Classifying Anxiety and Depression through LLMs Virtual Interactions: A Case Study with ChatGPT

1st Yongfeng Tao
School of Information Science and Engineering, Lanzhou University
Lanzhou, China
taoyf21@lzu.edu.cn

2nd Minqiang Yang*
School of Information Science and Engineering, Lanzhou University
Lanzhou, China
yangmq@lzu.edu.cn

3rd Hao Shen
School of Information Science and Engineering, Lanzhou University
Lanzhou, China
shenhaoshiki@163.com

4th Zhichao Yang
School of Information Science and Engineering, Lanzhou University
Lanzhou, China
yzhichao49@gmail.com

5th Ziru Weng
School of Information Science and Engineering, Lanzhou University
Lanzhou, China
220220943190@lzu.edu.cn

6th Bin Hu*
School of Information Science and Engineering, Lanzhou University
Lanzhou, China
bh@lzu.edu.cn

2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) | 979-8-3503-3748-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/BIBM58861.2023.10385305

***Abstract*—Mental health has long been studied, and many AI-based approaches have been proposed for diagnosis and adjunctive therapy. The emergence of Pre-trained Large Language Models (LLMs) has had a profound impact on various fields, but the potential of using ChatGPT for cognitive behavior therapy is largely unexplored. Therefore, there is an urgent need to build and design a virtual interactive framework for assisted diagnosis/treatment. In this paper, we present a virtual interaction framework based on LLMs that allows participants to engage in a dialogue with a virtual character, analyse mental health issues through augmented LLMs, and make suggestions during the dialogue to alleviate the psychological problems they are currently facing. Based on this framework, we develop a use case for the application of ChatGPT in the field of emotional disorders. Specifically, we use data from question-and-answer dialogues in real-life scenarios to populate the current exploration of ChatGPT's potential for depression and anxiety detection. The case study shows the great potential of ChatGPT in the analysis of depression and anxiety tests. The feasibility of a virtual interaction framework based on LLMs has been preliminarily demonstrated.**

***Index Terms*—affective disorders, depression detection, anxiety detection, virtual character**

I. Introduction

The assessment of mental disorders in the field of affective computing has been one of the research hotspots in recent years. Mental disorders, as a common mental health problem, have affected people's daily activities and most of them require treatment for improvement and recovery. There has also been a gradual increase in the prevalence of depression in recent years, and there is already a related social burden. Previous studies have shown that up to 85% of depressive disorders are associated with anxiety [1], but [2] have shown that there are also phenotypic differences between depression and anxiety, and that improving these phenotypic characteristics could improve clinical assessment. Prevention and treatment of mental disorders include both clinical diagnosis by professionals and treatment after diagnosis. Clinical diagnosis has the disadvantages of subjectivity of experts and uncooperativeness of participants. To alleviate this problem, some researchers in recent years have used physiological [3], [4] or behavioural [5], [6] data to model objective detection methods. Treatment of mental disorders consists mainly of medication and psychotherapy. However, studies have shown that psychotherapy has comparable effects to medication in the short term, but may be more effective in the long term [7]. Current psychotherapies include cognitive behavioural therapy (CBT), problem-solving therapy, etc., with CBT being the most widespread and common method [7]. CBT was also slowly being introduced into the remote treatment of abnormal mental states, with relatively good results. Psychotherapy based on asynchronous messaging has shown its viability as an adjunct treatment in combination with traditional counselling services [8]. A virtual voice coach, Lumen, has been proposed for the emotional treatment of subjects and has shown good results in the cognitive control of depression and anxiety symptoms [9]. With the development of avatar technology and the Pre-trained Large Language Models (LLMs), it has become possible to build a model with interactive capabilities for the adjunctive treatment of mental disorders. Therefore, we propose a virtual interaction framework based on enhanced LLMs, and the avatars in this framework is able to provide participants with some advice and methods to alleviate the mental health problems they are facing, based on the content of the dialogue.

