Digital Mental Health Discourse Among Singaporean Youth: A Topic Modeling Analysis and Comparison to International Platforms

Onno P. Kampman[1], Ye Sheng Phang[1], Jin Yan[1], Ikhoon Eom[2], Surya Appana[3], Jia Jun Tan[2], Victor Zhang[2], Hazirah Hoosainsah[1], Xinyi Hong[1], Charmaine Lim[1], Caleb Tan[1], Creighton Heaukulani[1], Robert JT Morris[1], Janice Huiqin Weng[1]

[1]MOH Office for Healthcare Transformation, Singapore

[2]Department of Mathematics, National University of Singapore, Singapore

[3]Department of Mathematics, UC Berkeley, United States of America

Corresponding Author:

Onno P. Kampman

MOH Office for Healthcare Transformation

onno.kampman@moht.com.sg

Journal options (impact factor)

1. Lancet Digital Health (24.1)
2. Npj Digital Medicine (15.1)
3. Npj Mental Health
4. PLOS Medicine (10.0)
5. JMIR Public Health (6.6)
6. JMIR (6.0)
7. JMIR Mental Health (5.8)
8. BMC Public Health (3.6)
9. BMC Psychiatry (3.6)
10. Frontiers in Digital Health (3.8)

Table of Contents

Abstract (250-300)

Background: Youth worldwide are increasingly experiencing mental health symptoms, prompting prioritization by care providers and public health institutions. However, the drivers of this trend remain largely unclear. Digital mental health forums provide unprecedented opportunities to understand youth mental health needs at scale, yet most analyses focus on Western populations, potentially missing culturally specific patterns that could inform targeted interventions.

Objective: To identify and characterize mental health discussion themes among Singaporean youth on a moderated digital peer support forum using automated topic modeling and to examine how cultural and demographic factors influence online mental health discourse.

Methods: We analyzed 1,921 discussions from *let's talk*, a youth-focused mental health forum in Singapore, automatically (through machine learning) uncovering, identifying, and extracting topics and their prevalences using BERTopic and LLM models. The relationships between topics and user engagement metrics are explored. To contextualize demographic and moderation differences, we conducted comparative analyses with the topic distributions inferred from mental health discussions on Reddit (Anglosphere-centric) and HardwareZone (Singapore-based, adult-centric) forums.

Results: The most frequently discussed topics on *let's talk* were "Stress and Anxiety", "Friendship Issues", and "Academic Stress", accounting for nearly 50% of posts. Topics related to interpersonal issues such as "Relationship Breakup" and "Family Dynamics" resulted in more lengthy discussions, while the highest number of replies were found for "Job Anxiety" and "Loneliness". Compared to international forums, Singaporean youth discussions showed significantly lower rates of substance abuse-related topics and used less clinical language. Youth on *let's talk* also expressed distinct stressor profiles, with academic pressure and family expectations featuring more prominently than employment-related stress. LLM results are broadly in line with BERTopic results, but the LLM identified relationship-specific and interpersonal functioning themes as most salient, whereas BERTopic highlighted stress- and anxiety-related clusters more strongly.

Conclusions: We automatically discovered and quantified topics discussed on *let's talk* using machine learning models, providing both a qualitative and quantitative description of stressors and mental health conversations amongst youth in Singapore. We additionally showed how themes on *let's talk* differ from those on forums with a different target audience, including international, adult-focused, and unmoderated (not tailored to mental health) forums, demonstrating the necessity of taking cultural and demographic differences into account. Our methodology provides a scalable framework for measuring and analyzing youth mental health needs in real-time as a complement to traditional, time-consuming, human-based thematic analysis. This approach can help to scale digital health services by extracting real-time, localized, actionable, and community-driven insights that may drive public health priorities and interventions.

Keywords: digital mental health, topic modeling, forums, public health, youth mental health, cross-cultural analysis, Singapore

Introduction (500-800)

Background

Mental health conditions among young people have risen sharply worldwide, placing unprecedented pressure on healthcare systems. Many individuals face treatment gaps, receiving inappropriate or no care (Chong et al., 2012; Gulliver et al., 2010). As digital natives, youth increasingly turn to online platforms for mental health support, making peer-to-peer forums a scalable complement to traditional services (Ali et al., 2015; Naslund et al., 2016; Park et al., 2018; Prescott et al., 2017; Smith-Merry et al., 2019; Stevens et al., 2022). Young people often find such forums more informative than static resources, and prefer discussing sensitive topics like self-harm anonymously rather than with family or friends (Jones et al., 2011). These platforms enable real-time emotional expression during distress and align with stepped-care models through accessible, lower-intensity interventions (Bond et al., 2023).

This digital shift also enables more naturalistic understanding of community mental health needs. Forum and social media content can be systematically analyzed to identify at-risk individuals (Chim et al., 2024) and derive population-level insights that complement traditional epidemiological studies. Such insights inform product and service enhancements, resource allocation, policy decisions, and mental health awareness campaigns, offering alternatives to resource-intensive public health studies.

Digital Mental Health Platforms

Mental health conversations occur across online platforms, where individuals build communities, improve mental health literacy, and exchange support (Gowen et al., 2012). Compared to general-purpose platforms, custom-built forums attract more targeted demographics and benefit from stronger moderation, producing more relevant and curated content (Li et al., 2021; Marshall et al., 2024). These platforms also allow for culturally nuanced expressions of distress and help-seeking.

In Singapore, the *let's talk* forum launched in October 2022 as an English-language community-based digital mental health and wellbeing platform targeted at individuals aged 17 to 35. Developed with government support and co-created with youth advocates (including those with lived experience) it provides a safe, anonymous, and moderated space for discussion with peers and trained professionals (Choudhury & De, 2014; Heaukulani et al., 2024; Phang et al., 2023; Samsuri, 2024; Weng et al., 2024). The platform facilitates authentic and culturally grounded discourse reflecting local mental health realities (Killacky, 2023).

Computational Approaches to Mental Health Analysis

Natural language processing (NLP) methods are increasingly used to analyze mental health discourse at scale, enabling automated screening, early detection, and thematic analysis (Cambria et al., 2024). These techniques identify patterns in large volumes of text that would be difficult to detect manually. While many studies focus on *screening* and *early detection* in individuals for proactive and preventative clinical outreach (Tsakalidis et al., 2022), which is believed to be crucial in effective mental health care (Cuijpers et al., 2012), this study instead aims to generate *population-level* insights into mental health needs.

A widely used NLP approach for corpus analysis is *thematic analysis*, a qualitative method that identifies recurring themes through systematic coding by following the six phases outlined by (Boyatzis, 1998; Braun & Clarke, 2006), from data familiarization to theme definition to report production. Although early and final stages typically require human involvement , intermediate phases can increasingly be supported by computational tools (De Paoli, 2024). Transformer-based embeddings and clustering algorithms, such as BERTopic (Grootendorst, 2022), automate theme extraction while maintaining conceptual clarity (Dan et al., 2025). More recently, large language models (LLMs) have enabled both *inductive* and *deductive* thematic analysis (Sankaranarayanan et al., 2025). This study employs BERTopic for theme extraction (phase 3), with deductive LLM-based classification using clinically meaningful themes providing comparative insights.

Research Gap and Cultural Context

Despite increasing interest in using social media data for mental health research, existing work remains largely focused on Western populations and global platforms. This may obscure culturally specific patterns and limit the relevance of findings for local interventions (Pendse et al., 2019). Help-seeking behaviors, symptom expression, and mental health discourse differ significantly across cultural contexts. Singapore's unique multicultural composition and distinct social pressures around academic achievement, family expectations, and enduring stigma shape how youth discuss and experience mental health challenges (Chong et al., 2007). However, non-Western datasets remain underrepresented in the literature, and there is limited understanding of culturally grounded digital expressions of psychological distress. Understanding these cultural and demographic differences is essential for designing effective, context-specific interventions that resonate with the local population.

Study Objectives

This study aims to advance understanding of youth mental health concerns in Singapore by applying automated topic modeling to analyze discussions on the *let's talk* forum. Through systematic analysis, we identify and characterize salient mental health themes and examine which topics generate the highest levels of user engagement. We further contextualize these findings through comparisons with the unmoderated forums of Reddit (representing global English-speaking users) and HardwareZone (representing Singapore's adult population). This study demonstrates how NLP methods can complement traditional epidemiological approaches and offer real-time community-specific insights for targeted public health interventions and resource allocation.

Methods (1000-1500)

In this section, we describe the dataset construction, thematic analysis methods, and evaluation studies.

Corpora Construction

We analyzed data from three sources: *let's talk* as our core dataset, with Reddit Mental Health and HardwareZone providing comparative mental health discussions on other social media platforms to investigate demographic and geographic differences, as well as differences potentially attributable to open versus monitored platforms.

The *let's talk* forum contains two major categories of threads: "Ask-a-Therapist" lets users post a question for a professional on-staff therapist who will respond for free within 24 hours, and "Hangouts" is the general space intended for more casual user-driven conversations. Both are professionally moderated, with toxic or sensitive posts removed and not included in our analysis, and assessed for risk types that may inform the triaging of users. The types of topics discussed and severity levels of symptoms are expected to differ between these two spaces, so we also analyze these two subcategories separately. Our dataset includes only original user posts (excluding community manager and counselor posts) to focus on grassroots, user-driven discussions, totaling 1,921 posts.

Reddit is an online forum with over 57 million daily active users, organized into subcommunities called "subreddits". We used the Reddit Mental Health dataset [1] containing anonymized conversations (original/initial posts only) from 28 subreddits, posted between 2018 and 2020, from 15 mental health support groups ("r/EDAnonymous", "r/addiction", "r/alcoholism", "r/adhd", "r/anxiety", "r/autism", "r/bipolarreddit", "r/bpd", "r/depression", "r/healthanxiety", "r/lonely", "r/ptsd", "r/schizophrenia", "r/socialanxiety", and "r/suicidewatch"), 2 broad mental health subreddits ("r/mentalhealth", "r/COVID19_support"), and 11 non-mental health subreddits (e.g., "r/conspiracy", "r/fitness", "r/jokes") that were included for comparison (Low et al., 2020). In this study, we considered the 15 mental health support and the "r/mentalhealth" subreddits, creating a combined dataset of ~300,000 posts. We conducted topic modeling on a uniform subsample of 10,000 posts. Additionally, we run the analysis on each subreddit individually. This dataset primarily represents Western, English-speaking mental health discourse patterns. Reddit's permissive data access policies make it common in research, exemplified by other open-source datasets like MentalHelp (Raihan et al., 2024) and RedditESS (Alghamdi et al., 2025).

HardwareZone is a Singaporean forum that evolved from hardware and electronics discussions into a general population forum. We focused specifically on the "Eat-Drink-Man-Woman" (EDMW) section, which is among the most active boards on the site. This subforum not only features high engagement and diverse conversations but also contains the majority of mental health–related posts identified in our exploratory analysis. To systematically identify relevant threads, we employed a structured keyword-matching approach on URLs (it was infeasible to scrape all data before filtering) informed by both thematic lexicons and core psychological terms. We first constructed a comprehensive set of keywords using GPT-4o by combining domain-specific phrases, covering topics such as emotional well-being, academic and work stress and relationships, with foundational mental health terms like "depression," "anxiety," "bipolar," "PTSD," "suicide," and "insomnia". The resulting vocabulary, composed of over 800 unique terms and expressions was lowercased and deduplicated to ensure consistent matching. We then filtered thread URLs based on whether their slugs (i.e., the final part of the URL, typically derived from the thread title) contained any of the predefined keywords. Multi-word phrases were matched as substrings, while single-word terms

[1] Dataset can be found at https://zenodo.org/records/3941387.

were tokenized and compared at the word level. 64 keywords were matched (see Appendix A8). Threads flagged as relevant were retained for scraping, allowing us to efficiently focus on discussions likely to involve mental health content. This approach resulted in 1,988 scraped posts.

Topic Modeling Analysis

We assigned each forum post (or *document*) to a *single* topic through automated topic extraction. While posts often contain multiple themes, there is typically one dominant theme that the user intends to discuss. Traditional topic modeling approaches include Latent Dirichlet Allocation (LDA) (Blei et al., 2003) and Non-negative Matrix Factorization (NMF) (Lee & Seung, 1999), but these methods represent topics through characteristic word lists using bag-of-words approaches that ignore semantic context (e.g., “depression” in economic vs. psychological context).

We used BERTopic, a modern topic modeling framework using *embedding-based* algorithms that convert text (after removing stopwords) to dense, fixed length vectors to capture contextual information. Our implementation used a pretrained BERT embedding model trained on Reddit data, aligning with our application’s context. BERTopic’s dimensionality reduction component, UMAP (McInnes et al., 2018), is stochastic, producing different topics and topic counts across runs. To address this variability, we developed *Consistent Clustering*, a novel algorithm that leverages this stochasticity to compute a distribution over topic assignments for each document across multiple iterations, identifying robust, reproducible topics (see Algorithm 1).

We averaged over the stochastic nature of BERTopic and UMAP by performing multiple independent clustering rounds. The consistency matrix from these 30 runs identifies semantically coherent topic representations that persist across random initializations. We then applied NMF to the resulting document-term matrix to uncover meta-clusters—higher-level groupings representing consistent semantic themes that emerge through lexically similar topic alignments across iterations. Finally, we used GPT-4o through KeyBERTInspired to generate interpretable topic labels for our 3-step pipeline.

The counted-based method can suffer from lexical dominance, where high-frequency but semantically broad words (e.g., “feel”, “stress”, and “sad”) appear across diverse topics and overshadow more discriminative signals. As a result, distinct subthemes may be merged into overly broad meta-topics. Rather than removing these emotionally generic words—which carry important affective signals in mental health contexts--we implemented a refinement step. We isolated posts assigned to broad meta-topics and reapplied BERTopic, using contextual embeddings and density-based clustering to reveal finer-grained subtopics. This selective deepening enhanced thematic granularity while preserving interpretability and consistency. The BERTopic will assign topic labels, where the label -1 typically denotes outliers that do not belong to any dense cluster in the embedding space. This is a consequence of HDBSCAN's density-based approach, which intentionally leaves sparse or isolated samples unassigned rather than forcing them into a potentially ill-fitting cluster. Here, we group these outlier posts under an "Others" category to help maintain thematic clarity and prevent the distortion of well-defined topics.

We also employed modern LLMs directly for topic modeling, which may be more expressive but less principled than BERTopic (Doi et al., 2024; Sankaranarayanan et al., 2025). GPT-4o (OpenAI et al., 2024) was instructed to generate semantically meaningful labels for posts using "few-shot" prompting with examples included (system prompt in Appendix X). LLMs enable both "open-ended" (*inductive*) and "restricted" (*deductive*) thematic extraction. Open-ended extraction allows any appropriate topic without limitations, enabling unexpected, explorative insights but risking inconsistent or overlapping themes (Naeem et al., 2023). Restricted extraction uses preselected topic lists, ensuring consistency and relevance but potentially missing novel themes. We deployed a restricted approach using themes from the Clinician Index of Clients Concerns (CLICC) (Pérez-Rojas et al., 2017), to maximize interpretability (topic list in Appendix X).

let's talk User Engagement Study

We investigated differences in user engagement across discovered mental health topics using metrics including likes, comments, total word count across discussions (including comments), and the number of reads of *let's talk* posts. These metrics are correlated, such as posts that receive more likes often also attract more reads and replies. It is also plausible that posts with more replies tend to have higher word counts, reflecting deeper discussions. However, different engagement behaviors may not always align. For instance, some posts may be widely read but provoke limited interaction, while others may receive substantial replies without garnering many likes. These possibilities underline the importance of using multiple metrics to capture distinct facets of user engagement.

Comparisons between Forums

To determine whether *let's talk* discussion patterns reflect broader demographic and cultural influences (e.g., dedicated vs. general-purpose platforms), we conducted comparative analyses with two contrasting communities. We compared with Reddit's mental health discussions (predominantly Western, English-speaking) to identify culturally-specific patterns, and with HardwareZone (Singapore's largest general forum, primarily adult users) to separate age-related from cultural factors within the same national context. This multi-forum comparison distinguished universal youth mental health concerns from those shaped by local cultural norms and expectations, validated whether observed patterns on *let's talk* reflect genuine demographic differences rather than platform biases, and it provided crucial context for public health practitioners and policymakers on whether interventions should be culturally tailored or draw from global practices based on universal patterns. Another key difference is that Reddit and HardwareZone are not moderated with regards to mental health discourse. We analyzed the differences in topic distributions (Park et al., 2018), using both BERTopic and LLM-extracted analyses.

Results (1000-1500)

Data Overview

Table 1 presents the characteristics of the three constructed corpora. The Reddit Mental Health dataset was approximately two orders of magnitude larger than *let's talk* and HardwareZone and was uniformly subsampled to 10,000 posts. Each subreddit varies in popularity, with *depression*, *suicidewatch*, and *anxiety* being more active than *addiction*, *alcoholism*, and *bipolarreddit*. This suggests differential prevalence or help-seeking patterns for these conditions within the online community.

