

Original Paper

Naturalistic Online Language as a Marker of Depression in Midlife and Older Adults: Computational Text Analysis of Bluesky and Reddit Posts

Lauren Rutter^{1,2}, PhD; Andy Edinger³, MS; Marijn ten Thij⁴, PhD; Lorenzo Lorenzo-Luaces¹, PhD; Danny Valdez⁵, PhD; Johan Bollen^{6,7}, PhD

¹Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, United States

²Indiana Alzheimer's Disease Research Center, Indiana University School of Medicine, Indianapolis, IN, United States

³Cognitive Science Program, Indiana University, Bloomington, IN

⁴Department of Intelligent Systems, Tilburg University, Tilburg, North Brabant, The Netherlands

⁵Department of Applied Health Sciences, Indiana University, Bloomington, IN

⁶Informatics Institute, Faculty of Science, University of Amsterdam, Amsterdam, North Holland, The Netherlands

⁷Department of Informatics, Computing, and Engineering, Luddy School of Informatics, Computing, and Engineering, Indiana University, IN, United States

Corresponding Author:

Lauren Rutter, PhD
Department of Psychological and Brain Sciences
Indiana University
1101 E 10th Street
Bloomington, IN 47405
United States
Phone: 1 (812) 856-9953
Email: larutter@iu.edu

Abstract

Background: Social media use among older adults continues to grow. Many people use social media to establish meaningful social ties and discuss their mental health. Depression in midlife and older adults is a critical public health concern, yet scalable, sensitive methods for early detection remain limited. Natural language processing offers new opportunities to examine sentiment and mental health through online language in typically understudied populations.

Objective: This study examined whether sentiment and cognitive features of naturalistic social media language are associated with depression diagnosis and symptom severity among midlife and older adults.

Methods: We constructed cohorts of adults aged 50 years and older from Reddit (n=688) and Bluesky (n=210). Participants who had ever had depression and never been diagnosed were recruited via a self-report survey on Prolific that collected depression history, age of diagnosis, current symptoms, and social media handles. Public posts were preprocessed and analyzed using a validated rule-based sentiment analysis (Valence Aware Dictionary and Sentiment Reasoner). Sentiment measures were aggregated at the user level and compared by depression status within and across platforms.

Results: On Bluesky, users with a history of depression exhibited lower Valence Aware Dictionary and Sentiment Reasoner compound sentiment compared to users who had never been diagnosed ($\eta^2 = -0.16$; $P < .001$). On Reddit, current depression severity showed a small association with lower compound sentiment, and although the bootstrap CI excluded zero, the association did not remain statistically significant after correction for multiple comparisons ($P > .05$). These findings indicate a platform-specific relationship between depression status and online sentiment.

Conclusions: Negative sentiment expressed on Bluesky was associated with depression history (ever diagnosed vs never diagnosed) among midlife and older adults. On Reddit, current depression severity showed a small association with negative sentiment that did not remain statistically significant after correction for multiple comparisons. Our findings suggest that sentiment-based language features may capture modest differences in online expression related to depression. Future work with larger samples and longitudinal symptom assessment is needed to determine whether sentiment reliably tracks with depression-related symptoms over time.

Keywords: naturalistic language; older adults; depression; midlife adults; cognitive distortions

Introduction

Over the past decade, digital traces from online platforms have become a dominant source of data, enabling researchers to study mental health phenomena naturalistically and at scale. Social media has emerged as a promising tool to understand processes related to emotions and mental health, offering insights into sleep [1], social relationships [2], markers of distress, and trajectories of well-being over time [3-6]. While activity on social media platforms has historically been dominated by young adults, there has been a steady increase in online engagement among midlife and older adults [7]. On the basis of a 2025 Pew Research Center survey, 90% of adults aged 50 to 64 years and 78% of adults aged 65 years and older own a smartphone [8]. As digital participation rises among midlife and older adults, so too does the opportunity to harness these platforms to assess psychological risk and identify early warning signs of disorders such as depression.

Prior work suggests that smartphone and social media use among older adults is relevant for social connection, loneliness, well-being, and mental health [9-14]. In a study of social media use among adults aged 65 years and older, active and passive social media use was differentially associated with depressive symptoms, with active use increasing odds of depressive symptoms and passive use lowering odds [10]. In another study of smartphone use in older adults, smartphones were used for social and nonsocial purposes, with social media and reading the news as the most common activities [11]. Social media is increasingly used not only to connect with others but also to disclose health experiences, including symptoms of depression, diagnoses, stress, and related concerns [15-17]. This is important because an estimated 10% to 15% of older adults have depressive symptoms. Depression symptoms, even those that are mild, are distressing [18, 19]. At the same time, adults in midlife remain less represented in social media-based mental health research. Studying adults aged 50 years and older provides an opportunity to examine whether computational markers of depression extend to midlife and older adult populations—populations that are both increasingly online and relevant to the study of depression in later life.

Natural language processing (NLP), deep learning, and other computational methods offer powerful tools for extracting information about mental health on social media at scale. In prior work, we extracted the hourly number of Twitter (subsequently rebranded X) posts from a large sample of users as an index of overall activity (ie, behavior) [1]. Content analyses revealed that the sample with depression showed greater levels of positive and negative emotionality when active and a steady rise in ruminative language and rigid thinking from midnight to dawn. In a follow-up study [20], we developed a theory-driven cognitive distortion schemata (CDS) lexicon to capture language reflecting negatively biased or rigid thinking patterns that are central in

cognitive models of depression. We refer to the CDS lexicon as theory-driven because it is based in cognitive theories of depression, particularly those articulated by Beck and Haigh [21]. In contrast to data-driven lexicons that are created from words that statistically distinguish groups or symptom-based depression lexicons that focus on depressive content, the CDS lexicon was designed to capture linguistic patterns consistent with theoretically defined cognitive distortions. Cognitive distortions refer to maladaptive thinking patterns that are negatively biased interpretations of the self, world, and future. These patterns include all-or-nothing thinking (eg, “Anything less than 100% is failure”), jumping to conclusions (eg, “The email from my boss requesting a meeting means I did something wrong”), catastrophizing (eg, “Because he didn’t text me back, he probably doesn’t care about me—I will be single and die alone”), and emotional reasoning (eg, “Because I feel anxious, I won’t have fun tonight”), for example. Using this lexicon, we found in our previous Twitter study that cognitive distortion language was significantly more prevalent in the messages posted by users with depression compared to a random sample, which was true across all age categories. Additional work using this CDS lexicon has examined broader temporal trends in cognitively distorted language and shown that there has been a rise in written distorted language in recent decades [22,23].

Previous research has shown that detailed indicators of behavior [1,24], mood and affect [25], cognitive vulnerability [26,27], and social relationships [28] can be extracted from social media content for millions of individuals. While our prior work used Twitter in all ages, in this study, we extended our work to target adults aged 50 years and older who were active on Reddit, a long-form social networking website, and Bluesky, a short-form social networking website with similar functionality to that of X. We created a novel dataset of Reddit and Bluesky users aged 50 years and older who reported a lifetime diagnosis of depression and an age-matched never-diagnosed comparison group. Our prior work validated the CDS lexicon using Twitter; however, changes in X API access limited our ability to collect new data from the populations of interest. Therefore, we selected Bluesky as a short-form alternative with similar functionality to that of X. Reddit was selected as a contrasting text-based platform. Publicly available posts were retrieved using the Bluesky API and Reddit API through user handles. Data collection was conducted in accordance with the platforms’ applicable terms and institutional review board approval. Together, Reddit and Bluesky allowed us to examine naturalistic language across two distinct, text-based social media platforms.

