Mental Distress in English Posts from *r/AmITheAsshole* Subreddit community with Large Language Models: A survey

Author1
Affiliation1
Address1
author1@xxx.yy

Abstract

Mental distress emerged as a significant area of concern. However, previous research predominantly concentrated on the detection and classification of mental issues, with limited attention given to comprehensive investigations of the interrelationships and contextual events associated with these issues. This research delves into the exploration of mental distress derived from user-generated content on the *r/AmItheAsshole* subreddit. The experimental data contains 1,888,423 words of 5051 English posts. Through the utilization of NLP techniques, including emotion classification, topic modeling, and semantic role labeling, the study revealed a significant prevalence of negative emotions (94%) such as sadness, anger, and fear in the analyzed content. A correlation between heightened intimacy and an increased occurrence of disagreements was discerned. Furthermore, a combined BERTopic and narrative analysis shed light on the trivial origins of these conflicts. As online communities become increasingly instrumental in people's daily lives, this paper emphasizes their potential in providing invaluable insights to psychologists and sociologists, thereby enabling the formulation of effective strategies and interventions. We recommend educational endeavors to nurture empathy and promote effective communication.

Keywords: language models; subreddit; mental distress; relationship; event analysis

1. Introduction

With high pressure from work, family, and school, a large number of people are stuck in psychosocial and mental health problems, which has been a concerning agenda in society. In recent years, the pandemic has amplified the effects of mental health difficulties, such as suicidal thinking, severe depression, and self-harm behaviors (Salimi, et al. 2023). In this article, we adopt the term mental distress to capture both mental disorders, high-pressure, and related symptomatology. Social media has inevitably evolved into an outlet for many individuals to express their suppressed emotions (Naslund, et al. 2020). Consequently, we witness numerous people online venting about their work-related stress and emotional challenges to find solace on social media; however, it conversely leads to the further pervasion of anxious sentiments because it is very likely that many people share similar experiences with others. Also, some extreme individuals may resort to sharing posts laden with racial (Thomas, et al. 2023) and gender (Chen, et al. 2022) discrimination as an outlet for their stress. Social media seems like a crucible containing different societal emotions where people share their current feelings with others on the platform at any time. As to an individual, diachronically, posts are a reflection of emotional changes within a time span. In the context of a particular event, social media posts serve as manifestations of the attitudes and viewpoints held by users. Therefore, the longitudinal and vertical study on emotions is of significance in monitoring mental health.

Considering its considerable research importance, scholars have investigated the impact of social media from various perspectives, including psychological analysis (Coyne, et al. 2020), adolescent growth (O'reilly 2020), and detection techniques (Chancellor and Choudhury, 2020; Kabir, et al. 2023), to name just a few. Researchers specializing in NLP dedicated considerable efforts to extracting important information by behavioral and linguistic cues from words. By using text mining approaches, we can predict the presence of mood distress, such as psychological stress states detection (Lin et al. 2017). Since 2017, amid the ongoing research on mental health via advanced techniques, researchers have become aware that NLP methods can be leveraged to analyze, predict, and timely prevent mental diseases (Garg 2023).

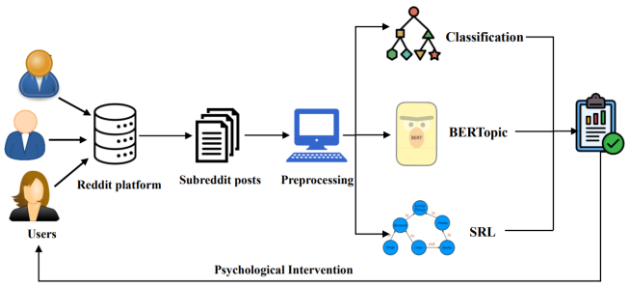

Figure 1: Pipeline of mental health detection

This study aims to explore the significant role of language in detecting mental distress. We have focused on a particular subreddit, *r/AmITheAsshole*, a platform where people frequently share personal experiences. These posts provide valuable insights as users often narrate their stories and can be assessed and appraised by other users. Consequently, the content and associated evaluations within such specialized subreddits can serve as invaluable datasets for the development of a classification model (Efstathiadis, et al. 2021; Haworth, et al. 2021). Such kind of model, once refined, holds potential utility in the context of judicial judgment (Jiang, et al. 2020), While exploring the selective sharing of information and moral judgment

is an interesting area for future research, our main goal in this study is to examine the storytelling aspect. We aim to analyze the root causes of mental pressure and the types of events that occur online through this case study.

We ground this analysis in theoretical approaches that focus on negative sentiment and its underlying sources (O'dea, et al. 2017; Vedula and Parthasarathy 2017), particularly stress expressed in posts, and investigate the following research questions:

RQ1: How frequently are different emotions including joy, sadness, anger, fear, love, and surprise? Are the posts with negative emotions correlated to mental distress?

RQ2: To what extent do individuals encounter distressing situations in various types of relationships? What are the attributed identities of individuals who experience such distress?

RQ3: Which specific events or circumstances are associated with the trouble individuals encounter?

