PNAS

Supporting Information for

Key language markers of depression on social media depend on race

Sunny Rai, Elizabeth C. Stade, Salvatore Giorgi, Ashley Francisco, Lyle H. Ungar, Brenda Curtis, Sharath C. Guntuku

Sunny Rai.
E-mail: sunnyrai@seas.upenn.edu

This PDF file includes:

Supporting text
SI References

Supporting Information Text

1. Methods

A. Participants. Our initial sample consisted of Black and White English speakers in the United States who provided demographic information including race, and access to their Facebook posts and completed a self-report measure of depression severity (Patient Health Questionnaire (PHQ)-9; n = 2,179). There were 67% female and 31.2% male participants, ranged in age from 18 to 81 (M = 39.88, SD = 12.63), with mild levels of depression, on average (M=7.29, SD=6.17).

A.1. Matching. Individuals with a cumulative word count of less than 300 words on Facebook were excluded as they lacked sufficient language to obtain reliable language estimates (1, 2). This resulted in 2,121 (438 Black) individuals. Another 8 individuals who marked their gender as *others* were dropped. We performed coarsened exact matching (3) using the *Matchit* package in R to minimize the confounding effect of age and gender on the regression analysis, resulting in a final sample (n = 868) with equal numbers of Black and White participants matched on age and gender. Age, a continuous variable, was categorized into five bins for matching.

A.2. Data Statistics. In the matched sample, participants were primarily female (76%), ranged in age from 18 to 72 (M = 38.10, SD = 10.89), with mild levels of depression, on average (M = 7.37, SD = 6.25). The matched sample aligns with the social media usage behavior of US adults i.e., more US female adults (77%) are known to use Facebook than US male adults (61%) and no significant difference exists in the percentage of Blacks and Whites who use Facebook (4). Although 75.5% of the US population are Whites and 13.6% are Blacks (5), our final sample is over-enriched for Black individuals vis-a-vis white individuals on purpose to increase our ability to detect an effect. The median 1-gram count is 14,298.5 for the participants identifying as White and 12,615.5 for the participants identifying as Black, reflecting a similar amount of language sample for training. The median number of statuses posted is 755 (and the median number of words per status is 11) for White individuals and 687.5 (and the median number of words per status is 10) for Black individuals.

B. Language markers associated with depression and race. Several prior works identified first-person pronoun usage and negative emotions (6, 7) as consistent language markers of depression. First-person pronoun usage is known to reflect self-focused attention (SFA), a focus on information generated internally (from the self) rather than externally (from the environment; (8, 9)). SFA involves preoccupation with one's negative thoughts and feelings and serves to maintain/amplify negative affect. In our analyses, we wanted to analyze these two categories. First person pronoun usage was obtained using Linguistic Inquiry Word Count (LIWC-22; (10)). We correlated a pre-trained set of 2000 Facebook topics (11), reweighted to the word distribution in our sample, and found 320 topics were significantly correlated ($p < 0.05$) with depression in our matched sample. One of the co-authors manually labeled these topics to identify topics expressing negative emotions, which was verified by a second member of the team independently. We studied the interaction effect of race with 23 topics that were found to be expressing negative emotions.

In a series of primary analyses, we regressed each language marker on depression severity; in secondary analyses focused on moderation by race, we introduced a dummy-coded race variable (White = 0, Black = 1) and a race by language marker interaction term into our regression equation. We probed significant interactions using simple-slopes analyses to understand how race moderated the use of each language marker across the range of depression severity. We applied Benjamini-Hochberg (BH) false discovery rate (FDR) correction and controlled for age and gender in all analyses. In contrast to Family-wise Error Rate (FWER) that controls for *any* false positives (i.e., the denominator is the total number of statistical tests conducted), FDR is used more widely because it controls for the overall rate of false positives and the denominator is the number of "discoveries" (i.e., rejections of the null) (12). We set the significance level as 0.05 that is, the proportion of false discoveries should be less than 5%. We validated the meaning of top keywords and assigned themes by manually analyzing the top messages associated with the topics that survived BH correction.

