The Hidden Toll of Social Media News on Psychosocial Wellbeing

Anonymous ACL submission

Abstract

News consumption on social media has become ubiquitous, yet how different forms of engagement shape psychosocial outcomes remains unclear. To address this gap, we leveraged a large-scale dataset of ∼26M posts and ∼45M comments on the *Bluesky* platform, and conducted a quasi-experimental study, matching 81,345 *Treated* users exposed to `News` feeds with 83,711 *Control* users using stratified propensity score analysis. We examined psychosocial wellbeing, in terms of *affective*, *behavioral*, and *cognitive* outcomes. Our findings reveal that news engagement produces systematic trade-offs: increased depression, stress, and anxiety, yet decreased loneliness and increased social interaction on the platform. Regression models reveal that `News` feed bookmarking is associated with greater psychosocial deterioration compared to commenting or quoting, with magnitude differences exceeding tenfold. These per-engagement effects accumulate with repeated exposure, showing significant psychosocial impacts. Our work extends theories of news effects beyond crisis-centric frameworks by demonstrating that routine consumption creates distinct psychological dynamics depending on engagement type, and bears implications for tools and interventions for mitigating the psychosocial costs of news consumption on social media.

1 Introduction

News occupies a distinct position within social media ecosystems, serving as an emotionally salient and socially consequential form of content that is encountered routinely and often incidentally as part of everyday online activity. News consumption has become a central form of engagement on social media platforms. Rather than occurring during isolated moments of intentional information seeking, news exposure now unfolds as part of everyday online activity. In 2025, more than half of U.S. adults (53%) reported that they at least sometimes get news from social media (Pew Research Center, 2025). These platforms further transform news consumption by enabling users to engage with news content in interactive ways through liking, commenting, and sharing news content (Hermida, 2010). As a result, users do not merely consume news, but actively shape their informational environments through repeated and varied forms of engagement. This shift raises important questions about how different modes and intensities of news engagement affect users' wellbeing.

In parallel, prior work has studied the psychosocial effects of news exposure—reflected in changes to emotional states, stress-related behaviors, and cognitive expressions—primarily, in the context of crises. For example, a survey of 2,251 adults found that more frequent information seeking about COVID-19 across television, newspapers, and social media was associated with higher levels of emotional distress (Hwang et al., 2021). Similar patterns have been observed following mass shootings, terrorist attacks, and natural disasters (Silver et al., 2013; Strasser et al., 2022). Thompson et al. further showed that psychological distress can drive increased news consumption, reinforcing negative outcomes (Thompson et al., 2019).

Notably, crisis-driven events account for only a fraction of the news users encounter online; everyday social media use involves ongoing exposure to political, economic, and local news (Boczkowski et al., 2018). Such sustained news consumption can contribute to emotional exhaustion and reduced wellbeing, with news avoidance emerging as a common coping response (McLaughlin et al., 2023; Soroka and McAdams, 2015; Boukes and Vliegenthart, 2017). These experiences are often termed in popular discourse as "headline stress disorder," which may be more pronounced on social media where news is encountered repeatedly and incidentally within everyday feeds (Huff, 2022).

Understanding the effects of social media news engagement is critical, as routine and repeated exposure may accumulate psychological effects distinct from crisis-driven media exposure and general social media use (Silver et al., 2013; Soroka and McAdams, 2015). Isolating these effects is essential for advancing media-effects theory and informing platform design and policy interventions aimed at mitigating unintended psychological con-

sequences (Thorson and Wells, 2016; Newman et al., 2025; Huff, 2022).

Existing research on news consumption and mental health has largely relied on survey-based studies and laboratory experiments (Boukes and Vliegenthart, 2017; Strasser et al., 2022). While informative, these approaches are limited in social media contexts: surveys are retrospective and susceptible to recall and subjective bias (Tourangeau et al., 2000), and experimental designs often capture short-term responses under controlled conditions that may not reflect everyday patterns of news engagement online. Consequently, it remains unclear how news engagement shapes psychosocial outcomes over extended periods across large populations in naturalistic settings, and whether these effects vary by the type and intensity of engagement. This challenge is compounded by the difficulty of observing users early in their engagement trajectories, as most platforms capture individuals only after years of accumulated habits and algorithmic personalization (Thorson and Wells, 2016). Without such longitudinal examinations, existing studies risk conflating the effects of news exposure with broader platform use, obscuring the mechanisms through which specific engagement modalities influence wellbeing.

To address these gaps, we conducted a large-scale quasi-experimental study of news engagement on *Bluesky*—a decentralized (Twitter-like) social media platform that was launched publicly in February 2024. *Bluesky* provides a unique opportunity to examine news engagement from its early stages, allowing observation of users as they develop engagement practices from the ground up (Failla and Rossetti, 2024). Unlike established platforms where algorithmic news exposure is often implicit, *Bluesky* users explicitly choose whether to bookmark algorithmic feeds, including a verified `News` feed that aggregates headlines from established news organizations. This paper is guided by the following research questions (RQs):

RQ1: How does news exposure on social media affect psychosocial wellbeing?

RQ2: How are the types and intensity of news engagement on social media associated with psychosocial outcomes?

To address our RQs, we adopted a causal-inference design, based on the potential outcomes framework (Rubin, 2005), on a large-scale dataset on *Bluesky*. We identified a *Treated* group consisting of users who engaged with the `News` feed through bookmarking, liking, commenting, reposting, quoting, or creating original posts. We compared these users to stratified propensity-score-matched *Control* users who were active on the platform but never engaged with news content. We examined psychosocial outcomes across *affective*, *behavioral*, and *cognitive* dimensions.

Our findings reveal a complex pattern of effects. News exposure is associated with increases in depression, stress, and anxiety, yet simultaneously linked to decreased loneliness and increased social participation, with *Treated* users commenting and quoting at substantially higher rates than *Control* users. These results highlight a fundamental trade-off: *everyday news engagement can enable social connection, but also strain emotional wellbeing.* Importantly, not all engagement forms carry equal risks: the quantity of bookmarking, has 1900% greater psychosocial effects than other forms of engagement (e.g., commenting, quoting).

Together, these findings underscore the need to move beyond binary assessments of whether news exposure is "good" or "bad" for mental health. Our study highlights the importance of engagement forms as a key design and behavioral lever, and bears implications for platform design and user behavior. For instance, platform design that shifts users from accumulated feed consumption toward more social engagement may enable users to draw benefits of news consumption, while limiting the psychosocial costs.

2 Related Work

Effects of Media and News Consumption. News consumption has long been a central focus of media-effects research. Early research established that news coverage shapes what issues people percieve as important (McCombs and Shaw, 1972) and that repeated exposure to television news—particularly violent crime coverage—can lead to heightened perceptions of risk and fear even when objective threat remains low (Gerbner et al., 2006). Today, over half of U.S. adults report getting news from social media platforms, such as Facebook and YouTube (Shearer and Mitchell, 2021), where algorithmic curation, personalized feeds, and real-time interaction, allowing users to encounter, share, and respond to news continuously as part of everyday online activity — blurring the boundary between intentional news seeking and incidental exposure.