To investigate these research questions, we initially undertook the training of a classification model, using DistilBERT as the foundational model architecture. Then, we calculated the frequency of each emotion, with the objective of conducting a preliminary assessment concerning the presence of negative emotions. We confirm that the prevalence of negative emotions is notably elevated, given that the subreddit *r/AmItheAsshole* primarily serves as a platform for individuals to seek solace and subject themselves to external evaluation (connected to **RQ1**). For **RQ2**, we use semantic role labeling techniques to extract the events and entities in order to find the conflict events and relationships. Conflict is a normal, inevitable part of any relationship. In the context of family and society, it seems that when people are closer, they may experience more quarrels and conflicts. This can be attributed to various factors, such as poor communication, lack of understanding, and unresolved issues. A study on communication between spouses during forced self-isolation found that the features of communication between spouses affected the degree of constructiveness of marital relationships. In stressful situations, disputes and unsolvable conflicts may arise, leading to quarrels between spouses and other family members (Sorokoumova, et al. 2020). We assume that interpersonal relationships characterized by intimacy, such as those between parents and children, friends, and spouses, are more likely to exhibit occurrences of conflict and discord. This potential proclivity towards relational conflict is anticipated to manifest and be discernible within the content of subreddit posts. Furthermore, we have employed topic modeling techniques, a valuable method for the examination of prevalent topics within a corpus. This approach gives us insights into the nature of events and the central focal points of the arguments (**RQ3**). This combined analytical framework enables us to gain a comprehensive understanding of the specific incidents or situations encountered by the authors.

The significant contributions of this paper can be summarized as:

(1) **new datasets:** While it is worth noting that prior research endeavors have engaged with the subreddit *r/AmItheAsshole* (Haworth et al., 2021; Giorgi et al., 2023), their targets are predominantly the development of classification models. To address the specific aims of our study, we undertook web scraping methods to acquire our datasets which are useful for psychologists and sociologists.

(2) **mental distress research:** Our research holds significant relevance and contributes to a nuanced understanding of emotional well-being, interpersonal dynamics, and the ways in which individuals navigate and cope with distressing situations online.

(3) **usage of NLP methods for social science study:** From a methodological perspective, our research carries significant implications of large language model techniques, such as BERTopic, BERT classification (specifically, DistilBERT), and semantic role labeling. Our analysis can provide social scientists with an insight into the usage of NLP techniques. Our codes of our research have been uploaded to GitHub[1].

2. Related Work

2.1 Mental Health Study in Social Media

Reddit, as an open-source platform where users can freely publish and comment on posts, has up to 13B+ posts and comments in 2022 [2] . Some subreddits address mental issues, such as anxiety, depression, and thoughts of suicide (Skaik and Inkpen 2020), mainly including *r/Anxiety*, *r/Bipolar*, and *r/Depression*. In these mental health-related subreddits, users can find a supportive environment where they can express feelings, ask questions, and receive guidance from other users who may have faced similar struggles. This critical situation underscores the need for mental intervention. For better study the mental health issues, some scholars (Losada and Crestani 2016; Yates, et al. 2017; Thorstad and Wolff 2019) created datasets containing subreddits full of depression and negative emotions. Inspired by previous research, we created our dataset as well.

2.2 Large Language Models in Mental Health Analysis

A language model is a statistical representation of a natural language that calculates the likelihood of a sequence of words, on the basis of corpora in one or multiple languages it was trained on. There are some commonly used language models in NLP that have gained widespread popularity and recognition, such as the N-gram model and skip-gram model. These language models are designed to understand

[1] https://github.com/xxx
[2] https://www.redditinc.com/

and generate human-like text, making them valuable tools for applications in psychological and sociological studies.

Nowadays, Large language models (LLMs) are very popular. LLMs are deep learning models trained on large amounts of text data enabled by AI accelerators thereby large in size. LLMs, such as GPT-4 (Bubeck, et al. 2023), PaLM (Chowdhery, et al. 2022), FLAN-T5 (Chung, et al. 2022), have shown impressive capabilities in understanding and producing text similar to human language. A noteworthy trend in recent years that emerged in NLP is contextualized pretrained models, especially transformer-based language models. This trend has gained immense popularity and widespread acclaim for its notable efficiency gains, which enable substantial savings in both time and resources. The remarkable work on the pretrained model is BERT by Devlin et al. (2019). Further, based on the architecture of BERT, several variants of BERT applied in the domain of biomedicine and clinics have been developed and released. The seminal ones include biomedical BERT (Lee, et al. 2020), clinical BERT (Alsentzer, et al. 2019; Huang, et al. 2019), and MentalBERT (Ji, et al. 2021; Xu, et al. 2023). Also, recently Greco et al. (2023) provided a detailed literature review on the usage of Transformer-based language models in mental health and found that "most of the mental health tasks are reduced to a classification problem". This study perpetuates the way previous research has paved and focuses on distressed emotions.

However, there is still a need for a more detailed analysis of mental distress. The gap our study intends to bridge is analyzing the emotions, relations, and issues by leveraging pre-trained language models. Theoretically, by analyzing user-generated content from the subreddit *r/AmItheAsshole*, the study offers valuable insights into how individuals express and cope with emotional distress. Practically, by identifying prevalent emotional states and distressing situations within the online community, this research can inform mental health professionals and support organizations about the specific challenges individuals face. Insights gained from this study could lead to more targeted and effective interventions for individuals experiencing emotional distress.

3. Methods

3.1 Subreddit Data Collection

3.1.1 r/AmItheAsshole[3]

"*r/AmItheAsshole*" is a popular subreddit within the online platform Reddit that serves as a dynamic and participatory space for individuals to seek ethical assessments of their actions and behaviors in various social scenarios. Believing that they have the moral high ground, users of this subreddit share real-life anecdotes, situations, or dilemmas they have encountered, and subsequently inquire whether their actions were morally appropriate or objectionable (Giorgi, et al 2023). The community engages in discussions where members provide opinions, judgments, and perspectives on the presented scenarios, evaluating the behavior of the redditor and determining if they were "the asshole" or not. Attributed to the anonymity, subreddit users on *r/AmItheAsshole* are liberated from the constraints of ethical or moral considerations, thereby fostering an environment conducive to the unfiltered sharing of authentic personal narratives. Moreover, the narratives featured on the *r/AmItheAsshole* platform frequently contain intricate and multifaceted interpersonal relationships. By scrutinizing these narratives, we are able to dissect and scrutinize the nuanced emotional triggers that emanate from these complex human relationships. Collectively, the aforementioned features underscore the reasoned basis for selecting posts from the *r/AmItheAsshole* subreddit as a dataset.

