

**“Broken Bones Heal Stronger” - Exploring Differences in the Discourse in CPTSD and
CPTSD Recovery Communities on Reddit using Large Language Models (LLMs)**

by

Selin Koşar (s2731908)

Department of Psychology, University of Twente

1st supervisor: Dr. Peter ten Klooster

2nd supervisor: Dr. Jorge Piano Simões

16th of July, 2025

Wordcount: 6863 words

Table of Contents

Abstract	4
“Broken Bones Heal Stronger” - Exploring Differences in the Discourse in CPTSD and CPTSD Recovery Communities on Reddit using Large Language Models (LLMs)	5
CPTSD and Validation in Online Support Communities	5
r/CPTSD and r/CPTSDNextSteps	7
Understanding CPTSD Validation and Recovery Communities	7
The CHIME Recovery Model and CPTSD Online Mental Health Communities.....	8
Previous Research on Trauma-Related Online Mental Health Communities	9
Text Mining	10
Present study.....	11
Methods	12
Data Collection	13
Data Preprocessing	14
GPT Text Mining Analyses.....	14
GPT-4o Sentiment and CHIME Analysis	15
Validation of GPT Scores against Human Ratings	18
Comparative Analysis of Sentiments and CHIME Model Alignment between Subreddits .	19
Results	19
Validation of GPT 4o-generated ratings against human ratings	20
Comparison of GPT-4o-based Sentiment and CHIME Scores Between r/CPTSD and r/CPTSDNextSteps.....	24

Sentiment Analysis	25
CHIME Components Analysis	26
Discussion	27
Interpreting the findings	28
Reliability of GPT-4o in Text Analysis.....	28
Alignment of the recovery subreddit with the CHIME Model.....	29
Sentiment and Emotional Tone.....	29
Strengths, Limitations & Future Research	29
Implications	31
Conclusion.....	32
References	33
Appendix A	41
Appendix B	42
Appendix C	43
Appendix D	45
Appendix E	47

Abstract

Objective: The present study investigated discourse in online mental health communities (OMHCs) for Complex Post-Traumatic Stress Disorder (CPTSD), comparing the validation-oriented subreddits r/CPTSD and recovery-oriented r/CPTSDNextSteps. Using the large language model (LLM) GPT-4o as a deductive text mining tool, it evaluated the alignment of popular posts with the Connectedness, Hope, Identity, Meaning, Empowerment (CHIME) model of personal recovery and assessed emotional tone.

Methods: A total of 991 posts from r/CPTSD and 1001 posts from r/CPTSDNextSteps were scraped on April 4, 2025. Zero-shot GPT-4o scores (1-7 Likert scale) were validated against independent human raters with intraclass correlation coefficients (ICC(2,1)). Comparative analyses tested differences in GPT-4o sentiment and CHIME scores across subreddits.

Results: r/CPTSDNextSteps posts scored significantly higher in sentiment ($p < .001$, $d = 1.06$), Hope ($p < .001$, $d = 1.68$) and Empowerment ($p < .001$, $d = 1.38$), while r/CPTSD posts scored higher on Meaning ($p < .001$, $d = -0.46$). GPT-4o showed good to excellent agreement with human ratings (ICCs ≥ 0.60) for sentiment, Connectedness, Hope and Empowerment across both subreddits, but only poor to fair agreement for Identity and Meaning in r/CPTSD.

Conclusion: The recovery-oriented r/CPTSDNextSteps indeed reflects more positive, hope-driven, and empowering discourse, while r/CPTSD emphasizes meaning. GPT-4o appears effective for analysing psychological constructs in OMHC texts, with limitations for more complex constructs. Findings highlight the distinct roles of OMHCs in CPTSD recovery and the potential of LLMs for analysing large-scale OMHC data.

Keywords: Complex Post-Traumatic Stress Disorder, online mental health communities, CHIME model, sentiment analysis, text mining, Reddit, recovery

“Broken Bones Heal Stronger” - Exploring Differences in the Discourse in CPTSD and CPTSD Recovery Communities on Reddit using Large Language Models (LLMs)

Life experiences play a critical role in shaping psychological development and well-being. When these experiences involve highly distressing situations that surpass the individual's ability to manage them, they can result in enduring and severe consequences. Complex Post-Traumatic Stress Disorder (CPTSD) is a fairly recent concept in psychological treatment and research. It is characterized as a mental health condition that arises in response to exposure to repeated trauma that often consist of interpersonal threats, also known as “complex trauma” (Herman, 1992). CPTSD is characterized by a set of symptoms such as emotional dysregulation, a negative self-concept and complications in interpersonal relationships. These symptoms occur alongside the core symptoms of Post-Traumatic Stress Disorder (PTSD), which include re-experiencing, avoidance, and hyperarousal (World Health Organization, 2025). Unlike single-event trauma in PTSD, CPTSD emerges from sustained trauma, such as childhood abuse or war captivity (Cloitre et al., 2013). While not recognized in the DSM-5, CPTSD is classified in the ICD-11 (American Psychiatric Association, 2013; World Health Organization, 2025). People with CPTSD can experience severe adverse long-term effects, impaired functioning and a significantly reduced quality of life (Maercker et al., 2022), especially when the symptoms remain undiagnosed or untreated.

CPTSD and Validation in Online Support Communities

Despite a growing awareness of CPTSD as a distinct clinical entity, recovery from CPTSD is hindered by barriers in help-seeking, such as stigma and provider knowledge (Blackie et al., 2024; Stadtmann et al., 2018). A key reason individuals with CPTSD may engage with online social media platforms is for emotional support. One aspect of coping is the need for validation; an affirmation of their emotions, identities and experiences by others (Ford, 2021; Marr et al., 2022), in which digital spaces continue to play an increasingly prominent role. According to a recent study by Wislocki et al. (2024), online mental health

communities (OMHCs) for trauma-related mental health disorders have become critical informal support spaces for survivors of CPTSD to seek peer-to-peer support (Prescott et al., 2020). Additionally, social media data from platforms such as Reddit offer distinct advantages over traditional research methods for studying experiences of people with mental health problems such as self-report surveys and interviews, as they provide a unique and potentially rich source for research in a naturalistic group setting (Wongkoblapp, Vadillo, & Curcin, 2017; Garg, 2023). This aligns with findings from De Choudhury (2013), who highlighted that social media platforms offer a new “lens” in understanding mental illness, as it expands sample sizes and reduces barriers for disclosure.

Much of the research on OMHCs to date has been directed towards social media networks and public online forums such as X (formerly known as Twitter) and Reddit (Naslund et al., 2020). This trend has been amplified by the increased usage of online-based research, particularly during the Covid-19 pandemic, in which researchers “scraped” Reddit threads to analyse its rich, user-generated conversations (Adams, 2022). Reddit, given its size, topic-oriented scope and easy accessibility of communities specifically been shown to be a particularly useful rich data source for studying the discourse on mental health issues (Kim et al., 2023), due to providing access to vast, publicly available, data enabling efficient analysis of user behaviours and sentiments (Adams, 2022). Reddit has hundreds of millions of active monthly visitors and hundreds of thousands of user-created communities with varying topics. These “subreddits” include OMHCs such as “r/ptsd” and “r/CPTSD” for (complex) posttraumatic stress disorder. Such subreddits have been found to align with clinical classifications of disorders. For example, Park et al. (2017) found that members in the PTSD Reddit community discussed several unique themes that were not shared by either the anxiety or depression communities. Additionally, Wislocki et al. (2024) found that venting, resources and coping tools were significantly more often found in r/CPTSD while advice-seeking and

discussions of non-interpersonal trauma and flashbacks were more commonly found in r/ptsd. Despite CPTSD being a relatively newly described clinical phenomenon, the r/CPTSD subreddit has double the number of users compared to r/ptsd. In addition to this community, complementary Reddit spaces, such as r/CPTSDNextSteps, also exist. While r/CPTSD caters mainly to individuals seeking validation and understanding, r/CPTSDNextSteps posits itself as focusing on proactive recovery and growth.

r/CPTSD and r/CPTSDNextSteps

r/CPTSD, with approximately 326 thousand members, defines itself as a “peer support community” (Reddit r/CPTSD, 2025) alongside several rules that posts must adhere to, including that all content must be CPTSD-related. The subreddit includes “CPTSD Victory & Vent” stories about celebrating achievements and expressing negative emotions and “Does Anybody Else (DAE)” posts, where users ask whether their CPTSD-specific experience is shared (Reddit r/CPTSD, 2025). Describing itself as a sister subreddit of r/CPTSD, r/CPTSDNextSteps (approximately 36 thousand members) differentiates itself from r/CPTSD as a space for those in an intermediate stage of recovery. The subreddit’s rules stress that r/CPTSDNextSteps posts must not include “venting, DAE, validation seeking, victory posts, despair-based posts, or “off my chest”-style posts.” The subreddit rules further stress that posts should be focused on recovery (Reddit r/CPTSDNextSteps, 2025).

Understanding CPTSD Validation and Recovery Communities

Understanding the content and dynamics of the postings shared on subreddits such as r/CPTSD and r/CPTSDNextSteps can provide relevant insight for several reasons. First, online CPTSD communities have been growing in their prominence as both informal extensions or substitutes for traditional therapeutic environments (Naslund et al., 2016). Second, online mental health communities (OMHCs) can offer a range of experiences for

users. For example, validation-oriented communities like r/CPTSD provide a space where individuals can share experiences and seek understanding, while recovery-oriented communities such as r/CPTSDNextSteps focus offer its users more structured and positive support, helping members to move towards healthier self-concepts (Corzine & Roy, 2024).

Despite their potential impact, subreddits on CPTSD, including both validation and recovery-focused communities, remain heavily understudied. Prior research on OMHCs has largely been centred on long-established disorders such as depression and PTSD, rather than CPTSD. This study addresses this gap by analysing and comparing the content shared in r/CPTSD, a validation-oriented community, and r/CPTSDNextSteps, a recovery-oriented community, to explore differences in emotional tone and alignment with recovery frameworks. Given the subreddit's distinct focuses, recovery-focused communities such as r/CPTSDNextSteps are expected to exhibit more positive sentiment and higher alignment with theoretical recovery frameworks, while validation-focused communities like r/CPTSD may display more negative sentiment and lower alignment with recovery frameworks due to the subreddit's emphasis on emotional validation and sharing struggles.

The CHIME Recovery Model and CPTSD Online Mental Health Communities

Personal recovery in mental health is a central concept in understanding how individuals navigate psychopathology, involving broader aspects of well-being and personal growth (Slade et al., 2008). It can be defined as a “deeply personal, unique process of changing one’s attitudes, values, feelings, goals, skills and/or roles,” which in turn allows individuals to live a satisfying life despite the limitations of mental illness (Anthony, 1993, p. 15). The CHIME model provides an overarching framework for understanding how individuals engage in recovery narratives. Originally developed as a conceptual model for personal recovery in severely mentally ill individuals, the model presents five components deemed to be essential for personal recovery: Connectedness, Hope and optimism, Identity,

Meaning in life and Empowerment (Leamy et al., 2011). The components of CHIME align with subjectively and self-defined goals, rather than clinical outcomes. For example, Kraiss et al. (2019) found that in bipolar disorder, a condition that bears similarities to CPTSD due to its chronicity and negative impact on mental health, personal recovery components such as hope strongly correlate with well-being.

