

# Exploring Mental Health Discourse on Nordic and Baltic Reddit: A Multilingual Computational Study

Aparup Khatua<sup>1</sup>

<sup>1</sup>Lulea University of Technology, Sweden  
aparup.khatua@ltu.se

## Abstract

Prior mental-health studies using online data primarily explored English-language dedicated support forums in single-country settings. In contrast, we investigate *high-visibility* mental-health-related discourse on Reddit spanning 7 Nordic and Baltic countries (Norway, Sweden, Denmark, Finland, Iceland, Estonia, Lithuania), 3 regional communities, and 8 languages collected over 12 months. Specifically, we study multilingual general-community subreddits and extend the analysis to community reply behavior. A two-stage approach (multilingual keyword lexicon followed by GPT-4o-mini annotation) yields 820 high-confidence disclosures. Counter-intuitively, we note no significant differences between winter and summer rates, and daylight models remain non-significant. Cross-country differences are statistically detectable but substantively tiny and are entangled with cross-lingual confirmation bias. Alarming, our community response analysis reveals that stress sharing attracts supportive responses, but suicidal-ideation disclosures received the most fragile response environment (36.5% of responses were harmful/dismissive).

## Introduction

Mental health disorders, such as depression, anxiety, burnout, loneliness, suicidal ideation, and substance misuse, are becoming a major and growing challenge across the world. Interestingly, the Nordic countries paradoxically combine high reported well-being (Helliwell et al. 2023) with some of Europe’s highest antidepressant prescription rates (OECD 2023) and self-reported depression diagnoses (World Health Organization 2023). A potential antecedent can be seasonality: at latitudes above 55N, winter daylight falls below six hours per day, and photoperiod stress is a well-established trigger for Seasonal Affective Disorder (SAD), which affects 1–3% of the Nordic population with subsyndromal “winter blues” affecting a further 10–20% (Lam and Levitan 2000; Rosenthal et al. 1984; Kurlansik and Ibay 2012). The Baltic states share a high-latitude climate but differ markedly in healthcare infrastructure, cultural attitudes toward mental illness, and historical context (Semrau et al. 2011), creating an ideal context to examine public discourse on mental health.

Traditional mental health surveillance relies on clinical records and periodic surveys, which are costly, time-consuming, and subject to social-desirability bias (Chancellor and De Choudhury 2020). Reddit offers a complementary signal. Here, users disclose mental health experiences pseudonymously, in long-form, and in real time (Coppersmith, Dredze, and Harman 2014; Gkotsis et al. 2017; Wolohan et al. 2018). Prior computational work has studied depression (Yates, Cohan, and Goharian 2017), suicidality (Ji et al. 2020), anxiety (Shen and Rudzicz 2017; Tsakalidis et al. 2022), and stress (Turcan and McKeown 2019; Birnbaum et al. 2017) on Reddit, and offered meaningful insights. These studies, however, focus primarily on English-language *dedicated support communities* such as r/depression and r/SuicideWatch in single-country datasets. Intuitively, the vast majority of Reddit users in non-English-speaking countries may discuss mental health not in specialist forums but embedded within general national subreddits, in their native language, along with everyday topics - a setting that has received almost no computational attention.

Conceptually, we do not treat discourse in general subreddits as a proxy of population mental health. Instead, it is a *platform-conditioned disclosure signal*. Hence, the temporal, geographic, and reply patterns should not be interpreted as direct proxies of mental health in these countries. Overall, this study makes four incremental contributions to the extant literature as follows:

1. **Temporal Analysis.** A systematic examination of LLM-filtered disclosure rates across monthly, seasonal, weekly, and diurnal scales in high-latitude communities.
2. **Cross-country Analysis.** Comparison across 7 countries together with within-country analyses of language choice and subreddit scale effects.
3. **Community Response Analysis.** A reply-level study of how Reddit users respond to confirmed disclosures, distinguishing supportive, uncritical, and harmful reply patterns across mental health categories.
4. **Dataset Preparation.** A multilingual Reddit corpus (8 languages, 29 subreddits, 220,544 comments in the reported corpus, twelve months) and a two-stage approach combining a 600-term multilingual keyword lexicon with full GPT-4o-mini annotation - as a discourse-measurement tool.

## Related Work

Research on mental health in social media has gained momentum over the past decade. Coppersmith, Dredze, and Harman (2014) showed that language patterns on Twitter discriminate users with diagnosed depression, PTSD, and bipolar disorder. Subsequently, Yates, Cohan, and Goharian (2017) extended this to predicting self-disclosed depression from posting history in Reddit. Gkotsis et al. (2017) also characterized the linguistic properties of different disorder categories in Reddit mental health posts. Similarly, Ji et al. (2020) demonstrated that transformer models trained on Reddit can detect suicide risk with clinically meaningful sensitivity. Sabri et al. (2024) also showed that burnout discourse can be inferred across heterogeneous communities. LLMs have more recently been validated as annotators for sensitive mental health content, achieving near-human agreement on depression and anxiety tasks (Amin, Cambria, and Schuller 2023). Several prior studies have foregrounded a humanitarian perspective, with particular attention to the lived experiences and hardships of migrants as well as their mental health outcomes (Khatua and Nejd 2021; Khatua 2025). Across this body of work, the dominant pattern is English-language data from dedicated clinical forums (r/anxiety, r/ptsd, r/stress) (Turcan and McKeown 2019), where mental health disclosure is the community norm. Consequently, cross-cultural and multilingual work are crucial research gaps (Rai et al. 2025).

Only a handful of studies explored the **temporal dimension** in this domain. De Choudhury et al. (2013) observed elevated suicidal language on Mondays and in early-morning hours. Dutta et al. (2021) also confirmed diurnal variation in a suicide-support forum. Other studies also echo this trend. For example, MacAvaney et al. (2018) showed that temporal annotation materially affects how mental health timelines should be interpreted. Ashokkumar and Pennebaker (2021) showed that city-level Reddit language tracks population-level psychological shifts over time.

Interestingly, the Nordic high-latitude context makes seasonality a natural antecedent. For instance, Haggarty et al. (2002) documented winter spikes in help-seeking for depression in Scandinavia. Hence, *whether SAD-consistent patterns survive in general subreddits, and whether Nordic and Baltic discourse communities respond differently to disclosures, are the unexplored questions this paper addresses*. Multilingual approaches using models such as XLM-R (Conneau et al. 2020) and multilingual sentence transformers (Reimers and Gurevych 2019) enable cross-lingual transfer, but these have rarely been applied to mental health specifically, and to our knowledge, none has covered the multi-lingual NordicBaltic context.

