# Inferring Mental Burnout Discourse Across Reddit Communities

**Nazanin Sabri§, Anh C. Pham†, Ishita Kakkar†, Mai ElSherief‡**
nsabri@ucsd.edu, acpham@umass.edu, ikakkar@umass.edu, m.elsherif@northeastern.edu
§: University of California San Diego (UCSD)
†: University of Massachusetts at Amherst
‡: Northeastern University

## Abstract

Mental burnout refers to a psychological syndrome induced by chronic stress that negatively impacts the emotional and physical well-being of individuals. From the occupational context to personal hobbies, burnout is pervasive across domains and therefore affects the morale and productivity of society as a whole. Currently, no linguistic resources are available for the analysis or detection of burnout language. We address this gap by introducing a dataset annotated for burnout language. Given that social media is a platform for sharing life experiences and mental health struggles, our work examines the manifestation of burnout language in Reddit posts. We introduce a contextual word sense disambiguation approach to identify the specific meaning or context in which the word "burnout" is used, distinguishing between its application in mental health (e.g., job-related stress leading to burnout) and non-mental health contexts (e.g., engine burnout in a mechanical context). We create a dataset of 2,330 manually labeled Reddit posts for this task, as well as annotating the reason the poster associates with their burnout (e.g., professional, personal, non-traditional). We train machine learning models on this dataset achieving a minimum F1 score of 0.84 on the different tasks. We make our dataset of annotated Reddit post IDs publicly available to help advance future research in this field.

## 1 Introduction

As a response to chronic interpersonal stressors, burnout syndromes develop through latent psychological erosion. At the individual level, burnout manifests in cardiovascular, mental, and physical problems, such as headaches, chronic fatigue, gastrointestinal disorders, and more (Schaufeli and Buunk, 1996; Chutko et al., 2019). Within workplaces, burnout can create interpersonal conflict and disrupt team productivity, resulting in the propagation of burnout and a greater rate of job turnover (Maslach and Leiter, 2016a). By conservative estimates, burnout among physicians alone costs the United States $4.6 billion annually (Han et al., 2019). Considering the negative affects of burnout, the World Health Organization has recognized it as an occupational phenomenon in the International Classification of Diseases (ICD-11) (Organization, 2019). Although it was initially believed that burnout only occurs in human-centered jobs (Weber and Jaekel-Reinhard, 2000), it has since been shown to develop among all professional (Edú-Valsania et al., 2022), non-professional (e.g., parental (Mikolajczak et al., 2019, 2021), e-sports athlete (Hong et al., 2022)) and historically marginalized groups (e.g., individuals with autism (Mantzalas et al., 2022)).

While questionnaires (e.g., Maslach Burnout Inventory (MBI) (Maslach and Leiter, 2016b)) are commonly used for the detection of burnout, they have limitations in terms of accessibility and scalability. Thus the lack of large-scale studies on burnout is a gap in the literature. We address this gap by creating a linguistic resource for the study of burnout. Developing linguistic resources for the language of burnout is crucial for several reasons. Firstly, it enables the creation of more accurate NLP models that can identify and assess burnout-related language, facilitating early detection and intervention, which in turn provides better support for those experiencing burnout. Additionally, these resources allow for a deeper analysis of how burnout is discussed across different contexts helping to understand its broader impact. By supporting cross-disciplinary research, linguistic resources also bridge fields like psychology, medicine, and occupational studies, leading to more comprehensive insights into burnout. Linguistic resources allow for the systematic collection and analysis of burnout-related language, leading to data-driven insights. These insights can inform public policy, workplace practices, and mental health in-



*Proceedings of the Third Workshop on NLP for Positive Impact*, pages 224–231
November 15, 2024 

terventions, ultimately contributing to better societal outcomes. However, if we attempt to collect data using burnout-related keywords, the resulting dataset will likely include instances from both mental health contexts (e.g., emotional exhaustion in the workplace) and non-mental health contexts (e.g., burnout in a physical or mechanical sense), due to the term's varied usage across different domains. This emphasizes the need for effective word sense disambiguation to accurately interpret the context in which "burnout" is used.

