Capturing the Dynamics of Mental Well-Being: Adaptive and Maladaptive States in Social Media

Anastasia Sandu[1], **Teodor Mihailescu**[1], **Ana Sabina Uban**[1,2], **Ana-Maria Bucur**[3,4]
[1]Faculty of Mathematics and Computer Science, [2]HLT Research Center,
[3]Interdisciplinary School of Doctoral Studies, University of Bucharest
[4]PRHLT Research Center, Universitat Politècnica de València
anastasiasandu777@gmail.com, teomihailescu@yahoo.com
auban@fmi.unibuc.ro, ana-maria.bucur@drd.unibuc.ro

Abstract

This paper describes the contributions of the BLUE team in the CLPsych 2025 Shared Task on Capturing Mental Health Dynamics from Social Media Timelines. We participate in all tasks with three submissions, for which we use two sets of approaches: an unsupervised approach using prompting of various large language models (LLM) with no fine-tuning for this task or domain, and a supervised approach based on several lightweight machine learning models trained to classify sentences for evidence extraction, based on an augmented training dataset sourced from public psychological questionnaires. We obtain the best results for summarization Tasks B and C in terms of consistency, and the best F1 score in Task A.2.

1 Introduction

The assessment of mental health through digital technologies is an increasingly important topic in both psychology and natural language processing. Digital mental health tools can support individuals in need and facilitate remote care, especially as the prevalence of mental disorders continues to rise while access to mental health services remains limited.[1] Most approaches for mental health assessment using online data are focused on performing binary classification for depression (Yates et al., 2017; Liu et al., 2023) or suicide risk (Coppersmith et al., 2018; Ramírez-Cifuentes et al., 2020; Lee et al., 2020), with few previous works focused on explainable mental health assessment (Wang et al., 2024; Bao et al., 2024; Uban et al., 2022). The CLPsych Workshop was the first to address the challenge of extracting evidence from social media data by proposing the task of highlighting evidence for mental disorders (Chim et al., 2024). This year's task builds on that foundation, shifting the focus to mental states, specifically adaptive and maladaptive states.

The Shared Tasks from CLPsych 2022 (Tsakalidis et al., 2022) and 2024 (Chim et al., 2024) focused on analyzing longitudinal user posts. The 2022 task focused on capturing moments of change from the social media timeline of a user, while the 2024 task aimed to extract evidence regarding the suicide risk of users. Similar to the shared task from this year, the 2024 edition included a summarization component, which required participants to provide textual summaries of the mental health dynamics throughout the entire timeline of the user. While the extraction of evidence from social media data is a relatively new task, it was previously modeled as a binary classification task for maladaptive states (Gollapalli et al., 2023, 2024).

In this paper, we present the contributions of the BLUE team to the CLPsych 2025 Shared Task: Capturing Mental Health Dynamics from Social Media Timelines (Tseriotou et al., 2025). Our approach relies on both classical machine learning algorithms and LLMs, merging established classification methods with recent advancements in the field. Our team achieved good results, scoring the highest in summarization Tasks B and C. Moreover, for highlighting evidence of adaptive and maladaptive states, as well as inferring the well-being score (Tasks A.1 and A.2), our team ranks fifth.

2 Data and Tasks

The data provided for this task consists of Reddit posts annotated by domain experts for self-states following the MIND Framework (Slonim, 2024). This dataset aligns with prior work in computational linguistics and clinical psychology, particularly studies on suicide risk assessment (Shing

[1]https://www.who.int/news/item/17-06-2022-who-highlights-urgent-need-to-transform-mental-health-and-mental-health-care

Proceedings of the 10th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2025), pages 225–234
May 3, 2025 ©2025 Association for Computational Linguistics

et al., 2018; Zirikly et al., 2019) and longitudinal mental health analysis (Tsakalidis et al., 2022). The dataset allows for the evaluation of self-states and well-being, facilitating our contributions to mental health analysis.

