The Use of Automatic AI-based Notes and Transcription Services in Qualitative Research: Ethical and Methodological Concerns

Anisah Herdiyanti[a,b]

[a]School of Information, Florida State University, USA

[b]Department of Information Systems, Institut Teknologi Sepuluh Nopember, Indonesia

aherdiyanti@fsu.edu

ABSTRACT

The rise of online interviewing using videoconferencing platforms has impacted the way qualitative researchers conduct their data collection. Previous studies have investigated the use of Zoom Meetings for online data collection (Archibald et al., 2019; Lobe et al., 2020), yet other studies also have outlined ethical challenges when collecting data remotely (Bamdad et al., 2022; Lim & Kaveri, 2024). In LIS studies, researchers have examined ethical considerations for qualitative research methods in a virtual setting (Newman et al., 2021), although the discussion surrounding the advancement of AI-based technology in online data collection has been limited. With the growing interest in the use of transcription features in Zoom Meetings to support data analysis (Handler et al., 2022), issues surrounding automatic AI-based notes and transcription services integrated with this platform may have been overlooked. This study reflected ethical and methodological concerns when collecting data on student mothers' use of mobile devices via Zoom Meetings synced with Otter.ai services. Ten participants were involved in online interviews from February to May 2023, resulting in nine transcripts from Zoom Meetings and Otter.ai services each. Ethical and methodological concerns surrounding AI-based meeting assistants' use were outlined, including data ownership and intellectual property management, data repository and flow strategies, real-time nature of interactions, implicit speech bias, and decision-making accountability. The study contributes to broadening the understanding of the ethical and methodological concerns for qualitative researchers when using AI meeting assistants and further calls for specific LIS guidelines in online research using AI technology.

ALISE RESEARCH TAXONOMY TOPICS

artificial intelligence; information ethics; research methods.

AUTHOR KEYWORDS

ethical; qualitative; AI-based assistants; Zoom; Otter.ai.

DOI: https://doi.org/10.21900/j.alise.2024.1717

INTRODUCTION

There has been a major shift in the way researchers interact with participants when collecting data remotely. From phone interviews to computer-based interviews, previous researchers have experimented with using different types of information and communication technologies (ICTs) to enable qualitative data collection strategies. For instance, Timbrook et al. (2018) compared the use of landline and cell phones for telephone interviews, Menzies et al. (2020) employed online chat-based interviews with screen reader users, de Villiers et al. (2022) discussed challenges and opportunities when conducting video-based interviews, and Żadkowska et al. (2022) conducted in-depth interviews with mothers using videoconferencing platforms during two first waves of the COVID-19 pandemic.

Previous research in library and information studies (LIS) referred to ICTs' integration into interview methods as technology-mediated interviewing (Kazmer & Xie, 2008) or online interviews (Oates et al., 2022). Online asynchronous interviews, such as via email, tend to generate electronic transcripts that are better thought-out than online synchronous interviews, such as via instant messaging (IM). Nowadays, researchers conduct online interviews using videoconferencing platforms, which enable them to manage interview schedules, facilitate online synchronous interviews, and produce born-digital data in the form of electronic recordings and transcripts. Lynn Silipigni and Marie (2021) pointed out that LIS researchers have used videoconferencing platforms, such as Skype, Teams and Zoom Meetings, for data collection.

With the advancement of artificial intelligence (AI) technology in videoconferencing platforms, automatic notes and transcriptions have undergone transformations in providing real-time captioning with remarkable accuracy and efficiency. Gray et al. (2020) explored the use of Zoom Meetings for online interviews and pointed out this approach's strengths, including ease of use, multi-platform accessibility, and easier rapport building. Seilkhanova et al. (2024) shared how their research team used an online platform to send gift cards (tremendous.com) to participants after the completion of the online interview and, therefore, integrate the Zoom Meeting's scheduling information. Corrente and Bourgeault (2022) discussed the use of Otter.ai services integrated with the Zoom Meetings platform for automatic notes and transcription in qualitative research.