Studies have linked heightened online news exposure— particularly in crisis contexts such as COVID-19, mass shootings, and terrorist attacks—to increased stress, anxiety, and emotional exhaustion (Silver et al., 2013; Strasser et al., 2022; Thompson et al., 2019)— with these experiences often described as "headline stress" (Huff, 2022). Everyday news exposure can similarly contribute to emotional exhaustion and news avoidance (McLaughlin et al., 2023; Soroka and McAdams, 2015; Boukes and Vliegenthart, 2017); however, the psychosocial effects of repeated and

incidental news exposure on social media remain understudied, which our work aims to address.

Social Media and Mental Health. Prior work has shown how social media platforms enable candid self-disclosure of mental health experiences (De Choudhury and De, 2014), and how behavioral and linguistic signals on social media can provide meaningful insights into mental health states, including predicting clinical depression (De Choudhury et al., 2016, 2013; Coppersmith et al., 2014). Accordingly, social media data has enabled large-scale study of psychological responses to crises (Palen, 2008; Saha and De Choudhury, 2017), wars and conflicts (Mark et al., 2012; De Choudhury et al., 2014), natural disasters (Starbird et al., 2010), crime and violence (Valdes et al., 2015), and terrorist attacks (Cohn et al., 2004; Lin and Margolin, 2014), with platforms functioning as spaces for self-expression and solidarity (Mark et al., 2012; Starbird et al., 2010).

Computationally, the LIWC lexicon (Tausczik and Pennebaker, 2010) and classifiers for stress, anxiety, and suicidal ideation (Saha et al., 2019b; Lin et al., 2014; Burnap et al., 2015) have demonstrated construct validity for operationalizing psychological constructs from social media language (Saha et al., 2022; Ernala et al., 2019). Building on these advances, we leverage established linguistic and behavioral measures to examine the psychosocial effects of everyday news consumption.

Causal Inference Approaches on Social Media and Observational Data. Ideally, causal inference is established through randomized trials; however, ethical considerations and practical constraints often limit their feasibility in observational data settings such as longitudinal social media data. Recent research has therefore drawn on the potential outcomes framework (Rubin, 2005) and matching-based quasi-experimental approaches to construct comparable treatment and control groups, examining outcomes including college experiences (Kıcıman et al., 2018), psychiatric medication effects (Saha et al., 2019b), crisis events (Saha and De Choudhury, 2017), counseling interventions (Saha et al., 2018), online support (De Choudhury and Kıcıman, 2017; Saha and Sharma, 2020), suicidal ideation (De Choudhury et al., 2016), mental health coping (Yuan et al., 2023), county-level health estimates (Culotta, 2014), and COVID-19 impacts (Saha et al., 2020; Jha et al., 2021; Vowels et al., 2023). Methodological advances have further enabled heterogeneous treatment effect estimation (Xie et al., 2012; Olteanu et al., 2019), with prior work examining ideologically diverse news exposure (Bakshy et al., 2015), misinformation (Guess et al., 2020), social endorsements (Messing and Westwood, 2014), and hateful speech exposure (Saha et al., 2019a). Building on this literature, our work adopts a quasi-experimental approach to estimate the psychosocial effects of news engagement on social media.

3 Data

We study the psychosocial effects of news engagement on social media, particularly on *Bluesky*.

Bluesky is a decentralized, Twitter-like social media platform launched publicly in February 2024 (Failla and Rossetti, 2024). Prior work characterizes *Bluesky* as exhibiting an activity distribution comparable to other social media platforms, with a higher proportion of original content and comparatively lower toxicity (Nogara et al., 2025).

We obtained our dataset from the *Bluesky* data collated by prior work (Failla and Rossetti, 2024), which contains data from over 80% of all registered accounts at the time of collection.

The *Bluesky* platform consists of feed generators — content recommendation algorithms — which users can subscribe to based on their preferences and interests. This dataset had data of 11 such thematic feeds available on the platform.

We adopted a quasi-experimental study design based on the potential outcomes framework (Rubin, 2005), where the *treatment* is operationalized as exposure to the `News` feed. The `News` feed is a verified feed generator on *Bluesky* that aggregates headlines from verified news organizations, with the most recent content appearing first (Nogara et al., 2025), providing users with a centralized stream of news content covering current events, politics, and public affairs from established journalistic sources.

We compiled two groups: a *Treated* group of users who engaged with the `News` feed, and a *Control* group of users who never engaged with it. For each user, we obtained the entire longitudinal data of all interactions, including bookmarking, posting, liking, commenting, reposting, and quoting posts. We describe our approach below:

Compiling Treated Dataset. For our study, we define treatment as the exposure to the `News` feed. We define *Treated* users to be a group of users who were exposed to the `News` feed, where exposure could be in the form of *bookmarking*, *liking*, *commenting*, *reposting*, *quoting*, or *posting* on this feed. We gathered the longitudinal timeline of all the users who integrated feed-specific post listings, feed bookmarks, post-level likes, and users' posting histories as well as users who indirectly interacted with the `News` feed through commenting, reposting, and quoting others' posts that appeared on the `News` feed. Finally, we obtained a total of 148,680 *Treated* users, and recorded the date of their first exposure to the `News` feed as the treatment date.

Compiling Control Dataset. We built a *Control* dataset consisting of *Bluesky* users who were exposed to algorithmic feeds, but never engaged with the News feed. Similar to the *Treated* users, we compiled the entire longitudinal timeline of these users. Given that the *Control* users never engaged with the News feed, their timelines did not have an actual treatment date. Therefore, for comparability with the *Treated* users, we assigned placebo dates. We assigned each *Control* user a placebo date, matching the non-parametric distribution of treatment dates of the *Treated* dataset, to mitigate the effects of any temporal confounds. For this, we ensured that the treatment and placebo dates follow a similar distribution by non-parametrically simulating placebo dates from the pool of treatment dates. A Kolmogorov–Smirnov test between treatment and placebo dates revealed an extremely low statistic of 0.06, indicating similarity in the probability distribution of dates (Figure A1a). Table A1 shows the descriptive statistics of our entire dataset including 2.37M users, 26M posts, and 45M comments.

4 Methods

To study the psychosocial effects of News exposure on *Bluesky*, we adopted a quasi-experimental design based on the potential outcomes framework (Rubin, 2005). Ideally, causal inference is established through randomized controlled trials (RCTs); however, RCTs are often impractical and raise ethical concerns when investigating naturally occurring social media exposures (Small, 2024; Moreno et al., 2013; Grimmelmann, 2015). This framework examines whether an outcome is caused by a treatment T by comparing two potential outcomes: 1) $Y_i(T = 1)$ when exposed to T, and 2) $Y_i(T = 0)$ if there was no T. Since it is impossible to obtain both outcomes for the same individual, the framework estimates the missing counterfactual based on outcomes of similar matched individuals. In particular, we employed stratified propensity score analysis to match and examine psychosocial outcomes in *Treated* and *Control* individuals by measuring the relative treatment effect (RTE) of news exposure (Saha et al., 2019b; Kıcıman et al., 2018).

For this purpose, we estimated each user's likelihood, i.e., propensity score of encountering news content, using an AdaBoost classifier trained on baseline covariates. Users are then stratified into propensity score bins, creating matched groups where *Treated* users (those exposed to news) and *Control* users (those not exposed). Within each matched stratum of the *Treated* and *Control* groups, we compared post-treatment psychosocial outcomes. This section elaborates our approach (see Figure 1 for a schematic overview).

