Tracking Life's Ups and Downs: Mining Life Events from Social Media Posts for Mental Health Analysis

Minghao Lv[1,2], Siyuan Chen[1,2], Haoan Jin[1,2], Minghao Yuan[3], Qianqian Ju[3], Yujia Peng[3,4,5], Kenny Q. Zhu[6], Mengyue Wu[1,2*]

[1] X-LANCE Lab, School of Computer Science, Shanghai Jiao Tong University, China
[2]MoE Key Lab of Artificial Intelligence, Jiangsu Key Lab of Language Computing, China
[3] School of Psychological and Cognitive Sciences, Peking University, China
[4] Institute for Artificial Intelligence, Peking University, China
[5] State Key Laboratory of General Artificial Intelligence, BIGAI, China
[6]University of Texas at Arlington, USA

Abstract

Social media platforms possess considerable potential in the realm of exploring mental health. Previous research has indicated that major life events can greatly impact individuals' mental health. However, due to the complexity and ambiguity nature of life events, shedding its light on social media data is quite challenging. In this paper, we are dedicated to uncovering life events mentioned in posts on social media. We hereby provide a carefully-annotated social media event dataset, **PsyEvent**, which encompasses 12 major life event categories that are likely to occur in everyday life. This dataset is human-annotated under iterative procedure and boasts a high level of quality. Furthermore, by applying the life events extracted from posts to downstream tasks such as early risk detection of depression and suicide risk prediction, we have observed a considerable improvement in performance. This suggests that extracting life events from social media can be beneficial for the analysis of individuals' mental health.

1 Introduction

Mental health constitutes a fundamental component of overall wellbeing, with mental disorders contributing significantly to the global disease burden (Prince et al., 2007). The National Institute of Mental Health reports that 23.1% of U.S. adults (59.3 million) experienced Any Mental Illness in 2022, including 6.0% (15.4 million) with Serious Mental Illness, yet only half received treatment[1]. This treatment gap underscores the critical need for automated mental state inference through life event analysis, which can reveal developmental trajectories of mental conditions. Social media platforms emerge as particularly valuable resources in this context (Cohan et al., 2018), offering sequential, temporally rich data from users - including substantial populations with mental health conditions - who document life changes and internal states through continuous self-disclosure.

User Post Example

Post:

My roommate and I had to move apartments at the beginning of the semester ... but **he was out of a job** and just partying the whole time ... **I actually had to call an ambulance for him because he ODed in front of our apartment complex** ....

When we got kicked out of our apartment, we were only given 10 days to move out and find a new place ...

At the same time, my laptop died on me ... **I decided to get another laptop instead** ... I had no access to the internet or a computer during this time. This put me behind several weeks in my courses ... The final grades came in and sure enough **I failed every course**.

Career **Health** **Relocation** **Financial** **Education**

Final diagnosed with: **Depression, ADHD**

Figure 1: One Example of Life Events Contained in Social Media Posts from SMHD (Cohan et al., 2018). Various life events are mentioned, the user is later diagnosed with Depression and ADHD.

Life events - major incidents altering identity, resources, and social position - have long been recognized in psychological research as critical determinants of mental health trajectories (Mandemakers, 2023; Rabkin and Struening, 1976). These transitions (e.g., marriage, career changes, relocation) profoundly influence psychological well-being, social functioning, and physical health (Luhmann et al., 2012; Marc et al., 2024), with particular relevance to depressive episodes and suicidal risk: meta-analyses demonstrate stressful events increase suicidal outcomes by 37%-45%, disproportionately affecting males and youth (Howarth et al., 2020; Liu and Miller, 2014).

However, applying life events from social media to analyze mental health trajectory data faces numerous challenges. Firstly, applying life events to mental health analysis requires a clear definition and taxonomy for life events. The definition of life events should be as comprehensive as possible, covering the major life events that individuals might experience and could impact mental health. Secondly, the vast amount of less critical information

*Corresponding author.

[1]https://www.nimh.nih.gov/health/statistics/mental-illness

Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 6950–6965
July 27 - August 1, 2025 ©2025 Association for Computational Linguistics

and noise in social media can negatively affect the performance of event detection models (Atefeh and Khreich, 2015). Despite the abundance of data on social media, the amount of annotated data available for mental health analysis research is quite limited (Cohan et al., 2018). Finally, determining life events requires extensive context. In other words, within a single post, the poster might use multiple sentences or even several discrete phrases to describe one life event. This poses difficulties for the annotation of life events. As illustrated in the example presented in Figure 1, the poster described multiple life events using several discrete sentences within the same post, which complicates the annotation of life events.

In this work, we adjusted a stress life event taxonomy for social media research and proposed an annotation system for 12 major life event categories. Our definition of life events, based on Haimson et al. (2021), modifies and supplements to constitute 12 categories with 127 distinct life events. Then, we leverage innovative annotation methods and the bespoke annotation system to propose a multi-life event classification dataset, **PsyEvent** (**Psy**chologically-related Life **Event**s) based on social media posts, encompassing a total of 7,965 Reddit post sentences labeled with 12 life event categories. We recruit annotators and provide them with rigorous training. The annotation approach we adopt is complex and incorporates numerous quality assurance measures. We also develop a web-based annotation system tailored to life events to implement these annotation methods and quality assurance techniques (For more details, please refer to the appendix A.). Furthermore, to filter posts more relevant to life events for annotation, we employ an embedding-based search method (Reimers and Gurevych, 2019) instead of traditional keyword-matching techniques to identify candidate posts. This approach allows annotators to focus more on posts related to life events, thereby enhancing annotation efficiency. We train two life event classification models using the aforementioned annotated dataset.

Finally, we utilize the trained life event classification models on two downstream tasks: early risk detection of depression and suicide risk prediction. We employ two different approaches to apply life events, and the downstream tasks with life events show performance improvements, demonstrating the effectiveness of life events in mental health analysis. Our contributions are three-fold:

- We are the first to formalize life events computationally by adopting and adapting an existing taxonomy of stress-related life events for social media analysis. Our dataset, *PsyEvent*, includes detailed annotations for 12 life event categories, covering 127 specific events. This comprehensive approach enables the development of a life event detection model with over 95% AUC.

- Our life event detection model can effectively trace a user's post history to identify and tag major life events that are critical to their mental health status. We hence innovatively adopt life events as a novel feature for mental health-related tasks, enchancing model performance, interpretability and clinical relevance.

- We demonstrate the effectiveness of life events by applying them to two critical downstream tasks: early risk detection and suicide risk prediction. Our results show significant performance improvements, validating the value of life events in enhancing predictive models.

2 Related Work

2.1 Events Extraction in Social Media

Social media event detection has evolved through two methodological paradigms. Early approaches predominantly employed clustering techniques: Weng and Lee (2011) proposed wavelet-based signal clustering (EDCoW) for noise reduction and event identification, while Li et al. (2012) developed Twevent using bursty segment detection and content-based clustering. Ritter et al. (2012) advanced open-domain categorization through latent type discovery in TwiCal.