TABLE 1: Overview of mental health forum corpora under study.

	let's talk	**Reddit Mental Health***	**HardwareZone**
Number of posts	1,921 (645 for "Hangouts", 1,276 for "Ask-a-Therapist")	10,000 (subsampled from 304,326)	1,988
Location of users	Singapore	Global	Singapore
Time period	2022-2025	2018-2020	2022-2025
Number of unique users	1,357	50,712	Unknown
Average words per post	186.98	187.48	95.86

**Includes posts from 16 subreddits: r\depression (59,242), r\suicidewatch (41,354), r\anxiety (35,872), r\mentalhealth (32,439), r\adhd (30,298), r\lonely (19,500), r\bpd (16,980), r\socialanxiety (16,015), r\EDAnonymous (11,989), r\autism (6,785), r\healthanxiety (6,681), r\ptsd (6,542), r\schizophrenia (6,144), r\addiction (5,882), r\alcoholism (4,515), r\bipolarreddit (4,088).*

Thematic Analysis of *let's talk* Forum Posts

Themes were extracted from *let's talk* forum posts using our custom BERTopic algorithm and through the CLICC-restricted LLM approach. The BERTopic-extracted topics are shown in Tab. 2, along with their prevalence and a list of representative words. These representative words were identified using class-based TF-IDF and reflect the most salient lexical features of each topic, though not all words necessarily appear in every post within the topic cluster. Characteristic posts for each topic can be found in Appendix A3.

TABLE 2: BERTopic-extracted topics discussed on *let's talk* (1,921 posts).

Topic	Post Count (Percentage)	Representation
Stress and Anxiety*	391 (20.35%)	'anxiety', 'just', 'feel', 'stress', 'stressed', 'feeling'
Friendship Issues	269 (14.00%)	'friend', 'friends', 'friendship', 'group', 'make', 'talk', 'best', 'social', 'relationship', 'betrayed', 'just', 'bestfriend', 'shes', 'trust', 'felt'
Academic Stress	242 (12.60%)	'study', 'exam', 'studying', 'motivation', 'grade', 'just', 'school', 'stress', 'stressed', 'quit', 'grading', 'table', 'exams', 'uni', 'join'
Mental Wellness Education	232 (12.08%)	'mental', 'health', 'resilience', 'talk', 'anxiety', 'wellness', 'depression', 'thinking', 'help', 'stress', 'normal', 'education', 'feeling', 'parents', 'illness'
Suicidal Thoughts	179 (9.32%)	'self', 'suicidal', 'thoughts', 'depression', 'suicide', 'harm', 'sad', 'don', 'depressed', 'negative', 'therapy', 'harming', 'just', 'feel', 'anxiety'
Family Dynamics	157 (8.17%)	'family', 'parent', 'mom', 'dad', 'mother', 'mum', 'sister', 'father', 'si', 'sibling', 'members', 'parents', 'want', 'abuse', 'jail'
Relationship Breakup*	130 (6.77%)	'breakup', 'relationship', 'bf', 'leave', 'feel'
Panic Attacks	88 (4.58%)	'panic', 'anxiety', 'pain', 'attacks', 'attack', 'symptom', 'having', 'palpitation', 'palpitations', 'nausea', 'anxious', 'chest', 'diagnosed', 'reflux', 'vomiting'
Social Interaction	79 (4.11%)	'ppl', 'person', 'affect', 'talk', 'say', 'counsellor', 'attack', 'psych', 'dis', 'people', 'feelings', 'stupid', 'abt', 'got', 'self'
Eating Disorders	33 (1.72%)	'weight', 'eating', 'eat', 'lost', 'binge', 'losing', 'lose', 'gaining', 'loss', 'skinny', 'stress', 'insecure', 'gain', 'diet', 'struggling'
Job Anxiety	23 (1.20%)	'anxious', 'nervous', 'anxiety', 'stressful', 'going', 'interview', 'phobia', 'feeling', 'job', 'quit', 'getting', 'start', 'matter', 'constantly', 'talk'
Sleep Problems	21 (1.11%)	'sleep', 'tired', 'hours', 'time', 'feeling', 'exhausted', 'adhd', 'sleeping', 'normal', 'hrs', 'waking', 'feel', 'sufficient', 'rest', 'trouble'
Therapy and Counseling	20 (1.04%)	'therapist', 'mental', 'therapy', 'counsellor', 'psychologist', 'mentor', 'healthcare', 'help', 'counselling', 'health', 'singapore', 'counselor', 'mistreatment', 'hospital', 'talking'
Anger Management	20 (1.04%)	'anger', 'angry', 'family', 'takes', 'feel', 'issues', 'emotions', 'moment', 'like', 'control', 'irritable', 'outbursts', 'im', 'overcome', 'quarrel'
Marital Issues	17 (0.88%)	'husband', 'divorce', 'wife', 'marriage', 'connection', 'pregnant', 'spouse', 'anger', 'management', 'birth', 'giving', 'emotional', 'emotions', 'married', 'care'
Others*	11 (0.57%)	'losing investments', 'unappreciated marriage', 'struggle cope', 'lied saving', 'live past'
Loneliness	9 (0.47%)	'lonely', 'feel', 'loneliness', 'talk', 'just', 'make', 'really', 'friends', 've', 'life', 'emptiness', 'single', 'want', 'friendship', 'person'

** Derived from the further decomposition of the "Emotional Distress" topic.*

The topic "Emotional Distress" (not shown in table) emerged as the most prevalent category (532 posts, 27.69% of posts). To increase granularity and uncover latent substructures within this general and dominant category, we isolated all "Emotional Distress" documents and applied BERTopic to this subset, which decomposed the category into three subtopics: "Stress and Anxiety", "Relationship Breakup" and "Others". The topic "Stress and Anxiety" is the most frequently discussed on *let's talk*. Other commonly discussed topics include "Academic Stress" and "Friendship Issues", reflecting the educational and social pressures faced by a younger demographic. The topics "Suicidal Thoughts" and "Panic Attacks" indicate that users also express more acute and severe concerns.

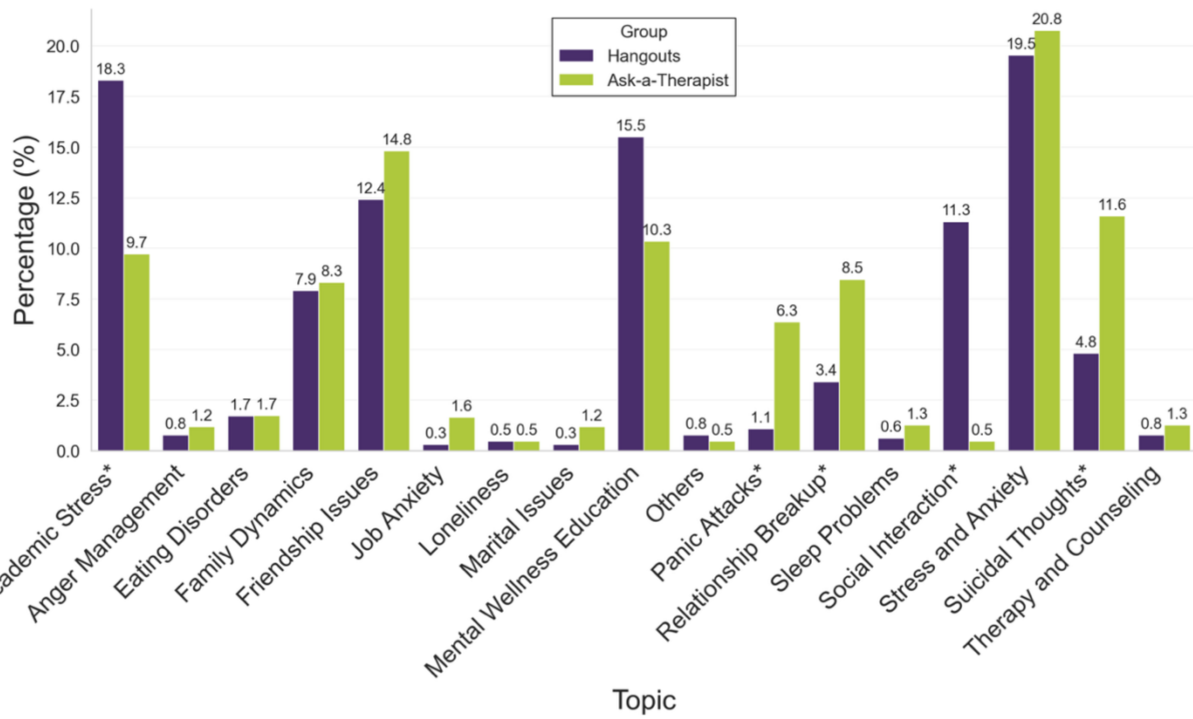


Figure 1: Topic distribution in the "Hangouts" and "Ask-a-Therapist" spaces of the let's talk forum, which respectively have 645 and 1,276 posts (see Tab. 1 and Tab. A11 for statistical test results). Topics marked with an asterisk (*) indicate statistically significant differences in distribution between the two spaces ($p < 0.05$), based on a chi-square test.

The differences in topic distribution between the *Hangouts* and *Ask-a-Therapist* spaces, as assessed using chi-square tests of independence, show distinct patterns (see Fig. 1). For each topic, a separate chi-square test was conducted to examine whether its distribution differed significantly between the two spaces, with the null hypothesis stating that the proportion of posts on the topic is the same in both. There are several significant differences between these spaces. In *Hangouts*, users engage relatively more in peer-oriented interactions such as "Social Interaction" and "Academic Stress", reflecting informal emotional sharing and community support. In contrast, *Ask-a-Therapist* features significantly higher proportions of posts on "Panic Attacks," "Suicidal Thoughts," and "Relationship Breakup," topics that often signal acute distress and the need for professional guidance.

User Engagement Analysis of let's talk Forum Posts

Platform engagement for each topic was measured according to: 1) number of likes; 2) number of reads; 3) number of replies; and 4) total word count of the first post and all replies. Fig. 2 illustrates user engagement as measured by the distribution of total word count across the original post and comments across different topics. Average number of words is sensitive to outliers, so we report the median as the "typical" word length. "Relationship Breakup" and "Family Dynamics" have the highest median word counts, suggesting users are more likely to engage in extended conversations when discussing interpersonal issues rather than more individualistic issues such as "Sleep Problems" and "Loneliness".

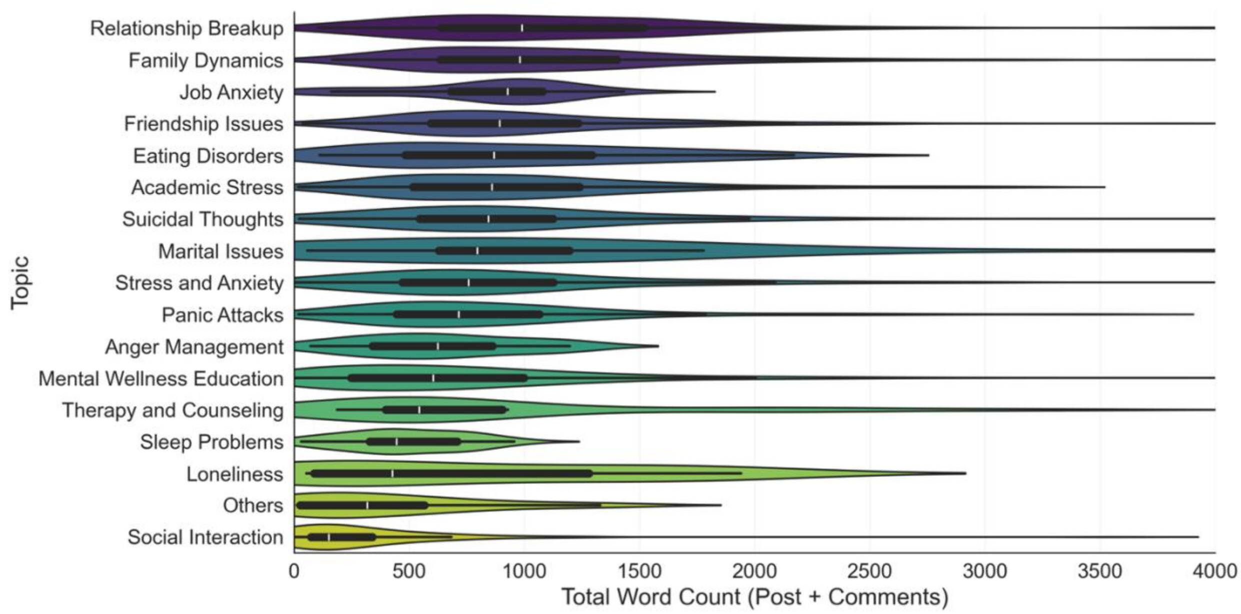


Figure 2: Engagement as measured by number of words across original posts and comments by topic. Topics are ranked by median number of words per thread.

The ridge plot in Fig. 3 reflects how likely a post is to generate interaction and community response, as measured by number of replies. Topics such as "Job Anxiety" and "Loneliness" receive the highest median number of replies, indicating that these issues may resonate more strongly with the community or elicit greater empathy and support from peers, although the word count of "Loneliness" is generally low, indicating shorter replies. Long-tailed distributions, seen in topics like "Loneliness", suggest certain posts attract high attention and engagement whereas others are mostly ignored, indicating that specific experiences *within* this topic may trigger particularly strong community responses.

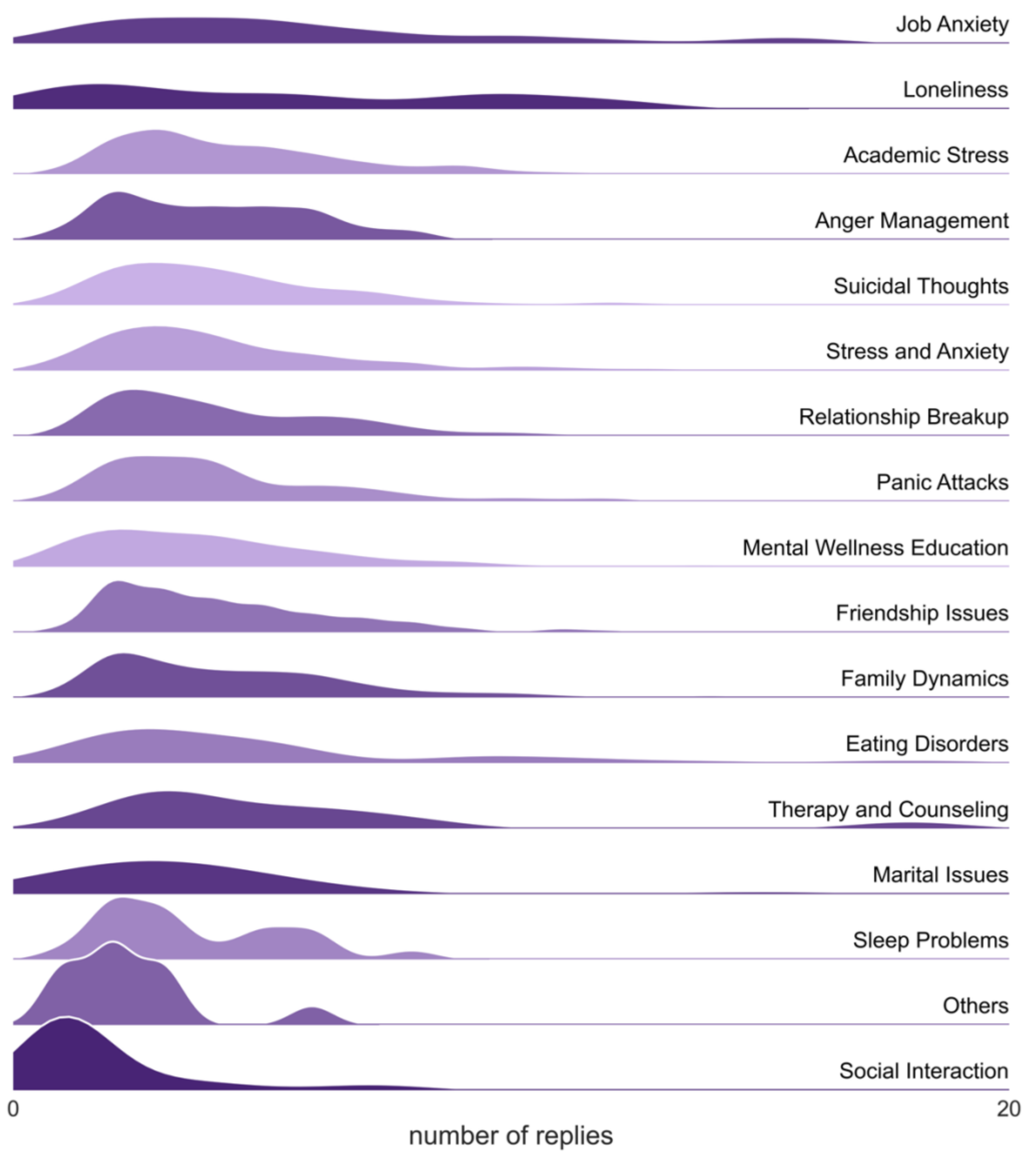


Figure 3: Distribution of number of replies by topic, ranked by median number of replied posts.