Our paper leverages discourse on Reddit surrounding mental burnout to infer instances of self-disclosure of burnout and the context they occur in. We collect Reddit posts across all subreddits that contain burnout-related keywords[1] for a nine year period (2014-2022). After cleaning[2], our dataset consists of 297,623 posts from across 23,519 subreddits. We then annotate 2,330 randomly sampled posts for disclosure and context[3] of burnout.[4] Following the annotation of posts, we build and evaluate models to detect whether burnout keywords are used in a mental health setting, and the context of burnout. Our best models for both classification tasks achieve an F1 score of 0.84 or higher. Our dataset will be made publicly available for use by the research community.[5] Deploying our trained models on the full Reddit collection, we observe an overall increase in the online disclosure of burnout. We also find that burnout attributed to issues outside of work makes up a considerable portion of the online discourse, indicating a need for more studies in non-professional settings.

## 2 Related Work

**Burnout** is defined by its three dimensions of exhaustion, cynicism, and professional inefficacy (Maslach and Leiter, 2016b). While majority of psychological studies have focused on occupational burnout, emerging research has also evaluated burnout in other populations such as parents (Mikolajczak et al., 2021). To the best of our knowledge, only one more study has investigated burnout using social media data (Wu et al., 2021). Studying posts by 1,532 burnt-out Weibo users, Wu, Ma, Wang, and Wang (2021) predicted burnout using user posting behavior such as time and interaction patterns. While Wu, Ma, Wang, and Wang (2021) focused on extended activity by a set of users and examined changes in behavior before and after their bursts of burnout, our work focuses on self-disclosure of burnout. Focusing on the specific posts in which users describe their experience of burnout allows us to understand burnout risk factors and user needs that might not be evident when posting behavior is studied. Additionally, the aforementioned study focuses on job burnout, while we consider all aspects, occupational and non-occupational of the burnout experience.

**Mental Health & Social Media.** Prior work has leveraged social media data to computationally predict mental health status and improve mental health outcomes of at-risk individuals (Chancellor and De Chourdhury, 2020).

Closely related to our work, Saha and De Choudhury (2017), Saha, Kim, Reddy, Carter, Sharma, Haimson, and De Choudhury (2019), and Cascalheira, Hamdi, Scheer, Saha, Boubrahimi, and Choudhury (2022) assess the self-disclosure of stress to develop a greater understanding of stigmatized topics within online discourse. More specifically, Saha, Kim, Reddy, Carter, Sharma, Haimson, and De Choudhury (2019) developed a machine learning classifier to scalably identify social media posts describing minority stress experiences, achieving an AUC of 0.80. Saha and De Choudhury (2017) is another example which examined expressions of stress from survivors of gun violence on college campuses within online discourse.

Our work contributes to the body of research on inferring mental health from language by developing a dataset and models for the detection of burnout and its context. Although chronic stress precedes and contributes to burnout syndrome (Maslach and Leiter, 2016b), we note that the general form of burnout is defined as a psychological reaction triggered by perceived environmental demands, while stress is more unspecific in origin. Prior work has specifically cautioned against mixing stress and burnout (Schaufeli and Buunk, 1996).

## 3 Methodology

**Data Collection & Cleaning.** The social media platform Reddit affords users a degree of pseudo-anonymity and allows longer posts relative to other platforms. As a result, individuals from around the world often use Reddit as a medium for discussing

[1] "burnout", "burn out", "burnt out", and "burned out"

[2] Explained in Appendix A.1.

[3] Classes: professional, personal, non-traditional

[4] Definitions of both tasks are provided in Table 1.