The tasks for CLPsych 2025 are as follows:

Task A.1 Identification of adaptive and maladaptive self-states within each post.

Task A.2 Prediction of well-being score for each post, ranging from 1 (low well-being) to 10 (high well-being).

Task B Generation of a summary describing the interplay between adaptive and maladaptive self-states within a post.

Task C Generation of a timeline-level summary encapsulating the evolution of self-states across multiple posts from the same user.

3 Method

3.1 Machine learning approach

We used a machine learning method for Task A.1 to perform sentence classification and identify adaptive and maladaptive states in text using text embeddings and supervised classifiers. For extracting text embeddings, we use TF-IDF (Sparck Jones, 1972), BERT (Devlin et al., 2019) and SentenceTransformer (Reimers and Gurevych, 2019). XGBoost (Chen and Guestrin, 2016) was used for identifying maladaptive states, while Logistic Regression (Cox, 1958) was used for detecting adaptive ones. The hyperparameters used for the two classifiers are presented in Appendix A.4 and A.5. For feature extraction, we applied TF-IDF using the default configuration (Appendix A.6).

These models are trained on the labeled data provided for this task. To augment the dataset used for training, we include external psychological resources for better generalization. The external data sources used for maladaptive states include the items from the Young Schema Questionnaire (YSQ) (Young, 2003b), the Young Schema Questionnaire-Revised (YSQ-R) (Young, 2003a), and annotated texts from Liu et al. (2022). To enhance the dataset for adaptive states, we expanded it by including items from the Young Positive Schema Questionnaire (YPSQ) (Louis et al., 2018). Since the data for adaptive states was limited, we supplemented these samples by generating additional samples[2] using GPT-o1 (OpenAI, 2024b) and GPT-4o (OpenAI, 2024a), based on the YPSQ items.

Classification is performed at the sentence level using majority voting across multiple predictions, with Gaussian noise added to reduce overfitting.

3.2 LLMs approach

While the machine learning approach was applied only for Task A.1, LLMs were used for all tasks from CLPsych 2025. For the LLM-based approach, we implement a structured processing pipeline that uses multiple models. This pipeline consists of the following steps:

Post-Level Processing: Each post undergoes the following steps:

- **Evidence Extraction:** The model identifies adaptive and maladaptive self-state evidence in the post text.
- **Well-Being Prediction:** A well-being score is assigned to the post on a scale of 1 to 10, based on predefined psychological criteria.
- **Post Summary Generation:** The model summarizes the interplay between adaptive and maladaptive self-states within the post.

Timeline-Level Processing: After processing individual posts, the full timeline is analyzed:

- **Aggregation of Posts:** All posts from a single user are compiled into a coherent timeline.
- **Timeline Summary Generation:** The model generates a high-level summary describing the evolution of self-states over time.

This structured approach allows for a detailed analysis of self-states across individual posts and entire timelines, providing insights into psychological well-being and behavioral patterns.

In this approach, we rely on LLM prompting to solve the tasks. We use models in various families: Gemma 2 9B (Team et al., 2024), Mistral 7B (AI, 2024b), Llama 2 7B (Touvron et al., 2023), Llama 3.1 8B (Grattafiori et al., 2024), and Llama 3.2 3B (AI, 2024a). These models were selected based on their ability to process complex psychological

[2]We make the generated data available on github, together with the code used for the submissions: https://github.com/Teo1230/clpsych25-task

text while balancing computational efficiency and accuracy.

We experiment with three different types of prompts in our work. The first is the *default prompt*, which provides instructions for extracting evidence, predicting well-being scores, and summarizing information without including additional context or definitions of concepts. The second prompt, referred to as the *expert prompt*, guides the model to evaluate Reddit posts as if it were a psychology expert. This prompt includes definitions of adaptive and maladaptive states and asks the model to generate summaries based on key emotional, cognitive, and behavioral patterns. The third prompt, *structured summarization prompt*, focuses on generating summaries using the structure proposed by the CLPsych Shared Task. This emphasizes determining the dominant self-state and describing the interplay between adaptive and maladaptive self-states. The complete prompts can be found in Appendix A.

