

Review Article

Automated Transcription of Interviews in Qualitative Research Using Artificial Intelligence: A Simple Guide

Zubair M Mojadeddi¹, Jacob Rosenberg^{1*}

¹Center for Perioperative Optimization, Department of Surgery, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark

*Correspondence author: Jacob Rosenberg, MD, DSc, Center for Perioperative Optimization, Department of Surgery, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark; Email: jacob.rosenberg@regionh.dk

Citation: Rosenberg J, et al. Automated Transcription of Interviews in Qualitative Research Using Artificial Intelligence: A Simple Guide. J Surg Res Prac. 2024;5(2):1-6.

<https://doi.org/10.46889/JSRP.2024/5204>

Received Date: 25-05-2024

Accepted Date: 17-06-2024

Published Date: 24-06-2024



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CCBY) license

(<https://creativecommons.org/licenses/by/4.0/>).

Abstract

Qualitative research often involves the transcription of interviews, a traditionally manual and time-consuming task. Recent advancements have introduced AI-driven transcription technologies, aiming to streamline this process. One such technology is OpenAI's Whisper, an automated speech recognition system capable of transcribing audio in multiple languages. This paper introduces Whisper and provides a guide on its utilization for research transcription. In conclusion, automated transcription of interviews in qualitative research using artificial intelligence is now possible with excellent accuracy and user-friendliness. While Whisper presents a promising solution to the transcription challenges in qualitative research, careful usage and data review are essential.

Keywords: Artificial Intelligence; Whisper; GPT Technology; Qualitative Research

Introduction

Qualitative research often involves interviews with patients or relevant parties. These interviews have traditionally been transcribed verbatim manually [1]. The transcription process is time-consuming; it has been reported that one hour of an interview can take from six to seven hours to transcribe for a skilled secretary [2]. In recent years, various transcribing technologies have appeared, such as Microsoft Word's own transcribing feature and other technologies such as Dragon Speech Recognition Solutions, Aiko and Whisper Transcription for iOS products [3,6].

Some transcribing technologies stem from Artificial Intelligence (AI) such as Aiko and whisper Transcription. AI is the ability of a digital computer or a computer-controlled robot to perform tasks commonly associated with intelligent beings [7]. The advancements in AI technologies over the years have been surprising. For instance, OpenAI introduced the Generative Pre-Trained Transformer (GPT) technology, which led to the development of ChatGPT. ChatGPT has shown human-like abilities in conversation and writing. Researchers have been using ChatGPT for different parts of the scientific process such as in the research reporting process, generating cover letters and helping write introduction sections for scientific papers [8-10]. Given the advancement of AI technologies over the years, it is worth suggesting that AI has the potential to alleviate the time-consuming nature of manual transcription in qualitative research. On September 21, 2022, OpenAI released its Whisper model, which is an automated speech recognition system [11]. This means that the Whisper model can transcribe audio in different languages, reportedly in 97 languages and in different accents.

In this study, we are going to provide a simple guide on how to use OpenAI's Whisper technology in transcribing your interviews. This will eventually help the researcher in a time-consuming task. With the help of whisper and other technologies that can transcribe, more focus will be on other important aspects of the research.

What is Whisper?

Whisper operates on a neural network framework known as the Transformer, enabling it to recognize and transcribe multiple languages [12,13]. Whisper is trained on 681,070 hours of multilingual hours of audio data [12]. The training data are diverse and divided in the following way: 117,113 hours of multilingual speech recognition (17%) which covers 96 other languages than English, 125,739 hours of translation (18%) and the remaining 438,218 hours of English speech recognition (65%) [12]. Whisper operates on a variety of models, each based on different parameters. The bigger the parameters, the more data it can work on. Additionally, the ability to recognize words is better the bigger the model. For parameter specification (Table 1).

As of 2024, several transcription tools are available, such as speak.ai, Trint, Beey and fireflies.ai [14-17]. Each of these tools offers multilingual transcription features. Our choice to use Whisper for transcription is based on several factors. Primarily, Whisper transparently conducts its research. Users can access this tool via an Application Programming Interface (API), which is a software bridge, an intermediary that allows programs to communicate and exchange information, much like how a translator helps people who speak different languages understand each other, thereby allowing anyone to use it, provided they possess the coding knowledge and understanding. While many transcription technologies exist, most support only a handful of languages. In contrast, Whisper provides support for a considerably larger number of languages. Lastly, unlike some other alternatives, Whisper doesn't come with any associated costs thus making it accessible to everyone. This combination of features makes it our current tool of choice.