For LLMs, such as ChatGPT [10], have been widely used in various domains since their release. ChatGPT has been used in text classification, text generation, education, medicine, etc., and has a wide range of applications in data visualisation, information extraction, data enhancement, quality assessment and multimodal data processing [11]. Overall, it can be divided into two main categories: fine-tuning the application of mod-

*Corresponding author.

979-8-3503-3748-8/23/$31.00 ©2023 IEEE

els [12], and testing the potential performance of models in different areas. Because ChatGPT can process text flexibly and has good contextual understanding and information inference, there are many studies testing its potential performance in sentiment analysis, personality recognition and affective disorders. In [13], the results of the zero-shot experiment suggest a potential application of the ChatGPT in mental health classification tasks. In [14], the potential for text classification in three affective computing problems was explored using ChatGPT. In order to better use ChatGPT for analysing relevant tasks, studies have explored the ability to zero-point classify text by introducing prompts. In [15], a variety of emotion-related cues were used to further explore the interpretability of ChatGPT on LLM for the 5-task emotional and mental health dataset using ChatGPT. For affective disorders, the study [16] used the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) depression indicators as prompts and fed some of the social media users' posts into the LLM model to determine whether they were depressed or not.

To explore the performance of LLM in addition to abnormal sentiment analysis, researchers have used LLM to classify depressed and non-depressed, depressed and suicidal on different datasets, as shown in Table I. These studies were mainly based on datasets of posts on social media platforms (Reddit, Twitter, Weibo, etc.), e.g. Depression_Reddit (DR) [17], CLPsych15 [18], the Twitter multimodal depression dataset (TMDD) [19], the Weibo user depression detection dataset (WU3D) [20], the depression/suicide cause detection dataset (CAMS) [21], etc., and most of these datasets were in English language. Based on the proposed framework, a case study of the potential of ChatGPT in affective disorders is investigated. To the best of our knowledge, no Chinese-language Question and Answer (Q&A) data from real-life scenarios have been used to validate the performance of LLM in anomalous sentiment detection. The main contributions of this paper are as follows:

- We have constructed a virtual interaction framework based on enhanced LLMs. We also discuss the outstanding issues in the framework and provide ideas for future research.
- We evaluated for the first time the potential of the ChatGPT model to detect depression and anxiety in real-life scenarios using a database of question-answer dialogues.
- The descriptive information prompted by speech enhances ChatGPT's ability to classify depression and anxiety, and this introduces an easy way to prompt input for multiple modalities.

The rest of the paper is structured as follows. Section II describes the virtual interaction framework based on enhanced LLMs, Section III describes the affective disorder case study, and Section IV discusses the challenges and open issues of the framework. Section V concludes this paper.

II. Virtual Interactive Framework based on LLMs

Physiological data, such as EEG, require specialised equipment to collect, so the use of visual and auditory data can be used to assess and support the treatment of abnormal mental states in a non-intrusive way. Our proposed virtual interaction framework is based on enhanced LLMs is shown in Fig. 1, and mainly consists of user-side and cloud-based augmented large model composition.

TABLE I
THE APPLICATION OF LLMS TO THE EXPLORATION OF POTENTIAL ON AFFECTIVE DISORDERS

Database	Condition	Use Post Num	Platform	Language
CLPsych15 [18]	DP / HC	300 [15]	Twitter	English
TMDD [19]	DP / HC	232,895/ 879,025 [16]	Twitter	English
WU3D [20]	DP / HC	39,595/ 80,167 [16]	Weibo	Chinese
DR [17]	DP / HC	406 [15]	Reddit	English
CAMS [21]	DP / SP	626 [15]	Reddit	English
[22]	DP / SP	- [14]	Reddit	English

Note: DP, HC, and SP denote depressed patients, healthy controls and suicidal patients, respectively.