[5] Our annotated dataset is available on GitHub: https://github.com/Computing-for-Social-Good-CSG/mental-burnout-disambiguation.

| Task | Classes | Definition | Excerpt from Dataset |
|---|---|---|---|
| Mental Burnout | 1 (N=1455, 62.4%) | Use of one of the burnout keywords in a manner related to mental health. The described experience is in the past or present. Hypothetical scenarios are not considered. | I'm burned out and anxious, and am on the verge of breaking down all the time. [...] I can't just pop by and ask questions and I feel totally detached from everybody. |
| | 0 (N=875, 37.6%) | Burnout keyword is used in contexts unrelated to mental health (i.e. without reference to psychological burnout), for example mechanical failure. | [...] my son's gaming laptop's power supply input on the motherboard burned out [...] Would it be possible to swap that specific part out? [...] |
| Context | Professional (N=609, 41.8%) | Mention of burnout occurring in the context of paid work or education. | [...] I am starting to get burnt out to the point of sleeping through a class twice in a two week span. I work crazy hours at my other job and have almost no life now that I'm doing both [...] |
| | Personal (N=616, 42.3%) | Mention of burnout in life outside of work, such as hobbies, relationships, and belief. | [...] I have no idea why my chest pain is worsening and why my exercise intolerance will not improve [...] I'm getting very burnt out. I'm in tears daily over the pain [...] |
| | Non-Traditional (N=230, 15.8%) | Mention of burnout occurring in the context of work not traditionally recognized by society. This includes unpaid work such as homemaker, and parenting, or paid work such as sex work. | [...] she knew all along that was what was going on, but didn't tell anyone in the family [...] I know that taking care of my grandmother has left my mom feeling exceedingly burnt out and that she is no longer taking care of herself because of it [...] |

Table 1: Example excerpts from posts in our dataset and their corresponding manual annotations for the 'Mental Burnout', and 'Context' categories. To maintain the privacy of posters, posts have been slightly paraphrased to avoid traceability.

sensitive topics, such as mental health (De Choudhury and De, 2014). With these considerations in mind, we collected all posts from Reddit written in English from January 1, 2014 to June 26, 2022 that contain at least one burnout-related keyword through the Pushshift API (Baumgartner et al., 2020) and Google BigQuery to use as our dataset[6].

We used different lexical variations of the term burnout (i.e., "burnout", "burnt out", "burned out", and "burn out") to collect our dataset. The initial collection included 379,371 posts. We performed a round of cleaning on the dataset, the details of which are explained in Appendix A.1. After cleaning, our dataset is comprised of 297,623 posts. These posts are written by 241,392 unique accounts across 23,519 subreddits.

**Qualitative Data Annotation.** After a qualitative inspection of the dataset, we found that many posts utilized a burnout keyword in a manner that is irrelevant to our mental burnout. For instance, posts used burnout-related keywords to describe electrical hardware damage. To distinguish between these different use cases, we employed a systematic annotation task, with the goal of distinguishing between mental burnout and non-mental burnout discourse. Two members of the research team labeled 2,330 randomly sampled posts. Our corpus size is in line with prior textual mental health corpora[7]. 1,455 posts (62.4%) were annotated as discussing burnout in a mental health context.

Having identified posts that discussed mental burnout, we wanted to understand the context in which each individual was experiencing burnout. In other words, we annotated the positively labeled posts (i.e., posts discussing mental burnout) with one of the three context classes: personal, professional, and non-traditional Table 1 displays definitions, number of posts, and examples for each of the labels within our dataset. Details of our annotation procedure are discussed in Appendix A.2.

**Automatic Detection.** We train a number of machine learning, deep learning, and in-context learning models for both tasks. We discuss our training and hyperparameter turning procedure in Appendix A.3 and A.4. The performance of our models is presented in Section 4.

## 4 Results

To study burnout at scale we train NLP models on our annotated dataset. Table 2 displays our top models within each category of models. We built three classes of models for each task: (1) classical machine learning models, (2) BERT-based deep learning models, and (3) in-context learning (ICL) methods. Details about how the dataset is split, as well as the model training and hyperparameter tuning, are discussed in Appendix A.3 and A.4 respectively. We observe reasonable performance on both the word sense disambiguation task, and the classification of context. Our performance is inline with other work on the detection of mental health conditions using social media data (De Choudhury et al., 2021).

