MentalBlend: Enhancing Online Mental Health Support through the Integration of LLMs with Psychological Counseling Theories

Ziyin Gu [1,2], Qingmeng Zhu[1,2,*]
[1]University of Chinese Academy of Sciences, Beijing, China
[2]Institute of Software, Chinese Academy of Sciences, Beijing, China
{ziyin2020,qingmeng}@iscas.ac.cn

Abstract

Online mental health support plays a crucial role in addressing the mental health issues faced by modern individuals. However, delivering high-quality online mental health support presents a significant challenge. In response to this challenge, we introduce MentalBlend, a framework that leverages psychological counseling theories, including Cognitive-Behavioral Therapy, Dialectical Behavior Therapy, Person-Centered Therapy, and Reality Therapy, to guide large language models (LLMs) in offering professional online mental health support to individuals seeking assistance. Experimental evidence validates that the collaboration between LLMs and the MentalBlend framework results in the generation of responses that align with professional standards for online mental health support. Overall, our research aims to contribute to advancing the capabilities of LLMs in understanding the emotional backgrounds of help-seekers and delivering professional mental health support effectively.

Keywords: mental health support; psychological counseling theories; large language models

Introduction

Mental health is a pressing public health concern, as many individuals grapple with emotional fluctuations and psychological distress(Good & Sambhanthan, 2014). With the rise of social media and internet technologies, online mental health support has become increasingly popular. Figure 1 illustrates an authentic example of online mental health support from the PsyQA dataset(Sun, Lin, Zheng, Liu, & Huang, 2021). Typically, posts from individuals seeking help include a title and a detailed description of their mental health concerns, and professional supporters provide online mental health support based on these texts. The recent advent of Large Language Models (LLMs) has significantly advanced text generation capabilities, showcasing expressive capacities similar to human (Pendse et al., 2019; Cao et al., 2019; Z. Xu, Pérez-Rosas, & Mihalcea, 2020; Chen & Liu, 2023), indicating their potential in the realm of online mental health support. However, as of now, there is a dearth of high-quality mental health support work based on these large language models. Therefore, this research aims to explore the integration of the latest LLMs and professional psychological counseling theories for application in online mental health support, providing individuals with more comprehensive and effective mental health support.

However, combining the latest Large Language Models (LLMs) in online mental health support poses several significant challenges. (1) Traditional LLMs may struggle to deeply understand users' psychological issues. Despite their remarkable progress in text generation, accurately grasping users' true feelings and needs in complex emotional and psychological contexts remains challenging. (2) Providing personalized support is also a challenge. Each individual facing mental health issues has unique backgrounds, experiences, and needs. Traditional LLMs may lack sufficient personalized elements, making it difficult to offer targeted support and advice for each user. (3) Traditional LLMs may face challenges in providing specialized knowledge support. Mental health issues often require profound expertise, including cognitive-behavioral therapy, dialectical behavior therapy, and others. While LLMs excel in text generation, their lack of in-depth understanding of specialized knowledge in the field of psychology may hinder their ability to provide support that meets professional standards.

Figure 1: An example post from help-seeker.

Therefore, to address the aforementioned challenges, we propose MentalBlend in this study. The primary objective of

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

MentalBlend is to fully leverage the potential of large language models in the realm of mental health support, aiming to generate personalized, professional, and empathetic responses and recommendations. MentalBlend comprises three main components, each addressing one of the aforementioned challenges. Firstly, to stimulate the large model's understanding of implicit information, we introduce prompts to compel the large language model to delve deeper into comprehending mental health issues, thereby enhancing (1). Additionally, some studies have demonstrated that knowledge from emexplars can enhance the understanding capabilities of LLMs (Zhao et al., 2023). Therefore, we consider introducing relevant emexplars to improve (2). In (3), we contemplate incorporating professional psychological counseling theories (Cognitive-Behavioral Therapy(Beck, 1979; Kaczkurkin & Foa, 2015; Hofmann, Sawyer, & Fang, 2010), Dialectical Behavior Therapy(Linehan, 1987), Person-Centered Therapy(Cooper & McLeod, 2011), Reality Therapy(Wubbolding, Casstevens, & Fulkerson, 2017)) to inspire the large model in delivering professional mental health support. MentalBlend first employs a classifier based on BERT model to discern the mental health issues of help-seekers and provides them with appropriate professional psychological counseling theories. Subsequently, the large language model, guided by the designed prompts, delivers personalized, professional, and empathetic responses to help-seekers.