3.1.2 Scraping via *asyncpraw*

We utilized the *asyncpraw* library to collect our data. This is an asynchronous interaction with the Reddit API. Posts are acquired based on attributes such as title, score, unique identifier, subreddit, URL, number of comments, body content, and timestamp of creation.

In light of the constraints associated with limited data acquisition capabilities, we curated the dataset over two months. Inevitably, our scraping efforts yielded some instances of duplicated data. To address this, we adopted a methodological strategy grounded in the comparison of post titles. Posts sharing identical titles were deemed duplicates, and as such, were excluded from the dataset. Our final dataset comprises 1,888,423 words from 5051 records, each of which has been verified as a valid and reliable source for our research endeavors.

column name	denotation
title	title of a post
id	every post has its own id
subreddit	the subreddit that a post belongs to
url	a permalink pointing to a post
num_comments	the number of comments
body	main content of a post

Table 1: Column names with their denotation

3.2 Fine-tuning Language Models

3.2.1 Emotion Classification Model

Our emotion classification model was inspired by Tunstall, et al. (2022: 21-54) where the foundation model is a DistilBERT-uncased model specifically tailored and fine-tuned for the nuances inherent to the emotion dataset. Our dataset for fine-tuning is from an article (Saravia, et al. 2018) that explored how emotions are represented in English Twitter messages. The dataset is randomly divided into training (16000 examples), validation (2000 examples), and test sets (2000 examples) with a ratio of 0.8:0.1:0.1. The hyperparameters include a learning rate of 2e-05, training and evaluation batch sizes of 64, utilization of the Adam optimizer with

[3] https://www.reddit.com/r/AmItheAsshole/

betas set at (0.9, 0.999) and epsilon of 1e-08, a linear learning rate scheduler, and a training duration spanning 2 epochs. Our fine-tuned model accepts a sequence of BPE tokens (Sennrich, et al. 2016) using the Sentence Piece model (Kudo and Richardson 2018) as an input. After training we used the fine-tuned model to detect emotions for all scraped posts.

3.2.2 Narrative Model

Narratives, whether in the form of stories, literature, or even personal anecdotes, hold significant cultural, psychological, and informational value. Rich information is hidden in the character development and plot progression. With the assistance of Relatio (Ash, et al. 2022) for narrative extraction and its capacity to facilitate result visualization through graphs, we have elected to employ this tool within the scope of our research. The Relatio begins with a plain-text corpus as input and segments a paragraph into sentences. These segmented sentences are then processed through three key stages: named entity recognition, semantic role labeling, and the phrase embedding model. The outputs from these initial operations include identified named entities, annotated semantic roles, and a phrase embedding model for phrase vectorization. In the subsequent steps, the pipeline tags the named entities and finalizes roles that contain named entities for output. Roles without named entities are vectorized using the phrase embedding model and are subsequently subjected to K-means clustering. This clustering process generates clustered entities. Finally, the pipeline constructs the final narrative statements by incorporating the named entities, verbs, and clustered entities.

3.2.3 Topic modelling with BERTopic

With the increasing popularity of transformer models in NLP, as aforementioned, Devlin et al. (2019) proposed BERT, which has shown good performance in encoding language. Furthermore, Grootendorst (2022) proposed BERTopic. BERTopic has an advantage over other topic models, especially in the aspects of coherence and diversity. Given BERTopic's strength in the automatic visualization and detection of the topic hierarchy (Egger and Yu 2022), we selected BERTopic instead of Top2vec or LDA. BERTopic generates topic representations in three steps: converting each document to its vector representation using a pre-trained language model, reducing the dimensionality of the vectors to optimize the clustering, and finally by means of the TF-IDF algorithm extracting topic representations from document clustering. In the first step, document embeddings are created through the utilization of the Sentence-BERT (SBERT) framework (Reimers and Gurevych 2019). Certainly, alternative embedding techniques can be employed, but we opted for the SBERT framework due to its ability to attain state-of-the-art results in Semantic Textual Similarity (STS) tasks and its strong computational efficiency in clustering. Next, we employ the UMAP (Uniform Manifold Approximation and Projection) algorithm to reduce the dimensionality of the output produced by SBERT. To cluster the dimensionally reduced document vectors, we use the HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) algorithm which is particularly useful when dealing with complex and unknown datasets. The final step in implementing BERTopic involves creating topic representations. Each cluster extracted by BERTopic represents a distinct topic, except outlier observations that are not assigned to any cluster. Within the same cluster, text documents are presumed to exhibit semantic similarity to one another. To identify topics from topic representations, BERTopic modified the classical TF-IDF "such that it allows for a representation of a term's importance to a topic instead" (Grootendorst 2022). The local representation of each topic is:

$$W_{t,d} = tf_{t,d} \cdot \log(1 + \frac{A}{tf_t}) \quad (1)$$

where the frequency of term *t* appearing in document *d* is represented as $tf_{t,d}$.

Pruning can be an alternative step wherein we can eliminate similar and redundant words from a given topic by employing the Maximal Marginal Relevance (MMR) technique on the top *n* words within that topic. This step serves to enhance the interpretability of the topics, and therefore, we incorporate this method into our topic processing pipeline. It is worth noting that to derive meaningful topic names, we use OpenAI API to help us generate names based on the keywords of each topic through prompting. The prompt is as follows:

> I have a topic that contains the following documents: [DOCUMENTS] The topic is described by the following keywords: [KEYWORDS] Based on the information above, extract a short but highly descriptive topic label of at most 5 words. Make sure it is in the following format: topic: <topic label>.