Guide to Using Whisper

Whisper can be accessed through both an online version and an offline version, Fig. 1. Before using the online version of Whisper to transcribe your interviews, it is important to anonymize your audio files or video files. Initially, before conducting interviews, some precautions are needed; inform participants that all personal details will be addressed at the start (so it can easily be deleted afterwards). Furthermore, request the interviewee to refrain from mentioning any identifiable information throughout the interview. The second step is editing your files to remove any personal information if needed. If necessary, this can be done through various video editing tools such as Audacity, Microsoft Movie Maker, Wavepad, iMovie, Oceanaudio and Garageband or any other preferred software programs. It is essential to edit out any personal information to comply with European Union GDPR rules and similar regulations that apply in other parts of the world [18-24]. After editing and complying with the rules and regulations of your home country, the online version of Whisper can be accessed and the transcription can begin. It is worth noting that the offline versions of Whisper do not require editing and data anonymization.

There are three main ways of getting access to Whisper, Fig. 1. The first way is through the coding program Python, but this approach requires coding knowledge and a powerful computer. Nonetheless, a guide for this is available through the OpenAI Github Page [25]. First, you will have to require a configuration on your PC where Python 3.8-3.11 is installed. You may choose any version within this range as it is compatible with all the specified versions. Secondly, PyTorch will need to be installed [26]. Whisper is compatible with all the recent PyTorch versions, the guide doesn't specify the exact versions. Next, you will need to install ffmpeg on your PC [27]. Although this method is available, it is quite time-consuming and potentially difficult, particularly if you are not familiar with coding, Python and software applications. Alternatively, you can find various step-by-step videos that explain how to install Whisper on Python, as well as how to use it.

The second way of using Whisper is through web pages that use the Whisper API, for instance freesubtitles.ai [28]. On this page, two different options are available, the first one is to use the free version for videos or audio with a maximum length of 1 hour; this uses the medium Whisper model (Table 1). The second option is one where you purchase access to the large model, which is a more accurate Whisper model for videos or audio with a maximum length of 10 hours. Both models can transcribe nearly 1 hour of audio or video in approximately 1-5 minutes greatly reducing transcription time for researchers. However, the quality may vary according to model and language, where English is best. After uploading the interview, you receive three files containing the transcribed version of your interview with different features: SubRip subtitle (SRT) file which is a file that contains the text with time codes, Video Text Tracks (VTT) file containing the text with time codes and lastly Text (TXT) file containing the text without time codes.

The last approach on how to access Whisper is to use applications such as Aiko and Whisper Transcription. Both Aiko and Whisper Transcription use Whisper's API to get access to its data and thereby transcribe audio. One advantage of these apps is

that they operate offline. After downloading the app, you download a data package that enables the app to transcribe without communicating with a server. Because all of this is done offline, it solves the problem of GDPR and other data protection rules as long as data are transferred and stored according to regulations.

Regardless of your chosen method, we recommend that after transcribing you listen to your interviews and correct any transcription errors line by line, especially for non-English content where there may be more errors than for English content. Furthermore, identify the interviewer and interviewee and separate the text line by line. You will then have a fully transcribed audio or video file to use for qualitative analysis.