The CHIME framework offers a structured approach in examining how individuals construct recovery narratives within OMHCs. In a recovery-oriented space such as r/CPTSDNextSteps the components of Connectedness and Identity may be more salient due to the emphasis on mutual support and encouragement. Furthermore, its subreddit rules align with Hope, Meaning and Empowerment, as they focus on agency, proactive recovery strategies and self-improvement, contrasting with the emotional venting found in r/CPTSD. The discussions in the subreddit can thus be expected to contribute to Identity reconstruction as the users shift to an identity that incorporates self-growth and healing. By deductively analysing the subreddits through the lens of the CHIME model, it becomes possible to assess how its narratives align with established dimensions of recovery. This comparison may reveal key distinctions between validation-focused and recovery-oriented spaces in personal recovery narratives.

Previous Research on Trauma-Related Online Mental Health Communities

Previous research into OMHCs has shown that platforms like Reddit are used for psychoeducation, coping strategies and emotional support. In particular, Wislocki et al. (2024) used automated text mining methods to explore thematic content across r/ptsd and r/CPTSD. However, no prior studies have explored how posts in these OMHCs differ in expressed sentiment and align with personal recovery frameworks. To address this gap, this study explores the presence of CHIME elements and expressed sentiments in the discourse in a

validation-oriented community, r/CPTSD, and a recovery-oriented community, r/CPTSDNextSteps, in order to understand their role in supporting CPTSD recovery.

Text Mining

While the amount of user-generated textual content available to the public in Reddit communities presents itself as an opportunity for research on OMHCs, the scale and unstructured nature of social media data poses significant challenges for traditional and manual research methods. Manually processing and analysing unstructured text data is labour-intensive and can limit both the reliability and generalizability and richness of the findings (Gu et al., 2022). One methodology that overcomes these limitations is text mining, a data-driven approach used to automate the process of extracting meaningful patterns and insights from large volumes of unstructured textual data (Hassani et al., 2020; Chegg & Shepherd, 2008). Text-mining leverages algorithms, often underpinned by machine learning, in order to automate topic modeling, sentiment analysis and thematic clustering processes. Yet, traditional topic modelling and sentiment analysis tend to be limited for psychological text analysis due to their inductive nature, as they may yield unfiltered associations or an overwhelming pattern variety (Kodratoff, 2001), and their reliance on predefined lexicons or extensive labelled data for training. Moreover, traditional bag-of-words (BoW) text-mining approaches often struggle to capture the contextual relationships and semantic nuances of language, as they treat words as independent units and lack the ability to understand sequential dependencies or the broader discourse context (Mu et al., 2024).

However, with the advent of generative AI models, computational systems are increasingly able to process and analyse human language effectively. Transformer-based large language models (LLMs) such as GPT and BERT (Bidirectional Encoder Representations from Transformers) can effectively model contextual relationships and semantic nuances (Vaswani et al., 2017). The effectiveness of LLMs stems from their ability to understand

context by processing all preceding, and sometimes surrounding (e.g. BERT), words in a text. This contextual awareness arises from their training on datasets using generative AI techniques, including transformer architectures (Mu et al., 2024; Rathje et al., 2024), allowing them to more accurately interpret complex emotional content. In fact, a recent study by Rathje et al. (2024) demonstrated how GPT outperformed traditional dictionary-based methods for multilanguage psychological text analysis in sentiment analysis and emotion detection. Importantly, Rathje et al. (2024) used simple prompts rather than extensive prompt seeding in ChatGPT, which required no additional training data, making it suitable for analysing subreddits' diverse and informal language. While LLMs such as GPT show promise for mental health applications, their specific use for OMHC texts in deductively analysing psychological constructs related to personal recovery models has not yet been explored. This approach could be particularly suited for processing large volumes of unstructured text data from OMHCs, offering a more comprehensive view of recovery narratives compared to traditional methods.

Present study

In this study, text mining analyses are applied to original postings on r/CPTSD and r/CPTSDNextSteps to explore the discourse about CPTSD in OMHCs. Specifically, this study evaluated the potential of ChatGPT's GPT-4o, a popular and efficient large language model, as a text mining tool for analysing and comparing expressed sentiments and the alignment with the psychological constructs of the CHIME model of personal recovery in both subreddits. As such, this exploratory study aimed to evaluate the distinct recovery-based topics and user sentiments that characterize these support and recovery-oriented communities by addressing the following research questions (RQs):

1. To what extent can GPT-4o accurately detect and estimate sentiment and CHIME model components in Reddit posts, as validated against human raters?

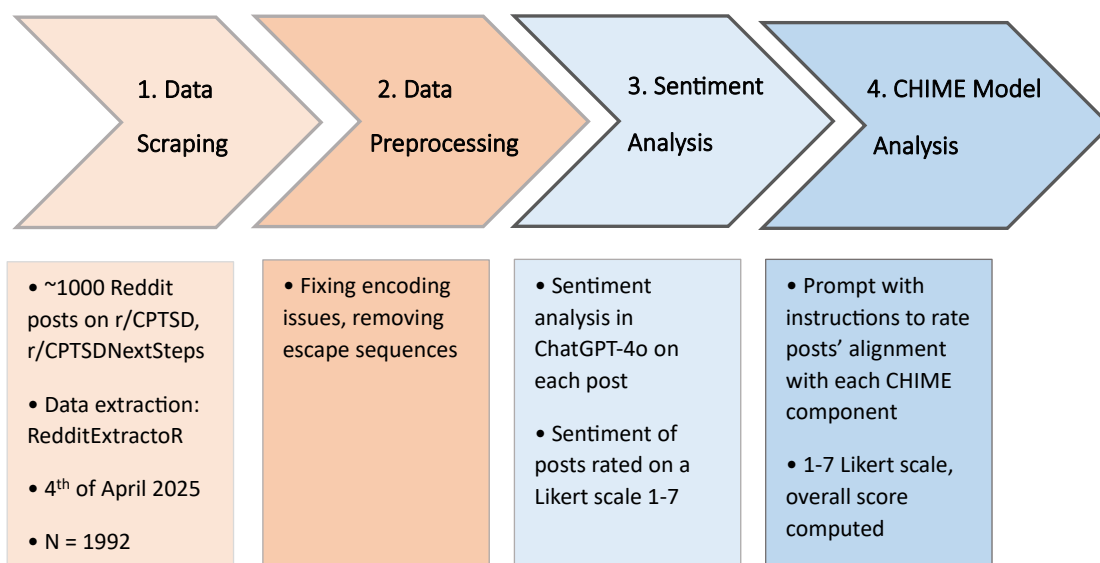
2. How well do posts in r/CPTSD and r/CPTSDNextSteps align with the CHIME model of personal recovery?
3. To what extent do r/CPTSD and r/CPTSDNextSteps differ in their alignment with the CHIME model?
4. What is the average sentiment expressed in posts from r/CPTSD and r/CPTSDNextSteps?
5. How do the sentiment scores on r/CPTSD and r/CPTSDNextSteps differ?

Methods

In order to answer the RQs, original posts from both subreddits were scraped, pre-processed and analysed. The text mining pipeline is visualised in Figure 1, starting with scraping the “top” Reddit posts from each subreddit, after which they were analysed on their sentiment and alignment with the CHIME model components. The study was approved by the Behavioural, Management and Social sciences Ethics Committee (approval number 251602).

Figure 1

Text mining pipeline in the present study.



Note. The study design is illustrated in four successive steps in the pipeline.

Data Collection

991 posts from r/CPTSD and 1001 posts r/CPTSDNextSteps were scraped on the 4th of April, 2025, forming the dataset for this study. The r/CPTSD subreddit, created in 2012, is a complex trauma peer-support community with approximately 326 thousand members as of early 2025 (Reddit r/CPSTD, 2025). r/CPTSDNextSteps, established in 2020, posits itself as an intermediate and advanced stage of healing recovery-oriented complex trauma community with approximately 36 thousand members (Reddit r/CPSTDNextSteps, 2025).

The R package Reddit-Extractor version 3.0.9 was used to extract data from the two subreddits (Rivera, 2023). The Application Programming Interface of Reddit allows to scrape a fixed maximum of approximately 1000 posts per subreddit. Comments on the posts were excluded from data scraping, because of this study's focus on the topics and sentiments present in the spontaneously posted original content. Thus, only original posts were included without any metadata. On Reddit, posts gain points by utilizing "upvotes" and "downvotes", where a higher amount of upvotes by users on a post make it a more popular post. The most popular posts of all time on any subreddit can be found by any user by filtering the posts by "top". Selecting top posts was considered more advantageous than the most recent posts for two reasons. First, top posts have the highest user engagements and visibility and are thus more likely to reflect the views and experiences that resonate with the community. Moreover, top posts are likelier to reflect higher quality content, such as valuable information or personal stories. Regarding representativeness, by selecting "top" posts, the study prioritized content that likely reflect core topics that resonate most with the community members. In the guidebook for text mining in social sciences, Ignatow and Mihalcea (2017) discuss how purposive sampling (such as selecting "top" posts) and prioritizing high-impact content, while potentially sacrificing depth, may be able to capture the dominant themes that resonate with the majority of the community, which aligns with the exploratory research goals.

Data Preprocessing

Unlike traditional Natural Language Processing (NLP) methods, which typically require extensive text preprocessing, this study leveraged GPT-4o's ability to interpret raw and unstructured textual data (Belal et al., 2023) for the sentiment and CHIME model analyses. Thus, preprocessing steps were minimal and limited to ensuring the readability of the text as well as its context. This included replacing certain characters (e.g. “/”, “@”) with spaces, and fixing any encoding issues (e.g., invalid characters such as \n) by converting the text into UTF-8 format by using the tm package in R (Feinerer et al., 2008). Stop words and emojis were retained and no tokenization, lemmatization or stemming was performed. GPT-4o is able to contextualize these naturally, without treating them as distinct tokens (Belal et al., 2023) and providing contextual cues that can improve the accuracy of GPT-4o's sentiment and CHIME model analyses. Posts with no text content or containing fewer than 5 words were removed. Additionally, since the study focuses on textual data analysis, posts with both fewer than 10 words and an image were also excluded, as they may rely heavily on visual context that cannot be interpreted. Posts with more than 10 words were considered to provide sufficient textual context, even if they included an image. Manual inspection confirmed that the raw textual data remained interpretable.

