

A Computational Diachronic Analysis of Gen Z Mental Health Discourse: A Large-scale Reddit Corpus Study from Pre- to Post-COVID

Felix Mao

Rye Country Day School
fy Mao27@gmail.com

Abstract

Generation Z's mental health discourse has been uniquely shaped by digital saturation and the COVID-19 pandemic. This study introduces a large-scale corpus of Gen Z mental health discourse on Reddit, comprising over 3 million posts across 11 subreddits (2017–2025), identified through behavioral cross-posting between mental health and Gen Z-identified communities. Using a hybrid methodology that integrates statistical corpus linguistics with NLP techniques, we conduct diachronic keyness analysis, Transformer-based sentiment tracking and topic modeling to examine lexical, syntactic, and semantic patterns across pre-, during-, and post-COVID periods. Our analysis reveals: (1) ritualized support exchanges more pronounced in Gen Z where highly negative self-disclosure functions as an authenticity signal; (2) a pandemic-induced reframing of existing mental health topics, particularly a rise in physical symptoms, followed by a sustained post-pandemic sentiment decline; and (3) a generational divergence where Gen Z favors abstract, existential concerns, unlike the pragmatic focus of non-Gen Z users. This study contributes a replicable approach for analyzing youth discourse and underscores the importance of culturally and linguistically informed digital mental health interventions, which can support Gen Z's modes of expressing distress rather than pathologizing them.

Keywords: Computational Linguistics, Corpus Analysis, Mental Health Discourse, Sentiment Analysis, Topical Modeling, Generation Z, Social Media, COVID-19

1. Introduction

Generation Z, a commonly identified age group born between the mid-1990s and the early 2010s, have experienced increased mental health issues, such as rising rates of anxiety, depression, self-harm, and loneliness (Twenge, 2020). This generation is also heavily influenced by its digital exposures, as they were growing up with internet and social media an integral part to their daily life from a very young age (Cismariu and Hosu, 2019). Furthermore, Gen Z's adolescence and early adulthood coincided with the COVID-19 pandemic. The resulted global level social isolation and disruption have dramatically intensified their mental health challenges, which also increasing their dependence on digital interactions for support. Thus, comparing to prior generations, these factors have profoundly shaped how Gen Z in adolescence and early adulthood express distress and seek support (Mastroianni, 2024), leading to increased usage of online platforms. Among these, Reddit has emerged as an important shared space for these conversations and community-driven interactions (Thukral et al., 2018). Its structure of topic-specific forums, known as subreddits, creates dedicated communities for mental health, providing spaces for self-disclosure, advice-seeking, and peer-to-peer support, that are an essential window to understand how this generation navigates and

validates its mental health experiences.

While increasing focus and studies documented prevalence and risk factors of social media to Gen Z mental health decline (Tang et al., 2024; Haidt, 2024), few studies examined how Gen Z's own linguistic practices, especially in social media contexts, mediate their experiences of distress and support using a corpus-driven analysis. These existing discourse analysis on mental health is predominantly qualitative and focused on the general population (Achterbergh et al., 2020; Crowe et al., 2015; Peters, 2010), thus there remains a significant gap in corpus driven computational investigation, specifically regarding the Gen Z cohort's linguistic behavior in social media. This gap is particularly critical given that the COVID-19 pandemic uniquely and profoundly disrupted this cohort's social development, making it a primary key to unlock not just what they experience, but how they linguistically construct and collectively navigate these experiences in a post-pandemic world. Computational methods, such as corpus linguistics and natural language processing (NLP), offer the opportunity to analyze millions of words of naturally occurring text (Guetterman et al., 2018), yet their potential regarding this field of research has been relatively underexplored.

To address these gaps, our study aims to collect a large-scale Reddit corpus to examine Gen Z mental health discourse, specifically the cohort's linguis-

tic and behavioral patterns in how they discuss mental health online. Additionally, our study also hopes to research how these patterns have shifted due to the impact of COVID-19, applying diachronic keyness, Transformer-based sentiment analysis and topic modeling to compare pre-, during-, and post-COVID periods, as well as create a comparison between Gen Z and non-Gen Z users. Leveraging rich linguistic resources and building on recent LLM technologies, we aim not only to detect patterns but also explain the strategies Gen Z uses to frame distress, claim authenticity, and negotiate support, offering new insights into how digitally native youth navigate well-being discursively in a rapidly evolving digital landscape. This paper is organized as follows. Section 2 will talk about the literary review, section 3 will explore the methodology, section 4-6 will explore our hypotheses, results, and relevant discussions. Section 7 will summarize the paper.