Another under-explored domain is **community response** - what happens after a mental health disclosure. Morini et al. (2025) showed that in dedicated Reddit mental health communities, receiving replies is critical for sustaining participant engagement, but community response dynamics differ substantially from general-purpose subreddits. Ferizaj et al. (2025) identified Yalom’s group therapeutic factors, particularly universality and altruism, as dominant mechanisms in anonymous Reddit peer-support discussions, measured us-

ing LLM-based thematic analysis. Accordingly, Rayland and Andrews (2023) also argued that peer support traverses shared lived experience and role modeling rather than information provision, thereby distinguishing it mechanistically from professional support. Donnelly et al. (2025) pointed out that negative and dismissive online responses are a measurable disadvantage for suicidal disclosures on social media. Notably, all of this work examines moderated, dedicated clinical communities. The response environment surrounding disclosures in general national subreddits, where no community norms or automated moderation guardrails exist, has not been studied to the best of our knowledge.

**Research Gaps:** As noted, extant literature on social-media mental health has three broad limitations: (1) it primarily uses English data exclusively (Yates, Cohan, and Goharian 2017; Ji et al. 2020; Gkotsis et al. 2017), (2) it focuses on dedicated clinical forums where disclosure is the community norm rather than from general communities where it is the exception (Turcan and McKeown 2019), and (3) it rarely examines how communities respond once a disclosure is made (Morini et al. 2025). These gaps lead to the following intriguingly interlinked research questions (RQ):

- RQ1. Frequency Analysis:** How often does mental health discourse appear in Nordic and Baltic subreddits?
- RQ2. Temporal Patterns:** How do they vary across seasonal, weekly, and diurnal scales in this high-latitude setting on general subreddits?
- RQ3. Cross-Country Patterns:** How do the discourses differ across countries, and how do language choice and subreddit scale shape disclosure context?
- RQ4. Community Response Patterns:** When mental health discourse receives direct replies, what forms of community support emerge, and how does the response environment vary by disclosure severity?

## Data Collection

**Subreddit Selection:** We explore Reddit communities associated with 8 Nordic and Baltic countries, plus 3 regional subreddits (r/Nordiccountries, r/BalticStates, r/scandinavia). We had to exclude Latvia due to data scarcity and insufficient keyword-lexicon coverage. For each retained country, we include the primary national subreddit and, where available, city-level or English-oriented subreddits. Table 1 summarizes the 29 subreddits across 8 country groups.

**Dataset Collection:** Data were collected using PRAW (Boe 2024) from February 2025 to January 2026. For each subreddit, we retrieved up to 200 top-ranked threads, and for each thread, all comments with at least 30 characters. Language was identified per comment using langdetect (Danilak 2014). Our initial corpus contained 230,321 comments from 31 subreddits across 8 countries. After excluding Latvia from the main 7-country analysis due to data scarcity and insufficient lexicon coverage, the reported analytical corpus comprises 220,544 comments from 29 subreddits.

**Sampling note:** Top-thread sampling over-represents high-engagement content. Thus, all findings should be interpreted as *high-visibility* discourse rather than baseline ac-

Table 1: Subreddits per country group.

Country	Count	Subreddits
Norway	5	r/norge, r/Norway, r/oslo, r/Bergen, r/trondheim
Sweden	4	r/sweden, r/stockholm, r/Gothenburg, r/malmo
Denmark	4	r/Denmark, r/copenhagen, r/Aarhus, r/odense
Finland	5	r/Suomi, r/Finland, r/helsinki, r-Turku, r/tampere
Iceland	2	r/Iceland, r/Reykjavik
Estonia	3	r/Eesti, r/Tallinn, r/Tartu
Latvia	2	r/latvia, r/Riga
Lithuania	3	r/lithuania, r/Vilnius, r/Kaunas
Regional	3	r/Nordiccountries, r/BalticStates, r/scandinavia

tivity. High-visibility posts also attract more replies, meaning the community-response analysis reflects the response environment for the most-seen disclosures, which may differ from disclosures in low-traffic threads. Table 2 gives the country-wise breakdown.

**Language Distribution:** Automatic language detection reveals 8 dominant languages. English accounts for 33.5% of all comments (77,128), across several subreddits. The 5 Nordic and 2 Baltic languages together cover the remaining majority, emphasizing the need for our multilingual approach.

Table 2: Dataset summary by country group.

Country	Threads	Comments	Subreddits
Norway	1,000	46,927	5
Sweden	659	35,598	4
Denmark	800	41,316	4
Finland	1,000	41,717	5
Iceland	254	7,218	2
Estonia	400	13,965	3
Lithuania	400	11,754	3
Regional	414	22,049	3
<b>Total</b>	<b>4,927</b>	<b>220,544</b>	<b>29</b>

## Methodology

**Multilingual Keyword Filtering:** We developed a hand-crafted mental health keyword lexicon covering six categories - *depression*, *anxiety*, *stress/burnout*, *loneliness*, *suicidal ideation*, and *substance abuse* - across 7 languages. Each comment is matched against the lexicon for its detected language, and, if a match is found, we label it as a *mental health candidate*.

**LLM-based Annotation:** Next, we employ GPT-4o-mini (OpenAI 2024) for multi-label classification. For each comment, we consider whether it contains a personal or proximate experiential reference to a mental health condition, ranging from first-person self-report to close-third-

person reference to a family member or intimate contact, as distinct from general discussion, news references, metaphorical usage, or moderation commentary. Each comment receives a binary label indicating its category, a confidence rating (high/medium/low), and a `none` flag for non-disclosure use.

Given the absence of corpus-wide human adjudication, we had to consider LLM-based classification. For robustness, we re-annotate all 759 confirmed disclosures using two independent LLMs, i.e., GPT-4o and Claude Haiku (Anthropic 2024), with the same prompt and taxonomy. Pairwise Cohen’s  $\kappa$  across the three annotators (GPT-4o-mini, GPT-4o, Claude Haiku) is reported in Table 3. Mean  $\kappa$  ranges from 0.833 (GPT-4o-mini vs. Claude) to 0.868 (GPT-4o vs. Claude), with an overall mean of 0.854 - in the “almost perfect” agreement band by Landis and Koch (Landis and Koch 1977). The lowest individual category is depression ( $\kappa = 0.753$ , “substantial”). Suicidal ideation and anxiety exceed  $\kappa = 0.87$  across all pairs. The primary 7-country analysis considers only “high” confidence comments.

In our robustness analysis, we explore the *stability of the filter*. Accordingly, we use multi-LLM agreement, threshold sensitivity, and a small qualitative manual audit to explore the reliability of our methodological approach - discussed elaborately in the limitations section. For instance, our threshold sensitivity check compares key findings under high-only versus high + medium inclusion.

