Extraction and Summarization of Suicidal Ideation Evidence in Social Media Content Using Large Language Models

Loitongbam Gyanendro Singh*, Junyu Mao*, Rudra Mutalik, Stuart E. Middleton
School of Electronics and Computer Science,
University of Southampton, Southampton, UK
{gyanendro.loitongbam, j.mao, rudra.mutalik, sem03}@soton.ac.uk

Abstract

This paper explores the use of Large Language Models (LLMs) in analyzing social media content for mental health monitoring, specifically focusing on detecting and summarizing evidence of suicidal ideation. We utilized LLMs *Mixtral7bx8* and *Tulu-2-DPO-70B*, applying diverse prompting strategies for effective content extraction and summarization. Our methodology included detailed analysis through Few-shot and Zero-shot learning, evaluating the ability of *Chain-of-Thought* and *Direct* prompting strategies. The study achieved notable success in the CLPsych 2024 shared task (ranked top for the evidence extraction task and second for the summarization task), demonstrating the potential of LLMs in mental health interventions and setting a precedent for future research in digital mental health monitoring.

1 Introduction

Large Language Models (LLMs) such as GPT (Generative Pre-trained Transformer) (Brown et al., 2020) have become cornerstones in the field of natural language processing domain. Their ability to process and generate human-like text, learned from extensive datasets, empowers them to recognize and interpret complex language patterns on various reasoning tasks, such as arithmetic, commonsense, and symbolic reasoning (Kojima et al., 2022; Wei et al., 2022). One of the critical abilities of LLMs is text span extraction from unstructured data, such as social media posts (e.g. Reddit, Twitter, Facebook, etc.) (Srivastava et al., 2023; Xu et al., 2023; Yang et al., 2023). This process involves identifying and extracting specific segments of text that contain relevant information or unique characteristics. In the mental health context, this capability becomes indispensable for spotting signs of mental illness, such as depression, anxiety, and particularly suicidal thoughts or tendencies in online conversations.

Given the increasing prevalence of mental health issues and the growing tendency of individuals to express their thoughts and emotions on social media platforms, accurately recognizing signs of suicidal ideation and other mental health concerns from individual posts shared online becomes imperative (Singh et al., 2024; Azim et al., 2022). Since LLMs can process and interpret such complex language patterns, they are essential for identifying early signs of mental health concerns, including suicidal ideation, and thus play a crucial role in mental health interventions (Xu et al., 2023; Yang et al., 2023).

Addressing the need for effective mental health monitoring on social media, this study attempts to make the best use of LLMs to scrutinize user-generated content focusing specifically on identifying and summarizing potential indicators of suicidal ideation. We employ two LLMs, *Mixtral7bx8* (Jiang et al., 2024) and *Tulu-2-DPO-70B* (Ivison et al., 2023), utilizing diverse prompting strategies to extract and summarize the text that signifies suicidal thoughts, thereby gaining insights into users' mental states. Our team (UoS_NLP) participated in the CLPsych 2024 shared task (Chim et al., 2024), where we excelled, securing first place in the evidence extraction task with an F1 score of 0.929 and second in the summarization task with a mean consistency score of 0.977. Our methodology encompassed in-depth post-by-post analysis, incorporating both Few-shot (Wei et al., 2022; Zhang et al., 2022) and Zero-shot (Kojima et al., 2022; Wei et al., 2021) learning techniques, and further evaluate the ability of *Chain-of-Thought* and *Direct* prompting strategies (Kojima et al., 2022; Zhang et al., 2022). For the summarization aspect, we adopted a Zero-shot approach, exploring the impact of including meta-information such as sentiments and user suicide risk labels in the prompts and showcasing the potent application of LLMs in mental health analysis and intervention.

*Equal contributions.