2. Literary Review

Research on youth mental health consistently shows that each generation’s well-being is shaped by its distinct social, cultural, and economic context (Reinherz et al., 1999). Linguistic expression plays a crucial role in the manifestation and evolution of mental health concerns over time (Collins and Baker, 2023). Corpus studies have well demonstrated the orthographic, grammatical, semantic, lexical and pragmatic change of linguistic expressions through time (Baker, 2017). Social media, as a primary communicative platform widely used by Gen Z, provides vast and authentic records of naturally occurring language (Kern et al., 2016). Within these digital spaces, Gen Z exhibits distinctive linguistic practices, such as brevity, slang, multimodal creativity, and syntactic looseness, that both mark generational identity and coalesce into distinct online communities (Sunday, 2025; Mitul, 2024). Despite this, most sociolinguistic research on social media has emphasized themes such as identity, humor, or group belonging, often through small-scale qualitative studies (Langman and Sayer, 2012). Large-scale, corpus-based studies of Gen Z’s mental health discourse remain limited.

In parallel, computational linguistics and NLP research have been increasingly utilized to study mental health through language data. Surveys and empirical studies have identified a range of linguistic and paralinguistic features associated with depression and anxiety, including elevated first-person pronoun usage, frequent negative affect terms, and topic clusters revealed through pre-trained emotion embeddings (Garg, 2023; Zhang et al., 2022). Computational methods have also scaled discourse analysis to detect patterns at population level. For instance, identifying thematic cor-

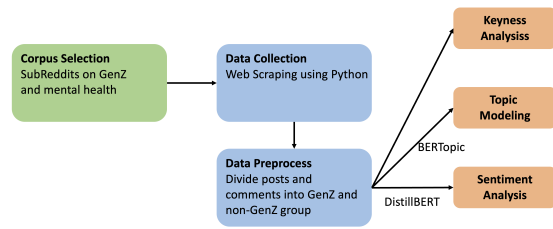


Figure 1: Overview Flowchart of This Study

relations of suicidal ideation during the pandemic (Kumar et al., 2022). Some recent research employed large language models (LLMs) to improve interpretability and domain-specific understanding in mental health prediction (Yang et al., 2024). However, most of these studies have challenges in domain adaptation and tend to overlook generational variation, often treating youth language as homogeneous rather than accounting for the distinct communicative norms of Gen Z.

The COVID-19 pandemic and its impact highlighted the relevance of studying Gen Z’s mental health in digital contexts. Numerous studies have documented the pandemic’s psychological impact on adolescents and students worldwide (Marciano et al., 2022; Chen and Lucock, 2022), highlighting increased levels of stress, anxiety, and loneliness. The resulting surge in online interaction has made social media discourse an even more crucial space for expressing and managing well-being. While prior research has noted the general prevalence and risk factors linking social media use to mental health decline (Tang et al., 2024; Haidt, 2024), relatively few studies have systematically compared linguistic patterns across pre-, during-, and post-pandemic periods, especially using corpus-driven approaches.

3. Methodology

An outline of our methodology is shown in Figure 1. It consists of three key components: corpus selection, corpus construction, and analytical framework. More details of our corpus of mental health support spaces are shown in Table 1.

3.1. Data Collection and Corpus Construction

To capture the key findings of Gen Z mental health discourse, we constructed a corpus¹ of user comments and posts from 11 selected subReddits from two main categories: mental health support

¹The dataset used in this study is publicly available at <https://github.com/FMB678/GenZ-Mental-Health-Reddit-Corpus>.