Table 3: Pairwise Cohens  $\kappa$  across three independent LLM annotators. All mean pairwise  $\kappa$  values exceed 0.80, indicating near-perfect agreement.

Annotator pair	Depr.	Anx.	Stress	Lone.	Suic.	Subs.	Mean $\kappa$
GPT-4o-mini vs GPT-4o	0.769	0.934	0.813	0.887	0.936	0.829	<b>0.861</b>
GPT-4o-mini vs Claude	0.753	0.896	0.807	0.879	0.831	0.832	<b>0.833</b>
GPT-4o vs Claude	0.833	0.910	0.785	0.931	0.870	0.878	<b>0.868</b>

Due to data scarcity, the reported 7-country corpus excludes Latvia for RQ1 and RQ2, leaving 5,817 keyword candidates and 759 high-confidence disclosures. However, the reply-level analysis (RQ4) considers the full 820 confirmed disclosures. We consider confirmed-disclosure rates normalized by total comment volume within the corresponding temporal unit (month, season, weekday, or local-time hour). To probe the daylight hypothesis, we consider the 7-country panel and employed regression analysis. For cross-country comparison we use two complementary summaries: a Kruskal–Wallis test on country-month confirmed rates, and a comment-level chi-square omnibus test with Cramér’s  $V$  as the effect-size measure. It is worth noting that comments are clustered within threads, subreddits, and countries, and confirmation rates vary by language. Therefore, these results should be interpreted carefully.

**LLM Prompt:** The reply-support analysis classifies each direct reply to a confirmed mental-health disclosure using GPT-4o-mini (OpenAI 2024), applying a structured response-style prompt. Each reply is assessed for: (1) *primary response type* from the nine-category taxonomy below, (2) *stance to poster* (supportive / neutral / dismissive), and (3) *support quality* (constructive / mixed / poor

/ harmful). Category-wise reply profiles are interpreted descriptively because disclosure categories are multi-label and some replied-category counts are modest.

*System:* You are annotating a reply to a confirmed mental health disclosure on Reddit. Classify the reply using the taxonomy below. Output JSON with keys: `primary_response_type`, `stance_to_poster`, `support_quality`.

*Taxonomy:* **shared\_experience** (responder discloses a similar personal struggle), **emotional\_support** (empathy or validation without personal disclosure), **informational\_support** (factual advice or coping strategies), **resource\_referral** (directs to professional help, hotlines, or services), **constructive\_challenge** (supportive disagreement or grounding), **uncritical\_validation** (mirrors the poster without grounding or nuance), **hostile\_dismissive** (minimises, mocks, or challenges the disclosure negatively), **harmful\_escalation** (reinforces risky, retaliatory, or hopeless framing), **neutral\_offtopic** (no substantive engagement with the mental health content).

## Findings & Limitations

### RQ1: Frequency Analysis

Unless otherwise stated, all reported mental-health disclosure rates are normalized by total comment volume for that respective category. Excluding Latvia, keyword filtering identifies 5,817 mental health disclosures (Stage 1 candidate rate: 2.64%). High-confidence GPT-4o-mini filtering retains only 759 disclosures (0.34% of the full corpus). This is substantially lower than rates typically observed in dedicated forums where disclosure is the community norm. (Morini et al. 2025). Table 4 presents category-wise prevalence within the retained 7-country confirmed-disclosure subset.

Table 4: Category label prevalence (multi-label, hence, percentages do not sum to 100%).

Category	Count	% of Confirmed
Substance abuse	298	39.3
Depression	165	21.7
Stress / burnout	130	17.1
Anxiety	116	15.3
Suicidal ideation	114	15.0
Loneliness	100	13.2

### RQ2: Temporal Patterns

Figure 1 reports the temporal pattern. We note that monthly trends vary within a narrow band, with the highest rate in December (0.53%) and a secondary elevation in late summer (August: 0.39%). Aggregated seasonally, Winter (0.38%) and Summer (0.36%) are close (ratio = 1.05). Furthermore, our regression analysis doesn't yield statistically significant results. In other words, we *do not detect a robust association with daylight*. Given the small number of countries and the possibility that the filter misses subtle or implicit cases, it would be problematic to conclude that no SAD-related effect exists.

Table 5: Cross-lingual confirmation rates for high-confidence disclosures (Latvia and Iceland excluded). The confirmation rate is defined as the proportion of keyword-matched candidates retained after LLM filtering. Differences across languages primarily reflect variation in the precision of the lexicon rather than true differences in disclosure prevalence.

Language	Candidates	Confirmed	Rate (%)
Finnish	150	51	34.0
Lithuanian	117	27	23.1
Norwegian	841	128	15.2
Swedish	1,175	159	13.5
English	1,715	231	13.5
Danish	1,236	120	9.7
Estonian	558	38	6.8

Category-level temporal structure is more differentiated. Anxiety remains winter-peaked, whereas suicidal ideation is relatively more prominent in Summer than Winter, consistent with the documented spring–summer peak in Scandinavian suicide rates (Partonen et al. 2004). Weekly and diurnal rhythms are clearer than the seasonal gradient: Monday has the highest confirmed rate (0.42%), confirming the Monday elevation in suicidal language reported by De Choudhury et al. (2013), and the hourly profile concentrates in overnight hours (01h–06h local time per country; peak at 06h: 0.71%), consistent with diurnal patterns observed in dedicated suicide-support forums by Dutta et al. (2021). Both patterns persist after country-specific time zone correction (UTC+0 to UTC+3), confirming that they are not artifacts of UTC.

### RQ3: Cross-Country Patterns

We find that Sweden has the highest observed disclosure rate (2.31%), followed by Estonia (1.78%) and Norway (1.71%), while Finland (1.24%), Lithuania (1.16%), and Iceland (0.68%) display lower disclosure rates. The effect is statistically detectable but substantively minimal. Also, Table 5 shows large cross-lingual confirmation differences, so these country contrasts are entangled with language-specific pipeline sensitivity. Country-category composition also varies within the retained subset. In the Nordic group, substance abuse is the most frequent confirmed label (37.9%), followed by depression (20.5%) and stress (17.5%). In Estonia and Lithuania combined, substance abuse is still dominant (49.3%), but depression (35.6%) and suicidal ideation (20.5%) are more prominent than in the Nordic group. However, these composition differences are best read as differences in the posts that survive the filter rather than direct cross-national comparisons of disclosure propensity. Figure 3 shows the full cross-country structure.