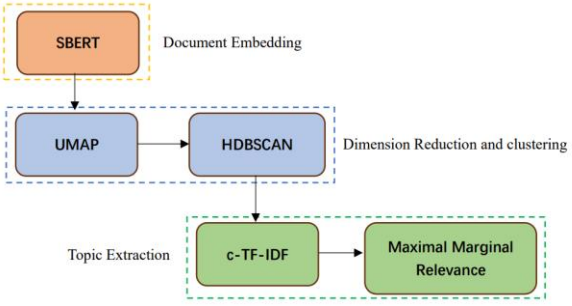

Figure 2: Flowchart of BERTopic: There are three steps in BERTopic pipeline, including document embedding, dimension reduction and clustering, and topic extraction and representation.

4. Results

In the following subsections, we delve into a detailed interpretation of the results obtained from each of the models discussed in the methodology section. Section 4.1 offers a comprehensive presentation of the detailed results derived from the fine-tuned BERT classification model. This model is specifically designed to address the question of whether

Redditors on the subreddit were experiencing mental distress issues (**RQ1**). Sections 4.2 and 4.3 show the results of BERT topic modeling and narrative modeling respectively. These approaches aim to provide insights into the types of events and relationships individuals with troubles are engaged in (**RQ2** and **RQ3**).

4.1 Emotion Analysis

Our dataset to fine-tune the DistilBERT is the "emotion" dataset downloaded from the Hugging Face Hub. This dataset contains six emotions, namely, joy, sadness, anger, fear, love, and surprise. Fine-tuning this model took around 15 hours and 34 minutes on the 11th Intel Core i5 CPU. In our test dataset, we plotted a confused matrix to visualize the accuracy. It shows high accuracy except for the label of surprise. The reason is that surprise is frequently mistaken for joy, or confused with fear.

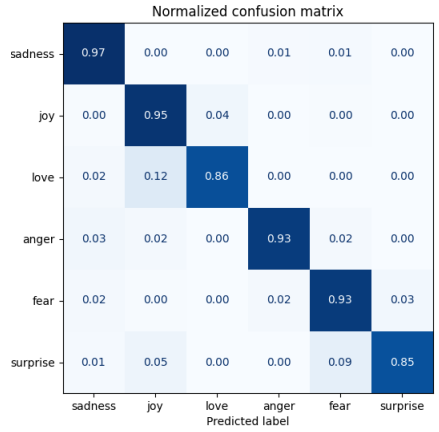

Figure 3: Confusion matrix in our test dataset

The F1 score is more than 93%. To verify the performance of this model, we compared our model with other common emotion classification models using machine learning and deep learning methods, including traditional machine learning models such as rule-based keyword searching (LIWC), Naïve Bayes, and deep learning models, such as LSTM, BERT, and RoBERTa. In the rule-based model, we created word lists related to different emotions, such as sadness (sad, grief, etc.) In Naïve Bayes, we did a multi-class classification based on a N-gram language model.

We observed that our model has the best performance in precision, F-measure, and accuracy.

Methods	Precision	F-score	Accuracy
LIWC	0.53	0.69	0.60
Naïve Bayes	0.49	0.55	0.46
LSTM	0.72	0.77	0.65
BERT	0.87	0.82	0.74
RoBERTa	0.89	0.83	0.90
Fine-tuned DistilBERT	**0.90**	**0.94**	**0.92**

Table 2: The results on the test set by different text classification methods

The classification results are as follows (See Table 3). We found a substantial portion, comprising 2,539 instances, is labeled with the emotion of anger. The emotion of fear is associated with 1,721 instances within the dataset, the second high-proportion emotion. There are 521 instances expressing the emotion of sadness. These three negative emotions account for 94.65% of the whole dataset, which, to some extent, indicates that subreddit users might have mental distress issues. In addition, 243 instances within the dataset convey the emotion of surprise. The joy emotion only has 15 instances. A relatively small number of instances, specifically 12, express the emotion of love.

Emotion Class	Frequency	Example
Anger	2539	We noticed behavior from my aunt, we didn't approve of - speaking a bit passive aggressively to/about service workers…My husband and I got angry and stressed. We vented to each other in my husband's language while walking back to the hotel.
Fear	1721	Normally I'm not the lying tricking type at all. But it scares me that I'll be so selfish just to keep my cat. Would I be the asshole for not calling them now?
Sadness	521	Right now I've just been slowly ghosting to keep myself from getting hurt more. I don't see a future where we reconnect the way we were, and this half-friendship and slow deterioration feels awful.
Surprise	243	I was surprised as I had'nt ever noticed inappropriate behavior from him, he was very devoted to his girlfriend and never showed any sign of liking me in another way.
Joy	15	I keep telling her that I'm glad we had a wedding and that it was really cool for what it was. […] I greatly preferred the rehearsal dinner.
Love	12	My husband loved it but I"e never told him it's not a random name I came up with. I love my husband and I obviously don't love my "x" just appreciate him, as he was a big part of my teenage life.

Table 3: Distribution and real-world examples of six emotions: joy, sadness, anger, fear, love, and surprise

We randomly selected 100 real-world examples for analysis, seeking to identify mental distress as reflected. We asked two annotators to independently

review these examples, with the instruction to annotate each example with a "1" if it was perceived to indicate mental distress. To assess the reliability of the annotations and the level of agreement between the annotators, we employed Gwet's AC1 (Gwet 2014) instead of other common agreement methods for the reason that Gwet's AC1 was shown to have "a more stable inter-rater reliability coefficient than Cohen's Kappa" (Wongpakaran, et al. 2013). The results revealed that the agreement scores for sadness, anger, and fear were 0.745, 1, and 1, respectively. These high levels of agreement highlight the consistency and reliability of the annotations. Further, we compared the proportion of texts annotated as "distress" to the total number of texts (100 examples) and found that in all three random-selected test sets, texts classified as "distress" constituted over 98% of the total. This finding significantly illustrates that the subreddit *r/AmITheAss*hole is replete with instances of distress. The connection between negative emotion and distress is evident, corroborating previous related research in this field (Chaurasia, et al. 2021; Shi, Y., et al. 2023)