GPT Text Mining Analyses

After the textual data of the two subreddits were prepared, the GPT-4o text analyses were conducted. GPT-4o, developed by OpenAI and accessible through ChatGPT, is a LLM that uses techniques from NLP and that is trained on a large amount of internet data. GPT-4o can execute tasks without much training and performs well on smaller datasets as well as subtle contexts in expressions (e.g. sarcasm) (Belal, 2023). In the present study, GPT-4o was used to conduct two types of text mining analyses; sentiment analysis and CHIME model alignment analysis. The use of LLMs for such analyses is supported by recent studies that

employed LLMs for similar tasks. For instance, Rathje et al. (2024) demonstrated that GPT models outperform traditional methods in multilingual psychological text analysis. Similarly, Belal et al. (2023) highlighted ChatGPT's capability as a text annotation tool for sentiment analysis, emphasizing its efficiency with unstructured data.

GPT-4o Sentiment and CHIME Analysis

Sentiment analysis is a computational technique that is used to determine the emotional tone of a text. A “sentiment” stands for the underlying views (e.g. feelings, opinions) that are communicated in the textual data (Cambria et al., 2013). Traditionally, the process involves assigning scores to words based on predefined lexicons or machine learning methods; posts are tokenized into individual words and they are then matched with the predefined lexicons. Traditional sentiment analysis may, however, misinterpret context, especially in emotionally nuanced discussions (Cambria et al., 2013). Transformer-based LLMs (e.g. GPT-4o) are able to capture sentiment in a more effective and subtle manner due to contextual understanding. Compared to lexicon-based tools such as VADER (Belal et al., 2023), GPT-4o may offer advantages for analysing the vast amount of Reddit data by interpreting its context, emotional nuances and slang without requiring extensive preprocessing (Belal et al., 2023). This is especially relevant given that Reddit-specific expressions (e.g. “DAE”) and complex narratives are common in complex trauma-focused discussions. LLM-based sentiment analysis could thus yield more accurate insight into the differences in the sentiment distribution between the validation-focused and recovery-focused communities.

GPT-4o was instructed from the 4th of April 2025 onwards to rate the sentiment of each post on a Likert scale from 1 to 7, consistent with prior research using similar rating scales to quantify psychological constructs using a standardized simple zero-shot prompt (Rathje et al., 2024) (see Appendix B for the exact prompt), namely to rely on brief

instructions rather than examples in the prompt. To evaluate the alignment of each of the Reddit posts with the CHIME model for personal recovery (Figure 2), GPT-4o was similarly prompted (Appendix B) to assign a CHIME alignment score on a Likert scale from 1 to 7 on each of the five components. No additional training of GPT-4o took place outside of the initial prompts (Appendix A and B) as GPT-4o's outputs demonstrated a (correct) understanding of sentiment and the CHIME model, removing the need for further few-shot or many-shot prompting. All output in ChatGPT was manually inspected in order to ensure that the content was properly processed. Due to ChatGPT token limitations, posts were fed to ChatGPT in batches of 50 at the time. In case of hitting token rate limitations or empty outputs, the post batch of 50 was reduced to 10 to 25, and manual inspection was undertaken again.

Figure 2

CHIME Model Components and Descriptions for Reddit Post Analysis

CHIME Component	Description	Example Quote from Subreddits
Connectedness	The extent to which the post reflects social support, relationships, or community.	Edit: thank you so much for all the responses! It's really comforting that so many people can relate. I'm learning so much from these comments.
Hope	The presence of optimistic statements about recovery and future improvement.	2 years ago I was suicidal and hospitalized for a couple months bc of depression and cptsd. Today I went to the gym for the first time in years ... Life really can get better :) don't you ever give up.

Identity	Discussions related to self-concept, personal growth, or reclaiming one's identity after trauma.	I'm just happy I'm no longer a people pleaser. Now the next step is to relax a little and let people in more and start making new friends again.
Meaning	Expression of purpose, existential reflections, or finding meaning in lived experiences.	We aren't glass. We are metal. ... I was explaining how I don't understand how anybody expects me to heal. ... The only way to get the strongest, most pure metal is to break it and melt it down over ... as many times as necessary ... The end product is significantly stronger than when the metal was in the beginning. This helped me see it differently.
Empowerment	Statements that convey self-efficacy, agency, and personal strength in the recovery process.	I committed to going to climbing with some new people This is the big victory for me. ... I would stay in, hide, deprive myself etc. It's quite a shift for me to not abandon my need for social connection. ... This really helped me to see how committing to a social behaviour when I don't want to can really pay off

Note: Adapted from Leamy et al. (2011)

Validation of GPT Scores against Human Ratings

Automated text analysis methods, such as ChatGPT-based assessments, are increasingly explored for their potential in academic research. However, their use for specific tasks, such as evaluating social media posts for specific psychological constructs, remains novel, necessitating the validation of their accuracy against humans. Thus, 100 random posts from each subreddit were manually scored on the same 7-point rating scales by the author in terms of sentiment and alignment to the CHIME model, and a second human rater independently rated all posts as well. Both raters were (under)graduate students in psychology with prior experience in content analysis. The raters viewed the same prompts given to ChatGPT on the 7-point Likert scale for sentiment and CHIME alignment in order to ensure consistent application of the criteria, after which they independently coded the posts.

Agreement between ratings was assessed using Intraclass Correlation Coefficient (ICCs), which measures the proportion of variance that can be attributed to differences between raters, with values ranging from 0 (no agreement) to 1 (perfect agreement). Specifically, ICC(2,1), a two-way random-effects model for single raters with absolute agreement, was chosen to measure the inter-rater reliability between the two human raters and between GPT-4o and (single and averaged) human ratings of sentiment and CHIME alignment. Agreement values were interpreted according to Cicchetti's (1994) recommendations: $ICC < 0.40$ indicates poor agreement, $0.40-0.59$ indicates fair agreement, $0.60-0.74$ indicates good agreement, and $0.75-1.00$ indicates excellent agreement. In the present study, an ICC of 0.60 or higher was considered indicative of sufficient agreement.

Results were visualized using scatter plots with regression lines, created in R using the *ggplot2* package (Wickham et al., 2024). Additional Bland-Altman plots were generated to

graphically assess the variability between the averaged human ratings and GPT-4o ratings for each psychological construct, and to screen for any systematic or proportional biases (Bland & Altman, 1986).

Comparative Analysis of Sentiments and CHIME Model Alignment between Subreddits

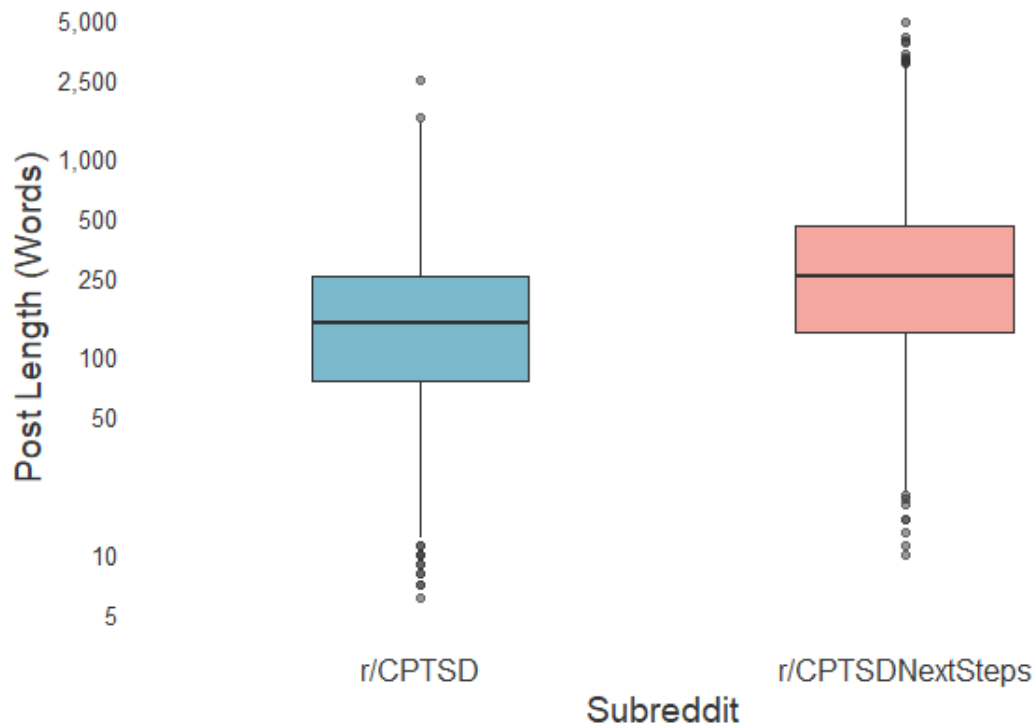
After obtaining GPT-4o sentiment and CHIME alignment scores, a comparative analysis between the two subreddits was conducted with the primary objective of determining whether expected differences exist in the emotional tone and recovery-oriented themes expressed in the two communities. In order to compare sentiment distributions, independent sample t-tests were performed to assess whether the mean GPT-4o-based sentiment scores and the mean scores of each of the five CHIME components (Connectedness, Hope, Identity, Meaning, Empowerment) differed between the two groups. Cohen's d effect sizes were computed to quantify the magnitude of the differences between the group means in standard deviation units. According to Cohen (1988), effect sizes of Cohen's d are interpreted as follows; a d of 0.2 indicating a small effect, 0.5 representing a moderate effect and 0.8 denoting a large effect. All statistical analyses were performed in R using the stats package (R Core Team., 2025), with significance set at $p < .05$.

Results

After preprocessing, 961 r/CPTSD posts and 884 r/CPTSDNextSteps posts remained available for text mining analysis. The distribution of post lengths between subreddits was visualized in the box plot in Figure 3. Posts in r/CPTSDNextSteps exhibited a higher median post length and more than double the average text length of r/CPTSD posts.

Figure 3

Box Plot of Post Lengths in r/CPTSD and r/CPTSDNextSteps



Note. Post length was measured as the numbers of words per post after preprocessing. A logarithmic scale (base 10) was used on the y-axis.

Validation of GPT 4o-generated ratings against human ratings

ICC agreement coefficients between human raters and GPT-4o across the total subset of 200 posts and the posts stratified by subreddit were summarized in Table 1. To visualize the relationship between averaged human and GPT-4o ratings, scatter plots with regression lines were created for each construct (Figure 4, Appendix C).

GPT-4o's accuracy in coding sentiment and CHIME components was evaluated against both individual human ratings and their averaged ratings, as summarized in Table 1. Human-GPT agreement was generally good to excellent for Sentiment, Connectedness, Hope, and Empowerment, with ICCs exceeding 0.60 for both raters for HRa v. GPT and for HRb v. GPT. However, agreement was poor to fair for Identity and Meaning in r/CPTSD, whilst being good to excellent in r/CPTSDNextSteps.