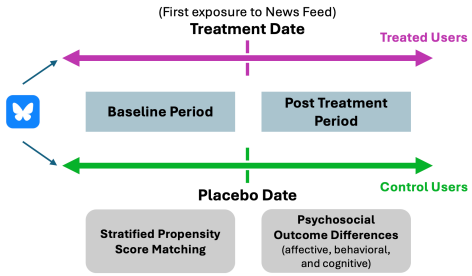


Figure 1: A schematic figure showing our causal-inference approach to analyze users' *Bluesky* timeline.

4.1 Operationalizing Psychosocial Outcomes

To measure the psychosocial effects of being exposed to the News feed on *Bluesky*, we draw from the tripartite model of psychological wellbeing (Breckler, 1984), which states that psychological states comprise three distinct components: *affective* (emotional experiences), *behavioral* (overt actions and social engagement), and *cognitive* (thoughts and mental processes). Inspired by this framework and prior research at the intersection of social media and mental health (Saha et al., 2018; Saha and Sharma, 2020; Yuan et al., 2023), we operationalized these categories as psychosocial outcomes, explained below:

Affective Outcomes. Affective outcomes capture individuals' emotional experiences and psychological states. We measured two types of affective outcomes: 1) **Symptomatic Mental Health Expressions.** We obtained the symptomatic mental health expressions of *depression*, *anxiety*, *stress*, *suicidal ideation*, and loneliness in users' language using SVM classifiers from Saha et al. (2019b). Each classifier uses n-gram (n=1,2,3) features and was trained on posts from specific mental health communities on Reddit. The training data includes posts from r/depression, r/anxiety, r/stress, r/SuicideWatch, and r/lonely, with non-mental-health posts serving as negative examples. These models achieve approximately 90% accuracy and have proved reliability across multiple social media platforms (Saha et al., 2019b; Saha and Sharma, 2020; Yuan et al., 2025). To assess cross-platform validity, the authors manually annotated 150 randomly sampled posts, finding overall classifier accuracy of 91–98% across all five conditions, consistent with prior cross-platform applications (Saha et al., 2022; Ernala et al., 2019). For each user and time period, we concatenated all their posts and applied the classifiers to obtain a normalized distribution of these symptomatic expressions. 2) **Psycholinguistic affect.** We extracted affect word frequencies using LIWC (Pennebaker et al., 2015). We counted words in the affect categories (positive emotion, negative emotion, anxiety, anger, and sad-

ness), normalizing by post length (per 100 words).

Behavioral Outcomes. Behavioral outcomes include an individual's overt actions, behavioral intentions, and verbal statements regarding behavior (Breckler, 1984). Prior studies quantified behavioral psychological wellbeing by measuring the shifts in social functioning and interests (Guntuku et al., 2019; Saha et al., 2018), and we obtained 1) **Activity.** To investigate whether news exposure shifts an individual's *Bluesky* posting activity, we calculated the average number of posts per day, quotes per day, and comments per day for every individual. We also calculated authored rows per day, which captures the total self-authored content each user has per day; 2) **Interactivity.** Interactivity is another indicator of an individual showing psychosocial changes (Saha and Sharma, 2020). We measured interactivity based on a user's *Bluesky* participation as a ratio of interactive posts (replies and quotes) to self-authored posts.

Cognitive Outcomes. Cognitive outcomes reflect an individual's thought patterns and information processing (Breckler, 1984). Drawing on psycholinguistic research showing that language reveals cognitive functioning (Pennebaker and Stone, 2003; Tausczik and Pennebaker, 2010), we operationalize cognitive outcomes as: 1) **Readability.** Readability reflects how easily text can be understood and is important for both expression and interpretation, measured using the Coleman-Liau Index (CLI) (Ernala et al., 2017; Park et al., 2018); 2) **Verbosity.** Verbosity captures elaboration in communication, measured as words per sentence (Saha and Sharma, 2020); 3) **Complexity and Repeatability.** Higher complexity and lower repetition are associated with better mental health (Ernala et al., 2017). We measure *repeatability* as word reuse frequency, and *complexity* as average word length, reflecting more precise and nuanced expression (Kolden et al., 2011); and 4) **LIWC Cognition and Social.** We extract *cognitive processes* (e.g., causation, certainty), *social focus* (e.g., pronouns), and *function words* (e.g., articles, prepositions, auxiliaries).

4.2 Matching for Causal Inference

Covariates. When conditioned on high-dimensional covariate data, matching is known to mitigate bias compared to naive correlational analyses (Rubin, 2005). We draw on prior work to identify a range of covariates: 1) **activity metrics**: *posting frequency* (number of posts per day) and *posting tenure* (number of days on *Bluesky*), 2) **symptomatic mental health expressions:** classifier-based measures of *depression, anxiety, stress, suicidal ideation*, and *loneliness* (Saha et al., 2019b; Yuan et al., 2025), 3) **content**: top 500 n-grams (n=2,3), and 4) **psycholinguistics**: all LIWC-22 attributes (Pennebaker et al., 2001). In total, we obtained 520 covariates and normalized the measures per user, based on each user's baseline period of a minimum of 30-days before the treatment date.

Stratified Propensity Score Matching. We used matching to find pairs (generalizable to groups) of *Treated* and *Control* users whose covariates are statistically similar in the baseline period. We employed the propensity score matching (PSM), which matches users based on the likelihood of news exposure given baseline covariates (Saha et al., 2019b). For this purpose, we built an Adaboost classifier using the SAMME algorithm with decision trees (max depth=2) as base learners (rate=0.05, estimators=200) ((Zhu et al., 2009)). This model predicted the propensity scores ranging between 0 and 1. Figure A1b shows a distribution of propensity scores.

Next, we adopted a stratification approach—stratified matching enables us to handle the bias-variance tradeoff by striking a balance between too biased (one-to-one matching) and too variant (unmatched) data comparisons (Kıcıman et al., 2018). In particular, we stratified all the users into 10 equal-width strata based on their propensity scores, where each stratum consisted of users with similar baseline covariates, i.e., matched *Treated* and *Control* users (Saha et al., 2019b).

We filtered out users with no posting activity, users with baseline periods shorter than 30 days, and users with no post-treatment activity. Additionally, we used a minimum threshold of at least 10 users per group and dropped any strata with fewer users per group. Accordingly, we dropped three strata: the remaining seven strata, consisting of 83,711 *Treated* users and 81,345 *Control* users, were used for our ensuing analysis.

Quality of Matching. To ensure that our matching obtained comparable *Treated* and *Control* users, we evaluated the balance of covariates. We obtained the standardized mean differences (SMD) for all the covariates across the *Treated* and *Control* users. For each covariate, SMD is the difference in the mean values between the two groups as a fraction of the pooled standard deviation between the two groups. Two groups can be considered to be balanced if the SMD is lower than 0.25 (Kıcıman et al., 2018). We find a significant reduction in the SMD from 0.042 (max=1.34) in the unmatched datasets to 0.037 (max=0.21) in the matched datasets (Figure A1c). Using a conservative threshold of 0.15 for adequate covariate balance, the number of imbalanced covariates decreased from 16 (3.1%) before matching to 4 (0.8%) after matching, reinforcing that the stratification successfully created comparable treatment and control groups.