The engagement metrics of number of likes and number of reads are highly correlated with the number of replies (number of likes & number of reads, $\rho = 0.73$, number of likes & number of replies, $\rho = 0.66$, number of reads & number of replies, $\rho = 0.83$, respectively). Their results can be found in the Appendix A5. “Therapy and Counseling” and “Job Anxiety” receive both high likes and reads, suggesting more user attractions and identifications. A cluster of topics including “Suicidal Thoughts”, “Friendship Issues” and “Stress and Anxiety” share remarkably similar like patterns, highlighting shared norms in responding to common psychological challenges. “Social Interaction” consistently exhibits the lowest levels of user engagement, both in terms of likes and reads. This may imply that, compared to emotionally charged or support-seeking topics, discussions around general social interaction may fail to capture user attention.

LLM-extracted topics

The LLM-extracted CLICC topics from *let's talk* are shown in Fig. 4a. Frequently discussed themes encompassed mental health conditions ("anxiety", "stress", "depression", "suicidality"), interpersonal dynamics ("relationship problem", "family", "interpersonal functioning", "social isolation", "social anxiety"), identity-related concerns particularly relevant for youths ("self-esteem/confidence", "eating/body image"), and performance-related stressors ("academic performance", "career"). The results are broadly similar to the BERTopic results. However, friendship is not an isolated topic, a theme similar to "Mental Wellness Education" is not present here, and neither are "Sleep Problems" and "Therapy and Counseling". Suicidality and depression are considered two separate topics here, whereas they are merged into one topic with BERTopic.

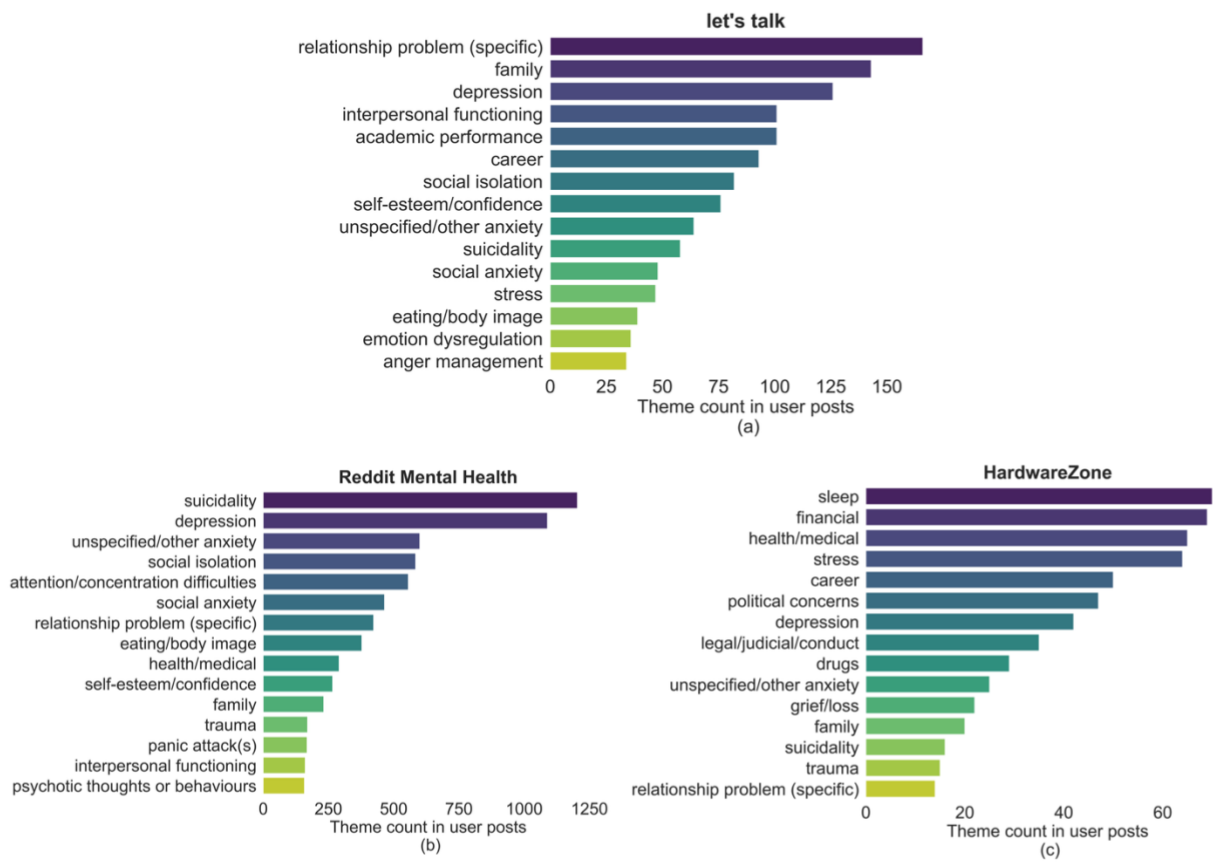


Figure 4: Top 15 LLM-extracted top CLICC concern themes from different forums posts and their prevalence: a) let's talk, b) Reddit Mental Health, c) HardwareZone.

Comparison with Other Forums

We compared the distribution of topics discussed on *let's talk* with mental health subreddits on Reddit and mental health discussions on HardwareZone.

The BERTopic thematic analysis results of Reddit are shown in Tab. 3. We found many similar results as compared to *let's talk*: depression, anxiety, suicidal ideation, and friendships are frequently discussed topics. However, loneliness features more prominently on Reddit, and family and romantic relationships are less frequently discussed. This could be due to the user base being more individualistic, older, or less likely to live with family. Reddit users also often discuss alcohol addiction, medication, drugs, and clinical diagnoses ("anxiety disorders", "ADHD", "Bipolar Disorder", "PTSD", "Borderline Personality Disorder"). They discuss academic stressors less than *let's talk* users. Interestingly, BERTopic considered "Birthday Celebration" worthy of a separate topic, indicating that many users talk about mental health in relation to birthdays. The LLM-extracted themes (Fig. 4b) show that the top concerns are similar to the BERTopic themes: depression, anxiety, social isolation, and suicidality. One key difference is that family and relationships do show up as topics in the LLM analysis, while substance usage does not. BERTopic-extracted topics from the individual 16 Reddit subreddits are shown in Fig. A4.

TABLE 3: Top 15 topics of concern from BERTopic analysis of Reddit (10,000 posts).

Topics	Post Count (Percentage)	Representation
Depression	2,055 (20.55%)	'depression', 'depressed', 'feel', 'mental', 'life', 'job', 'self', 'help', 'motivation', 'hate', 'health', 'just', 'makes', 'people', 'think'
Anxiety Disorders	1,638 (16.38%)	'anxiety', 'panic', 'anxious', 'attack', 'attacks', 'social', 'job', 'health', 'having', 'sleep', 'symptoms', 'feel', 'stress', 'phobia', 'ocd'
Emotional Loneliness	1,203 (12.03%)	'lonely', 'feel', 'feeling', 'depressed', 'just', 'self', 'life', 'like', 'talk', 'love', 'emptiness', 'sad', 'don', 'want', 'hug'
Suicidal Ideation	857 (8.57%)	'die', 'life', 'kill', 'want', 'just', 'hate', 'death', 'live', 'hug', 'living', 'wanna', 'way', 'alive', 'dying', 'painless'
Alcohol Addiction	742 (7.42%)	'alcoholic', 'drinking', 'sober', 'alcohol', 'problem', 'father', 'stop', 'functioning', 'bf', 'addiction', 'quit', 'don', 'day', 'withdrawal', 'years'
ADHD Medication	687 (6.87%)	'adhd', 'adderall', 'diagnosed', 'help', 'just', 'adult', 'meds', 'ritalin', 'think', 'taking', 'medication', 'having', 'know', 'reading', 'prescribed'
Suicide Prevention	619 (6.19%)	'suicide', 'suicidal', 'thoughts', 'commit', 'think', 'thinking', 'self', 'option', 'having', 'committed', 'people', 'help', 'dying', 'family', 'prevention'
Friendship Building	563 (5.63%)	'friends', 'talk', 'social', 'make', 'friendship', 'group', 'making', 'lonely', 'new', 'friend', 'want', 'close', 'looking', 'need', 'help'
Bipolar Disorder	469 (4.69%)	'bipolar', 'diagnosed', 'disorder', 'just', 'meds', 'ii', 'hard', 'years', 'think', 'having', 'type', 'taking', 'mental', 'schizophrenia', 'help'

Eating Disorders	444 (4.44%)	'weight', 'eating', 'disorder', 'loss', 'lose', 'anorexia', 'losing', 'gained', 'diet', 'binge', 'overweight', 'gaining', 'habits', 'anorexic', 'lost'
PTSD and Trauma	418 (4.18%)	'ptsd', 'therapist', 'therapy', 'diagnosed', 'having', 'does', 'symptoms', 'defines', 'trauma', 'say', 'doing', 'mental', 'suffers', 'abusive', 'abused'
Borderline Personality Disorder	161 (1.61%)	'bpd', 'symptoms', 'diagnosed', 'diagnosis', 'relationship', 'know', 'feel', 'disorder', 'having', 'personality', 'mental', 'told', 'abandonment', 'therapy', 'self'
Stimulant Medication	88 (0.88%)	'vyvanse', '50mg', 'taking', 'adderall', 'prescribed', 'effects', 'anxiety', '100mg', '30mg', '40mg', 'brain', '60mg', 'medication', 'started', 'stimulants'
Loneliness	60 (0.66%)	'loneliness', 'feeling', 'lonely', 'advice', 'deal', 'rate', 'true', 'depression', 'finding', 'feels', 'solution', 'unhealthy', 'today', 'emotional', 'intense'
Birthday Celebration	56 (0.56%)	'birthday', 'tomorrow', 'today', 'celebrate', 'fun', 'yesterday', 'happy', 'just', 'party', 'birthdays', '21st', 'depressed', '61', 'life', 'holiday'

Tab. 4 shows the most frequently discussed topics across 1,988 HardwareZone posts. The dominant theme, “Workplace Stress” (28.52%), highlights widespread concerns about job pressures and burnout. This is followed by “Singaporean Youth Issues” (13.48%), reflecting mental health and socioeconomic anxieties among younger users. HardwareZone features a wider range of discourse, including controversial or toxic topics like Drug Trafficking (5.43%) and Singapore Military (1.91%) that are largely absent from *let’s talk*. Although clinically oriented topics such as “Cancer Treatment” and “Suicide and Death” suggest that the forum may at times serve as a space for expressing acute psychological distress, the overall topic distribution remains closer to the Hangouts section of *let’s talk*, focusing on informal, peer-oriented discussions of everyday struggles rather than therapeutic intervention or clinical diagnosis. Compared to *let’s talk,* academic stressors and family are rarely discussed, whereas drugs are more frequently discussed.

Notably, “Singlish Expressions” (9.51%) reveals the forum’s cultural specificity. Frequent use of Singlish (a blend of English and Chinese/Malay) terms like "sian," "liao," "guat," and "kwang" fosters local identity but complicates topic modeling, given that embeddings are not trained on such linguistic variants. In contrast, *let’s talk* users tend to use more formal and standardized English, allowing for cleaner topic delineation.

For HardwareZone, LLM-extracted themes (Fig. 4c) like depression, stress, anxiety, sleep, and work-related issues feature prominently again, but others like health/medical, financial themes, and political concerns (e.g., cost of living) are also frequently discussed. Unlike with the BERTopic analysis, topics related to Singapore-specific cultural experiences such as military training and Singlish are not extracted, because such themes are not part of the restricted themes list that the LLM uses.

TABLE 4: BERTopic-extracted topics discussed on HardwareZone (1,988 posts).

Topic	Post Count (Percentage)	Representation
Workplace Stress	567 (28.52%)	'stress', 'work', 'burnout', 'depressed', 'feelings', 'stressed', 'employees', 'significant', 'employee', 'anxiety', 'psle', 'weekend', 'day', 'emotional', 'constant'
Singaporean Youth Issues	268 (13.48%)	'singapore', 'youth', 'mental', 'singaporean', 'loan', 'disorders', 'samaritans', 'reported', 'food', 'people', 'health', 'financial', 'fiscal', 'singaporeans', 'debt'
Psychological Disorders	238 (11.97%)	'mental', 'depression', 'health', 'depressed', 'pile', 'anxiety', 'illness', 'state', 'people', 'disorders', 'kindness', 'patients', 'youth', 'wellbeing', 'suicidal'
Sleep Issues	190 (9.56%)	'sleep', 'hours', 'wake', 'like', 'early', 'quality', 'average', 'time', 'feel', 'long', 'work', 'insomnia', 'apnea', 'sleeping', 'asleep'
Singlish Expressions	189 (9.51%)	'liao', 'sian', 'kwang', 'tio', 'think', 'messi', 'loh', 'guat', 'tried', 'need', 'siao', 'lang', 'waiting', 'hui', 'sia'
Drug Trafficking	108 (5.43%)	'drug', 'ecstasy', 'drugs', 'police', 'suspected', 'arrested', 'tablets', 'methamphetamine', 'abuse', 'driver', 'ketamine', 'car', 'vehicle', 'trafficking', 'seized'
Suicide and Death	76 (3.82%)	'suicide', 'funeral', 'committed', 'cher', 'director', 'assisted', 'deaths', 'death', 'coroner', 'autopsy', 'boyfriend', 'prevention', 'euthanasia', 'dying', 'abetting'
Cancer Treatment	57 (2.87%)	'therapy', 'cancer', 'treatment', 'patients', 'proton', 'centre', 'cost', 'treatments', 'products', 'cell', 'gene', 'chemotherapy', 'patient', 'lymphoma', 'diagnosed'
Fatigue	52 (2.62%)	'tired', 'fatigue', 'feel', 'feeling', 'monday', 'muscle', 'sunday', 'day', 'sleepy', 'caffeine', 'rest', 'lifting', 'heavy', 'ps', 'like'
Alcohol Consumption	41 (2.06%)	'alcohol', 'drink', 'drank', 'bottle', 'substance', 'thailand', 'gulp', 'blood', 'drinking', 'poisoning', 'test', 'stating', 'coke', 'amounts', 'challenged'
Singapore Military	38 (1.91%)	'sg', 'ranger', 'like', 'saf', 'outlets', 'jp', 'trainee', 'needs', 'girls', 'tkl', 'km', 'ladies', 'far', 'trainees', 'men'
Heat Stress	28 (1.41%)	'heat', 'temperature', 'stress', 'risk', 'heater', 'hot', 'new', 'note', 'humid',

		'prevailing', 'temp', 'weather', 'levels', 'deal', 'electricity'
Panic Disorder	26 (1.31%）	'panic', 'anxiety', 'related', 'disorder', 'attacks', 'attack', 'mental', 'diagnosis', 'health', 'illness', 'stress', 'compulsive', 'anxious', 'financial', 'help'
Infectious Diseases	24 (1.21%)	'covid', 'illness', 'vaccine', '19', 'influenza', 'infectious', 'vaccines', 'diseases', 'respiratory', 'outbreak', 'vaccination', 'severe', 'fatigue', 'cause', 'china'
Narcissistic Traits	22 (1.11%)	'personality', 'narcissistic', 'traits', 'values', 'certain', 'individuals', 'person', 'narcissism', 'empathy', 'trait', 'humility', 'expressive', 'disorders', 'disorder', 'characteristics'
Marriage and Dating	18 (0.91%)	'single', 'marry', 'married', 'like', 'ppl', 'know', 'pretty', 'tinder', 'marriage', 'suddenly', 'hope', 'girls', 'judging', 'lifestyle', 'marrying'
Online Forum Discussion	17 (0.86%)	'edmw', 'mention', 'oni', 'opposite', 'preference', 'help', 'edmwer', 'support', 'edmwers', 'alot', 'devote', 'lol', 'solutions', 'tmr', 'forumers'
Forum Rules	15 (0.75%)	'rules', 'forum', 'hwz', 'netflix', 'read', 'rule', 'watching', 'main', 'dose', 'upz', 'suddenly', 'posting', 'cny', 'raja', 'chsjason'
Weekend Relaxation	14 (0.70%)	'mood', 'relaxed', 'things', 'work', 'friday', 'taking', 'able', 'excited', 'weekend', 'successfully', 'free', 'feel', 'got', 'day', 'depress'

Discussion (1000-1500)

Summary of results

Thematic analysis of *let's talk* revealed meaningful patterns in Singaporean youth mental health discourse. Our BERTopic approach identified "Stress and Anxiety", "Friendship Issues", and "Academic Stress" as dominant themes, accounting for nearly half of all posts. The consistent clustering method preserved semantic similarity across related topics while disentangling overlapping psychological dimensions, yielding interpretable and robust topic groupings.