Instruction prompt	As a mental health assistant, your task is to [TASK] directly from the provided input text to highlight the mental health issues. For the [TASK] task, consider the following aspects: • **Emotions**: Evaluate expressed emotions, from sadness to intense psychological pain, as they may influence the assigned risk level. • **Cognitions**: Explore the individual's thoughts and perceptions about suicide, including the level and frequency of suicidal thoughts, intentions of suicide, and any existing plans. • **Behavior and Motivation**: Evaluate the user's actions related to suicide, such as access to means and concrete plans. Consider their ability to handle difficult/stressful situations and the motivations behind their desire to die. • **Interpersonal and Social Support**: Investigate the individual's social support or stable relationships, and understand their feelings toward significant others. • **Mental Health-Related Issues**: Consider psychiatric diagnoses associated with suicide such as schizophrenia, bipolar, anxiety, eating disorder, previous suicidal attempts, and others. • **Additional Risk Factors**: Consider other factors like socioeconomic and demographic factors, exposure to suicide behavior by others, chronic medical conditions, etc.
Meta-information	The opinion holder has an indication of [Risk] suicidal risk, with probable [Emotion] emotion and [Sentiment] sentiment.
Input	[USER POST]
Output	[TASK OUTPUT]

Table 1: Instruction prompt for Mental Health Analysis Task Using Large Language Models (LLMs). The [TASK] placeholder is adapted based on whether the focus is on evidence extraction or summarization. The [TASK OUTPUT] is considered for the *Few-shot* prompting strategy.

The rest of the paper is organized as follows: Section 2 provides an overview of the shared task. Section 3 discusses the experiment designs. Section 4 presents the results and discussion, and finally, the study concludes in Section 5.

2 CLPsych 2024 Shared Task and Dataset

The CLPsych 2024 Shared Task (Chim et al., 2024) centers on employing Large Language Models (LLMs) to identify supporting evidence of an individual's suicide risk level from their social media posts. The challenge requires using an LLM to extract and coherently present evidence from posts that align with the pre-assigned risk levels of low, moderate, or severe suicide risk(Zirikly et al., 2019). This task aims to utilize the generative capabilities of LLMs in producing supportive evidence for clinical assessments.

For this shared task, we were provided UMD Suicidality Dataset (Shing et al., 2018; Zirikly et al., 2019). This dataset includes social media posts from users on the Reddit platform, specifically from the r/SuicideWatch subreddit. These posts have been annotated with suicide risk levels by experts and crowdsource workers, categorizing them into no, low, moderate, or severe risk. Participants are tasked with using LLMs to identify and extract text spans from these posts that support the assigned risk levels. This dataset provides a unique opportunity to explore the application of LLMs in mental health analysis, particularly in assessing and understanding suicide risk from online interactions.

3 Prompting Strategies

In this section, we explore various prompting strategies for text span extraction and summarization in the realm of mental health analysis, utilizing Large Language Models (LLMs). Focusing specifically on two LLMs, *Mixtral7bx8* (Jiang et al., 2024) and *Tulu-2-DPO-70B* (Ivison et al., 2023), this part of the study examines how diverse prompting techniques can enhance the extraction and summarization of relevant information from large datasets in the context of mental health.

3.1 Zero-shot Prompting

This approach utilizes the inherent knowledge of the LLMs without relying on task-specific training. We assess its effectiveness by providing the LLMs with carefully crafted instruction prompts that include the context of the task. Table 1 presents the instruction prompt used for this study. The aim is to guide the LLMs to concentrate on six crucial aspects when identifying text spans related to suicide risk: Emotions, Cognitions, Behavior and Motivation, Interpersonal and Social Support, Mental Health-Related Issues, and Additional Risk Factors. The instruction is designed to guide the LLMs in utilizing their pre-trained knowledge to identify key text spans that may indicate mental health issues. By giving precise and contextualized prompts, we aim to measure the inherent capabilities of the LLMs in extracting meaningful information without additional training or examples.

3.2 Few-Shot Prompting

To enhance the understanding of Large Language Models (LLMs) beyond Zero-shot prompting, this approach incorporates context examples (referred to as demonstrations), enabling the use of few-shot prompting for *In-Context Learning* (ICL). We integrate k-number of input-output pairs with the instruction prompts in Table 1 for effective ICL. Our methodology involves selecting posts that display a range of sentiments, emotions, and levels of user-suicidal risk for annotation, ensuring a comprehensive coverage of contexts for ICL. This selection aids the LLMs in gaining a deeper grasp of the task.