Scatter plots (Figure 4) confirmed a good agreement overall between averaged human and GPT-4o ratings, with an average ICC across measures of 0.63 for r/CPTSD and 0.76 for r/CPTSDNextSteps and slightly narrower confidence intervals in r/CPTSDNextSteps, which indicated a greater consistency in ratings for the recovery-focused subreddit.

Table 1

Intraclass Correlation Coefficients for Human Rater and Human-GPT Agreement across Subreddits

Component	Subreddit	HRa vs GPT	HRb vs GPT	HRa/b vs GPT
		ICC [95% CI]	ICC [95% CI]	ICC [95% CI]
Sentiment	Total (n = 200)	0.87 [0.83, 0.90]	0.87 [0.82, 0.90]	0.91 [0.89, 0.93]
	r/CPTSD	0.89 [0.84, 0.92]	0.88 [0.82, 0.92]	0.92 [0.89, 0.95]
	r/CPTSDNextSteps	0.74 [0.63, 0.82]	0.79 [0.69, 0.85]	0.84 [0.78, 0.89]
Connectedness	Total (n = 200)	0.75 [0.68, 0.81]	0.66 [0.57, 0.73]	0.78 [0.72, 0.83]
	r/CPTSD	0.78 [0.65, 0.85]	0.68 [0.39, 0.82]	0.79 [0.53, 0.89]
	r/CPTSDNextSteps	0.69 [0.56, 0.78]	0.60 [0.43, 0.73]	0.74 [0.58, 0.84]
Hope	Total (n = 200)	0.93 [0.91, 0.95]	0.84 [0.75, 0.90]	0.93 [0.89, 0.95]
	r/CPTSD	0.91 [0.87, 0.94]	0.83 [0.68, 0.91]	0.91 [0.83, 0.95]
	r/CPTSDNextSteps	0.82 [0.74, 0.87]	0.60 [0.45, 0.72]	0.80 [0.71, 0.86]
Identity	Total (n = 200)	0.57 [0.38, 0.69]	0.41 [-0.01, 0.66]	0.54 [0.08, 0.75]
	r/CPTSD	0.54 [0.25, 0.72]	0.31 [-0.09, 0.65]	0.46 [-0.10, 0.76]
	r/CPTSDNextSteps	0.61 [0.45, 0.72]	0.60 [0.38, 0.74]	0.69 [0.47, 0.81]
Meaning	Total (n = 200)	0.53 [0.38, 0.64]	0.19 [-0.04, 0.39]	0.37 [0.03, 0.59]
	r/CPTSD	0.38 [0.10, 0.59]	0.10 [-0.07, 0.32]	0.22 [-0.09, 0.53]

	r/CPTSDNextSteps	0.74 [0.64, 0.82]	0.64 [0.44, 0.76]	0.78 [0.67, 0.85]
Empowerment	Total (n = 200)	0.86 [0.81, 0.90]	0.85 [0.79, 0.89]	0.90 [0.85, 0.94]
	r/CPTSD	0.83 [0.73, 0.89]	0.90 [0.85, 0.93]	0.91 [0.85, 0.94]
	r/CPTSDNextSteps	0.79 [0.70, 0.85]	0.58 [0.39, 0.71]	0.76 [0.61, 0.85]

Note. ICC < 0.40 = poor, 0.40–0.59 = fair, 0.60–0.74 = good, 0.75–1.00 = excellent. An ICC of

less than 0.40 was deemed as insufficient. HRa = human rater A, HRb = human rater B.

HRa/b = average score between human rater A and human rater B.

Bland-Altman plots of differences between averaged human ratings (HRa/b) and GPT-4o ratings for each construct in the total sample were presented in Figure 5. On average, GPT-4o tended to provide higher ratings than humans, with the largest mean differences observed for Identity (-0.74) and Meaning (-0.78). In contrast, sentiment and Connectedness showed less systematic bias between human and GPT-4o ratings. As seen in Appendix D, GPT-4o tended to overestimated ratings more in r/CPTSD (mean difference = -0.39) than in r/CPTSDNextSteps posts (mean difference = 0.02).

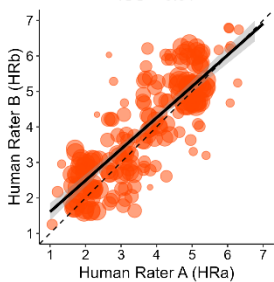
Overall, GPT-4o demonstrated good to excellent accuracy for coding sentiment, Connectedness, Hope, and Empowerment in both subreddits, and for Meaning and Identity in r/CPTSDNextSteps. Reliability was especially limited for Identity and Meaning in r/CPTSD, with variability observed across components and subreddits. Notably, GPT-4o demonstrated greater alignment with human ratings in r/CPTSDNextSteps compared to r/CPTSD.

Figure 4

Scatterplots of human-rater and human-GPT comparisons per measure and across the total sample.

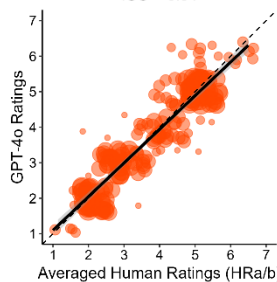
Human Rater Agreement: Sentiment

ICC = 0.81



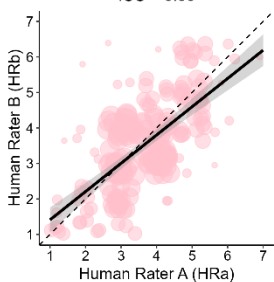
Human-GPT Comparison: Sentiment

ICC = 0.91



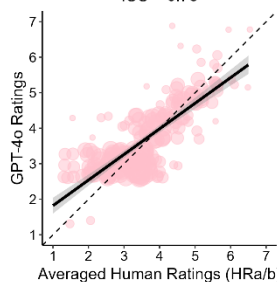
Human Rater Agreement: Connectedness

ICC = 0.66



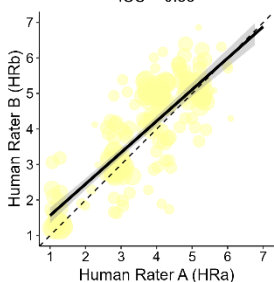
Human-GPT Comparison: Connectedness

ICC = 0.78



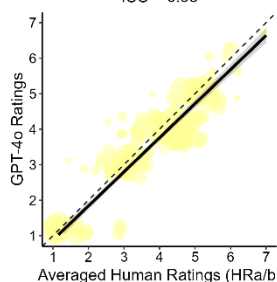
Human Rater Agreement: Hope

ICC = 0.83



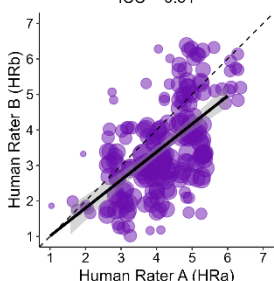
Human-GPT Comparison: Hope

ICC = 0.93



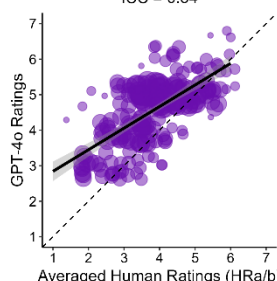
Human Rater Agreement: Identity

ICC = 0.51



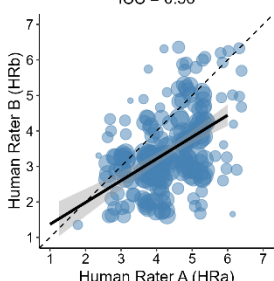
Human-GPT Comparison: Identity

ICC = 0.54



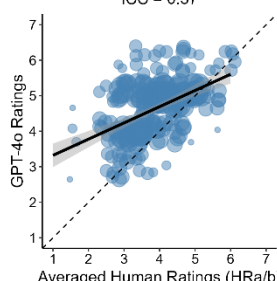
Human Rater Agreement: Meaning

ICC = 0.36



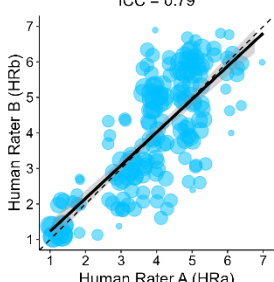
Human-GPT Comparison: Meaning

ICC = 0.37



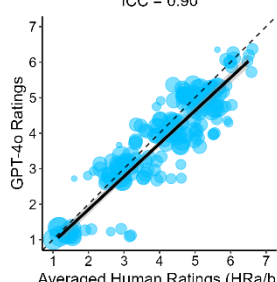
Human Rater Agreement: Empowerment

ICC = 0.79



Human-GPT Comparison: Empowerment

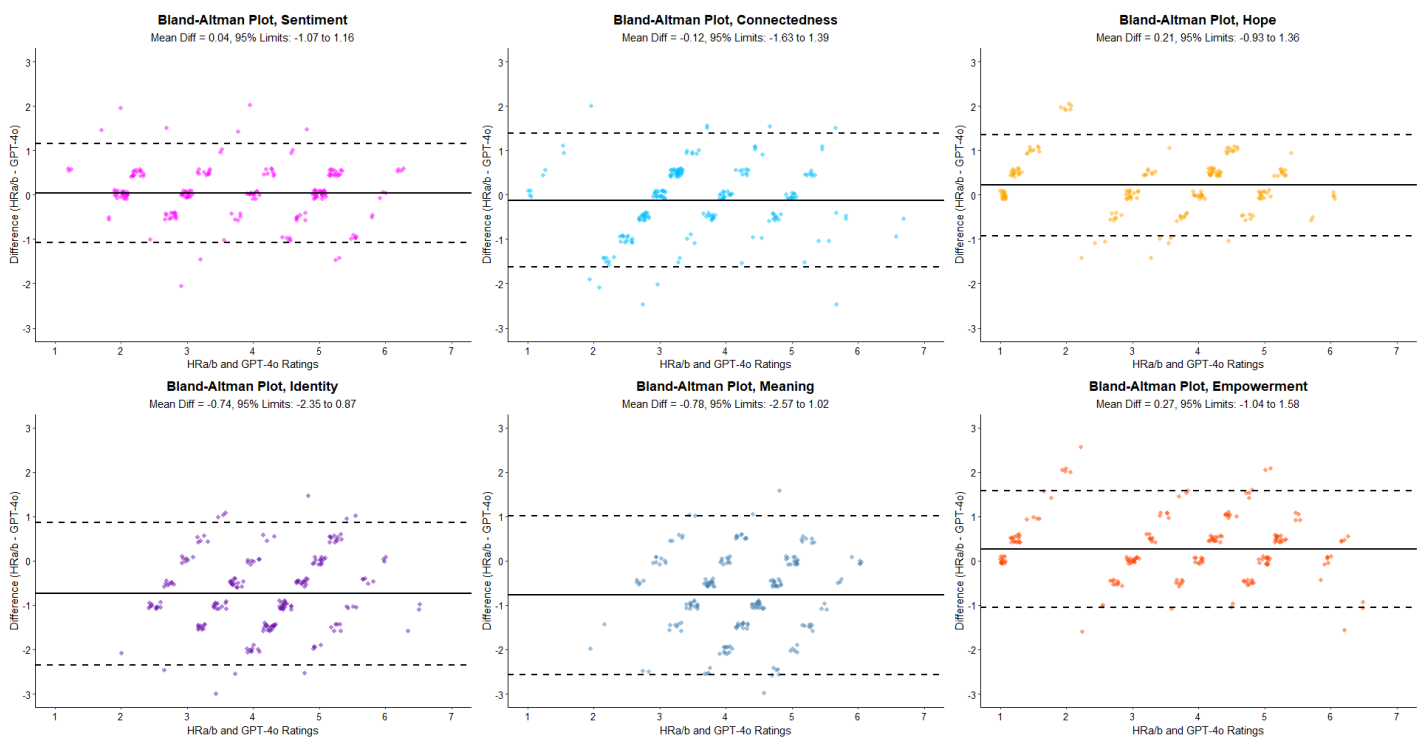
ICC = 0.90



Note. Linear regression lines were plotted with 95% confidence intervals. Dashed identity lines ($y = x$) indicated perfect agreement.

