Transformer-Based Analysis of Adaptive and Maladaptive Self-States in Longitudinal Social Media Data

Abhin B and Renukasakshi V Patil
Department of Information Technology
National Institute of Technology Karnataka, Surathkal
Mangalore, India
{abhinb.211ai003@nitk.edu.in, renukasakshivpatil.211ai030@nitk.edu.in}

Abstract

The CLPsych workshop, held annually since 2014, promotes the application of computational linguistics to behavioral analysis and neurological health assessment. The CLPsych 2025 shared task, extending the framework of the 2022 iteration, leverages the MIND framework to model temporal fluctuations in mental states. This shared task comprises three subtasks, each presenting substantial challenges to natural language processing (NLP) systems, requiring sensitive and precise outcomes in analyzing adaptive and maladaptive behaviors. In this study, we employed a range of modeling strategies tailored to the requirements and expected outputs of each subtask. Our approach mostly utilized traditional language models like BERT, LongFormer and Pegasus diverging from the prevalent trend of prompt-tuned large language models. We achieved an overall ranking of 13th, with subtask rankings of 8th in Task 1a, 13th in Task 1b, 8th in Task 2, and 7th in Task 3. These results highlight the efficacy of our methods while underscoring areas for further refinement in handling complex behavioral data.

1 Introduction

Understanding mental health through digital footprints has become a critical area of research, with social media providing a unique lens into users' psychological states over time. The CLPsych 2025 Shared Task builds upon prior research efforts by integrating longitudinal modeling with evidence generation The CLPsych 2025 Shared Task builds upon prior research efforts (Tsakalidis et al., 2022; Tseriotou et al., 2025; Zirikly et al., 2019) by integrating longitudinal modeling with evidence generation..., focusing on adaptive and maladaptive self-states in user timelines. Our work in this shared task contributes to the growing field of computational mental health assessment by employing **state-of-the-art transformer-based models** across various subtasks. A key preprocessing step in our approach was **removing posts with null values**, ensuring cleaner and more informative datasets for analysis. Unlike previous studies that retained all posts for completeness, our decision aimed to enhance the **signal-to-noise ratio**, thereby improving model performance. By applying a combination of specialized NLP models, we effectively extracted relevant psychological markers, assigned well-being scores, and summarized self-state dynamics in both individual posts and entire timelines.

For well-being score prediction, we fine-tuned MentalBERT, a model specifically trained for mental health applications, to enhance accuracy in assessing psychological functioning. For post-level summarization, we leveraged Longformer, which excels in capturing contextual dependencies in lengthy texts, ensuring comprehensive self-state summaries. At the timeline level, we utilized Pegasus-X-Large, a model optimized for abstractive summarization, to generate coherent narratives capturing self-state transitions. Lastly, for evidence extraction, we employed Mistral, a robust transformer model capable of identifying relevant spans with high precision. Our results indicate that MentalBERT achieved state-of-the-art performance in well-being score prediction, Longformer provided detailed and context-aware post summaries, while Pegasus-X effectively distilled timeline-level insights, and Mistral demonstrated high recall in extracting meaningful evidence spans. These findings reinforce the potential of advanced NLP techniques in modeling dynamic mental health patterns and offer promising directions for future clinical and computational research.

2 Background

Toxicity detection in NLP has focused on span-level identification of harmful content. SemEval-

Proceedings of the 10th Workshop on Computational Linguistics and Clinical Psychology (CLPsych 2025), pages 292–299
May 3, 2025 ©2025 Association for Computational Linguistics

2021 Task 5 (Ji et al., 2021a) highlighted challenges in detecting toxic spans using token classification and span prediction. Transformer-based models like BERT, RoBERTa, and SpanBERT improved performance by combining these approaches. In mental health, domain-specific models like MentalBERT (Chhablani et al., 2021) have enhanced social media text analysis for early detection of mental disorders and suicidal ideation, demonstrating the value of contextualized representations in both toxicity detection and mental health assessment (Shing et al., 2018; Zirikly et al., 2019).

Timeline summarization distills event evolution from timestamped documents, requiring coherence and diversity. Evolutionary Timeline Summarization (ETS) (Yan et al., 2011) optimizes relevance and coverage, while graph-based methods enhance abstractive and extractive summarization. Handling long-context dependencies remains a challenge for transformer models (Qin, 2024), with research on dynamic-resolution encoding (e.g., Nugget) improving efficiency.