Subreddit	GenZ Users			Submissions		Comments	
	Number	AvgWords	AvgPosts	#Word	AvgWord	#Word	AvgWord
r/depression	158669	81.48	7.04	195537	143.58	919371	44.05
r/Anxiety	84525	73.52	8.51	82190	154.54	634038	47.16
r/mentalhealth	92132	104.65	6.99	108350	183.56	534471	50.87
r/socialanxiety	49741	55.94	8.06	34061	150.49	362309	41.12
r/bipolar	25239	51.81	10.26	36006	121.55	465362	43.89
r/bipolar2	7716	58.71	23.90	9401	152.17	173857	49.01

Table 1: Corpus Dataset Overview

spaces (as shown in Table 1) and Gen Z identity forums (r/Teenagers, r/GenZ, r/teenagersnew, r/TeenagersButBetter, and r/teenagersbuthot) reflecting both online mental health discourse of broader population and those specific of Gen Z. This dual-category design reflects both the online mental health discourse of the broader population and that which is specific to Gen Z. Crucially, it enabled us to extract "Gen Z" discourse using a behavioral inference method based on community participation. Following established approaches in computational social science (Zhang et al., 2021), we identified users active in both categories as a proxy for Gen Z membership, under the rationale that sustained engagement in age-specific identity forums is a reliable indicator of cohort affiliation. These users were designated as the Gen Z cohort, while those active only in mental health subreddits served as the non-Gen Z control group.

In total, the corpus consists of more than 3 million posts and comments from over 320 thousand unique users from 2017-2025. This data time frame was selected to enable comparison across three key phases: pre-pandemic (2017-2019), pandemic (2020-2022), and post pandemic (2023-2025), allowing for the assessment of the impact of COVID-19 on Gen Z mental health discourse.

3.2. Analytical Framework

To address our hypothesis, we combined multiple computational methodologies including corpus linguistics techniques and NLP approaches. Specifically, we applied keyness analysis, topic modeling, and sentiment analysis to conduct a comprehensive examination of Gen Z mental health discourse. Keyness analysis was used to identify statistically overrepresented lexical items among our corpus, highlighting thematic and sentiment trends, as well as how they have shifted over time. To study keyness, we programmed to create word clouds to display a list of words occurring in the corpus, allowing for a comparison of keywords. Furthermore, we employed Transformer-based topic modeling using BERTopic (Grootendorst, 2022), a powerful topic modeling tool using Transformers and c-TF-IDF to generate interpretable topics from text

with high accuracy and flexibility, to uncover overall thematic structures in the corpus. We implemented BERTopic using its official Python package, with `all-MiniLM-L6-v2` as the sentence-transformer embedding model, to uncover latent thematic structures in the corpus. Finally, sentiment analysis was conducted using DistilBERT (Sanh et al., 2019), a distilled, efficient variant of BERT optimized for general-purpose language representation. We utilized the `distilbert-base-uncased-finetuned-sst-2-english` model from Hugging Face² to classify the affective tone of posts and comments, enabling large-scale tracking of emotional valence across time and user groups. For validation purposes, we also experimented with traditional approaches including LDA topic modeling and VADER sentiment analysis, which produced comparable patterns; however, due to space constraints, we report only the Transformer-based results here.

4. Hypothesis Development

This study employs a computational linguistic approach to analyze mental health discourse on Reddit, focusing on the distinct communication patterns of Generation Z. Three sets of hypotheses were developed to examine how this cohort expresses mental stress, how these expressions were reshaped by the COVID-19 pandemic, and how they differ linguistically from older generations.

4.1. Gen Z mental health online discourse

With the increasing integration of social media into daily interaction, prior research suggests that self-disclosure and authentic emotional expression are central mechanisms in online communication and support exchange (De Choudhury and De, 2014). Within the specific context of online platforms such as Reddit, this often creates a clear discursive division: posts initiating a topic are typically framed through personal vulnerability, while replies are

²<https://huggingface.co/distilbert/distilbert-base-uncased-finetuned-sst-2-english>

Subreddit	Authenticity		Distress		First-person		Sentiment	
	lvsR	GvN	lvsR	GvN	lvsR	GvN	lvsR	GvN
Anxiety	221.92***	13.62	225.78***	-8.24	221.48***	13.77	-157.25***	7.71***
depression	293.89***	3.74	65.82***	-2.53	292.18***	3.63	-222.05***	6.35***
mentalhealth	257.31***	-0.07	192.35***	-16.14	254.85***	-1.68	-143.71***	59.78***
Socialanxiety	138.62***	8.63	108.95***	-4.96	138.12***	8.20	-41.26***	11.37***
bipolar	121.63***	4.66	81.79***	-6.33	120.27***	4.32	-53.79***	6.59***
bipolar2	63.82***	6.02	41.83***	-2.56	62.88***	5.80	-26.09***	5.41***