Notable differences in topic distribution emerged between forum spaces: more severe and urgent topics (e.g., “Suicidal Thoughts” and “Panic Attacks”) concentrated in the therapist-moderated *Ask-a-Therapist* space, while everyday emotional concerns (e.g., “Academic Stress” and “Social Interaction”) dominated the peer-led *Hangouts* space. This suggests that professional moderation reduces disclosure barriers for crisis-level issues, by offering a sense of safety and legitimacy that encourages users, while peer spaces facilitate routine stress management and emotional expression.

User engagement patterns revealed that interpersonal themes--particularly romantic and family relationships--generated the lengthiest engagements, while “Job Anxiety”, “Loneliness”, and “Academic Stress” received the most comments.

Themes extracted using LLMs largely corroborated these BERTopic findings, suggesting potential to augment traditional methods. However, restricted LLMs are limited to predefined themes of interest, and less effective for discovering unexpected themes. There are notable differences between the two approaches, and without manual validation it remains unclear which approach is most suitable.

Cross-forum comparisons revealed distinct cultural, demographic, and moderation patterns. *let's talk* discussions emphasized academic pressure (12.6% of all posts) more frequently compared to Reddit and HardwareZone, consistent with its youth focus. Reddit hosted higher rates of medication and substance use (e.g., 7.42% of posts are about alcohol addiction) discussions, potentially reflecting Singapore’s lower substance abuse prevalence and stringent drug policies. Clinical terms and psychiatric diagnoses were also more commonly mentioned on Reddit. HardwareZone users more frequently discussed workplace stress, in line with its older demographic. These findings highlight the critical importance of demographic-specific analyses in mental health research. Many of these contrasts are also explained by the fact that *let’s talk* is moderated as a safe environment.

Our methodology provides a scalable framework for obtaining timely, naturalistic insights into community mental health needs, complementing traditional epidemiological approaches. These insights can directly inform targeted public health interventions, such as Singapore’s Beyond the Label stigma-reducing campaign (Subramaniam et al., 2016).

Comparison to Other Studies

We posited this study as complementary to traditional population mental health research. While not directly comparable, our findings can be contextualized against local studies including the Youth Epidemiology and Resilience (YEAR) study by the NUS Mind Science Centre (Wong et al., 2024) and the Singapore National Youth Mental Health Study by IMH (Chang et al., 2025; Samari et al., 2025; Subramaniam et al., 2025).

Furthermore, while similar international datasets are not available, we may also compare our results to comparable studies.

Comparison with the YEAR study revealed both convergent and divergent patterns. YEAR's findings that perceived stress peaks around age 16 and strongly correlates with anxiety and depression, particularly regarding academic expectations, aligned with *let's talk* data where "Academic Stress" and "Stress and Anxiety" formed a cluster among the most discussed themes. Notable differences emerged due to demographic scope. While YEAR examined 10-18-year-olds, *let's talk* includes older contributions, potentially explaining discussions of career decisions, romantic and family relationships, beyond typical adolescent experiences.

Notably, despite YEAR findings on social media use and its link to internalizing and externalizing (aggression, rule-breaking) behaviors, and a growing body of evidence on the associations between smartphone use and youth mental health (Abi-Jaoude et al., 2020), technology-related concerns were infrequently discussed on *let's talk*. This suggests that public discourse may not always reflect prevailing research concerns, echoing arguments that moral panic over technology's harms may be overstated (Orben, 2020). Similarly, ADHC, identified in YEAR as the third most prevalence condition, was rarely mentioned on *let's talk*—possibly due to stigma, lack of interest, low self-recognition, or its typical identification by others. This reinforces the point that digital mental health discourse complements, but does not substitute, clinical prevalence data.

The National Youth Mental Health Study (NYMHS) of 15-35-year-olds also showed convergence with *let's talk*. Body image concerns were common in both datasets, with "Eating Disorders" frequently discussed and associated with internalizing symptoms like anxiety and depression (Rodgers et al., 2023). However, bullying (including cyberbullying) and addiction (including alcohol and drugs) were identified in NYMHS as strong predictors of anxiety symptoms, but these topics were rarely discussed on *let's talk*, possibly due to anticipatory discrimination or self-censorship. NYMHS focused on stigma through perceptions of others, often highlighting fear and social distance. In contrast, *let's talk*, as an anonymous forum, captured more candid self-disclosure, highlight the orle of anonymity in reducing internalized stigma.

Comparison with a large-sample Wysa chatbot study in Singapore (Sinha et al., 2024), revealed substantial thematic overlap despite differing interfaces. Both platforms featured frequent discussions of relationship rejection, interpersonal conflict, academic pressure, and future-related anxiety. Wysa users' expressions of responsibility and self-criticism aligned with self-esteem themes identified through our CLICC analysis.

Both Kooth, a comparable forum in the UK, and *let's talk* identified emotional difficulties, sadness, and family-related problems as key youth concerns. Kooth reported a notable rise in self-harm and suicidal ideation during the early phase of the COVID-19 pandemic, particularly among male and younger users, but do to the young age of *let's talk* such temporal analyses could not (yet) be confirmed (*Kooth-Pulse-2021-Report*). Findings from TalkLife (Williams et al., 2024) revealed that young users frequently disclosed abuse experiences through emotionally indirect language, highlighting a need for NLP methods in this domain to be sufficiently sensitive and aware of nuances.

Finally, the Center for Collegiate Mental Health (CCMH) study of 61,473 clients during the 2023–2024 academic year[2] revealed "stress", "depression", and "anxiety" as predominant CLICC concerns, consistent with *let's talk*. However, "Relationship problem (specific)" emerged as the most frequently extracted theme in *let's talk* but ranked only sixth (7.7%) in CCMH data, suggesting possible differences in how interpersonal distress manifests across clinical versus peer-support contexts.

Limitations

Our study considered *let's talk* as a bellwether for Singaporean youth mental health discourse, but this sample may be biased towards English-speaking, digitally engaged individuals who are willing to share personal experience online. Generalization to the broader (youth) population should be approached with caution. Due to the anonymity of the platform, detailed demographic information about users is lacking. For this reason, we consider the current analysis to be *complementary* to traditional epidemiological studies or in-depth qualitative research such as interviews or surveys. Furthermore, our analysis does not include posts that were removed by moderators due to risks or sensitive content, which may skew our findings.

Several methodological limitations warrant consideration. Our single-topic approach may oversimplify posts that span multiple thematic categories; future work should explore multi-topic modeling approaches. Additionally, extracted topics and labels extracted lack clinical validation. The theme "depression" does not imply any clinical diagnosis. While our analysis identified commonly discussed topics that can inform public health initiatives, less frequently discussed topics such as rare mental health conditions may be overshadowed. The "long tail" of topics merits separate investigation. Importantly, the frequency of topics discussed does not indicate severity or treatment gaps, requiring caution in policy applications of these findings.

While the consistent clustering framework enhances topic stability and robustness through integration of BERTopic, NMF and GPT-based summarization, it has several limitations. BERTopic relies on standard English embeddings and performs poorly with informal or culturally specific language such as Singlish, leading to vague or misgrouped clusters when dealing with Singapore forums.

[2] https://ccmh.psu.edu/assets/docs/CCMH%202024%20Annual%20Report.pdf

The use of the CLICC also has several limitations, such as being therapist-centric, Western-centric, adult-focused, and based on clinical settings rather than forum discussions. A framework like HEADSS may be more appropriate, as it is focused on youths and aligns better with our population and context.

In our user engagement study, we did not control for confounding variables such as post length, time since posting (some posts had less time to receive comments or likes), or day of week.

The comparison between *let's talk* and the other forums has several limitations. In terms of the comparison between forums, the collection period of the posts differs between the data sources. This may reduce the validity of comparison. For example, the data from Reddit overlaps with the COVID-19 period, which may have brought specific mental health issues with it (Low et al., 2020). Furthermore, the platform structures and user incentives are not identical. Data from Reddit and HardwareZone was not overseen by mental health professionals. Their posts may contain trolling or sarcasm, which may bias our results. Running a data clean step before our analysis could reduce this effect. Moreover, the users and mental health conversations can be fundamentally different depending on the data we seek from Reddit and HardwareZone (i.e., the posts we choose as being mental health related and our search and inclusion/exclusion criteria), which may compromise our ability to make meaningful comparisons. Beyond forums, future work should also compare to other data sources, such as social media platforms like Twitter, Q&A, or conversational datasets. For example, DatD (Owen et al., 2020) is a manually annotated corpus of 5,550 tweets, labeled for indicators of depression and anxiety. Several benchmark datasets have also emerged to NLP in mental health, including SNAP (Althoff et al., 2016), Psych8K (Liu et al., 2023), and HOPE (Malhotra et al., 2022). In addition, the recent MentalChat16K dataset (Xu et al., 2025) provides 16,000 high-quality mental health question–answer pairs, combining real anonymized intervention transcripts and synthetic counseling conversations.

Ethical Considerations

This study meets established criteria for IRB exemption under standard research ethics guidelines. All data analyzed consisted of publicly available, anonymous, and nonclinical forum posts with no personal identifiers collected or stored. Users of *let's talk* explicitly consent to data usage for "research and service enhancement" through the platform's terms of use and data protection policy. The research involved no direct participant contact, intervention, or compensation, and employed only observational analysis of pre-existing public discourse.

The study design incorporated several ethical safeguards. All analyses were conducted at aggregate levels without individual user identification or tracking. No attempt was made to re-identify users or link posts to individuals. The research focused on understanding population-level mental health discourse patterns rather than individual behaviors or outcomes. Data handling procedures ensured complete anonymization throughout the analysis pipeline.

We acknowledge ongoing ethical discussions in digital mental health research regarding content exposure effects and user privacy expectations. However, this study's approach—analyzing anonymous, consented, public data at population scale to inform public health understanding—aligns with established ethical frameworks for observational research using publicly available online content. The potential public health benefits of understanding youth mental health discourse patterns, particularly in underrepresented populations, support the ethical justification for this research approach.

Acknowledgements

Funding to create *let's talk* and to support the writing of this article was provided by the MOH Office for Healthcare Transformation (Singapore).

Conflicts of Interest

All authors were employees of the MOH Office for Healthcare Transformation (Singapore), which created the mindline.sg and *let's talk* platforms, at the time of writing.

Abbreviations

AI: artificial intelligence

CLICC: Clinician Index of Client Concerns

LLM: large language model

NLP: natural language processing

NMF: non-negative matrix factorization

References

Abi-Jaoude, E., Naylor, K. T., & Pignatiello, A. (2020). Smartphones, social media use and youth mental health. *CMAJ*, *192*(6), E136–E141. https://doi.org/10.1503/cmaj.190434

Åkerstedt, T., Knutsson, A., Westerholm, P., Theorell, T., Alfredsson, L., & Kecklund, G. (2002). Sleep disturbances, work stress and work hours: A cross-sectional study. *Journal of Psychosomatic Research*, *53*(3), 741–748. https://doi.org/10.1016/S0022-3999(02)00333-1

Alghamdi, Z., Kumarage, T., Agrawal, G., Karami, M., Almuteb, I., & Liu, H. (2025). *RedditESS: A Mental Health Social Support Interaction Dataset -- Understanding Effective Social Support to Refine AI-Driven Support Tools* (No. arXiv:2503.21888). arXiv. https://doi.org/10.48550/arXiv.2503.21888

Ali, K., Farrer, L., Gulliver, A., & Griffiths, K. M. (2015). Online Peer-to-Peer Support for Young People With Mental Health Problems: A Systematic Review. *JMIR Mental Health*, *2*(2), e19. https://doi.org/10.2196/mental.4418

Almojali, A. I., Almalki, S. A., Alothman, A. S., Masuadi, E. M., & Alaqeel, M. K. (2017). The prevalence and association of stress with sleep quality among medical students. *Journal of Epidemiology and Global Health*, *7*(3), 169–174. https://doi.org/10.1016/j.jegh.2017.04.005

Althoff, T., Clark, K., & Leskovec, J. (2016). Large-scale Analysis of Counseling Conversations: An Application of Natural Language Processing to Mental Health. *Transactions of the Association for Computational Linguistics*, *4*, 463–476. https://doi.org/10.1162/tacl_a_00111

Blei, D. M., Ng, A., & Jordan, M. (2003). *Latent Dirichlet Allocation*.

Bond, R. R., Mulvenna, M. D., Potts, C., O'Neill, S., Ennis, E., & Torous, J. (2023). Digital transformation of mental health services. *Npj Mental Health Research*, *2*(1), 13. https://doi.org/10.1038/s44184-023-00033-y

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

Boyatzis, R. E. (1998). *Transforming Qualitative Information: Thematic Analysis and Code Development*. SAGE.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cambria, E., Gulyás, B., Pang, J. S., Marsh, N. V., & Subramaniam, M. (2024). Explainable AI for Stress and Depression Detection in the Cyberspace and Beyond. *PAKDD 2024 Workshops*.

Chang, S., Vaingankar, J. A., Tan, B., Tan, Y. W. B., Samari, E., Archana, S., Chua, Y. C., Lee, Y. P., Tang, C., Verma, S., & Subramaniam, M. (2025). Prevalence and correlates of nonsuicidal self-injury among youths in Singapore: Findings from the National Youth Mental Health Study. *Child and Adolescent Psychiatry and Mental Health*, *19*(1), 27. https://doi.org/10.1186/s13034-025-00885-6

Chim, J., Tsakalidis, A., Gkoumas, D., Atzil-Slonim, D., Ophir, Y., Zirikly, A., Resnik, P., & Liakata, M. (2024). Overview of the CLPsych 2024 Shared Task: Leveraging Large Language Models to Identify Evidence of Suicidality Risk in Online Posts. In A. Yates, B. Desmet, E. Prud'hommeaux, A. Zirikly, S. Bedrick, S. MacAvaney, K. Bar, M. Ireland, & Y. Ophir (Eds.), *Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024)* (pp. 177–190). Association for Computational Linguistics. https://aclanthology.org/2024.clpsych-1.15

Chong, S. A., Abdin, E., Sherbourne, C., Vaingankar, J., Heng, D., Yap, M., & Subramaniam, M. (2012). Treatment gap in common mental disorders: The Singapore perspective. *Epidemiology and Psychiatric Sciences*, *21*(2), 195–202. https://doi.org/10.1017/S2045796011000771

Chong, S. A., Verma, S., Vaingankar, J. A., Chan, Y. H., Wong, L. Y., & Heng, B. H. (2007). Perception of the public towards the mentally ill in developed Asian country. *Social Psychiatry and Psychiatric Epidemiology*, *42*(9), 734–739. https://doi.org/10.1007/s00127-007-0213-0

Choudhury, M. D., & De, S. (2014). Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 71–80. https://doi.org/10.1609/icwsm.v8i1.14526

Cuijpers, P., Beekman, A. T. F., & Reynolds, C. F. (2012). Preventing Depression: A Global Priority. *Jama*, *307*(10), 1033–1034. https://doi.org/10.1001/jama.2012.271

Dan, E., Zhu, J., & Jin, R. (2025). Exploring Suicide Factors in Online Discourse: Sentiment and Thematic Analysis of Reddit. *ACM Transactions on the Web*, 3716546. https://doi.org/10.1145/3716546

De Paoli, S. (2024). Performing an Inductive Thematic Analysis of Semi-Structured Interviews With a Large Language Model: An Exploration and Provocation on the Limits of the Approach. *Social Science Computer Review*, *42*(4), 997–1019. https://doi.org/10.1177/08944393231220483

Doi, T., Isonuma, M., & Yanaka, H. (2024). *Comprehensive Evaluation of Large Language Models for Topic Modeling* (No. arXiv:2406.00697). arXiv. https://doi.org/10.48550/arXiv.2406.00697

Gowen, K., Deschaine, M., Gruttadara, D., & Markey, D. (2012). Young adults with mental health conditions and social networking websites: Seeking tools to build community. *Psychiatric Rehabilitation Journal*, *35*(3), 245–250. https://doi.org/10.2975/35.3.2012.245.250

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure* (No. arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794

Gulliver, A., Griffiths, K. M., & Christensen, H. (2010). Perceived barriers and facilitators to mental health help-seeking in young people: A systematic review. *BMC Psychiatry*, *10*, 113. https://doi.org/10.1186/1471-244X-10-113

Heaukulani, C., Phang, Y. S., Weng, J. H., Lee, J., & Morris, R. J. T. (2024). Deploying AI Methods for Mental Health in Singapore: From Mental Wellness to Serious Mental Health Conditions. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4935469

Ji, S., Zhang, T., Ansari, L., Fu, J., Tiwari, P., & Cambria, E. (2022). MentalBERT: Publicly Available Pretrained Language Models for Mental Healthcare. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 7184–7190). European Language Resources Association. https://aclanthology.org/2022.lrec-1.778

Jones, R., Sharkey, S., Ford, T., Emmens, T., Hewis, E., Smithson, J., Sheaves, B., & Owens, C. (2011). Online discussion forums for young people who self-harm: User views. *The Psychiatrist*, *35*(10), 364–368. https://doi.org/10.1192/pb.bp.110.033449

Kooth-Pulse-2021-Report.pdf. (n.d.). Retrieved June 27, 2025, from https://explore.kooth.com/wp-content/uploads/2021/05/Kooth-Pulse-2021-Report.pdf

Lee, D. D., & Seung, H. S. (1999). *Algorithms for Non-negative Matrix Factorization*.