Facial behaviour includes facial expression data and eye movement data, and are the most commonly used forms of non-verbal communication. Studies of eye movement data in depressed patients have shown a negative attentional bias [23]. The collection of facial behaviour data does not require much cooperation from the participants, and it can be used as an input data to the virtual interaction framework is based on enhanced LLMs to correct for effective changes in the virtual character's facial emotions during the interaction with the participants. Significantly different body movements, such as slower arm swinging and walking, can be observed in depressed people [24], which is associated with fewer brain neurotransmitters in their negative mood. Similarly, the acquisition of body movement data can be captured by a high-resolution camera in a distraction-free environment. Participants with abnormal emotions showed different speech characteristics compared to normal individuals [25]. Depressed patients mainly showed longer speech pauses, lack of rhythm, slower speech rate, etc. The main task of the user terminal is to collect the participant's behavioural data, pre-process it, and provide real-time feedback on some suggestions and necessary intervention methods of the virtual character during the interaction, such that these methods are used to alleviate the current participant's abnormal mental state. In our framework, we have designed an assisted treatment scheme specifically for patients with mental disorders, which mainly uses LLMs and avatar generation technology by taking the behavioural data captured by the camera and microphone and pre-processing them before inputting them as prompts to the corresponding modules in the framework.

The framework can provide real-time feedback on the user's psychological state and gradually guide the user to solve the psychological problems faced in response to bad emotions. The extracted effective emotional features are processed by processing multimodal data such as video, audio and text. These features not only serve as a reference for the virtual character's emotional adjustment, but also help to prompt the

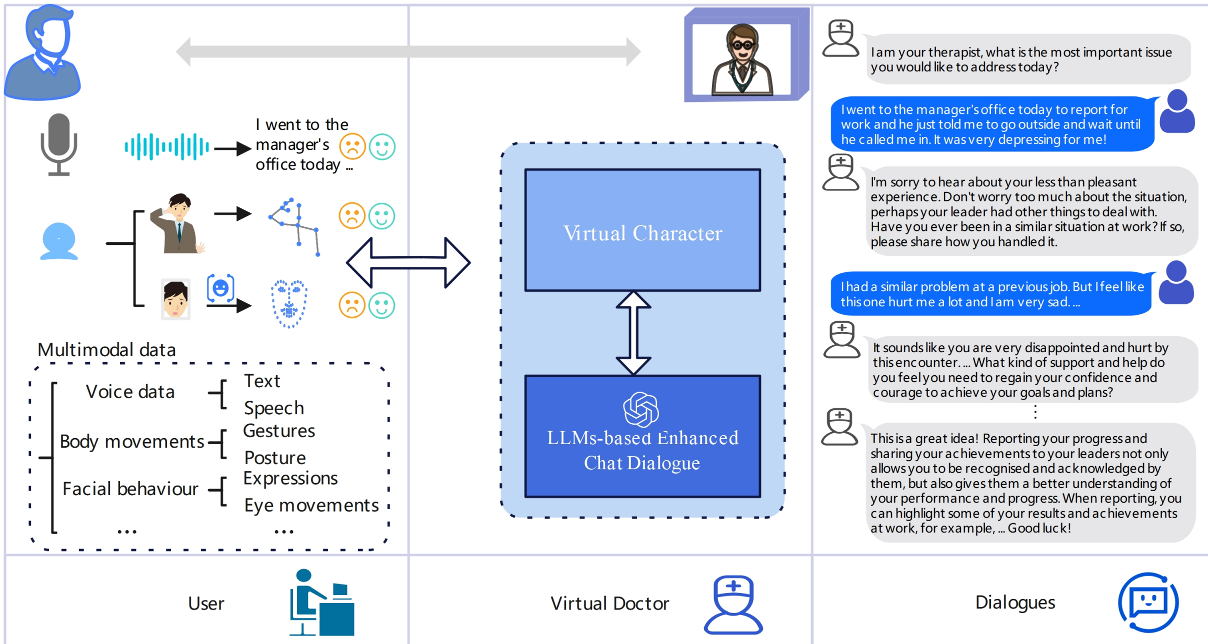

Fig. 1. The virtual interaction framework is based on enhanced LLMs. The main multimodal data collected by camera, microphone, etc. are processed and transmitted to the virtual doctor in real time. The virtual doctor takes into account the textual content of the speech and the emotion conveyed by the patient in order to provide timely feedback.