To ensure a balance between determinism and variability in generation, we experimented with different temperature settings across models. The temperature parameter controls the randomness of the model's responses: lower values make outputs more deterministic, while higher values introduce more diversity. For our experiments, we used a temperature of 0.7 for the Gemma 2 9B, Llama 3.1 8B, Llama 3.2 3B, and a temperature of 0.5 for Mistral 7B and Llama 2 7B. These settings were selected based on preliminary trials to balance consistency and adaptability in handling complex psychological discourse. All models used a top-k sampling of 40 and a top-p of 0.9. Our methodology relies on prompting rather than fine-tuning for several reasons:

Domain-Specific Adaptability Fine-tuning requires large domain-specific datasets and extensive computational resources, which may not generalize well to unseen cases. Prompting allows leveraging LLMs' broad pretraining without retraining.

Flexibility in Task Definitions By designing different prompts, we can easily modify task instructions without retraining models. This is crucial for a field like mental health, where criteria may evolve.

Clinical Interpretability Using well-defined prompts provides clearer interpretability compared to fine-tuned black-box models, making the approach more suitable for clinical applications where transparency is essential.

While fine-tuning could improve model specialization, it introduces several challenges, such as the need for large, annotated datasets specific to adaptive or maladaptive self-states, increased computational costs for training and inference, and the potential loss of generalizability across different domains. Given these limitations, our focus remains on optimizing prompting strategies while leveraging pre-trained LLMs.

3.3 Evaluation Metrics

Our system is evaluated using established metrics from prior CLPsych shared tasks (Zirikly et al., 2019; Tsakalidis et al., 2022):

Task A.1: Recall and Weighted Recall for adaptive and maladaptive self-state identification. The Recall metric is represented by the maximum recall-oriented BERTScore (Zhang et al., 2019).

Task A.2: Mean Squared Error (MSE) across different well-being categories and macro F1-score.

Task B: Consistency, maximum contradiction, and maximum entailment scores.

Task C: Mean consistency and maximum contradiction.

3.4 Submissions

In this section, we present our submissions for the CLPsych 2025 Shared Task (Table 1).

Submission 1 This submission consists of multiple LLMs within a common processing approach to analyze user timelines in Reddit posts, extracting psychological insights through structured tasks. **Gemma 2** is used for evidence extraction and well-being scoring (Tasks A.1 and A.2) using default prompts (Appendix A.1). Meanwhile, **LLaMA 3.2** generates summaries for the post (Task B) and timeline levels (Task C) using the default prompts.

Submission 2 This submission integrates both machine learning classifiers and LLMs. For evidence extraction (Task A.1), we employ supervised classifiers trained on labeled data. **XGBoost** is utilized to identify maladaptive states, while **Logistic Regression** is used for classifying adaptive states. TF-IDF is used for text representation. **Mistral** handles well-being scoring (Task A.2) with the default prompt. Summarization at both the post (Task B) and timeline levels (Task C) is performed using

Team BLUE	Task A.1 - Adaptive	Task A.1 - Maladaptive	Task A.2	Task B	Task C
Submission 1	Gemma 2_D	Gemma 2_D	Gemma 2_D	LLaMA 3.2_D	LLaMA 3.2_D
Submission 2	TF-IDF & LR	TF-IDF & XGB	Mistral_D	LLaMA 3.2_{SS}	LLaMA 3.2_{SS}
Submission 3	Gemma 2_E	Gemma 2_E	Gemma 2_E	LLaMA 3.1_D	LLaMA 3.1_D
Submission 4	TF-IDF & LR	BERT & XGBoost	$\text{Mistral}_{E,L}$	Mistral_E	Mistral_E
Submission 5	MiniLM & LR	MiniLM & XGBoost	LLaMA 3.1_D	LLaMA 2_D	LLaMA 2_D
Submission 6	Gemma $2_{D,L}$	Gemma $2_{D,L}$	LLaMA $3.2_{E,L}$	Gemma 2_D	Gemma $2_{D,L}$

Table 1: Approaches used for our team's submissions across all tasks. D denotes *default prompt*, E denotes *expert prompt*, SS - *structured summarization prompt*, and L denotes that LangChain was used for the prompt template.