The model that achieved the best performance in both cases was the fine-tuned *Distilbert-base* (Sanh et al., 2019) model (F1 = 0.86 and 0.84). We then applied this model to the entire dataset to label the 297,623 posts we had collected from Reddit. The posts were automatically labeled to indicate whether the use of burnout was to discuss mental health issues. 185,129 (62%) posts were identified as using burnout in a mental health related context. We visualize the trend of the number of posts

[6]Reddit data is available under BSD 2-Clause License: licensed under a permissive license allowing redistribution and modification with the retention of copyright and disclaimer notices.

[7]$|D_{stress}| = 1402$ (Saha and De Choudhury, 2017)

| Category | Model and Features | Acc | Precision | Recall | F1 | Test Class Distribution | # |
|---|---|---|---|---|---|---|---|
| Mental Burnout | SVM (TF-IDF) | 0.79 | 0.78 | 0.76 | 0.76 | Mental Burnout | 364 |
| | Distilbert-base | **0.87** | **0.87** | **0.85** | **0.86** | | |
| | gpt-4 zero shot COT | 0.82 | 0.83 | 0.78 | 0.80 | No Burnout | 219 |
| Context | Logistic Regression (TF-IDF) | 0.80 | 0.79 | 0.75 | 0.77 | Professional | 152 |
| | Distilbert-base | **0.85** | **0.84** | **0.84** | **0.84** | Non-traditional | 58 |
| | gpt-4 zero shot COT | 0.82 | 0.81 | **0.84** | 0.81 | Personal | 154 |

Table 2: Top classical, deep learning, and in-context learning models for each category with their corresponding accuracy, precision, recall, and macro F1 scores. The distribution of posts within the test set are also displayed in the last column.

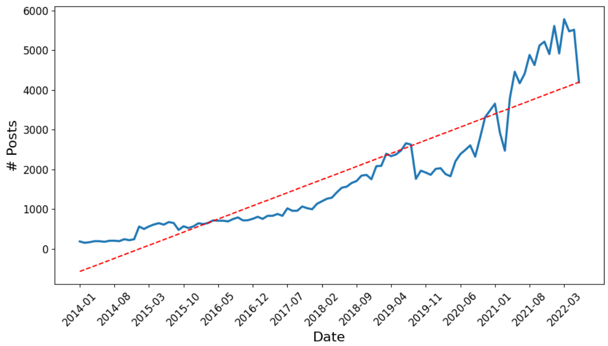


Figure 1: Monthly number of posts within our dataset that were classified as discussing mental burnout (Mental burnout = 1) during the 2014-2022 period. The red line displayed in the figure represents the general trend of the number of posts derived from numpy's *polyfit* function.

within each month that discusses mental burnout, displayed in Figure 1. This figure reveals a general upward trend over time, indicating an increase in burnout discourse across time. While this increase could be partially due to a general increase in Reddit use, this trend is in line with prior work such as Rasdi, Zaremohzzabieh, and Ahrari (2021) which found elevated levels of burnout and work disengagement among people who worked multiple jobs when also experiencing financial insecurity during the pandemic.

| Context | # Posts | Class % |
|---|---|---|
| Professional | 92,649 | 50.1% |
| Personal | 88,022 | 47.5% |
| Non-Traditional | 4,458 | 2.4% |

Table 3: Number of posts within the 185,129 posts discussing mental burnout that were labeled with each context by our best model (fine-tuned Distilbert base).

We further labeled our collection with the context burnout occurred in. The number of posts classified into each context is displayed in Table 3. We find that burnout due to personal stressors make up a considerable portion of the online discourse (47.5%). The discussion of burnout in non-professional contexts encourages additional research in settings other than occupational burnout. Trends over time are discussed in Appendix A.5. We also provide brief descriptive statistics of posts with burnout language in Appendix A.6.