Table 3: Hypothesis 1 Analysis: T-statistics for linguistic differences. lvsR: GenZ initiators vs respondents. GvN: GenZ vs Non-GenZ difference. ***p < 0.001. Positive lvsR values indicate higher initiator scores; negative lvsR sentiment reflects more negative emotion. Positive GvN values indicate phenomenon is stronger in GenZ.

$$auth = fpp * 0.4 + self_ref * 0.3 + cog * 0.3$$

$$sent = sent * (-0.4) + neg * 0.3 + anxiety * 0.3$$

The weighting was selected to balance various measures and their expressiveness in terms of Authenticity or Distress, for example, direct reference to first-person pronouns tends to show more keenly self-expression subconsciously, thus increase the authenticity, where self-reference phrases or Cognitive process words also improve such connection, though to a less degree (Garg, 2023).

Results. As shown in Table 3, Hypothesis 1 received strong but nuanced support across our analyses. For **H1a**, we found consistent and highly significant linguistic differences between GenZ initiators and respondents across all mental health subreddits. Authenticity scores showed the largest effects ($t = 63.82-293.89$), followed by first-person usage ($t = 62.88-292.18$), distress ($t = 41.83-225.78$), and negative sentiment ($t = -26.09$ to -222.05). The depression subreddit exhibited the strongest initiator-respondent differences across all dimensions, while bipolar2 showed the smallest but still highly significant effects.

For **H1b**, examining generational differences revealed a more complex pattern. While sentiment polarity divergence between initiators and respondents was significantly more pronounced in Gen Z communities across all subreddits, differences in authenticity, distress, and first-person usage showed no statistical significance between generations. This pattern was remarkably consistent, with sentiment showing strong GenZ effects ($t = 5.41-59.78$) while other dimensions remained non-significant. Notably, the mentalhealth subreddit displayed the largest generational difference in sentiment ($t = 59.78$), suggesting this platform exhibits the strongest GenZ-specific emotional polarization.

Therefore, **H1a** receives strong support with consistent initiator-respondent differences across all linguistic dimensions, while **H1b** receives partial support: GenZ demonstrates uniquely strong emotional polarization in support exchanges, but patterns of authentic self-disclosure and distress expression appear consistent across generations.

Category	Top 10 Key Words
Abstract	purpose, meaning, life death, reality, don't
Existential	know, why, feeling, thought, empty
Physical Symptom	body, heart, headache, pain, hurt, dizzy, nausea, fatigue, exhaust, panic
Realistic Trouble	college, money, family, work, friend, home, living, deadline, project, responsibility

Table 4: Topic Category Definition

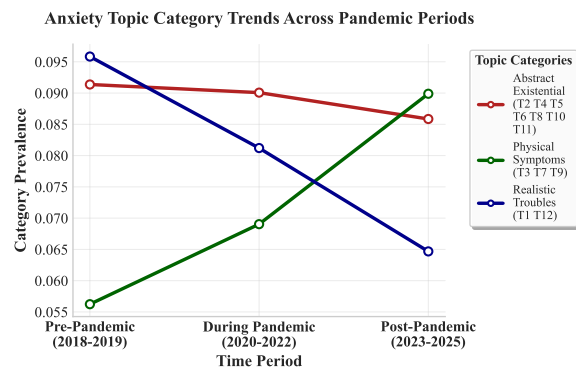


Figure 3: H2a Topic Trends for r/Anxiety

5.2. Change in Discussion of Impact Across the COVID-19 Pandemic Period

Feature engineering: To examine topic reorganization (**H2a**), we applied BERTopic topic modeling with top 12 components to the corpus segmented into three periods: pre-pandemic (2017-2019), during-pandemic (2020-2022), and post-pandemic (2023-2025). A topic modeling results can be found in Figure 2. The resulting topics were qualitatively categorized into 3 thematic groups based on keyword analysis: Abstract Existential, Physical Symptoms, and Realistic Troubles (as shown in Table 4). Topic prevalence was tracked across periods based on the categorized results. For **H2b**, we computed sentiment scores for all posts and aggregated by period.