Meanwhile, conducting ethically sound internet-based research has raised concerns among researchers who used online data collection in their studies. Bamdad et al. (2022) presented grey areas when conducting online data collection, especially in cross-border studies, including distrust in collaboration and data flow ambiguities. Lim and Kaveri (2024) reflected on their experience in online interviews with children using Zoom Meetings and shared how they dealt with parent involvement in acting as an agency during the interview process with their children. The issue of transparency, data bias, data privacy and data ownership is in line with the discussion of AI ethical considerations in interpreting (Horváth, 2022). However, there is still limited discussion in LIS studies on ethical and methodological concerns when using automatic AI-based notes and transcription services in online interviews.

This ongoing study will reflect on the use of two automatic transcription services in online interviews with ten participants in a study to explore student mothers' use of mobile devices. The discussion will look back on the ethical and methodological concerns when collecting data via Zoom Meetings synced with Otter.ai services in online interviews between February and May 2023. This article will present the ethical and methodological concerns by recollecting perspectives and experiences from the researcher's perspective when using the caption-enabled Zoom Meeting platform integrated with Otter.ai notes and transcription services during online interviews.

AI-BASED NOTES AND TRANSCRIPTION SERVICES

Prior to the establishment of the Zoom AI companion feature, the Zoom Meeting platform offered closed captioning features that can be activated during the meeting session. If the meeting is recorded, then not only will the users get the video and audio recordings of the session but also the time-stamped transcriptions from the session (McCarron, 2021). The Otter.ai platform offered notes and transcription services during a meeting, including recording of the meetings (Oates et al., 2022). Both the Zoom Meeting and Otter.ai platforms can be included in the category of automatic notes and transcriptions using AI-based technology. Automatic notes and transcription services are valuable features in internet-based research that require data to be collected remotely (Keen et al., 2022), and AI-based technology is the key to future videoconferencing platforms (Suduc & Bizoi, 2022).

Meanwhile, previous studies have discussed ethical concerns when conducting online data collection. Bamdad et al. (2022) discussed challenges in cross-border online data collection in social science research, discussing two ethical concerns: (a) data ownership and intellectual property management and (b) distrust in collaboration. The first ethical concerns looked at three levels of perspectives, which are participant's, institutional, and country levels. The perspective of participants is mainly concerned with ensuring personal data safety and security following the UK General Data Protection Regulation (GDPR), especially when EU citizens' data are being transferred and processed outside the EU. The institutional perspective is concerned with ensuring that the intellectual property policies of the researchers' employers are being followed, indicating a fair collaboration across institutions. The country's perspective is concerned with whether or not there are any requirements to recruit a local researcher as a principal or a co-

investigator of the research project when the study involves online data collection. Furthermore, the second ethical concern about distrust in collaboration could occur when a research team is formed with a diverse range of people in terms of ethnic background and identities and where the team members are distanced due to different geographical locations.

Other studies highlighted implicit speech bias (Goldenthal et al., 2021) and decision-making accountability among ethical concerns in conducting internet-based research using AI-based technology. Goldenthal et al. (2021) shared responses from non-native and native English speakers alike, who suggested that research participants may be willing to use AI-mediated communication tools (e.g., voice assistants) if these tools offer services in the participants' native languages and are capable of understanding various accents. As pointed out by Horváth (2022), systems using AI-based technology "are gaining in autonomy and are increasingly taking decisions" (p. 5). In a previous study, Long and Magerko (2020) argued that giving decision-making power to any system using AI-based technology can result in ethical dilemmas, such as trolley problems of whether to sacrifice one person in exchange for saving a larger population and unexpected results due to misinterpretation between the actual instructions and the purpose of doing the instruction, e.g., self-driving car driving above the speed levels because the instruction was getting to a specific destination as fast as possible.