Li, H., Kraut, R. E., & Zhu, H. (2021). Technical Features of Asynchronous and Synchronous Community Platforms and their Effects on Community Cohesion: A Comparative Study of Forum-based and Chat-based Online Mental Health Communities. *Journal of Computer-Mediated Communication*, *26*(6), 403–421. https://doi.org/10.1093/jcmc/zmab016

Liu, J. M., Li, D., Cao, H., Ren, T., Liao, Z., & Wu, J. (2023). *ChatCounselor: A Large Language Models for Mental Health Support* (No. arXiv:2309.15461). arXiv. https://doi.org/10.48550/arXiv.2309.15461

Low, D. M., Rumker, L., Talkar, T., Torous, J., Cecchi, G., & Ghosh, S. S. (2020). Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study. *Journal of Medical Internet Research*, *22*(10), e22635. https://doi.org/10.2196/22635

Malhotra, G., Waheed, A., Srivastava, A., Akhtar, M. S., & Chakraborty, T. (2022). Speaker and Time-aware Joint Contextual Learning for Dialogue-act Classification in Counselling Conversations. *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 735–745. https://doi.org/10.1145/3488560.3498509

Marshall, P., Booth, M., Coole, M., Fothergill, L., Glossop, Z., Haines, J., Harding, A., Johnston, R., Jones, S., Lodge, C., Machin, K., Meacock, R., Nielson, K., Puddephatt, J.-A., Rakic, T., Rayson, P., Robinson, H., Rycroft-Malone, J., Shryane, N., … Lobban, F. (2024). Understanding the Impacts of Online Mental Health Peer Support Forums: Realist Synthesis. *JMIR Mental Health*, *11*, e55750. https://doi.org/10.2196/55750

McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, *3*(29), 861. https://doi.org/10.21105/joss.00861

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research. *International Journal of Qualitative Methods*, *22*, 16094069231205789. https://doi.org/10.1177/16094069231205789

Naslund, J. A., Aschbrenner, K. A., Marsch, L. A., & Bartels, S. J. (2016). The future of mental health care: Peer-to-peer support and social media. *Epidemiology and Psychiatric Sciences*, *25*(2), 113–122. https://doi.org/10.1017/S2045796015001067

OpenAI, Hurst, A., Lerer, A., Goucher, A. P., Perelman, A., Ramesh, A., Clark, A., Ostrow, A. J., Welihinda, A., Hayes, A., Radford, A., Mądry, A., Baker-Whitcomb, A., Beutel, A., Borzunov, A., Carney, A., Chow, A., Kirillov, A., Nichol, A., … Malkov, Y. (2024). *GPT-4o System Card* (No. arXiv:2410.21276). arXiv. https://doi.org/10.48550/arXiv.2410.21276

Orben, A. (2020). Teenagers, screens and social media: A narrative review of reviews and key studies. *Social Psychiatry and Psychiatric Epidemiology*, *55*(4), 407–414. https://doi.org/10.1007/s00127-019-01825-4

Owen, D., Collados, J. C., & Espinosa-Anke, L. (2020). *Towards Preemptive Detection of Depression and Anxiety in Twitter* (No. arXiv:2011.05249). arXiv. https://doi.org/10.48550/arXiv.2011.05249

Park, A., Conway, M., & Chen, A. T. (2018). Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach. *Computers in Human Behavior*, *78*, 98–112. https://doi.org/10.1016/j.chb.2017.09.001

Pendse, S. R., Niederhoffer, K., & Sharma, A. (2019). Cross-Cultural Differences in the Use of Online Mental Health Support Forums. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–29. https://doi.org/10.1145/3359169

Pérez-Rojas, A. E., Lockard, A. J., Bartholomew, T. T., Janis, R. A., Carney, D. M., Xiao, H., Youn, S. J., Scofield, B. E., Locke, B. D., Castonguay, L. G., & Hayes, J. A. (2017). Presenting concerns in counseling centers: The view from clinicians on the ground. *Psychological Services*, *14*(4), 416–427. https://doi.org/10.1037/ser0000122

Phang, Y. S., Heaukulani, C., Martanto, W., Morris, R., Tong, M. M., & Ho, R. (2023). Perceptions of a Digital Mental Health Platform Among Participants With Depressive Disorder, Anxiety Disorder, and Other Clinically Diagnosed Mental Disorders in Singapore: Usability and Acceptability Study. *JMIR Human Factors*, *10*, e42167. https://doi.org/10.2196/42167

Prescott, J., Hanley, T., & Ujhelyi, K. (2017). Peer Communication in Online Mental Health Forums for Young People: Directional and Nondirectional Support. *JMIR Mental Health*, *4*(3), e29. https://doi.org/10.2196/mental.6921

Raihan, N., Puspo, S. S. C., Farabi, S., Bucur, A.-M., Ranasinghe, T., & Zampieri, M. (2024). MentalHelp: A Multi-Task Dataset for Mental Health in Social Media. In N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 11196–11203). ELRA and ICCL. https://aclanthology.org/2024.lrec-main.977

Rodgers, R. F., Laveway, K., Campos, P., & de Carvalho, P. H. B. (2023). Body image as a global mental health concern. *Cambridge Prisms: Global Mental Health*, *10*, e9. https://doi.org/10.1017/gmh.2023.2

Samari, E., Vaingankar, J. A., Tan, B., Chang, S., Tan, Y. W. B., Archana, S., Chua, Y. C., Tang, C., Lee, Y. P., Chew, C. S. E., Davis, C., Verma, S., & Subramaniam, M. (2025). Gender-based analysis of body dissatisfaction among youths in Singapore: Findings from the National Youth Mental Health Study. *Frontiers in Psychiatry*, *16*, 1505161. https://doi.org/10.3389/fpsyt.2025.1505161

Samsuri, S. (2024, February 16). *New online mental health forum inspired by Reddit offers "safe space" for youths to seek advice, professional help*. TODAY. https://www.todayonline.com/singapore/new-online-mental-health-peer-support-forum-reddit-youths-2362751

Sankaranarayanan, S., Borchers, C., Simon, S., Tajik, E., Ataş, A. H., Celik, B., Balzan, F., & Shahrokhian, B. (2025). *Automating Thematic Analysis with Multi-Agent LLM Systems*. Center for Open Science. https://doi.org/10.35542/osf.io/kq8zh_v1

Sinha, C., Dinesh, D., Heaukulani, C., & Phang, Y. S. (2024). Examining a brief web and longitudinal app-based intervention [Wysa] for mental health support in Singapore during the COVID-19 pandemic: Mixed-methods retrospective observational study. *Frontiers in Digital Health*, *6*, 1443598. https://doi.org/10.3389/fdgth.2024.1443598

Smith-Merry, J., Goggin, G., Campbell, A., McKenzie, K., Ridout, B., & Baylosis, C. (2019). Social Connection and Online Engagement: Insights From Interviews With Users of a Mental Health Online Forum. *JMIR Mental Health*, *6*(3), e11084. https://doi.org/10.2196/11084

Stevens, M., Cartagena Farías, J., Mindel, C., D'Amico, F., & Evans-Lacko, S. (2022). Pilot evaluation to assess the effectiveness of youth peer community support via the Kooth online mental wellbeing website. *BMC Public Health*, *22*(1), 1903. https://doi.org/10.1186/s12889-022-14223-4

Subramaniam, M., Abdin, E., Picco, L., Pang, S., Shafie, S., Vaingankar, J. A., Kwok, K. W., Verma, K., & Chong, S. A. (2016). Stigma towards people with mental disorders and its components – a perspective from multi-ethnic Singapore. *Epidemiology and Psychiatric Sciences*, *26*(4), 371–382. https://doi.org/10.1017/S2045796016000159

Subramaniam, M., Vaingankar, J., Tan, B., Abdin, E., Chang, S., Tan, Y., Samari, E., Archana, S., Chua, Y., Lee, J., Tang, C., Lee, Y., Chong, S., & Verma, S. (2025). Examining psychological distress among youth in Singapore: Insights from the National Youth Mental Health Study. *Asian Journal of Psychiatry*, *105*, 104405. https://doi.org/10.1016/j.ajp.2025.104405

Tsakalidis, A., Nanni, F., Hills, A., Chim, J., Song, J., & Liakata, M. (2022). *Identifying Moments of Change from Longitudinal User Text* (No. arXiv:2205.05593). arXiv. https://doi.org/10.48550/arXiv.2205.05593

Weng, J. H., Hu, Y., Heaukulani, C., Tan, C., Chang, J. K., Phang, Y. S., Rajendram, P., Tan, W. M., Loke, W. C., & Morris, R. J. T. (2024). Mental Wellness Self-Care in Singapore With mindline.sg: A Tutorial on the Development of a Digital Mental Health Platform for Behavior Change. *Journal of Medical Internet Research*, *26*, e44443. https://doi.org/10.2196/44443

Williams, A. H., Williams, A. W., Renner, L., PettyJohn, M. E., Cash, S. J., & Schwab-Reese, L. M. (2024). "I Need to Talk to Someone…What Do I Do?": Peer-to-Peer Disclosures of Child Maltreatment on Social Media. *Journal of Family Violence*. https://doi.org/10.1007/s10896-024-00697-w

Wong, J. C. M., Lei, N. C. L., Kom, D. M. Y., Fee, V. H. F., Yap, N. H., Teoh, J. Y., Shen, L., Yap, Q. V., Wan, M. J. S., Du, R., Kroneman, L., Fung, D. S. S., Ong, S. H., Tian, C. S., Bin Sachiman, M. N. S., Sii, N. E.-P., Teng, J. Y., Ng, T. P., & Verhulst, F. (2024). Youth Epidemiology and Resilience (YEAR) in a student population: Prevalence and predictors of mental health symptoms. *Frontiers in Psychiatry*, *15*. https://doi.org/10.3389/fpsyt.2024.1454484

Xu, J., Wei, T., Hou, B., Orzechowski, P., Yang, S., Jin, R., Paulbeck, R., Wagenaar, J., Demiris, G., & Shen, L. (2025). *MentalChat16K: A Benchmark Dataset for Conversational Mental Health Assistance* (No. arXiv:2503.13509). arXiv. https://doi.org/10.48550/arXiv.2503.13509

Appendix

A1. Algorithm

Algorithm 1 Consistent Clustering Algorithm via Meta-Clustering over Topic Representations

Input: Dataset of forum posts, number of iteration (iter)
Output: Consistent clusters and posterior distribution on each forum post over topics extracted from the algorithm
for $i = 1$ to iter **do**
    Run BERTopic and record topic representation in dictionary
end for
Lemmatize vocabularies in the list of words in individual topic representations
Join the preprocessed vocabularies as a sentence and vectorize the data using countervectorizer
M = matrix from countervectorizer
Run NMF algorithm on the preprocessed data
W, H = NMF over matrix M
Obtain the top 15 words representing each meta-cluster
Use GPT-4 prompt to obtain (summarized) labels and tag the posts to the label generated accordingly

A2. CLICC concerns

The Clinician Index of Client Concerns is a validated framework for determining presenting concerns. Our restricted list of themes and their explanations can be found in Tab. A1.

TABLE A1: CLICC concerns and their explanations.

Concern	Explanation	Concern	Explanation
academic performance	Worry over grades, school performance, or school-related stress	autism spectrum	References to being autistic or challenges related to neurodivergence in social or sensory processing.
addiction (not drugs or alcohol)	Dependency on substances or behaviors (e.g. gambling, gaming, pornography). Exclude alcohol and drugs	bipolar disorder	References to being bipolar or challenges related to bipolar symptoms (eg. extreme mood swings between mania and depression)
adjustment to new environment	Difficulties adapting to a new school, country, job, or life change	burnout	Exhaustion from prolonged stress (often academic or professional)
adulting	Struggles with independence, responsibility, or transitions into adulthood	career	Job dissatisfaction, unemployment, or career decision-making stress
alcohol	Use of alcohol or related concerns, including dependency, binge drinking, or impact on functioning	depression	Feelings of sadness, emptiness, hopelessness, or loss of interest in activities
anger management	Struggles to manage anger in healthy ways	discrimination	Feeling unfairly treated based on identity or traits (e.g., gender, disability)

attention/concentration difficulties	Problems with focus or attention (possibly related to ADHD)	dissociative experiences	Feeling detached from reality or oneself (e.g., “watching life from outside my body”)
drugs	Non-medical or recreational use of substances (other than alcohol), including concerns about dependency, usage patterns, or impact	financial	Worries about money, debt, or basic needs
eating/body image	Concerns about weight, food habits, or dissatisfaction with appearance	gender identity	Concerns or exploration related to one's internal sense of gender, which may differ from the sex assigned at birth
emotion dysregulation	Difficulty managing intense or unstable emotions	generalised anxiety	Persistent worry, nervousness, or fear not specific to any particular contexts
family	Issues or dynamics involving parents, siblings, or extended family	gratitude	Expression of appreciation or reflection on positive things
grief/loss	Mourning a death or significant loss (e.g., relationship, pet)	interpersonal functioning	General difficulties in getting along with others or maintaining relationships
harassment/emotional abuse (victim)	Repeated emotional mistreatment or manipulation	learning disorder/disability	Challenges due to conditions like dyslexia, dyscalculia, etc
health/medical	Physical health issues affecting well-being	legal/judicial/conduct	Legal troubles, fear of legal action, or criminal behavior
health service dissatisfaction	Dissatisfaction or negative experiences with medical, mental health, or counselling services, including access, quality, or interactions with providers	mood instability	Frequent or intense mood swings, irritability, or emotional unpredictability not tied to a clear cause or diagnosis
identity development	Exploration or confusion around personal identity (e.g., gender, values)	obsessions or compulsions	Intrusive thoughts or repetitive behaviors driven by anxiety
panic attack(s)	Sudden episodes of intense fear or discomfort with physical symptoms (e.g., heart palpitations, shortness of breath) not tied to a specific threat	racial, ethnic or cultural concerns	Discrimination, cultural identity issues, or race-related distress
perfectionism	Rigid self-imposed standards, fear of failure, or chronic dissatisfaction with performance	relationship problem (specific)	Romantic relationship struggles (e.g., breakups, dating anxiety)
physical abuse/assault (victim)	Physical harm or threat of harm by another person	religion/spirituality	Issues related to faith, religious conflict, or spiritual meaning

political concerns	Political events, ideologies, government policies, or civic issues	suicidality	Thoughts of ending one's life, whether passive or active
pregnancy related	Pregnancy, including unplanned pregnancy, parenting readiness, or abortion, etc.	self-esteem/confidence	Negative self-perception or lack of belief in oneself
psychotic thoughts or behaviours	Hallucinations, delusions, or breaks from reality	self-injurious thoughts or behaviors	Mention of self-harm (e.g., cutting, burning) without suicidal intent
setting boundaries	Learning to say no or manage expectations from others	sleep	Insomnia, oversleeping, or disrupted sleep
sexual abuse/assault (victim)	Disclosure of sexual violence or coercion	social anxiety	Fear or discomfort in social situations due to worry of being judged or embarrassed
sexual concern	Issues or confusion related to sexual experiences, sexual function, desire, behavior, or consent, excluding trauma cases	social comparison	Comparing oneself to peers, often negatively
sexual orientation	Mentions of one's romantic or sexual attraction toward others that is non-heterosexual in nature, including questioning or coming out	social isolation	Loneliness or avoiding contact with others
social withdrawal	Persistent/Increasing avoidance of social interaction or participation	stress	Emotional strain or tension due to life demands or responsibilities
specific phobia	Intense fear of a specific object/situation (e.g., spiders, heights, flying)	teenage pregnancy	Concerns or experiences involving pregnancy during adolescence, including stigma, decision-making, and emotional impact
stalking (victim)	Unwanted, persistent attention or surveillance	test-taking anxiety	Significant distress or fear specifically related to exams or assessments, which impairs performance or well-being
therapy experiences	Reflections on past/current therapy or seeking help	unspecified/other anxiety	General anxiety-related concerns that do not clearly fit into a specific subtype, but still cause noticeable distress
time management	Feeling overwhelmed due to time constraints or unable to manage tasks efficiently	none	No concerns
trauma	Emotional or psychological responses to distressing life events	other	Other concerns not listed but is a distinct factor on its own (please name the concern in

			parenthesis after indicating as 'other')
violent thoughts or behaviors towards others	Mention of wanting or attempting to hurt others		

A3. LLM System Prompt for Themes Extraction

You are a helpful assistant that extracts Clinician Index of CLient Concerns (CLICC) concerns from forum posts about mental health.