This study addressed 2 research questions. First, we examined whether adults aged 50 years and older with a lifetime diagnosis of depression differed from an age-matched never-diagnosed comparison group in sentiment across Reddit and Bluesky. As a second, exploratory question, we examined whether depression diagnosis was related to cognitive

distortion prevalence in adults aged 50 years and older. Prior work has linked depression with more negative sentiment and greater use of cognitive distortion language [29], but it remains unclear whether these patterns extend to midlife and older adults on Reddit and Bluesky. By examining sentiment and cognitive distortion language in naturalistic posts across two public, text-based social media platforms, this study addressed gaps in the literature on depression-related language in a digitally active but understudied age group.

Methods

Ethical Considerations

This research was reviewed by the Indiana University Bloomington Institutional Review Board. The broader grant under which this work was conducted was approved under protocol 21317, and the Prolific study reported here was reviewed under protocol 27266 and determined to be exempt research. Participants recruited through Prolific were provided with a study information sheet describing the research prior to participation, and a signed informed consent form was not required.

Data Privacy and Handling

We aimed to recruit social media users with depression who were aged 50 years and above and a comparison group of users aged 50 years and older who had never had depression. As described below, we recruited users with depression from a Prolific study called “Seeking Social Media Users with Diagnosed Depression.” Prolific is an online research participant platform. Prolific users were recruited between May 2025 and September 2025.

To protect participant privacy, social media usernames and handles were removed from the analytic post files and were not stored with age, sex, or depression status information. Posts were saved as text files grouped by depression history without usernames or demographic information included. Because social media posts may contain self-disclosed personal information, we refer to these files as deidentified rather than anonymized. This study was not preregistered.

Cohort Construction

We identified users who had ever had depression and never had depression through a Prolific study asking individuals aged 50 years and older who used social media to fill out a short self-report survey about whether they had ever been diagnosed with depression and their current depression symptoms. Subsequently, we asked them to provide their Reddit and Bluesky handles. From the provided social media handles, we retrieved the available posting history for each user profile. We removed cases where individuals had no posts or provided invalid usernames. Non-English-language posts were excluded, and reposts were removed as these were not concordant with our prior approach and did not reflect the individual poster’s original thoughts. For each account, we retrieved all posts available throughout the time of data collection. Our final analytic sample consisted of 154 users

who had ever had depression diagnoses and 56 users who had never had depression diagnoses from Bluesky and 506 users who had ever had depression diagnoses and 182 users who had never had depression diagnoses from Reddit.

Sentiment Analysis

Sentiment analysis refers to approaches that measure the valence or emotional tone of text using NLP. In this work, we used the Valence Aware Dictionary and Sentiment Reasoner (VADER) [30] to estimate the sentiment of posts. VADER is a rule-based sentiment tool developed specifically for social media text, incorporating a large lexicon of common words, abbreviations, and idiomatic expressions along with heuristics that account for punctuation, negation, hedging, and intensification. We have previously used VADER to study sentiment in social media posts of adults with depression and anxiety [4,20,29]. VADER assigns each word with emotional content a sentiment valence using a validated lexicon and then adjusts these values based on linguistic context, including negation, degree modifiers such as intensifiers or diminishers, punctuation, capitalization, and emojis. VADER computes a compound score, which ranges from -1 , indicating maximally negative sentiment, to $+1$, indicating maximally positive sentiment. In the present analyses, the compound score was used as the primary indicator of overall sentiment, with lower values reflecting more negative sentiment.

Cognitive Distortions

The conceptualization and assessment of cognitive distortion language is described in detail in prior work [20]. Because the current study applied the previously developed CDS lexicon, portions of our description are adapted from the prior publication. The Committee on Publication Ethics and Text Recycling Research Project describes text recycling as the reuse of material from an author’s prior work in a new document, including prose, visuals, or equations. In the current manuscript, any reused text is limited to methodological description of the previously developed CDS lexicon.

The CDS lexicon was developed to identify short words and word sequences that mark the building blocks of cognitive distortions as described in cognitive and cognitive behavioral models of depression. The categories of cognitive distortions that we chose included catastrophizing, dichotomous reasoning, disqualifying the positive, emotional reasoning, fortune-telling, labeling and mislabeling, magnification and minimization, mental filtering, overgeneralizing, personalizing, and “should” statements. In the original CDS development study [20], the lexicon was developed by a panel of psychologists with expertise in cognitive behavioral therapy, depression, cognitive distortions, and language-based assessment of depression. The panel engaged in a process of collaborative design followed by a consensus voting procedure requiring unanimous decision to identify 241 CDS n-grams ranging from 1-gram to 5-grams, each geared to express at least one type of cognitive distortion.

The candidate n-grams were selected to capture context-independent language patterns that could express distorted

thinking, avoiding expressions that are specific to depression-related topics such as poor sleep, fatigue, or health issues. For example, the common 3-gram “I am a” was included as a building block of expressing a variety of labeling and mislabeling cognitive distortions because it would be a highly likely (and nearly unavoidable) n-gram to express many self-referential (“I”) expressions of labeling (“am a”). Where possible, higher-order n-grams were chosen to capture as much of the semantic structure of one or more distorted schemata as possible; for example, the 3-gram “everyone will believe” captures both overgeneralizing and fortune-telling. We did include 1-grams such as “nobody” and “everybody” as they strongly correspond to the expression of dichotomous reasoning. For the purposes of the current study, we combined all CDS categories to create an overall cognitive distortion language prevalence measure. Cognitive distortion prevalence was calculated by counting the number of posts that contained at least one CDS n-gram divided by the total number of posts, as described in the Data Analysis section.

Depression Assessment

We used the Patient Health Questionnaire–8 (PHQ-8) [31], a widely used measure of depression symptoms with excellent psychometric properties. The PHQ-8 is an 8-item scale where 8 symptoms of depression are rated from 0 (“not at all”) to 3 (“nearly every day”). Scores range from 0 to 24, with 0 to 4 indicating no significant depression symptoms, 5 to 9 indicating mild depression, 10 to 14 indicating moderate depression, 15 to 19 indicating moderately severe depression, and 20 to 24 indicating severe depression. We selected the PHQ-8 because the study was conducted remotely without real-time clinical monitoring, and therefore, we did not administer the Patient Health Questionnaire–9 suicidal ideation item. Of note, PHQ-8 scores reflected depressive symptom severity during the 2 weeks before study enrollment and, therefore, may not reflect participants’ depressive symptom severity at the time their posts were written. Accordingly, associations between the PHQ-8 and language should be interpreted as exploratory examinations of the relationship between current depressive symptom severity and aggregated language history rather than evidence of contemporaneous associations between depressive symptoms and language use.

Data Analysis

Overview

After creating our final analytic sample, described above in the Cohort Construction section, we normalized text across timelines by applying deterministic cleaning. Posts were lowercased; links, mentions, hashtags, and emojis were removed; white space was collapsed; and punctuation was normalized. We then conducted a user-level analysis for each cohort and compared emotional language and prevalence ratios of cognitive distortion between cohorts to test our hypotheses. We began by testing associations between variables using the Spearman correlation. Bluesky and Reddit users were kept as separate analytic groups. Cognitive distortion language was noted as present (1) or absent (0), and

user-level analyses were conducted based on the following formula: user-level cognitive distortion prevalence = number of posts containing at least one CDS n-gram/total number of posts.

To quantify uncertainty in bivariate Spearman correlations and account for variance in posting volume across users, we applied bootstrap resampling ($k=10,000$ iterations) to all pairwise correlations as well as the linear model correlations. At each iteration, users were resampled with replacement, and Spearman correlations were computed across all variable pairs simultaneously. Bootstrap means and 95% CIs were derived from the resulting distributions (Tables S1 and S2 in [Multimedia Appendix 1](#)).