Relative Treatment Effect. To measure the psychosocial effects of exposure to News, we computed the Relative Treatment Effect (RTE) for each outcome (Kıcıman et al., 2018). We first calculated the RTE per outcome measure in every stratum, as a ratio of the mean outcome value in the *Treated* group to that in the *Control* group. Next, using a weighted average across the strata, we obtained RTE per outcome measure. An outcome RTE greater than 1 would indicate that the outcome increased in *Treated* users, whereas an RTE lower than 1 would indicate that it decreased in *Treated* users, following the treatment (news exposure). We obtained effect size (Cohen's d) and conducted t-tests to assess statistical significance in differences. Further, in order to examine the duration of these effects, we also plotted a temporal representation of the RTE over a two-month period following treatment.

Individual Treatment Effect. To quantify differences in psychosocial effects across forms of news engagement, within each stratum, we computed the Individual Treatment Effect (ITE) for each user as the difference between their post-treatment outcome and the mean outcome of matched controls in the same stratum. We then modeled these ITEs using linear regressions. Our engagement types are: 1) *Posting*, 2) *Commenting*, 3) *Quoting*, 4) *Reposting*, 5) *Bookmarking*, and 6) *Liking*. For each outcome, we built a separate model with ITE as the dependent variable, and engagement types, baseline mental health (depression, anxiety, stress, suicidal ideation, loneliness), linguistic features (readability, verbosity, repeatability), and interactivity as independent variables. To address multicollinearity, we excluded covariates with VIF > 20 (Ó́brien, 2007).

5 Results

5.1 RQ1: Psychosocial Effects

Table 1 summarizes the psychosocial changes observed between matched *Treated* and *Control* users following exposure to News feed, and Figure A2 shows how these effects evolve over the next two months following treatment. In summary, we note significant changes across several measures in both the directions of increase (RTE>1) and decrease (RTE<1) following News feed exposure. We elaborate on these observations below:

Affective Outcomes. The *Treated* group showed significantly higher symptomatic mental health expressions, most notably 51% higher depression (RTE=1.51), followed by anxiety (RTE=1.36) and stress (RTE=1.28). Additionally, the *Treated* group showed 56% higher negative affect (RTE=1.56) and 63% higher anger (RTE=1.63). Further, Figure A2a reveals that depression, anxiety, stress, and suicidal ideation expressions increase sharply among *Treated* users immediately after News feed exposure and then gradually attenuate before stabilizing. The sharp post-exposure increases followed by gradual attenuation align with prior research on the models of stress and habituation, in which initial emotional arousal partially subsides while maintaining an elevated baseline (Garfin et al., 2015).

Metric	Tr.	Ct.	RTE	d	t-test
Affective Outcomes					
Depression	0.74	0.49	1.51	0.86	174.03***
Anxiety	0.55	0.41	1.36	0.47	95.90***
Stress	0.66	0.51	1.28	0.46	93.27***
Suicidal Ideation	0.47	0.40	1.16	0.22	45.01***
Loneliness	0.20	0.34	0.59	-0.70	-141.90***
LIWC: Anger	0.77	0.47	1.63	0.21	38.61***
LIWC: Neg. Affect	1.86	1.19	1.56	0.31	56.45***
LIWC: Pos. Affect	4.97	4.71	1.05	0.05	9.45***
LIWC: Sadness	0.30	0.22	1.37	0.09	17.62***
Behavioral Outcomes					
Posting Frequency	1.15	0.38	3.01	0.12	23.15***
Quotes Frequency	0.78	0.02	39.56	0.07	14.44***
Comments Frequency	2.04	0.28	7.20	0.17	34.77***
Interactivity	2.73	0.85	3.20	0.04	7.06***
Cognitive Outcomes					
Verbosity	18.44	11.57	1.59	0.65	132.52***
Readability	8.83	14.86	0.59	-0.36	-64.32***
Repeatability	0.43	0.19	2.23	0.99	201.51***
Complexity	4.23	4.45	0.95	-0.07	-13.94***
LIWC: Cog. Proc.	8.08	5.91	1.37	0.39	71.66***
LIWC: Social Context	6.47	5.23	1.24	0.24	42.88***
LIWC: Function Words	36.26	28.37	1.28	0.39	72.92***
LIWC: Personal Pronouns	5.88	5.42	1.08	0.10	17.54***

Table 1: Summary of comparisons of *Treated* and *Control* datasets in terms of Relative Treatment Effect (RTE), effect size (Cohen's d), and paired t-test (*p<0.05, ** p<0.01, ***p<0.001).

In contrast, *Treated* users' loneliness expressions were 41% lower (RTE=0.59), and loneliness expressions sharply dropped after treatment, and continued to decrease over the next two months (Figure A2a). The sustained decrease in loneliness suggests that news engagement may facilitate shared attention and social connection, echoing prior findings that participation in collective online discourse can reduce perceived isolation (Boczkowski et al., 2018; Saha and De Choudhury, 2017).

Behavioral Outcomes. *Treated* users show significantly higher behavioral outcomes than *Control* users. *Treated* users show a significantly higher quantity of quoting behavior (0.78 vs. 0.02 quotes per day; RTE=39.56) and commenting behavior (2.04 vs. 0.28 comments per day; RTE=7.20). Further, *Treated* users show 220% higher overall interactivity (RTE=3.20) than *Control* users.

Following first exposure to the News feed, posting frequency, commenting, quoting, and overall interactivity rise sharply in the immediate post-exposure period and then plateau, indicating sustained but stable engagement levels over the subsequent 60 days (Figure A2b). These patterns are consistent with an initial surge in participation that stabilizes over time, but RTE continues to remain above 1, indicating shared attention and collective discourse. The sharp post-exposure increase followed by a plateau is consistent with theories of attention dynamics, in which salient informational

Outcome	Bookmark	Comment	Like	Post	Quote	Repost	R^2
Affective Outcomes							
Depression	4.60	0.30	0.03	-3E-3	0.17	0.07	0.250
Anxiety	5.60	0.02	8.7E-3	-4.6E-3	0.36	-0.10	0.416
Stress	4.50	0.19	0.01	-4.9E-3	0.20	-0.09	0.439
Suicidal Ideation	-2.80	-0.18	1.6E-3	-2.1e-05	-0.37	0.03	0.283
Loneliness	-4.40	-0.15	-0.02	3.8E-3	-0.03	-0.01	0.412
LIWC: Anger	-11.80	0.88	3.9E-3	1.9e-05	0.58	-0.26	0.044
LIWC: Neg. Affect	-11.60	1.60	1.5E-3	8.2e-05	0.95	-0.53	0.109
LIWC: Pos. Affect	32.00	0.27	0.05	0.0002	-1.40	-0.90	0.229
LIWC: Sadness	0.0093	0.14	-2.1E-4	2.5E-3	-6E-3	-0.07	0.016
Behavioral Outcomes							
Posting Frequency	-5.90	4.60	0.04	0.85	5.30	0.86	0.011
Quotes Frequency	-0.290	-4.3E-3	-0.14	-0.08	18.00	0.06	0.007
Comments Frequency	33.70	17.20	0.49	-0.33	18.80	-0.65	0.028
Interactivity	-573.50	-24.30	-0.59	-0.0022	167.40	37.90	0.022
Cognitive Outcomes							
Verbosity	-1.053	-2.80	-0.33	0.16	6.50	-1.50	0.422
Readability	-21.90	-1.70	-0.06	0.0010	6.70	0.73	0.520
Repeatability	4.70	0.41	0.04	0.02	0.58	0.05	0.325
Complexity	-0.028	0.04	0.03	0.01	0.86	0.22	0.273
LIWC: Cog. Proc.	53.40	3.40	0.02	-0.0014	1.30	-2.20	0.417
LIWC: Social Context	37.90	3.60	0.01	-0.0004	1.50	-1.40	0.280
LIWC: Function Words	182.60	11.00	0.14	-0.33	-0.16	-8.50	0.535
LIWC: Personal Pronouns	0.262	0.019	0.0005	-0.0014	-0.024	-0.016	0.356