Team BLUE	Task A.1: Recall	Task A.2: MSE	Task B: Mean Consistency	Task C: Mean Consistency
Submission 1	0.555	2.390	**0.910**	**0.946**
Submission 2	0.539	2.900	0.328	0.854
Submission 3	0.538	**2.260**	0.393	0.911
Submission 4	0.444	3.164	0.908	0.913
Submission 5	0.422	3.025	0.918	0.897
Submission 6	**0.569**	3.842	0.890	0.900
Ranking Team	5	5	1	1

Table 2: Final evaluation results for our team's submissions across all tasks.

LLaMA 3.2, leveraging the structured summarization prompt (Appendix A.3).

Submission 3 In this method, we again use multiple LLMs to analyze user timelines in Reddit posts. **Gemma 2** is used for evidence extraction and well-being scoring (Tasks A.1 and A.2) with the expert prompts (Appendix A.2). In addition, **LLaMA 3.1** generates summaries for the post (Task B) and timeline levels (Task C) using default prompts.

Submission 4 This submission combines machine learning and LLMs. For evidence extraction (Task A.1), we use **BERT** (bert-base-uncased) (Devlin et al., 2019) to generate sentence embeddings, which are then classified by **XGBoost** for detecting maladaptive states. Adaptive states are identified using **TF-IDF** features with **Logistic Regression**. For well-being scoring (Task A.2), we use **Mistral** with the expert prompt and LangChain (Chase, 2022). The same model and prompt are used for summarizing posts (Task B) and timelines (Task C).

Submission 5 For evidence extraction (Task A.1), we use the same method as in Submission 4, but instead of generating sentence embeddings with BERT, we experiment with **all-MiniLM-L6-v2**[3]. For well-being scoring (Task A.2), we use **LLaMA 3.1** with the default prompt, while **LLaMA 2** handles summarization (Tasks B and C), also employing the default prompt.

Submission 6 For evidence extraction (Task A.1), we use **Gemma 2** with the default prompt, implemented with LangChain. Well-being scoring (Task A.2) is handled by **LLaMA 3.2** with the expert prompt, also using LangChain. Summarization at both the post (Task B) and timeline levels (Task C) is performed with **Gemma 2**, using the default prompt for Task B and the default prompt with LangChain for Task C.

4 Results

Our system demonstrates strong performance across the tasks, particularly in summarization (Tasks B and C). The results of our submissions are presented in Table 2. The first three submissions in Table 2 are the official submissions for the CLPsych 2025 Shared Task, while the remaining three submissions are additional runs that were not submitted officially.

We performed best in summarization (Tasks B and C), achieving top consistency scores. For well-being scoring (Task A.2), Submission 3 achieved the smallest MSE of 2.260, placing us at rank 5. Submission 6 reached the highest recall on Task A.1 but had a higher MSE, indicating that while our methods are strong at summarizing timelines, they still struggle with the finer details of post-level scoring—especially when detecting adaptive signals.