LLMs to accurately assess the participants' psychological state in real time and provide assessment results, psychological counselling methods, and so on. The feedback information is fed back to the user through the virtual character's voice, enabling him or her to effectively reduce stress and improve the quality of life and work. At the end of the interaction with the user, the framework summarises the problems faced by the user and assigns follow-up tasks that need to be attended to and completed in life.

With the development of LLMs and virtual character technology, the virtual interaction framework based on enhanced LLMs is technically feasible. This framework allows patients to be treated regardless of time and place, and better protects their privacy without fear of being discovered or commented on by others, allowing them to be more relaxed and focused on the therapeutic process. the virtual interaction framework is designed to simulate conversations in real-life scenarios and uses CBT to achieve the purpose of assisting in the treatment of psychiatric disorders. By interacting with the virtual character, patients can provide feedback on their needs and preferences to create a more personalised treatment plan and improve outcomes. This framework can alleviate the problem of low doctor-patient ratios to some extent and save time and effort for the therapist, thereby reducing the cost of treatment for the patient.

III. Case Study on Affective Disorders

A. Materials and Methods

1) Participants and Preprocessing: To investigate the potential performance of ChatGPT for the detection of anxiety and depression in real-life dialogue scenarios. Seventy-five patients with anxiety and 64 patients with depression were recruited from the Peking University Sixth Hospital in China, and demographic information is shown in Table II. Diagnosis recommended by at least one clinical psychiatrist. The local ethics committee approved consent forms and studies designed for biomedical research at the Peking University Sixth Hospital in China, following the Code of Ethics of the World Medical Asso-ciation (Declaration of Helsinki), and written informed consent forms were signed by all participants before the experiment.

The collected speech data consisted of nine Q&A tasks, and six of the Q&A dialogue tasks were related to the participants' daily activities and did not involve personal privacy, and were therefore chosen for this experiment. To input the speech data into ChatGPT, the text data in the speech is extracted using a combination of the Whisper [26] model and manual methods. The text data contains some private information about places, people, time and so on, we have done the privacy protection processing and used "xxx" to replace the original text.

TABLE II
Demographic variables for participants with depressive disorder (DP) and anxiety patient (AP).

Variables	DP	AP
N (M/F)	64 (19/45)	75 (33/42)
Age (m+SD)	32.09 $\pm$ 10.55	37.83 $\pm$ 11.31

2) Feature Extraction: Previous studies have shown significant differences in speech characteristics between people with different affective disorders, e.g. depressed people speak more slowly, have longer pauses, etc [27]. We found texts that appeared to have relatively normal descriptions, but the speech sounded as if it was spoken in a choked, sighing voice. Therefore, we believe that there is a limit to the amount of emotional information that can be conveyed by text alone. To characterise speech, we extracted speech rate (words/sec), average pitch (HZ), speech rhythm (BPM), and pauses (/sec), which are entered as prompts in ChatGPT. Features are extracted from speech using open source Python packages, e.g. librosa, etc. In addition, as many current LLMs can only process text, the following prompt description is given for speech speed features ($speed$):

If $speed > 200$, make $\mathcal{D}_{speed}$ = "faster speech speed, $\{speed\}$ words per minute". If $speed < 120$, make $\mathcal{D}_{speed}$ = "slower speaking speed, $\{speed\}$ words per minute". Otherwise, $\mathcal{D}_{speed}$ = "moderate speaking speed, $\{speed\}$ words per minute". The description of pitch of speech is similar to the above.