## 5 Conclusion

In this paper, we examined burnout language through computational techniques for the detection and characterization of Reddit posts containing burnout-related keywords. Our dataset consisted of 297,623 Reddit posts with at least one occurrence of a burnout-related keyword collected from across 23,519 different subreddits. Following the annotation process, our best classification models perform at an minimum F1 of 0.84. We showcase high-level trends of burnout in our nine-year collection of online discourse on Reddit.

**Implications & Future Work.** Through our work, we developed a burnout word sense disambiguation model. This model could be utilized for personalized interventions, public health monitoring, and policy development. For instance, these models could be used to provide early intervention for individuals experiencing mental burnout. By identifying relevant posts, support systems can reach out to those in need, offering assistance and resources. Additionally, the model can assist the research community in gaining insights into the prevalence and trends of burnout in various contexts. Our results demonstrated a growth in the number of individuals struggling with burnout over time. The increasing prevalence of this psychological phenomenon signals a need for a deeper understanding of the causes and manifestations of this issue at scale, which our classification model could assist in. Future work could look into identifying and characterizing the language of exhaustion, cynicism, and inefficacy which are the dimensions of mental burnout. They could also examine how the manifestation of these dimensions differ across contexts.

**Limitations.** Our dataset is limited to Reddit posts with self-identified and ecologically valid self-expressions of burnout. While we believe this data is valuable to analyze, this limitation could influence the effectiveness of our models on textual instances in which the author does not explicitly disclose burnout. For instance, posts that discuss signs and symptoms of burnout without explicitly using the term "burnout" would not be detectable using our models. Moreover, not all individuals have the same level of comfort when it comes to sharing their mental health struggles online. As a result, the analyzed discourse could be more representative of this subset of the public. Additionally, our approach and classification models should not be used to make diagnostic claims, as questionnaires, such as the Maslach Burnout Inventory (MBI), are the only clinically validated means of assessing burnout. Rather, our work seeks to broadly capture burnout discourse across different domains. Finally, it is worth noting that social media platforms can act as "echo chambers" (Cinelli et al., 2021), where specific viewpoints could be amplified. It is important to keep this effect in mind when analyzing experiences of burnout using social media data.

**Ethical Considerations.** To protect the poster's of content on Reddit, we only release Reddit post IDs within our dataset. Doing so ensures that content that is removed from Reddit would no longer be accessible for future research, thus protecting user agency if they decide to delete or remove the content. To reduce the potential misuse of automated models, we focused on self-disclosure to ensure that the model is targeting content where individuals voluntarily share their experiences with burnout. In other words, the model is trained to respond to explicit signals of burnout rather than making assumptions about a user's mental state. It is also important to note the cost of misclassification. False positives may lead to unwarranted interventions for those incorrectly identified as experiencing burnout. Conversely, false negatives could result in overlooking those who genuinely need support. It is also worth noting that while Reddit is pseudo anonymous, when annotating the dataset user IDs were removed to preserve the identity of the poster. Additionally, we did not observe any offensive content or offensive language use in the posts we annotated.

## References


Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):830–839.

Leo Breiman. 2001. Random forests. *Machine Learning*, 45:5–32.

Cory J. Cascalheira, Shah Muhammad Hamdi, Jillian R. Scheer, Koustuv Saha, Soukaina Filali Boubrahimi, and Munmun De Choudhury. 2022. Classifying minority stress disclosure on social media with bidirectional long short-term memory. *Proceedings of the International AAAI Conference on Web and Social Media*, 16(1):1373–1377.

Stevie Chancellor and Munmun De Chourdhury. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *npj Digitial Medicine*.

LS Chutko, AV Rozhkova, S Yu Surushkina, TI Anisimova, and MD Didur. 2019. Clinical manifestations of burnout. *Zhurnal Nevrologii i Psikhiatrii Imeni SS Korsakova*, 119(1):14–16.

Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9):e2023301118.

Munmun De Choudhury and Sushovan De. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):71–80.

Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2021. Predicting depression via social media. *Proceedings of the International AAAI Conference on Web and Social Media*, 7(1):128–137.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Sergio Edú-Valsania, Ana Laguía, and Juan A Moriano. 2022. Burnout: A review of theory and measurement. *International journal of environmental research and public health*, 19(3):1780.

Shasha Han, Tait D Shanafelt, Christine A Sinsky, Karim M Awad, Liselotte N Dyrbye, Lynne C Fiscus, Mickey Trockel, and Joel Goh. 2019. Estimating the attributable cost of physician burnout in the united states. *Annals of internal medicine*, 170(11):784–790.

Hee Jung Hong, Guy Wilkinson, and Claudio M. Rocha. 2022. The relationship between basic needs satisfaction, self-determined motivation, and burnout in korean esports players. *Journal of Gambling Studies*.

David G Kleinbaum, K Dietz, M Gail, Mitchel Klein, and Mitchell Klein. 2002. *Logistic regression*. Springer.

Jane Mantzalas, Amanda L Richdale, and Cheryl Dissanayake. 2022. A conceptual model of risk and protective factors for autistic burnout. *Autism Research*.

C. Maslach and M.P. Leiter. 2016a. Chapter 43 - burnout. In George Fink, editor, *Stress: Concepts, Cognition, Emotion, and Behavior*, pages 351–357. Academic Press.

Christina Maslach and Michael P. Leiter. 2016b. Understanding the burnout experience: recent research and its implications for psychiatry. *World Psychiatry*, 15(2):103–111.

Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282.

Moïra Mikolajczak, James J Gross, and Isabelle Roskam. 2019. Parental burnout: What is it, and why does it matter? *Clinical Psychological Science*, 7(6):1319–1329.

Moïra Mikolajczak, James J Gross, and Isabelle Roskam. 2021. Beyond job burnout: Parental burnout! *Trends in Cognitive Sciences*, 25(5):333–336.

World Health Organization. 2019. Burn-out an "occupational phenomenon": International classification of diseases. Accessed: 2022-12-23.

Roziah Mohd Rasdi, Zeinab Zaremohzzabieh, and Seyedali Ahrari. 2021. Financial insecurity during the covid-19 pandemic: Spillover effects on burnout–disengagement relationships and performance of employees who moonlight. *Frontiers in Psychology*, 12:610138.

Koustuv Saha and Munmun De Choudhury. 2017. Modeling stress with social media around incidents of gun violence on college campuses. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–27.

Koustuv Saha, Sang Chan Kim, Manikanta D Reddy, Albert J Carter, Eva Sharma, Oliver L Haimson, and Munmun De Choudhury. 2019. The language of lgbtq+ minority stress experiences on social media. *Proceedings of the ACM on human-computer interaction*, 3(CSCW):1–22.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Wilmar B Schaufeli and Bram P Buunk. 1996. Professional burnout. *Handbook of work and health psychology*, 1:383–425.

Armin Shmilovici. 2005. *Support Vector Machines*, pages 257–276. Springer US, Boston, MA.

Yisheng Song, Ting Wang, Puyu Cai, Subrota K Mondal, and Jyoti Prakash Sahoo. 2023. A comprehensive survey of few-shot learning: Evolution, applications, challenges, and opportunities. *ACM Computing Surveys*, 55(13s):1–40.

Wei Wang, Vincent W Zheng, Han Yu, and Chunyan Miao. 2019. A survey of zero-shot learning: Settings, methods, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–37.

Andreas Weber and A Jaekel-Reinhard. 2000. Burnout syndrome: a disease of modern societies? *Occupational medicine*, 50(7).

Jue Wu, Junyi Ma, Yasha Wang, and Jiangtao Wang. 2021. Understanding and predicting the burst of burnout via social media. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3):1–27.