Zero-shot prompting in LLMs faces challenges in generating concise, coherent summaries. Chain-of-event (CoE) prompting (Wei et al., 2022) structures summarization into four steps, improving abstraction and coherence. Models like Pegasus (Zhang et al., 2020) further enhance contextual understanding. We also reference the Self-Other Dynamics (SOD) framework (Slonim, 2024) for analyzing adaptive and maladaptive self-states.

3 Methodology

Our methodology focuses on exploiting pre-trained language models and fine-tuning them effectively with the provided data. We employ various language models like BERT, Longformer, and Pegasus across multiple tasks.

3.1 Task A.1: Evidence Extraction for Adaptive and Maladaptive Self-States

Task A.1 involves extracting text spans indicating adaptive or maladaptive self-states. Adaptive self-states support fundamental needs fulfillment, while maladaptive self-states obstruct these needs(Slonim, 2024). The objective is identifying spans evidencing adaptive, maladaptive or neither self-states.

Initially, we considered BERT for token classification (Devlin et al., 2019), labeling each token as part of an adaptive, maladaptive, or neutral span. However, this risked producing fragmented spans with misclassified portions, compromising contextual integrity.

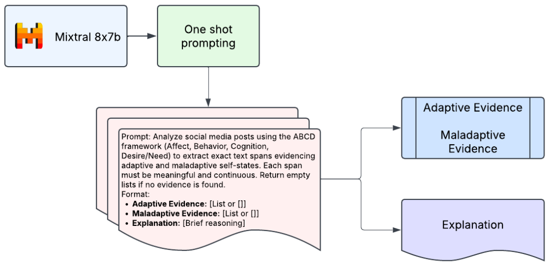

Figure 1: (Task 1a) Methodology diagram showing the prompt tuning of Mistral 8x7b for the purpose of accurate extraction of adaptive and maladaptive spans with addition of explanation component.

Consequently, we adopted Mixtral 8x7B (Jiang et al., 2024), a large language model renowned for robust natural language understanding. This MoE model was selected for its proficiency in interpreting complex natural language and extracting coherent spans, overcoming BERT's fragmentation issues.

The model was guided by a carefully designed prompt leveraging one-shot learning to capitalize on Mistral 8x7B's pre-trained knowledge. The prompt structure can be found in A which was finalized after a careful effort in experimentation based on general research of what adaptive and maladaptive behaviors are.

To enhance performance, we integrated the "explain then annotate" strategy (Lee et al., 2020). This technique requires providing a rationale before finalizing annotations, improving accuracy through deeper contextual understanding. In our implementation, Mixtral 8x7B generated brief explanations for each classification, refining the span extraction process, as illustrated in Figure 1.

3.2 Task 1b: Well-Being Score Prediction

Task 1b involves predicting a well-being score for each post as a classification task. We leveraged embeddings from transformer-based encoder models as input features for a Random Forest Classifier (Breiman, 2001).

For contextual embeddings extraction, we employed various models including BERT (Devlin et al., 2019), MiniLM (Wang et al., 2020), RoBERTa (Liu et al., 2019), ClinicalBERT (Huang et al., 2020), and MentalBERT (Ji et al., 2021b).

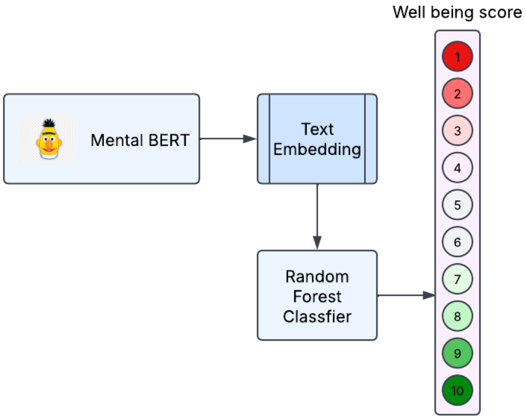

Figure 2: (Task 1b) Methodology of well-being score prediction using Mental BERT and Random Forest Classifier

The [CLS] token embedding from each model's final layer served as post representation, feeding into a Random Forest Classifier for multi-class classification.

Performance evaluation using accuracy as the primary metric revealed MentalBERT as the best performer. This superior performance likely stems from MentalBERT's pre-training on mental health-related text, enabling better capture of domain-specific patterns relevant to well-being assessment, as shown in Figure 2.