We found it more challenging to extract adaptive

[3]https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

	Recall			Weighted Recall		
Team BLUE	**Overall**	**Adaptive**	**Maladaptive**	**Overall**	**Adaptive**	**Maladaptive**
Submission 1	0.555	**0.472**	0.639	0.392	0.400	0.384
Submission 2	0.539	0.298	**0.779**	0.239	0.285	0.192
Submission 3	0.538	0.414	0.662	0.389	0.351	**0.428**
Submission 4	0.444	0.298	0.589	0.326	0.286	0.365
Submission 5	0.422	0.303	0.540	0.334	0.291	0.376
Submission 6	**0.569**	0.457	0.681	**0.393**	**0.403**	0.382

Table 3: Evaluation results for Task A.1 for adaptive and maladaptive self-states.

evidence than maladaptive evidence, mostly because people tend to describe distress with clearer cues, while adaptive statements are often subtle and less standardized. Another factor is data imbalance: our original training dataset leaned heavily toward maladaptive examples, as prior work in mental health analysis has traditionally focused on distress or at-risk behaviors. To address this, we added data from the Young Positive Schema Questionnaire (YPSQ) (Louis et al., 2018) and generated more adaptive statements with GPT-o1 and GPT-4o. Although this helped balance the data and improve recall, identifying adaptive language is still challenging, which is reflected in the results presented in Table 3.

5 Conclusions and Future Work

The CLPsych 2025 Shared Task on Capturing Mental Health Dynamics from Social Media Timelines proposed a novel problem that has not been approached computationally in the past, related to adaptive and maladaptive states, in a variety of different tasks. Our team participated with three submissions using two sets of approaches: one based on prompting various LLMs for all tasks and a supervised approach based on classical machine learning models trained on the provided training data as well as external data, including expert-generated data (from relevant psychological questionnaires) and AI-generated. While we experiment and include in our submissions different kinds of LLM models and prompts, as well as different machine learning models for the second approach, our methods are relatively cheap and accessible, and our good results across tasks confirm that relatively simple approaches can be effective for identifying adaptive and maladaptive states in social media texts. In-context learning was minimal, with the only external knowledge provided to the models including a description of the scoring scheme for some of the prompts. The supervised approaches include classical machine learning algorithms (which performed better in this setting than pretrained transformers according to our preliminary experiments). All LLM models are general domain, with only Llama2 7B, Llama3.1 8B, Llama3.2 3B, Mistral 7B, Gemma 2 9B parameters, run using modest infrastructure. Using these relatively accessible approaches, we obtain competitive results compared to the other participants, with the best mean consistency score out of all teams in both summarization tasks (Tasks B and C), the 5th MSE score for well-being score (Task A.2) and the 5th Recall for Task A.1 related to evidence highlighting.

Future research should look at more diverse datasets to make sure our approaches work for different populations. Also, exploring specialized or fine-tuned LLMs that incorporate domain knowledge from psychology or clinical practice could further enhance both performance and interpretability in mental health tasks. Another direction is investigating more advanced or ensemble-based machine learning methods to improve the detection and classification of adaptive and maladaptive states.

Limitations

The data for this task primarily consists of Reddit posts, which may not accurately reflect the broader population. Social media often reveals biases related to factors such as gender and socioeconomic status, meaning our findings might not be applicable to all groups, particularly beyond American males (Gottfried, 2024). Furthermore, the collected posts may contain incomplete or misleading information, as users do not always provide factual or comprehensive details online. While the performance of our LLM-based solutions was limited by our infrastructure, our good results show that reasonable performance is achievable for this task, even with relatively small, generic LLMs.

Ethical Considerations

The data we used was obtained through a strict data agreement to ensure we adhered to ethical guidelines for handling sensitive information. We prioritize individual privacy and confidentiality by conducting all analyses locally, without using any external APIs that could compromise data security. We follow ethical research guidelines from Benton et al. (2017) for the sensitive data provided for this shared task. We recognize the potential impact of our findings on individuals facing mental health challenges. It is crucial to approach these analyses with sensitivity and to consider the broader societal implications of our work. Our goal is to make a positive contribution to mental health research while upholding ethical integrity.

Acknowledgements

This research was partially supported by the project "Romanian Hub for Artificial Intelligence - HRIA", Smart Growth, Digitization and Financial Instruments Program, 2021-2027, MySMIS no. 334906.

