AI Analysis of Social Media Language Predicts Depression Severity for White but Not Black Americans

Study also found Black people with depression used different language compared with white people to express their thoughts on Facebook.

April 16, 2024 By National Institutes of Health

Researchers were able to predict depression severity for white people, but not for Black people using standard language-based computer models to analyze Facebook posts. Words and phrases associated with depression, such as first-person pronouns and negative emotion words, were around three times more predictive of depression severity for white people than for Black people.

The study, published today in the Proceedings of the National Academy of Sciences, is co-authored by researchers at the University of Pennsylvania, Philadelphia, and the National Institute on Drug Abuse (NIDA), part of the National Institutes of Health (NIH), which also funded the study.

While previous research has indicated that social media language could provide useful information as part of mental health assessments, the findings from this study point to potential limitations in generalizing this practice by highlighting key demographic differences in language used by people with depression. The results also highlight the importance of including diverse pools of data to ensure accuracy as machine learning models, an application of artificial intelligence (AI) language models, are developed.

“As society explores the use of AI and other technologies to help deliver much-needed mental health care, we must ensure no one is left behind or misrepresented,” said Nora Volkow, MD, NIDA director. “More diverse datasets are essential to ensure that healthcare disparities are not perpetuated by AI and that these new technologies can help tailor more effective health care interventions.”

The study, which recruited 868 consenting participants who identified themselves as Black or white, demonstrated that models trained on Facebook language used by white participants with self-reported depression showed strong predictive performance when tested on the white participants. However, when the same models were trained on Facebook language from Black participants, they performed poorly when tested on the Black participants, and showed only slightly better performance when tested on white participants.

While depression severity was associated with increased use of first-person singular pronouns (“I,” “me,” “my”) in white participants, this correlation was absent in Black participants. Additionally, white people used more language to describe feelings of belongingness (“weirdo,” “creep”), self-criticism (“mess,” “wreck”), being an anxious-outsider (“terrified,” “misunderstood”), self-deprecation (“worthless,” “crap”), and despair (“begging,” “hollow”) as depression severity increased, but there was no such correlation for Black people. For decades, clinicians have been aware of demographic differences in how people express depressive symptoms, and this study now demonstrates how this can play out in social media.

Language-based models hold promise as personalized, scalable, and affordable tools to screen for mental health disorders. For example, excessive self-referential language, such as the use of first-person pronouns, and negative emotions, such as self-deprecating language, are often regarded as clinical indicators of depression. However, there has been a notable absence of racial and ethnic consideration in assessing mental disorders through language, an exclusion that leads to inaccurate computer models. Despite evidence showing that demographic factors influence the language people use, previous studies have not systematically explored how race and ethnicity influence the relationship between depression and language expression.

Researchers set up this study to help bridge this gap. They analyzed past Facebook posts from Black and white people who self-reported depression severity through the Patient Health Questionnaire (PHQ-9) — a standard self-report tool used by clinicians to screen for possible depression. The participants consented to share their Facebook status updates. Participants were primarily female (76%) and ranged from 18 to 72 years old. The researchers matched Black and white participants on age and sex so that data from the two groups would be comparable.

The study’s findings challenge assumptions about the link between the use of certain words and depression, particularly among Black participants. Current clinical practices in mental health that have not accounted for racial and ethnic nuances may be less relevant, or even irrelevant, to populations historically excluded from mental health research, the researchers note. They also hypothesize that depression may not manifest in language in the same way for some Black people — for example, tone or speech rate, instead of word selection, may relate more to depression among this population.

“Our research represents a step forward in building more inclusive language models. We must make sure that AI models incorporate everyone’s voice to make technology fair for everyone,” said Brenda Curtis, Ph.D., MsPH, chief of the Technology and Translational Research Unit in the Translational Addiction Medicine Branch at NIDA’s Intramural Research Program and one of the study’s senior authors. “Paying attention to the racial nuances in how mental health is expressed lets medical professionals better understand when an individual needs help and provide more personalized interventions.”

Future studies will need to examine differences across other races and demographic features, using various social media platforms, the authors say. They also caveat that social media language is not analogous to everyday language, so future work on language-based models must take this

into account.

“It’s important to note that social media language and language-based AI models are not able to diagnose mental health disorders — nor are they replacements for psychologists or therapists — but they do show immense promise to aid in screening and informing personalized interventions,” said the study’s lead author, Sunny Rai, Ph.D., a postdoctoral researcher in Computer and Information Science at the University of Pennsylvania. “Many improvements are needed before we can integrate AI into research or clinical practice, and the use of diverse, representative data is one of the most critical.”

For more information on substance and mental health treatment programs in your area, call the free and confidential National Helpline 1-800-662-HELP (4357) or visit FindTreatment.gov(link is external). Anyone who needs assistance with the first steps in pursuing help can find guidance at FindSupport.gov.

If you or someone you know is in crisis and needs immediate help, call the 988 Suicide & Crisis Lifeline at 988. Learn more about suicide prevention and ways you can help someone who might be at risk for self-harm.

This news release was published by the National Institutes of Health on March 26, 2024.

© 2026 Smart + Strong All Rights Reserved.
http://beta.docker.sanemag.com/article/ai-analysis-social-media-language-predicts-depression-severity-white-black-americans