Within-country contextual analyses reveal mixed effects. On language of disclosure (Figure 4), Estonia and Sweden show higher native-language confirmed rates, while Finland shows a reverse gaps, offering no support for a universal native-language advantage. On subreddit scale (Figure 5), city subreddits exceed national ones in Denmark, Estonia,

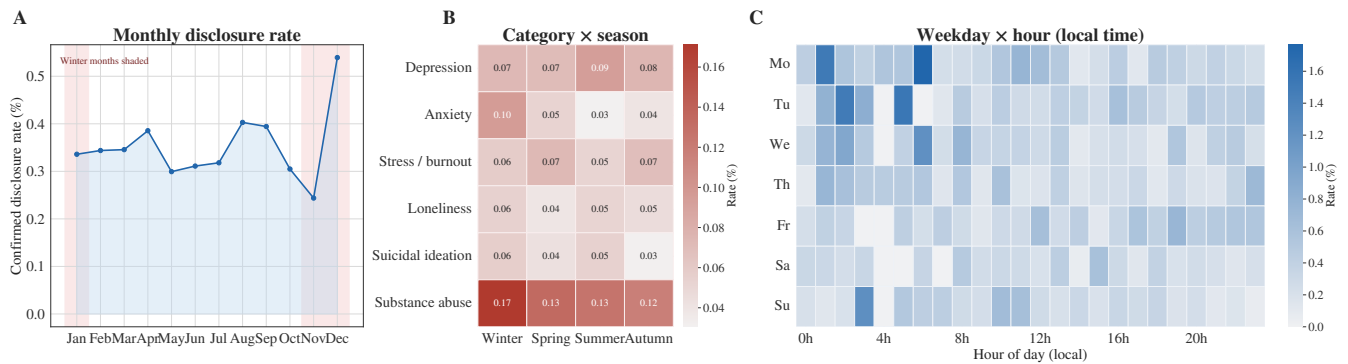


Figure 1: Temporal profile of mental-health disclosures; all timestamps converted to each country’s local time (UTC+0 to UTC+3). (A) Monthly disclosure rate. December is the highest, but the annual range is modest. (B) Category–season profile normalised by total seasonal volume. Anxiety and substance abuse show the strongest winter elevation, while suicidal ideation peaks in summer. (C) Weekday-hour heatmap (local time per country, UTC+0 to UTC+3) of confirmed disclosure rate. Overnight hours (01h-06h local time) and Monday show the strongest concentration.

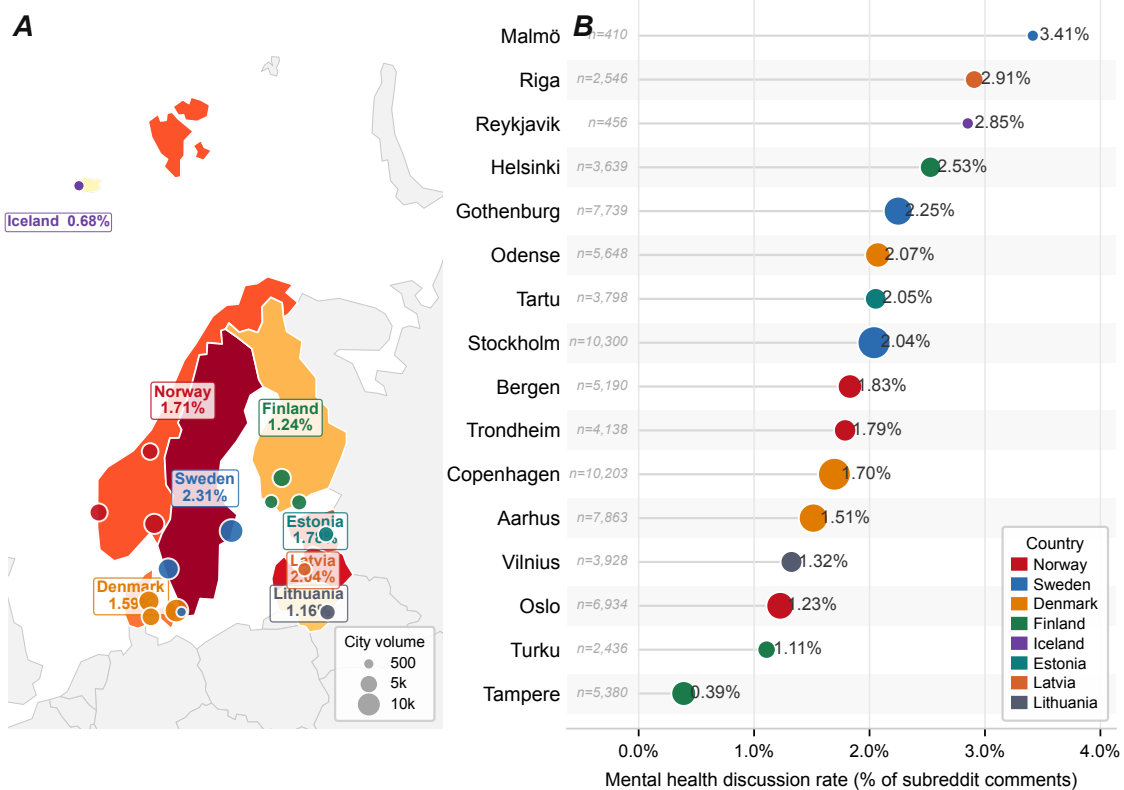


Figure 2: Geographic overview of mental-health disclosure rates. (a) Country-level rate across 8 countries. (b) Ranked rates for city subreddits with total comment volume.

Finland, Iceland, and Norway, but Sweden and Lithuania show the opposite pattern. Both analyses point to platform context as a moderator of observed disclosure behaviour rather than a universal driver.

### RQ4: Community Response Patterns

Prior work on mental health in social media primarily focuses on the *disclosure* side: detecting who discloses, when, and to what extent. The *response* side - what happens once a disclosure is made - has received much less attention, and what exists focuses on moderated, dedicated clinical com-

