requiring up to 4-fold the recording time.⁹ Urban and Mehrotra suggested that the research team should consider additional training and customisation of the tool if WER >30%, though not all teams have such resources.¹³

Choosing an AI transcription tool can be a complicated process, especially for the uninitiated. Post-marketing assessments of commercial AI tools are inherently biased, and researchers should aim to assess tools independently. This study proposes a

Table 1. Main considerations in selecting an AI transcription tool for interviews in qualitative research.

Criteria	Considerations	Key queries in selecting an AI transcription tool
Accuracy and reliability	Error rate	What is the language-specific word error rate of the tool? Can it recognise technical or field-specific terminology with ease?
	Speaker differentiation	Can the tool recognise and tag multiple speakers accurately, especially in group settings? Are the strategies in place to improve speaker attribution (e.g. editing speaker attribution during live transcription to improve accuracy)?
	Languages, dialects, accents and speech patterns	How many languages/dialects can it support? Can it recognise multilingual conversations with ease? Can it recognise accented speech or different speech patterns with ease?
	Challenging speech conditions	Can the tool effectively handle interruptions, overlapping speech, unclear speech, background noise?
	Complex speech	Can the tool handle complex and layered dialogue? Is the tool able to identify underlying emotional tones or sentiment in the speech, and code them accordingly (e.g. sentiment analysis)?
	Biases	Is the transcription accuracy affected by systemic biases (e.g. gender, racial bias), which can result in lower transcription accuracy in these groups?
Ease of use	Ready to use	Is the tool ready to use, or is further programming needed?
	Hardware requirements	Is a separate microphone needed? Does the platform run locally or on the cloud? Is internet or intranet required? Does the platform run on mobile phone, tablet, laptop and/or desktop? Does the platform have high hardware requirements (e.g. RAM, memory space)?
	Collaboration	Is there a way for users on the platform to collaborate on the same project? Can varying degrees of rights be assigned to the different collaborators?

Table 1. Main considerations in selecting an AI transcription tool for interviews in qualitative research. (Cont'd)

Criteria	Considerations	Key queries in selecting an AI transcription tool
	Storage	Will the outputs be saved on the platform? How long will the outputs be available on the platform?
	User interface	Is the interface intuitive? Is complex training required? Are tutorials readily available? Is there an easy way to get technical assistance if problems are encountered?
Inputs and outputs	Inputs supported	Can it support live transcription? Is there an input limit for live transcription? Can it support uploading of audio and video files? Is there a limit to the size of audio or video file that can be uploaded? Is there a limit as to how many files can be uploaded?
	Outputs supported	What are the output formats supported (text, audio)? Can the outputs be exported with ease? Can the outputs be integrated seamlessly with other qualitative research software (e.g. NVivo [Lumivero, Denver, CO, US]) or is further processing needed?
Customisability	Transcription adjustments	Can the transcript be edited directly on the platform? If live transcription is possible, can editing be done during live transcription? Is there a way to easily review audio sections of the corresponding transcription (e.g. time stamping, replaying the audio when corresponding transcription is clicked)? If the tool does not allow for editing, can the output be edited easily using another software?
	Custom vocabulary	Can the tool be trained to recognise technical terms or jargon to improve transcription accuracy?
Time efficiency	Turnaround time	How long does it take to perform live transcription? How long does it take to transcribe an uploaded audio or video file? How long does it take to edit and correct the transcript?
	Batch processing	Can the tool handle large volumes of audio or video files simultaneously? How long will it take?
Cost effectiveness	Pricing structure	Is the tool free to use? How much does it cost per word/minute of transcription? Is the pricing on a subscription basis? Is the pricing transparent? Are there any hidden costs? Are there discounts for academic use? Is there a free trial available?
	Additional costs	Are there any additional costs incurred (e.g. purchasing hardware, editing transcript, obtaining relevant approvals)?
Data security	Data privacy	Is the tool local and does not require cloud services? Does the tool harness your data for large language model training, or reserve the right to use it for future purposes? What are the security measures in place to safeguard data? Has the platform obtained necessary approvals and passed relevant security reviews?
	Encryption	Does the service use encryption to protect data during transmission and storage?

framework of major considerations in selecting an AI transcription tool in Table 1.