To further account for high variation in participant posting volume, we reran Spearman correlations between depression indicators (PHQ-8 total score and diagnosis status) and linguistic markers (CDS prevalence and VADER compound sentiment) across a range of minimum post thresholds (1, 5, 10, 25, and 50 posts per user). Correlations were computed separately for the Bluesky and Reddit samples (Tables S5 and S6 in [Multimedia Appendix 1](#)). Associations between CDS prevalence and VADER components should be interpreted cautiously because both measures were derived from the same text and may partly reflect lexical or mathematical dependence rather than distinct psychological processes.

To address potential confounding by demographic factors, we additionally computed partial Spearman correlations between depression indicators (PHQ-8 score and diagnosis status) and linguistic markers (CDS prevalence and VADER compound sentiment) controlling for age and biological sex. Analyses were conducted separately for Bluesky and Reddit samples. *P* values were corrected for multiple comparisons using the false discovery rate (FDR) correction procedure across all partial correlation tests ($k=8$; Table S7 in [Multimedia Appendix 1](#)).

Linear Modeling of Bluesky User VADER Scores

After computing VADER sentiment scores for each post in the Bluesky dataset, we used linear models to fit each user’s sentiment scores as a function of time, defined as the time elapsed since the user’s first observed post. Models were fit independently for each user using ordinary least squares regression. Only users with at least 2 posts with valid time stamps were included in the analysis. From each model, we extracted the estimated sentiment slope, intercept, and coefficient of determination (R^2) as summary measures of longitudinal sentiment. The slope parameter thus measures within-user trends in affect over time, whereas the intercept gives some indication of baseline affect.

Results

From Bluesky, our final analytic sample comprised 210 users with 117,801 posts. A total of 73.3% (154/210) of these users, with 97,031 posts, reported a depression diagnosis, whereas 26.7% (56/210) of the users, with 20,770 posts,

reported no diagnosis. From Reddit, our final analytic sample comprised 688 users with 261,057 posts. Of these 688 users, 506 (73.5%) with 173,868 posts reported a prior diagnosis of depression, whereas 182 (26.5%) with 87,189 posts reported no depression diagnosis. Demographic characteristics of the analytic sample and descriptive statistics are shown in Table 1. Race and ethnicity data were not collected in the current study. To test our primary hypothesis that there would be

differences in sentiment between groups who had ever had depression and never been diagnosed with depression, we examined the relationships between VADER and depression status and severity using Spearman correlations. Figure 1 and Figure 2 show the Bluesky and Reddit results, respectively, of heat mapped Spearman correlations with 10,000 bootstrap resampling and FDR comparisons applied.

Table 1. Descriptive information of the analytic sample of midlife and older adult Bluesky and Reddit users with and without depression (N=898).

	Bluesky	Reddit
Users, n (%)	210 (23.4)	688 (76.6)
Diagnosed, n (%)	154 (17.1)	506 (56.3)
Posts, mean (SD; range)	560.96 (3263.25; 1-44,697)	379.44 (669.63; 1-3850)
Age (y), mean (SD; range)	57.43 (5.82; 50-80)	56.72 (5.95; 50-80)
Diagnosis age (y), mean (SD; range)	36.10 (12.66; 12-66)	34.33 (12.73; 6-70)
PHQ-8 ^a total score (0-24), mean (SD; range)	8.26 (5.77; 0-24)	8.44 (6.12; 0-24)

^aPHQ-8: Patient Health Questionnaire–8.

Figure 1. Pairwise Spearman correlation between natural language processing indexes and participant features in a sample of Bluesky users (n=210). CDS: cognitive distortion schemata; PHQ-8: Patient Health Questionnaire–8. *Correlations remained significant after false discovery rate correction.

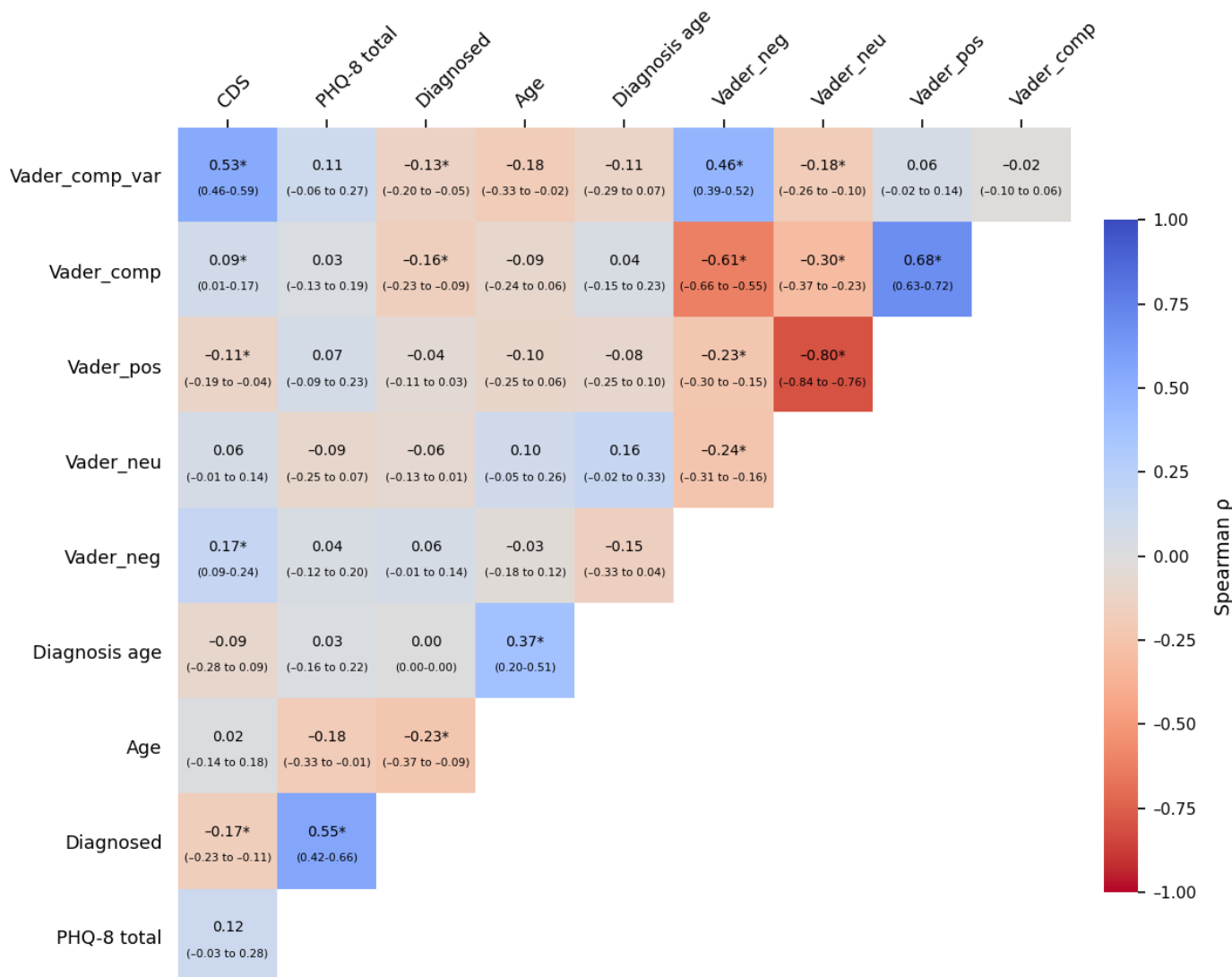
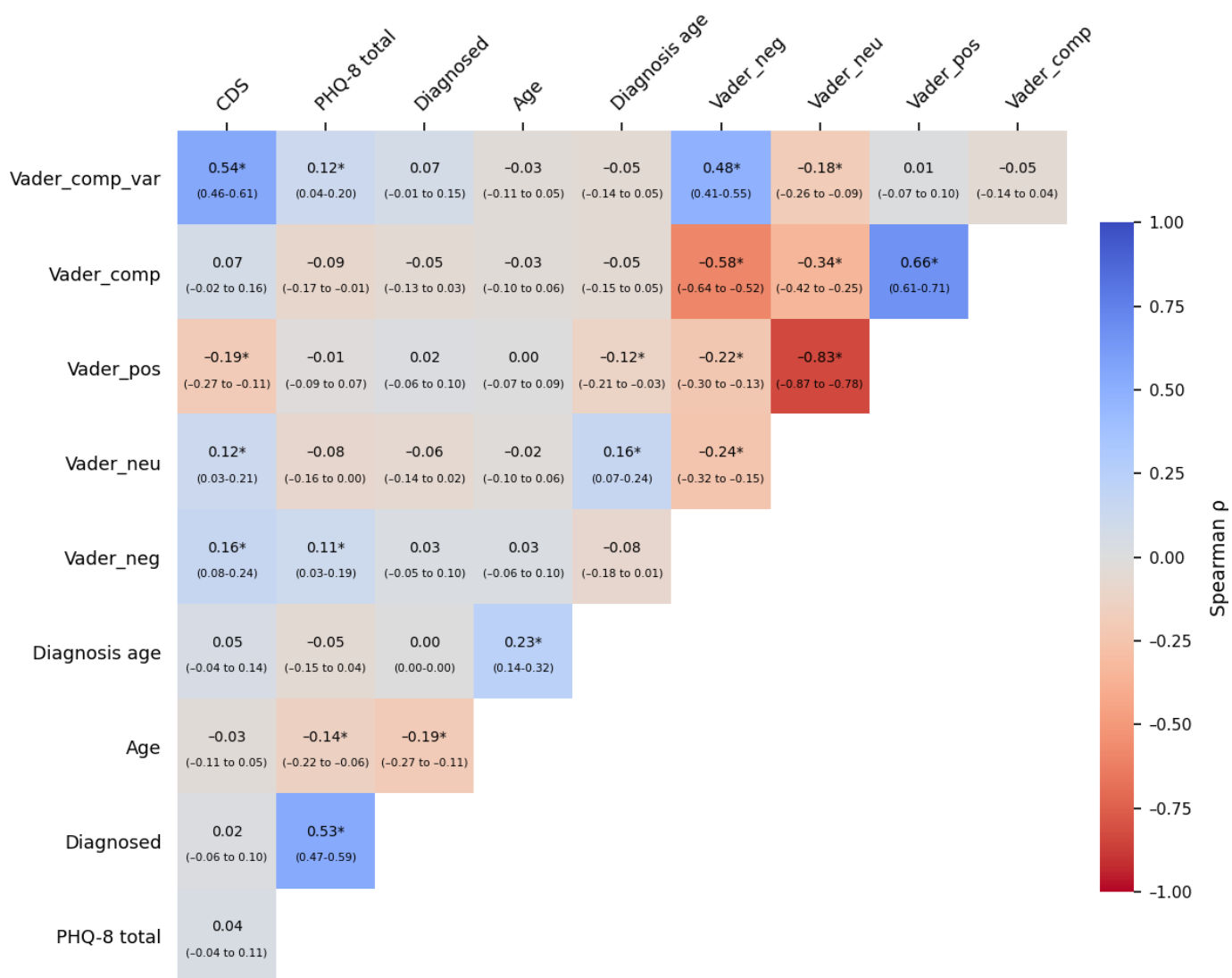