Emotion	Gwet's AC1	Annotator A (distress%)	Annotator B (distress %)
Sadness	0.745	98%	99%
Anger	1.000	100%	99%
Fear	1.000	98%	98%

Table 4: Results of agreement between two annotators and the proportion of texts annotated as "distress" to the total number of 100 examples

4.2 Topic Analysis

BERTopic yields 60 topics, and their spatial distribution in a 2D representation is visually depicted in Fig. 4. This figure illustrates the topic assignments for each document within our dataset, as determined by the BERTopic algorithm. Each document is linked to one or more topics, indicating the extent of its relevance to those topics. Additionally, the proportions of a document's association with different topics are visually represented through the use of distinct colors, quantifying the strength of the document's connections to each topic. With each color corresponding to a unique topic, this approach facilitates the analysis of content within the same topic, aiding in our understanding of the topic's implications and connotations. Furthermore, it assists in the process of assigning meaningful names to each topic. As a result, the distribution of documents across topics stands as a pivotal component of BERTopic-based topic modeling, facilitating the interpretation of generated topics and their alignment with the content of our documents.

The topic distance map (upper) with the topic clustering (lower) results are shown in Fig. 4, we are able to scrutinize the content of topics and identify overarching topics prevalent in subreddit posts. This approach enhances our ability to detect underlying trends and dominant topics in subreddit discussions.

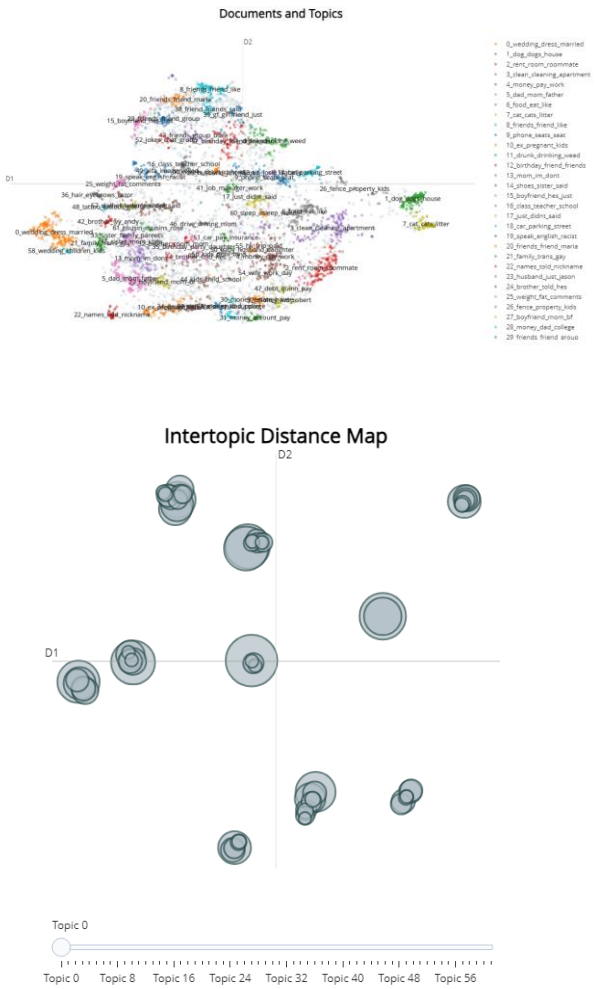

Figure 4: Upper: Documents distribution: Different colors denote different topics and each data point is one paper Lower: It shows the clustering results of topics, each of them is one topic.

The details of each topic and examples are in Table 5. Topics 46, 18, and 51 pertain to various aspects of car driving and associated payments. These discussions revolve around topics such as the eligibility of Redditors to operate and park vehicles, as well as the monthly costs of car insurance. These topics involve motherhood. Topics 53, 55, and 40 are about tips and bills. The typical scenario is the payment of tips and bills for a party between friends. Topics 60, 61, 48, 36, 57, and 49 are centering on arguments related to appearance, such as tattoos, hair cutting, or clothes painting, to name just a few. The predominant theme recurring in these topics primarily revolves around the relationship between parents and children. Topics 54, 44, 34, 35, 50, 56, and 58 focus on child-rearing, such as picking up kids after school, celebrating children's birthday, and so forth. Therefore, the relationships in these topics are couples mainly including wife and husband. The overarching theme that prominently emerges across a multitude of topics (specifically, topics 52, 25, 39, 19, 16, 9, 41, 23, 0, 12, 11, 17, 15, 38, 29, 8, 20, 43) is that of friendship, including various facets of shared experiences and events with friends, such as friends' wedding, birthday, party, etc. Heed that

friends' events serve as a blasting fuse for an argument thus we can notice topics involving relationships such as those related to husbands (Topic 23). Topics 42, 32, 22, 27, 13, 5, 10, 14, 21, 24, 33, and 45 are concerning family-related discussions, with a particular focus on the dynamics between parents and siblings, such as kids' pregnant, gay brother, and so on. The typical relations are parenthood and siblings. The last cluster is associated with house renting and family spending. From the perspective of relationships, basically, house renting and family spending are topics that family members talk about. The keywords confirm that, such as "mom pay work", "money account pay" and "child support".