Researchers should consider post-transcription strategies to “scrub” the AI-generated transcript. Some studies promote using AI transcription tools as a “good enough” first step, with manual cleaning for accuracy.^{1,2} Others suggest using another AI tool, such as ChatGPT, to clean the transcript, though

with caution.⁹ Researchers may opt for manual cleaning for recordings that are challenging for AI tools, such as those involving colloquial or technical language, which is usually the case for interviews conducted in Singapore as they often involve Singlish.

In conclusion, AI transcription tools offer significant time and cost savings in qualitative research, but researchers should address challenges such as

transcription quality, biases, security and trade-offs when selecting a tool. The authors hope that researchers can use Table 1 as a guide to selecting the most appropriate tool, but acknowledge that the considerations are not exhaustive, especially given the evolving nature of AI. Stricter guidelines to identify acceptable AI tools in research from regulatory bodies might aid in addressing these challenges.

Ethics statement

Not applicable as there are no study participants.

Declaration

The authors declare there are no affiliations with or involvement in any organisation or entity with any financial interest in the subject matter or materials discussed in this manuscript. This work was supported by the SingHealth Polyclinics-National Medical Research Council (NMRC) Centre Grant (CG21APR3006 [NMRC/CG3/001/2022-SHP]). During the preparation of this work, the authors used ChatGPT to improve readability of the text. After using this tool, the authors reviewed and edited the content and take full responsibility for the content of the published article.

Keywords: artificial intelligence, interview, qualitative research, tool, transcription

REFERENCES

1. Tolle H, Castro MDM, Wachinger J, et al. From voice to ink (Vink): development and assessment of an automated, free-of-charge transcription tool. *BMC Res Notes* 2024;17:95.
2. Bokhove C, Downey C. Automated generation of 'good enough' transcripts as a first step to transcription of audio-recorded data. *Methodol Innov* 2018;11:2059799118790743.
3. Eftekhari H. Transcribing in the digital age: qualitative research practice utilizing intelligent speech recognition technology. *Eur J Cardiovasc Nurs* 2024;23:553-60.
4. Bircan T, Ceylan D. Machine Discriminating: Automated Speech Recognition Biases in Refugee Interviews. *J Immigr Refug Stud* 2024;1-16.
5. Meanwell E, Walsh A, Berg P, et al. Automated Transcription Service: Enabling Transcription at Scale for Social Science Research. In: *Practice and Experience in Advanced Research Computing 2024: Human Powered Computing*. Providence RI USA: ACM; 2024.
6. Pfeifer VA, Chilton TD, Grilli MD, et al. How ready is speech-to-text for psychological language research? Evaluating the validity of AI-generated English transcripts for analyzing free-spoken responses in younger and older adults. *Behav Res Methods* 2024;56:7621-31.
7. Höhne JK, Lenzner T, Claassen J. Automatic speech-to-text transcription: evidence from a smartphone survey with voice answers. *Int J Soc Res Methodol* 2025;1-8.
8. Adianto R, Satriawan CH, Lestari DP. Transcriber: An Android application that automates the transcription of interviews in Indonesian. In: *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)*. Denpasar: IEEE; 2017.
9. Helin M. Evaluating the Accuracy of Automated Transcription using Open AI's Whisper Model for Swedish IT-Related Interviews: A Comparative Analysis with Manual Transcription, 2024. <https://urn.kb.se/resolve?urn=urn:nbn:se:du-49741>. Accessed 22 January 2025.
10. Da Silva J. Producing 'good enough' automated transcripts securely: Extending Bokhove and Downey (2018) to address security concerns. *Methodol Innov* 2021;14:2059799120987766.
11. Ong JCL, Chang SYH, William W, et al. Ethical and regulatory challenges of large language models in medicine. *Lancet Digit Health* 2024;6:e428-32.
12. Haberl A, Fleiß J, Kowald D, et al. Take the aTrain. Introducing an Interface for the Accessible Transcription of Interviews. *J Behav Exp Finance* 2024;41:100891.
13. Urban E, Mehrotra N. Test accuracy of a custom speech model, 2024. <https://learn.microsoft.com/en-us/azure/ai-services/speech-service/how-to-custom-speech-evaluate-data?pivot=speech-studio>. Accessed 22 January 2025.