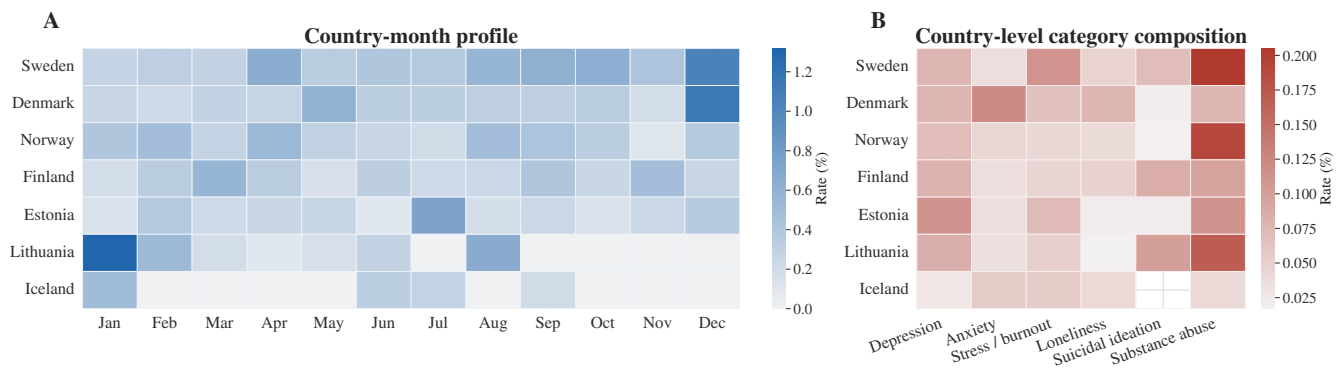


Figure 3: Cross-country structure of confirmed disclosures. (A) Country-month profile. (B) Country-level category composition, illustrating differences in retained category mix between the Baltic and Nordic subsets.

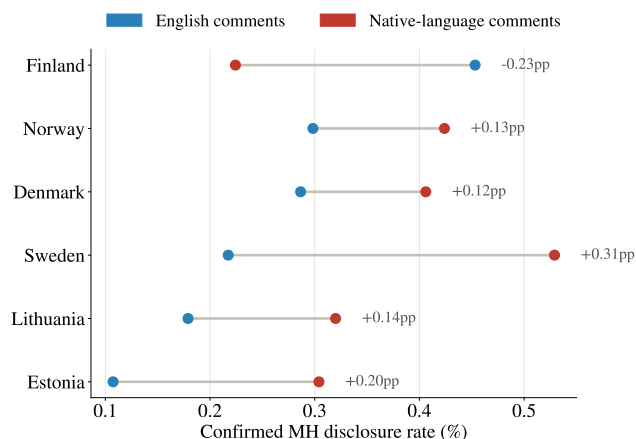


Figure 4: Mental-health disclosure rate by language (English vs. native) within each country. Native-language rates are higher in Sweden and Estonia. Finland shows a reverse gap, indicating that language-of-disclosure effects are context-dependent.

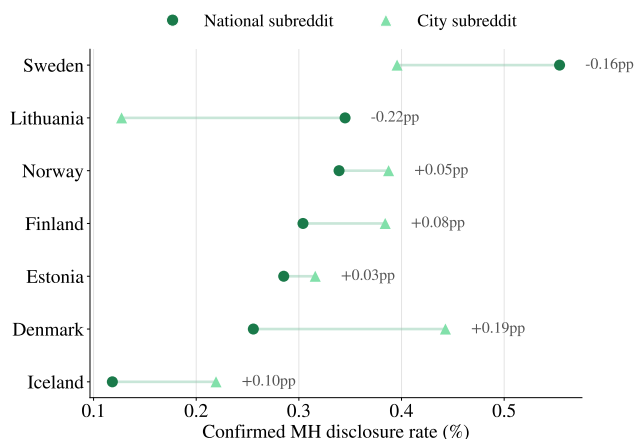


Figure 5: Mental-health disclosure rate in national versus city subreddits within each country. City subreddits are higher in Denmark, Estonia, Finland, Iceland, and Norway. Sweden and Lithuania show the reverse, indicating that subreddit-scale effects are country-dependent rather than universal.

munities such as *r/depression* and *r/SuicideWatch*, where community norms actively encourage supportive reply behavior (Morini et al. 2025; Ferizaj et al. 2025). General national subreddits are structurally different environments: mental health disclosure is incidental and off-norm, and there are no community moderation rules or auto-posted crisis links to shape replies. Hence, whether these spaces function as impromptu peer-support networks or expose disclosers to harmful responses is an unanswered question with direct implications for platform safety design.

To address this, we analyze the direct-reply chains of all 820 confirmed disclosures. The sub-50% reply rate (409 of 820 disclosures received at least one direct reply) is itself informative: in dedicated support forums, virtually every post receives a reply, because reply-giving is the community norm (Morini et al. 2025). Conversely, here, mental health content competes with the full range of everyday discussion topics, and a post that receives no reply simply goes unno-

ticed rather than being ignored. The 654 classified replies therefore represent *organic* peer responses to incidental disclosures in ordinary public discourse, structurally different from studies that considered dedicated forums.

Among replied disclosures, 72.4% receive at least one supportive reply and 18.1% receive at least one harmful or dismissive reply. At the reply level, LLM-classified response patterns suggest that the dominant primary type is *shared experience* (41.1% of replies), followed by *constructive challenge* (14.5%), *hostile/dismissive* (13.6%), *neutral/off-topic* (13.1%), *uncritical validation* (9.2%), and *informational support* (7.0%). *Emotional support* (0.6%), *resource referral* (0.5%), and *harmful escalation* (0.3%) are rare as primary reply styles.

As a robustness check, collapsing to a binary scheme (supportive = *shared\_experience* + *emotional\_support* + *informational\_support* + *resource\_referral* + *construc-*

tive\_challenge, non-supportive = hostile\_dismissive + harmful\_escalation + neutral\_offtopic + uncritical\_validation) yields 63.7% supportive primary replies overall, and the suicidality gap remains the largest cross-category contrast. Peer response is expressed primarily through personal testimony rather than overt counseling or resource referral - consistent with Yalom's *universality* factor (Ferizaj et al. 2025) and with the peer-vs-professional distinction (Rayland and Andrews 2023). The very low resource-referral rate (0.5%) indicates that these communities rarely route disclosers to formal services (Morini et al. 2025). The category profile is uneven (Figure 6). Stress disclosures attract the most supportive reply environment (95.2% receive at least one supportive reply, only 6.0% receive a harmful/dismissive reply), whereas suicidal-ideation disclosures receive the most fragile response mix (60.3% supportive, 36.5% harmful/dismissive). Overall, the social response environment varies materially by disclosure severity - a pattern that holds under both the fine-grained taxonomy and the binary collapsed scheme.

## Limitations

At the outset, we acknowledge that online platforms may not reflect reality. For example, Reddit users are primarily younger and more educated (Pew Research Center 2021), making them demographically unrepresentative. Despite this limitation, online platforms provide a timely measure of how mental-health-related disclosure and reply behavior appear in sampled high-visibility threads.