Bamdad et al. (2022) further discussed methodological concerns regarding data repository and flow strategies where researchers need to review their research data management plan in internet-based research. For instance, considering whether or not collecting data via online surveys in another country and storing the results in the country where the researcher's team institution is located, sounds like a good data management plan. Lim and Kaveri (2024) conducted an online data collection with children via Zoom Meetings and highlighted the need for the researchers who collected data to observe the real-time nature of interactions with participants, such as the children's body language throughout the data collection process, as the participants may withdraw their consent at any point of time. They provided an example where they observed one participant reading a book during a group discussion and checked whether he would rather read than participate in the discussion, and withdrew his consent when he answered yes. Another example was provided when they exercised researcher *withitness*, referring to the skill of knowing what is happening in the participant's environment, such as managing distractions of toys at the participants' home during the online data collection and being able to redirect children's attention to the data collection process.

RESEARCH METHODOLOGY

This ongoing study reflected on the perspectives and experiences of the researcher when collecting data using online interviews with student mothers on their use of mobile devices. While the results from exploring student mothers' use of mobile devices will not be the focus of this article's discussion, ethical and methodological concerns from doing online synchronous interviews with automatic AI-based notes and transcription services will be presented. Two notes and transcription services were employed during the online interviews, i.e., (a) Zoom transcription services – which were activated through closed captioning features (which were the

only feature available in Zoom for transcriptions services before the birth of Zoom AI companion) and (b) Otter.ai services – which were activated through the integration of email accounts used for the Zoom Meetings and for using the Otter.ai services. The Otter.ai platform allows users to perform this integration by including the Otter.ai notetaker as a user in any Zoom Meetings, and therefore, the notetaker can automatically record and transcribe interviews when the Zoom Meetings take place.

This ongoing study will use ethical and methodological concerns identified in previous studies involving internet-based research as a reference to the discussion on ethical and methodological concerns regarding the use of AI-based notes and transcription services in online interviews. Ethical concerns identified from previous studies include data ownership and intellectual property management (Bamdad et al., 2022), implicit speech bias (Goldenthal et al., 2021), and decision-making accountability (Long & Magerko, 2020). Moreover, methodological concerns from previous studies include data repository and flow strategies (Bamdad et al., 2022), the need to exercise the real-time nature of interactions between the researcher and the participants (Lim & Kaveri, 2024).

CURRENT FINDINGS

This study reflected on the data collection process using online interviews in research involving doctoral student mothers during February and May 2023. The online interviews were conducted using closed-captioning enabled Zoom Meetings, which were also synchronized with Otter.ai services for automatic notes and transcriptions. Ten individual online interviews were conducted, resulting in nine transcripts from the Zoom Meetings platform and nine transcripts from the Otter.ai platform. On one occasion, one transcript from Zoom Meetings did not go through because closed-captioning was not activated due to human error, but the Otter.ai transcript for this particular meeting was working. On another occasion, one transcript from the Otter.ai platform did not go through due to a technical error in Otter.ai and Zoom Meetings account integration, but the Zoom transcript for that meeting was available. A lesson learned from this practice is that using two different notes and transcription services resulted in higher chances of not losing the interview transcripts. If either of the services did not go through, then the other one can serve as a backup purpose.

The lens of ethical concerns on data ownership and intellectual property management was not clearly visible when collecting data during the research, as mentioned in the previous paragraph. Two researchers were involved at that time, including the researcher who wrote this article, and both of them were affiliated with the same institution. It is, however, more practical to also look at data security and privacy measures used by the data collection tools used in the study. For instance, the aforementioned study used the following software for data collection: Qualtrics, Zoom Meetings, and the Otter.ai platforms. Qualtrics is compliant with EU-based data privacy law, i.e., GDPR (General Data Protection Regulation). Zoom Meetings are also GDPR and HIPAA-compliant and follow the California Consumer Privacy Act (CCPA), which is among the U.S. state data privacy laws. Otter.ai transcription services follow best practices for

data security and privacy laws, including GDPR, CCPA, and Voluntary Product Accessibility Template (VPAT) for accessibility requirements compliance.