On a real-world dataset of online mental health support, we conducted tests on MentalBlend, and the results indicate not only outstanding performance in automated evaluations but also recognition in human assessments. The introduction of MentalBlend opens up new possibilities for better integration of future online mental health support tasks with artificial intelligence and large language models. This advancement contributes to providing users with more professional, personalized, and emotionally enriched support.

Related Work

NLP for Mental Health Detection and Therapy

In recent years, research related to mental health detection and treatment has flourished, driven by the intersection of cognitive science and technological innovations such as artificial intelligence. Firstly, cognitive science theories and methods have played a crucial role in psychological treatments, with cognitive-behavioral therapy (CBT) and other cognition-based therapeutic approaches becoming predominant(Beck, 1979; Kaczkurkin & Foa, 2015; Hofmann et al., 2010; Linehan, 1987; Cooper & McLeod, 2011; Wubbolding et al., 2017). These methods emphasize the interplay between an individual's thoughts and behaviors, providing a solid theoretical foundation for understanding and intervening in mental health issues. Secondly, advancements in technologies such as artificial intelligence have opened up new possibilities in the field of mental health. Particularly, the application of natural language processing (NLP) techniques has enhanced the accuracy and precision of analyzing individual language, behaviors, and emotions. Early research has predominantly focused on the analysis of text data related to mental health using NLP technology (Ridout & Campbell, 2018; Kruzan et al., 2022; Livingston, Cianfrone, Korf-Uzan, & Coniglio, 2014), including posts and blogs on social media platforms (Park, McDonald, & Cha, 2013; Tsugawa et al., 2015; X. Xu et al., 2021). These studies have successfully explored various mental health issues, such as depression, suicidal ideation, and others (Burnap, Colombo, & Scourfield, 2015; Coppersmith, Leary, Crutchley, & Fine, 2018; Tadesse, Lin, Xu, & Yang, 2019; De Choudhury, Kiciman, Dredze, Coppersmith, & Kumar, 2016). Additionally, researchers have continually innovated and improved tools for supporting online mental health, aiming to provide users with more extensive and convenient mental health services (Martinengo, Lum, & Car, 2022; Yang, Ji, Zhang, Xie, & Ananiadou, 2023). However, it is worth noting that existing AI-based online mental health support efforts have yet to integrate cognitive science fully, limiting user experience and treatment efficacy. Therefore, this paper proposes the MentalBlend framework, which more effectively leverages LLMs to address the lack of cognitive science knowledge in current mental health detection and treatment efforts. MentalBlend not only enhances the understanding of psychological distress but also enables the generation of more professional and empathetic mental health counseling through the provision of specialized conversational prompts.

LLMs and Chain-of-Thought Prompting

In recent years, Large Language Models (LLMs) have garnered widespread attention in the field of natural language processing. Renowned for their massive parameter sizes and outstanding performance across various language tasks, these models have been a subject of extensive research (Zhao et al., 2023). Specifically, prompting has emerged as an innovative approach within the LLMs domain, aiming to guide LLMs in generating more flexible and efficient outputs by providing specific inputs or prompts (Madaan et al., 2023; Shinn, Labash, & Gopinath, 2023; Yao et al., 2022). Driven by prompting techniques, LLMs exhibit potential unique advantages in producing emotionally rich responses and engaging in effective interactions with users. Among these prompting methods, Chain- of-thought (CoT) stands out as a representative work enhancing the reasoning capabilities of LLMs (Wei et al., 2022; Kojima, Gu, Reid, Matsuo, & Iwasawa, 2022). Some improvements based on CoT have also been proposed, such as self-improvement (Madaan et al., 2023), which focuses on enabling LLMs to engage in dialogue with themselves, providing feedback on generated content, and iteratively revising answers. Additionally, ReAct synergizes reasoning and action in LLMs, further enriching the model's knowledge acquisition approach (Yao et al., 2022). However, despite their excellent performance in handling complex problems, these methods have not been extensively explored in the context of mental health applications.

However, the application of Large Language Models (LLMs) in the field of mental health encounters specific challenges, primarily stemming from a lack of domain-specific knowledge and the limitations of LLMs in comprehending intricate mental health issues. In response to these challenges, we present MentalBlend in this study.