Figure 5

Bland-Altman plots for each construct (Sentiment, Connectedness, Hope, Identity, Meaning, Empowerment) in the total sample ($n = 200$), comparing averaged human ratings (HRa/b) to GPT-4o ratings.



Comparison of GPT-4o-based Sentiment and CHIME Scores Between r/CPTSD and r/CPTSDNextSteps

Means and standard deviations (SDs) for sentiment and CHIME component scores in r/CPTSD and r/CPTSDNextSteps were displayed in Table 2, alongside the results of independent-samples Welch corrected t -tests and Cohen's d effect sizes.

Table 2

Descriptive Statistics and Independent-Samples t-Tests for GPT-4o computed Sentiment and CHIME scores by Subreddit

Variable	r/CPTSD	/rCPTSDNextSteps	<i>t</i> (df)	<i>p</i>	Cohen's <i>d</i>
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)			
Sentiment	3.01 (1.46)	4.34 (1.03)	22.97 (1729)	< .001*	1.06
Connectedness	3.55 (1.20)	3.54 (0.78)	-0.32 (1666)	.627	-0.01
Hope	2.19 (1.51)	4.25 (0.80)	36.92 (1488)	< .001*	1.68
Identity	4.67 (1.03)	4.67 (0.88)	-0.13 (1833)	.552	-0.01
Meaning	4.93 (0.95)	4.51 (0.84)	-9.90 (1839)	< .001*	-0.46
Empowerment	2.52 (1.68)	4.42 (0.94)	30.26 (1526)	< .001*	1.38

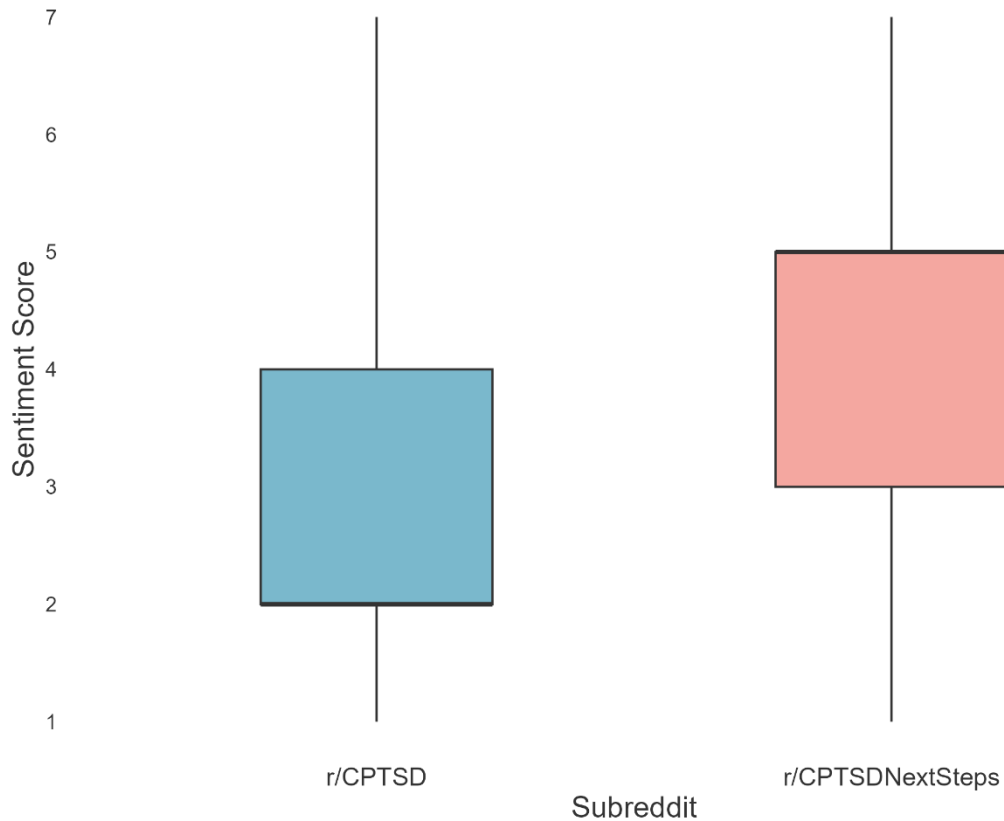
*Indicates $p < 0.05$.

Sentiment Analysis

As expected, r/CPTSDNextSteps demonstrated significantly more positive sentiment scores, with a large effect size (Cohen's $d = 1.06$). This difference was also reflected in the box-plots, with many scores centred at 5 for r/CPTSDNextSteps and at 2 for r/CPTSD (Figure 6).

Figure 6

Box Plot of Sentiment Scores Across Subreddits. Median sentiment score for r/CPTSD and r/CPTSDNextSteps are 2 and 5, respectively.

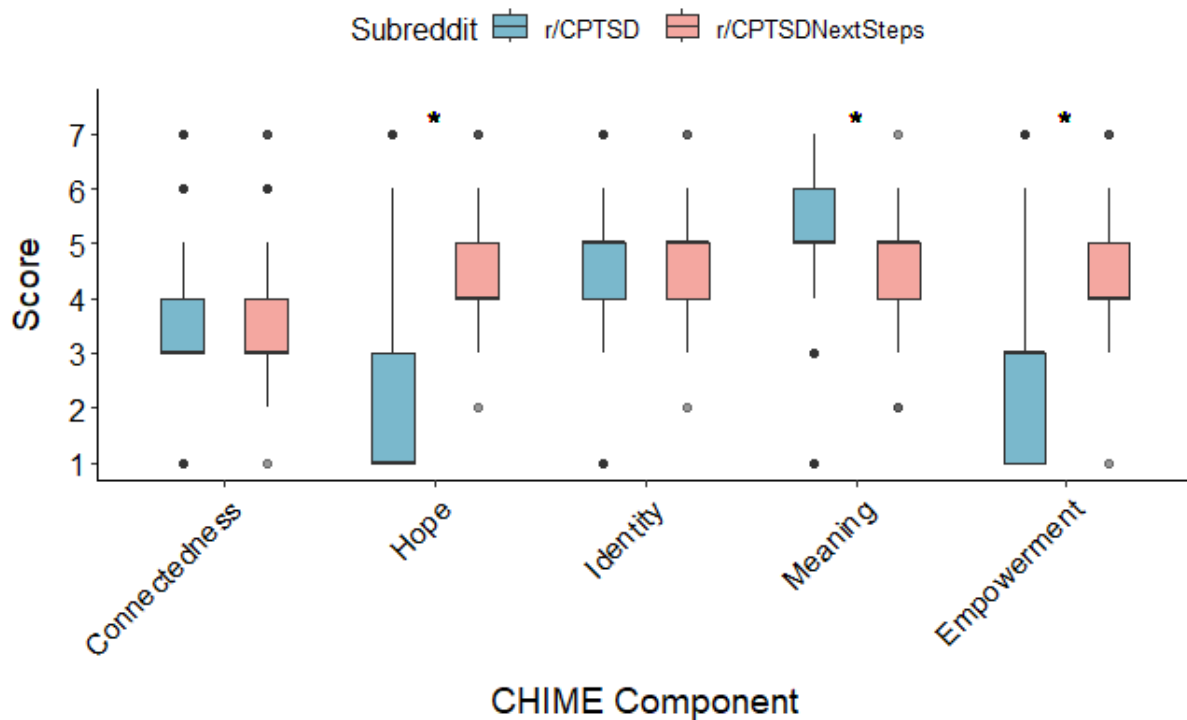


CHIME Components Analysis

Consistent with the hypothesis regarding the CHIME dimensions (Figure 7 and Table 2), r/CPTSDNextSteps posts scored significantly higher on Hope ($M = 4.25$, $SD = 0.80$) and Empowerment ($M = 4.42$, $SD = 0.94$) than in r/CPTSD ($M = 2.19$, $SD = 1.51$ and $M = 2.52$, $SD = 1.68$), both with large effect sizes. Unexpectedly, r/CPTSD posts scored significantly higher on Meaning $M = 4.93$, $SD = 0.95$ v. $M = 4.51$, $SD = 0.84$ in r/CPTSDNextSteps with a medium effect size. No group differences emerged for Connectedness or Identity (see Figure 7 and Table 2).

Figure 7

Box Plot of Mean GPT-4o computed CHIME Components Scores.



*Indicates significant differences between subreddits ($p < .05$)

Discussion

The present study investigated the discourse on CPTSD within 1992 posts in two online mental health communities, r/CPTSD and r/CPTSDNextSteps, using text mining methods. Posts were analysed with GPT-4o, through ChatGPT, to assess sentiment and alignment with the CHIME model, as well as to evaluate the reliability of LLMs in rating psychological constructs. Overall, the study showed that GPT-4o was effective in capturing sentiment and most CHIME components with good to excellent reliability, although it showed lower accuracy for Meaning and Identity in r/CPTSD. GPT-based sentiment and CHIME scores indicated that posts in the recovery-oriented r/CPTSDNextSteps scored significantly higher on sentiment and Hope and Empowerment, but lower on Meaning, compared to r/CPTSD, with no differences in Identity or Connectedness.

Interpreting the findings

Reliability of GPT-4o in Text Analysis

GPT-4o demonstrated good to excellent reliability in scoring sentiment, Connectedness, Hope and Empowerment in r/CPTSD posts, and sentiment, Connectedness, Hope, Identity, Meaning and Empowerment in r/CPTSDNextSteps posts. These findings are largely consistent with and expand on the findings of Rathje et al. (2024), who demonstrated that GPT models could accurately detect psychological constructs such as sentiment and emotions across texts using zero-shot prompting, outperforming traditional dictionary-based methods in construct detection as well as reliability.