The deep learning revolution introduced neural architectures addressing three key limitations of clustering methods: context sensitivity, feature engineering dependency, and cross-domain adaptability. Nguyen et al. (2017) pioneered hybrid CNN-RNN frameworks for noise-resilient detection in Twitter streams. Syntactic hierarchy modeling saw breakthroughs with Yan et al. (2019)'s MOGANED, integrating multi-order graph attention networks for trigger word identification. Scalability challenges were addressed by Afyouni et al. (2022)'s Deep-Eware, combining incremental NLP pipelines with geospatial processing for real-time event tracking. The life event taxonomy used in Chen et al. (2024)'s study, includes 11 categories and 43 life events-based on the Holmes-Rahe Stress Inventory (Holmes and Rahe, 1967). Furthermore, a life event classifier based on BERT was trained for causal relationship computation. The current study differs from previous works by formerly defining a 121 life-event paradigm and innovatively investigates life events as effective mental health indicators, with dual paradigms (feature-based vs. embedding fusion) proposed for depression and suicide prediction.

2.2 Prior Life Event Inventories

The foundational Social Readjustment Rating Scale (SRRS) (Holmes and Rahe, 1967) established quantitative life event analysis through 43 stress-inducing events rated by 394 participants based on social adaptation demands. While pioneering stress measurement (Scully et al., 2000), SRRS faced critiques regarding categorical granularity and stressor calibration (Monroe, 1982). Dohrenwend et al. (1978) addressed these limitations through methodological refinements in their Psychiatric Epidemiology Research Interview (PERI) scale, expanding to 102 events with population-specific validation across three dimensions: event selection criteria, rater diversity, and consensus metrics. Contemporary analyses (Haimson et al., 2021) reveal remaining gaps in capturing sociotechnical transitions critical for modern digital behavior studies.

3 PsyEvent Dataset Construction

3.1 Life Events Definition

Selecting the appropriate definition of life events is crucial for effectively conducting mental health analysis from social media data. In light of the aforementioned issues, we aim to select life events that more closely align with the major occurrences in people's daily lives. Our definition of life events is based on the taxonomy proposed by Haimson et al. (2021), which identifies 121 life events organized into 12 categories. This taxonomy is grounded in a comprehensive survey conducted with a representative sample of 554 participants from the U.S. general population. We adopted most of the life event taxonomy proposed by Haimson et al. (2021), with modifications driven by overlapping categories, annotation clarity, and pilot annotation feedback. These adjustments resulted in a refined taxonomy comprising 12 life event categories and 127 sub-events. The finalized taxonomy reflects daily life events that influence mental health and is detailed in Appendix D, with an example of the *"New Birth in Family"* category shown in Table 1.

Life Event Category	Life Events
New Birth in Family	gain of new family member; gave birth / became a parent; adopted a child; became a grandparent; became a great-grandparent; became an aunt/uncle

Table 1: Life events in the adjusted *"New Birth in Family"* category

3.2 Dataset Annotation

We used a subset of the SMHD dataset (Zhang et al., 2022; Cohan et al., 2018), which includes Reddit posts related to mental health, making it suitable for detecting life events. A preliminary screening process was conducted to select relevant posts using sentence-BERT embeddings and cosine similarity. Posts were filtered for length to enhance annotation efficiency. We adopted iterative annotator training methods and a serious quality control protocol to ensure the quality. Annotators used a bespoke website system for annotation (for more detailed descriptions, please refer to the appendix A.3.). The dataset includes both life event and control sentences to improve model training. We asked annotators to identify 127 sub-life events within the posts and, if present, label the sentences with the broader category to which the sub-life event belongs. This method resulted in more specific labels for life events.

During the online annotator training sessions, we openly disclosed and discussed our compensation structure, which exceeded the local minimum wage requirements. We recruited annotators through poster advertisements. All participants were well-educated, including those majoring in psychology. The age distribution was as follows: 75% of annotators were 18-25 years old, 18.75% were 26-30, and 6.25% were 31-40 years old. Regarding educational background, 37.5% held bachelor's degrees while 62.5% possessed graduate-level degrees or higher.

We also attempted automatic annotation with Large Language Models (LLMs, including GPT-4o and GPT-4) for life events extraction. However, it did not excel at certain event tagging: such context-dependent annotation still seems challenging even with the application of advanced context-learning and CoT techniques. We had three annotators labelling 220 data points to create a small test set for evaluating the annotation accuracy of GPT-4o. The inter-annotator agreement was 0.86. The average annotation accuracy of GPT-4o on this small test set was found to be 0.70. Therefore, human annotation is solely adopted to ensure a high-quality dataset.

3.3 Data Statistics

After annotating (for the definition of life events and the annotation process of the dataset, please refer to Appendix A.), our final product, PsyEvent, containing 12 categories of life event labels, is divided into three parts: training/validation/test set. After the aforementioned annotation method, we ultimately obtained 7,965 sentences, which were then allocated to the test set, validation set, and training set in a ratio of 7:1:2. Due to natural distribution, there are certain life events less mentioned in these posts and we tried to exclude vague, ambiguous posts for more precise recognition. The number of positive samples in the dataset is shown in Table 2.

Life Event Categories	Positive Num.
Health	909
Financial	344
Relocation	165
Legal	226
Relationship Changes	323
New Birth in Family	178
Death	171
Career	377
Education	199
Lifestyle Change	294
Identity	262
Societal	136
Total	3,584

Table 2: Number of positive samples across 12 categories

4 Life Events Recognition System

We intruduce the models trained to extract life events from social media and self-status determination, along with their respective results.

4.1 Life Events Detection

An individual may experience multiple life events throughout their life, and when posting, the poster may describe several life events simultaneously. As a result, there may be instances where a single post contains multiple life events. Therefore, we treat life event detection as a multi-label binary classification task. We employ a BERT-based encoder (Devlin et al., 2019) to train the life events detection model. Since we incorporate control posts, a data balancing sampler is used to ensure that each training batch samples an equal amount of annotated data and control sentences.

Life Event Categories	AUC (%)
Health	92.1
Financial	95.7
Relocation	97.7
Legal	96.1
Relationship Changes	95.0
New Birth in Family	92.6
Death	99.7
Career	93.5
Education	99.2
Lifestyle Change	87.9
Identity	95.5
Societal	97.4
Avg.	95.2

Table 3: Life event detection results on *PsyEvent*.

Results As illustrated in Table 3, the average classification performance is over 95%, through varies across the 12 categories of life events due to disparities in the inherent difficulty of classifying the events themselves. We can observe that among the 12 life event categories, the *"Death"* category achieved the best classification performance, while the *"Lifestyle Change"* category had the poorest classification outcome. During the annotation process, we also noticed that annotators had the most questions and discussions about the *"Lifestyle Change"* category. It is possible that the varying understandings of this type of life event among annotators led to less successful classification results for this category. For a 12-category multi-class task, extracting life events is inherently a challenging task (Zhang et al., 2022). More detailed information is provided in Appendix B.1.