Event name	Topic involved	Example
car	46, 18, 51	The other reasons I didn't want to drive are that my mother has been constantly complaining about my driving for the last few days
tips and bill	53, 55, 40	The bill came. It was 770Euros? I told my friend we should all split evenly, he was cool with it
appearance	60, 61, 48, 36, 57, 49	I can't even talk about the tattoo without her getting upset because in my view, you don't have to be invited to every single thing
children rearing	54, 44, 34, 35, 50, 56, 58	the problem that has been bothering me since our kid is very little is that my husband is quite a strict parent
friendship	52, 25, 39, 19, 16, 9, 41, 23, 0, 12, 11, 17, 15, 38, 29, 8, 20, 43	normally update my friends on my group chat letting them know whether or not I'll be going to school that day
parents and sibling	42, 32, 22, 27, 13, 5, 10, 14, 21, 24, 33, 45	on new years I guess my brother didn't give his gf enough attention at our family party and she left angry saying he always ignores her etc.
rent and spending	59, 30, 28, 31, 37, 47, 6, 4, 2, 3, 7, 1, 26	I'm in my last year at university and have saved up pounds 1,400 from working and any bits of maintenance money my parents let me have. However we've constantly gotten into disagreements over how much money I can take.

Table 5. Topic classifications and examples

On the basis of the above results of topic modeling, we may conclude that the relationships where Redditors faced mental distress mainly involve children-parents, husband-wife, boyfriend-girlfriend, brother and sister (sibling), and friendship (**RQ2**).

4.3 Narrative Analysis

Our narrative extraction process is enhanced by the utilization of Relatio. Relatio not only facilitates the extraction of events and entities from textual data but also offers a visualization method that enables the clear representation of the relationships between events and entities in the form of a graph. Fig. 5 visualizes the entities and actions in our dataset. The core entities involved are "people", "thing", "cat", "friend", "name" and "mom". Actions between "people" and "things" are verbs that involve communication, such as "tell", "say", "suggest", and "ask". The common thread among these words is their involvement in conveying, expressing, or making known information, whether it's through spoken or written communication or through actions that result in communication. Actions between "mom" and "thing" represent various aspects of human interaction, cognition, and action, whether it is related to communication, planning, emotional states, physical activity, or knowledge. There are some verbs between "mom" and "name" that all represent different actions, states, or processes related to human activities, emotions, cognition, and communication, including manage, put, lose, etc. Verbs between "friend" and "name" are expressing opinions or judgments to engaging in physical or mental activities, such as "criticize", "play", "consider", and "change". The entity "name" here refers to any pronoun or unknown, such as $mom \xrightarrow{lose} name$, which, in original texts, is "my mom lost me and my brother's birth certificate and Social Security Card".

This figure verifies the main relations that subreddit Redditors have trouble with are motherhood, friendship, etc. (**RQ2**). Additionally, we can identify several recurring situations that often lead to disagreements or conflicts, including activities like "cleaning the pet," interactions involving the criticism of friends, or incidents such as "hitting <name>," among others. These are just a few examples, and there exists a broad spectrum of events and circumstances that can potentially trigger arguments or disputes (**RQ3**).

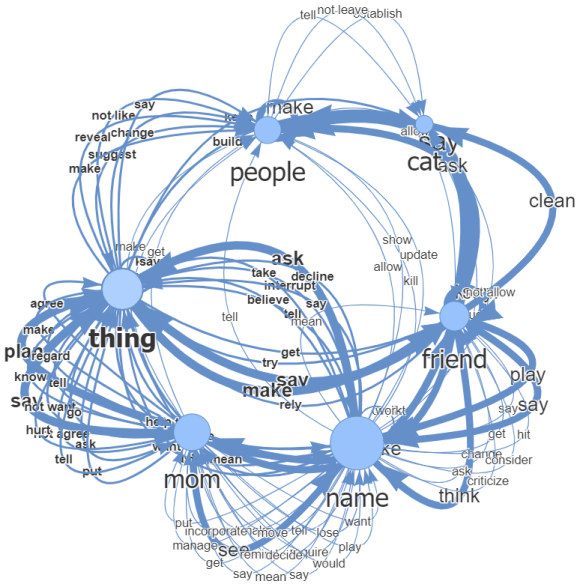

Figure 5: Top 100 most frequent narratives in *r/AmItheAsshole*

5. Discussion and Conclusion

In the preceding sections of our research paper, we introduced three key research questions. We employed language models to perform emotion classification, topic modeling, and semantic role labeling to derive answers to these questions. Our emotion detection results unveiled a predominant presence of negative emotions, including sadness, anger, and fear. After the supervision of two annotators, we derive that users in the *r/AmItheAsshole* subreddit community grapple with various mental distress troubles in their daily lives. This result corresponds to the findings of Choudhury and De (2014), "the subreddit contents shared by anonymous redditors have negative or caustic nature". Moreover, utilizing BERTopic, we discerned that heightened intimacy correlates with a heightened likelihood of encountering frequent disagreements and conflicts. Finally, the combined analysis of BERTopic and narrative modeling revealed that these conflicts often stem from seemingly trivial matters such as childcare duties, pet care, or splitting the cost of tips, to name just a few. The findings align with our initial hypothesis, indicating a prevalence of mental distress issues among participants in the *r/AmItheAsshole* subreddit. We observed the valuable role that online communities can play in shedding light on the psychological troubles individuals suffer from, thereby offering psychologists crucial insights to inform their strategies and interventions. On the other hand, these digital fora and support groups create a supportive environment for individuals to connect with others who share similar struggles, encompassing feelings of distress, pain, and adversity (Eysenbach, et al. 2004). Consequently, these online spaces serve as cost-effective and accessible platforms for seeking guidance and advice related to health challenges.

Our overall goal was to use voluntarily shared posts from *r/AmItheAsshole* users in order to enhance our understanding of mental distress and ultimately provide suitable mental care for those people with troubles. In accordance with our findings, we recommend that educational institutions offer lectures or workshops aimed at fostering empathy among both children and parents. This proactive approach can help mitigate unnecessary conflicts and reduce the incidence of mental health issues within families. Additionally, it is essential to recognize the role of effective communication in conflict resolution (Cupach 1980). We encourage individuals to practice active listening, articulate their thoughts calmly and clearly, refrain from accusatory language, and make a concerted effort to understand the perspectives of others. Then, in relations with friends, family members, and other intimacy, individuals can mitigate numerous conflicts and maintain positive sentiments.