References

Meta AI. 2024a. Llama 3.2: Open foundation language models. https://github.com/meta-llama/llama-models.

Mistral AI. 2024b. Mistral models: High-performance open-weight language models.

Eliseo Bao, Anxo Pérez, and Javier Parapar. 2024. Explainable depression symptom detection in social media. *Health Information Science and Systems*, 12(1):47.

Adrian Benton, Glen Coppersmith, and Mark Dredze. 2017. Ethical research protocols for social media health research. In *Proceedings of the first ACL workshop on ethics in natural language processing*, pages 94–102.

Harrison Chase. 2022. LangChain. Software. Released on 2022-10-17.

Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 785–794. ACM.

Jenny Chim, Adam Tsakalidis, Dimitris Gkoumas, Dana Atzil-Slonim, Yaakov Ophir, Ayah Zirikly, Philip Resnik, and Maria Liakata. 2024. Overview of the clpsych 2024 shared task: Leveraging large language models to identify evidence of suicidality risk in online posts. In *Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024)*, pages 177–190.

Glen Coppersmith, Ryan Leary, Patrick Crutchley, and Alex Fine. 2018. Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights*, 10:1178222618792860.

D. R. Cox. 1958. The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2):215–242.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, pages 4171–4186.

Sujatha Gollapalli, Beng Ang, and See Kiong Ng. 2023. Identifying early maladaptive schemas from mental health question texts. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 11832–11843.

Sujatha Das Gollapalli, Beng Heng Ang, Mingzhe Du, and See-Kiong Ng. 2024. Counseling responses for mental health forum questions with early maladaptive schema prediction. In *ECAI 2024*, pages 2556–2563. IOS Press.

Jeffrey Gottfried. 2024. Americans' social media use. *Pew Research Center*, 31.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Daeun Lee, Soyoung Park, Jiwon Kang, Daejin Choi, and Jinyoung Han. 2020. Cross-lingual suicidal-oriented word embedding toward suicide prevention. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2208–2217.

X. Liu, Y. Zhang, and C. Tan. 2022. Identifying early maladaptive schemas from mental health question texts. *Proceedings of EMNLP*.

Yujian Liu, Laura Biester, and Rada Mihalcea. 2023. Improving mental health classifier generalization with pre-diagnosis data. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, pages 566–577.

J. P. Louis, A. M. Wood, G. Lockwood, M.-H. R. Ho, and E. Ferguson. 2018. Positive clinical psychology and schema therapy (st): The development of the young positive schema questionnaire (ypsq) to complement the young schema questionnaire 3 short form (ysq-s3). *Psychological Assessment*, 30(9):1199–1213.

OpenAI. 2024a. Gpt-4o system card.

OpenAI. 2024b. Openai o1 system card.

Diana Ramírez-Cifuentes, Ana Freire, Ricardo Baeza-Yates, Joaquim Puntí, Pilar Medina-Bravo, Diego Alejandro Velazquez, Josep Maria Gonfaus, and Jordi Gonzàlez. 2020. Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis. *Journal of medical internet research*, 22(7):e17758.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *Proceedings of EMNLP*, page 3982–3992.

Han-Chin Shing, Suraj Nair, Ayah Zirikly, Meir Friedenberg, Hal Daumé III, and Philip Resnik. 2018. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic*, pages 25–36.

Dana Atzil Slonim. 2024. Self-other dynamics (sod): A transtheoretical coding manual.

Karen Sparck Jones. 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1):11–21.

Gemma Team, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, et al. 2024. Gemma 2: Improving open language models at a practical size. *arXiv preprint arXiv:2408.00118*.

Hugo Touvron, Louis Martin, Kevin Stone, Max Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Siddhartha Batra, Pallavi Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Adam Tsakalidis, Jenny Chim, Iman Munire Bilal, Ayah Zirikly, Dana Atzil-Slonim, Federico Nanni, Philip Resnik, Manas Gaur, Kaushik Roy, Becky Inkster, et al. 2022. Overview of the clpsych 2022 shared task: Capturing moments of change in longitudinal user posts. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, pages 184–198.