For the preparation of ICL, we utilize Sentence-BERT (Thakur et al., 2021) and pre-trained RoBERTa-base models (Barbieri et al., 2020) to represent user posts in a vector space, incorporating semantic, emotional, and sentiment dimensions. These post representations are then categorized into eight clusters via K-means clustering[1]. In each cluster, the top three posts nearest to the centroid are identified for further analysis. These posts are manually reviewed to determine the user's suicidal risk levels, and three are manually selected for the annotation process in ICL. Our study considers two prompting strategies for ICL: *Direct Prompting* and *Chain-of-Thoughts Prompting*, to evaluate their effectiveness in this context.

3.2.1 Direct Prompting

In *Direct Prompting* strategy, the focus is on presenting clear, explicit instructions or queries that directly correspond with the text span extraction task. This method hinges on the clarity of the prompt to effectively guide the model's response. Additionally, we incorporate few-shot demonstrations within these prompts. These examples are intended to provide more context, thereby enhancing the LLM's ability to discern and extract the relevant text spans accurately. An example of a direct prompting instruction template with input and output is shown in Appendix Table 4.

[1] https://en.wikipedia.org/wiki/K-means_clustering

3.2.2 Chain-of-Thought Prompting

In the *Chain-of-Thoughts Prompting* (CoT) strategy, we direct LLMs through a step-by-step logical reasoning process, thereby enhancing their ability to handle complex tasks. This method involves designing prompts that not only present a problem but also guide the model in a structured thought process towards a solution, exemplified by the phrase *"Let's think step by step"*. The goal is to encourage the LLMs to effectively identify text spans that are relevant to critical aspects such as Emotions, Cognitions, Behavior and Motivation, Interpersonal and Social Support, Mental Health-Related Issues, and Additional Risk Factors. It is important to note that some text spans might be relevant to multiple aspects. For example, the phrase *"I just want to die"* could be indicative of both an emotional state and a cognitive condition. In these instances, we carefully avoid redundant text spans in our final extraction task to maintain a clear and focused representation of each aspect. An example of a CoT template with one demonstration example is shown in Appendix Table 5.

By implementing these prompting strategies, we aim to harness the full potential of LLMs in the domain of text span extraction, addressing both simple and complex extraction requirements with high precision and contextual relevance.

4 Result analysis

In this section, we conduct a comprehensive analysis comparing Zero-shot and few-shot prompting strategies, assessing their effectiveness in enabling in-context learning for text span extraction tasks. Additionally, we examine the performance of direct prompting and chain-of-thought prompting strategies, highlighting their impact on precision and coherence in text span extraction. We also investigate how the inclusion of meta-information such as sentiment, emotion, and suicide risk levels influences the LLMs' ability to identify and summarize mental health-related content. This evaluation provides insights into the strengths and weaknesses of these strategies in the evidence extraction and summarization tasks. Table 2 and 3 showcase the performance of our LLMs in the evidence extraction and summarization tasks, along with the top-ranked competitors from the shared task.

LLMs	Prompt Strategies	Evidences extraction task			
		Recall	Precision	W-Recall	F1-score
Mixtral7bx8	0-shot	0.914	0.911	0.675	0.912
Mixtral7bx8	Direct	0.914	0.907	0.651	0.910
Tulu-2-DPO-70B	CoT	**0.943**	0.916	0.527	**0.929**
BERT-finetuned	–	**0.944**	0.906	0.489	0.924

Table 2: Performance Comparison of Large Language Models (LLMs) in Evidence Extraction Task using various prompting strategies. *sophiaADS* is noted as the top competitor against our models in the evidence extraction task.

LLMs	Include Meta information?	Evidences summarization	
		Mean Consistency	Max Contradict
Mixtral7bx8	–	0.951	0.127
Mixtral7bx8	✓	**0.977**	**0.079**
Tulu-2-DPO-70B	✓	0.966	0.107
mistral-7b-instruct-v0.2	–	**0.979**	**0.064**

Table 3: Performance Comparison of Large Language Models (LLMs) in Evidence Summarization Tasks Using Zero-Shot Prompting, with and without Meta-Information. ✓indicates the inclusion of meta-information like sentiments and suicide risk labels in the prompts. *UZH_CLyp* is noted as the top competitor against our model in the summarization task.