Method

In this section, we have introduced the MentalBlend framework, which consists of three main stages. These stages are dedicated to driving Large Language Models (LLMs) to a deeper understanding of subtle emotional nuances and mental health issues, selecting appropriate psychological counseling theories for users, and equipping LLMs with specialized knowledge for addressing specific mental health problems. Finally, we discuss the process of generating online mental health support responses based on these three stages.

Stage One

In this phase, MentalBlend directs Large Language Models (LLMs) to delve into the mental health concerns of help-seekers through the formulation of prompts that facilitate the understanding of emotional backgrounds and the analysis of mental health issues. This stage comprises two pivotal steps. Firstly, MentalBlend stimulates LLMs to comprehend the emotional states of patients, providing a cognitive foundation for further problem analysis. MentalBlend achieves this by guiding LLM to focus on emotional words, analyze contextual information, and consider emotional intensity and variations. Specific prompts include:

> Given the question from a help-seeker [X], what are the words expressing the emotional state of the help-seeker?

> Carefully read descriptions related to the words expressing emotional states; what is the background that triggers these emotions?

> Starting from the intensity and trends of the identified emotions, what is the complete emotional state of the help-seeker?

Following the conclusions drawn from the previous step, MentalBlend continues to guide LLM to concentrate on keywords related to mental health issues:

> Continue analyzing the keywords related to mental health issues mentioned in the question from a help-seeker.

MentalBlend then proceeds with the analysis of mental health issues:

> Based on all the analyses, what is the mental health issue that the help-seeker possesses?

This stage fully leverages the robust reasoning and generative capabilities of LLMs, progressively deepening the LLM's understanding of mental health issues. It establishes a more solid cognitive foundation for subsequent personalized and professional mental health support responses.

Stage Two

At this stage, MentalBlend employs a deep learning model based on BERT (Bidirectional Encoder Representations from Transformers)(Devlin, Chang, Lee, & Toutanova, 2018) with the aim of automatically selecting the most suitable professional psychological counseling theory for each help-seeker. These theories encompass Cognitive-Behavioral Therapy, Dialectical Behavior Therapy, Person-Centered Therapy, and Reality Therapy. BERT, renowns for its widespread application and success in extracting comprehensive language features from text, has achieved notable success in tasks such as sentiment analysis through fine-tuning pre-trained BERT models.

In this phase, MentalBlend selects a pre-trained BERT model and appended a fully connected layer after BERT. This model outputs the appropriate professional psychological counseling theory category and corresponding probability scores for each help-seeker. We conduct fine-tuning training for the proposed model using 1,000 QA pairs from the PsyQA dataset, and two human annotators inspected and labeled the responses of professional supporters in the PsyQA dataset with the corresponding professional psychological counseling theory categories (Cognitive-Behavioral Therapy labeled as 1, Dialectical Behavior Therapy as 2, Person-Centered Therapy as 3, and Reality Therapy as 4). The original data had quantities of 189, 317, 196, and 298 for categories 1, 2, 3, and 4, respectively. The ground truth dataset was divided into training and testing sets in an 8:2 ratio, and the fine-tuned BERT model was trained using the training set. Our model exhibited superior performance on the testing set compared to the original research model, with an accuracy of 86.2% in the professional psychological counseling theory category, surpassing the original model's accuracy of 82.5%. This indicates that the model proposed in this study accurately classifies the appropriate professional psychological counseling theory category for help-seekers. Through the utilization of these fine-tuned models, we ultimately labeled the professional psychological counseling theory category for all data points in the PsyQA dataset.

Stage Three

In this phase, MentalBlend employs a retrieval approach, incorporating relevant exemplar knowledge (Zhao et al., 2023), to address potential gaps in specialized knowledge within LLMs. Prior research has robustly demonstrated that LLMs can acquire task-specific knowledge through contextual examples. We employ SimCSE to retrieve examples from real-world psychological counseling case datasets that are semantically most similar to the mental health issues the help-seeker has.