Talia Tseriotou, Jenny Chim, Ayal Klein, Aya Shamir, Guy Dvir, Iqra Ali, Cian Kennedy, Guneet Singh Kohli, Anthony Hills, Ayah Zirikly, Dana Atzil-Slonim, and Maria Liakata. 2025. Overview of the clpsych 2025 shared task: Capturing mental health dynamics from social media timelines. In *Proceedings of the 10th Workshop on Computational Linguistics and Clinical Psychology*. Association for Computational Linguistics.

Ana-Sabina Uban, Berta Chulvi, and Paolo Rosso. 2022. Explainability of depression detection on social media: From deep learning models to psychological interpretations and multimodality. In *Early Detection of Mental Health Disorders by Social Media Monitoring: The First Five Years of the eRisk Project*, pages 289–320. Springer.

Yuxi Wang, Diana Inkpen, and Prasadith Kirinde Gamaarachchige. 2024. Explainable depression detection using large language models on social media data. In *Proceedings of the 9th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2024)*, pages 108–126.

Andrew Yates, Arman Cohan, and Nazli Goharian. 2017. Depression and self-harm risk assessment in online forums. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2968–2978.

J. E. Young. 2003a. Young schema questionnaire - revised (ysq-r).

J. E. Young. 2003b. Young schema questionnaire (ysq).

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.

Ayah Zirikly, Philip Resnik, Ozlem Uzuner, and Kristy Hollingshead. 2019. Clpsych 2019 shared task: Predicting the degree of suicide risk in reddit posts. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology*, pages 24–33.

A Appendix

In this appendix, we provide the prompts used for analyzing self-states in Reddit posts. These prompts are designed to facilitate the extraction of adaptive and maladaptive self-state evidence, predict well-being scores, and generate summaries. The following prompts are used to guide the models in processing Reddit posts:

A.1 Default Prompts

- **Extract Evidence:**

  Given the following Reddit post, identify evidence of adaptive and maladaptive self-states. Extract text spans as JSON lists.
  Post: "{post_text}"
  Response format: {
  "adaptive_evidence": [<adaptive text spans>],
  "maladaptive_evidence": [<maladaptive text spans>]
  }

- **Predict Well-Being:**

  Given the following Reddit post, assign a well-being score from 1 (low) to 10 (high).

- **1**: The person is in persistent danger of severely hurting self or others...
- **2**: In danger of hurting self or others...
- **3**: A person experiences delusions or hallucinations...
- **4**: Some impairment in reality testing or communication...
- **5**: Serious symptoms (e.g., suicidal thoughts)...
- **6**: Moderate symptoms (e.g., panic attacks)...
- **7**: Mild symptoms (e.g., depressed mood)...
- **8**: If symptoms are present, they are temporary...
- **9**: Absent or minimal symptoms...
- **10**: No symptoms and superior functioning...

Post: "post_text"
Response format: { "wellbeing_score": <score> }

- **Summarize Post:**

  Given the following Reddit post, summarize the interplay between adaptive and maladaptive self-states.
  Post: "post_text"
  Response format: { "summary": "<post-level summary>" }

- **Summarize Timeline:**

  Given the following series of Reddit posts from one user, generate a timeline-level summary. Begin by determining which self-state is dominant (adaptive/maladaptive) and describe it first.
  Timeline: "timeline_text"
  Response format: { "summary": "<timeline-level summary>" }

A.2 Expert Prompts

- **Extract Evidence:**

  You are an expert in **psychological self-states and mental health analysis**. Your task is to analyze the Reddit post below and extract textual evidence that indicates **adaptive and maladaptive self-states**.
  - **Adaptive self-states**: Indicate resilience, coping, self-awareness, or positive cognitive and behavioral patterns.
  - **Maladaptive self-states**: Indicate distress, negative cognitive distortions, emotional dysregulation, or harmful behaviors.