# A Appendix

## A.1 Data Cleaning Process

We clean the dataset by removing posts in which: (1) the title or body indicates the post was deleted (e.g., '[removed]'), (2) the body has less than 20 characters or does not have any identifiable words (e.g., the body only contains emojis or special characters) , or (3) the same content was posted across multiple subreddits by the same author with no changes.[8]

## A.2 Qualitative Annotation

The classes and definitions for both tasks are displayed in Table 1. An example belonging to each class is also shown in the table.

The annotation procedure for both tasks was as follows: the first 2,000 posts within the data were split into multiple batches, and annotators labeled the posts individually. For each post, annotators would first select whether the poster was using burnout in a mental health context or not. If the post was annotated as using burnout in a mental health context, they would then annotate what the context of burnout was: personal, professional, or non-traditional.

In between every two batches, annotators would discuss their disagreements and reach consensus

[8]In these cases we only keep one copy of the post

before beginning the next batch. The average batch agreement percentage was 98.68% and 93.36% for burnout detection and context detection, respectively. Cohen's Kappa (McHugh, 2012) was on average 0.89 and 0.86 for burnout detection and context detection, respectively. This indicates strong agreement. The annotation concluded with a complete agreement between the two annotators. Once agreement was reached on the first 2,000 posts, an additional 330 posts were labeled independently.

Both annotators were undergraduate computer science students. Both annotators obtained a background in burnout through studying the literature prior to labeling the data.

### A.3 Dataset Train-Test Split

We split our dataset into train and test sets using the *sklear*, test-train splitting function with 20% of the data being set aside for testing. Our test set was created using *stratified* sampling.

### A.4 Hyperparameter Tuning

We created a validation set from our training set (20% of the training data sampled through stratified sampling). All hyperparameter tuning efforts explained below were performed on the validation set with only the best model being deployed on the test data. Once the best parameters were found, the model was trained on the entire training set using those parameters and then evaluated using the test set.

All models were trained on Google Colab, with deep learning models using the *L4 GPU* runtime, and other models using the *CPU* runtime. Evaluation of models was performed through sklearn's *classification report* function as well as individual scoring functions of the package.

#### A.4.1 Classical Machine Learning

For this subset of models, we experimented with three models: Support Vector Machines (SVM) (Shmilovici, 2005), logistic regression (Kleinbaum et al., 2002), and Random Forests (Breiman, 2001). To vectorize our data, we experimented with Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) vectorizers with n-gram ranges of (1,1), (1, 2), and (1, 3). The values of 20,000 and 50,000 were also tested as the maximum number of features for these vectorizers. For each model, we test the combination of the following hyperparameters:

- **SVM**
  - $kernel \in \{linear, poly, rbf, sigmoid\}$
  - $gamma \in \{scale, auto\}$
- **Random Forest**
  - $n\_estimators = 100$
  - $max\_features \in \{None, sqrt, log2\}$
  - $max\_depth \in \{4, 6, 8, None\}$
  - $criterion \in \{gini, entropy\}$
- **Logistic Regression**
  - $penalty \in \{l2, None\}$
  - $c \in \{0.25, 1, 5\}$

The best model for burnout was SVM with $gamma = auto, kernel = linear$. The best model for context was Logistic regression with $penalty = None, C = 0.25$.

#### A.4.2 Deep Learning Models

We fine-tune two pre-trained models of *bert-base-uncased* (Devlin et al., 2019) and *distilbert/distilbert-base-uncased* (Sanh et al., 2019) to detect burnout and context. For both cases, we examine the combination of the following hyperparameters:

- $learning\_rate \in \{2e^{-5}, 3e^{-5}, 4e^{-5}\}$
- $per\_device\_batch\_size \in \{8, 16, 32\}$
- $num\_train\_epochs \in \{2, 4, 6\}$
- $weight\_decay = 0.001$

The best model for the detection of burnout was *distilbert* with the following hyperparameter: $learning\_rate = 4e^{-5}$, $per\_device\_batch\_size = 8$, $num\_train\_epochs = 4$. The same fine-tuned model also performed best for the classification of context, with the following hyperparameter: $learning\_rate = 4e^{-5}$, $per\_device\_batch\_size = 8$, $num\_train\_epochs = 6$.