Table 2: Linear regression estimates ($\beta \times 10^{-2}$) showing associations between engagement frequency (columns) and psychosocial outcomes (rows). Shading indicates statistical significance: $p<0.05$, $p<0.01$, $p<0.001$.

stimuli trigger an initial surge in participation that stabilizes as norms and routines of engagement emerge (Lehmann et al., 2012). Again, the sustained elevation in interactivity suggests that news feeds may serve as focal points for ongoing social coordination and shared interpretation, rather than producing only short-lived bursts of activity (Shamma et al., 2009).

Cognitive Outcomes. The *Treated* users' posted 59% more verbose content (18.44 vs. 11.57 words per sentence, RTE=1.59), but 123% more repeatable content (0.43 vs. 0.19, RTE=2.23) than *Control* users This indicates that *Treated* wrote longer but more reused content. Figure A2c further reveals how verbosity, repeatability, and linguistic complexity increased immediately after first exposure to the News feed and then gradually attenuated before stabilizing, while readability remained consistently lower over time.

Treated users also showed higher use of cognitive (RTE=1.37) and social (RTE=1.24) language, along with greater function word and pronoun usage (RTE=1.28, 1.08). They used fewer first-person singular pronouns (RTE=0.94) but more third-person singular (RTE=1.34) and plural pronouns (RTE=1.91), indicating reduced self-focus and greater emphasis on others, consistent with prior work (Pennebaker et al., 2003).

5.2 RQ2: Degree of News Engagement and Psychosocial Outcomes

For RQ2, we examine how the degree of news engagement is associated with psychosocial outcomes. For each outcome, we fit a regression model in which the dependent variable was the outcome's individual treatment effect (ITE), and the independent variables were the frequencies of various engagement types: bookmarking, commenting, liking, posting, quoting, and reposting. Table 2 summarizes the β coefficients of the independent variables (columns) for each psychosocial outcome (rows). Across outcomes, several engagement measures show significant associations, suggesting that the degree of engagement meaningfully relates to psychosocial responses. This pattern is consistent with a "dosage" interpretation, in which greater levels of engagement correspond to stronger psychosocial effects. Notably, the frequency of bookmarking showed the strongest association, especially with affective outcomes, and quoting showed the strongest association with behavioral outcomes.

Affective Outcomes. For *depression*, bookmarking showed the largest positive association (β=4.60), far exceeding commenting (β=0.30), quoting (β=0.17), reposting (β=0.07), and liking (β=0.03). Similar patterns appear for *anxiety* (β=5.60) and *stress* (β=4.50), with bookmarking effects substantially larger than other behaviors. Reposting showed small negative associations with stress (β=-0.09) and anxiety (β=-0.10). In contrast, for *suicidal ideation* and *loneliness*, quoting showed the strongest negative association (β=-0.37), followed by commenting (β=-0.18) and bookmarking (β=-2.80), while reposting showed a small positive association (β=0.03). For *loneliness*, bookmarking showed the largest negative association (β=-4.40), followed by commenting (β=-0.15), quoting (β=-0.03), and liking (β=-0.02), indicating that greater news engagement is associated with reduced loneliness. Similar patterns emerge for LIWC-based affect. Bookmarking showed strong negative associations with *anger* (β=-11.80) and *negative affect* (β=-11.60), while commenting and quoting were positively associated with both. *Positive affect* showed the opposite pattern, with bookmarking strongly positive (β=32.00) and quoting and reposting negative (β=-1.40, β=-0.90).

Behavioral Outcomes. Quoting showed the strongest association with *posting activity* (β=5.30), followed by commenting (β=4.60) and reposting (β=0.86). Overall, engaging with news through quoting or commenting was associated with higher posting frequency. For *interactivity* (comments-to-post ratio), quoting showed a strong positive association (β=167.40), followed by reposting (β=37.90). In contrast, bookmarking (β=-573.50) and commenting (β=-24.30) were negatively associated, indicating lower interactivity. The large negative effect for bookmarking suggests that users who primarily save content engage less, while those who quote or share news exhibit higher interactivity. Similar patterns emerge for *activity frequency*: quoting frequency shows the strongest positive associations with quotes/day (β=18.00) and comments/day (β=18.80), while commenting frequency is also strongly linked to comments/day

(β=17.20). Bookmarking shows negligible or negative associations.

Cognitive Outcomes. For *readability*, quoting News content showed the strongest positive association (β=6.70), followed by reposting (β=0.73), while commenting was negatively associated (β=-1.70). For *complexity*, quoting and reposting were both positive, suggesting that these behaviors are linked to more complex and readable language, possibly reflecting uptake of news-related discourse. Additionally, *verbosity* varied by engagement type: posting (β=0.16) and quoting (β=6.50) were associated with longer content, while reposting was negatively associated (β=-1.50), suggesting that content creation encourages more elaborate expression than simple redistribution. For *complexity*, quoting showed the strongest positive association (β=0.86), followed by reposting (β=0.22), indicating that these behaviors are linked to more complex language. Again, for *repeatability*, all engagement types showed positive associations, with bookmarking having the largest effect (β=4.70), followed by quoting (β=0.58), commenting (β=0.41), reposting (β=0.05), liking (β=0.04), and posting (β=0.02). Overall, greater news engagement is associated with more repetitive language.

6 Discussion and Conclusion

This work examines the psychosocial effects of social media news exposure, revealing nuanced patterns in affective, behavioral, and cognitive well-being. Rather than uniform outcomes, news engagement reflects trade-offs that depend on how users engage. We discuss the implications of our work.

Theoretical Implications. Our findings extend existing theories of media and news effects in three key ways. *First*, we extend news effects research beyond crisis-centric frameworks by demonstrating psychosocial consequences of routine engagement. Prior work has focused on acute events such as pandemics or natural disasters, where distress follows temporally bounded exposure (Thompson et al., 2019). Our results show that everyday news engagement embedded in routine social media use can alter psychosocial well-being, suggesting that theories must account not only for episodic shocks but also for the cumulative burden of persistent informational environments. *Second*, our findings highlight engagement heterogeneity in affecting wellbeing. Research often conceptualizes exposure in terms of volume (Garfin et al., 2020; Thompson et al., 2019), assuming greater engagement leads to stronger effects. Our findings challenge this by showing that different engagement forms are associated with varying outcomes—bookmarking was associated with greater increases in depression, anxiety, and stress than active forms such as commenting or quoting. This calls for models that distinguish not only how much users consume news, but how they participate in the news ecosystem. *Third*, our findings reveal coexisting affective costs and social benefits of news engagement, complicating narratives that frame news consumption as uniformly harmful or beneficial. This suggests that news engagement facilitates social connection and collective sense-making even as it imposes emotional strain (De Choudhury and De, 2014; Chancellor et al., 2016).