It is noteworthy that the retrieval query concatenates not only the original questions and descriptions but also incorporates the cognitive foundations established in Stage One for

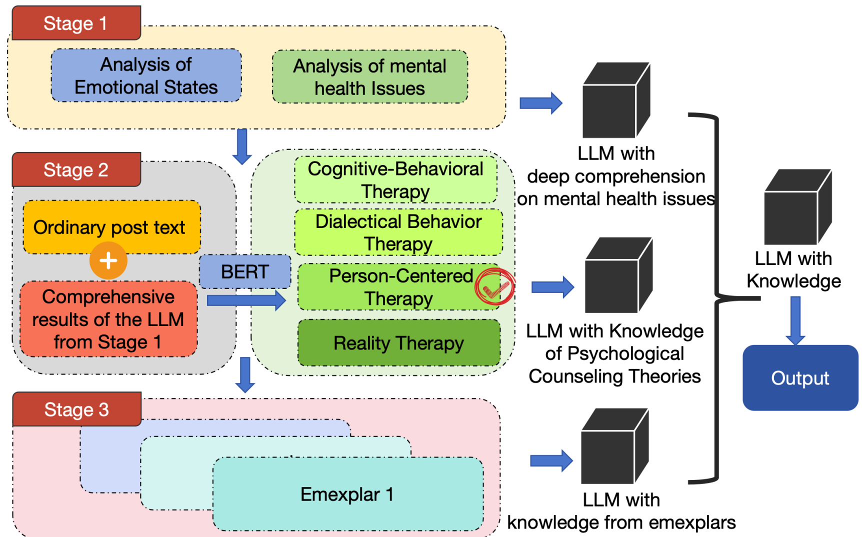

Figure 2: The overview of MentalBlend. MentalBlend comprises three primary stages. In stage 1, MentalBlend introduces two-step prompting to obtain a deep understanding of mental health issues. In stage 2, MentalBlend employs a BERT-based classifier to recommend a suitable psychological counseling theory for help-seekers. In stage 3, MentalBlend retrieves relevant exemplars, providing the large language model with professional problem-solving approaches for the specific type of mental health issue. The large language model then generates personalized and professional responses based on results of these stages.

understanding mental health issues and the cognitive foundations acquired in Stage Two for suitable psychological counseling theories. This augmentation aims to facilitate the retrieval of exemplars more tailored to the characteristics of the help-seeker.

Ultimately, MentalBlend directly provides the retrieved similar exemplars to the LLM in the format:

> Original statement: [q], Similar question: [Q], Solution: [S]···.

Supportive Response Generation

Table 1 presents the types of mental health supportive strategy labeled in the PsyQA dataset. Following the classification process in Stage Two, we conduct a statistical analysis of the mainstream support strategy chains and their respective proportions for the four categories. Subsequently, two annotators with a background in psychology (who also participate in the classification labeling in Stage Two) independently fine-tuned the mainstream strategies. During the fine-tuning process, the two annotators engage in iterative discussions until a consensus was reached. This resulted in the final support strategies for the four professional psychological counseling theories, as shown in Table 2. Early researches state that LLMs tend to follow fixed patterns of thinking and reasoning(Zhao et al., 2023). To provide professional and personalized responses, we pre-define the typical forms of support strategies corresponding to the selected professional psychological counseling theories. This allows LLMs to engage in fixed-pattern reasoning and mental health support responses based on the chosen type of professional psychological counseling theory. The typical supportive strategy chains MentalBlend uses corresponding to different professional psychological counseling theories are listed in Table 2. Due to the robust understanding capabilities of LLMs(Zhao et al., 2023), we can extract the results of the analyses from stages one and three from historical interaction with LLMs. This information serves as the cognitive foundation for generating online mental health support and is incorporated into the process of online mental health support generation.

Simultaneously, MentalBlend combines the classification results from stage two with the strategies outlined in Table 2, explicitly formulating the strategies for LLMs to generate mental health support. In summary, MentalBlend generates online mental health support responses with the following prompt:

> Given [the support strategy for the selected professional psychological counseling theory], follow this support strategy, and based on your analysis of the mental health issues the help-seeker is dealing with, generate an online mental health support response.

Experiments

Dataset

We conduct our research based on the PsyQA dataset (Sun et al., 2021). PsyQA is an authoritative Chinese mental health support dataset. Each data consists of a question, a detailed description, and keyword labels presented by anonymous help-seekers. Answers are asynchronously provided by

Table 1: Supportive strategy from PsyQA dataset.