3) Question Formulation: The format of the question whether to add a voice message description is shown in the following snippets. The question in the question and answer dialogue is replaced by the section [*question*], the participant's answer is replaced by [*text*] and the voice message description is replaced by $\mathcal{D}$. The formula for calculating the prompt information for these two questions is as follows:

a) For the question without adding voice description information, we pose such a question:

Analyze the dialogue to determine whether the respondent's emotional state is depression or anxiety. Question: [question]*, Answer:* [text]*,..., Question:* [question]*, Answer:* [text]*, tell me the respondent's emotion in the following format: "anxiety" or "depression". Just give me the final word, no further analysis.*

b) For questions with voice descriptions added, we ask the following question:

Analyse the conversation to determine whether the respondent's emotional state is depression or anxiety. Question: [question]*, Answer:* [text]*, where* [$\mathcal{D}$]*,..., Question:* [question]*, Answer:* [text]*, where* [$\mathcal{D}$]*. Tell me the mood of the respondent in the following format "anxiety" or "depression". Just give me the final word, no further analysis.*

ChatGPT's response answers may appear in arbitrary formats, although we have fixed the output answers. Sometimes the given answers have results that are inconsistent with the given answer, such as extra punctuation, "my guess is ...", and we adjust the output according to the principle of semantic similarity.

B. Results

In this section, we explored the potential of ChatGPT for anxiety and depression classification, mainly following different prompts. First, as the experimental data used in this paper were six task questions about participants' daily activities, we need to explore the single question and answer dialogues for depression and anxiety classification effects. Secondly, we chose two of the questions with high accuracy for anxiety and depression classification to subsequently explore the effect on classification between different descriptions of speech information.

1) Compare Single Question: Before exploring multiple questions together, it is important to explore the potential of ChatGPT to discriminate between depression and anxiety using each individual question. We explored individual questions using the prompt information question a) in Question Formulation. Classification accuracy is best for question Q4 and second best for question Q2, as shown in Table III. And the values of Rec, F1, and Pre under questions Q2 and Q4 are relatively stable compared to the evaluation metrics under other questions, for example the Rec evaluation metrics of Q1 and Q6 are as high as 96%, suggesting that ChatGPT is more inclined to categorise the population into one category through these questions. We found that Q2 and Q4 are questions about the participant's recent mood and 'what to do if the participant is unhappy', and we chose these two question dialogue data for the subsequent experiments. In order to explore the contextual information between question and answer dialogues, we use the prompt information from question a) in Question Formulation to determine the input data for the union of the two questions. As shown in the last row of Table III, the evaluation metrics of Acc, Rec, F1, and Pre are 67.62%, 72.00%, 70.58%, and 69.23%, respectively, which are significantly higher than the categorical metrics for single questions. This indicates that the contextual information in the question-answer dialogue is beneficial for improving the classification effect.

TABLE III
ACCURACY, RECALL, F1-SCORE, AND PRECISION UNDER SINGLE PROBLEMS.

Question	Acc(%)	Rec(%)	F1(%)	Pre(%)
Q1	56.83	96.00	70.59	55.81
Q2	57.55	76.00	65.89	58.16
Q3	52.52	88.00	66.66	53.66
Q4	60.43	80.00	68.57	60.00
Q5	53.23	96.00	68.89	53.73
Q6	53.96	92.00	68.31	54.33
Q2/Q4	67.62	72.00	70.58	69.23

2) Compare Speech Descriptions: Several studies have extracted audio features in speech for use in mental disorders, and speech features can be used as objective biomarkers for assessing the severity of depression, suggesting that speech features may reflect abnormal mental states in people [27]. Compared to healthy controls, depressed patients had slower speech, longer pauses and longer reaction times, indicating the presence of a psychomotor disorder. To explore the effectiveness of speech features and speech descriptions in prompting ChatGPT to categorise depression and anxiety, we set up the input data according to the prompting message question b) in Question Formulation. As shown in Table IV, the classification results for the prompted input with added speech description information are significantly improved in every evaluation index compared to the input without added speech description prompts. This shows that ChatGPT is better able to distinguish between the two categories for text data with speech description information. Better classification results are achieved with the addition of rhythm and speech rate, which can reach 79.14% and 76.97% accuracy respectively. The confusion matrix for pause and speech rhythm is shown in Fig. 2. From the figure, it is evident that compared to adding pause prompts, incorporating speech rate prompts is more beneficial for the classification of anxiety. It is worth noting that of these four speech feature descriptions, pitch has the lowest classification accuracy.