The poor agreement for Meaning scores in r/CPTSD posts for both human raters and Human-GPT is consistent with the finding of Rathje et al. (2024) that GPT may struggle with constructs that are hard to clearly define. Specifically, unlike the concrete markers for positive or negative sentiment, Meaning, by definition, is shaped by personal narratives, consistent with the finding of Sakaguchi et al. (2025) that ChatGPT-4's accuracy on nuanced themes dropped significantly compared to descriptive themes. Prior work on qualitative coding with GPT-4 has shown that hard-to-define constructs for humans, also result in poor agreement for GPT-4 (Liu et al., 2018), which may explain the poor inter-rater reliability on Meaning.

Despite these areas of weak alignment, the findings overall indicate that GPT-4o can be a viable tool for psychological text analysis of recovery-oriented constructs in OMHCs. Nonetheless, GPT-4o's performance varies by construct and context, still necessitating validation against human ratings, especially for detecting complex or nuanced psychological constructs such as some recovery-based concepts in trauma-based discourse.

Alignment of the recovery subreddit with the CHIME Model

As hypothesized, r/CPTSDNextSteps posts scored significantly higher on Hope and Empowerment compared to r/CPTSD with large effect sizes (Table 2). Unexpectedly, r/CPTSD posts scored significantly higher on Meaning than r/CPTSDNextSteps with a medium effect size. This finding suggests that validation-oriented spaces such as r/CPTSD may also facilitate reflection and purpose-finding. However, the poor inter-rater and human-GPT agreement for Meaning in r/CPTSD and the tendency of GPT-4o to overestimate Meaning in r/CPTSD suggests that the higher scores may be driven by variable interpretations of “Meaning” rather than thematic differences across subreddits.

Sentiment and Emotional Tone

Consistent with expectations, posts in the r/CPTSDNextSteps subreddit exhibited significantly more positive sentiment compared to r/CPTSD, with a large effect size. This finding supports the hypothesis that recovery-oriented communities foster more positive discourse, possibly due to the subreddit’s rules emphasizing self-help strategies and goal-setting over venting (Griffiths et al., 2012; Prescott et al., 2020; Reddit, r/CPTSDNextSteps, 2025). In contrast, r/CPTSD’s focus on validation and trauma-sharing may contribute to its more negative sentiment, as users may more often express distress and emotional pain, consistent with prior research on validation-seeking behaviours reinforcing rumination (Marr et al., 2022; Park et al., 2017).

Strengths, Limitations & Future Research

A strength of the study lies in the novel use of LLMs for this purpose, which enabled the analysis of large volumes of unstructured text data without extensive preprocessing and capturing contextual nuances. The validation of GPT-4o’s ratings against human raters, using ICCs and Bland-Altman plots confirmed good to excellent reliability for most constructs. As

such, this study adds to the emerging evidence that supports the accuracy and feasibility of LLMs in psychological text analysis, particularly in recovery mental health contexts. This novel approach aligns with recent research highlighting LLMs' superiority over traditional methods for psychological text analysis (Mu et al., 2024; Rathje et al., 2024). Furthermore, the study's focus on "top" posts ensured high-engagement content, likely reflecting dominant themes that resonate with community members, thus also enhancing the representativeness of findings for exploratory purposes (Ignatow & Mihalcea, 2017).

However, several limitations must be acknowledged. First, regarding the sampling strategy of posts, the "top" scoring algorithm may have introduced a bias towards more active and influential users, potentially overlooking less popular but important discussions. In addition, the study did not explore longitudinal trends, as the "top" post selection may not fully capture the evolution of discourse over time, a critical aspect for understanding recovery trajectories in CPTSD (Wislocki et al., 2024). Considering the study's exploratory nature, it may be a reasonable compromise given the limitations set by the Reddit Application Programming Interface. The comparative design of the two subreddits may also not generalize to other platforms (e.g., X, TikTok) or less structured OMHCs, where community dynamics and user demographics might differ.

The present study also found notable differences in post lengths between the two subreddits, with r/CPTSDNextSteps posts being, on average, more than twice as long as those in r/CPTSD. This variability in post lengths may pose challenges for text analysis. Longer posts may have provided richer context, making it more feasible for GPT-4o and human raters to rate the constructs, whereas rating agreement may have been reduced in r/CPTSD due to the immediacy of posts introducing more ambiguity. Future research could explore the relationship between post length and coding accuracy.

Additionally, the manual validation sample size (200 posts, 100 per subreddit) was modest. While this sample size provided a sufficient basis for estimating the inter-rater reliability with reasonable accuracy (Bonett, 2002), it may not have fully captured the variability across all post types. Additionally, GPT-4o's training data lacks transparency, raising concerns about potential biases, as evidenced by its overestimation in r/CPTSD compared to human ratings. This variability may stem from the "black box" nature of LLMs, as their training on unsupervised datasets can introduce biases that are difficult to interpret or control (Belal et al., 2023). While this limitation was mitigated by validating against human ratings, it underscores the need for cautious interpretation of LLM-generated scores in psychological research.

Implications

Given the exploratory nature of this research, its implications are primarily theoretical, with potential practical applications for understanding and supporting CPTSD recovery in OMHCs. The study demonstrates and further confirms the utility of GPT-4o as a deductive psychological text mining tool for analysing large-scale, unstructured Reddit data, effectively capturing sentiment and CHIME model components with good to excellent for most constructs, such as sentiment and Hope. This aligns with prior research highlighting the scalability and contextual sensitivity of large language models (LLMs) in psychological text analysis (Rathje et al., 2024). However, the observed variability in accuracy, in particular the poor performance for Meaning and Identity in r/CPTSD, underscores the need for careful validation when applying LLMs to nuanced psychological constructs, mirroring challenges noted in other transformer-based analyses (Belal et al., 2023; Pugh et al., 2024) such as providing inconsistent results without careful prompting, suggesting that future applications of LLMs in mental health research may need to prioritize construct-specific validation and transparency in model training to mitigate biases.

The findings also offer theoretical insights into the role of OMHCs in CPTSD recovery. The significant differences in sentiment and CHIME alignment demonstrate how community focus can help shape discourse and recovery narratives. Future studies can potentially build on this research by exploring how specific CHIME components evolve over time within these communities by using longitudinal designs.

Conclusion

This study sought to explore the discourse within r/CPTSD and r/CPTSDNextSteps, two online mental health communities (OMHCs) dedicated to individuals with Complex Post-Traumatic Stress Disorder (CPTSD), by examining their alignment with the CHIME model of personal recovery, sentiment profiles, and the accuracy of GPT-4o in coding psychological constructs. The validation of GPT-4o as a scalable tool for mental health analysis, despite its limitations in detecting and scoring nuanced constructs, addresses a gap in applying LLMs to trauma-related discourse. This research contributes to the field by offering the first CHIME-based comparison of validation- (r/CPTSD) and recovery-oriented (r/CPTSDNextSteps) CPTSD communities. It has demonstrated how recovery-oriented spaces reflect hope and empowerment and validation-oriented spaces reflect trauma-centric reflection.

References

- Adams, N. N. (2022). 'Scraping' Reddit posts for academic research? Addressing some blurred lines of consent in growing internet-based research trend during the time of Covid-19. *International Journal Of Social Research Methodology*, 27(1), 47–62. <https://doi.org/10.1080/13645579.2022.2111816>
- American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. <https://doi.org/10.1176/appi.books.9780890425596>
- Anthony, W. A. (1993). Recovery from mental illness: The guiding vision of the mental health service system in the 1990s. *Psychosocial Rehabilitation Journal*, 16(4), 11–23. <https://doi.org/10.1037/h0095655>
- Belal, M., She, J., & Wong, S. (2023). Leveraging ChatGPT as text annotation tool for sentiment analysis. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2306.17177>
- Blackie, M., De Boer, K., Seabrook, L., Bates, G., & Nedeljkovic, M. (2024). Digital-Based Interventions for Complex Post-Traumatic Stress Disorder: A Systematic Literature Review. *Trauma Violence & Abuse*, 25(4), 3115–3130. <https://doi.org/10.1177/15248380241238760>
- Bland, J. M., & Altman, D. (1986). STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. *The Lancet*, 327(8476), 307–310. [https://doi.org/10.1016/s0140-6736\(86\)90837-8](https://doi.org/10.1016/s0140-6736(86)90837-8)
- Bonett, D. G. (2002). Sample size requirements for estimating intraclass correlations with desired precision. *Statistics in Medicine*, 21(9), 1331–1335. <https://doi.org/10.1002/sim.1108>

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2), 15–

21. <https://doi.org/10.1109/mis.2013.30>

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–

290. <https://doi.org/10.1037/1040-3590.6.4.284>

Clegg, A. B., & Shepherd, A. J. (2008). Text mining. *Methods in Molecular Biology*, 471–491. https://doi.org/10.1007/978-1-60327-429-6_25

Cloitre, M., Garvert, D. W., Brewin, C. R., Bryant, R. A., & Maercker, A. (2013). Evidence for proposed ICD-11 PTSD and complex PTSD: a latent profile analysis. *European Journal of Psychotraumatology*, 4(1). <https://doi.org/10.3402/ejpt.v4i0.20706>

Corzine, A., & Roy, A. (2024). Inside the black mirror: current perspectives on the role of social media in mental illness self-diagnosis. *Discover Psychology*, 4(1).

<https://doi.org/10.1007/s44202-024-00152-3>

De Boer, K., Arnold, C., Mackelprang, J. L., & Nedeljkovic, M. (2022). Barriers and facilitators to treatment seeking and engagement amongst women with complex trauma histories. *Health & Social Care in the Community*, 30(6). <https://doi.org/10.1111/hsc.13823>

De Choudhury, M. (2013). Role of social media in tackling challenges in mental health. *ACM Digital Library*. <https://doi.org/10.1145/2509916.2509921>

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text Mining Infrastructure in R. *Journal of Statistical Software*, 25(5), 1–54. <https://doi.org/10.18637/jss.v025.i05>

Ford, J. D. (2021). Progress and Limitations in the Treatment of Complex PTSD and Developmental Trauma Disorder. *Current Treatment Options in Psychiatry*, 8(1), 1–

17. <https://doi.org/10.1007/s40501-020-00236-6>

Garg, M. (2023). Mental health analysis in social media Posts: a survey. *Archives Of Computational Methods in Engineering*, 30(3), 1819–1842. <https://doi.org/10.1007/s11831-022-09863-z>

Griffiths, K. M., Mackinnon, A. J., Crisp, D. A., Christensen, H., Bennett, K., & Farrer, L. (2012). The Effectiveness of an Online Support Group for Members of the Community with Depression: A Randomised Controlled Trial. *PLoS ONE*, 7(12), e53244. <https://doi.org/10.1371/journal.pone.0053244>

Gu, D., Li, M., Yang, X., Gu, Y., Zhao, Y., Liang, C., & Liu, H. (2022). An analysis of cognitive change in online mental health communities: A textual data analysis based on post replies of support seekers. *Information Processing & Management*, 60(2), 103192. <https://doi.org/10.1016/j.ipm.2022.103192>

Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text mining in big data analytics. *Big Data and Cognitive Computing*, 4(1), 1. <https://doi.org/10.3390/bdcc4010001>

Herman, J. L. (1992). Complex PTSD: A syndrome in survivors of prolonged and repeated trauma. *Journal of Traumatic Stress*, 5(3), 377–391. <https://doi.org/10.1002/jts.2490050305>

Ignatow, G., & Mihalcea, R. (2018). *An introduction to text mining: research design, data collection, and analysis*. <https://doi.org/10.4135/9781506336985>

Kim, S., Cha, J., Kim, D., & Park, E. (2023). Understanding mental health issues in different subdomains of social networking services: Computational analysis of Text-Based Reddit posts. *Journal of Medical Internet Research*, 25, e49074.