It is important to acknowledge several limitations that may impact the interpretation and generalizability of the findings. We observed issues with the classification model's ability to distinguish between "surprise" and "fear". Upon reviewing the raw dataset, we found that some posts could be categorized into two or more emotion types simultaneously. However, the classification model selects the most probable emotion based on the maximum likelihood, which can lead to an increase in the confusion score. Below is an example.

> She also threatened to message him and tell him all about us. I was confused. [...] I felt awful cus i guess i just didn't see it that way, and now i've lost my best friend and feel like i betrayed them both, but i also feel like i don't owe anybody an explanation seen as im single. I have nobody i can talk to about this because if i do they will figure out that she's gay and i could never do that, so i'm here asking u guys, am i the asshole?

This example contains elements of fear (*threatened*), surprise (*confused*), and sadness (*I've lost my best friend*). The co-occurrence of these emotions in a single narrative poses a challenge for the classification model, highlighting the need for more nuanced approaches in emotion detection, such as aspect-based emotion classification (Brauwers and Frasincar 2022).

Harnessing advanced NLP models has the potential to revolutionize our understanding of psychological phenomena and enhance the depth of insights derived from text-based data in the field. In recent days, innovations such as ChatGPT have captured the public's attention to NLP. ML and NLP methods may sometimes be impressive as tools to support clinical practice and medical research because of their intelligence. However, Language models may lack the ability to fully understand the context and nuances of human communication, leading to incorrect assessments. Mental health assessment often requires understanding beyond the text, including non-verbal cues and broader situational context. Consequently, the accuracy and reliability of such models in complex and nuanced applications remain a subject of ongoing research and development within the NLP community.

In conclusion, this paper explores the usage of large language models as well as semantic role labeling in mental healthcare. The exploration of employing NLP techniques within psychological research remains a compelling avenue for investigation and innovation. This research is just an exploratory analysis of online posts to provide some insights into the NLP usage in mental distress for psychologists and sociologists and wishes related managers can propose effective strategies and interventions.

6. Ethics

The data utilized in this research has been sourced from Reddit, and the user identifications have been made anonymous. Additionally, all the example posts showcased in this study have been altered, rephrased, and anonymized to safeguard the privacy of users and to avoid any potential misuse.

7. Acknowledgments

I would like to express my sincere gratitude to xxx for her invaluable assistance with web scraping codes. Her expertise and dedication significantly contributed to the success of this paper.

8. Appendices

Anonymous

9. Bibliographical References

Alsentzer, E., Murphy, J. R., Boag, W., Weng, W. H., Jin, D., Naumann, T., and McDermott, M. (2019). Publicly available clinical BERT embeddings. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pages 72–78.

Ash, E., Gauthier, G. Philine W. (2022). *Text Semantics Capture Political and Economic Narratives*. [S.l.]: SSRN.

Bonial, C., Hwang, J., Bonn, J., Conger, K., Babko-Malaya, O., and Palmer, M. (2012). English propbank annotation guidelines. *Center for Computational Language and Education Research Institute of Cognitive Science University of Colorado at Boulder*, 48.

Brauwers, G., and Frasincar, F. (2022). A survey on aspect-based sentiment classification. *ACM Computing Surveys*, *55*(4), 1-37.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee P., Lee Y. T., Li Y., Lundberg S., Nori H. Palangi H., Ribeiro M. T., and Zhang Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv preprint arXiv:2303.12712*.

Chancellor, S., and De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine*, *3*(1), 43.

Chaurasia, A., Prajapati, S. V., Tiru, P. A., Kumar, S., Gupta, R., and Chauhan, A. (2021). Predicting mental health of scholars using contextual word embedding. In *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 923-930). IEEE.

Chen, X. Y., Zhuge, Y., Feng, J. S., and Guo, L. K. (2022). Invisible culture dimension of gender discrimination: Speech cyberbullying against women on Chinese social media. In *Computational Social Science* (pp. 87-93). Routledge.

Chowdhery, A., Narang, S., Devlin, J., et al. (2022). Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*.

Chung, H. W., Hou, L., Longpre, S., et al. (2022). Scaling instruction-finetuned language models. *arXiv preprint arXiv:2210.11416*.

Coyne, S. M., Rogers, A. A., Zurcher, J. D., Stockdale, L., and Booth, M. (2020). Does time spent using social media impact mental health?: An eight year longitudinal study. *Computers in human behavior*, 104, 106160.

Cupach, W. R. (1980). Interpersonal Conflict: Relational Strategies and Intimacy. In *the Annual Meeting of the Speech Communication Association*.

De Choudhury, M., Gamon, M., Counts, S., and Horvitz, E. (2013). Predicting depression via social media. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 7, No. 1, pp. 128-137).

De Choudhury, M., and De, S. (2014). Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 71-80).

Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Efstathiadis, I. S., Paulino-Passos, G., and Toni, F. (2022). Explainable patterns for distinction and prediction of moral judgement on reddit. In *1st Workshop on Human and Machine Decisions (WHMD)*.

Egger, R., and Yu, J. (2022). A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Frontiers in sociology*, *7*, 886498.

Garg, M. (2023). Mental health analysis in social media posts: a survey. *Archives of Computational Methods in Engineering*, *30*(3), 1819-1842.

Giorgi, S., Zhao, K., Feng, A. H., and Martin, L. J. (2023). Author as character and narrator: Deconstructing personal narratives from the r/amitheasshole reddit community. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 17, pp. 233-244).