Results: For **H2a**, our analysis revealed that

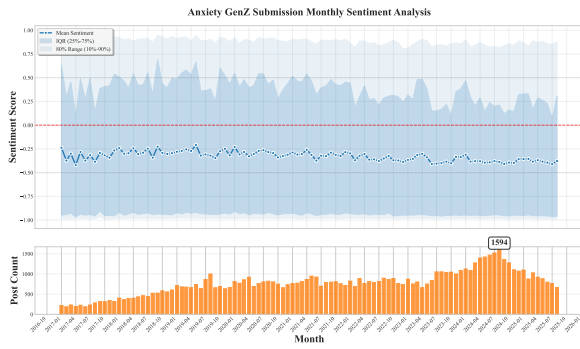


Figure 4: Temporal Sentiment Analysis for Gen Z in r/Anxiety from pre- to post-COVID

Subreddit	Pre vs During	During vs Post
Anxiety	4.41***	9.98**
depression	7.93***	8.41***
mentalhealth	6.24***	-3.87*
socialanxiety	0.44	1.90
bipolar	0.09**	4.57
bipolar2	-0.28	2.21

Table 5: Hypothesis 2 Analysis: T-statistics for sentiment changes across pandemic periods. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Positive values indicate increased scores in the later period.

while the pandemic did not fundamentally transform the thematic structure of GenZ mental health discourse, it did induce significant topic reorganization. As shown in Figure 3, three distinct patterns emerged across topic categories:

- **Abstract Existential** topics remained largely stable throughout the period. Abstract Existential is seen as most stable, with modest reduction during pandemic but largely flat post-Pandemic, demonstrating resilience in how Gen Z engages with fundamental questions of identity and meaning.
- **Realistic Troubles** topics were barely impacted during pandemic, but seem some notable drop post-pandemic (~30%), potentially indicated some delayed effect of the pandemic impact.
- **Physical Symptoms** topics show a consistent and substantial increase (~80%) across all three periods, indicating a growing somatic focus in mental health discussions.

This reorganization without structural transformation confirms **H2a**, demonstrating the pandemic's role as a redistributive rather than transformative force in mental health discourse.

For **H2b**, as shown in Table 5, sentiment analysis revealed significant pandemic-related emotional shifts across mental health communities. The

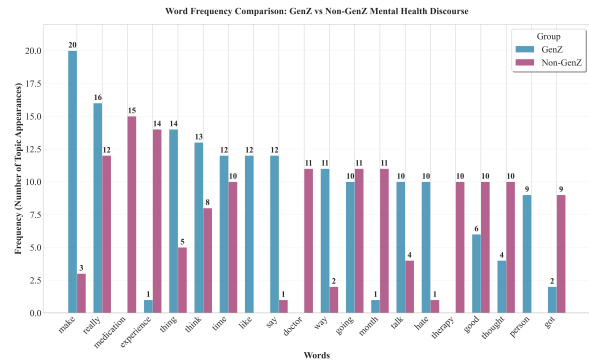


Figure 5: Gen Z vs. non-Gen Z Topic Word Frequency Comparison

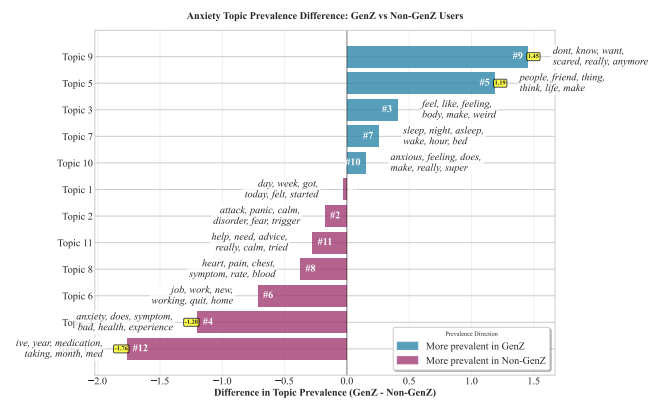


Figure 6: Gen Z vs. non-Gen Z Topic Prevalence Difference in r/Anxiety

depression and anxiety subreddits exhibited the strongest and most consistent negative trends, with highly significant sentiment declines in both pre-to-during ($t=4.47-7.93$, $p < 0.001$) and during-to-post comparisons ($t=8.41-9.98$, $p < 0.01$). However, the mental health community showed a significant initial decline during the pandemic ($t=6.24$, $p < 0.001$) followed by partial recovery post-pandemic ($t=-3.87$, $p < 0.05$), suggesting adaptive resilience. In contrast, socialanxiety and bipolar2 communities demonstrated no significant sentiment changes, indicating their emotional responses to the pandemic were less pronounced than in communities dealing with more severe conditions like depression and anxiety.