#### A.4.3 In-Context Learning

We use the two LLM models of *gpt-3.5-turbo* and *gpt-4* with $temperature = 0$. Higher temperature values result in more creative responses, while lower temperature makes the output more deterministic. As our classification task does not require creativity, we elect to set temperature to zero. Maximum output tokens for the models are set to 100.

We examine zero-shot (Wang et al., 2019) and few-shot learning (Song et al., 2023) approaches for the detection of both classes. For our burnout detection task, our prompt is as shown in Table 4. The prompt for our context task follows a similar format with definitions of context classes being used.

| *Zero-shot prompt* |
|---|
| **Title**: "Classification of mental burnout in text" |
| **Definition**: In this task, we ask you to classify the input text into two options: |
| (A): Mental burnout: the poster discussed burnout related to their own mental health in the past or present. The context of burnout can be related to school, work, personal life, hobbies, and games. |
| (B): No mental burnout: burnout used in a context unrelated to mental health. Or mental burnout in hypothetical situations when the poster is not discussing their own experience in the past and present. |
| **Emphasis** & **Caution**: Discussions of hypothetical situations such as fear of burnout or future/imaginary circumstances should NOT be labeled as (A). |
| **Things to avoid**: All input must be classified into one of the options. If you cannot pick then choose the option with higher probability. The output must be either (A) or (B) but not both. |
| **Input**: {text} |
| **Output**: |

Table 4: Zero-shot prompt.

In the few-shot learning cases, we include 3 random examples from the dataset. We further test COT prompting by including the phrase "let's think about it step by step" at the end of both the zero-shot and few-shot learning prompts. Recognizing the token limit of the models, we cut off the texts of the posts to fit this token limit.

### A.5 Context Trends Over Time

Figure 2 shows how the context in which burnout is used has changed over time on Reddit. We can see that while professional and personal context have had a general upward trend throughout the years, non-traditional context have been consistently low until recently.

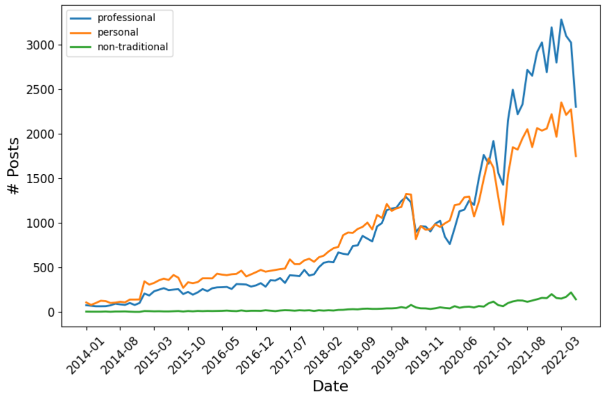


Figure 2: Monthly number of posts within our dataset that were classified into one of three contexts during the 2014-2022 period.

### A.6 Descriptive Statistics of Burnout Language

As discussed in the paper 185,129 Reddit posts were classified as discussing burnout in a mental health context. The top 10 subreddits that discussed mental burnout were as follows: *r/Advice* (4,513), *r/relationship_advice* (4,467), *r/offmychest* (3,527), *r/depression* (3,520), *r/careerguidance* (3,178), *r/ADHD* (3,004), *r/relationships* (2,599), *r/jobs* (2,440), *r/cscareerquestions* (2,101), and *r/antiwork* (2,002). The existence of subreddits dedicated to relationships and personal issues within the top 10 emphesizes the importance of studying burnout in non-professional contexts.

| Category | # Characters | # Words |
|---|---|---|
| Mental Burnout | 1,255 | 274 |
| Context: Professional | 1,320 | 287 |
| Context: Personal | 1,151 | 252 |
| Context: Non-Traditional | 2,144 | 364 |

Table 5: Median length of posts classified into each of our classes.

Table 5 displays the median length of posts classified into each of our classes. This length reflects the total length of the title and body of the post.