Methodological, Practical, and Design Implications. Our findings highlight a fundamental tension between user wellbeing and platform incentives. Patterns associated with adverse psychosocial outcomes—such as accumulating multiple news feeds and sustained passive consumption—are those that maximize time-on-platform. Interventions that reduce these patterns would counter platforms' core economic objectives, raising questions about the feasibility of platform-led solutions and suggesting that the burden of mitigation may fall on users, researchers, and civic technologists outside platform governance.

In this context, user-facing and post-hoc interventions become salient. Tools independent of platform ranking algorithms—such as browser extensions or third-party overlays—could enable users to monitor cumulative news exposure, and reorganize or filter feeds. Recent work on teachable feeds (Choi and Chandrasekharan, 2025) and user-driven curation (Malki et al., 2025) on *Bluesky* presents promising directions. Beyond reducing exposure, interventions that encourage commenting, discussion, or contextualization may preserve social benefits while mitigating emotional costs.

Our findings also highlight the role of community-driven platforms, where governance and incentive structures differ from dominant commercial platforms. These platforms may be better positioned to experiment with designs that limit harmful exposure while promoting contextualized engagement, serving as testbeds for translating insights into design practices that are difficult to implement within profit-driven ecosystems.

Finally, our work underscores the value of quasi-experimental methods for measuring psychosocial outcomes in new platforms. By observing early engagement trajectories and distinguishing between participation types, researchers can better isolate the effects of specific design features and behaviors, complementing survey and laboratory studies by capturing longitudinal, large-scale dynamics in naturalistic settings.

7 Ethical Considerations

This research employs computational and data-driven approaches to study the psychosocial effects of routine news exposure on social media users. Given the sensitivity of mental health–related outcomes and the potential vulnerability of affected populations, the ethical implications of this work are multifaceted and warrant careful consideration. In particular, computational models used to infer symptomatic mental health expressions or affective states from social media language can be misused or misappropriated in ways that may cause harm. For example, such models could be leveraged for targeted advertising, commercial profiling, surveillance, or insurance-related decision-making, rather than for research or wellbeing-oriented purposes. We caution against interpreting these models as diagnostic tools or as providing clinical assessments of mental health conditions. Our findings reflect population-level patterns inferred from behavioral and linguistic expressions and should not be used to draw conclusions about individual users' mental health status. Any deployment of tools inspired by this work would require substantial additional safeguards, including clear limitations on use, transparency, and oversight by domain experts. Further, social media expressions are shaped by cultural, social, and contextual factors, and mental health experiences and coping mechanisms vary widely across individuals and communities. Applying these methods without sensitivity to such differences risks reinforcing stereotypes or biases about particular groups or forms of expression. Researchers and practitioners should therefore take care to avoid overgeneralization and to consider how differences in language use, cultural norms, and platform practices may influence observed patterns. Overall, this work is intended to inform empirical understanding of psychosocial dynamics in online news environments, not to enable monitoring, profiling, or intervention at the individual level. We emphasize the importance of ethical reflection, responsible use, and contextualized interpretation when studying mental health–related phenomena using computational approaches.

8 Limitations and Future Directions

Our study has limitations, which also suggest interesting future directions. We focus on how users engage with news rather than on the substantive topics or narrative frames of the news itself. While our results show that engagement modality and intensity shape psychosocial outcomes, different types of news—such as political conflict, public health crises, or local community events—may have distinct psychological consequences even under similar engagement patterns. Future work should examine how topic-specific news exposure interacts with forms of engagement to influence mental health.

Although our quasi-experimental design strengthens causal inference relative to purely observational approaches, it does not establish *true causality*. Moreover, our findings are not clinical in nature and are inferred from users' behavioral and linguistic expressions on *Bluesky*, rather than from validated mental health assessments. This inspires future work with natural experiments and mixed-methods designs, and validated mental health surveys to further strengthen both causal inference and clinical interpretability.

Finally, our analysis is limited to a single platform, *Bluesky*, raising questions about generalizability. Studying news (and other online) engagement across platforms is increasingly difficult due to restricted data access, which often limits external researchers to a narrow set of observable interactions, such as commenting, while obscuring passive consumption behaviors that may be most consequential for mental health. As a result, some forms of research can realistically be conducted only within platforms. Addressing these constraints may therefore require new research models, including data donation initiatives, privacy-preserving data sharing, and stronger industry–academia partnerships, to enable broader and more representative analyses while maintaining user privacy.

9 AI Involvement Disclosure

AI-assisted language editing (e.g., Grammarly, ChatGPT, Claude) was used exclusively to improve grammar and readability. The study design, analyses, interpretations, and experiments were conducted fully by the authors.

References

Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on facebook. *Science*.

Pablo J Boczkowski, Eugenia Mitchelstein, and Mora Matassi. 2018. "news comes across when i'm in a moment of leisure": Understanding the practices of incidental news consumption on social media. *New media & society*, 20(10):3523–3539.

Mark Boukes and Rens Vliegenthart. 2017. News consumption and its unpleasant side effect. *Journal of Media Psychology*.

Steven J Breckler. 1984. Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of personality and social psychology*, 47(6):1191.

Pete Burnap, Walter Colombo, and Jonathan Scourfield. 2015. Machine classification and analysis of suicide-related communication on twitter. In *Proc. ACM conference on hypertext & social media*.

Stevie Chancellor, Zhiyuan Lin, Erica L Goodman, Stephanie Zerwas, and Munmun De Choudhury. 2016. Quantifying and predicting mental illness severity in online pro-eating disorder communities. In *CSCW*.

Frederick Choi and Eshwar Chandrasekharan. 2025. Designing usable controls for customizable social media feeds. *arXiv preprint arXiv:2509.19615*.

Michael A Cohn, Matthias R Mehl, and James W Pennebaker. 2004. Linguistic markers of psychological change surrounding september 11, 2001. *Psychol. Sci.*

Glen Coppersmith, Mark Dredze, and Craig Harman. 2014. Quantifying mental health signals in twitter. In *Proc. ACL CLCP Workshop*.

Aron Culotta. 2014. Estimating county health statistics with twitter.

Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *ICWSM*.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting depression via social media. In *ICWSM*.

Munmun De Choudhury and Emre Kıcıman. 2017. The language of social support in social media and its effect on suicidal ideation risk. In *ICWSM*.

Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *CHI*.

Munmun De Choudhury, Andres Monroy-Hernandez, and Gloria Mark. 2014. Narco emotions: affect and desensitization in social media during the mexican drug war. In *CHI*, pages 3563–3572. ACM.

