Manual transcription still beats AI: A comparative study on transcription services

April 5 2024

From Hashes to Ashes – A Comparison of Transcription Services. Credit: CISPA

A research team from the Empirical Research Support (ERS) at CISPA Helmholtz Center for Information Security has conducted a systematic comparison of the most popular transcription services. The comparison involved 11 providers of manual as well as AI-based transcriptions.

It shows that, good quality notwithstanding, the latter still have problems with speaker attribution and that there are discrepancies between recording and transcription that distort meaning. Whisper AI from OpenAI delivered the best results among the AI providers.

Interviews are a popular method for collecting scientific data. There is a basic distinction between quantitative and qualitative interviews. While the former is designed to obtain statistically usable information from a large number of participants with the help of standardized questionnaires, the latter is aimed at obtaining interview data that allow for interpretation by the researchers.

A special type is the guided interview, in which there is a prepared list of questions, which can, however, be deviated from during the interview. "In cybersecurity research, these interviews are utilized when exploring the patterns of action and interpretation of actors who operate through digital means," explains sociologist Dr. Rafael Mrowczynski from CISPA's Empirical Research Support (ERS) team. The ERS team advises the Center's researchers on methodological issues.

Converting an audio file into text

Transcription is a crucial step in qualitative data analysis. "The standard procedure is to convert the audio recordings of the interviews into text. It is important for the quality of the data that the transcriptions are adequate," Mrowczynski explains. Depending on the scientific field, there are different standards for transcription.

"In cybersecurity research, we usually work with transcripts that precisely reproduce the content of the conversation," says Mrowczynski. An adequate transcript, therefore, only contains the relevant spoken words. The researchers can obtain the transcript in two ways: Either it is created by the research team itself, or the task is outsourced to third-

party providers.

Among the third-party providers, besides manual transcription, there has recently been real hype about automated, AI-based transcription. This is due to the exponential leaps in development and quality that AI applications have experienced in many areas over the last two years.

The researchers from CISPA's ERS team wanted to know which provider on the market achieves the best results and how automated, AI-based transcription performs in comparison with manual transcription. The goal was to be able to provide the researchers at CISPA and the cybersecurity community with a recommendation for working with qualitative interviews.

The approach of the ERS team

For their research project, Mrowczynski and his colleagues Dr. Maria Hellenthal, Dr. Rudolf Siegel, and Dr. Michael Schilling created a test dataset. This consisted of individual interviews lasting about ten minutes and group discussions with CISPA researchers in German and English. The content focused on the research field of cybersecurity.

"It was important that technical terms from the community were included so that the precision of the transcription could be assessed," Mrowczynski explains. Some of the interviews were additionally enhanced with background noise in order to reflect real settings in everyday research better.

The data were sent to eleven providers in December 2022. Among those were the transcription services Amberscript, GoTranscript, QualTranscribe, Rev, and Scribbl, as well as the AI-based transcription providers Amazon Transcribe, AssemblyAI, Audiotranskription.de, Google Cloud, Microsoft Azure, and Whisper by OpenAI.

For the assessment of the obtained transcripts, Mrowczynski and his colleagues created a reference transcript that served as the basis for the comparative analysis. The analysis itself then focused on two central criteria. First, the researchers assessed the word error rate, which indicates by how many words a transcript differs from the reference transcript. Second, the qualitative deviation from the reference transcript was coded manually.

Manual transcription services beat AI

In their paper, Mrowczynski and his colleagues conclude that, in general, "most of the manual transcription services achieve a commendable level of performance, while AI-based services often show meaning-distorting discrepancies between recording and transcription."

The distortion of meaning can be clearly seen in technical terms; Mrowczynski explains, "In the transcript, for example, the term 'hashes' became 'ashes." That is how we came up with the title of the paper."

OpenAI's Whisper achieved the best results among the AI-based providers. Most providers handled English better than German. Three providers did not offer transcription for German at all. Background noise generally had a negative effect on the result. The AI-based providers particularly had problems with speaker assignments.

In addition, the transcripts created by an AI had to be reformatted before it was possible to further process them in software for qualitative data analysis. However, the researchers point out that their analysis reflects the state of the art as of December 2022 and that current developments could not be taken into account.

The research was presented at the 2023 CCS ACM Conference on Computer and Communications Security.

Provided by CISPA Helmholtz Center for Information Security