Some of the limitations of our study open avenues for future research. For instance, we mostly rely on a multi-LLM agreement that demonstrates consistency. However, we also conducted an informal **manual annotation** of two small random samples from the annotated pool: 60 LLM-rejected candidates and 30 medium-confidence positives. Across the 60 randomly sampled rejected candidates, we note generic policy/news discussion, and idiomatic or otherwise non-self-disclosure uses of terms such as *depression*, *anxiety*, *addiction*, or *loneliness*. Across the 30 sampled medium-confidence positives, we mostly find borderline cases in which the context was too weak or figurative to justify a high-confidence label. This indicates that the high-only threshold is conservative: it removes a large amount of obvious keyword noise and many borderline cases, while also likely undercounting mild, implicit, or weakly contextualized disclosure.

We note **cross-lingual pipeline bias** (Table 5). Confirmation rates within keyword-matched disclosures vary substantially by detected language (Table 5). Finnish (34.0%) and Lithuanian (23.1%) show markedly higher within-candidate confirmation rates than Danish (9.7%) and Estonian (6.8%), suggesting that keyword-lexicon precision differs across languages. This variation means that cross-language and cross-country comparisons of confirmed-disclosure counts may partly reflect differences in detection sensitivity. Therefore, cross-country comparisons should be interpreted with this possible limitation in mind.

Additionally, keyword detection fails to detect implicit distress expressions lacking recognized vocabulary. Also,

top-thread sampling over-represents high-engagement content and under-represents low-visibility disclosures. Intuitively, our proposed approach may be more applicable for *high-visibility discourse*. Hence, we conducted a **threshold sensitivity analysis**. To assess whether the choice of confidence threshold drives the main findings, we repeat the key analyses, including medium-confidence labels (high+medium: 1,550 disclosures vs. 759 high-only). The core qualitative conclusions are stable: Monday remains the peak weekday under both thresholds, and the winter-to-summer ratio remains consistent - 1.05 (high-only) versus 1.01 (high+medium) - both near-null and directionally consistent. Combined with the audit of medium-confidence cases above, this suggests that the high-only threshold mainly trades recall for a cleaner retained set rather than flipping the substantive temporal conclusions. Finally, hourly and weekday analyses use country-specific local time (UTC+0 to UTC+3), and monthly and seasonal analyses use UTC, where a  $\leq 3$  h offset does not significantly affect month or seasonal pattern.

## Discussion

**What the LLM Filter Measures:** Our multi-LLM agreement indicates that the retained labels are stable, but our manual audit suggests that this approach is not impeccable. This LLM-based approach produces a consistent *measurement* of explicit, high-confidence disclosure under stated assumptions, but may ignore weakly contextualized first-person distress. We also note that the temporal pattern is slightly counterintuitive. Winter and summer rates are broadly similar (ratio = 1.05), and the 7-country daylight tests remain non-significant. We therefore do not observe a robust daylight association in this dataset, but a 'no SAD effect' conclusion might be problematic because the country panel is small and subtle or implicit disclosures may be filtered out. What appears more stable is a December spike, a Monday concentration, and consistent overnight elevation (01h–06h local time). The December anomaly may reflect holiday-related discussion or year-end reflection rather than photoperiod, whereas the weekly and diurnal effects may serve as a (weak) indicator of underlying societal distress.

### Cross-Country Differences and Contextual Factors:

As noted, our analysis based on community-specific Reddit data may not reflect the offline reality. Despite this limitation, Sweden's higher observed rate and Iceland's lower one are descriptively true within this corpus, but the effect size is nominal, and cross-lingual confirmation rates differ sharply. Hence, we may conclude that general-community disclosure looks different across language communities and subreddit ecologies, while country-level magnitude differences remain highly confounded by measurement sensitivity.

Interestingly, online and offline rankings did not converge. However, this does not make the Reddit signal useless, but it means the signal is better understood as public expression in a platform setting than as a proxy for reality. Language choice and subreddit scale also matter, but not uniformly. The absence of a consistent native-language advantage and the reversal of the city-effect sign across countries suggest that context moderates disclosure behaviour in

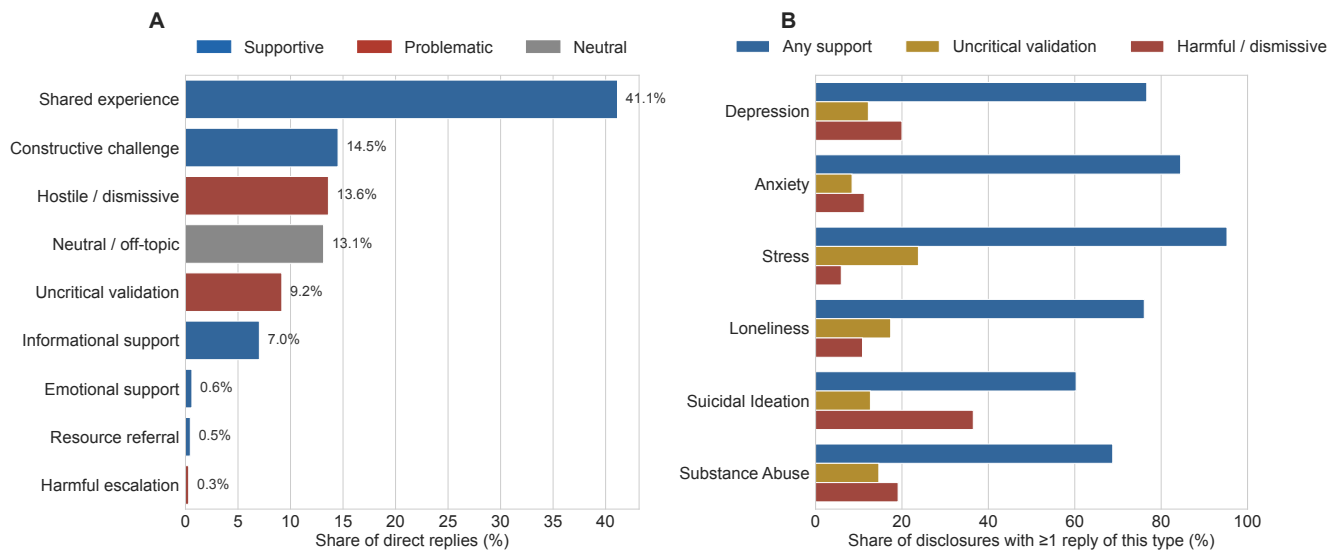


Figure 6: Community response to disclosures. (A) Distribution of primary reply response types. Shared experience dominates, formal emotional support and resource referral are rare. (B) Share of replied disclosures receiving at least one supportive, uncritical, or harmful/dismissive reply by MH category. Stress attracts the safest response environment, suicidality the least reliable.

ways that aggregate country-level analysis cannot fully capture.