4.2 Self-Status Determination

In addition to the life event detection model, we developed a life event self-status determination model. When individuals recount life events on social media, they may refer to past experiences rather than current ones, or even events involving others rather than themselves. To address these complexities, we specifically trained a model to distinguish whether the narrated events pertain to the individual's own experiences or those of others.

Such a self-status determination model is to determine whether the *posters themselves* are currently experiencing the life events detected in sentences. Thus, we define this task as a single-label

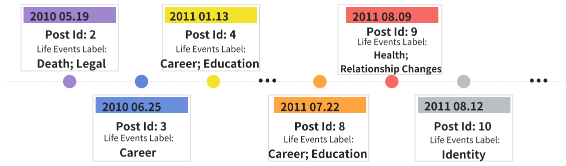

Figure 2: The results after a user's posts have been processed by the Life Events Detection model, showing a portion of the user's posts along with their life event labels.

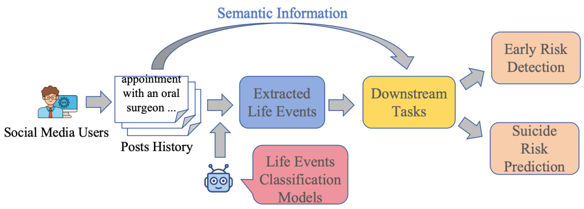

Figure 3: The overall architecture of our work involves applying life event classification models trained on *PsyEvent* to posts published by social media users to extract life events and utilize them in two downstream tasks: early risk detection and suicide risk detection.

binary classification task. We employ the same model used for the life event detection task to accomplish the self-status determination task. Finally, the self-status determination model yields an AUC of 76.8%. This status determination model is later used in combination with the Life Event Detection model to infer life events status in downstream tasks. Detailed information is provided in Appendix B.2.

4.3 Case Study

Figure 2 illustrates the information of a portion of a user's posts after being processed by the Life Events Recognition model, including the posting time, post ID, and corresponding life event labels. Each user has a complete posting history and life event history. With the two Life Event Classification models, we can capture long-term life events change which is critical to one's mental health status.

5 Downstream Tasks

In this section, we describe the application of the above model to two downstream tasks: early detection of depression risk and suicide risk prediction. The overall architecture is illustrated in Figure 3.

5.1 Early Risk Detection of Depression (ERDD)

Early risk detection is designed to identify mental disorders at an early stage, thereby achieving the goal at a lower cost (Losada and Crestani, 2016). The early detection task is defined as follows: For a user U_i, we define their posting history as $[P_{i,1}, P_{i,2}, ..., P_{i,n}]$ (where n is the total number of posts made by user U_i and $P_{i,j}$ is the j-th post published by U_i. Traditional depression detection models predict whether a user has depression based on their entire posting history $[P_{i,1}, P_{i,2}, ..., P_{i,n}]$. However, in early risk detection, posts emerge one after another, so at time t, the model can only see t posts $[P_{i,1}, P_{i,2}, ..., P_{i,t}]$. Once the model has sufficient confidence, it can make an early prediction on y_i at time $t(where\ t \leq n)$, allowing the prediction to strike a good balance between accuracy and earliness.

5.2 Suicide Risk Prediction (SRP)

Suicide is one of the leading causes of death globally (Värnik, 2012). Suicide risk prediction (SRP) using social media posts offers a promising approach to proactively identify at-risk individuals, enabling early intervention and potentially reducing mortality rates. While binary classification (i.e., distinguishing between individuals with and without suicidal intent) is a common approach, it fails to capture the nuances in the severity of suicidal intent. Therefore, this paper adopts a five-class classification framework inspired by Gaur et al. (2019). Three classes are derived from the Columbia Suicide Severity Rating Scale (C-SSRS): Suicidal Ideation (ID), Suicidal Behavior (BR), and Actual Attempt (AT), representing increasing levels of risk. Two additional categories are included for a more comprehensive assessment: Suicide Indicator (IN), for users with at-risk language but no active distress symptoms, and Supportive (SU), for individuals discussing suicide without expressing personal risk. Thus, SRP is framed as a multi-class classification problem, using the user's social media history to determine their risk category.

5.3 Applying Life Events

Life event features are critical in depression detection and suicide risk prediction as they provide contextual information that enhances model interpretability and clinical relevance. By identifying specific life stressors such as job loss or relationship breakdown, these features contextualize the onset of mental health issues and improve prediction accuracy. This contextual understanding is valuable for developing targeted interventions and informing clinical decision-making.

We integrate the life event detection model and the life event self-status determination model into two tasks: early risk detection of depression (ERDD) and suicide risk prediction (SRP). We explore two approaches to leverage life events in these tasks.

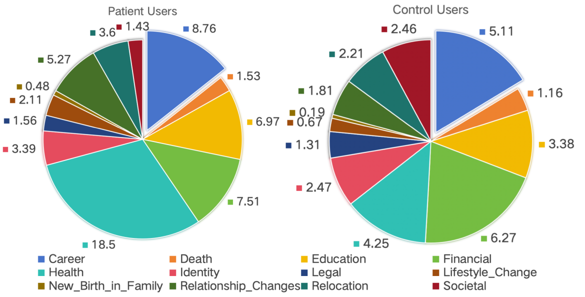

Figure 4: The average number of life events in patient users (left) and control users (right).

Life event as features We apply the life event classification models to the ERDD dataset, generating sequences of life event detection probabilities and self-status determination probabilities for each user's posts. These probability sequences are used as features alongside depression labels to train a depression detection model.

Concatenating life event features Since depression-related posts inherently contain semantic information related to depression, this semantic information can also be leveraged to train a depression classification model. Taking it a step further, since the life event features and post semantic information can both be used to train a depression classification model, why not use both together? Therefore, we concatenate the life event detection probability sequences, self-status determination sequences, and semantic information from the same posts to train a classifier model, hoping to achieve the best possible depression detection results.

6 Experiments

We present the outcomes of two downstream tasks: (1) integrating life events into early risk detection of depression using the two approaches described above, and (2) utilizing life events in suicide risk prediction.

6.1 Early Risk Detection of Depression (ERDD)

Dataset We utilize the ERD dataset proposed by Chen et al. (2023). They extracted users and posts from a public Reddit corpus. They selected depression users by detecting patterns that consist of two parts: one matching self-reported diagnoses (e.g., "diagnosed with") and the other mapping relevant keywords to depression (e.g., major depressive disorder). Control users (i.e., healthy individuals) were randomly selected from those who have never posted or commented in mental health-related subreddits. The dataset comprises 3,105 depression users and 17,209 control users. In this dataset, the average number of life events mentioned by users is shown in Figure 4. We can observe that patient users experience a higher number of life events, including health, career, education, and relationship changes.