Implicit speech bias situation can be reflected when the researcher in this ongoing study looked at the transcripts from the Zoom Meetings and the Otter.ai platform. The researcher, who is a non-English native speaker, found that the transcriptions worked better when transcribing the participants' responses who happened to be English native speakers. A few transcriptions were not clear enough and, therefore, should be changed manually during the data cleaning. One way to increase the accuracy of the Otter-ai transcribing services is to record the researcher's voiceprint in the Otter.ai platform prior to the online interviews, as discussed in Corrente and Bourgeault (2022). Regarding the notes and transcription results, the researcher only made use of the transcription for data cleaning and further data analysis. The summary of the meetings was not included in both processes. Overviewing decision-making accountability of AI-based notes and transcription services can be reflected during data cleaning when the researcher found out about the wrong assignment of names in the transcripts, where there was one piece of response that should have been assigned as the researcher's response, but it was assigned to the participant's response.

In the case of data repository as a part of methodological concerns, all subscriptions to Zoom Meetings and Otter.ai services were associated with the researcher's institutional email account. The aforementioned study on doctoral student mothers' use of mobile devices enabled the Qualtrics account associated with the researcher's institutional account, and the export data were stored at the institution's Onedrive. Zoom Meetings associated with the institution's Onedrive were used, and the Zoom recordings and transcripts were downloaded from the cloud and stored at the institution's Onedrive. The Otter.ai recordings and transcripts were also downloaded from the cloud and stored at the institution's Onedrive. Furthermore, the methodological concern regarding the real-time nature of interactions between the researcher and the participants can be shown in one example where a participant needs to bring along her young child next to her during the interview. When the discussion was halted, the researcher paused the recording and continued when the participant was ready to continue. The Otter.ai platform also enables sending a copy of the interview transcripts to the participants right after the transcription from this platform is ready if the participants' email address is included in the Otter.ai conversation.

FUTURE PLANS

This ongoing study's reflection has opened up the discussion of ethical and methodological concerns using AI-based notes and transcription services in LIS research using online interviews. LIS researchers performing qualitative studies using these services can use this information as a basis for finding strategies to minimize the impact of using AI technology in internet-based research. The study also further calls for specific LIS guidelines in internet-based research using AI technology, especially when collecting data remotely.

REFERENCES

Archibald, M. M., Ambagtsheer, R. C., Casey, M. G., & Lawless, M. (2019). Using Zoom videoconferencing for qualitative data collection: Perceptions and experiences of researchers and participants. *International Journal of Qualitative Methods*, *18*. https://doi.org/10.1177/1609406919874596

Bamdad, S., Finaughty, D. A., & Johns, S. E. (2022). 'Grey areas': Ethical challenges posed by social media-enabled recruitment and online data collection in cross-border, social science research. *Research Ethics*, *18*(1), 24-38. https://doi.org/10.1177/17470161211045557

Corrente, M., & Bourgeault, I. (2022). Innovation in transcribing data: Meet Otter.ai. In *Sage research methods: Doing research online*. SAGE Publications. https://doi.org/10.4135/9781529799033

de Villiers, C., Farooq, M. B., & Molinari, M. (2022). Qualitative research interviews using online video technology – challenges and opportunities. *Meditari Accountancy Research*, *30*(6), 1764-1782. https://doi.org/10.1108/medar-03-2021-1252

Goldenthal, E., Park, J., Liu, S. X., Mieczkowski, H., & Hancock, J. T. (2021). Not all AI are equal: Exploring the accessibility of AI-mediated communication technology. *Computers in Human Behavior*, *125*, Article 106975. https://doi.org/10.1016/j.chb.2021.106975