3.3 Task 2: Post-Level Summaries of Self-State Dynamics

Task 2 involves generating summaries describing the interplay between adaptive and maladaptive self-states, focusing on the dominant state and its driving ABCD aspect. The summary must outline how this central aspect influences other components and address any complementary self-state.

Initially, we utilized BART (Lewis et al., 2019) for summarization. However, BART's 1024 token context limitation proved insufficient for longer posts. We switched to Longformer (Beltagy et al., 2020) with its 4096 token capacity, allowing effective processing of extended posts with their evidence spans.

We leveraged outputs from Task 1a—specifically, the extracted evidence spans—as additional input to inform summarization. The data was structured as:

```
post: {post}
adaptive_evi: [list of spans]
maladaptive_evi: [list of spans]
```

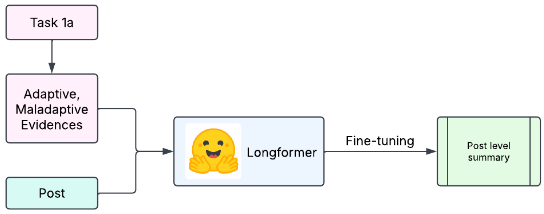

Figure 3: (Task 2) Methodology diagram depicting the finetuning process of LongFormer for the task of post-level summary generation.

Figure 4: (Task 3) Methodology diagram for the task 3 showing the finetuning of pegasus-X-large for timeline summary generation.

This structured input enabled Longformer to contextualize posts with identified self-state evidence, generating summaries accurately reflecting self-state interplay, as depicted in Figure 3.

3.4 Task 3: Timeline-Level Summaries

Task 3 required generating timeline-level summaries reflecting an individual's self-state dynamics over time. We combined multiple posts from the same timeline into cohesive text, preserving chronological order to maintain timeline integrity for effective self-state progression analysis.

Analysis revealed a mean word count of 1224 words, with the longest text reaching 5555 words. This necessitated a model capable of processing long sequences. We selected Pegasus-X-large (Zhang et al., 2020), which accommodates up to 16,000 tokens, eliminating truncation concerns.

Pegasus-X-large was fine-tuned using combined posts as input and expected timeline summaries as targets, enabling the generation of summaries effectively capturing temporal dynamics and transitions, as shown in Figure 4.

4 Experimental Results

In this section, we present results for each task in the CLPsych 2025 shared task. Experiments were conducted using Kaggle's free 2xT4 GPUs, providing sufficient computational resources.

The initial dataset contained 343 rows. After removing null values from well-being scores and post summaries, 199 rows remained for Tasks 1b and 3. For Tasks 1a and 2, we utilized the full dataset as these tasks were unaffected by the missing data.

4.1 Task 1a: Evidence Extraction

For Task 1a, we employed Mixtral-8x7b (Jiang et al., 2024), a mixture of experts model quantized to 4-bit to optimize memory usage. A temperature of 0.2 was selected after experimentation, balancing creativity and precision for effective extraction of adaptive and maladaptive self-state indicators.

4.2 Task 1b: Well-Being Score Prediction

In Task 1b, we tested various models including BERT (Devlin et al., 2019) and LLaMA (Touvron et al., 2023) for sequence classification. Optimal performance was achieved by combining MentalBERT (Ji et al., 2021b) embeddings with a Random Forest Classifier (Breiman, 2001) (100 estimators). Embeddings were extracted using the Sentence Transformers library, capturing domain-specific nuances relevant to well-being assessment.

4.3 Task 2: Post-Level Summaries

For Task 2, we utilized Longformer (Beltagy et al., 2020), chosen for its long-sequence handling capability. Input data combined post content with Task 1a outputs (adaptive/maladaptive evidence spans). The model was fine-tuned using the Seq2SeqTrainer from Transformers (Wolf et al., 2019) for 10 epochs, generating coherent summaries reflecting self-state interplay.

4.4 Task 3: Timeline-Level Summaries

Task 3 involved timeline-level summaries by concatenating posts from the same timeline. Posts were separated chronologically using a delimiter (`\n\n\—-\n\n`). Pegasus-X-large (Zhang et al., 2020), with its 16,000-token context window, was fine-tuned for 10 epochs using Seq2SeqTrainer.

The performance across all tasks is summarized in Table 1, presenting our team's (ResBin) scores.