Baseline We employ the model proposed by Zhang et al. (2022) as baseline, which utilizes CNN of various kernel sizes as backbone that has yielded good results on the task, and we use `RoBERTa-large` (Liu et al., 2019), `BERT-base` (Devlin et al., 2019) and `Qwen2-7B` (Yang et al., 2024) as encoders to extract semantic information from depression posts as input. More specifically, for `RoBERTa` and `BERT`, we use sentence-BERT (Reimers and Gurevych, 2019) to extract embeddings from these posts, and apply the obtained embeddings to train classification models.

Evaluation Metric We adopt the official metrics $ERDE_5$ and $ERDE_{50}$ for Early Detection task proposed by Losada and Crestani (2016). The lower $ERDE_5$ and $ERDE_{50}$, the better model performs early detection, and $ERDE_5$ has a higher penalty than $ERDE_{50}$ for late detection. Detailed introduction of these metrics can be found in Appendix C.

Experiment Results The results for the ERDD task are presented in Table 4. We conduct three experiments for each method using different seed values, and the results shown in the table are the average of the three experiments. It can be observed that using the `RoBERTa-large` as an encoder yields better results than using `BERT-base`, and `Qwen2-7B` achieved the best results. Therefore, we used `RoBERTa-large` and `Qwen2-7B` as encoder in subsequent experiments. Naturally, using *life events as features* alone did not yield satisfactory results since by its nature is sparse feature. However, it works well when in combination with ordinary embeddings, where the model can efficiently learn the contextual semantics as well as the relations between life events and mental status. The most effective approach in this task was *concatenating life event features*. This method performed the best on both evaluation metrics using `Qwen2-7B` and showed a more significant improvement on $ERDE_5$ while using `RoBERTa-large`. For `RoBERTa`, adding life events brings about 25 relative percentage improvement on $ERDE_5$. Since $ERDE_5$ is more sensitive to delays, this also indicates that using *concatenated life event features* can predict depression at an earlier stage. For this method, the improvement using `Qwen2-7B` is not as significant as that with `RoBERTa-large`. This may be because the embedding dimension of `Qwen2-7B` is sufficiently high, inherently containing rich semantic information. For the lower-dimensional embeddings of `RoBERTa-large`, the role of life events becomes more pronounced.

Method	$ERDE_5(\%)\downarrow$	$ERDE_{50}(\%)\downarrow$
BERT-base	12.73 ± 0.12	6.09 ± 0.28
Roberta-large	12.33 ± 1.28	5.67 ± 0.59
Qwen2	9.51 ± 0.91	4.67 ± 0.79
LE as feature	13.56 ± 0.02	6.30 ± 0.06
LE concat(Roberta[1])	9.77 ± 1.33	5.41 ± 0.07
LE concat(Qwen2-7B)	$\mathbf{9.26} \pm 0.62$	$\mathbf{4.11} \pm 0.18$

Table 4: Experiment results of Early Risk Detection of Depression (ERDD). Roberta[1] means `Roberta-large`.

6.2 Suicide Risk Prediction (SRP)

Dataset We utilize the suicide risk prediction dataset proposed by Gaur et al. (2019), a human-annotated dataset encompassing 500 users categorized based on varying levels of suicide risk. The data comprises of 22% control users, 20% users with some *suicidal indication*, 34% users with *suicidal ideation*, 15% users with *suicidal behaviors*, and 9% users have made an *attempt to commit suicide*. On average, each user has 31.5 posts, spanning a timeframe from 2005 to 2016. The entire dataset is randomly divided into training, validation, and test sets in a 7:1:2 ratio. Additionally, the sum of life event probabilities in this dataset is shown in Figure 5. Interestingly, severe conditions including Behavior and Attempt users exhibit larger proportion of Health-related life events in the total distrubution.

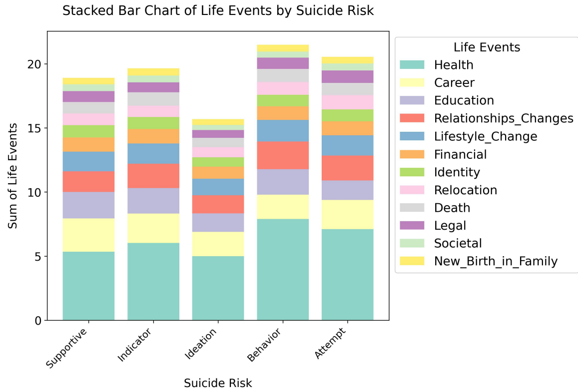

Figure 5: The sum of life event probabilities mentioned by users in the SRP dataset by suicide risk.

Baseline We utilize the same baseline model as used in ERDD. And we use `RoBERTa-large` (Liu et al., 2019), `BERT-base` (Devlin et al., 2019) and `Qwen2-7B` (Yang et al., 2024) as encoders to extract semantic information from depression posts as input.

Experiment Results The results of the SRP task are demonstrated in Table 5. Notably, the approach of concatenating life event features with semantic features significantly improves SRP performance, demonstrating the value of incorporating life events into predictive models. However, life event features, by their nature, are sparse and thus insufficient on their own to drive effective predictions. In contrast, when life event features are combined with any contextual semantic features, the performance improves by over 2%. This suggests that while life event features alone may not be effective enough, their combination with semantic features derived from advanced embeddings can significantly enhance model performance. This integration not only enhances model accuracy but also maintains interpretability, as life events offer clear, contextual explanations for predictions. For instance, life events can directly link specific stressors to mental health outcomes, making the model's decisions more transparent and clinically actionable.

However, we observed that the performance gains from integrating life event features with large language model embeddings varied across the two tasks (ERDD and SRP). This indicates that the effectiveness of combining life event features with semantic embeddings may depend on the specific task and model architecture. Therefore, further exploration is needed to optimize the integration of life event features with large language model embeddings for different predictive tasks.

Method	F_1 Score (%) ↑
BERT-base	33.50 ± 0.38
Roberta-large	34.52 ± 0.75
Qwen2	36.44 ± 2.17
LE as feature	12.49 ± 0.00
LE concat(Roberta-large)	36.00 ± 0.83
LE concat(Qwen2-7B)	$\mathbf{39.13} \pm 2.12$

Table 5: Experiment results of Suicide Risk Prediction (SRP)

6.3 Analysis

The extraction of life events can also provide explanations for the diagnosis of mental illnesses. In this section, we will specifically demonstrate an example, starting from a post where a user was diagnosed, and looking backward to show the life events experienced by this user before being diagnosed with depression to illustrate the auxiliary role of life event identification in disease diagnosis.

User Post No.: 33
Predicted life events label: *Education*, *Health*, *Relationship Changes*
Post: This is the second semester I've blown it ... My parents were going through a rough divorce and my life was a mess ... **I was diagnosed with mild depression** ... I got into community college and I bombed my first year there ...

Table 6: This user's 33rd post was labeled with *Education*, *Health*, and *Relationship Changes* by the aforementioned life event classification model.