Your goal is to identify what users are discussing and why they are seeking support — *not* to diagnose them.

Each post has a main, top concern, and an optional list of additional presented concerns.

The top concern should be the main issue the user seems most focused on.

Guidelines

- Understand the content contextually—read the entire post and identify the psychological or emotional themes expressed.

- Identify the main issue the original poster (OP) is discussing or struggling with. This becomes the top concern.

- Identify any additional issues mentioned or implied that are relevant to the OP’s experiences. These become the presenting concerns.

- Choose only concerns **explicitly supported** by the post.

- Do **not** over-pathologise or infer diagnoses.

- Use only concerns from this list:

[{concerns_list:s}]

- The **top concern** is the user's **main issue** — either the most discussed or clearly first stated.

- Separate concerns with commas, starting with the top concern.

- Return `'none'` if no concerns can be reliably extracted or if the post is not centred around a mental health issue.

- Maintain consistency in terminology and formatting across all extractions

Concerns

The possible presented concerns are described here:

{concerns_list_with_descriptions:s}

Output format

- Return only the concerns themselves.

- Each concern label must exactly match the wording from this list.

- Do NOT include labels like "Top concern:" or "Presenting concerns:" in your response.

- Do NOT include newline characters or extra formatting.

- Return concerns in a **flat, comma-separated list**, all enclosed in **single quotes**.

- Start the list with the **top concern**, followed by additional concerns.

- If no concerns can be identified, return `'none'` (as a string, not a list).

Correct format:

'family', 'self-esteem/confidence', 'suicidality'

Incorrect format:

- ["Top concern: family", "Presenting concerns: ..."]

- ['family\n\npresenting concerns: ...']

- "family, self-esteem/confidence, suicidality"

Examples

Example 1

User post:

Sometimes I just want to disappear. It's like I need a long break that I cannot find to do. Im too busy with work, I need to work. Sometimes I just want to pause the time, or worse I just want to disappear from everyone I know. Sometimes, I have the curiosity on life after death, on what it truly feels like. I bet it's the beginning of experiencing true peace. Im so disappointed about life because I think im not capable in doing everything. I just dont see my potential in my job. Im so tired already. I feel so dumb and stupid.

Top concern:

social withdrawal

Presenting concerns:

social isolation, career

Response:

'social withdrawal', 'social isolation', 'career'

Example 2

User post:

To be honest, idk which place is a better category to post this, under "trigger warning" or "my parents". My parent used to abuse my physically ever since I was small. And ever since I grew into a teenager it switched to verbal abuse. But the physical abuse didnt stop either, just reduced because she use her words towards me.

Its hard for me because no one ever came forward to help me, even though alot of people witness my getting abuse with their own 2 eyes. Back then, I thought it was normal, not until I start sharing with my *friends my experiences about it. Worse part of it, I was outcasted by those "friends" because they think that I was lying about my experiences.*

Suicide was always an option to me especially when I was told that I was never wanted and that I was such a difficult child to the extend my parent would rather go to jail for murdering me than to continue living with me. I never had problem with my appearance, it didnt matter if I gained weight or not having perfect skin, but when my own parent compares me with other girls and keep mentioning that I NEED to lose weight and dress up more, it definitely made me upset.

And after experiencing being an outcast at school for sharing my experiences to people I trust, I started to have trust issues. But things were so bad at home that there are times I go to school crying uncontrollably (like early in the morning before assembly back in sec sch) and I have no choice to let things out to my close friends.

Now that I am 18, I have been thinking of leaving home, but there is a constant fear of getting caught by authorities and having to return back home because then it would make the situation worst. However, its also not easy to live with someone that constantly bring you down and caused you so much trauma in your life. Thoughts?

Top concern:

family

Presenting concerns:

harassment/emotional abuse (victim), physical abuse/assault (victim), self-esteem/confidence, suicidality

Response:

'family', 'harassment/emotional abuse (victim)', 'physical abuse/assault (victim)', 'self-esteem/confidence', 'suicidality'

Example 3

User post:

As per previous posts, my boyfriend depression has relapsed and our relationship dynamic has changed.

These past few weeks have been tough on both of us as it is my first time dealing with his depression and he has been busy with work and school.

There is a conflict inside of me where I want to break up with him but I still really like him.

I don't feel as happy and as emotionally secured in the relationship; When we meet, I am almost always crying because of his emotional distance.

I still really like him and I have thought of building a family with him but i know i am not emotionally strong to support his depression and I am fearful of the future.

He does not go for therapy or take medication because he tells me they don't work.

At the same time, I also fear if we break up, how would it impact him.

I don't want his depression to get worsened.

I am hoping that when we are free in a few days time, we coulf do activities that we both enjoy and he would feel better.

But in the long run, I am not confident in our future.

Top concern:

relationship problem (specific)

Presenting concerns:

depression

Response:

'relationship problem (specific)', 'depression'

A4. Characteristic Posts for Each Topic

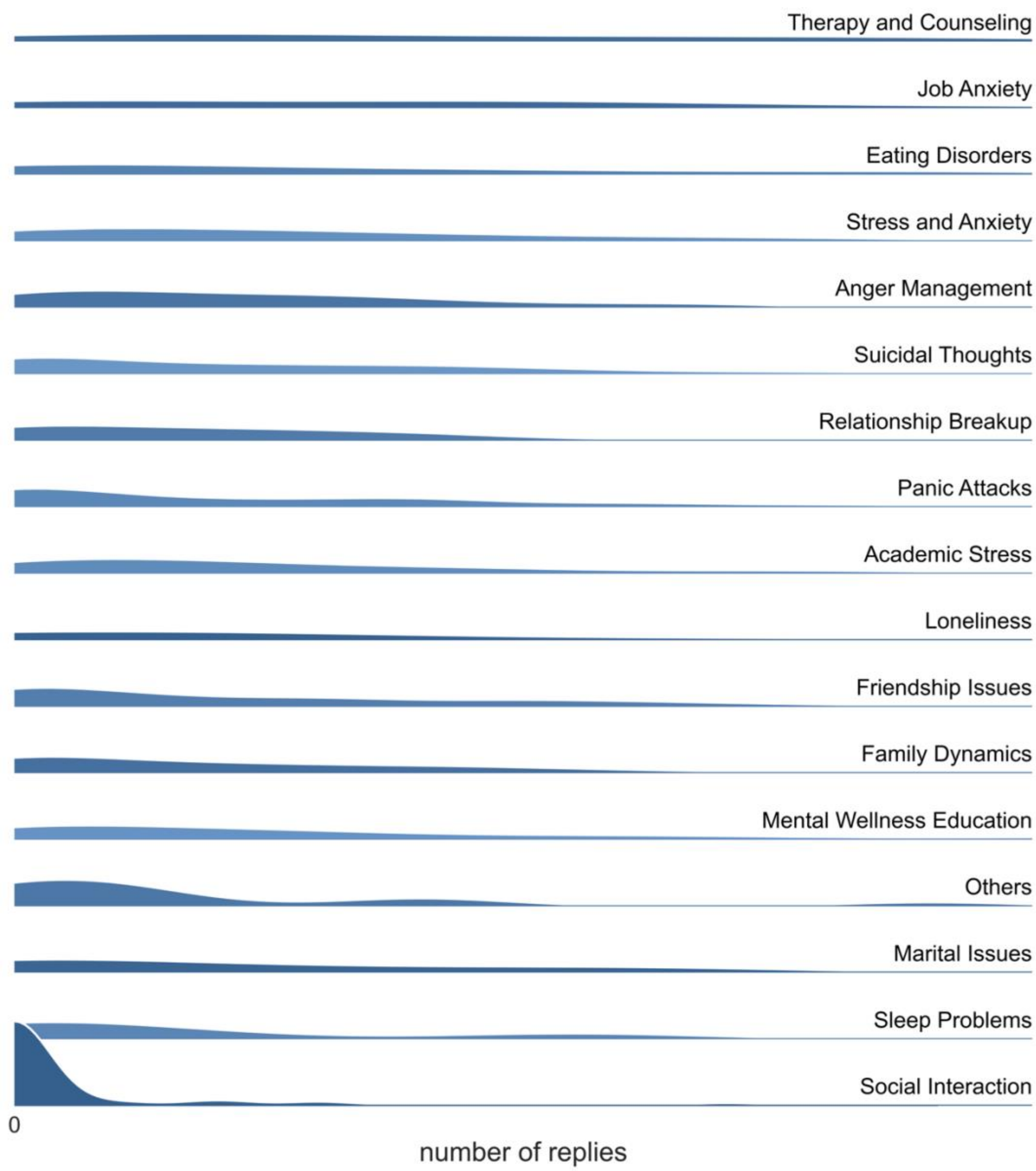
TABLE A2: Characteristic posts on *let's talk* for each BERTopic topic.

Topic	Representation	Characteristic Posts
Stress and Anxiety*	'anxiety', 'just', 'feel', 'stress', 'stressed', 'feeling'	recently alot of **stress** has been piling up work family relationship and money not sure if its the **stress** thats causing me to be annoyed at everything and everyone or am i actually having signs of depression **anxiety** issue
Mental Wellness Education	'mental', 'health', 'resilience', 'talk', 'anxiety', 'wellness', 'depression', 'thinking', 'help', 'stress', 'normal', 'education', 'feeling', 'parents', 'illness'	my **mental health** is an on and off issue i am currently on medication my medication has cause me a weight gain which makes me feel depressed i tried to stop taking my medicine but i feel unsafe when i never take it so in the end i take it i tried to workout and take weight loss supplement…
Friendship Issues	'friend', 'friends', 'friendship', 'group', 'make', 'talk', 'best', 'social', 'relationship', 'betrayed', 'just', 'bestfriend', 'shes', 'trust', 'felt'	i feel frustrated by a person at school she used to be my **friend** but she often made me feels less and puts a negative influence on me i always feels like i have to compete with other people for her attention in able to keep her as my **friend** after i realised that this **relationship** is toxic…
Academic Stress	'study', 'exam', 'studying', 'motivation', 'grade', 'just', 'school', 'stress', 'stressed', 'quit', 'grading', 'table', 'exams', 'uni', 'join'	hi all how would you try to stop yourself from procrastinating assignment week submission is going to happen in weeks time and I **feel** a bit **stressed** that I might procrastinate and struggle to complete my work on time what will you do
Suicidal Thoughts	'self', 'suicidal', 'thoughts', 'depression', 'suicide', 'harm', 'sad', 'don', 'depressed', 'negative', 'therapy', 'harming', 'just', 'feel', 'anxiety'	i feel heavy every single day i have **suicidal** ideations and urges to **harm** myself for the past years ive been like this…
Family Dynamics	'family', 'parent', 'mom', 'dad', 'mother', 'mum', 'sister', 'father', 'si', 'sibling', 'members', 'parents', 'want', 'abuse', 'jail'	my **mother** ended our ties over a phone call and blocked me after the call i do not know what to do it opened up my old wounds again
Relationship Breakup*	'breakup', 'relationship', 'bf', 'leave', 'feel'	i need help i m going through a really bad **breakup** i lost the person i used to share my whole life with things are really tough right now i m getting worse day by day i don t think i ll ever be okay unless he comes back and i hope he comes back one day i have no one to share how i **feel** there s no one there for me really i m just at my worst

Panic Attacks	'panic', 'anxiety', 'pain', 'attacks', 'attack', 'symptom', 'having', 'palpitation', 'palpitations', 'nausea', 'anxious', 'chest', 'diagnosed', 'reflux', 'vomiting'	i have been experiencing similar to **anxiety** and depression symptoms daily for a few years symptoms such as constant breathlessness impending sense of doom heart **palpitations** suicide thoughts heavy feelings of feeling anxious i have tried many methods to calm myself down but it s not effective
Social Interaction	'ppl', 'person', 'affect', 'talk', 'say', 'counsellor', 'attack', 'psych', 'dis', 'people', 'feelings', 'stupid', 'abt', 'got', 'self'	seriously no one can u trust even ur own **ppl**
Eating Disorders	'weight', 'eating', 'eat', 'lost', 'binge', 'losing', 'lose', 'gaining', 'loss', 'skinny', 'stress', 'insecure', 'gain', 'diet', 'struggling'	I feel like im not deep into ed enough to get help ive been not able to **eat** without thinking about calories and sometimes I don’t **eat** just because im scared of **gaining weight** I don’t know how to explain it
Job Anxiety	'anxious', 'nervous', 'anxiety', 'stressful', 'going', 'interview', 'phobia', 'feeling', 'job', 'quit', 'getting', 'start', 'matter', 'constantly', 'talk'	i recently **quit** from a **stressful job** and looking for a new one for about month plus but there isnt any calls for interviews i engaged a career coach from gov this thu jun hopefully can get more tips on **getting** my next **job**
Sleep Problems	'sleep', 'tired', 'hours', 'time', 'feeling', 'exhausted', 'adhd', 'sleeping', 'normal', 'hrs', 'waking', 'feel', 'sufficient', 'rest', 'trouble'	i **feel** like no matter how much i **sleep** i still wanna **sleep** all the time my mom has complained to me that sleeping this much will cause issues is it true
Therapy and Counseling	'therapist', 'mental', 'therapy', 'counsellor', 'psychologist', 'mentor', 'healthcare', 'help', 'counselling', 'health', 'singapore', 'counselor', 'mistreatment', 'hospital', 'talking'	hi therapists out there i am wondering if there is anyway for a minor to visit a **therapist** without parental knowledge i feel like i am someone that has a lot of emotional turmoil inside and i would like an experts…
Anger Management	'anger', 'angry', 'family', 'takes', 'feel', 'issues', 'emotions', 'moment', 'like', 'control', 'irritable', 'outbursts', 'im', 'overcome', 'quarrel'	how to **control** and maintain my **anger** whenever something goes wrong in the groupwork i do have groupmates who are not contributing at the start of a group project it gets better now during that period i was really stressed and whenever there s group project discussion and they are not contributing my tone gets really **angry** how do i control my **anger**
Marital Issues	'husband', 'divorce', 'wife', 'marriage', 'connection', 'pregnant', 'spouse', 'anger', 'management', 'birth', 'giving', 'emotional', 'emotions', 'married', 'care'	over the last mths i seem to be running out of patience with my **spouse** i am male yrs old over the last decade i tolerated with her tyranny and verbal abuse i therefore moved out of master bedroom to avoid more interaction with her interaction is already minimal to avoid friction is this normal help thks
Others*	'losing investments', 'unappreciated marriage', 'struggle cope', 'lied saving', 'live past'	eversince my past relationship losing k on investments i feel so worthless useless and feel like just giving up on life i have a

		boyfriend that i can confide to but i don t have the guts to do…
Loneliness	'lonely', 'feel', 'loneliness', 'talk', 'just', 'make', 'really', 'friends', 've', 'life', 'emptiness', 'single', 'want', 'friendship', 'person'	i love being alone i m out of words when i meet people so i don t want to **talk** with people and i don t wanna make **friends** i love doing something myself without helps i don t think i **want** anyone rn that s why i m being end up still **single** no **friends** no family but i m doing good is it alright

A5. More user engagement results



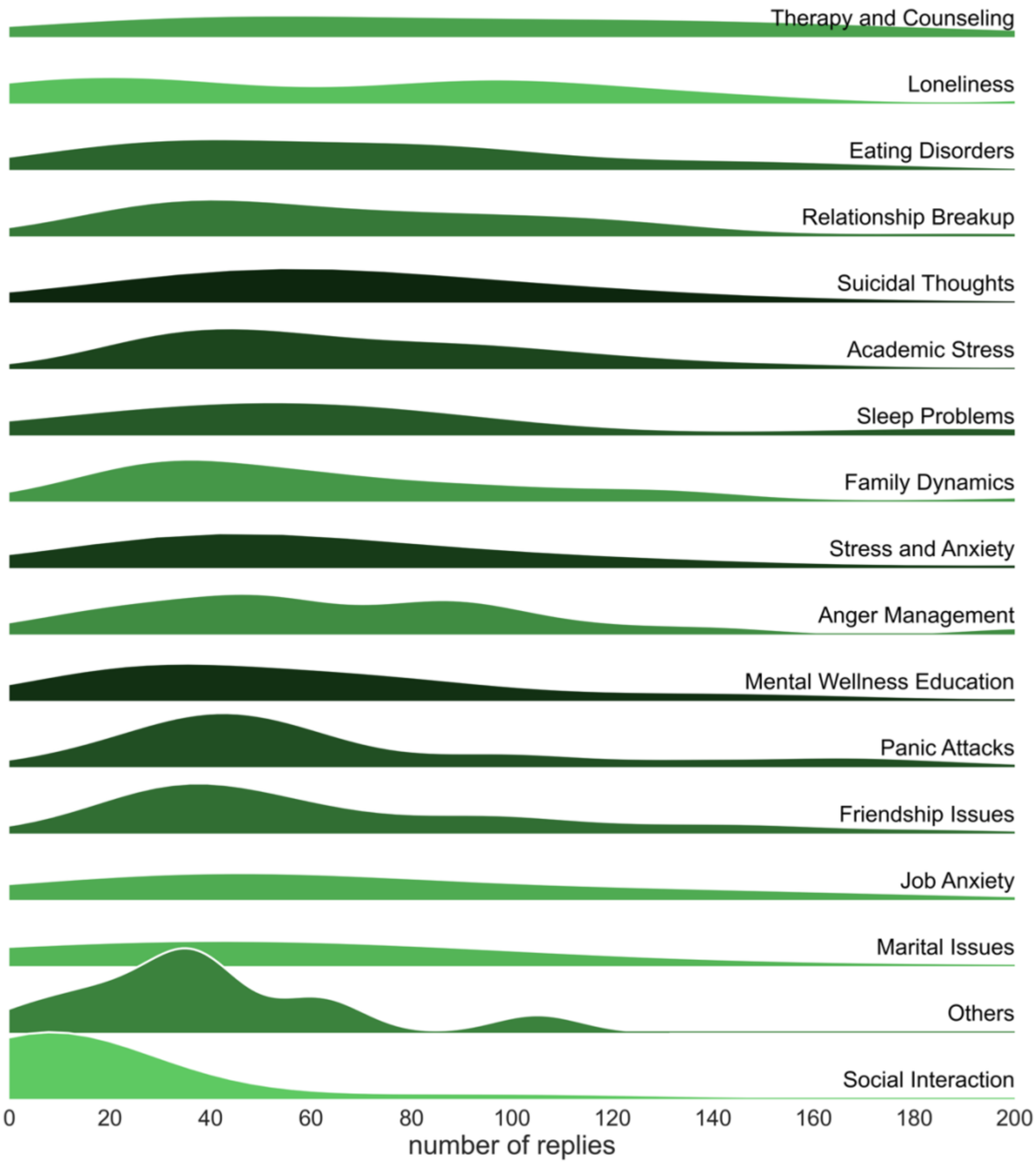


Figure A1: Distribution of likes and reads by topic. Topics are ranked by median number.