4.1 Evidence extraction task

In the evidence extraction task, the *Mixtral7bx8* model shows no significant difference between its Few-Shot (Direct) and Zero-shot prompting strategies, with both achieving a recall and F1-score of 0.914 and 0.912, respectively shown in Table 2. This suggests that the model's performance in identifying relevant textual evidence does not depend on the additional context provided by a Few-Shot approach. On the other hand, the *Tulu-2-DPO-70B* model benefits from a Few-Shot (CoT) strategy, leading to the highest recall of 0.943 and F1-score of 0.929 among the listed approaches, indicating that the Chain-of-Thought prompting substantially enhances its evidence extraction capabilities. This is expected as *Tulu-2-DPO-70B* is an instruction-tuned model that incorporates chain-of-thought data within its training mixture dataset, endowing it with robust reasoning abilities. While the BERT-finetuning model by *sophiaADS* (Tanaka and Fukazawa, 2024), a competitor in this evidence extraction task, marginally surpasses the *Tulu-2-DPO-70B* in recall. However, the *Tulu-2-DPO-70B* model achieves a higher F1 score. This superior F1 score indicates a more optimal balance between precision and recall, underscoring the *Tulu-2-DPO-70B* model's enhanced effectiveness in evidence extraction when compared to its competitor.

4.2 Evidence summarization task

For the evidence summarization task, we utilized a Zero-shot prompting strategy with a focus on incorporating specific types of meta-information: emotion, sentiment, and user risk label, derived from multiple posts of a user, as outlined in Appendix Table 6. This process involved aggregating all posts from a single user, analyzing the emotion and sentiment for each post, and then selecting the most prevalent emotion and sentiment to represent the overall state of the user. The user risk label was determined based on an assessment of all the user's posts. Table 3 presents the effectiveness of LLMs using Zero-shot strategies both with and without this meta-information. Notably, the *Mixtral7bx8* model incorporating meta-information attains a high mean consistency score of 0.977, indicating its strong capability in generating accurate and coherent summaries aligned with the extracted evidence. This underscores the effectiveness of Zero-shot strategies in mental health analysis on social media. Notably, the *mistral-7b-instruct-v0.2* used by *UZH_CLyp* (Uluslu et al., 2024), as our top competitor in the shared task, slightly outperforms the *Mixtral7bx8* model with meta-information in terms of mean consistency, highlighting their marginally superior summarization reliability.

5 Conclusion

This study demonstrates the significant potential of Large Language Models (LLMs) in mental health analysis, particularly in identifying and summarizing suicidal ideation from social media content. Through the CLPsych 2024 Shared Task, we successfully applied advanced LLMs using both Few-shot and Zero-shot prompting strategies, achieving notable performance in evidence extraction and summarization tasks. Our findings emphasize the efficacy of LLMs in handling complex mental health data and highlight the impact of Chain-of-Thought (CoT) prompting on evidence extraction. The study also highlights the importance of incorporating meta-information to enhance the evidence summarization tasks. Our success in the CLPsych 2024 shared task highlights the practical application of LLMs in mental health interventions, paving the way for future advancements in digital mental health monitoring.

6 Limitation

The study has a few key limitations. First, it focuses mainly on two models, *Mixtral7bx8* and *Tulu-2-DPO-70B*, which might not represent how other Large Language Models would perform. Second, it only looks at Few-Shot and Zero-shot prompting, missing out on other possible methods. Third, the datasets used for this study might not be diverse or large enough, which could make the findings less applicable to real-world situations. Additionally, the reliance on manual annotation for Few-Shot prompting introduces subjectivity and potential scalability issues. Lastly, analyzing only two competitor models may not fully capture the competitive landscape, potentially leading to an incomplete comparative assessment. These factors highlight the need for a more inclusive approach in model selection, diverse prompting strategies, comprehensive datasets, and objective competitor analysis in future research.

Ethical Statement

We adhere to the ethical standards of the CLPsych 2024 Shared Task by ensuring a secure environment and user anonymity. All parties involved have signed agreements to protect data privacy and prohibiting data sharing with any unauthorized entities. Our approach is designed to prioritize responsible use, with strict access controls and guidelines to prevent misuse.