Analysing negative words in speech texts through word cloud (Fig. 3) and word frequency (Fig. 4) visualisation, we observed differences in negative words between people with anxiety and depression. Words such as 'anxious' and 'angry' appear frequently in the speech texts of people with anxiety disorders, whereas people with depression use the term 'bad mood' more often. Fig. 3 shows that, overall, anxiety has a more negative vocabulary, which is related to differences in the verbalisation of anxiety and depression. As shown in Fig. 4, the highest frequency of the negative word 'insomnia' is found in the anxiety and depression speech texts, which indicates that both groups have some degree of sleep disorders.

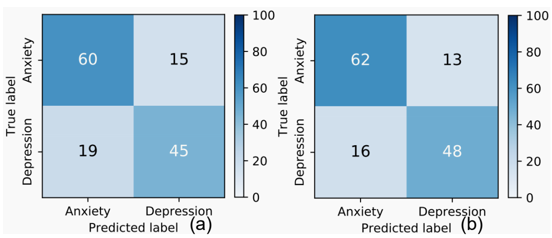

Fig. 2. Subfigure a and subfigure b represent confusion matrices of pauses (/sec) and speech rhythm (BPM), respectively.

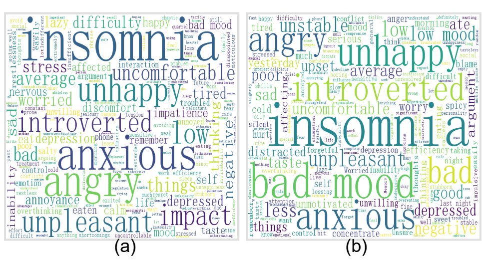

Fig. 3. Subfigure a and subfigure b represent the word cloud distribution of negative words in the speech texts of anxiety and depression patients, respectively.

IV. Research Challenges and Open Issues

We conducted a case study of chatgpt's potential in depression and anxiety using the framework, which proved to have significant potential. Combined with the various modules used

TABLE IV

Accuracy, recall, F1-score, and precision of different speech features as prompts under the union of two problems.

Acoustic	Acc(%)	Rec(%)	F1(%)	Pre(%)
Pause	75.74	80.00	77.92	75.94
Pitch	74.10	80.00	76.92	74.07
Rhythm	79.14	82.66	81.05	79.49
Speed	76.97	78.66	78.66	78.67

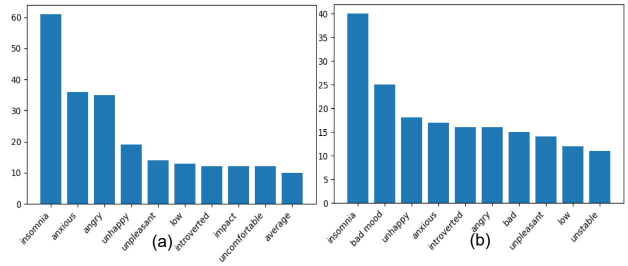

Fig. 4. Subfigure a and subfigure b represent the top ten negative words in terms of word frequency for negative words in the speech texts of anxious and depressed patients, respectively.

in the framework to collaborate on virtual interactions, we believe that the following aspects are worth exploring further:

A. Generation of Virtual Characters

The behavioral capabilities of virtual characters include facial expressions, body movements, and language. Facial expressions are the core of constructing virtual characters as they enhance the sense of immersion for patients. The challenge in creating a virtual doctor system lies in achieving realistic, smooth, and emotionally rich facial expressions and movements. To reduce the cognitive load on participants, the language, facial expressions, and body movements of virtual characters must be natural and fluid during interactions. Additionally, it is necessary to design language and behavioral guidelines for medical assistance, such as determining which behaviors should be prohibited. To accomplish these objectives, extensive experimentation and consultation with industry experts are required to develop various interaction strategies and control mechanisms.