Strategy	Strategy Definitions
Information	Supply information in the form of data, facts, opinions and resources.
Direct Guidance	Provide suggestions, directives, instructions, or advice about what the help-seeker should do to change.
Approval and Reassurance	Emotional support reassurance,encouragement and reinforcement.
Restatement	A simple repeating or rephrasing of the content or meaning of the question,usually in a more concrete and clear way
Interpretation	Go beyond what the help-seeker has overtly stated or recognized and give a new meaning, reason or explanation.
Self-disclosure	Reveal something personal about the helper's non-immediate experiences or feelings.

Table 2: Supportive strategy chains of different psychological counseling theories.

Type	Core objective	Mainstream support strategy chains	Proportions	Fine-tuned strategies
CBT	Focusing on the client's specific issues and negative thoughts	[Approval and Reassurance][Direct Guidance]	18.26%	[Restatement][Approval and Reassurance][Direct Guidance]
DBT	Emphasizing emotion regulation and coping skills	[Approval and Reassurance][Direct Guidance]	11.32%	[Approval and Reassurance][Self-disclosure][Direct Guidance]
PCT	Giving positive attention and empathy, addressing problems	[Approval and Reassurance][Direct Guidance][Approval and Reassurance]	14.60%	[Approval and Reassurance][Restatement][Approval and Reassurance]
RT	Providing practical and feasible solutions	[Approval and Reassurance][Restatement][Direct Guidance]	16.35%	[Restatement][Direct Guidance]

Table 3: Automatic evaluation results.

Method	BLEU Avg	PBERT	RBERT	FBERT	Dist-2
GPTft+strategy	16.58	72.08	71.92	72.50	21.23
Cot+ChatGPT	18.12	73.20	73.68	73.42	20.64
Cot+GPT4	18.40	73.75	73.93	73.82	23.42
React	17.79	73.50	73.74	72.57	22.67
Cue-Cot	17.93	72.97	73.72	73.52	24.90
Chameleon	18.67	73.38	74.01	73.49	20.60
MentalBlend	**24.17**	**75.73**	**75.68**	**75.79**	**23.75**

professional counselors, offering detailed analysis and guidance for the help-seekers' questions.

Task Definition

We define the task as follows: given a question (SQ) describing a mental health issue and a detailed description (SD) of the mental health issue, the objective is to generate a supportive response that emulates advice from a mental health counselor. This response aims to provide comfort or guidance for the help-seekers, addressing their mental health concerns.

Baselines

We utilize ChatGPT , employing it through API. Initially, we benchmark against classic methods for this task, including: 1) GPTft+strategy – utilizing GPT2 to generate responses with support strategies aimed at providing mental health support and assistance. We then turn our attention to comparisons with other LLM-based approaches, such as: (2) CoT – a method that encourages the generation of reasonable responses through the prompting "Let's think step by step." Additionally, we compare with other typical LLM-based methods, including: (3) ReAct(Yao et al., 2022), (4) Chameleon(Lu et al., 2023), and (5) Cue-CoT(Wang et al., 2023). These approaches share similarities with our study as they all involve leveraging and incorporating external knowledge when invoking Large Language Models (LLMs) to enhance their performance across various downstream tasks.

Metrics

To assess the performance of MentalBlend, we employ automatic evaluation metrics, including BLEU Avg (average of BLEU-1, BLEU-2, BLEU-3, and BLEU-4), BERTscore (PBERT, RBERT, FBERT) and Distinct-2 (Dist-2). BLEU and BERTscore mainly measure the similarity between generated text and responses from professional mental health counselors, while Dist-2 gauges the richness and diversity of vocabulary in the responses.

Automatic Evaluation

Table 3 displays the experimental results. MentalBlend exhibits higher BLEU scores and BERTscores compared to other methods, indicating a more comprehensive and in-depth understanding of help-seekers' needs through the implementation of emotional states understanding and mental health issues. Consequently, the generated supportive responses demonstrates higher content relevance compared to reference responses, a characteristic not present in GPTft+strategy and COT. The experiments reveal that, compared to other LLM-based methods incorporating external knowledge, such as ReAct, Cue-CoT, and Chameleon, MentalBlend performs superiorly, even when ReAct involves knowledge from Wikipedia. Furthermore, MentalBlend surpasses Cue-CoT in BLEU avg and BERTscore, suggesting that despite Cue-CoT's ability to reason and plan across the emotional, psychological, and personality aspects, similar to MentalBlend's stage one, the integration of psychological counseling theories in MentalBlend further deepens the understanding of acquired information into a more profound comprehension of mental health issues, making it superior to Cue-CoT. Finally, MentalBlend outperforms Chameleon, indicating that, compared to Chameleon, which assembles a set of tools to enhance LLMs, MentalBlend integrates a better understanding of mental health issues and the introduction of knowledge from professional psychological counseling theories, resulting in a more effective improvement in mental health support. Moreover, MentalBlend's superior performance in diversity suggests it has

Table 4: Human evaluation results.