Table 6 shows the post published by the user at the time of diagnosis. This post, after being processed by the two life event classification models, was labeled with *"Education"*, *"Health"* and *"Relationship Changes"*. The poster was diagnosed with depression while also experiencing stress from school and family upheaval. These are significant stressor events contributing to the poster's depression. For students, exams are a significant source of stress, and the pressure felt by students after entering college is particularly pronounced among first-year students (Robotham, 2008). Moreover, even before the breakdown of a marriage, children whose parents later divorce exhibit higher levels of anxiety/depression and antisocial behavior compared to children whose parents remain married; the divorce itself can further exacerbate the child's anxiety/depression (Strohschein, 2005).

User Post No.: 31
Predicted life events label: *Death*
Post: Friend died the other night. Not really too upset over it ... I just feel kinda cold hearted because I'm not crying over it. I'm just bummed ... It's not that I don't care, but it's not something that I'm gonna sit around and cry over.
User Post No.: 32
Predicted life events label: *Death*, *Relationship Changes*
Post: Cut a friend out of my life. Getting the vibe that she hates me ... I was getting tired of that also .... I was at another friend's funeral the other day and barely any of them bothered to show up .... they couldn't even bother themselves to show up at his funeral ... As of right now I don't really plan on reaching out to her ...

Table 7: Life events mentioned in the user's 31st and 32nd posts.

Additionally, before this, the poster had experienced events such as deteriorating relationships with friends and the death of a friend, as shown in Table 7. Peer relational issues may be a common cause of worry symptoms in depression and generalized anxiety disorders, or this risk factor may facilitate the association between these symptoms (Konac et al., 2021). Therefore, life events can provide an explainable pathway for the development of mental disorders, aiding in the diagnosis of these conditions. Moreover, Figure 6 illustrates the principle and function of ERDD. The ERDD model diagnosed the poster with depression at the 27th post, while the individual did not mention their own diagnosis until the 33rd post.

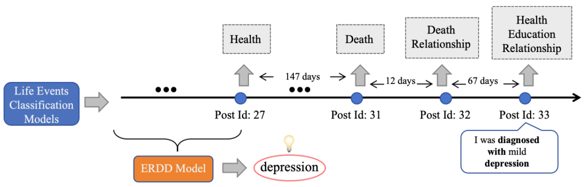

Figure 6: ERDD aims to predict the onset of depression in posters using a minimal number of posts, thereby detecting the condition at an earlier stage.

7 Conclusion

In this study, we demonstrated the potential of leveraging social media data to explore mental health through the lens of major life events. Recognizing that life events are often scattered throughout social media posts, we adopted a sentence-level annotation approach. To ensure high-quality data, we established a meticulous annotation guideline and incorporated rigorous testing and quality control processes throughout the annotation workflow. Using this carefully curated dataset, *PsyEvent*, we introduced a structured methodology for identifying life events from social media posts. Our experiments revealed that integrating life event information into downstream tasks, such as Early Risk Detection of Depression (ERDD) and Suicide Risk Prediction (SRP), largely improves task performance. These findings underscore the value of life event analysis as a powerful tool in mental health research. Future studies can build upon our dataset and methodologies to further enhance the applications of social media analysis in mental health interventions.

Ethical Statement

Dataset To protect personal privacy during data collection, we replaced usernames with random identifiers to prevent user identification without external information. All datasets used in this study are publicly available or used under appropriate licenses. We comply with data use agreements to prevent privacy violations and misuse.

Annotation We ensure that annotators are compensated with a fair wage above the minimum requirement. Any questions or concerns they have will be addressed promptly. Given that the content involves sensitive topics such as stressful life events, we acknowledge the potential emotional impact on annotators. Therefore, they are free to take breaks or discontinue the task at any time. We also interviewed some annotators about their feeling after annotation. They only reported slight discomfort at the time of reading sad or frightening posts due to empathy, and they found no long-term negative effects on them.

Application This study carefully considered the application of social media for the detection of mental illnesses. The purpose of this work is not to replace psychiatrists. Instead, we hope our model will be used as an effective supportive tool by experienced psychiatrists in the future.

Limitations

Although in this work we demonstrated the potential of life events for mental health research, we did not explore the pathways through which life events exert their influence, nor the differences in the impact of different life event categories. Such research would be of greater assistance to mental health studies. Moreover, other mental health research tasks could be explored to investigate the differences in the role of life events across various tasks.

Further research is needed to identify better ways to apply life event features, as different tasks may require different approaches to using life events to achieve optimal results. Additionally, the integration of large language models with life events also needs further exploration to optimize the combination of large language model embeddings and life events for different prediction tasks.

Acknowledgement

This work has been supported by the China NSFC project (No. U23B2018) and Guangxi major science and technology project (No. AA23062062).

References

Imad Afyouni, Aamir Khan, and Zaher Al Aghbari. 2022. Deep-eware: spatio-temporal social event detection using a hybrid learning model. *Journal of big Data*, 9(1):86.

Farzindar Atefeh and Wael Khreich. 2015. A survey of techniques for event detection in twitter. *Computational Intelligence*, 31(1):132–164.

Siyuan Chen, Meilin Wang, Minghao Lv, Zhiling Zhang, Juqianqian Juqianqian, Dejiyangla Dejiyangla, Yujia Peng, Kenny Q. Zhu, and Mengyue Wu. 2024. Mapping long-term causalities in psychiatric symptomatology and life events from social media. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 5472–5487.

Siyuan Chen, Zhiling Zhang, Mengyue Wu, and Kenny Q. Zhu. 2023. Detection of multiple mental disorders from social media with two-stream psychiatric experts. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 9071–9084, Singapore. Association for Computational Linguistics.

Arman Cohan, Bart Desmet, Andrew Yates, Luca Soldaini, Sean MacAvaney, and Nazli Goharian. 2018. Smhd: a large-scale resource for exploring online language usage for multiple mental health conditions. In *27th International Conference on Computational Linguistics (COLING 2018)*, pages 1485–1497. Association for Computational Linguistics (ACL).

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.

Barbara Snell Dohrenwend, Larry Krasnoff, Alexander R. Askenasy, and Bruce P. Dohrenwend. 1978. Exemplification of a method for scaling life events: the peri life events scale. *Journal of health and social behavior*, 19 2:205–29.

Joseph L Fleiss, Bruce Levin, Myunghee Cho Paik, et al. 1981. The measurement of interrater agreement. *Statistical methods for rates and proportions*, 2(212-236):22–23.

Manas Gaur, Amanuel Alambo, Joy Prakash Sain, Ugur Kursuncu, Krishnaprasad Thirunarayan, Ramakanth Kavuluru, Amit Sheth, Randy Welton, and Jyotishman Pathak. 2019. Knowledge-aware assessment of severity of suicide risk for early intervention. In *The World Wide Web Conference*, WWW '19, page 514–525, New York, NY, USA. Association for Computing Machinery.