Figure 2. Pairwise Spearman correlation between natural language processing indexes and participant features in a sample of Reddit users (n=688). CDS: cognitive distortion schemata; PHQ-8: Patient Health Questionnaire–8. *Correlations remained significant after false discovery rate correction.



Results showed that, for the Bluesky dataset, there was a significant negative correlation between diagnosis status and VADER compound sentiment score: users who reported a previous diagnosis of depression were found to use more negative sentiment ($\rho=-0.16$; $P<.001$; Figure 1). However, while diagnostic status was associated with more negative overall sentiment, the proportional negative sentiment was not significant. This pattern was expected because the compound score captures the intensity and contextual weighting of negative language, whereas the negative score reflects only the relative frequency of negative tokens and can be diluted by neutral language. In the Reddit dataset, higher PHQ-8 scores were modestly associated with VADER compound sentiment in the unadjusted analysis ($\rho=-0.09$, bootstrap 95% CI -0.17 to -0.01; $P=.03$; Figure 2 and Table S4 in Multimedia Appendix 1); however, this association did not remain significant after FDR correction ($P=.07$) and is

therefore interpreted as suggestive rather than statistically significant.

We next examined the correlations between sentiment and distorted thinking in our sample by conducting Spearman correlations to test the relationships between VADER scores and cognitive distortion use. The results were significant for positive and negative VADER scores on both Reddit and Bluesky and are shown in Table 2. Unexpectedly, on Bluesky, depression status was associated with lower CDS prevalence ($\rho=-0.17$; $P<.001$), which remained consistent across minimum post thresholds but was not significant after controlling for age and sex (Table S1 in Multimedia Appendix 1). For Reddit, CDS prevalence was not associated with diagnostic status ($\rho=0.02$; $P=.64$). CDS prevalence was not associated with PHQ-8 scores on either platform.

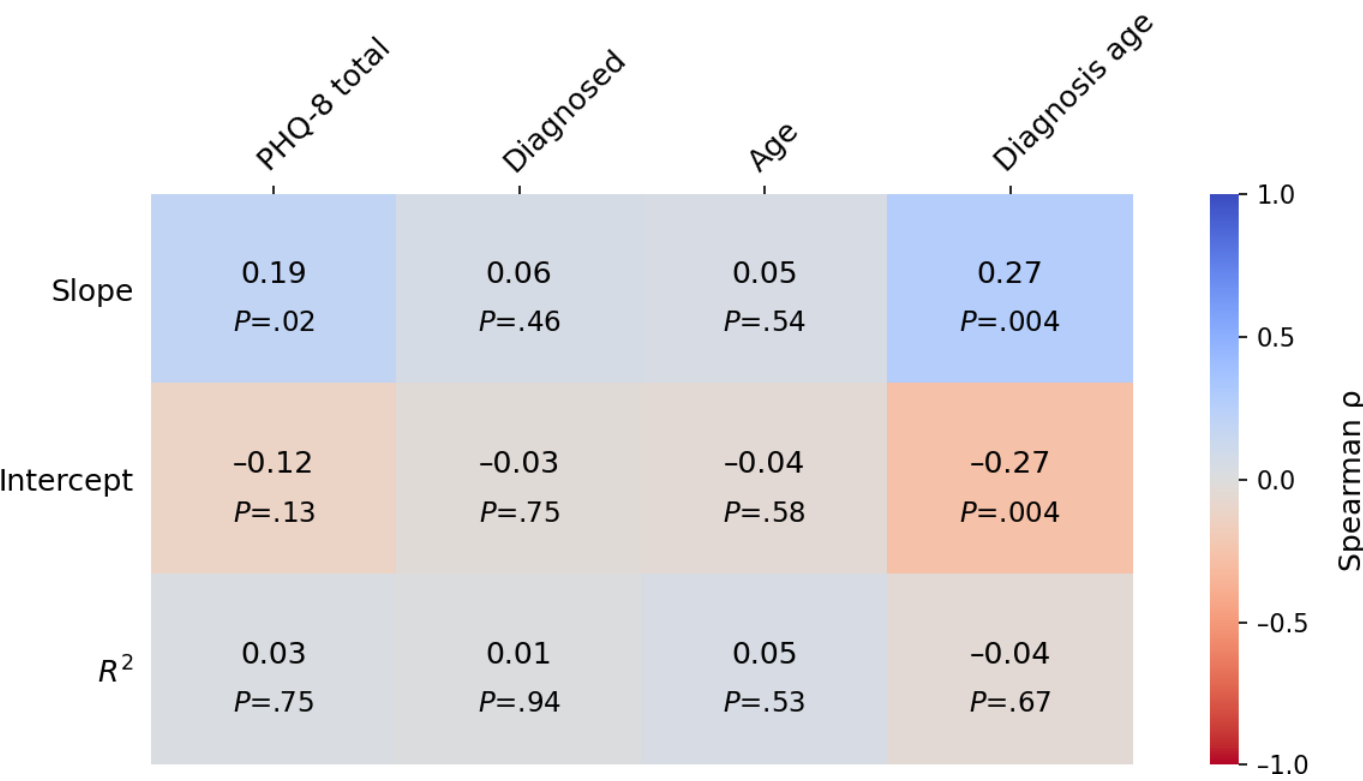
Table 2. Associations between sentiment (Valence Aware Dictionary and Sentiment Reasoner) and cognitive distortions in a sample of midlife and older adult social media users by platform sampled (N=898).