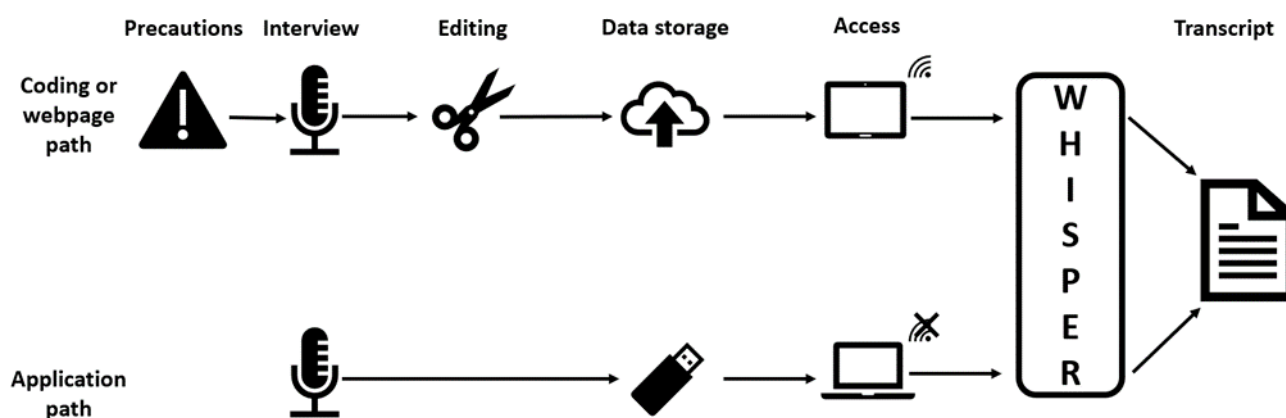


Figure 1: Accessing Whisper. Interview: Conducting interviews. Precautions: Precautions before conducting the interview. Editing: Editing files for anonymization. Data storage: This can be done in the cloud or on encrypted USB. Access: Online version or offline version. Transcript: Fully transcribed interview.

• Model Tiny: based on 39M parameters
• Model Base: 74M parameters
• Model Small: 244M parameters
• Medium: 769M Parameters
• Model large: 1550M parameters
• Model Large V2 1550M parameters. But more efficient in translation and language recognition.

Table 1: Whisper parameter details. M: Million.

Discussion

Traditionally, manual transcription of interviews has been the norm, which is a time-consuming task. However, with the advancement of AI technology, different AI solutions reduce transcription time to a few minutes with very few errors. Thus, AI-powered transcription automates the process, significantly reducing time and effort. Traditionally, the transcription of one hour of audio demands approximately six to seven hours of manual labor and with AI this time-consuming task can be reduced to minutes [2]. This efficiency significantly opens doors for more researchers to engage in qualitative research [29]. However, the advantages of AI in research extend beyond time savings. It redirects the focus of researchers toward the analysis process and other crucial aspects of their research. For years, researchers have used research assistants' help for data collection and statisticians for complex analysis. Similarly, AI technology should not be viewed as a replacement for researchers but as a tool that facilitates their work [30-32]. The use of AI for transcription is replacing some secretarial tasks, but it does not diminish the intellectual efforts required by the researcher. The intellectual effort to interpret the data and derive meaningful conclusions still rests with the researcher. Therefore, there is no question of compromising the intellectual contribution of the researcher.

Another proposed approach to analyzing qualitative interview data employs a method where notes or analyses are made without transcribing the entire interview, making it less time-consuming than traditional manual transcription [33,34]. Our method aligns with this approach but is even more efficient. With our method, audiotapes are already transcribed, facilitating easier analysis. While non-transcription methods often require researchers to repeatedly listen to audio recordings and review their notes, this can still be time-consuming. Therefore, our method offers a good balance for analyzing qualitative research.

Even though it can greatly help researchers, limitations exist. Transcription is commonly categorized into two forms: naturalized and denaturalized. Naturalized transcription involves adapting oral speech to written norms, while denaturalized transcription includes everything, including utterances, mistakes, repetitions and grammatical errors [35]. Whisper appears to use the denaturalized transcription verbatim approach. Furthermore, there is a potential for text prediction bias in Whisper, where the model can incorrectly predict words and transcribe the wrong words. This can be avoided by having the researcher listen to the transcriptions first and then read the text it generates, allowing them to customize it according to their own preferences. Furthermore, another significant limitation to consider is the potential risk of uncontrolled data-sharing and anonymity breach of the interviewee. While many organizations and online platforms, including those dealing with AI, claim to delete datasets, the complete and permanent erasure of data cannot always be guaranteed. As such, it is critical for researchers to exercise caution when uploading audio files that could potentially identify interviewees. OpenAI is dedicated to ensuring the safety of AI and they uphold data usage policies. They assert that data aren't used for commercial purposes such as selling services, advertising or building individual profiles. Instead, they claim to use data to refine their AI models, making them more beneficial to users. Furthermore, OpenAI affirms its commitment to privacy by working to remove any personal data from individuals present in their training datasets. They also fine-tune their models to deny requests for personal information, reinforcing their commitment to safeguarding user privacy. A big spectrum of policies has been taken by OpenAI for their APIs [36]. Additionally, OpenAI states that they will not use API data to train or improve their models unless this is explicitly shared with them for that purpose [37,28]. Lastly, any data sent through API will be retained for abuse and misuse monitoring purposes for 30 days and deleted afterward [36]. It seems, however, that all these issues may be solved by the use of one of the available offline systems and we will most probably see more of these in the near future.