Method	Fluency	Helpfulness	Relevance	Empathy	Professionalism
GPTft+strategy	3.08	3.17	3.17	3.25	2.92
Cot+ChatGPT	3.59	3.42	3.25	3.25	3.08
Cot+GPT4	3.67	3.67	3.42	3.42	3.25
React	3.25	3.00	2.92	3.00	3.17
Cue-Cot	3.42	3.42	3.58	3.17	3.42
Chameleon	3.58	3.25	3.60	3.42	3.25
MentalBlend	**3.75**	**3.80**	**4.17**	**3.83**	**4.10**

greater potential to provide services to individuals seeking help, catering to diverse needs and traits compared to other methods.

Human Evaluation

To evaluate the quality of generated supportive responses, we also conduct a human evaluation. We recruit 12 graduate students, major in psychology, to annotate the responses. These expert annotators were asked to rate the responses based on five criteria: fluency, relevance, helpfulness and empathy, and topic relevance (Liu et al., 2021).

Fluency: The response is smooth and easily understandable.

Helpfulness: The response provides comfort or advice for mental health issues and is generally helpful for alleviating mental health concerns.

Relevance: The response closely revolves around specific mental health issues.

Empathy: The response demonstrates warmth, sympathy, and concern.

Professionalism: The response exhibits professionalism similar to that of a human mental health counselor.

The questionnaire includes the following options: 1 - Strongly Disagree, 2 - Disagree, 3 - Neutral, 4 - Agree, 5 - Strongly Agree.

Table 4 provides a comprehensive overview of the outcomes from the manual evaluations. Broadly speaking, the responses crafted by Large Language Models (LLMs) under the guidance of MentalBlend garner more favor from human evaluators when compared to alternative baseline methods. Notably, MentalBlend demonstrates a substantial edge in terms of relevance, showcasing its exceptional ability to concentrate on the psychological concerns of those seeking assistance and delve into profound analyses of the presented problems. Furthermore, the responses generated by MentalBlend are also favored for their professionalism. We posit that this preference can be ascribed to MentalBlend endowing LLMs with a professional approach to thinking and responding, facilitated by the integration of advanced psychological counseling theories.

Limitations and Future Work

Despite MentalBlend's use of prompt engineering to encourage large language models to delve into the emotions and mental health issues of help-seekers, it is undeniable that challenges persist in accurately reasoning about users' emotions, needs, or mental health issues. This is primarily due to the fact that mental health issues often require a profound understanding of users' personal experiences and contexts. The complexity of these issues may lead to instances where MentalBlend encounters challenges in accurately reasoning about users' complex mental health issues. To address the challenge, our next efforts will focus on in-depth research into the precise fine-tuning of large language models (LLMs), with a specific emphasis on mitigating issues related to inaccurate reasoning, thereby enhancing the model's performance in the field of mental health.

Furthermore, we aim to broaden the application scope of MentalBlend, especially in other relevant research domains that face challenges related to emotional complexity and expressive intricacy.

Conclusion

In this study, we introduce MentalBlend as a framework aimed at addressing the challenges that LLMs face in providing high-quality online mental health support. Through experiments, we demonstrate that prompts centered around deepening the understanding of mental health issues, provided by MentalBlend, effectively enhance LLMs' comprehension of users' psychological concerns. Furthermore, human evaluation results indicate that, under the guidance of professional psychological counseling theories, ChatGPT based on MentalBlend can generate psychologically supportive responses similar to those provided by professional counselors. Simultaneously, the framework's automatic classification feature ensures that these responses closely align with the unique needs of users.

Overall, our research contributes to advancing the capabilities of LLMs in the domain of online mental health support.