Oliver L Haimson, Albert J Carter, Shanley Corvite, Brookelyn Wheeler, Lingbo Wang, Tianxiao Liu, and Alexxus Lige. 2021. The major life events taxonomy: Social readjustment, social media information sharing, and online network separation during times of life transition. *Journal of the Association for Information Science and Technology*, 72(7):933–947.

Thomas H. Holmes and Richard H. Rahe. 1967. The social readjustment rating scale. *Journal of Psychosomatic Research*, 11(2):213–218.

Emma J. Howarth, Daryl B. O'Connor, Maria Panagioti, Alexander Hodkinson, Sarah Wilding, and Judith Johnson. 2020. Are stressful life events prospectively associated with increased suicidal ideation and behaviour? A systematic review and meta-analysis. *Journal of Affective Disorders*, 266:731–742.

Deniz Konac, Katherine S Young, Jennifer Lau, and Edward D Barker. 2021. Comorbidity between depression and anxiety in adolescents: Bridge symptoms and relevance of risk and protective factors. *Journal of psychopathology and behavioral assessment*, 43:583–596.

Chenliang Li, Aixin Sun, and Anwitaman Datta. 2012. Twevent: segment-based event detection from tweets. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 155–164.

Richard T. Liu and Ivan Miller. 2014. Life events and suicidal ideation and behavior: A systematic review. *Clinical Psychology Review*, 34(3):181–192.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

David E Losada and Fabio Crestani. 2016. A test collection for research on depression and language use. In *International conference of the cross-language evaluation forum for European languages*, pages 28–39. Springer.

Maike Luhmann, Wilhelm Hofmann, Michael Eid, and Richard E. Lucas. 2012. Subjective well-being and adaptation to life events: A meta-analysis. *Journal of Personality and Social Psychology*, 102(3):592–615.

Jornt Mandemakers. 2023. Life Events. In Filomena Maggino, editor, *Encyclopedia of Quality of Life and Well-Being Research*, pages 3841–3845. Springer International Publishing, Cham.

Gabriela Marc, Laurenţiu Mitrofan, and Camelia-Iulia-Maria Vlad. 2024. The relationship between critical life events, psycho-emotional health and life satisfaction among youths: Coping mechanisms and emotional regulation. *Frontiers in Psychology*, 14.

Scott M Monroe. 1982. Life events assessment: Current practices, emerging trends. *Clinical Psychology Review*, 2(4):435–453.

Van Quan Nguyen, Hyung-Jeong Yang, Young-chul Kim, Soo-hyung Kim, and Kyungbaek Kim. 2017. Social media based real-time event detection by using deep learning methods. *Smart Media Journal*, 6(3):41–48.

Martin Prince, Vikram Patel, Shekhar Saxena, Mario Maj, Joanna Maselko, Michael R Phillips, and Atif Rahman. 2007. No health without mental health. *The lancet*, 370(9590):859–877.

Judith G. Rabkin and Elmer L. Struening. 1976. Live Events, Stress, and Illness. *Science*, 194(4269):1013–1020.

Nils Reimers and Iryna Gurevych. 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)*, pages 3982–3992.

Alan Ritter, Mausam, Oren Etzioni, and Sam Clark. 2012. Open domain event extraction from twitter. In *Knowledge Discovery and Data Mining*.

David Robotham. 2008. Stress among higher education students: Towards a research agenda. *Higher education*, 56:735–746.

Judith A. Scully, Henry Tosi, and Kevin Banning. 2000. Life event checklists: Revisiting the social readjustment rating scale after 30 years. *Educational and Psychological Measurement*, 60(6):864–876.

Anselm Strauss and Juliet Corbin. 1998. Basics of qualitative research techniques.

Lisa Strohschein. 2005. Parental divorce and child mental health trajectories. *Journal of marriage and family*, 67(5):1286–1300.

Peeter Värnik. 2012. Suicide in the world. *International journal of environmental research and public health*, 9(3):760–771.

Jianshu Weng and Bu-Sung Lee. 2011. Event detection in twitter. In *Proceedings of the international aaai conference on web and social media*, volume 5, pages 401–408.

Haoran Yan, Xiaolong Jin, Xiangbin Meng, Jiafeng Guo, and Xueqi Cheng. 2019. Event detection with multi-order graph convolution and aggregated attention. 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)*, pages 5766–5770.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, and et al Chang Zhou. 2024. Qwen2 technical report. *ArXiv*, abs/2407.10671.

Zhiling Zhang, Siyuan Chen, Mengyue Wu, and Kenny Q. Zhu. 2022. Symptom identification for interpretable detection of multiple mental disorders on social media. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9970–9985. Association for Computational Linguistics.

A PsyEvent Dataset Construction

A.1 Life Events Definition

Selecting the appropriate definition of life events is crucial for effectively conducting mental health analysis from social media data. In light of the aforementioned issues, we aim to select life events that more closely align with the major occurrences in people's daily lives. Our definition of life events is based on the taxonomy proposed by Haimson et al. (2021), which identifies 121 life events organized into 12 categories. This taxonomy is grounded in a comprehensive survey conducted with a representative sample of 554 participants from the U.S. general population.

In the survey, participants were asked to describe, in open-ended paragraphs, the major life events they had experienced. The researchers analyzed these responses using qualitative open coding (Strauss and Corbin, 1998), a method for identifying patterns and themes in textual data. Through this process, they determined which life events to include in the taxonomy and how to categorize them. To refine the taxonomy further, the researchers engaged in in-depth discussions to address any surprising or ambiguous data points. They prioritized events that participants explicitly identified as having a significant impact on their lives. Additionally, the researchers incorporated insights from the life transitions literature, ensuring that less common but widely recognized major transitions were also included. To validate and enhance the practical application of the taxonomy, they conducted experiments, confirming its relevance and reliability for analyzing life events.

We adopted most of the life event taxonomy proposed by Haimson et al. (2021), with modifications driven by overlapping categories, annotation clarity, and pilot annotation feedback.

To enhance annotation consistency, we merged similar categories, such as *"Family Relationships"* and *"Relationships"*, and renamed the former as *"New Birth in Family"*, focusing on family member changes. Events related to interpersonal relationship shifts were grouped under the renamed *"Relationship Changes"* category. Stressful life events, namely "exams" and "substance abuse and addiction" were also added to account for their potential impact on mental health.

These adjustments resulted in a refined taxonomy comprising 12 life event categories and 127 sub-events. The finalized taxonomy reflects daily life events that influence mental health and is detailed in Appendix D.

A.2 Annotation Data Selection

Raw Dataset The raw dataset we started from is a subset of SMHD (Zhang et al., 2022; Cohan et al., 2018), whose data originated from subreddits related to mental illnesses on Reddit. Consequently, many users in the dataset had been diagnosed with one or more mental disorders and had experienced various stressful life events, making it a highly suitable initial dataset for detecting life events.