Greco, C. M., Simeri, A., Tagarelli, A., and Zumpano, E. (2023). Transformer-based language models for mental health issues: A survey. *Pattern Recognition Letters*, *167*, 204-211.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC.

Haworth, E., Grover, T., Langston, J., Patel, A., West, J., and Williams, A. C. (2021). Classifying reasonability in retellings of personal events shared on social media: A preliminary case study with/r/amitheasshole. In *Proceedings of the*

International AAAI Conference on Web and Social Medi (Vol. 15, pp. 1075-1079).

Huang, K., Altosaar, J., and Ranganath, R. (2019). Clinicalbert: Modeling clinical notes and predicting hospital readmission. *arXiv preprint arXiv:1904.05342*.

Ji, S., Zhang, T., Ansari, L., Fu, J., Tiwari, P., and Cambria, E. (2021). Mentalbert: Publicly available pretrained language models for mental healthcare. In *Proceedings of the Language Resources and Evaluation Conference (LREC)*.

Jiang, Z. P., Levitan, S. I., Zomick, J., and Hirschberg, J. (2020). Detection of mental health from reddit via deep contextualized representations. In *Proceedings of the 11th international workshop on health text mining and information analysis* (pp. 147-156).

Kabir, M., Ahmed, T., Hasan, M. B., Laskar, M. T. R., Joarder, T. K., Mahmud, H., and Hasan, K. (2023). DEPTWEET: A typology for social media texts to detect depression severities. *Computers in Human Behavior*, 139, 107503.

Kudo, T. and Richardson, J. (2018). SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 66–71. Brussels, Belgium: Association for Computational Linguistics.

Le Glaz, A., Haralambous, Y., Kim-Dufor, et al. (2021). Machine learning and natural language processing in mental health: systematic review. *Journal of Medical Internet Research*, *23*(5), e15708.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., and Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, *36*(4), 1234-1240.

Lin, H., Jia, J., Qiu, J., Zhang, Y., Shen, G., Xie, L., Tang, J., Feng, L. and Chua, T.S. (2017). Detecting stress based on social interactions in social networks. *IEEE Transactions on Knowledge and Data Engineering*, 29(9), 1820-1833.

Losada, D. E., and Crestani, F. (2016). A test collection for research on depression and language use. In *International conference of the cross-language evaluation forum for European languages* (pp. 28-39). Cham: Springer International Publishing.

Naslund, J. A., Bondre, A., Torous, J., and Aschbrenner, K. A. (2020). Social media and mental health: benefits, risks, and opportunities for research and practice. *Journal of Technology in Behavioral Science*, 5, 245-257.

O'dea, B., Larsen, M. E., Batterham, P. J., Calear, A. L., and Christensen, H. (2017). A linguistic analysis of suicide-related Twitter posts. *Crisis*. 38(5), 319–329.

O'reilly, M. (2020). Social media and adolescent mental health: the good, the bad and the ugly. *Journal of Mental Health*, *29*(2), 200-206.

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

Salimi, N., Gere, B., Talley, W., and Irioogbe, B. (2023). College students mental health challenges: Concerns and considerations in the COVID-19 pandemic. *Journal of College Student Psychotherapy*, *37*(1), 39-51.

Saravia, E., Liu, H. C. T., Huang, Y. H., Wu, J., and Chen, Y. S. (2018). Carer: Contextualized affect representations for emotion recognition. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 3687-3697).

Sennrich, R., Haddow, B., and Birch, A. (2015). Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), 1715–1725. Berlin, Germany: Association for Computational Linguistics.

Shi, Y., et al. (2023). Detect Depression from Social Networks with Sentiment Knowledge Sharing. *arXiv preprint arXiv:2306.14903*.

Skaik, R., and Inkpen, D. (2020). Using social media for mental health surveillance: a review. *ACM Computing Surveys (CSUR)*, *53*(6), 1-31.

Sorokoumova, E. A., Matveeva, N. E., Cherdymova, E. I., Puchkova, E. B., Temnova, L. V., Chernyshova, E. L., and Ivanov, D. V. (2020). Features of communication between spouses during long-term forced self-isolation as a factor of constructive marital relationships. *EurAsian Journal of BioSciences*, *14*(2).

Stevens, H. R., Acic, I., and Rhea, S. (2021). Natural language processing insight into LGBTQ+ youth mental health during the COVID-19 pandemic: Longitudinal content analysis of anxiety-provoking topics and trends in emotion in LGBTeens microcommunity subreddit. *JMIR public health and surveillance*, *7*(8), e29029.

Thorstad, R., and Wolff, P. (2019). Predicting future mental illness from social media: A big-data approach. *Behavior research methods*, *51*, 1586-1600.

Thomas, A., Jing, M., Chen, H. Y., and Crawford, E. L. (2023). Taking the good with the bad?: Social

Media and Online Racial Discrimination Influences on Psychological and Academic Functioning in Black and Hispanic Youth. *Journal of youth and adolescence*, 52(2), 245-257.

Tunstall, L., Von Werra, L., and Wolf, T. (2022). *Natural Language Processing with Transformers*. O'Reilly Media, Incorporated.

Vedula, N., and Parthasarathy, S. (2017). Emotional and linguistic cues of depression from social media. In *Proceedings of the 2017 International Conference on Digital Health* (pp. 127-136).

Wongpakaran, N., Wongpakaran, T., Wedding, D., and Gwet, K. L. (2013). A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC medical research methodology*, *13*, 1-7.

Xu, X., Yao, B., Dong, Y., Yu, H., Hendler, J., Dey, A. K., and Wang, D. (2023). Leveraging Large Language Models for mental health prediction via online text data. *arXiv preprint arXiv:2307.14385*.

Yates, A., Cohan, A., and Goharian, N. (2017). Depression and self-harm risk assessment in online forums. *arXiv preprint arXiv:1709.01848*.