Ethical Considerations

Despite our use of datasets containing anonymously posted content in this study, we strictly adhere to robust privacy protocols (Benton, Coppersmith, & Dredze, 2017; Nicholas, Onie, & Larsen, 2020)to minimize privacy impact. This precaution is crucial, as social media datasets may potentially reveal users' thoughts and contain sensitive personal information. The social posts we utilize are explicitly public and sourced from the Yixinli platform. To prevent misuse, all examples mentioned in our paper have been paraphrased and obfuscated using a moderate disguising scheme (Bruckman, 2002). It's important to note that we do not access user profiles on social media, identify users, or engage in any interactions with them. The primary objective of our study is to leverage social media as an early source of information, assisting researchers in exploring how LLMs can provide more professional mental health support, especially in non-clinical settings. However, it must be emphasized that the mental health support responses generated by MentalBlend cannot replace professional mental health support services and psychiatric diagnoses.

Acknowledgments

This work is supported by the Science and Technology Innovation 2030 - Major Project "New Generation Artificial Intelligence (2030)" (2022ZD0116407) and the National Natural Science Foundation of China.

References

Beck, A. T. (1979). *Cognitive therapy and the emotional disorders*. Penguin.

Benton, A., Coppersmith, G., & Dredze, M. (2017). Ethical research protocols for social media health research. In *Proceedings of the first acl workshop on ethics in natural language processing* (pp. 94–102).

Bruckman, A. (2002). Studying the amateur artist: A perspective on disguising data collected in human subjects research on the internet. *Ethics and Information Technology*, *4*, 217–231.

Burnap, P., Colombo, W., & Scourfield, J. (2015). Machine classification and analysis of suicide-related communication on twitter. In *Proceedings of the 26th acm conference on hypertext & social media* (pp. 75–84).

Cao, L., Zhang, H., Feng, L., Wei, Z., Wang, X., Li, N., & He, X. (2019). Latent suicide risk detection on microblog via suicide-oriented word embeddings and layered attention. *arXiv preprint arXiv:1910.12038*.

Chalnick, A., & Billman, D. (1988). Unsupervised learning of correlational structure. In *Proceedings of the tenth annual conference of the cognitive science society* (pp. 510–516). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chen, Q., & Liu, D. (2023). Dynamic strategy chain: Dynamic zero-shot cot for long mental health support generation. *arXiv preprint arXiv:2308.10444*.

Cooper, M., & McLeod, J. (2011). Person-centered therapy: A pluralistic perspective. *Person-Centered & Experiential Psychotherapies*, *10*(3), 210–223.

Coppersmith, G., Leary, R., Crutchley, P., & Fine, A. (2018). Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights*, *10*, 1178222618792860.

De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., & Kumar, M. (2016). Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of the 2016 chi conference on human factors in computing systems* (pp. 2098–2110).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Feigenbaum, E. A. (1963). The simulation of verbal learning behavior. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought.* New York: McGraw-Hill.

Good, A., & Sambhanthan, A. (2014). Accessing web based health care and resources for mental health: interface design considerations for people experiencing mental illness. In *Design, user experience, and usability. user experience design for everyday life applications and services: Third international conference, duxu 2014, held as part of hci international 2014, heraklion, crete, greece, june 22-27, 2014, proceedings, part iii 3* (pp. 25–33).

Hill, J. A. C. (1983). A computational model of language acquisition in the two-year old. *Cognition and Brain Theory*, *6*, 287–317.

Hofmann, S. G., Sawyer, A. T., & Fang, A. (2010). The empirical status of the "new wave" of cognitive behavioral therapy. *Psychiatric Clinics*, *33*(3), 701–710.

Kaczkurkin, A. N., & Foa, E. B. (2015). Cognitive-behavioral therapy for anxiety disorders: an update on the empirical evidence. *Dialogues in clinical neuroscience*, *17*(3), 337–346.

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large language models are zero-shot reasoners. *Advances in neural information processing systems*, *35*, 22199–22213.

Kruzan, K. P., Williams, K. D., Meyerhoff, J., Yoo, D. W., O'Dwyer, L. C., De Choudhury, M., & Mohr, D. C. (2022). Social media-based interventions for adolescent and young adult mental health: A scoping review. *Internet interventions*, 100578.