	Composite		Positive sentiment		Neutral sentiment		Negative sentiment	
	Spearman ρ	<i>P</i> value	Spearman ρ	<i>P</i> value	Spearman ρ	<i>P</i> value	Spearman ρ	<i>P</i> value
Reddit	−0.01	.82	−0.18	<.001	0.09	.03	0.17	<.001
Bluesky	0.00	.94	−0.12	<.001	0.04	.22	0.17	<.001

To characterize longitudinal sentiment on Bluesky, we fit user-specific linear models of VADER sentiment scores as a function of time and examined associations between extracted model parameters and clinical and demographic variables (Figure 3). Comparable time-stamped data were not available for Reddit text files in the current study, and thus, longitudinal models could not be estimated for Reddit. For Bluesky, sentiment slopes were significantly associated with depressive symptom severity as measured using the PHQ-8 ($\rho=0.19$, bootstrap 95% CI 0.03-0.34; $P=.02$), indicating that individuals with higher depression symptom burden exhibited modestly more negative sentiment trajectories over time; however, this association did not remain statistically significant after FDR correction ($P=.07$) and should therefore be interpreted as suggestive

and requiring replication in a larger sample. Sentiment slopes were also positively associated with age at depression diagnosis ($\rho=0.27$; $P=.004$). In contrast, baseline affect (intercept) was negatively associated with age at diagnosis ($\rho=-0.27$; $P=.004$), suggesting lower average sentiment among individuals diagnosed later in life. Diagnosis status itself was not significantly associated with sentiment slope ($P=.46$), intercept ($P=.75$), or model fit ($P=.94$). Coefficients of determination (R^2) were small (Table S3 in Multimedia Appendix 1). Together, these findings suggest that longitudinal sentiment dynamics captured from social media are weakly related to current symptom severity and timing of diagnosis but not to diagnostic status per se, underscoring the importance of modeling sentiment patterns rather than relying on static clinical labels.

Figure 3. Spearman correlations between per-user linear model estimates of Valence Aware Dictionary and Sentiment Reasoner sentiment over time and participant depression-related features. PHQ-8: Patient Health Questionnaire–8.



Discussion

The online language of midlife and older adults on social media can offer insights into their health and social relationships, which are both important aspects of healthy aging [22,23]. In the current study, using the online language of midlife and older adults on Reddit and Bluesky, we compared a sample who had ever been diagnosed with depression to a never-diagnosed comparison sample in terms of VADER

sentiment scores and prevalence of cognitive distortions across timelines. Results showed that, for Bluesky users in our sample, ever having had a depression diagnosis was associated with more negative sentiment compared to users who had never had depression. In Reddit data, depression severity but not diagnosis showed a small association with more negative sentiment in the unadjusted analysis, but this did not remain statistically significant after FDR correction. To the best of our knowledge, this is the first study to use

Bluesky and Reddit to examine the emotional content of online language of adults aged 50 years and older using NLP in the context of depression. The results have implications for detecting linguistic features of depression and uncovering links among depression, language, and cognition in midlife and older adults.

Building on these findings, and given the associations among diagnosis age, symptom severity, and sentiment, we conducted additional exploratory analyses to examine whether longitudinal patterns of sentiment provided complementary information beyond cross-sectional sentiment measures. These analyses of longitudinal sentiment trajectories on Bluesky revealed that changes in sentiment over time were associated with current depressive symptom severity and age at diagnosis but not with diagnostic status itself; however, these results for depressive symptom severity did not remain statistically significant after FDR correction, whereas the associations with age at time of diagnosis remained significant. In particular, later age of depression diagnosis was associated with both lower baseline sentiment and steeper sentiment change over time, whereas binary diagnostic status showed little correspondence with sentiment dynamics. These exploratory findings suggest that longitudinal patterns of sentiment may capture depression heterogeneity that is not reflected by symptom measures or diagnostic labels alone [32,33]. Although linear trends explained limited variance in within-person sentiment, the pattern we observed is consistent with the episodic and context-dependent nature of sentiment in naturalistic social media data and underscores the importance of modeling trajectories rather than relying solely on aggregate sentiment measures [34].

Our findings also showed differences between platforms that should be considered for future research. Bluesky, which offers a microblogging format where individuals may post their beliefs and interact with other users, yielded slightly different patterns from those of Reddit, which offers a long-form, topic-centered discussion forum. Prior research has shown that the format of social media impacts how much time people spend on it, and individuals can become addicted to social networking platforms [35-38]. Prior work also shows a significant relationship between time spent on social media and declining mental health [39], although most of this work has been conducted in adolescent samples. Furthermore, research shows that negative thinking is contagious [40,41]. The dynamic relationships among distorted content, contagious negative thinking, and depression symptoms remain to be tested in midlife and older adults across different social media platforms.

On Bluesky, users who had ever had depression showed lower VADER compound sentiment and lower CDS prevalence than users who had never had depression in unadjusted analyses. However, the CDS association did not remain significant after adjustment for age and sex. On Reddit, depression diagnosis was not associated with CDS prevalence. The platform-specific CDS findings may reflect differences in platform structure, language use, or sample composition. First, most work on social media involves young adults and not adults aged 50 years and older, whose

online language use and disclosure may be systematically different from those of young people. For example, in one study using MySpace comparing teenagers (aged 13-19 years) to older adults (aged ≥ 60 years), teenagers used more self-references and negative emotions [42]. However, results likely vary by platform. Next, there are sampling challenges in recruiting adults who use social media and are willing to share their data, which leads to self-selected groups of people with more symptoms or who more willingly self-disclose their social media timelines [43,44]. Self-disclosure on social media is widely studied among adolescents and young people, but less is known about self-disclosure among midlife and older adults [45,46]. While some studies have been conducted, little is known about how midlife and older adults express psychological distress online or how their online language can serve as an early warning sign of risk. Some recent work suggests that there is a positive association between internet use and mental health in midlife and older adults [47], but more work is needed to understand this.

This work has broader implications for cognitive aging research in the digital age. Epidemiological studies suggest an association between depression and Alzheimer disease and related dementias (ADRD) [48]. Even earlier-life depression increases risk of dementia later in life [49]. Unrecognized or untreated depression in older adults is an urgent public health concern and critical intervention target. An initial step in examining the relationship between depression and ADRD is to understand depression in midlife and older adults using digital tools to identify and quantify depression risk factors. Digital language markers may offer a scalable and low-burden tool to characterize depressive symptoms in midlife and older adults and potentially identify early vulnerability to cognitive decline. However, the clinical validity and predictive utility of these digital language markers require additional study. The current work provides a step toward integrating digital phenotyping with established epidemiological and clinical approaches to an advanced understanding of the relationship between depression and ADRD.