Acknowledgements

This work was supported by the Natural Environment Research Council (NE/S015604/1), Economic and Social Research Council (ES/V011278/1), and Engineering and Physical Sciences Research Council (EP/S024298/1), and the Alan Turing Institute's Turing Enrichment Scheme. The authors acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, which significantly contributed to the completion of this work. The authors are particularly grateful to the anonymous users of Reddit whose data feature in this year's shared task dataset, to the clinical experts from Bar-Ilan University who annotated the data, and the assistance of the American Association of Suicidology for making the dataset available.

References

Tayyaba Azim, Loitongbam Singh, and Stuart Middleton. 2022. Detecting moments of change and suicidal risks in longitudinal user texts using multi-task learning. In *Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology*, pages 213–218.

Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa Anke, and Leonardo Neves. 2020. TweetEval: Unified benchmark and comparative evaluation for tweet classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1644–1650, Online. Association for Computational Linguistics.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

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 Ninth Workshop on Computational Linguistics and Clinical Psychology*. Association for Computational Linguistics.

Hamish Ivison, Yizhong Wang, Valentina Pyatkin, Nathan Lambert, Matthew Peters, Pradeep Dasigi, Joel Jang, David Wadden, Noah A. Smith, Iz Beltagy, and Hannaneh Hajishirzi. 2023. Camels in a changing climate: Enhancing lm adaptation with tulu 2.

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, et al. 2024. Mixtral of experts. *arXiv preprint arXiv:2401.04088*.

Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. *Advances in neural information processing systems*, 35:22199–22213.

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.

Loitongbam Gyanendro Singh, Stuart E Middleton, Tayyaba Azim, Elena Nichele, Pinyi Lyu, and Santiago De Ossorno Garcia. 2024. Conversationmoc: Encoding conversational dynamics using multiplex network for identifying moment of change in mood and mental health classification. In *Proceedings of the Machine Learning for Cognitive and Mental Health Workshop (ML4CMH) @ AAAI 2024*.

Aarohi Srivastava, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam Fisch, Adam R Brown, Adam Santoro, Aditya Gupta, Adrià Garriga-Alonso, et al. 2023. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *Transactions on Machine Learning Research*.

Rika Tanaka and Yusuke Fukazawa. 2024. Integrating supervised extractive and generative language models for suicide risk evidence summarization. In *Proceedings of the Ninth Workshop on Computational Linguistics and Clinical Psychology*. Association for Computational Linguistics.

Nandan Thakur, Nils Reimers, Johannes Daxenberger, and Iryna Gurevych. 2021. Augmented SBERT: Data augmentation method for improving bi-encoders for pairwise sentence scoring tasks. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 296–310, Online. Association for Computational Linguistics.

Ahmet Yavuz Uluslu, Andrianos Michail, and Simon Clematide. 2024. Utilizing large language models to identify evidence of suicidality risk through analysis of emotionally charged posts. In *Proceedings of the Ninth Workshop on Computational Linguistics and Clinical Psychology*. Association for Computational Linguistics.

Jason Wei, Maarten Bosma, Vincent Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. 2021. Finetuned language models are zero-shot learners. In *International Conference on Learning Representations*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837.

Xuhai Xu, Bingshen Yao, Yuanzhe Dong, Hong Yu, James Hendler, Anind K Dey, and Dakuo Wang. 2023. Leveraging large language models for mental health prediction via online text data. *arXiv preprint arXiv:2307.14385*.

Kailai Yang, Tianlin Zhang, Ziyan Kuang, Qianqian Xie, and Sophia Ananiadou. 2023. Mentalllama: Interpretable mental health analysis on social media with large language models. *arXiv preprint arXiv:2309.13567*.

Zhuosheng Zhang, Aston Zhang, Mu Li, and Alex Smola. 2022. Automatic chain of thought prompting in large language models. In *The Eleventh 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

The subsequent Tables 4 and 5 provide detailed examples illustrating the process of inputting instruction prompts into LLM models for the purpose of evidence extraction and summarization tasks related to suicidal thought ideation. These tasks are executed using both direct and Chain-of-Thought (CoT) prompting strategies, incorporating meta-information alongside the given input. Furthermore, Table 6 specifically displays the outputs generated by the LLM for the summarization tasks, highlighting the effectiveness of the applied prompting strategies.