<https://doi.org/10.2196/49074>

Kodratoff, Y. (2002). Rating the interest of rules induced from data and within texts. *12th International Workshop On Database And Expert Systems Applications (DEXA 2001)*, 265–269. <https://doi.org/10.1109/dexa.2001.953073>

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal Of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>

Kraiss, J. T., Klooster, P. M. T., Chrispijn, M., Stevens, A. W., Kupka, R. W., & Bohlmeijer, E. T. (2019). Measuring personal recovery in people with bipolar disorder and exploring its relationship with well-being and social role participation. *Clinical Psychology & Psychotherapy*, 26(5), 540–549. <https://doi.org/10.1002/cpp.2371>

Leamy, M., Bird, V., Boutillier, C. L., Williams, J., & Slade, M. (2011). Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *The British Journal of Psychiatry*, 199(6), 445–452.

<https://doi.org/10.1192/bjp.bp.110.083733>

Liu, X., Zambrano, A. F., Baker, R. S., Barany, A., Ocumpaugh, J., Zhang, J., Pankiewicz, M., Nasiar, N., & Wei, Z. (2025). Qualitative Coding with GPT-4. *Journal of Learning Analytics*, 12(1), 169–185. <https://doi.org/10.18608/jla.2025.8575>

Maercker, A., Cloitre, M., Bachem, R., Schlumpf, Y. R., Khoury, B., Hitchcock, C., & Bohus, M. (2022). Complex post-traumatic stress disorder. *The Lancet*, 400(10345), 60–72.

[https://doi.org/10.1016/s0140-6736\(22\)00821-2](https://doi.org/10.1016/s0140-6736(22)00821-2)

Marr, N. S., Zainal, N. H., & Newman, M. G. (2022). Focus on and venting of negative emotion mediates the 18-year bi-directional relations between major depressive disorder and generalized anxiety disorder diagnoses. *Journal Of Affective Disorders*, 303, 10–17. <https://doi.org/10.1016/j.jad.2022.01.079>

Mehmood, K., et al. (2020). Systematic review on text normalization techniques and its approach to non-standard words. *International Journal of Computer Applications*, 185(33), 1-4.

Miner, G., Delen, D., Elder, J., Fast, A., Hill, T., Nisbet, R., & Balakrishnan, K. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press.

Mu, Y., Dong, C., Bontcheva, K., & Song, X. (2024). Large Language Models Offer an Alternative to the Traditional Approach of Topic Modelling. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2403.16248>

Muller, K., & Cohen, J. (1989). Statistical Power Analysis for the Behavioral Sciences. *Technometrics*, 31(4), 499. <https://doi.org/10.2307/1270020>

Murphy, D., Karatzias, T., Busuttil, W., Greenberg, N., & Shevlin, M. (2021). ICD-11 posttraumatic stress disorder (PTSD) and complex PTSD (CPTSD) in treatment seeking veterans: risk factors and comorbidity. *Social Psychiatry and Psychiatric Epidemiology*, 56(7), 1289–1298. <https://doi.org/10.1007/s00127-021-02028-6>

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>

Neuendorf, K. A. (2017). *The Content Analysis Guidebook*. <https://doi.org/10.4135/9781071802878>

Park, A., Conway, M., & Chen, A. T. (2017). 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>

Prescott, J., Rathbone, A. L., & Hanley, T. (2020). Online mental health communities, self-efficacy and transition to further support. *Mental Health Review Journal*, 25(4), 329–344. <https://doi.org/10.1108/mhrj-12-2019-0048>

Pugh, S. L., Chandler, C., Cohen, A. S., Diaz-Asper, C., Elvevåg, B., & Foltz, P. W. (2024). Assessing Dimensions of Thought Disorder with Large Language Models: The Tradeoff of Accuracy and Consistency. *Psychiatry Research*, 341, 116119. <https://doi.org/10.1016/j.psychres.2024.116119>

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Rathje, S., Mirea, D., Sucholutsky, I., Marjeh, R., Robertson, C. E., & Van Bavel, J. J. (2024). GPT is an effective tool for multilingual psychological text analysis. *Proceedings Of The National Academy Of Sciences*, 121(34). <https://doi.org/10.1073/pnas.2308950121>

Reddit r/CPTSDNextSteps. (2025). *CPTSDNextSteps* [Online forum]. Reddit. <https://www.reddit.com/r/CPTSDNextSteps/>

Reddit r/CPSTD. (2025). *CPTSD* [Online forum]. Reddit. <https://www.reddit.com/r/CPTSD>

Rivera, I. (2023). *RedditExtractoR: Reddit data extraction toolkit* (Version 3.0.9) [R package]. <https://CRAN.R-project.org/package=RedditExtractoR>

Sakaguchi, K., Sakama, R., & Watari, T. (2025). Evaluating ChatGPT in Qualitative Thematic Analysis With Human Researchers in the Japanese Clinical Context and Its Cultural Interpretation Challenges: Comparative Qualitative Study. *Journal Of Medical Internet Research*, 27, e71521. <https://doi.org/10.2196/71521>

Slade, M., Amering, M., & Oades, L. (2008). Recovery: an international perspective. *Epidemiologia E Psichiatria Sociale*, 17(2), 128–137. <https://doi.org/10.1017/s1121189x00002827>

Stadtmann, M. P., Maercker, A., Binder, J., & Schnepf, W. (2018). Why do I have to suffer? Symptom management, views and experiences of persons with a CPTSD: a grounded theory approach. *BMC Psychiatry*, 18(1). <https://doi.org/10.1186/s12888-018-1971-9>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. *arXiv (Cornell University)*, 30, 5998–6008. <https://arxiv.org/pdf/1706.03762v5>

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., & Yutani, H. (2024). *ggplot2: Create elegant data visualisations using the grammar of graphics* (Version 3.5.0) [R package]. <https://ggplot2.tidyverse.org>

Wislocki, K. E., Naderi, G., Schueller, S. M., & Zalta, A. K. (2024). Assessing similarities and differences in thematic content across online mental health communities dedicated to trauma-related mental health conditions. *Psychological Trauma Theory Research Practice and Policy*. <https://doi.org/10.1037/tra0001817>

Wongkoblapp, A., Vadillo, M. A., & Curcin, V. (2017). Researching Mental Health Disorders in the Era of Social Media: Systematic Review. *Journal Of Medical Internet Research*, 19(6), e228. <https://doi.org/10.2196/jmir.7215>

World Health Organization. (2025, January). *6B41 Complex Post Traumatic Stress Disorder*. <https://icd.who.int/browse/2025-01/mms/en#585833559>

Appendix A

ChatGPT Prompt to Generate Sentiment Ratings with ChatGPT 4o

I am examining the emotional tone of Reddit posts in the r/CPTSD and r/CPTSDNextSteps communities. To this end, I am conducting sentiment analysis with ChatGPT. Your task is to evaluate the sentiment of a given post on a 1-7 Likert scale, which is defined as follows:

1 = extremely negative

4 = neutral

7 = extremely positive

Please review the Reddit posts in the attachment and assign each one of them a sentiment rating from 1 to 7. Provide your response in the following format:

Sentiment Rating: [number]

Appendix B

CHIME Model Analysis using ChatGPT 4o

I am examining alignment with the CHIME model of personal recovery of Reddit posts in the r/CPTSD and r/CPTSDNextSteps communities. To this end, I need you to assess the given posts' alignment with each of the five CHIME components:

Connectedness: The extent to which the post reflects social support, relationships, or community.

Hope: The presence of optimistic statements about recovery and future improvement.

Identity: Discussions related to rebuilding a positive self-concept, personal growth, or reclaiming one's identity after trauma.

Meaning: Expression of purpose, existential reflections, or finding meaning in lived experiences.

Empowerment: Statements that convey self-efficacy, agency, and personal strength in the recovery process.

For the following posts in the attachment, rate its alignment with each component on a 1-7 Likert scale (1 = no alignment, 7 = strong alignment). Provide your responses in the following format:

Connectedness: [number]

Hope: [number]

Identity: [number]