These findings provide strong but nuanced support for **H2b**, revealing that while the pandemic corresponded with significant negative sentiment shifts in most mental health communities, the magnitude and persistence of these effects varied substantially across different mental health conditions.

5.3. Gen Z Versus Non-Gen Z Discourse Characteristics

Feature engineering: similarly to analysis used in 5.2, we applied topic modeling with 12 components to the corpus across 3 periods, and conducted a keyness analysis to identify lexical items and topic clusters that were statistically overrepresented in each generational group. To quantify group-level differences, we calculated topic prevalence for each generational cohort by assigning each post to its most probable topic based on the BERTopic model's posterior probability distribution, then aggregating the count of posts assigned to each topic per group. This approach allowed us to compare the relative frequency with which Gen Z and non-Gen Z users engaged with specific thematic content across the corpus.

Results: Our analysis revealed systematic generational differences from keyness analysis as shown in Figure 5: GenZ vocabulary emphasizes personal feelings ("make", "think", "like", "hate") while non-GenZ language centers on clinical and practical terms ("medication", "therapy", "doctor", "experience",). Supporting evidence from topic comparison further demonstrates this divergence. Take r/Anxiety for example (as shown in Figure 6, GenZ discourse is significantly overrepresented in topics characterized by existential and interpersonal concerns (Topic 9 & 5). Conversely, non-GenZ discourse shows stronger prevalence in topics focused on practical management and clinical experiences (Topic 4 & 11).

These findings strongly support **H3**, demonstrating that generational identity systematically shapes both the thematic focus and linguistic framing of mental health discourse, with GenZ emphasizing internal emotional experiences and non-GenZ focusing on practical management and clinical intervention.

6. Discussion

Our analysis reveals three interconnected patterns in how Generation Z articulates mental health concerns online. The examples in Table 6, all drawn from r/Anxiety posts, illustrate how digital nativity, pandemic disruption, and developmental life stage collectively shape a distinct communicative ecology.

The examples demonstrate the high-authenticity, negatively-valenced initiator discourse identified in **H1**. GenZ initiators consistently employ intensive first-person focus ("I," "me," "my") and emotional vulnerability ("paranoid," "insecure," "scared") that characterizes their initiating posts. This pattern aligns with our quantitative findings showing significantly higher authenticity scores and negative

sentiment in GenZ initiators compared to respondents and non-Gen Z. The raw emotional exposure in these posts, such as the 16-year-old's detailed social anxiety description, exemplifies the authentic distress expression that distinguishes GenZ initiator discourse.

H2 shows that COVID-19 acted as a psychological intensifier rather than thematic disruptor for GenZ. The pandemic-era example shows how health crises amplified existing anxieties about social connection and identity, while the post-pandemic example demonstrates recovery and adaptation. This pattern reflects H2's finding that sentiment declines were most pronounced in mental health related communities.

For **H3**, we observe marked generational differences in how mental health is conceptualized. GenZ discourse centers on internal states and identity formation, characterized by deep self-reflection and social anxiety. This contrasts sharply with non-GenZ initiator discourse, which emphasizes practical management and clinical intervention. Notably, while both generations show authentic initiator expression, the content and focus of that authenticity differs fundamentally: GenZ initiators disclose emotional and existential struggles, while non-GenZ initiators share systematic treatment experiences.

These patterns collectively reveal how life stage shapes mental health discourse. GenZ's intensive first-person focus reflects both developmental identity exploration and the platform's function as a space for working through contemporary digital-age challenges. The high engagement and reciprocal support observed in GenZ communities suggest these platforms serve as crucial developmental sanctuary where identity work, social connection, and emotional support intersect, representing a significant evolution from previous generations' more instrumental use of support communities.

The consistency across hypotheses demonstrates that GenZ's mental health discourse is not merely a translation of traditional concerns into digital spaces, but represents a fundamental shift in their mental health expression, shaped by their unique developmental position at the intersection of digital immersion and global crisis.

7. Conclusion

This study presents a corpus-based, computational linguistics review of Gen Z mental-health discourse on Reddit from 2017 to 2025. By integrating linguistic feature analysis, NPL sentiment analysis, and topic modeling across multiple subreddits, we closely examine how youth construct and respond to mental distress through the online interaction.