Conclusion

AI technologies like Whisper can now help overcome the time-consuming task of transcribing interviews in qualitative research. We've provided guidance on how users can access the Whisper model, whether they prefer using Python or by easy access through the Internet or already released apps that use Whisper API. While Whisper possesses significant capabilities that can assist researchers in transcription, it's crucial to be mindful of data anonymization when using the online versions of Whisper. This ensures we don't infringe on patient rights or violate the rules and regulations of the countries where research is conducted. We therefore suggest anonymizing the audio tapes before online transcription. The offline versions, however, would not need data anonymization. Once transcription is complete, we recommend carefully listening to the audio tape while reviewing and correcting the transcribed interview for maximum accuracy. This approach can save researchers valuable time, enabling them to give more attention to other critical aspects of their work such as data analysis and interpretation.

Conflict of Interests

The authors have no conflict of interest to declare.

Trial Registration and Ethical Approval

Not relevant

Funding Declaration

No funding was used in this study.

Acknowledgements

The full text was written by the authors, while the abstract was generated using OpenAI's GPT-4 model. We would also like to thank Siv Fonnes for her assistance in creating Figure 1.

Authors' Contribution

Study concept and design: ZM, JR.

Data analysis and interpretation: ZM, JR.

Drafting the article: ZM, JR.

Critical revision: ZM, JR.

Final approval: ZM, JR.

Patient Consent Statement

Not relevant.

Permission to Reproduce Material from Other Sources

Not relevant.

References

1. ten Have P. Transcribing talk in interaction. Doing conversation analysis. A practical guide. London: SAGE Publications, Ltd. 1999:75-98.
2. Britten N. Qualitative interviews in medical research. BMJ. 1995;311:251.
3. Microsoft. Dictate your documents in Word. [Last accessed on June 17, 2024]
<https://support.microsoft.com/en-us/office/dictate-your-documents-in-word-3876e05f-3fcc-418f-b8ab-db7ce0d11d3c>
4. Nuance. Dragon. 2024. [Last accessed on June 17, 2024]
<https://www.nuance.com/dragon.html>
5. Apple. Aiko. [Last accessed on June 17, 2024]
<https://apps.apple.com/us/app/aiko/id1672085276>
6. Apple. Whisper Transcription. [Last accessed on June 17, 2024]
<https://apps.apple.com/us/app/whisper-transcription/id1668083311?mt=12>
7. Encyclopedia Britannica. Artificial intelligence. [Last accessed on June 17, 2024]
<https://www.britannica.com/technology/artificial-intelligence>
8. Mojadeddi ZM, Rosenberg J. Artificial intelligence can help in the research process. Ugeskr Læger. 2024.
9. Deveci CD, Baker JJ, Sikander B, Rosenberg J. A comparison of cover letters written by ChatGPT-4 or humans. Dan Med J. 2023;70:A06230412.
10. Sikander B, Baker JJ, Deveci CD, Lund L, Rosenberg J. ChatGPT-4 and human researchers are equal in writing scientific introduction sections: a blinded, randomized, non-inferiority-controlled study. Cureus. 2023;15:e49019.
11. OpenAI. Whisper. [Last accessed on June 17, 2024]
<https://openai.com/research/whisper>
12. Radford A, Kim JW, Xu T. Robust speech recognition via large-scale weak supervision. arXiv:2212.04356v1.
13. Vaswani A, Shazeer N, Parmar N. Attention is all you need. arXiv:1706.03762v7.
14. SpeakAI. Automated transcription. [Last accessed on June 17, 2024]
<https://speakai.co/automated-transcription/>
15. Trint. Plans. [Last accessed on June 17, 2024]
<https://app.trint.com/plans?tid=%24device%3A18857b1481a480-0170a8708d95c1-7b515477-232800-18857b1481a480>
16. Beey. Price List. [Last accessed on June 17, 2024]
<https://www.beey.io/en/price-list/>
17. Fireflies. Pricing. [Last accessed on June 17, 2024]
<https://fireflies.ai/pricing>
18. General Data Protection Regulation (GDPR) Info. [Last accessed on June 17, 2024]
<https://gdpr-info.eu/>
19. The Personal Information Protection and Electronic Documents Act (PIPEDA) - Office of the privacy commissioner of Canada. [Last accessed on June 17, 2024]
<https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/>
20. The Privacy Act - Office of the Australian information commissioner. [Last accessed on June 17, 2024]
<https://www.oaic.gov.au/privacy/privacy-legislation/the-privacy-act>
21. California Consumer Privacy Act (CCPA) - California office of the attorney general. [Last accessed on June 17, 2024]
<https://oag.ca.gov/privacy/ccpa>
22. Personal Information Protection Law (PIPL) - Personal information protection law. [Last accessed on June 17, 2024]
<https://doi.org/10.46889/JSRP.2024/5204>