Despite its strengths, including a novel focus on sentiment and cognitive distortions in adults 50 years and older and a vetted, theory-driven approach for cognitive distortion detection, there are limitations to the current work. First, the samples were small in comparison to our prior work with Twitter and large-scale adolescent studies, so further work is needed to generalize our findings. Second, our sample was obtained via a self-report survey where individuals disclosed a diagnosis of depression and reported on their current symptoms of depression. Thus, there is a sampling bias [50] whereby we only recruited users who were willing to disclose their diagnostic status and share their social media timelines. Moreover, some participants provided handles for both platforms and, therefore, contributed language to both the Reddit and Bluesky datasets. Because analyses were conducted separately within each platform, this is unlikely to impact our results, but the overlap of platform samples may reduce the independence of our comparisons across platforms and should be considered when interpreting differences between Reddit and Bluesky. Additionally, because

time-stamped post-level data were unavailable for Reddit at the time of analysis, longitudinal sentiment analysis could only be conducted for Bluesky, which limited our ability to compare temporal patterns in sentiment across platforms.

Future work should prioritize several directions. First, linking social media timelines to independently assessed clinical data would allow for stronger validation of linguistic markers of depression in older adults. Clinical data could include clinician-verified depression diagnoses, symptom history, treatment history, and cognitive performance. Linking social media timelines to clinical data could help determine whether language-based markers correspond to clinical indicators. Second, studying variation in the language used by individuals online could help understand cognitive decline. By integrating NLP and deep learning, this work could test whether language-based markers (cognitive distortions and perplexity) are enriched among older adults at elevated depression risk, positioning them as candidate indicators of cognitive vulnerability for future validation [51].

Still, there are many trade-offs between machine learning and deep learning for mental illness detection on social media, including accuracy, interpretability, and computational efficiency [52]. We believe that the current approach is innovative and scalable and leverages naturally occurring online data to identify early warning signs in at-risk populations.

In summary, the current study provides a proof of concept for using social media language from Bluesky and Reddit to detect language sentiment and cognitive distortions in midlife and older adults, a group that remains comparatively underrepresented in digital mental health research. This study lays the groundwork for future work that links online behavior to cognitive and clinical outcomes. Our findings support further study of how sentiment in online language relates to depression history and current depressive symptoms among adults aged 50 years and older. Larger longitudinal samples are needed before these features can be considered for detection or prediction.

Acknowledgments

The authors thank Harsh Patel, Vihaan Shah, and Suraj Muthyam for their assistance at the early stages of this project. During the revision of this manuscript, the authors used generative AI tools (ChatGPT; OpenAI) for editorial assistance and organization of revision materials. Generative AI was not used to generate data, conduct statistical analyses, or produce study results. The authors reviewed and edited all AI outputs and take full responsibility for the content of the manuscript.

Funding

This work is supported by the National Institute on Aging (K01 AG086598 to LR).

Data Availability

Analysis code will be made available on GitHub [53]. Deidentified data may be available from the corresponding author on reasonable request, subject to institutional approvals and data sharing agreements.

Authors' Contributions

Conceptualization: LR

Formal analysis: LR, AE, MtT

Funding acquisition: LR

Investigation: LR

Methodology: LR

Project administration: LR

Software: AE, MtT

Supervision: LR

Visualization: AE, MtT

Writing—original draft: LR

Writing—review and editing: LR, AE, MtT, LL-L, DV, JB

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplemental materials.

[DOCX File (Microsoft Word File), 45 KB-Multimedia Appendix 1]

References

1. ten Thij M, Bathina K, Rutter LA, et al. Depression alters the circadian pattern of online activity. *Sci Rep*. Oct 14, 2020;10(1):17272. [doi: [10.1038/s41598-020-74314-3](https://doi.org/10.1038/s41598-020-74314-3)] [Medline: [33057099](https://pubmed.ncbi.nlm.nih.gov/33057099/)]
2. Riley TN, Thompson HM, Howard J, Lorenzo-Luaces L, Rutter LA. Seeking connectedness through social media use: associations with adolescent empathic understanding and perspective-taking. *Curr Psychol*. Dec 2023;42:31227-31239. [doi: [10.1007/s12144-022-04096-6](https://doi.org/10.1007/s12144-022-04096-6)]

3. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. *Curr Opin Behav Sci*. 2017;18:43-49. [doi: [10.1016/j.cobeha.2017.07.005](https://doi.org/10.1016/j.cobeha.2017.07.005)]
4. Rutter LA, ten Thij M, Lorenzo-Luaces L, Valdez D, Bollen J. Negative affect variability differs between anxiety and depression on social media. *PLoS One*. 2024;19(2):e0272107. [doi: [10.1371/journal.pone.0272107](https://doi.org/10.1371/journal.pone.0272107)] [Medline: [38381769](https://pubmed.ncbi.nlm.nih.gov/38381769/)]
5. Kelley SW, Gillan CM. Using language in social media posts to study the network dynamics of depression longitudinally. *Nat Commun*. Feb 15, 2022;13(1):870. [doi: [10.1038/s41467-022-28513-3](https://doi.org/10.1038/s41467-022-28513-3)] [Medline: [35169166](https://pubmed.ncbi.nlm.nih.gov/35169166/)]
6. Riehm KE, Feder KA, Tormohlen KN, et al. Associations between time spent using social media and internalizing and externalizing problems among US youth. *JAMA Psychiatry*. Dec 1, 2019;76(12):1266-1273. [doi: [10.1001/jamapsychiatry.2019.2325](https://doi.org/10.1001/jamapsychiatry.2019.2325)] [Medline: [31509167](https://pubmed.ncbi.nlm.nih.gov/31509167/)]
7. Share of those 65 and older who are tech users has grown in the past decade. Pew Research Center. 2022. URL: <https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/> [Accessed 2023-05-27]
8. McClain C, Bishop W. What we know about internet use, smartphone ownership and digital divides in the U.S. Pew Research Center. 2026. URL: https://www.pewresearch.org/short-reads/2026/01/08/internet-use-smartphone-ownership-digital-divides-in-u-s/?utm_source=chatgpt.com [Accessed 2026-06-23]
9. Cotten SR, Schuster AM, Seifert A. Social media use and well-being among older adults. *Curr Opin Psychol*. Jun 2022;45:101293. [doi: [10.1016/j.copsyc.2021.12.005](https://doi.org/10.1016/j.copsyc.2021.12.005)] [Medline: [35065352](https://pubmed.ncbi.nlm.nih.gov/35065352/)]
10. Lewin KM, Meshi D, Schuster AM, Cotten SR. Active and passive social media use are differentially related to depressive symptoms in older adults. *Aging Ment Health*. Jan 2023;27(1):176-183. [doi: [10.1080/13607863.2022.2068133](https://doi.org/10.1080/13607863.2022.2068133)] [Medline: [35470731](https://pubmed.ncbi.nlm.nih.gov/35470731/)]
11. Busch PA, Hausvik GI, Ropstad OK, Pettersen D. Smartphone usage among older adults. *Comput Human Behav*. Aug 2021;121:106783. [doi: [10.1016/j.chb.2021.106783](https://doi.org/10.1016/j.chb.2021.106783)]
12. Guzman AA, Brecht ML, Doering LV, Macey PM, Menten JC. Social media use and depression in older adults: a systematic review. *Res Gerontol Nurs*. 2023;16(2):97-104. [doi: [10.3928/19404921-20230220-05](https://doi.org/10.3928/19404921-20230220-05)] [Medline: [36944173](https://pubmed.ncbi.nlm.nih.gov/36944173/)]
13. Meshi D, Cotten SR, Bender AR. Problematic social media use and perceived social isolation in older adults: a cross-sectional study. *Gerontology*. 2020;66(2):160-168. [doi: [10.1159/000502577](https://doi.org/10.1159/000502577)] [Medline: [31522178](https://pubmed.ncbi.nlm.nih.gov/31522178/)]
14. Beneito-Montagut R, Rosales A, Fernández-Ardèvol M. Emerging digital inequalities: a comparative study of older adults' smartphone use. *Soc Media Soc*. 2022;8(4). [doi: [10.1177/20563051221138756](https://doi.org/10.1177/20563051221138756)]
15. Balani S, De Choudhury M. Detecting and characterizing mental health related self-disclosure in social media. In: CHI EA '15: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. Association for Computing Machinery; 2015. [doi: [10.1145/2702613.2732733](https://doi.org/10.1145/2702613.2732733)]
16. Chu TH, Sun M, Crystal Jiang L. Self-disclosure in social media and psychological well-being: a meta-analysis. *J Soc Pers Relat*. 2023;40(2):576-599. [doi: [10.1177/02654075221119429](https://doi.org/10.1177/02654075221119429)]
17. Luo M, Hancock JT. Self-disclosure and social media: motivations, mechanisms and psychological well-being. *Curr Opin Psychol*. Feb 2020;31:110-115. [doi: [10.1016/j.copsyc.2019.08.019](https://doi.org/10.1016/j.copsyc.2019.08.019)] [Medline: [31563107](https://pubmed.ncbi.nlm.nih.gov/31563107/)]
18. Lavretsky H, Kumar A. Clinically significant non-major depression: old concepts, new insights. *Am J Geriatr Psychiatry*. 2002;10(3):239-255. [Medline: [11994211](https://pubmed.ncbi.nlm.nih.gov/11994211/)]
19. Alexopoulos GS, Borson S, Cuthbert BN, et al. Assessment of late life depression. *Biol Psychiatry*. Aug 1, 2002;52(3):164-174. [doi: [10.1016/s0006-3223\(02\)01381-1](https://doi.org/10.1016/s0006-3223(02)01381-1)] [Medline: [12182923](https://pubmed.ncbi.nlm.nih.gov/12182923/)]
20. Bathina KC, ten Thij M, Lorenzo-Luaces L, Rutter LA, Bollen J. Individuals with depression express more distorted thinking on social media. *Nat Hum Behav*. Apr 2021;5(4):458-466. [doi: [10.1038/s41562-021-01050-7](https://doi.org/10.1038/s41562-021-01050-7)] [Medline: [33574604](https://pubmed.ncbi.nlm.nih.gov/33574604/)]
21. Beck AT, Haigh EA. Advances in cognitive theory and therapy: the generic cognitive model. *Annu Rev Clin Psychol*. 2014;10:1-24. [doi: [10.1146/annurev-clinpsy-032813-153734](https://doi.org/10.1146/annurev-clinpsy-032813-153734)] [Medline: [24387236](https://pubmed.ncbi.nlm.nih.gov/24387236/)]
22. Bollen J, ten Thij M, Breithaupt F, et al. Historical language records reveal a surge of cognitive distortions in recent decades. *Proc Natl Acad Sci U S A*. Jul 27, 2021;118(30):e2102061118. [doi: [10.1073/pnas.2102061118](https://doi.org/10.1073/pnas.2102061118)] [Medline: [34301899](https://pubmed.ncbi.nlm.nih.gov/34301899/)]
23. Bollen J, ten Thij M, Breithaupt F, et al. Reply to Schmidt et al.: a robust surge of cognitive distortions in historical language. *Proc Natl Acad Sci U S A*. Nov 9, 2021;118(45):e2115842118. [doi: [10.1073/pnas.2115842118](https://doi.org/10.1073/pnas.2115842118)] [Medline: [34725168](https://pubmed.ncbi.nlm.nih.gov/34725168/)]
24. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*. Sep 30, 2011;333(6051):1878-1881. [doi: [10.1126/science.1202775](https://doi.org/10.1126/science.1202775)] [Medline: [21960633](https://pubmed.ncbi.nlm.nih.gov/21960633/)]
25. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. *Proc Int AAAI Conf Web Soc Media*. 2021;7(1):128-137. [doi: [10.1609/icwsm.v7i1.14432](https://doi.org/10.1609/icwsm.v7i1.14432)]