There are clear indications that COVID impact on various focus groups was nuanced and also

Topic	Generation	Example Excerpt	Date
Social and Existential	GenZ	"I feel like this pandemic has made me forget how to make friends... I get SO paranoid and insecure... I constantly think everyone will hate me..." [16-year-old discussing social anxiety]	2020-10
	Non-GenZ	"I thought about suicide as a way to relieve the uncomfot and panic I get from interacting with people... I have no idea how to summarize this..." [Adult reflecting on long-term struggles]	2017-04
Pandemic Impact	GenZ	"Anyone else feel like this pandemic has made them forget how to make friends?... Zoom college as a first year undergrad, with no friends..." [Direct pandemic impact]	2020-10
	Non-GenZ	"I grew up as a shy quiet person and take forever to let my guard down. I moved to CA from NC during Covid which was my junior year, and it was hard for me to make friends when everyone already had pre established friendships. I thrived in my freshman and sophomore year..." [Post-pandemic recovery]	2023-01
Medical Treatment	GenZ	"I've been having success with GABA supplements... I've become close with after a year of therapy and medication... I've had ups and downs..." [Young adult discussing treatment journey]	2023-02
	Non-GenZ	"I've dealt with generalized anxiety for years... I've spent 4-8 weeks on the below medicines... I've not found relief. Is this normal?" [Detailed medication history]	2022-02

Table 6: Representative Examples from r/Anxiety Illustrating Generational and Pandemic-Era Discourse Patterns in Mental Health Communities

differed based on their context and focusing topics. However we did observe several consequential themes: Gen Z initiators likely would express authentic, more negatively valenced emotion, while respondents more often used positive, supportive language; and compared with non-Gen Z, they tended to be more positive in response showing the support in the online communication. Furthermore, through examination of the impacts of COVID-19 pandemic, we haven't found overwhelming evidence that pandemic had become a topic materially impacting GenZ in general, however, the increased topic emphasis toward physical aspects was observed. Lastly, when conducting the generational comparison, the results demonstrated a clear trend that Gen Z likely focused more on topics about abstract and existential framing, centered on emotion, reflection, and meaning, while non-Gen Z users emphasized practical experiences of distress. These findings demonstrate that Gen Z's mental health discourse is deeply shaped by the increase of digital use and the sociohistorical context.

From a methodological perspective, this study illustrates the benefits of bridging social media corpus-based computational linguistics with mental-health discourse analysis, and laid out a reproducible approach for future interdisciplinary research connecting language, psychology, and social science to better understand how emerging generations articulate distress and in the online community. Future study can introduce a qualitative discourse analysis with embedding-based se-

mantic modeling to enhance interpretability, potentially extending to multilingual and cross-platform corpora (e.g., TikTok, Discord).

8. Bibliographical References

- L. Achterbergh, A. Pitman, M. Birken, et al. 2020. [The experience of loneliness among young people with depression: a qualitative meta-synthesis of the literature](#). *BMC Psychiatry*, 20:415.
- P. Baker. 2017. *American and British English: Divided by a common language?* Cambridge University Press.
- L.W. Barsalou, A. Santos, W.K. Simmons, and C.D. Wilson. 2008. Language and simulation in conceptual processing. In *Symbols, embodiment, and meaning*, pages 245–283. Oxford University Press.
- D. Biber and J. Egbert. 2018. *Register Variation on the Web*. Cambridge University Press.
- M. Brysbaert, E. Keuleers, and P. Mandera. 2014. A plea for more interactions between psycholinguistics and natural language processing research. *Computational Linguistics in the Netherlands Journal*, 4:209–222.
- T. Chen and M. Lucock. 2022. The mental health of university students during the covid-19 pan-