- <https://personalinformationprotectionlaw.com/>
23. Personal Data Protection Bill (PDPB) - Ministry of electronics and information technology, government of India. [Last accessed on June 17, 2024]
 24. <https://www.meity.gov.in/data-protection-framework>
 25. Consumer Data Protection Act (CDPA) - Virginia general assembly. [Last accessed on June 17, 2024]
<https://law.lis.virginia.gov/vacodefull/title59.1/chapter53/>
 26. OpenAI. Whisper GitHub. [Last accessed on June 17, 2024]
<https://github.com/openai/whisper>
 27. PyTorch. [Last accessed on June 17, 2024]
<https://pytorch.org/>
 28. Ffmpeg. [Last accessed on June 17, 2024]
<https://ffmpeg.org/>
 29. FreeSubtitles.ai. [Last accessed on June 17, 2024]
<https://freesubtitles.ai/>
 30. Bokhove C, Downey C. Automated generation of “good enough” transcripts as a first step to transcription of audio-recorded data. *Method Innov.* 2018;11.
 31. Altman DG, Goodman SN, Schroter S. How statistical expertise is used in medical research. *JAMA.* 2002;287:2817-20.
 32. Masuadi E, Mohamud M, Almutairi M. Trends in the usage of statistical software and their associated study designs in health sciences research: a bibliometric analysis. *Cureus.* 2021;13:e12639.
 33. Mojadeddi ZM, Rosenberg J. The impact of AI and ChatGPT on research reporting. *N Z Med J.* 2023;136:60-4.
 34. Greenwood M, Kendrick T, Davies H. Hearing voices: comparing two methods for analysis of focus group data. *Appl Nurs Res.* 2017;35:90-3.
 35. Halcomb EJ, Davidson PM. Is verbatim transcription of interview data always necessary? *Appl Nurs Res.* 2006;19:38-42.
 36. McMullin C. Transcription and qualitative methods: implications for third sector research. *Voluntas.* 2023;34:140-53.
 37. OpenAI. API data usage policies. [Last accessed on June 17, 2024]
<https://openai.com/policies/api-data-usage-policies>
 38. OpenAI. Usage policies. [Last accessed on June 17, 2024]
<https://openai.com/policies/usage-policies>

Journal of Surgery Research and Practice



Publish your work in this journal

Journal of Surgery Research and Practice is an international, peer-reviewed, open access journal publishing original research, reports, editorials, reviews and commentaries. All aspects of surgery health maintenance, preventative measures and disease treatment interventions are addressed within the journal. Medical surgeons and other researchers are invited to submit their work in the journal. The manuscript submission system is online and journal follows a fair peer-review practices.

Submit your manuscript here: <https://athenaeumpub.com/submit-manuscript/>