Instruction prompt	As a mental health assistant, your task is to extract evidence directly from the provided input text to highlight the mental health issues. For the evidence extraction task, consider the following aspects: • **Emotions**: Evaluate expressed emotions, from sadness to intense psychological pain, as they may influence the assigned risk level. • **Cognitions**: Explore the individual's thoughts and perceptions about suicide, including the level and frequency of suicidal thoughts, intentions of suicide, and any existing plans. • **Behavior and Motivation**: Evaluate the user's actions related to suicide, such as access to means and concrete plans. Consider their ability to handle difficult/stressful situations and the motivations behind their desire to die. • **Interpersonal and Social Support**: Investigate the individual's social support or stable relationships, and understand their feelings toward significant others. • **Mental Health-Related Issues**: Consider psychiatric diagnoses associated with suicide such as schizophrenia, bipolar, anxiety, eating disorder, previous suicidal attempts, and others. • **Additional Risk Factors**: Consider other factors like socioeconomic and demographic factors, exposure to suicide behavior by others, chronic medical conditions, etc.
Meta-information	The opinion holder indicates Severe suicidal risk, with probable sad emotion and negative sentiment.
Input	Today marks a particularly challenging point in my life. My relationship is going through significant dificulties, and academically, I'm failing to keep up in five subjects. The mornings bring with them a profound struggle to even leave my bed. At 16 years old, a high school student, I'm begining to feel the weight of life's hardships acutely. This period marks two years of battling depression , and after trying eight diferent medications without success, I've reached a state of despair. In my pain, I've contemplated ending my life, beleiving it to be a release from this seemingly endless suffering. I've even gone as far as to plan for it in a way that minimizes the impact on my loved ones, whom I care deeply about. I've prepared a farewell note . This post is a manifestation of my internal conflict; part of me questions if it's truely my time to go. I recognize that suicide is not the solution, yet the immense pain I'm experiencing makes it appear as the only option for temporary relief.
Output	[*relationship is going through significant dificulties*, *failing to keep up in five subjects*, *profound struggle to even leave my bed*, *two years of battling depression*, *I've reached a state of despair.*, *contemplated ending my life*, *endless suffering*, *prepared a farewell note*, *suicide is not the solution*, *the immense pain I'm experiencing*, *the only option for temporary relief*]

Table 4: Illustration of Instruction Prompt for evidence extraction task using Direct Prompting strategy. The output text spans are provided as a demonstration.

As a mental health assistant, your task is to extract relevant spans directly from the provided input text to support the assigned Suicide Risk Level. When highlighting evidence, consider the following aspects:
Emotions: Evaluate expressed emotions, from sadness to intense psychological pain, as they may influence the assigned risk level.
Cognitions: Explore the individual's thoughts and perceptions about suicide, including the level and frequency of suicidal thoughts, intentions of suicide, and any existing plans.
Behavior and Motivation: Evaluate the user's actions related to suicide, such as access to means and concrete plans. Consider their ability to handle difficult/stressful situations and the motivations behind their desire to die.
Interpersonal and Social Support: Investigate the individual's social support or stable relationships, and understand their feelings toward significant others.
Mental Health-Related Issues: Consider psychiatric diagnoses associated with suicide such as schizophrenia, bipolar, anxiety, eating disorder, previous suicidal attempts and others.
Additional Risk Factors: Consider other factors like socioeconomic and demographic factors, exposure to suicide behaviour by others, chronic medical condition, etc.