B.1. I-usage variance. We performed Shapiro-Wilk Normality test to test Gaussian distribution (statistic = 0.98; p-value = 0.88) for I-usage by Black individuals in our sample. Further, an overlaid histogram was created to compare race-wise variance in I-usage (Provided in OSF Data repository.).

C. Performance of Language-Based Depression Models Across Race. We built language-based depression prediction models and compared their performance across racial subgroups. To do so, we input a large set of language markers (i.e., 1-3 grams[*], 102 LIWC categories, 2,000 LDA topics) to ridge regression models ($\alpha = 10^4$) trained to predict depression severity using 5-fold cross-validation. Following these methods, we trained two separate language-based depression prediction models on exclusively White (Model M_{white}) and exclusively Black (Model M_{black}) samples, respectively, then tested their performance on exclusively White and exclusively Black samples, respectively, resulting in a 2x2 (train/test) design. We also repeated this analysis using concatenated BERT embeddings (13) from the last two layers (Layer 11 and 12) as features. Both models are tested on the held-out sets namely White Test set and Black Test set.

[*] n-gram refers to a sequence of n consecutive words

2. Differential measurement of depression by race

The risk of differential measurement error (i.e., the idea that the relationship between latent depression and observed depression systematically varies by some demographic factor, like race), is assessed by testing for measurement invariance of the instrument in question (e.g., PHQ-9) as a function of the demographic factor in question (e.g., racial group). Existing literature provides evidence for measurement invariance of the PHQ-9 across racial groups, (14, 15) allaying concerns that the PHQ-9 has reduced accuracy for black individuals. This evidence bolsters our confidence that our findings are attributable to the explanation that race moderates the relationship between depression and language rather than that race moderates the relationship between latent depression and observed depression.

References

1. K Jaidka, S Guntuku, L Ungar, Facebook versus twitter: Differences in self-disclosure and trait prediction in *Proceedings of the International AAAI Conference on Web and Social Media.* Vol. 12, (2018).
2. JC Eichstaedt, et al., Closed-and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychol. Methods* **26**, 398 (2021).
3. S Iacus, G King, G Porro, Cem: Software for coarsened exact matching. *J. statistical Softw.* **30**, 1–27 (2009).
4. Pew Research Center, Social media use in 2021. *NA* (2021).
5. United States Census Bureau, Population estimates, july 1, 2023. *NA* (2023).
6. LB Fisher, JC Overholser, J Ridley, A Braden, C Rosoff, From the outside looking in: Sense of belonging, depression, and suicide risk. *Psychiatry* **78**, 29–41 (2015).
7. F Rhodewalt, S Agustsdottir, Effects of self-presentation on the phenomenal self. *J. Pers. Soc. Psychol.* **50**, 47 (1986).
8. RE Ingram, Self-focused attention in clinical disorders: review and a conceptual model. *Psychol. bulletin* **107**, 156 (1990).
9. E Stade, LH Ungar, G Sherman, AM Ruscio, , et al., Depression and anxiety have distinct and overlapping language patterns: Results from a clinical interview. (2023).
10. RL Boyd, A Ashokkumar, S Seraj, JW Pennebaker, The development and psychometric properties of liwc-22. *Austin, TX: Univ. Tex. at Austin* pp. 1–47 (2022).
11. HA Schwartz, et al., Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one* **8**, e73791 (2013).
12. MA Lindquist, A Mejia, Zen and the art of multiple comparisons. *Psychosom. medicine* **77**, 114 (2015).
13. J Devlin, MW Chang, K Lee, K Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
14. BT Keum, MJ Miller, KK Inkelas, Testing the factor structure and measurement invariance of the phq-9 across racially diverse us college students. *Psychol. assessment* **30**, 1096 (2018).
15. JS Patel, et al., Measurement invariance of the patient health questionnaire-9 (phq-9) depression screener in us adults across sex, race/ethnicity, and education level: Nhanes 2005–2016. *Depress. anxiety* **36**, 813–823 (2019).