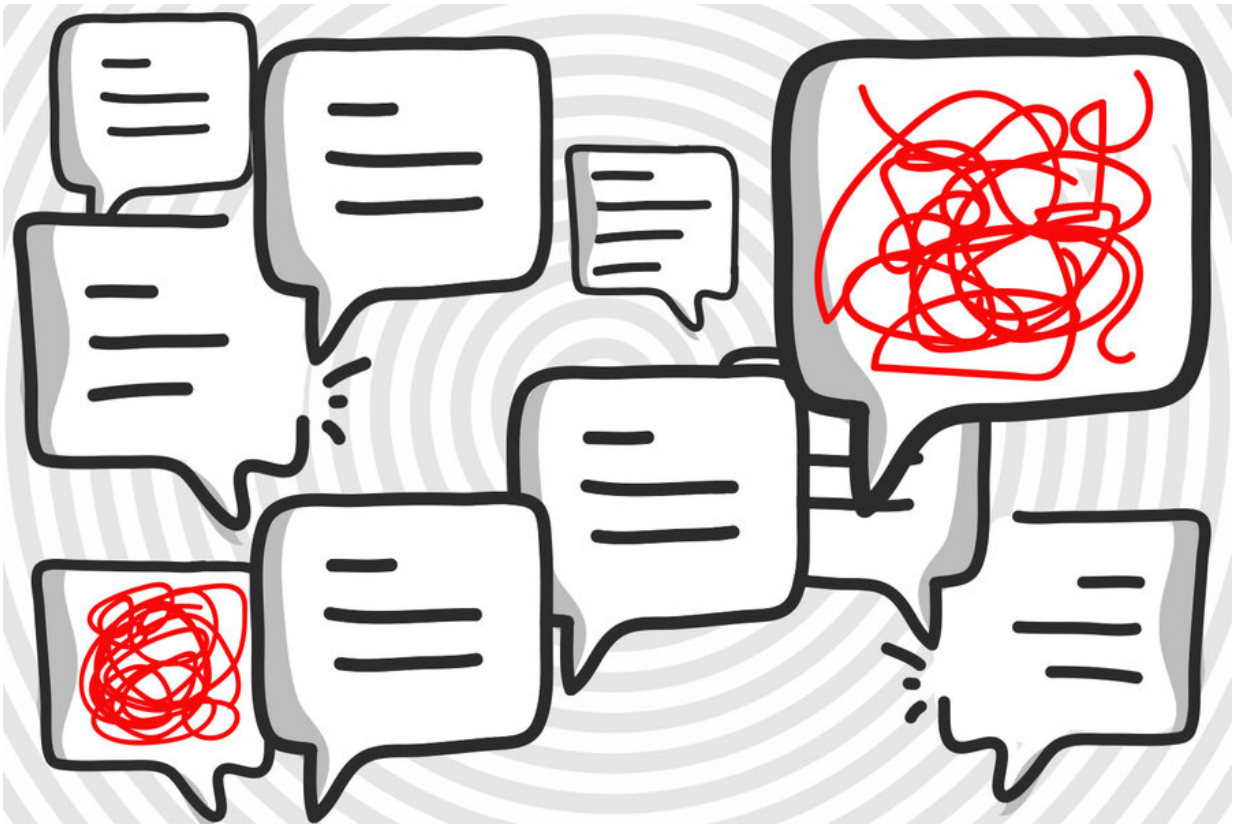# Using machine learning to track the pandemic's impact on mental health

November 5 2020, by Anne Trafton

Researchers used machine learning to analyze the text of hundreds of thousands of Reddit posts, revealing several key impacts on mental health, including an overall increase in anxiety and in thoughts about suicide. Credit: Christine Daniloff, MIT

Dealing with a global pandemic has taken a toll on the mental health of

millions of people. A team of MIT and Harvard University researchers has shown that they can measure those effects by analyzing the language that people use to express their anxiety online.

Using machine learning to analyze the text of more than 800,000 Reddit posts, the researchers were able to identify changes in the tone and content of language that people used as the first wave of the Covid-19 pandemic progressed, from January to April of 2020. Their analysis revealed several key changes in conversations about mental health, including an overall increase in discussion about anxiety and suicide.

"We found that there were these natural clusters that emerged related to suicidality and loneliness, and the amount of posts in these clusters more than doubled during the pandemic as compared to the same months of the preceding year, which is a grave concern," says Daniel Low, a graduate student in the Program in Speech and Hearing Bioscience and Technology at Harvard and MIT and the lead author of the study.

The analysis also revealed varying impacts on people who already suffer from different types of mental illness. The findings could help psychiatrists, or potentially moderators of the Reddit forums that were studied, to better identify and help people whose mental health is suffering, the researchers say.