A6. Topics and Emotions Associations on *let's talk*

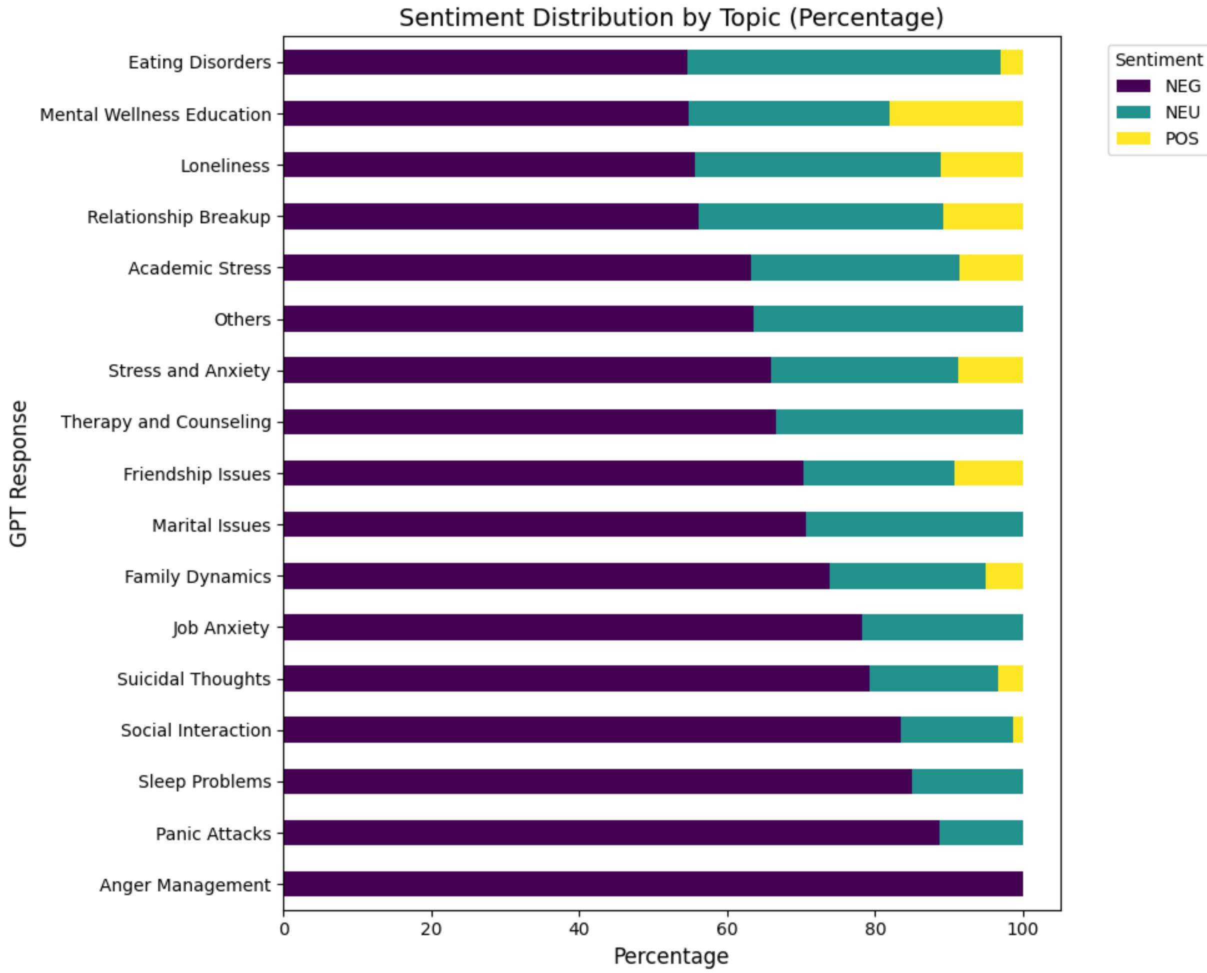


Figure A2: Sentiment Distribution across BERTopic.

A7. Multi-theme Extraction Results

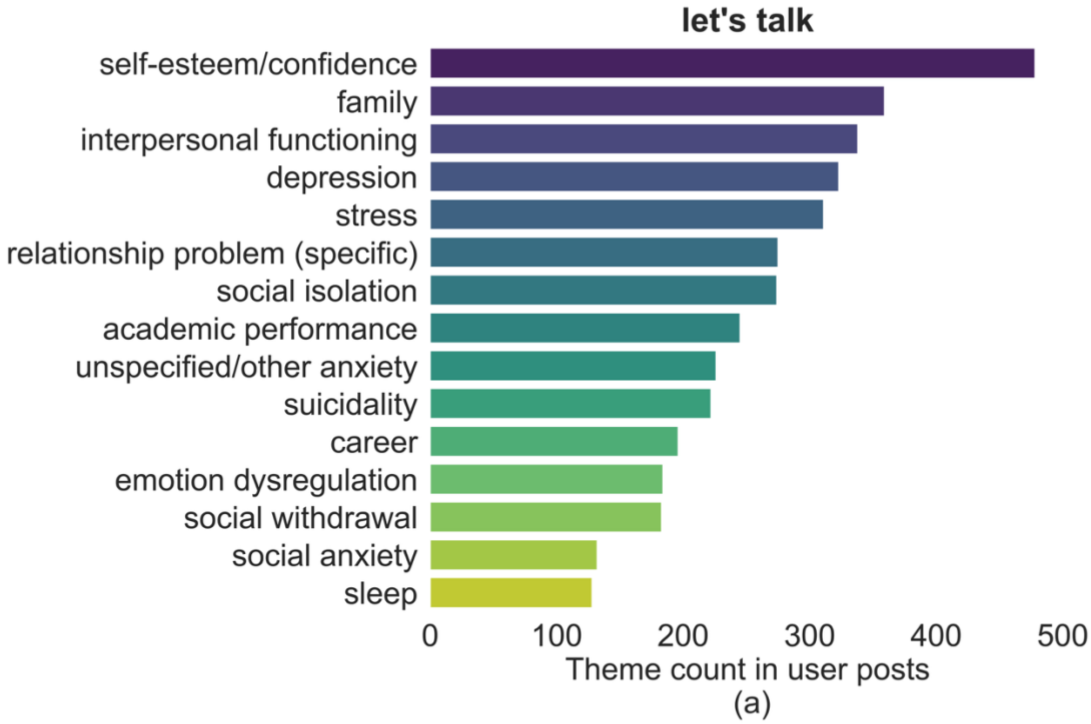
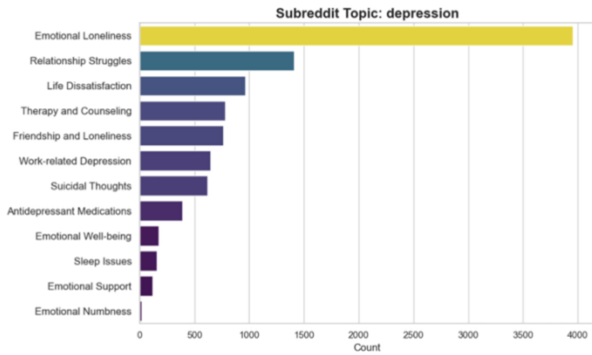


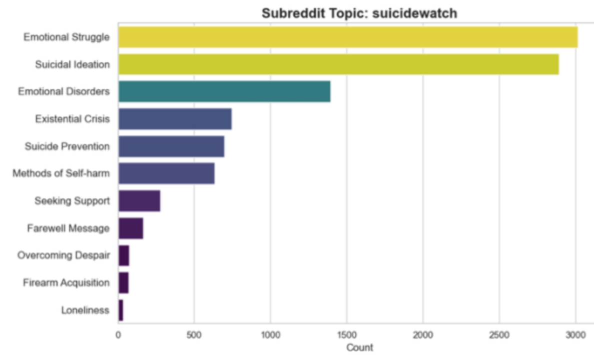
Figure A3: Top 15 LLM-extracted CLICC themes allowing for multiple themes per post.

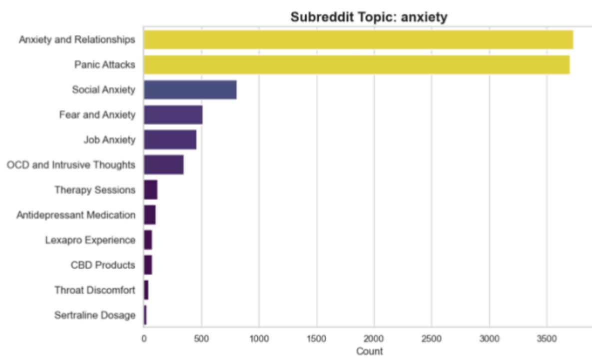
A8. Hardwarezone Matched Keywords Lists

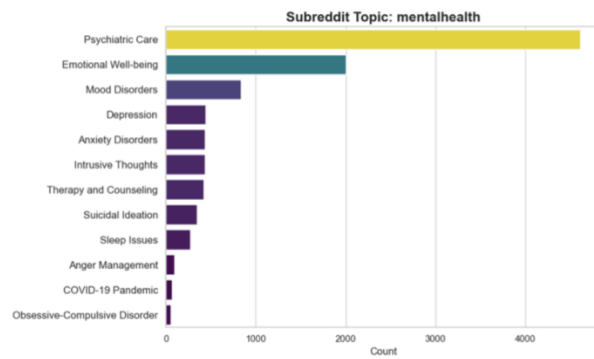
'academic stress', 'addiction', 'adhd', 'alcohol', 'alcohol recovery', 'antidepressants', 'anxiety', 'anxious', 'autism', 'binge', 'binge eating', 'bipolar', 'borderline', 'bulimia', 'coming out', 'compassion fatigue', 'counselling', 'counselor', 'cutting', 'cyberbullying', 'delusion', 'depressed', 'depression', 'doomscrolling', 'drugs', 'fatigue', 'favoritism', 'financial stress', 'flashback', 'flashbacks', 'hallucination', 'hyperventilation', 'illness', 'insomnia', 'insomnia and anxiety', 'living alone', 'mania', 'mental', 'mood', 'narcissism', 'nightmares', 'ocd', 'panic', 'panic attacks', 'personality', 'phobia', 'psych', 'psychiatrist', 'ptsd', 'schizophrenia', 'side effects', 'sleep', 'social media anxiety', 'stigma', 'stress', 'substance', 'suicidal', 'suicidal ideation', 'suicide', 'therapy', 'tired', 'trauma', 'wellness', 'worry'

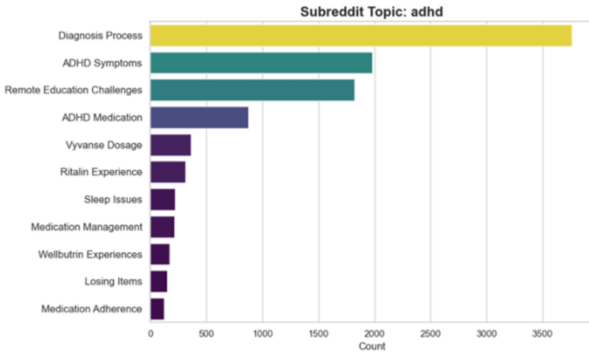
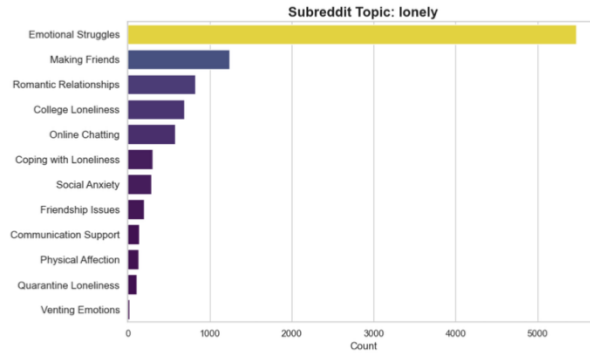
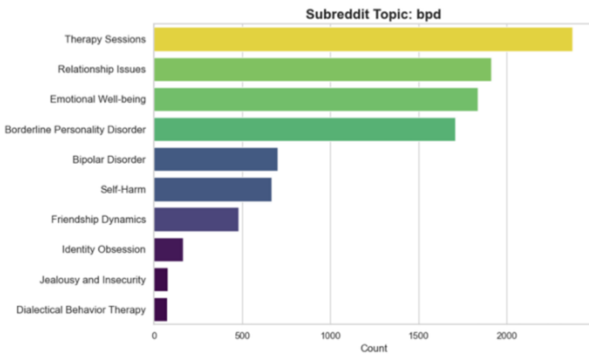
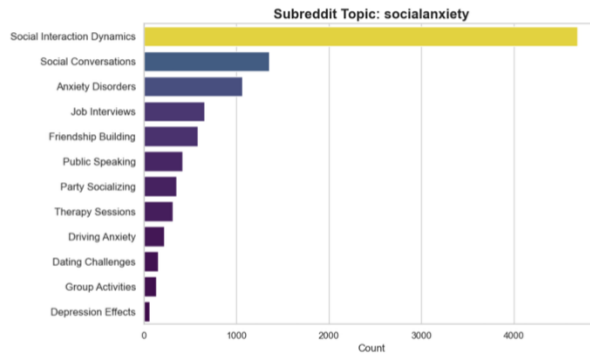
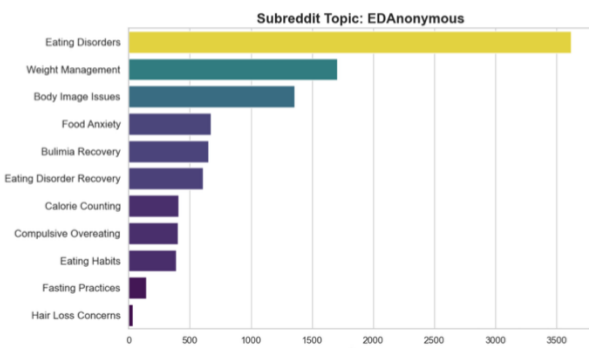
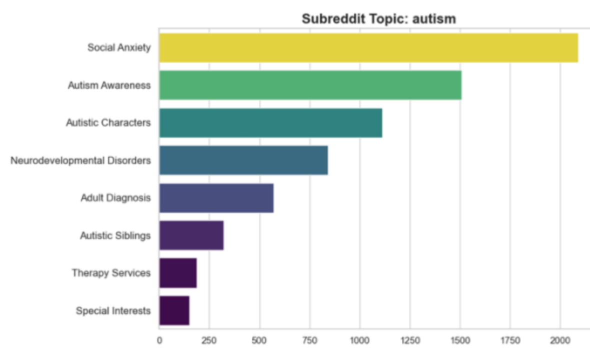
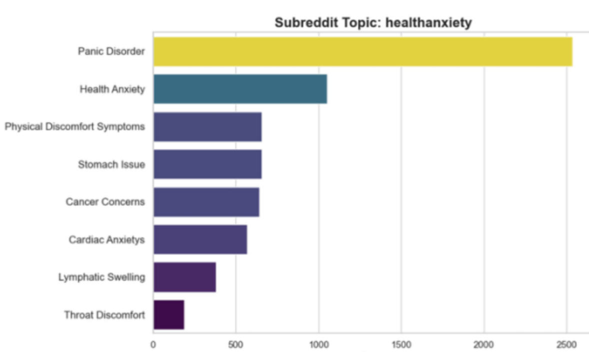
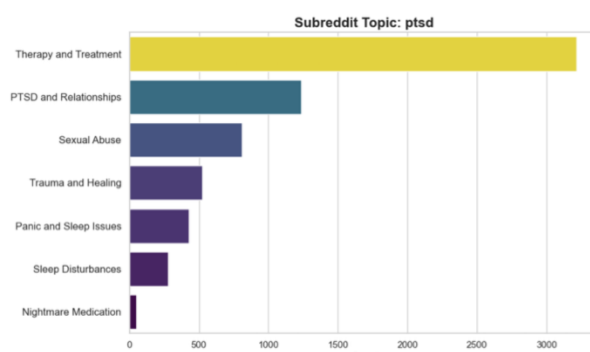
A9. Subreddit-specific results










