Researchers accurately identify people with PTSD through text data alone

April 7 2022, by Adrianna MacPherson

A study participant is interviewed by Ellie, an artificial character, to gather text data. Credit: Jonathan Gratch, USC Institute for Creative Technologies

University of Alberta researchers have trained a machine learning model to identify people with post-traumatic stress disorder with 80 percent

accuracy by analyzing text data. The model could one day serve as an accessible and inexpensive screening tool to support health professionals in detecting and diagnosing PTSD or other mental health disorders through telehealth platforms.

Psychiatry Ph.D. candidate Jeff Sawalha, who led the project, performed a sentiment analysis of text from a dataset created by Jonathan Gratch at USC's Institute for Creative Technologies. Sentiment analysis involves taking a large body of data, such as the contents of a series of tweets, and categorizing them—for example, seeing how many are expressing positive thoughts and how many are expressing negative thoughts.

"We wanted to strictly look at the sentiment analysis from this dataset to see if we could properly identify or distinguish individuals with PTSD just using the emotional content of these interviews," said Sawalha.

The text in the USC dataset was gathered through 250 semi-structured interviews conducted by an artificial character, Ellie, over video conferencing calls with 188 people without PTSD and 87 with PTSD.

Sawalha and his team were able to identify individuals with PTSD through scores indicating that their speech featured mainly neutral or negative responses.

"This is in line with a lot of the literature around emotion and PTSD. Some people tend to be neutral, numbing their emotions and maybe not saying too much. And then there are others who express their negative emotions."

The process is undoubtedly complex. For example, even a simple phrase like "I didn't hate that" could prove challenging to categorize, explained Russ Greiner, study co-author, professor in the Department of

Computing Science and founding scientific director of the Alberta Machine Intelligence Institute. However, the fact that Sawalha was able to glean information about which individuals had PTSD from the text data alone opens the door to the possibility of applying similar models to other datasets with other mental health disorders in mind.

"Text data is so ubiquitous, it's so available, you have so much of it," Sawalha said. "From a machine learning perspective, with this much data, it may be better able to learn some of the intricate patterns that help differentiate people who have a particular mental illness."

Next steps involve partnering with collaborators at the U of A to see whether integrating other types of data, such as speech or motion, could help enrich the model. Additionally, some neurological disorders like Alzheimer's as well as some mental health disorders like schizophrenia have a strong language component, Sawalha explained, making them another potential area to analyze.

The research was published in *Frontiers in Psychiatry*.

Provided by University of Alberta