26. Tauscher JS, Lybarger K, Ding X, et al. Automated detection of cognitive distortions in text exchanges between clinicians and people with serious mental illness. *Psychiatr Serv*. Apr 1, 2023;74(4):407-410. [doi: [10.1176/appi.ps.202100692](https://doi.org/10.1176/appi.ps.202100692)] [Medline: [36164769](https://pubmed.ncbi.nlm.nih.gov/36164769/)]
27. Al-Mosaiwi M, Johnstone T. In an absolute state: elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clin Psychol Sci*. Jul 2018;6(4):529-542. [doi: [10.1177/2167702617747074](https://doi.org/10.1177/2167702617747074)] [Medline: [30886766](https://pubmed.ncbi.nlm.nih.gov/30886766/)]
28. Bollen J, Gonçalves B, van de Leemput I, Ruan G. The happiness paradox: your friends are happier than you. *EPJ Data Sci*. Dec 2017;6:4. [doi: [10.1140/epjds/s13688-017-0100-1](https://doi.org/10.1140/epjds/s13688-017-0100-1)]
29. Rutter LA, Edinger A, Lorenzo-Luaces L, Thij MT, Valdez D, Bollen J. Anxiety and depression are associated with more distorted thinking on social media: a longitudinal multi-method study. *Cognit Ther Res*. Aug 2025;49(4):712-720. [doi: [10.1007/s10608-025-10580-7](https://doi.org/10.1007/s10608-025-10580-7)] [Medline: [40874036](https://pubmed.ncbi.nlm.nih.gov/40874036/)]
30. Hutto C, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. *Proc Int AAAI Conf Web Soc Media*. 2014;8(1):216-225. [doi: [10.1609/icwsm.v8i1.14550](https://doi.org/10.1609/icwsm.v8i1.14550)]
31. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. *J Affect Disord*. Apr 2009;114(1-3):163-173. [doi: [10.1016/j.jad.2008.06.026](https://doi.org/10.1016/j.jad.2008.06.026)] [Medline: [18752852](https://pubmed.ncbi.nlm.nih.gov/18752852/)]
32. Hamilton LJ, Lakhan P, Rutter LA. A latent variable approach to affect variability in daily life accurately predicts psychopathology, especially depression symptoms in a non-clinical sample. *J Emot Psychopathol*. Apr 24, 2025;1(2):569-588. [doi: [10.55913/joep.v1i2.82](https://doi.org/10.55913/joep.v1i2.82)] [Medline: [40370940](https://pubmed.ncbi.nlm.nih.gov/40370940/)]
33. Goldberg D. The heterogeneity of "major depression". *World Psychiatry*. Oct 2011;10(3):226-228. [doi: [10.1002/j.2051-5545.2011.tb00061.x](https://doi.org/10.1002/j.2051-5545.2011.tb00061.x)] [Medline: [21991283](https://pubmed.ncbi.nlm.nih.gov/21991283/)]
34. Panaite V, Rottenberg J, Bylsma LM. Daily affective dynamics predict depression symptom trajectories among adults with major and minor depression. *Affect Sci*. 2020;1(3):186-198. [doi: [10.1007/s42761-020-00014-w](https://doi.org/10.1007/s42761-020-00014-w)] [Medline: [36043207](https://pubmed.ncbi.nlm.nih.gov/36043207/)]
35. Xie J, Gallo K. Is TikTok more addictive than other social media platforms: perception versus reality. *Am J Soc Sci Hum*. 2024;9(2):38-48. [doi: [10.55284/ajssh.v9i2.1147](https://doi.org/10.55284/ajssh.v9i2.1147)]
36. Aichner T, Grünfelder M, Maurer O, Jegeni D. Twenty-five years of social media: a review of social media applications and definitions from 1994 to 2019. *Cyberpsychol Behav Soc Netw*. Apr 2021;24(4):215-222. [doi: [10.1089/cyber.2020.0134](https://doi.org/10.1089/cyber.2020.0134)] [Medline: [33847527](https://pubmed.ncbi.nlm.nih.gov/33847527/)]
37. Cotten SR. Technologies and aging: understanding use, impacts, and future needs. In: *Handbook of Aging and the Social Sciences*. Academic Press; 2021:373-392. [doi: [10.1016/B978-0-12-815970-5.00023-1](https://doi.org/10.1016/B978-0-12-815970-5.00023-1)]
38. Kuss DJ, Griffiths MD. Social networking sites and addiction: ten lessons learned. *Int J Environ Res Public Health*. Mar 17, 2017;14(3):311. [doi: [10.3390/ijerph14030311](https://doi.org/10.3390/ijerph14030311)] [Medline: [28304359](https://pubmed.ncbi.nlm.nih.gov/28304359/)]
39. Naslund JA, Bondre A, Torous J, Aschbrenner KA. Social media and mental health: benefits, risks, and opportunities for research and practice. *J Technol Behav Sci*. Sep 2020;5(3):245-257. [doi: [10.1007/s41347-020-00134-x](https://doi.org/10.1007/s41347-020-00134-x)] [Medline: [33415185](https://pubmed.ncbi.nlm.nih.gov/33415185/)]
40. Rozin P, Royzman EB. Negativity bias, negativity dominance, and contagion. *Pers Soc Psychol Rev*. Nov 2001;5(4):296-320. [doi: [10.1207/S15327957PSPR0504_2](https://doi.org/10.1207/S15327957PSPR0504_2)]
41. Haefl GJ, Hames JL. Cognitive vulnerability to depression can be contagious. *Clin Psychol Sci*. 2014;2(1):75-85. [doi: [10.1177/2167702613485075](https://doi.org/10.1177/2167702613485075)]
42. Pfeil U, Arjan R, Zaphiris P. Age differences in online social networking – a study of user profiles and the social capital divide among teenagers and older users in MySpace. *Comput Human Behav*. May 2009;25(3):643-654. [doi: [10.1016/j.chb.2008.08.015](https://doi.org/10.1016/j.chb.2008.08.015)]
43. Rutter LA, Howard J, Lakhan P, Valdez D, Bollen J, Lorenzo-Luaces L. "I haven't been diagnosed, but I should be" - insight into self-diagnoses of common mental health disorders: cross-sectional study. *JMIR Form Res*. Jan 13, 2023;7:e39206. [doi: [10.2196/39206](https://doi.org/10.2196/39206)] [Medline: [36637885](https://pubmed.ncbi.nlm.nih.gov/36637885/)]
44. Lorenzo-Luaces L, Howard J, Edinger A, et al. Sociodemographics and transdiagnostic mental health symptoms in SOCIAL (Studies of Online Cohorts for Internalizing Symptoms and Language) I and II: cross-sectional survey and botometer analysis. *JMIR Form Res*. Oct 20, 2022;6(10):e39324. [doi: [10.2196/39324](https://doi.org/10.2196/39324)] [Medline: [36264616](https://pubmed.ncbi.nlm.nih.gov/36264616/)]
45. Kohl G, Koh WQ, Scior K, Charlesworth G. Self-disclosure and social media use among younger and older people with dementia: an internet-mediated mixed-methods study. *Int J Hum Comput Interact*. 2024;40(22):7364-7374. [doi: [10.1080/10447318.2023.2265728](https://doi.org/10.1080/10447318.2023.2265728)]
46. Xu X, Zhao Y, Zhu Q. The effect of social media use on older adults' loneliness-the moderating role of self-disclosure. In: Gao Q, Zhou J, editors. *Human Aspects of IT for the Aged Population*. Springer; 2020. [doi: [10.1007/978-3-030-50232-4_10](https://doi.org/10.1007/978-3-030-50232-4_10)]

47. Luo Y, Yip PS, Zhang Q. Positive association between internet use and mental health among adults aged ≥ 50 years in 23 countries. *Nat Hum Behav*. Jan 2025;9(1):90-100. [doi: [10.1038/s41562-024-02048-7](https://doi.org/10.1038/s41562-024-02048-7)] [Medline: [39558112](https://pubmed.ncbi.nlm.nih.gov/39558112/)]
48. Tsuno N, Homma A. What is the association between depression and Alzheimer's disease? *Expert Rev Neurother*. Nov 2009;9(11):1667-1676. [doi: [10.1586/ern.09.106](https://doi.org/10.1586/ern.09.106)] [Medline: [19903025](https://pubmed.ncbi.nlm.nih.gov/19903025/)]
49. Byers AL, Yaffe K. Depression and risk of developing dementia. *Nat Rev Neurol*. May 3, 2011;7(6):323-331. [doi: [10.1038/nrneurol.2011.60](https://doi.org/10.1038/nrneurol.2011.60)] [Medline: [21537355](https://pubmed.ncbi.nlm.nih.gov/21537355/)]
50. Cao Y, Dai J, Wang Z, et al. Machine learning approaches for depression detection on social media: a systematic review of biases and methodological challenges. *J Behav Data Sci*. 2025;5(1):67-102. [doi: [10.35566/jbds/caoyc](https://doi.org/10.35566/jbds/caoyc)]
51. Frankenberg C, Weiner J, Schultz T, et al. Perplexity – a new predictor of cognitive changes in spoken language? – Results of the Interdisciplinary Longitudinal Study on Adult Development and Aging (ILSE). *Linguist Vanguard*. 2019;5(s2). [doi: [10.1515/lingvan-2018-0026](https://doi.org/10.1515/lingvan-2018-0026)]
52. Ding Z, Wang Z, Zhang Y, et al. Trade-offs between machine learning and deep learning for mental illness detection on social media. *Sci Rep*. 2025;15:14497. [doi: [10.1038/s41598-025-99167-6](https://doi.org/10.1038/s41598-025-99167-6)]
53. Edinger A. aedinger7. GitHub. URL: <https://github.com/aedinger7> [Accessed 2026-08-19]

Abbreviations

ADRD: Alzheimer disease and related dementias
CDS: cognitive distortion schemata
FDR: false discovery rate
NLP: natural language processing
PHQ-8: Patient Health Questionnaire–8
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by Matthew Balcarras; peer-reviewed by David Owen, Yuchen Cao; submitted 10.Mar.2026; final revised version received 22.Jul.2026; accepted 30.Jul.2026; published 26.Aug.2026

Please cite as:

Rutter L, Edinger A, ten Thij M, Lorenzo-Luaces L, Valdez D, Bollen J
Naturalistic Online Language as a Marker of Depression in Midlife and Older Adults: Computational Text Analysis of Bluesky and Reddit Posts
J Med Internet Res 2026;28:e95023
URL: <https://www.jmir.org/2026/1/e95023>
doi: [10.2196/95023](https://doi.org/10.2196/95023)

© Lauren Rutter, Andy Edinger, Marijn ten Thij, Lorenzo Lorenzo-Luaces, Danny Valdez, Johan Bollen. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 26.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.