Data Preliminary Screening To enhance annotation efficiency and ensure that the data to be annotated contains as many life events as possible, rather than irrelevant information, we conducted a preliminary screening of the data that needed to be annotated. We initially utilized sentence-BERT (Reimers and Gurevych, 2019) to encode both posts and life events into embeddings, calculating the cosine similarity between them. If the similarity exceeded our predefined threshold, we selected that post as a candidate for annotation. Subsequently, to facilitate annotation, we filtered out posts that were either too short or excessively long.

A.3 Iterative Annotator Training

We invited well-educated volunteers to conduct the annotation, including students majoring in psychology. Our entire annotation process is as follows:

Training Session: We provided detailed training documents to instruct participants on the annotation methods. We personally explained the training documents through video conferences and demonstrated some annotation examples.

Annotator Practice: We collected test questions from samples where the authors had reached a consensus in the preliminary annotations, created a question bank, and prepared it for practice. We extracted a certain number of practice questions from the pre-established question bank to create tests. After each test, annotators were able to see where their decisions differed from ours. We provided explanations for each question to help them better understand our annotation requirements. Annotators are only allowed to participate in formal annotation after achieving a sufficient score in their practice sessions.

Formal Annotation Annotators who have passed the practice session can proceed to formal annotation. Additionally, annotators will continuously discuss any issues encountered during the annotation process with us. Annotators conducted their practice and annotation on the system that we had established. The introduction to the annotation system is provided in Appendix E.

A.4 Quality Control

In addition to rigorous training for annotators, we have also implemented other methods to ensure the quality of the annotation. We conducted checks on the annotators' annotated content. We regularly selected 10% of the completed annotations for review and scored them. At the same time, we promptly corrected any annotations we deemed inappropriate. If the weighted score of the reviewed content was below the threshold we had set, all annotations in that batch we checked were invalidated and redistributed to the annotators for re-annotation.

Additionally, each post underwent two rounds of annotation, meaning that the same post was annotated by two different individuals to ensure the data quality. We calculated the Fleiss' kappa (Fleiss et al., 1981) between the two rounds of annotation to measure the inter-annotator agreement. Finally, the average inter-annotator agreement for the two rounds of annotation was 0.67 falling within substantial agreement.

A.5 Annotation Platform Development

Due to the extensive context required to identify life events, we designed a specific annotation method. We built a website to serve as an annotation system; a screenshot of the website can be found in Appendix E. The system presents each annotator with a post and highlights one sentence within that post. Annotators can select any sentence from the post and label it with a life event tag. Additionally, posts that fail the checks are redistributed to annotators by the system.

A.6 Labels and Data Splits

The vast majority of social media content is unrelated to life events. To better enable the model to recognize life events in posts, we have incorporated a subset of annotated sentences that do not contain life events as control sentences into the dataset as well.

The data underwent two rounds of annotation, resulting in multiple annotations for each post, and we accept any single annotation as valid. To consolidate the multiple annotations for each sentence into a single golden label, we adopt the following approach (for further explanation of the annotation labels, please refer to Appendix E): we consider a sentence to contain a life event if either of the two annotators identifies it as such, and we determine that the poster is experiencing a life event only if both annotators agree that the poster is going through a life event.

B Life Events Recognition System

B.1 Life Events Detection

An individual may experience multiple life events throughout their life, and when posting, the poster may describe several life events simultaneously. As a result, there may be instances where a single post contains multiple life events. Therefore, we treat life event detection as a multi-label binary classification task. When extracting life events from posts, we need the model to determine whether each life event is associated with each sentence in the post. We employ a BERT-based encoder (Devlin et al., 2019) and apply a linear layer with a sigmoid activation function to the hidden state of the [CLS] token to predict the probabilities of all life events. Additionally, we use binary cross-entropy as the loss function to train the model. Because we incorporate control posts, we use a data balancing sampler to ensure that each training batch samples an equal amount of annotated data and control sentences.

Results We trained the life event detection model on *PsyEvent* using the `BERT-large` (Devlin et al., 2019) model (340M parameters). We follow a previous symptom extraction study (Zhang et al., 2022), where `BERT-large` model excels at various symptom recognition, and the results are shown in Table 3.

B.2 Self-Status Determination

In addition to the life event detection model, we have also designed a life event self-status determination model. When people narrate life events on social media, they might be referring to experiences from the past rather than current experiences or even events that happened to others around them rather than themselves. Therefore, we have developed a life event self-status determination model to account for these possibilities.

The purpose of the self-status determination model is to determine whether the *posters themselves* are currently experiencing the life events detected in sentences. Thus, we define this task as a single-label binary classification task. During annotation, we ask annotators to label whether the life events in the posts are those that the poster is personally and currently experiencing. We used the labels obtained from the processing method described in Section A.6 to train the self-status determination model. Additionally, to enable the model to better learn relevant textual features, we treat samples with life events that are not currently experienced by the poster as negative samples and those with life events that are currently experienced by the poster as positive samples, using this to train the self-status determination model (for detailed annotation labels, see Appendix E).

We employ the same model used for the life event detection task to accomplish the self-status determination task. We utilize a BERT-based model with a linear layer on top of the [CLS] representation, employing a sigmoid activation function to predict the probability of the life event status, and finally, we use binary cross-entropy as the loss function. Finally, we train the self-status determination model, which yields an AUC of 76.8. This status determination model is later used in combination with the Life Event Detection model to infer life events status in downstream tasks.

C Evaluation Metrics

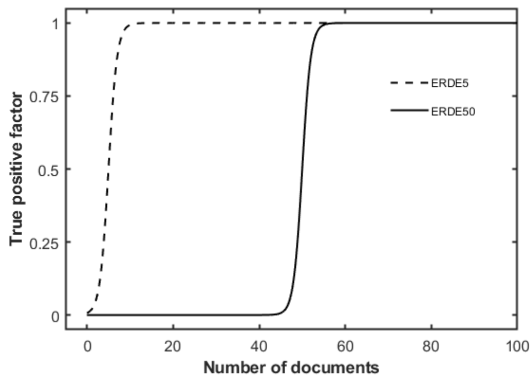

Figure 7: the cost factor $lc_o(k)$ for $ERDE_5$ and $ERDE_{50}$.

The following section introduces the evaluation metrics for two downstream tasks:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}$$

where TP represents true positives (samples correctly classified as positive), TN represents true negatives (samples correctly classified as negative), FP represents false positives (samples incorrectly classified as positive), and FN represents false negatives (samples incorrectly classified as negative).

For suicide risk prediction, the F1 score is used as the evaluation metric, providing a harmonic mean of precision and recall:

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

For the task of early risk detection of depression, the **Early Risk Detection Error (ERDE)** is defined as:

$$ERDE_o(d,k) = \begin{cases} c_{fp}, & \text{for } FP \\ c_{fn}, & \text{for } FN \\ lc_o(k) \cdot c_{tp}, & \text{for } TP \\ 0, & \text{for } TN \end{cases}$$

Here, c_{fp} and c_{fn} adjust the penalties for false positives (FP) and false negatives (FN), respectively. c_{fn} is set to 1, while c_{fp} is calculated as the ratio of positive cases in the dataset to the total number of users. The term $lc_o(k) \in [0, 1]$ represents the cost of delayed detection for true positives (TP), and c_{tp} quantifies the penalty for delays in detecting TP. Setting c_{tp} to 1 implies that a delayed detection is treated as equivalent to a missed detection.