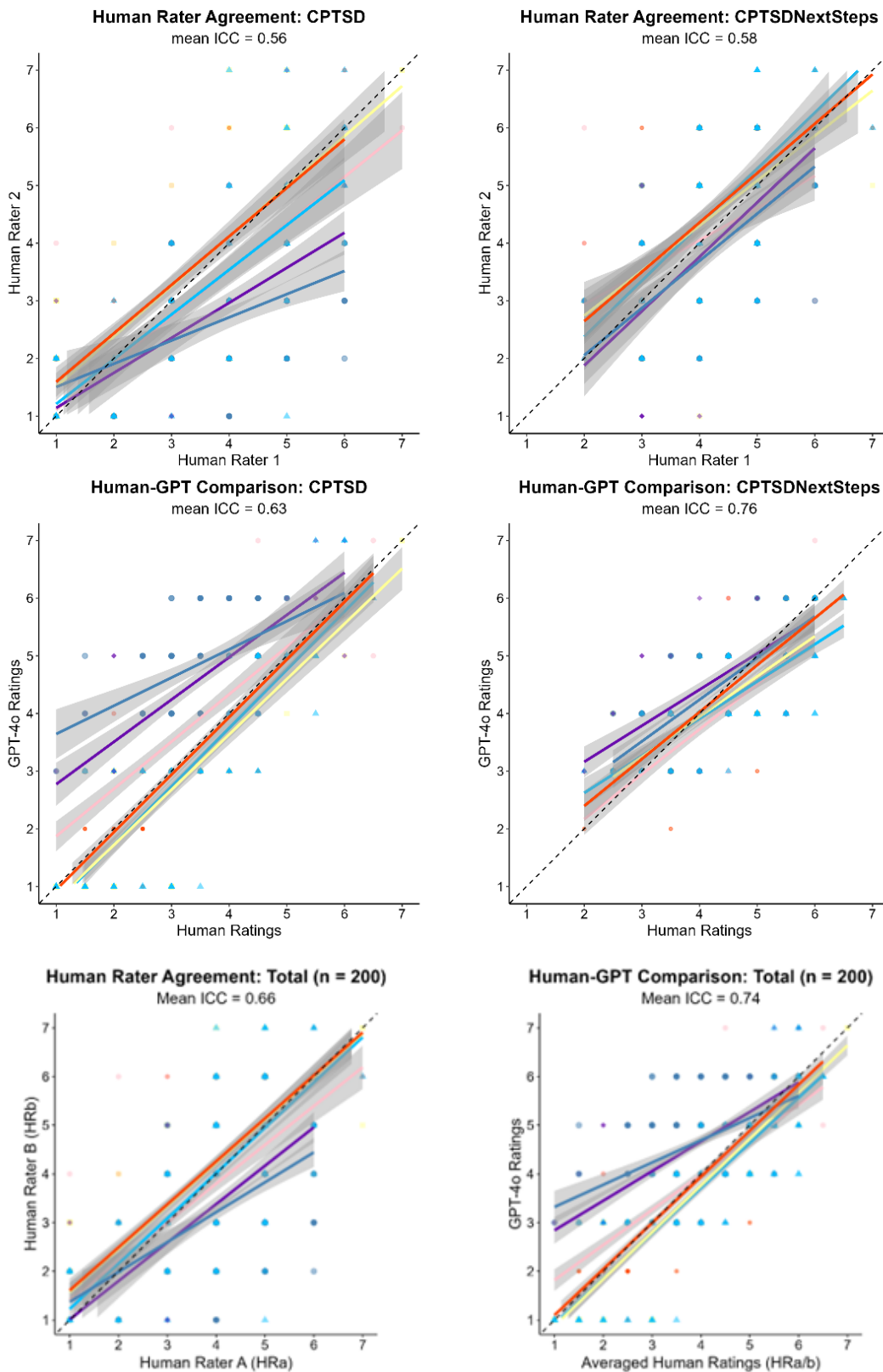
Meaning: [number]

Empowerment: [number]

Appendix C

Figure 8

Scatterplots of human-human and human-GPT rating comparisons across both subreddits



Measure	 Connectedness	 Hope	 Meaning
	 Empowerment	 Identity	 Sentiment

Appendix D

Bland-Altman Plots for r/CPTSD and r/CPTSDNextSteps

Figure 9

Bland-Altman plots for each construct (Sentiment, Connectedness, Hope, Identity, Meaning, Empowerment) in the r/CPTSD sample ($n = 100$), comparing averaged human ratings (HRa/b) to GPT-4o ratings.

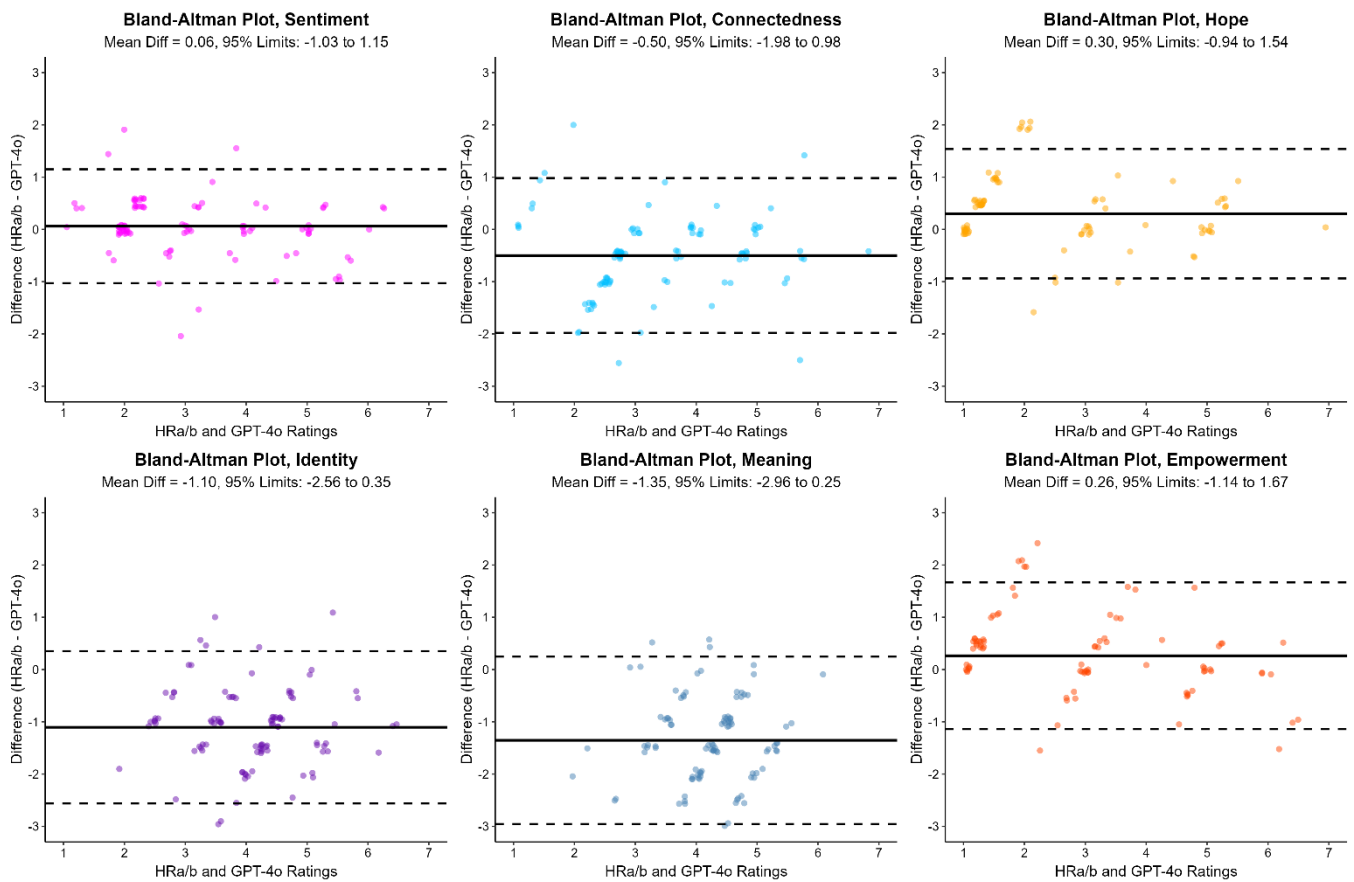
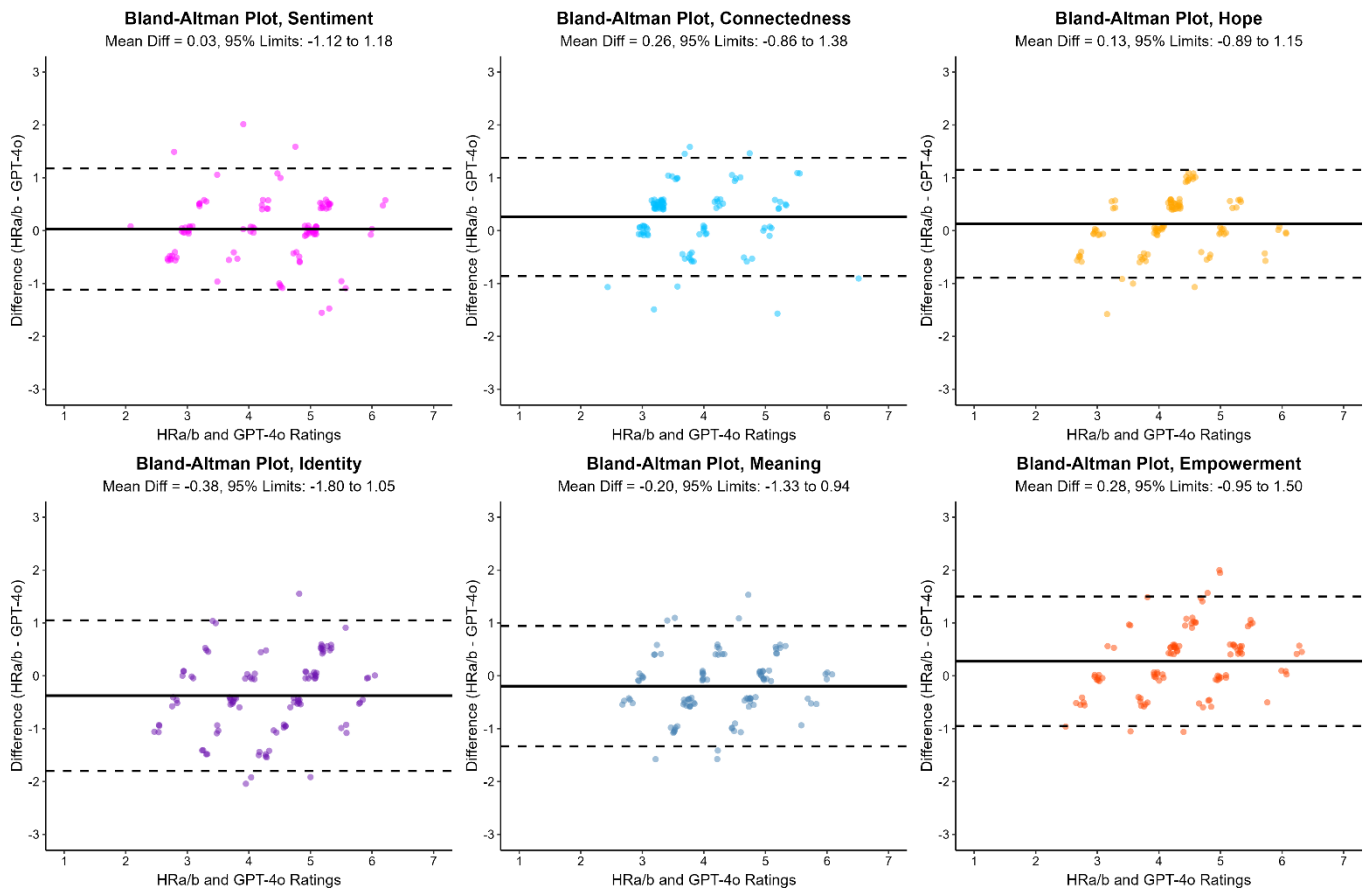


Figure 10

Bland-Altman plots for each construct (Sentiment, Connectedness, Hope, Identity, Meaning, Empowerment) in the r/CPTSDNextSteps sample (n = 100), comparing averaged human ratings (HRa/b) to GPT-4o ratings



Appendix E

R Code

The R code for this study is available on GitHub, at

https://github.com/selin1202/CPTSD_code_data.git