**Community Response and the Suicidality Gap:** To our knowledge, this is the first reply-level analysis of mental health disclosure responses in multilingual, general-community Reddit across Nordic and Baltic countries. As noted, prior work on community response is confined to dedicated clinical subreddits (Morini et al. 2025; Ferizaj et al. 2025), where reply-giving is a community norm and the response environment is shaped by explicit moderation policies. Without moderation policies, the 41.1% shared-experience dominance is not an artifact of a help-seeking community but a recurring response mode to *incidental* disclosure in ordinary public discourse. The result aligns with prior work on universality and peer testimony (Ferizaj et al. 2025; Rayland and Andrews 2023), but it also shows the limits of informal peer responses: explicit resource referrals are extremely rare.

The more actionable finding is the suicidality gap. While stress disclosures attract a largely safe response environment (95.2% supportive), suicidal-ideation disclosures attract the most harmful/dismissive replies (36.5%). Donnelly et al. (2025) documents a comparable pattern in broader social-media suicidal disclosure data, identifying negative responses, including dismissal and the framing of disclosures as attention-seeking, as a measurable disadvantage of public online disclosure. In our setting, community support is least reliable where disclosure severity is highest. Dedicated clinical subreddits such as r/SuicideWatch use moderation rules and auto-posted crisis resources to buffer that risk, but general national subreddits do not. Therefore, the platform safety tooling should pay particular attention to suicidal disclosure in ordinary community spaces rather than

only in specialist forums.

## Conclusion

We explored a multilingual corpus of mental-health-related discourse in Nordic and Baltic Reddit, covering 220,544 comments in 8 languages across 7 countries over a full annual cycle. Unlike prior studies, we have not considered focused subreddits, but generic subreddits. The core contribution of our study is community response: peer sharing dominates, but the reply environment for suicidal-ideation disclosures is a concern for platform safety tooling. Counterintuitively, daylight-linked seasonality is not evident in our analysis, and raw cross-country differences are both tiny in effect size and confounded by cross-lingual confirmation bias. However, Monday and overnight concentrations (01h-06h local time) survive as the more stable temporal patterns.

Overall, our findings suggest that general-purpose subreddits in local languages contain an interpretable, platform-conditioned signal of high-visibility mental-health disclosure and reply behavior, but this signal should be treated as a discourse measure rather than as a prevalence proxy or a country-burden ranking. Reddit is therefore best used as a complementary source for monitoring public expression and response behavior in multilingual, high-latitude communities.

## Ethical Statement

This study uses publicly accessible Reddit content in read-only mode and analyzes it only at the aggregate level. We do not infer individual diagnoses, identities, or treatment status, and we avoid reproducing identifiable user content. Findings should inform aggregate monitoring and hypothesis generation, not automated decision-making about individual users.

## Acknowledgments

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

## GenAI Usage Disclosure

LLMs were used in the research process to improve code for executing experiments, clarify language, correct grammatical errors, and create figure layouts. LLMs include ChatGPT and Claude Code, with different versions being used for different steps. We also use LLM for the classification task.

## References

- Amin, M.; Cambria, E.; and Schuller, B. W. 2023. Will Affective Computing Emerge from Foundation Models and General Artificial Intelligence? A First Evaluation of ChatGPT. *IEEE Intelligent Systems*, 38(2): 15–23.
- Anthropic. 2024. Claude Haiku: Fast and compact AI model. <https://www.anthropic.com/claude>.
- Ashokkumar, A.; and Pennebaker, J. W. 2021. Social Media Conversations Reveal Large Psychological Shifts Caused by COVID-19's Onset Across U.S. Cities. *Science Advances*, 7(39).
- Birnbaum, M. L.; Ernala, S. K.; Rizvi, A. F.; De Choudhury, M.; and Kane, J. M. 2017. A Collaborative Approach to Identifying Social Media Markers of Schizophrenia by Employing Machine Learning and Clinical Appraisals. *Journal of Medical Internet Research*, 19(8).
- Boe, B. 2024. PRAW: The Python Reddit API Wrapper. <https://praw.readthedocs.io/>. Version 7.x.
- Chancellor, S.; and De Choudhury, M. 2020. Methods in Predictive Techniques for Mental Health Status on Social Media: A Critical Review. *npj Digital Medicine*, 3: 43.
- Conneau, A.; Khandelwal, K.; Goyal, N.; Chaudhary, V.; Wenzek, G.; Guzmán, F.; Grave, .; Ott, M.; Zettlemoyer, L.; and Stoyanov, V. 2020. Unsupervised Cross-lingual Representation Learning at Scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, ACL 2020, 8440–8451. Association for Computational Linguistics.
- Coppersmith, G.; Dredze, M.; and Harman, C. 2014. Quantifying mental health signals in Twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, 51–60.
- Danilak, M. 2014. langdetect: Port of Google's language-detection library to Python. <https://pypi.org/project/langdetect/>.
- De Choudhury, M.; Gamon, M.; Counts, S.; and Horvitz, E. 2013. Predicting Depression via Social Media. In *Proceedings of the 7th International AAAI Conference on Weblogs and Social Media*, ICWSM 2013, 128–137. AAAI Press.
- Donnelly, H. K.; Solberg, V. S. H.; Green, J. G.; Wijaya, D.; and Jobes, D. A. 2025. Patterns of Suicidal Stress Disclosure on Social Media: Integrating Computational and Qualitative Approaches. medRxiv preprint.
- Dutta, R.; Gkotsis, G.; Velupillai, S.; Bakolis, I.; and Stewart, R. 2021. Temporal and Diurnal Variation in Social Media Posts to a Suicide Support Forum. *BMC Psychiatry*, 21: 259.
- Ferizaj, D.; Lalk, C.; Lahmann, N.; Strube-Lahmann, S.; and Rubel, J. 2025. Identifying Yalom's group therapeutic factors in anonymous mental health discussions on Reddit: a mixed-methods analysis using large language models, topic modeling and human supervision. *Frontiers in Psychiatry*, 16.
- Gkotsis, G.; Oellrich, A.; Velupillai, S.; Liakata, M.; Hubbard, T. J. P.; Dobson, R. J. B.; and Dutta, R. 2017. Characterisation of Mental Health Conditions in Social Media Using Informed Deep Learning. In *Scientific Reports*, volume 7.
- Haggarty, J. M.; Cernovsky, Z.; Husni, M.; Minor, K.; Kermeen, P.; and Merskey, H. 2002. Seasonal Affective Disorder in an Arctic Community. *Acta Psychiatrica Scandinavica*, 105(5): 378–384.
- Helliwell, J. F.; Layard, R.; Sachs, J. D.; and De Neve, J.-E. 2023. *World Happiness Report 2023*. New York: Sustainable Development Solutions Network.
- Ji, S.; Pan, S.; Li, X.; Cambria, E.; Long, G.; and Huang, Z. 2020. Suicidal ideation detection: A review of machine learning methods and applications. *IEEE Transactions on Computational Social Systems*, 8(1): 214–226.
- Khatua, A. 2025. Policies for Refugee Mental Health Using Generative Agents. In *Companion Proceedings of the ACM on Web Conference 2025*, 1062–1066.
- Khatua, A.; and Nejdil, W. 2021. Struggle to settle down! Examining the voices of migrants and refugees on Twitter platform. In *Companion publication of the 2021 conference on computer supported cooperative work and social computing*, 95–98.
- Kurlansik, S. L.; and Ibay, A. D. 2012. Seasonal Affective Disorder. *American Family Physician*, 86(11): 1037–1041.
- Lam, R. W.; and Levitan, R. D. 2000. Pathophysiology of Seasonal Affective Disorder: A Review. *Journal of Psychiatry & Neuroscience*, 25(5): 469–480.
- Landis, J. R.; and Koch, G. G. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1): 159–174.
- MacAvaney, S.; Desmet, B.; Cohan, A.; Soldaini, L.; Yates, A.; Zirikly, A.; and Goharian, N. 2018. RSDD-Time: Temporal Annotation of Self-Reported Mental Health Diagnoses. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 168–173. New Orleans, LA: Association for Computational Linguistics.
- Morini, V.; Sansoni, M.; Rossetti, G.; Pedreschi, D.; and Castillo, C. 2025. Participant behavior and community response in online mental health communities: Insights from Reddit. *Computers in Human Behavior*, 165.
- OECD. 2023. Health at a Glance 2023: OECD Indicators.
- OpenAI. 2024. GPT-4o Technical Report. <https://openai.com/research/gpt-4o>.