The function $lc_o(k)$ governs the cost increase as the number of posts k grows and is defined as:

$$lc_o(k) = 1 - \frac{1}{1 + e^{k-o}}$$

The parameter o controls the inflection point along the x-axis where the cost escalates sharply. We use $ERDE_5$ and $ERDE_{50}$ as evaluation metrics for early depression detection, as shown in Figure 7.

D Detailed Life Events Taxonomy

Table 8 and 9 and 10 display the detailed life events and the 12 major categories.

Life Event Categories	Life Events
Health	personal injury , accident or illness; got violently attacked (including sexual assault); became disabled; mental illenss; recovery from mental health struggles; major surgery; hospitalization; pregnancy; pregnancy loss & abortion; menopause; abuse (including sexual abuse); began to self-harm; suicide attempt; substance abuse and addiction; recovery from addiction; oss of healthcare; physical fitness milestone; sex difficulties; change in health of family member; taking medicine
Financial	change in financial state; loan; home purchase; car purchase; other major purchase; home improvement; paid off debt; major financial difficulty; major financial gain; claimed bankruptcy; foreclosure; Mortgage; personal property damaged or stolen
Relocation	move within same town/city; move to a different town/city; move to a different state; move to a different country; move to a different country as a refugee; became a permanent resident or citizen of a new country; move out of parent's home; move in with family; family member moved into household; family member moved out of household; moved into assisted living; lost home / became homeless; major travel

Table 8: Life events category 1-3, with its respective sub-events.

Life Event Categories	Life Events
Legal	got arrested; lawsuit or legal action; turned over power of attorney; loss of driver's license / DUI; went to jail or prison; released from jail or prison; minor violations of the law
Relationship Changes	began serious romantic relationship; ended serious romantic relationship; Engagement; ended engagement; marriage; divorce; marital separation; marital reconciliation; relationship became abusive; serious argument with neighbor or friend; change in number of arguments; serious argument with family member or relative; trouble with in-laws; family betrayal; parenting difficulties
New Birth in Family	gain of new family member; gave birth / became a parent; adopted a child; became a grandparent; became a great-grandparent; became an aunt/uncle
Death	death of spouse; death of child; death of parent; death of pet; death of a friend; death of a loved one; death of extended family member

Table 9: Life events category 4-7, with its respective sub-events.

E Annotation System

The context required to determine life events in social media posts is relatively long. Hence, we have designed a set of annotation methods tailored for life events. We need to provide annotators with a complete post and enable them to flexibly select any sentence within the post for annotation. The example of the annotation website is shown in Figure 8. The annotation interface displays the post that the website user is currently annotating on the left side, with the sentence currently being annotated highlighted. Users can click buttons to select the previous or next sentence. On the right side of the annotation interface are the life event annotation buttons, which users can click to tag sentences with one of the 12 life event categories. We provide three annotation options of life events for users: the sentence does not contain the life event, the sentence contains the life event but the poster is not currently experiencing it, and the sentence contains the life event that the poster is currently experiencing.

F Model and Computation Resource

The models used in our work include: bert-base (110M parameters), bert-large (340M parameters)

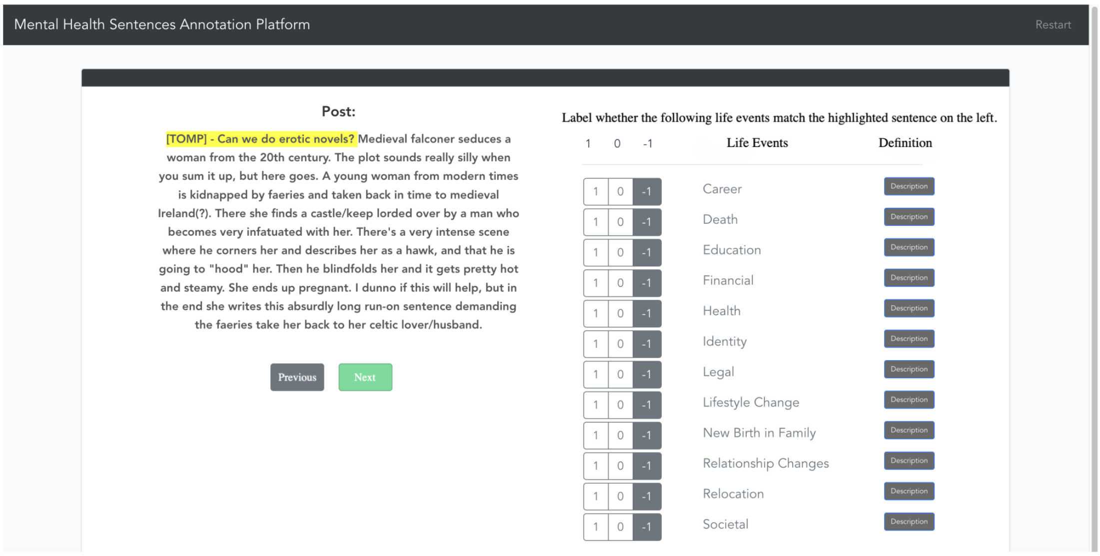

Figure 8: Website Annotation Interface Example.

(Devlin et al., 2019), roberta-large (355M parameters) (Liu et al., 2019), and qwen2-7B (Yang et al., 2024). The GPU resources we used are two NVIDIA RTX 4090 cards.

G Experiment Setting

When training life event classification models, we set the learning rate to 2e-3, max length to 1024, and batch size to 64. To prevent overfitting, the early stopping patience was configured at 4 epochs. In the ERDD and SRP experiments, we set the batch size to 64, the learning rate to 5e-3, the max length to 1024, and the early stopping patience to 4.

Life Event Categories	Life Events
Career	started a new job; change in responsibilities at work; change to a different line of work; change in work hours or conditions; business readjustment; promotion; demotion; significant success at work; troubles at work; workplace discrimination or harassment; voluntary job loss (e.g., quit); involuntary job loss (e.g., fired); became a business owner / entrepreneur; retirement; unable to find work; spouse begins or stops work
Education	begin or end school/college; change in school/college; left school (without graduating); denied entry into school; obtained a certification; examination
Lifestyle Change	change in physical habits; change in responsibilities in personal life; new pet; joined the military; returned to civilian life after military; change in living conditions; revision of personal habits; change in recreation; change in social activities; vacation
Identity	identified sexual preference; identified gender; came out as LGBTQ+; gender transition; change in political beliefs; change in religious/spiritual beliefs or practices; coming of age ceremony; new sexual experience; another major identity shift
Societal	natural disaster; war; major political event that had personal impact; met a celebrity

Table 10: Life events category 8-12, with its respective sub-events.