"When the mental health needs of so many in our society are inadequately met, even at baseline, we wanted to bring attention to the ways that many people are suffering during this time, in order to amplify and inform the allocation of resources to support them," says Laurie Rumker, a graduate student in the Bioinformatics and Integrative Genomics PhD Program at Harvard and one of the authors of the study.

Satrajit Ghosh, a principal research scientist at MIT's McGovern Institute for Brain Research, is the senior author of the study, which

appears in the *Journal of Internet Medical Research*. Other authors of the paper include Tanya Talkar, a [graduate student](#) in the Program in Speech and Hearing Bioscience and Technology at Harvard and MIT; John Torous, director of the digital psychiatry division at Beth Israel Deaconess Medical Center; and Guillermo Cecchi, a principal research staff member at the IBM Thomas J. Watson Research Center.

A wave of anxiety

The new study grew out of the MIT class 6.897/HST.956 (Machine Learning for Healthcare), in MIT's Department of Electrical Engineering and Computer Science. Low, Rumker, and Talkar, who were all taking the course last spring, had done some previous research on using machine learning to detect mental health disorders based on how people speak and what they say. After the Covid-19 pandemic began, they decided to focus their class project on analyzing Reddit forums devoted to different types of mental illness.

"When Covid hit, we were all curious whether it was affecting certain communities more than others," Low says. "Reddit gives us the opportunity to look at all these subreddits that are specialized [support groups](#). It's a really unique opportunity to see how these different communities were affected differently as the wave was happening, in real-time."

The researchers analyzed posts from 15 subreddit groups devoted to a variety of mental illnesses, including schizophrenia, depression, and bipolar disorder. They also included a handful of groups devoted to topics not specifically related to mental health, such as personal finance, fitness, and parenting.

Using several types of natural language processing algorithms, the researchers measured the frequency of words associated with topics such

as anxiety, death, isolation, and substance abuse, and grouped posts together based on similarities in the language used. These approaches allowed the researchers to identify similarities between each group's posts after the onset of the pandemic, as well as distinctive differences between groups.

The researchers found that while people in most of the support groups began posting about Covid-19 in March, the group devoted to health anxiety started much earlier, in January. However, as the pandemic progressed, the other mental health groups began to closely resemble the health anxiety group, in terms of the language that was most often used. At the same time, the group devoted to personal finance showed the most negative semantic change from January to April 2020, and significantly increased the use of words related to economic stress and negative sentiment.

They also discovered that the mental health groups affected the most negatively early in the pandemic were those related to ADHD and eating disorders. The researchers hypothesize that without their usual social support systems in place, due to lockdowns, people suffering from those disorders found it much more difficult to manage their conditions. In those groups, the researchers found posts about hyperfocusing on the news and relapsing back into anorexia-type behaviors since meals were not being monitored by others due to quarantine.

Using another algorithm, the researchers grouped posts into clusters such as loneliness or substance use, and then tracked how those groups changed as the pandemic progressed. Posts related to suicide more than doubled from pre-pandemic levels, and the groups that became significantly associated with the suicidality cluster during the pandemic were the support groups for borderline personality disorder and post-traumatic stress disorder.

The researchers also found the introduction of new topics specifically seeking mental health help or social interaction. "The topics within these subreddit support groups were shifting a bit, as people were trying to adapt to a new life and focus on how they can go about getting more help if needed," Talkar says.

While the authors emphasize that they cannot implicate the pandemic as the sole cause of the observed linguistic changes, they note that there was much more significant change during the period from January to April in 2020 than in the same months in 2019 and 2018, indicating the changes cannot be explained by normal annual trends.

Mental health resources

This type of analysis could help mental health care providers identify segments of the population that are most vulnerable to declines in mental health caused by not only the Covid-19 pandemic but other mental health stressors such as controversial elections or natural disasters, the researchers say.

Additionally, if applied to Reddit or other social media posts in real-time, this analysis could be used to offer users additional resources, such as guidance to a different support group, information on how to find mental health treatment, or the number for a suicide hotline.

"Reddit is a very valuable source of support for a lot of people who are suffering from mental health challenges, many of whom may not have formal access to other kinds of mental health support, so there are implications of this work for ways that support within Reddit could be provided," Rumker says.

The researchers now plan to apply this approach to study whether posts on Reddit and other social media sites can be used to detect mental

health disorders. One current project involves screening posts in a social media site for veterans for suicide risk and post-traumatic stress disorder.

More information: Daniel M Low et al. Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study, *Journal of Medical Internet Research* (2020). DOI: 10.2196/22635

Provided by Massachusetts Institute of Technology