Sindhu Kiranmai Ernala, Michael L Birnbaum, Kristin A Candan, Asra F Rizvi, William A Sterling, John M Kane, and Munmun De Choudhury. 2019. Methodological gaps in predicting mental health states from social media: triangulating diagnostic signals. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–16.

Sindhu Kiranmai Ernala, Asra F Rizvi, Michael L Birnbaum, John M Kane, and Munmun De Choudhury. 2017. Linguistic markers indicating therapeutic outcomes of social media disclosures of schizophrenia. *PACM HCI*, (CSCW).

Andrea Failla and Giulio Rossetti. 2024. "i'm in the bluesky tonight": Insights from a year worth of social data. *PloS one*.

Dana Rose Garfin, E Alison Holman, and Roxane Cohen Silver. 2015. Cumulative exposure to prior collective trauma and acute stress responses to the boston marathon bombings. *Psychological science*.

Dana Rose Garfin, Roxane Cohen Silver, and E Alison Holman. 2020. The novel coronavirus (covid-2019) outbreak: Amplification of public health consequences by media exposure. *Health psychology*.

George Gerbner, Larry Gross, Michael Morgan, and Nancy Signorielli. 2006. The "mainstreaming" of america: Violence profile no. 11. *Journal of Communication*, 30(3):10–29.

James Grimmelmann. 2015. The law and ethics of experiments on social media users. *Colo. Tech. LJ*, 13:219.

Andrew M Guess, Michael Lerner, Benjamin Lyons, Jacob M Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. 2020. A digital media literacy intervention increases discernment between mainstream and false news in the united states and india. *Proceedings of the National Academy of Sciences*, 117(27):15536–15545.

Sharath Chandra Guntuku, J Russell Ramsay, Raina M Merchant, and Lyle H Ungar. 2019. Language of adhd in adults on social media. *Journal of attention disorders*.

Alfred Hermida. 2010. Twittering the news: The emergence of ambient journalism. *Journalism practice*, 4(3):297–308.

Charlotte Huff. 2022. Media overload is hurting our mental health. here are ways to manage headline stress. *Monit. Psychol*, 53:20.

Juwon Hwang, Porismita Borah, Dhavan Shah, and Markus Brauer. 2021. The relationship among covid-19 information seeking, news media use, and emotional distress at the onset of the pandemic. *International Journal of Environmental Research and Public Health*, 18(24):13198.

Indra Prakash Jha, Raghav Awasthi, Ajit Kumar, Vibhor Kumar, and Tavpritesh Sethi. 2021. Learning the mental health impact of covid-19 in the united states with explainable artificial intelligence: observational study. *JMIR mental health*.

Emre Kıcıman, Scott Counts, and Melissa Gasser. 2018. Using longitudinal social media analysis to understand the effects of early college alcohol use. In *ICWSM*, pages 171–180.

Gregory G Kolden, Marjorie H Klein, Chia-Chiang Wang, and Sara B Austin. 2011. Congruence/genuineness. *Psychotherapy*, 48(1):65.

Janette Lehmann, Bruno Gonçalves, José J Ramasco, and Ciro Cattuto. 2012. Dynamical classes of collective attention in twitter. In *Proceedings of the 21st international conference on World Wide Web*, pages 251–260.

Huijie Lin, Jia Jia, Quan Guo, Yuanyuan Xue, Qi Li, Jie Huang, Lianhong Cai, and Ling Feng. 2014. User-level psychological stress detection from social media using deep neural network. In *ACM Multimedia*. ACM.

Yu-Ru Lin and Drew Margolin. 2014. The ripple of fear, sympathy and solidarity during the boston bombings. *EPJ Data Science*.

Omar El Malki, Marianne Aubin Le Quéré, Andrés Monroy-Hernández, and Manoel Horta Ribeiro. 2025. Bonsai: Intentional and personalized social media feeds. *arXiv preprint arXiv:2509.10776*.

Gloria Mark, Mossaab Bagdouri, Leysia Palen, James Martin, Ban Al-Ani, and Kenneth Anderson. 2012. Blogs as a collective war diary. In *CSCW*.

Maxwell E McCombs and Donald L Shaw. 1972. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2):176–187.

Bryan McLaughlin, Melissa R Gotlieb, and Devin J Mills. 2023. Caught in a dangerous world: Problematic news consumption and its relationship to mental and physical ill-being. *Health communication*.

Solomon Messing and Sean J Westwood. 2014. Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication research*.

Megan A Moreno, Natalie Goniu, Peter S Moreno, and Douglas Diekema. 2013. Ethics of social media research: Common concerns and practical considerations. *Cyberpsychology, behavior, and social networking*.

Nic Newman, A Ross Arguedas, Craig T Robertson, Rasmus Kleis Nielsen, and Richard Fletcher. 2025. *Digital news report 2025*. Reuters Institute for the study of Journalism.

Gianluca Nogara, Erfan Samieyan Sahneh, Matthew R. DeVerna, Nick Liu, Luca Luceri, Filippo Menczer, Francesco Pierri, and Silvia Giordano. 2025. A longitudinal analysis of misinformation, polarization and toxicity on bluesky after its public launch. *Preprint*, arXiv:2505.02317.

Robert M Ó́brien. 2007. A caution regarding rules of thumb for variance inflation factors. *Quality & quantity*, 41(5):673–690.

Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kiciman. 2019. Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data*, 2:13.

Leysia Palen. 2008. Online social media in crisis events. *Educause Quarterly*.

Albert Park, Mike Conway, and 1 others. 2018. Harnessing reddit to understand the written-communication challenges experienced by individuals with mental health disorders: analysis of texts from mental health communities. *Journal of medical Internet research*, 20(4):e8219.

James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015. Technical report, University of Texas at Austin.

James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71:2001.

James W Pennebaker, Matthias R Mehl, and Kate G Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, 54(1):547–577.

James W Pennebaker and Lori D Stone. 2003. Words of wisdom: language use over the life span. *Journal of personality and social psychology*, 85(2):291.

Pew Research Center. 2025. Social media and news fact sheet. Survey of U.S. adults conducted Aug. 18-24, 2025.

Donald B Rubin. 2005. Causal inference using potential outcomes: Design, modeling, decisions. *J. Am. Stat. Assoc.*

Koustuv Saha, Eshwar Chandrasekharan, and Munmun De Choudhury. 2019a. Prevalence and psychological effects of hateful speech in online college communities. In *ICWSM*.

Koustuv Saha and Munmun De Choudhury. 2017. Modeling stress with social media around incidents of gun violence on college campuses. *PACM HCI*, (CSCW).

Koustuv Saha and Amit Sharma. 2020. Causal factors of effective psychosocial outcomes in online mental health communities. In *ICWSM*.

Koustuv Saha, Benjamin Sugar, John Torous, Bruno Abrahao, Emre Kıcıman, and Munmun De Choudhury. 2019b. A social media study on the effects of psychiatric medication use. In *ICWSM*.

Koustuv Saha, John Torous, Eric D Caine, and Munmun De Choudhury. 2020. Psychosocial effects of the covid-19 pandemic: large-scale quasi-experimental study on social media. *JMIR*.

Koustuv Saha, Ingmar Weber, and Munmun De Choudhury. 2018. A social media based examination of the effects of counseling recommendations after student deaths on college campuses. In *ICWSM*.