Partonen, T.; Haukka, J.; Viilo, K.; Hakko, H.; Pirkola, S.; Isometsä, E.; Lönnqvist, J.; Särkioja, T.; Väisänen, E.; and Räsänen, P. 2004. Cyclic time patterns of death from suicide in northern Finland. *Journal of affective disorders*, 78(1): 11–19.

Pew Research Center. 2021. Social Media Use in 2021. <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/>.

Rai, S.; Shelat, K.; Jain, D.; Kishen, A.; Cho, Y. M.; Redkar, M.; Hardikar-Sawant, S.; Ungar, L.; and Guntuku, S. C. 2025. Cross-Cultural Differences in Mental Health Expressions on Social Media. In *Proceedings of the 3rd Workshop on Cross-Cultural Considerations in NLP (C3NLP 2025)*, 132–142. Albuquerque, New Mexico: Association for Computational Linguistics.

Rayland, A.; and Andrews, J. 2023. From Social Network to Peer Support Network: Opportunities to Explore Mechanisms of Online Peer Support for Mental Health. *JMIR Mental Health*, 10.

Reimers, N.; and Gurevych, I. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019*, 3982–3992. Association for Computational Linguistics.

Rosenthal, N. E.; Sack, D. A.; Gillin, J. C.; Lewy, A. J.; Goodwin, F. K.; Davenport, Y.; Mueller, P. S.; Newsome, D. A.; and Wehr, T. A. 1984. Seasonal Affective Disorder: A Description of the Syndrome and Preliminary Findings with Light Therapy. *Archives of General Psychiatry*, 41(1): 72–80.

Sabri, N.; Pham, A. C.; Kakkar, I.; and ElSherief, M. 2024. Inferring Mental Burnout Discourse Across Reddit Communities. In *Proceedings of the Third Workshop on NLP for Positive Impact, 224–231*. Miami, Florida, USA: Association for Computational Linguistics.

Semrau, M.; Barley, E. A.; Law, A.; and Thornicroft, G. 2011. Lessons Learned in Developing Community Mental Health Care in Europe. *World Psychiatry*, 10(3): 217–225.

Shen, J. H.; and Rudzicz, F. 2017. Detecting Anxiety through Reddit. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality*, 58–65. Vancouver, BC: Association for Computational Linguistics.

Tsakalidis, A.; Chim, J.; Bilal, I. M.; Zirikly, A.; Atzil-Slonim, D.; Nanni, F.; Resnik, P.; Gaur, M.; Roy, K.; Inkster, B.; et al. 2022. Overview of the CLPsych 2022 shared task: Capturing moments of change in longitudinal user posts. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, 184–198.

Turcan, E.; and McKeown, K. 2019. Dreddit: A Reddit Dataset for Stress Analysis in Social Media. In *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, 97–107. Association for Computational Linguistics.

Wolohan, J.; Hiraga, M.; Mukherjee, A.; Sayyed, Z. A.; and Millard, M. 2018. Detecting linguistic traces of depression in topic-restricted text: Attending to self-stigmatized depression with NLP. In *Proceedings of the first international workshop on language cognition and computational models*, 11–21.

World Health Organization. 2023. Mental Health Atlas 2023.

Yates, A.; Cohan, A.; and Goharian, N. 2017. Depression and Self-Harm Risk Assessment in Online Forums. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017*, 2968–2978. Association for Computational Linguistics.

## Paper Checklist

### 1. For most authors...

- (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? **No**
- (b) Do your main claims in the abstract and introduction accurately reflect the paper’s contributions and scope? **Yes**
- (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? **Yes**
- (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? **Yes**
- (e) Did you describe the limitations of your work? **Yes**
- (f) Did you discuss any potential negative societal impacts of your work? **NA**
- (g) Did you discuss any potential misuse of your work? **NA**
- (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? **NA**
- (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? **Yes**

### 2. Additionally, if your study involves hypotheses testing...

- (a) Did you clearly state the assumptions underlying all theoretical results? **NA**
- (b) Have you provided justifications for all theoretical results? **NA**
- (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? **NA**
- (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? **NA**
- (e) Did you address potential biases or limitations in your theoretical framework? **NA**
- (f) Have you related your theoretical results to the existing literature in social science? **NA**

- (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
- (a) Did you state the full set of assumptions of all theoretical results? NA
  - (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
- (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA
  - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
  - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
  - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
  - (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA
  - (f) Do you discuss what is “the cost“ of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity**...
- (a) If your work uses existing assets, did you cite the creators? NA
  - (b) Did you mention the license of the assets? NA
  - (c) Did you include any new assets in the supplemental material or as a URL? NA
  - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? NA
  - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? NA
  - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR ? NA
  - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset ? NA
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity**...
- (a) Did you include the full text of instructions given to participants and screenshots? NA
  - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? NA
  - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
- (d) Did you discuss how data is stored, shared, and de-identified? NA