5 Limitations

Our approach faces several constraints despite promising results. Using traditional language models rather than prompt-tuned LLMs may have limited our performance on tasks requiring nuanced psychological inferences. Effective pre-processing of the data was missing and no sort of data augmentation was carried out. Because of limited data the pre-trained transformer based models could not generalize to the extent LLMs are capable of generalizing. Furthermore fine-tuning of LLMs was not carried out in our work which could have possibly given even better results with extreme level of natural language understanding. Additionally, our models may inadequately capture long-term temporal dependencies in user posts, potentially missing subtle shifts in mental states.

Table 1: Experimental Results for Team ResBin Across Different Tasks

Task	Metric	Score	Rank
Task 1a	Recall	0.470	8
Task 1b	Mean Consistency	0.764	13
Task 2	MSE	8.020	8
Task 3	Mean Consistency	0.898	7

6 Ethics

Our research adheres to strict ethical guidelines protecting data privacy and dignity. We store data exclusively on local machines and google drives with team-restricted access and will delete all dataset files and derived models after the CLPsych 2025 workshop. We commit to not redistributing the dataset and to not submitting any part of it to platforms that might use it as training data. We acknowledge that models developed in this research are for computational research purposes only, not for direct clinical application.

7 Conclusions

In this work, we focused on fine-tuning natural language models like BERT, Longformer, and Pegasus (Tseriotou et al., 2025) for the CLPsych 2025 shared task, building on prior methodologies (Tsakalidis et al., 2022; Zirikly et al., 2019). Our approach effectively addressed evidence span extraction, well-being score prediction, and summary generation at both post and timeline levels. While we prioritized task-specific models over large language models (LLMs) for interpretability, future work could explore supervised LLM fine-tuning with sufficient computational resources to enhance prediction and summarization capabilities. This direction may further bridge the gap between general-purpose language models and domain-specific mental health analysis tasks.

References

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *Preprint*, arXiv:2004.05150.

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

Gunjan Chhablani, Abheesht Sharma, Harshit Pandey, Yash Bhartia, and Shan Suthaharan. 2021. Nlrg at semeval-2021 task 5: Toxic spans detection leveraging bert-based token classification and span prediction techniques. *arXiv preprint arXiv:2102.12254*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *Preprint*, arXiv:1810.04805.

Kexin Huang, Jaan Altosaar, and Rajesh Ranganath. 2020. Clinicalbert: Modeling clinical notes and predicting hospital readmission. *Preprint*, arXiv:1904.05342.

Shaoxiong Ji, Tianlin Zhang, Luna Ansari, Jie Fu, Prayag Tiwari, and Erik Cambria. 2021a. Mentalbert: Publicly available pretrained language models for mental healthcare. *arXiv preprint arXiv:2110.15621*.

Shaoxiong Ji, Tianlin Zhang, Luna Ansari, Jie Fu, Prayag Tiwari, and Erik Cambria. 2021b. Mentalbert: Publicly available pretrained language models for mental healthcare. *Preprint*, arXiv:2110.15621.

Albert Q Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, and 1 others. 2024. Mixtral of experts. *arXiv preprint arXiv:2401.04088*.

Dong-Ho Lee, Rahul Khanna, Bill Yuchen Lin, Jamin Chen, Seyeon Lee, Qinyuan Ye, Elizabeth Boschee, Leonardo Neves, and Xiang Ren. 2020. Lean-life: A label-efficient annotation framework towards learning from explanation. *arXiv preprint arXiv:2004.07499*.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *Preprint*, arXiv:1910.13461.

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. *Preprint*, arXiv:1907.11692.

Guanghui Qin. 2024. *Towards Efficient Long-Context Natural Language Processing*. Ph.D. thesis, Johns Hopkins University.

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.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, and 1 others. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Adam Tsakalidis, Jenny Chim, Iman Munire Bilal, Ayah Zirikly, Dana Atzil-Slonim, Federico Nanni, Philip Resnik, Manas Gaur, Kaushik Roy, Becky Inkster, and 1 others. 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.

Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. 2020. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *Preprint*, arXiv:2002.10957.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, and 1 others. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and 1 others. 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*.

Rui Yan, Xiaojun Wan, Jahna Otterbacher, Liang Kong, Xiaoming Li, and Yan Zhang. 2011. Evolutionary timeline summarization: a balanced optimization framework via iterative substitution. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 745–754.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. 2020. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. *Preprint*, arXiv:1912.08777.