  Post:
  "post_text"
  Response format (strict JSON):
  {
  "adaptive_evidence": [<text spans that show adaptive self-states>],
  "maladaptive_evidence": [<text spans that show maladaptive self-states>]
  }

- **Predict Well-Being:**

  You are a clinical expert in **mental health assessment**. Your task is to assign a **well-being score (1-10)** to the Reddit post below based on its emotional, cognitive, and behavioral indicators.
  - **1**: The person is in persistent danger of severely hurting self or others...
  - **2**: In danger of hurting self or others...
  - **3**: A person experiences delusions or hallucinations...
  - **4**: Some impairment in reality testing or communication...
  - **5**: Serious symptoms (e.g., suicidal thoughts)...
  - **6**: Moderate symptoms (e.g., panic attacks)...
  - **7**: Mild symptoms (e.g., depressed mood)...
  - **8**: If symptoms are present, they are temporary...
  - **9**: Absent or minimal symptoms...
  - **10**: No symptoms and superior functioning...

Post:
"post_text"
Response format (strict JSON):
{ "wellbeing_score": <integer between 1 and 10> }

- **Summarize Post:**

  You are a **psychological expert analyzing self-states** in text. Your task is to **summarize by determining which self-state is dominant (adaptive/maladaptive) and describe it first, then how adaptive and maladaptive self-states interact within this post**.
  - Identify **key emotional, cognitive, and behavioral patterns**.
  - Highlight **contrasts between adaptive and maladaptive self-states**.
  - Provide an **objective, clinical-style summary**.
  Post:
  "post_text"
  Response format (strict JSON):
  { "summary": "<concise analysis of self-states in the post>" }

- **Summarize Timeline:**

  You are a **clinical psychologist analyzing mental health trends over time**. Given the following series of Reddit posts from a single user, summarize their **self-state trajectory**.
  - Identify **patterns of emotional and cognitive change**.
  - Note **shifts between adaptive and maladaptive self-states**.
  - Highlight **any signs of improvement, deterioration, or instability**.
  Timeline:
  "timeline_text"
  Response format (strict JSON):
  { "summary": "<timeline-level psychological summary>" }

A.3 Structured Summarization Prompts

- **Summarize Post:**

  Analyze the following post in a clinical, objective manner. Identify both adaptive and maladaptive self-states, capturing the interplay between them and provide a clear, concise summary.
  Post:
  "post_text"
  Response format (strict JSON):
  { "summary": "<post-level summary>" }

- **Summarize Timeline:**

  Given the following series of Reddit posts from one user, generate a concise timeline-level summary of the evolution of self-states.
  Instructions:
  - Determine the overall dominant self-state (adaptive or maladaptive) and describe it first.
  - Describe how the interplay between adaptive and maladaptive self-states changes over time.
  - Emphasize any transitions, improvements, or deteriorations in emotional, cognitive, and behavioral aspects without referring to internal codes.
  - Ensure the summary is clear, natural, and coherent.

  Timeline:
  "timeline_text"
  Response format (strict JSON):
  { "summary": "<timeline-level summary>" }

A.4 Logistic Regression Parameters

- **Class weight balancing:** Enabled (`class_weight="balanced"`)
- **Maximum iterations:** 1000 (`max_iter=1000`)
- **Random seed:** 42 (`random_state=42`)

A.5 XGBoost Parameters

- **Number of estimators:** 200 (`n_estimators=200`)
- **Learning rate:** 0.1 (`learning_rate=0.1`)
- **Maximum tree depth:** 4 (`max_depth=4`)

A.6 TF-IDF Parameters

- **Lowercasing:** Enabled (`lowercase=True`)
- **Stop words:** None (`stop_words=None`)
- **N-gram range:** Unigrams only (`ngram_range=(1,1)`)
- **Max features:** None (`max_features=None`)