- demic: An online survey in the uk. *PLoS one*, 17(1):e0262562.
- C.D. Cismariu and I. Hosu. 2019. Generation z and social media. In *Proceedings of the 14th International Conference on Virtual Learning*.
- L. Collins and P. Baker. 2023. *Language, discourse and anxiety*. Cambridge University Press Cambridge.
- M. Crowe, M. Inder, and R. Porter. 2015. [Conducting qualitative research in mental health: Thematic and content analyses](#). *Australian & New Zealand Journal of Psychiatry*, 49(7):616–623.
- M. De Choudhury and S. De. 2014. Mental health discourse on Reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, pages 71–80.
- M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar. 2016. Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 2098–2110.
- M. Garg. 2023. Mental health analysis in social media posts: a survey. *Archives of Computational Methods in Engineering*, 30(3):1819–1840.
- M. Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.
- T.C. Guetterman, T. Chang, M. DeJonckheere, T. Basu, E. Scruggs, and V.V. Vydiswaran. 2018. [Augmenting qualitative text analysis with natural language processing: methodological study](#). *Journal of Medical Internet Research*, 20(6):e231.
- J. Haidt. 2024. *The Anxious Generation: How the Great Rewiring of Childhood Is Causing an Epidemic of Mental Illness*. Penguin Press.
- M.L. Kern, J.C. Eichstaedt, H.A. Schwartz, L. Dzurzynski, L.H. Ungar, D.J. Stillwell, M. Kosinski, S.M. Ramones, and M.E. Seligman. 2016. The online social self: An open vocabulary approach to personality. *Assessment*, 21(2):158–169.
- S.F. Kiesling. 2019. *Language, Gender, and Sexuality: An Introduction*. Routledge.
- R. Kumar, S. Mukherjee, T.M. Choi, and L. Dhamotharan. 2022. Mining voices from self-expressed messages on social-media: Diagnostics of mental distress during COVID-19. *Decision Support Systems*, 162:113792.
- J. Langman and P. Sayer. 2012. Qualitative sociolinguistics research. In *The Encyclopedia of Applied Linguistics*. Wiley-Blackwell.
- L. Marciano, M. Ostroumova, P.J. Schulz, and A.L. Camerini. 2022. Digital media use and adolescents’ mental health during the COVID-19 pandemic: a systematic review and meta-analysis. *Frontiers in Public Health*, 9:793868.
- C. Mastroianni. 2024. *Social Support in the Time of Social Media: Predicting Gen Zers’ Preferences for In-Person and Virtual Support*. Ph.D. thesis, Seton Hall University.
- R.T. Mitul. 2024. Social media language of Gen Z and perceptions of millennials. Master’s thesis, Brac University.
- S. Peters. 2010. Qualitative research methods in mental health. *Evidence Based Mental Health*, 13(2):35–40.
- H.Z. Reinherz, R.M. Giaconia, A.M.C. Hauf, M.S. Wasserman, and A.B. Silverman. 1999. Major depression in the transition to adulthood: Risks and impairments. In *Historical and Geographical Influences on Psychopathology*, pages 141–161. Lawrence Erlbaum Associates.
- V. Sanh, L. Debut, J. Chaumond, and T. Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.
- Y. Shen, T. Lu, J. Chang, and X. Zhang. 2025. Digitalized social infrastructure and mental health during the COVID-19 crisis: Evidence from communities in shanghai. *Cities*, 159:105745.
- E.V. Sunday. 2025. Language variation: A study of Gen Z language on selected social media platforms. *Nigerian Journal of Arts and Humanities*, 5(1):45–62.
- K. Tang, C. Hilton, R. Yunita, and R. Walvinson. 2024. Association between social media and mental health changes among Gen-Z in batam, indonesia: A cross-sectional study. *Journal of Community Mental Health and Public Policy*, 7(1):12–25.
- S. Thukral, H. Meisheri, T. Kataria, A. Agarwal, I. Verma, A. Chatterjee, and L. Dey. 2018. [Analyzing behavioral trends in community driven discussion platforms like Reddit](#). In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 662–669.
- J.M. Twenge. 2020. Increases in depression, self-harm, and suicide among US adolescents after 2012 and links to technology use: possible

mechanisms. *Psychiatric Research and Clinical Practice*, 2(1):19–25.

K. Yang, T. Zhang, Z. Kuang, Q. Xie, J. Huang, and S. Ananiadou. 2024. MentaLLaMA: Interpretable mental health analysis on social media with large language models. In *Proceedings of the ACM Web Conference 2024*, pages 125–136.

S. Zhang, M. Liu, Y. Li, and J.E. Chung. 2021. Teens' social media engagement during the covid-19 pandemic: a time series examination of posting and emotion on reddit. *International Journal of Environmental Research and Public Health*, 18(19):10079.

T. Zhang, A.M. Schoene, S. Ji, and S. Ananiadou. 2022. Natural language processing applied to mental illness detection: a narrative review. *NPJ Digital Medicine*, 5(1):46.