Gray, L. M., Wong-Wylie, G., Rempel, G. R., & Cook, K. (2020). Expanding qualitative research interviewing strategies: Zoom video communications. *The Qualitative Report*, *25*(5), 1292-1301. https://doi.org/10.46743/2160-3715/2020.4212

Handler, A., Mahyar, N., & O'Connor, B. (2022). ClioQuery: Interactive query-oriented text analytics for comprehensive investigation of historical news archives. *ACM Transactions on Interactive Intelligent Systems*, *12*(3), 1-49. https://doi.org/10.1145/3524025

Horváth, I. (2022). AI in interpreting: Ethical considerations. *Across Languages and Cultures*, *23*(1), 1-13. https://doi.org/10.1556/084.2022.00108

Kazmer, M. M., & Xie, B. (2008). Qualitative interviewing in internet studies: Playing with the media, playing with the method. *Information, Communication & Society*, *11*(2), 257-278. https://doi.org/10.1080/13691180801946333

Keen, S., Lomeli-Rodriguez, M., & Joffe, H. (2022). From challenge to opportunity: Virtual qualitative research during COVID-19 and beyond. *Int J Qual Methods*, *21*, 1-11. https://doi.org/10.1177/16094069221105075

Lim, C. A. M., & Kaveri, G. (2024). Listening to children's voices: Reflections on methods, practices and ethics in researching with children using zoom video interviews. *Qualitative Research Journal*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/QRJ-10-2023-0156

Lobe, B., Morgan, D., & Hoffman, K. A. (2020). Qualitative data collection in an era of social distancing. *International Journal of Qualitative Methods*, *19*. https://doi.org/10.1177/1609406920937875

Long, D., & Magerko, B. (2020). *What is AI literacy? Competencies and design considerations* Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, USA.

Lynn Silipigni, C., & Marie, L. R. (2021). Individual and focus group interviews. In *Research methods in library and information science* (7th ed.). Libraries Unlimited.

McCarron, L. (2021). Creating accessible videos: Captions and transcripts. *Communications of the Association for Information Systems*, *48*. https://doi.org/10.17705/1CAIS.04819

Menzies, R., Gorman, B. M., & Tigwell, G. W. (2020, 2020-10-26). Reflections on using chat-based platforms for online interviews with screen-reader users. Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility,

Newman, P. A., Guta, A., & Black, T. (2021). Ethical considerations for qualitative research methods during the COVID-19 pandemic and other emergency situations: Navigating the virtual field. *International Journal of Qualitative Methods*, *20*. https://doi.org/10.1177/16094069211047823

Oates, M., Crichton, K., Cranor, L., Budwig, S., Weston, E. J. L., Bernagozzi, B. M., & Pagaduan, J. (2022). Audio, video, chat, email, or survey: How much does online interview mode matter? *Plos One*, *17*(2), 1-43. https://doi.org/10.1371/journal.pone.0263876

Seilkhanova, T., Ledford, T. D., & Schneider, J. (2024, 2024-06-18). COVID-19 information spaces, boundaries, and information sharing: An interview study. Information Research, Aalborg, Denmark.

Suduc, A. M., & Bizoi, M. (2022). AI shapes the future of web conferencing platforms. *Procedia Computer Science*, *214*, 288-294. https://doi.org/10.1016/j.procs.2022.11.177

Timbrook, J., Olson, K., & Smyth, J. D. (2018). Why do cell phone interviews last longer? A behavior coding perspective. *Public Opinion Quarterly*, *82*(3), 553-582. https://doi.org/10.1093/poq/nfy022

Żadkowska, M., Dowgiałło, B., Gajewska, M., Herzberg-Kurasz, M., & Kostecka, M. (2022). The sociological confessional: A reflexive process in the transformation from face-to-face to online interview. *International Journal of Qualitative Methods*, *21*, 1-12. https://doi.org/10.1177/16094069221084785