Subreddit Topic: adhd
Diagnosis Process
ADHD Symptoms
Remote Education Challenges
ADHD Medication
Vyvanse Dosage
Ritalin Experience
Sleep Issues
Medication Management
Wellbutrin Experiences
Losing Items
Medication Adherence
0 500 1000 1500 2000 2500 3000 3500
Count

Subreddit Topic: lonely
Emotional Struggles
Making Friends
Romantic Relationships
College Loneliness
Online Chatting
Coping with Loneliness
Social Anxiety
Friendship Issues
Communication Support
Physical Affection
Quarantine Loneliness
Venting Emotions
0 1000 2000 3000 4000 5000
Count

Subreddit Topic: bpd
Therapy Sessions
Relationship Issues
Emotional Well-being
Borderline Personality Disorder
Bipolar Disorder
Self-Harm
Friendship Dynamics
Identity Obsession
Jealousy and Insecurity
Dialectical Behavior Therapy
0 500 1000 1500 2000
Count

Subreddit Topic: socialanxiety
Social Interaction Dynamics
Social Conversations
Anxiety Disorders
Job Interviews
Friendship Building
Public Speaking
Party Socializing
Therapy Sessions
Driving Anxiety
Dating Challenges
Group Activities
Depression Effects
0 1000 2000 3000 4000
Count

Subreddit Topic: EDAnonymous
Eating Disorders
Weight Management
Body Image Issues
Food Anxiety
Bulimia Recovery
Eating Disorder Recovery
Calorie Counting
Compulsive Overeating
Eating Habits
Fasting Practices
Hair Loss Concerns
0 500 1000 1500 2000 2500 3000 3500
Count

Subreddit Topic: autism
Social Anxiety
Autism Awareness
Autistic Characters
Neurodevelopmental Disorders
Adult Diagnosis
Autistic Siblings
Therapy Services
Special Interests
0 250 500 750 1000 1250 1500 1750 2000
Count

Subreddit Topic: healthanxiety
Panic Disorder
Health Anxiety
Physical Discomfort Symptoms
Stomach Issue
Cancer Concerns
Cardiac Anxietys
Lymphatic Swelling
Throat Discomfort
0 500 1000 1500 2000 2500
Count

Subreddit Topic: ptsd
Therapy and Treatment
PTSD and Relationships
Sexual Abuse
Trauma and Healing
Panic and Sleep Issues
Sleep Disturbances
Nightmare Medication
0 500 1000 1500 2000 2500 3000
Count

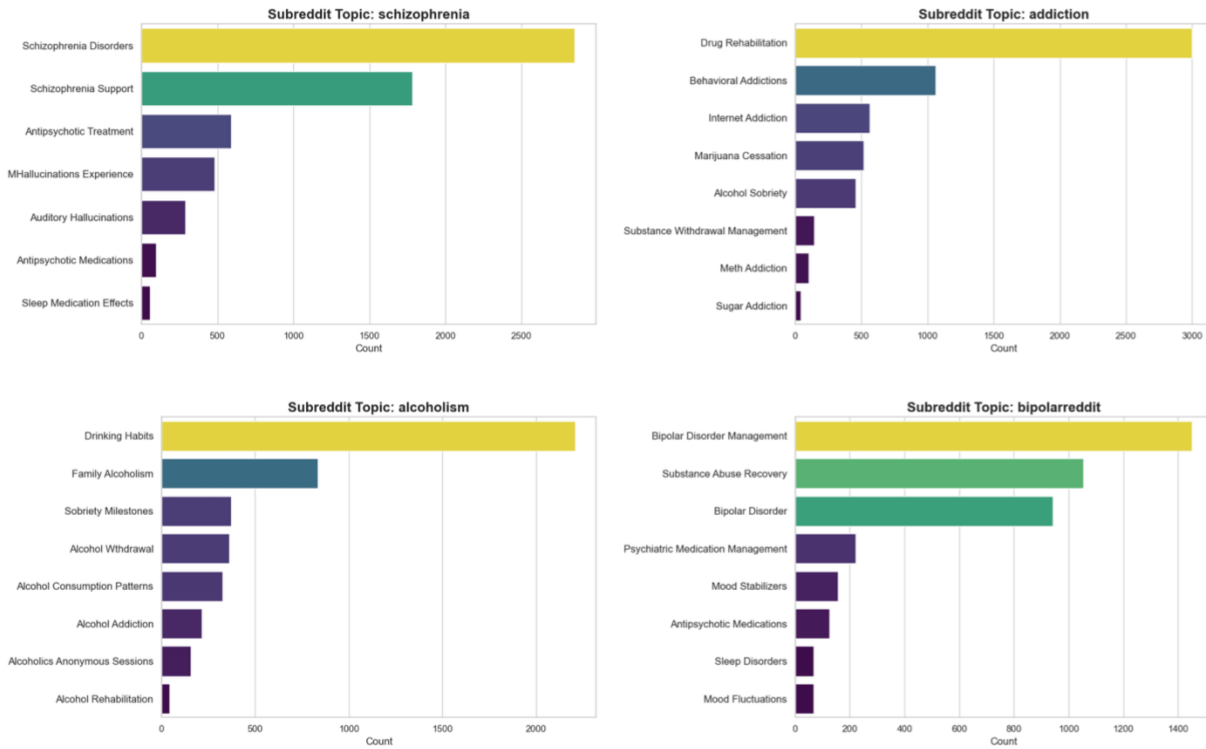


Figure A4: BERTopic-extracted topics from Reddit subreddits.

A10. Topic cluster analysis

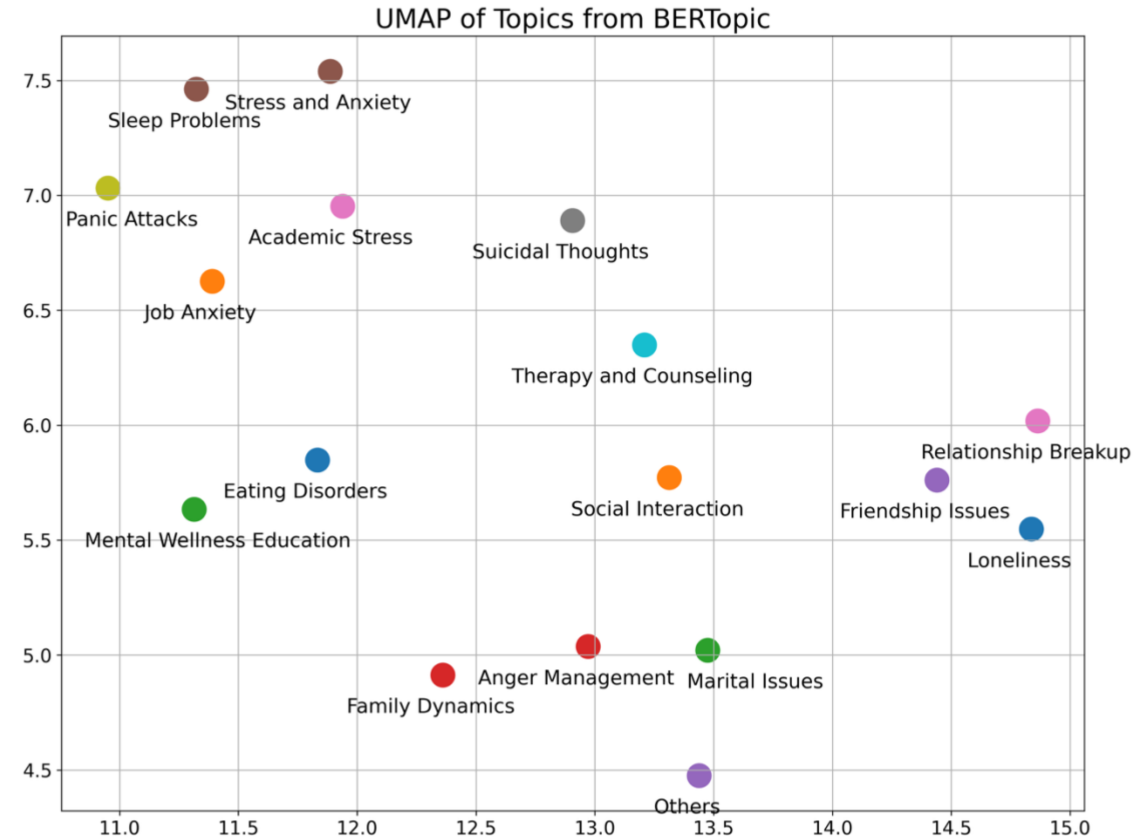


Figure A5: Two-dimensional UMAP projection of mental health topics extracted with BERTopic consistent clustering.

To explore the relationship between topics, Fig. A5 presents a two-dimensional UMAP projection of topic embeddings derived from BERTopic consistent clustering. The spatial proximity reflects semantic similarity based on the embedding space, highlighting the interconnected nature of mental health issues (Borsboom, 2017). Three distinct clusters emerged. The topics “Relationship Breakup”, “Friendship Issues”, and “Loneliness” cluster together, reflecting users’ *interpersonal* concerns and emotionally driven discussions around romance, social connection, and isolation. Second, “Stress and Anxiety”, “Sleep Problems”, “Academic Stress”, “Job Anxiety” and “Panic Attacks” formed a coherent grouping, consistent with established bidirectional relationships between stress and sleep quality (Åkerstedt et al., 2002; Almojali et al., 2017) and highlights how stress and anxiety are commonly expressed when discussing academic and occupational issues. Third, “Family Dynamics”, “Marital Issues” and “Anger Management” formed a cluster that reveals family tensions, including those between partners or relatives, are closely linked to discussion of anger and emotional outbursts.

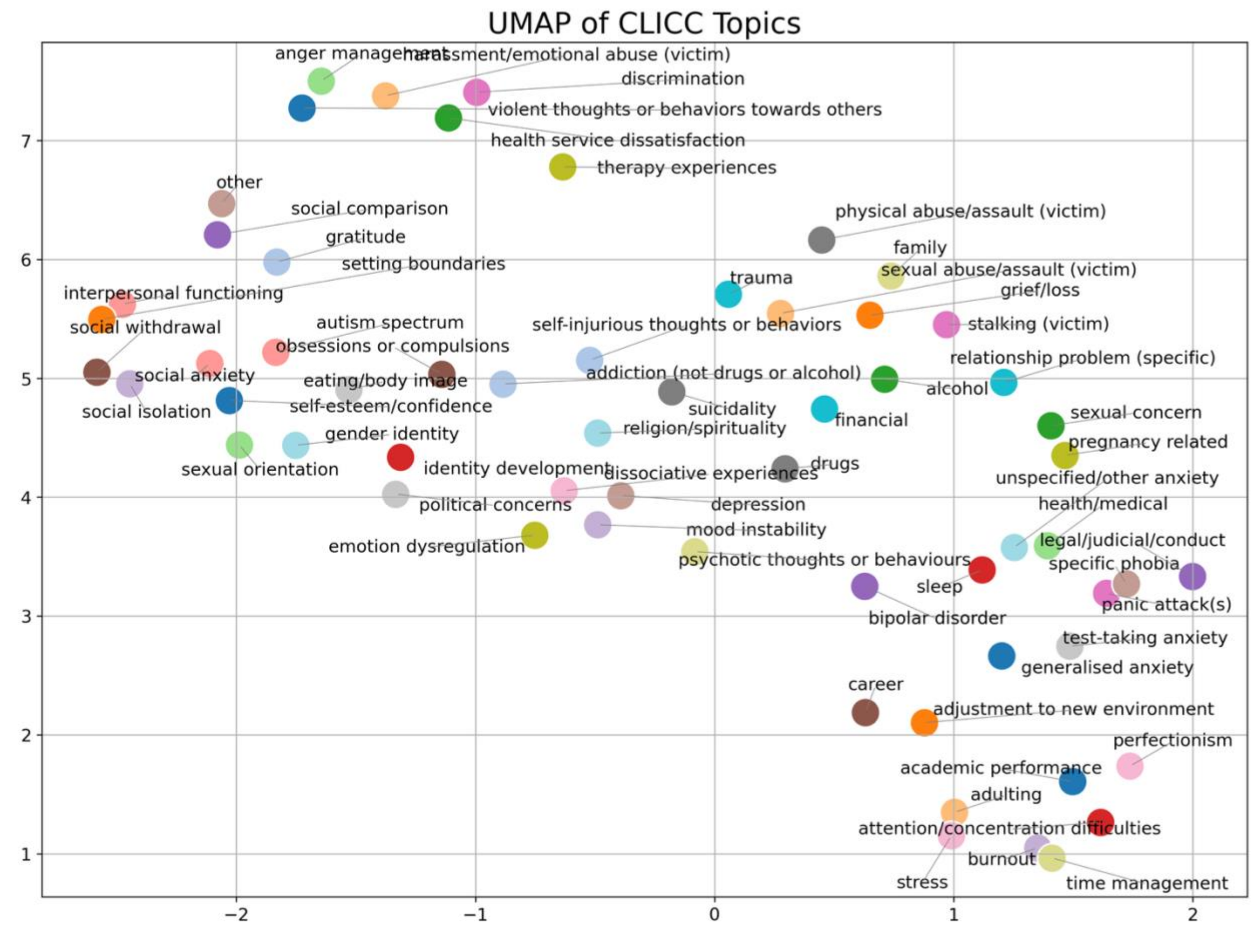


Figure A6: Two-dimensional UMAP projection of mental health topics generated via LLM-extracted CLICC topics.

Figure A6 illustrates a two-dimensional UMAP projection of LLM-extracted mental health topics. Academic- and performance-related topics (e.g., "academic performance", "career", "time management") co-located in the lower right quadrant. Interestingly, stress and adulting are positioned closely together, reflecting how youth often perceive the pressures of transitioning to adult responsibilities as strongly tied to feelings of stress. Discussions of aggression, violence, or mistreatment (like "harassment", "discrimination", "anger", and "violent thoughts") tend to appear semantically close to conversations about mental health service dissatisfaction and therapy experiences. In other words, people discussing experiences of abuse or expressing anger might simultaneously talk about their struggles accessing or benefiting from therapy, linking personal aggression and victimization to system-level dissatisfaction.

A11. Statistical test results

TABLE A3: Comparison of Topics between "Hangouts" and "Ask-a-Therapist" using the Chi-square Test.

Topic	Post Count		Adjusted p-value
	Hangouts	**Ask-a-Therapist**	
Social Interaction	73	6	8.18e-28***
Academic Stress	7	81	2.23e-06***
Panic Attacks	118	124	5.94e-06***
Suicidal Thoughts	31	148	3.40e-05***
Relationship Breakup	23	107	8.07e-04***
Mental Wellness Education	100	132	2.31e-02**
Job Anxiety	2	21	3.46e-01
Sleep Problems	4	16	1.00e+00
Eating Disorders	11	22	1.00e+00
Friendship Issues	80	189	1.00e+00
Loneliness	3	6	1.00e+00
Others	5	6	1.00e+00
Anger Management	5	15	1.00e+00
Family Dynamics	51	106	1.00e+00
Therapy and Counseling	5	16	1.00e+00
Marital Issues	2	15	1.00e+00
Stress and Anxiety	126	265	1.00e+00