Linehan, M. M. (1987). Dialectical behavioral therapy: A cognitive behavioral approach to parasuicide. *Journal of Personality disorders*, *1*(4), 328–333.

Livingston, J. D., Cianfrone, M., Korf-Uzan, K., & Coniglio, C. (2014). Another time point, a different story: one year effects of a social media intervention on the attitudes of young people towards mental health issues. *Social psychiatry and psychiatric epidemiology*, *49*, 985–990.

Lu, P., Peng, B., Cheng, H., Galley, M., Chang, K.-W., Wu, Y. N., ... Gao, J. (2023). Chameleon: Plug-and-play compositional reasoning with large language models. *arXiv preprint arXiv:2304.09842*.

Madaan, A., Tandon, N., Gupta, P., Hallinan, S., Gao, L., Wiegreffe, S., ... others (2023). Self-refine: Iterative refinement with self-feedback. *arXiv preprint arXiv:2303.17651*.

Martinengo, L., Lum, E., & Car, J. (2022). Evaluation of chatbot-delivered interventions for self-management of depression: content analysis. *Journal of affective disorders*, *319*, 598–607.

Matlock, T. (2001). *How real is fictive motion?* Doctoral dissertation, Psychology Department, University of California, Santa Cruz.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nicholas, J., Onie, S., & Larsen, M. E. (2020). Ethics and privacy in social media research for mental health. *Current psychiatry reports*, *22*, 1–7.

Ohlsson, S., & Langley, P. (1985). *Identifying solution paths in cognitive diagnosis* (Tech. Rep. No. CMU-RI-TR-85-2). Pittsburgh, PA: Carnegie Mellon University, The Robotics Institute.

Park, M., McDonald, D., & Cha, M. (2013). Perception differences between the depressed and non-depressed users in twitter. In *Proceedings of the international aaai conference on web and social media* (Vol. 7, pp. 476–485).

Pendse, S. R., Karusala, N., Siddarth, D., Gonsalves, P., Mehrotra, S., Naslund, J. A., … Sharma, A. (2019). Mental health in the global south: challenges and opportunities in hci for development. In *Proceedings of the 2nd acm sigcas conference on computing and sustainable societies* (pp. 22–36).

Ridout, B., & Campbell, A. (2018). The use of social networking sites in mental health interventions for young people: systematic review. *Journal of medical Internet research*, *20*(12), e12244.

Shinn, N., Labash, B., & Gopinath, A. (2023). Reflexion: an autonomous agent with dynamic memory and self-reflection. *arXiv preprint arXiv:2303.11366*.

Shrager, J., & Langley, P. (Eds.). (1990). *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.

Sun, H., Lin, Z., Zheng, C., Liu, S., & Huang, M. (2021). Psyqa: A chinese dataset for generating long counseling text for mental health support. *arXiv preprint arXiv:2106.01702*.

Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of depression-related posts in reddit social media forum. *Ieee Access*, *7*, 44883–44893.

Tsugawa, S., Kikuchi, Y., Kishino, F., Nakajima, K., Itoh, Y., & Ohsaki, H. (2015). Recognizing depression from twitter activity. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 3187–3196).

Wang, H., Wang, R., Mi, F., Wang, Z., Xu, R., & Wong, K.-F. (2023). Chain-of-thought prompting for responding to in-depth dialogue questions with llm. *arXiv preprint arXiv:2305.11792*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., … others (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, *35*, 24824–24837.

Wubbolding, R. E., Casstevens, W. J., & Fulkerson, M. H. (2017). Using the wdep system of reality therapy to support person-centered treatment planning. *Journal of Counseling & Development*, *95*(4), 472–477.

Xu, X., Chikersal, P., Dutcher, J. M., Sefidgar, Y. S., Seo, W., Tumminia, M. J., … others (2021). Leveraging collaborative-filtering for personalized behavior modeling: a case study of depression detection among college students. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *5*(1), 1–27.

Xu, Z., Pérez-Rosas, V., & Mihalcea, R. (2020). Inferring social media users' mental health status from multimodal information. In *Proceedings of the twelfth language resources and evaluation conference* (pp. 6292–6299).

Yang, K., Ji, S., Zhang, T., Xie, Q., & Ananiadou, S. (2023). On the evaluations of chatgpt and emotion-enhanced prompting for mental health analysis. *arXiv preprint arXiv:2304.03347*.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., … others (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*.