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

A.1 Prompt used in Task 1a

You are a mental health expert analyzing social media posts from the given posts to identify evidence of adaptive and maladaptive self-states. Each self-state is characterized by **Affect (A)**, **Behavior (B)**, **Cognition (C)**, and **Desire/Need (D)** (ABCD framework). Your task is to extract continuous spans of text from the post that directly evidence these self-states.

Definitions

Adaptive Self-State

Aspects of Affect, Behavior, Cognition, and Desire/Need that support the fulfillment of basic needs like relatedness (connection with others), autonomy (independence), and competence (feeling capable). These are healthy coping mechanisms or positive mental states. Examples include:

- **Affect (A)**: Positive emotions like happiness, hope, or pride (e.g., "I'm feeling really hopeful about my future").
- **Behavior (B)**: Healthy actions towards self or others, such as:
  - Support: Seeking or appreciating help from others (e.g., "I talked to my friend about my stress").
  - Physical Activity: Engaging in exercise like walking or yoga (e.g., "I went for a run to clear my mind").
  - Relaxation: Calming activities like meditation or listening to music (e.g., "I listened to soft music to relax").
  - Problem-Solving: Actively addressing issues (e.g., "I made a plan to tackle my workload").
- **Cognition (C)**: Positive perceptions of self or others, like self-acceptance or viewing others as supportive (e.g., "I believe I can improve with help").
- **Desire/Need (D)**: Healthy intentions or needs, like seeking relatedness or autonomy (e.g., "I want to connect with others to feel supported").

Maladaptive Self-State

Aspects of Affect, Behavior, Cognition, and Desire/Need that hinder the fulfillment of basic needs, reflecting unhealthy coping mechanisms or negative mental states. Examples include:

- **Affect (A)**: Negative emotions like depression, anxiety, shame, or hopelessness (e.g., "I feel so hopeless and sad").
- **Behavior (B)**: Unhealthy actions towards self or others, such as:
  - Escape: Withdrawing socially or over-engaging in solitary activities (e.g., "I stayed in my room all day scrolling online").
  - Unhealthy Self-Soothing: Overeating, binge drinking, or excessive internet use (e.g., "I binged on snacks to feel better").

- Numbing: Using substances to numb emotions (e.g., "I drank to forget my problems").
- Self-Harm: Engaging in self-injury (e.g., "I want to hurt myself").
- Compulsions/Risk-Taking: Seeking adrenaline through risky behaviors (e.g., "I drove recklessly to feel something").

- **Cognition (C)**: Negative perceptions, like self-criticism or expecting rejection (e.g., "I think I'm a failure and no one cares").
- **Desire/Need (D)**: Unhealthy expectations or fears, like expecting failure or rejection (e.g., "I feel like I'll never be good enough").

Task A.1 - States' Evidence

- Identify continuous spans of text in the post that directly provide evidence of adaptive self-states and maladaptive self-states.
- Each span should be a complete, meaningful segment of text (e.g., a full sentence or a phrase) that clearly reflects an adaptive or maladaptive self-state.
- A post may contain evidence for one self-state (adaptive or maladaptive), both, or neither.
- If no evidence is found for a self-state, return an empty list ([]) for that category.
- Ensure the spans are exact substrings of the post, preserving the original wording and punctuation.

Post to Analyze

{post}

Output Format

Provide the following in a structured format:

- **Adaptive Evidence**: [List of continuous text spans (strings) showing adaptive self-states, or [] if none]
- **Maladaptive Evidence**: [List of continuous text spans (strings) showing maladaptive self-states, or [] if none]
- **Explanation**: [Brief explanation of your evidence selection, referencing the ABCD framework and coping mechanisms]

Example

Post: "My friend went for a gym session to relieve stress, but he sometimes gets dissapointed and feels hopeless about his situation."
Output:

- **Adaptive Evidence**: ["My friend went for a gym session to relieve stress"]
- **Maladaptive Evidence**: ["he sometimes gets dissapointed and feels hopeless about his situation"]
- **Explanation**: The span "My friend went for a gym session to relieve stress" reflects an adaptive self-state through Behavior (B) - Physical Activity, as walking is a healthy coping mechanism to relieve stress. The span "he sometimes gets dissapointed and feels hopeless about his situation" indicates a maladaptive self-state through Affect (A) - hopelessness, a negative emotion that hinders well-being.

Disclaimer

Paraphrased representative data is being used in the prompt and not real data from training data provided during the shared task.