Input: Today marks a particularly challenging point in my life. My relationship is going through significant dificulties, and academically, I'm failing to keep up in five subjects. The mornings bring with them a profound struggle to even leave my bed. At 16 years old, a high school student, I'm begining to feel the weight of life's hardships acutely. This period marks two years of battling depression , and after trying eight diferent medications without success, I've reached a state of despair. In my pain, I've contemplated ending my life, beleiving it to be a release from this seemingly endless suffering. I've even gone as far as to plan for it in a way that minimizes the impact on my loved ones, whom I care deeply about. I've prepared a farewell note . This post is a manifestation of my internal conflict; part of me questions if it's truely my time to go. I recognize that suicide is not the solution, yet the immense pain I'm experiencing makes it appear as the only option for temporary relief.
The opinion holder indicates Severe suicidal risk, with probable sad emotion and negative sentiment.
Output: Let's think step by step.
Emotions:

* "I've reached a state of despair."
* "endless suffering"
* "the immense pain I'm experiencing"

Cognitions:

* "contemplated ending my life"
* "prepared a farewell note"
* "suicide is not the solution"
* "the only option for temporary relief"

Behaviour and Motivation:

* "relationship is going through significant dificulties"
* "failing to keep up in five subjects"
* "profound struggle to even leave my bed"
* "prepared a farewell note"

Interpersonal and Social Support:

* "relationship is going through significant dificulties"

Mental Health-Related Issues:

* "two years of battling depression"

Additional Risk Factors:

* (Not applicable or lacking information)

...

Table 5: Illustration of Instruction Prompt for evidence extraction task using Chain-of-Thought Prompting with one demonstration.

Instruction prompt	As a mental health assistant, your task is to summarize directly from the provided input text that highlights the mental health issues. For the summarization task, consider the following aspects: • **Emotions**: Evaluate expressed emotions, from sadness to intense psychological pain, as they may influence the assigned risk level. • **Cognitions**: Explore the individual's thoughts and perceptions about suicide, including the level and frequency of suicidal thoughts, intentions of suicide, and any existing plans. • **Behavior and Motivation**: Evaluate the user's actions related to suicide, such as access to means and concrete plans. Consider their ability to handle difficult/stressful situations and the motivations behind their desire to die. • **Interpersonal and Social Support**: Investigate the individual's social support or stable relationships, and understand their feelings toward significant others. • **Mental Health-Related Issues**: Consider psychiatric diagnoses associated with suicide such as schizophrenia, bipolar, anxiety, eating disorder, previous suicidal attempts, and others. • **Additional Risk Factors**: Consider other factors like socioeconomic and demographic factors, exposure to suicide behavior by others, chronic medical conditions, etc.
Meta-information	The opinion holder indicates Severe suicidal risk, with probable sad emotion and negative sentiment.
Input	Today marks a particularly challenging point in my life. My relationship is going through significant dificulties, and academically, I'm failing to keep up in five subjects. The mornings bring with them a profound struggle to even leave my bed. At 16 years old, a high school student, I'm begining to feel the weight of life's hardships acutely. This period marks two years of battling depression , and after trying eight diferent medications without success, I've reached a state of despair. In my pain, I've contemplated ending my life, beleiving it to be a release from this seemingly endless suffering. I've even gone as far as to plan for it in a way that minimizes the impact on my loved ones, whom I care deeply about. I've prepared a farewell note . This post is a manifestation of my internal conflict; part of me questions if it's truely my time to go. I recognize that suicide is not the solution, yet the immense pain I'm experiencing makes it appear as the only option for temporary relief.
Summary output without Meta-information	The individual, a 16-year-old high school student, is grappling with severe depression, which has persisted for two years despite multiple medication trials. Academic struggles in five subjects and relationship difficulties exacerbate their emotional distress. They have considered suicide, planning it to minimize harm to loved ones, and have drafted a farewell note. Despite understanding that suicide is not a solution, the intensity of their pain leads them to consider it as a potential escape..
Summary output with Meta-information	The user is at Severe suicidal risk because they have been struggling with depression for two years, despite trying various medications. They are currently experiencing academic difficulties and relationship issues, which exacerbate their feelings of hopelessness. Their morning struggles suggest possible sleep disturbances or lack of motivation, common in depression. Despite recognizing that suicide is not a solution, they have planned for it, indicating a serious level of distress. Their use of language like 'endless suffering' and 'temporary relief' suggests intense emotional pain.

Table 6: Illustration of Instruction Prompt Variations for the Suicidal Ideation Evidence Summarization Task. This table compares the results of instruction prompts when incorporating meta information against those without it.