Koustuv Saha, Asra Yousuf, Ryan L Boyd, James W Pennebaker, and Munmun De Choudhury. 2022. Social media discussions predict mental health consultations on college campuses. *Scientific reports*.

David A Shamma, Lyndon Kennedy, and Elizabeth F Churchill. 2009. Tweet the debates: understanding community annotation of uncollected sources. In *Proceedings of the first SIGMM workshop on Social media*, pages 3–10.

Elisa Shearer and Amy Mitchell. 2021. News use across social media platforms in 2020. Pew Research Center.

Roxane Cohen Silver, E Alison Holman, Judith Pizarro Andersen, Michael Poulin, Daniel N McIntosh, and Virginia Gil-Rivas. 2013. Mental-and physical-health effects of acute exposure to media images of the september 11, 2001, attacks and the iraq war. *Psychological science*, 24(9):1623–1634.

Dylan S Small. 2024. Protocols for observational studies: Methods and open problems. *Statistical Science*, 39(4):519–554.

Stuart Soroka and Stephen McAdams. 2015. News, politics, and negativity. *Political communication*, 32(1):1–22.

Kate Starbird, Leysia Palen, Amanda L Hughes, and Sarah Vieweg. 2010. Chatter on the red: what hazards threat reveals about the social life of microblogged information. In *CSCW*, pages 241–250. ACM.

Michelle A Strasser, Philip J Sumner, and Denny Meyer. 2022. Covid-19 news consumption and distress in young people: A systematic review. *Journal of affective disorders*, 300:481–491.

Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.

Rebecca R Thompson, Nickolas M Jones, E Alison Holman, and Roxane Cohen Silver. 2019. Media exposure to mass violence events can fuel a cycle of distress. *Science advances*, 5(4):eaav3502.

Kjerstin Thorson and Chris Wells. 2016. Curated flows: A framework for mapping media exposure in the digital age. *Communication theory*.

Roger Tourangeau, Lance J Rips, and Kenneth Rasinski. 2000. *The psychology of survey response*.

Jose Manuel Delgado Valdes, Jacob Eisenstein, and Munmun De Choudhury. 2015. Psychological effects of urban crime gleaned from social media. In *ICWSM*.

Laura M Vowels, Matthew J Vowels, Katherine B Carnelley, Abigail Millings, and Jilly Gibson-Miller. 2023. Toward a causal link between attachment styles and mental health during the covid-19 pandemic. *Br. J. Clin. Psychol.*

Yu Xie, Jennie E Brand, and Ben Jann. 2012. Estimating heterogeneous treatment effects with observational data. *Sociol. Methodol.*

Yunhao Yuan, Koustuv Saha, Barbara Keller, Erkki Tapio Isometsä, and Talayeh Aledavood. 2023. Mental health coping stories on social media: a causal-inference study of papageno effect. In *TheWebConf*.

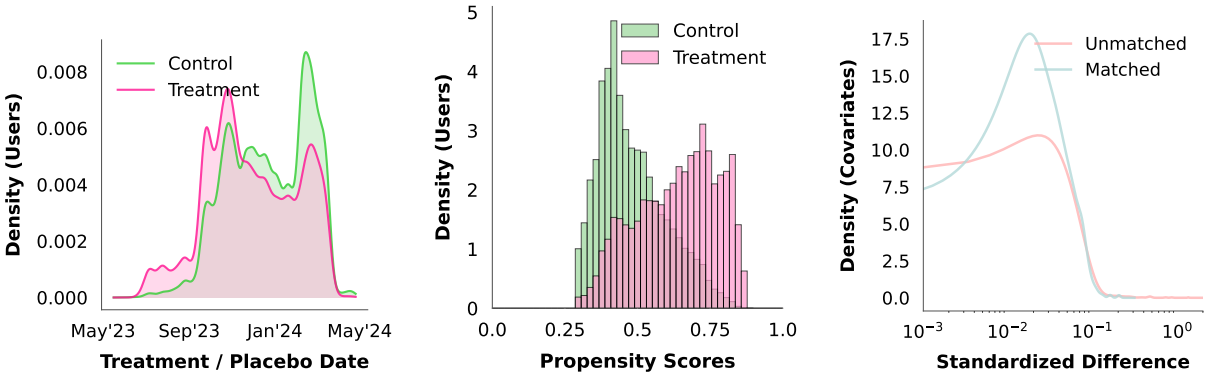
Yunhao Yuan, Jiaxun Zhang, Talayeh Aledavood, Renwen Zhang, and Koustuv Saha. 2025. Mental health impacts of ai companions: Triangulating social media quasi-experiments, user perspectives, and relational theory. *arXiv preprint arXiv:2509.22505*.

Ji Zhu, Hui Zou, Saharon Rosset, Trevor Hastie, and 1 others. 2009. Multi-class adaboost. *Statistics and its Interface*, 2(3):349–360.

A Appendix

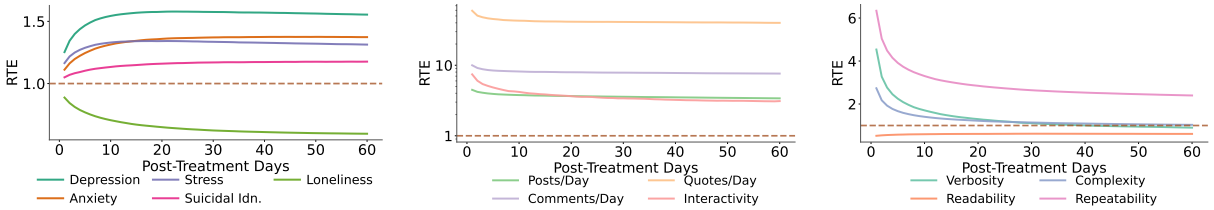
Dataset →	***Treated***		***Control***	
Metric ↓	**Count**	**Mean**	**Count**	**Mean**
Users	148,680	–	2,221,748	–
Posts	19,245,131	225.02	7,227,387	56.64
Quotes	8,102,224	94.73	540,409	4.23
Comments	39,370,440	460.33	6,303,635	49.40
Reposts	33,467,258	391.31	4,181,332	32.77
Likes	4,034,615	47.17	54,854	0.43

Table A1: Descriptive statistics of the *Bluesky* dataset.



(a) (b) (c)

Figure A1: Distribution of a) treatment/placebo dates, b) propensity scores, and b) standardized mean differences (SMD) across the *Treated* and *Control* users' datasets.



(a) Affective Outcomes (b) Behavioral Outcomes (c) Cognitive Outcomes

Figure A2: Aggregated relative treatment effects (RTE) on psychosocial outcomes measured over two months following exposure.

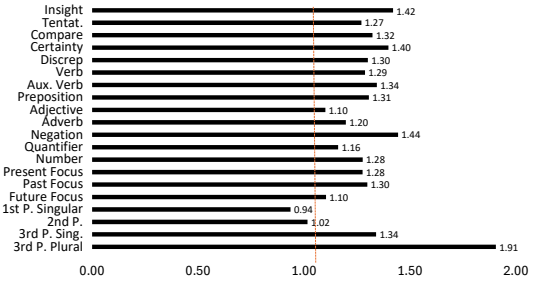


Figure A3: RTEs of LIWC cognitive and social attributes.