B. Privacy and Ethics

In the proposed framework, the collection of large amounts of multimodal data from participants poses ethical challenges to participant privacy. How to introduce federated machine learning into the framework is a concern. As CBT therapy is a long-term treatment, the framework must fully respect the patient's wishes without coercing or inducing the patient to make unwanted or inappropriate decisions. Following professional knowledge and ethical guidelines during the treatment process to ensure the reliability, controllability and trustworthiness of the virtual treatment is an important challenge.

C. Robust Model

To provide a robust CBT process, a multimodal fusion model is required that combines different types of data such as speech, facial behaviour, etc. to provide more accurate feedback. The challenge is how to effectively represent the fusion of different modal data, exploiting their complementarities and redundancies to improve the generalisation and robustness of the model, and how to accurately align the different modal data to improve the understanding and inference of the model. The patient's mental state also affects speech habits such as speech rate, pauses and accents, which may lead to incorrect or missing recognition results from existing speech models, affecting dialogue comprehension and generation. How to use large language models in relation to the condition is also a problem that needs to be overcome.

V. Conclusion

This paper introduces a virtual interaction framework based on LLMs to help alleviate negative psychological states in participants. By analysing the data obtained from the Q&A dialogues, the findings have revealed the significant potential of ChatGPT in classifying depression and anxiety. To improve the classification performance, four language features were extracted as additional clues, with prosodic and speech rate features showing positive effects on classification. However, the current research cases still have some limitations, including a small number of tasks and recruited participants. Future studies will continue to recruit more participants and increase the number of tasks to further optimize this framework. Despite the challenges in fine-grained classification, the virtual interaction framework based on LLM is worth exploring and implementing based on current technological feasibility and potential demand.

Acknowledgment

This work was supported in part by the National Key Research and Development Program of China (Grant No. 2019YFA0706200), in part by the National Natural Science Foundation of China (Grant No.62227807), in part by the Natural Science Foundation of Gansu Province, China(Grant No. 22JR5RA488), in part by the Fundamental Research Funds for the Central Universities (Grant No. lzujbky-2023-16). The authors would like to acknowledge Yanfeng Zhang at Georgetown University for providing the relevant computing resources.

References

[1] J. M. Gorman, “Comorbid depression and anxiety spectrum disorders,” *Depression and anxiety*, vol. 4, no. 4, pp. 160–168, 1996.

[2] Y. N. Gao, B. Coombes, E. Ryu, V. Pazdernik, G. Jenkins, R. Pendegraft, J. Biernacka, and M. Olfson, “Phenotypic distinctions in depression and anxiety: a comparative analysis of comorbid and isolated cases,” *Psychological Medicine*, pp. 1–9, 2023.

[3] D. Wang, C. Lei, X. Zhang, H. Wu, S. Zheng, J. Chao, and H. Peng, “Identification of depression with a semi-supervised gcn based on eeg data,” in *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2021, pp. 2338–2345.

[4] X. Zhang, W. Liang, T. Ding, J. Pan, J. Shen, X. Huang, and J. Gao, “Individual similarity guided transfer modeling for eeg-based emotion recognition,” in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2019, pp. 1156–1161.

[5] Y. Tao, M. Yang, Y. Wu, K. Lee, A. Kline, and B. Hu, “Depressive semantic awareness from vlog facial and vocal streams via spatio-temporal transformer,” *Digital Communications and Networks*, 2023.

[6] B. Hu, Y. Tao, and M. Yang, “Detecting depression based on facial cues elicited by emotional stimuli in video,” *Computers in Biology and Medicine*, vol. 165, p. 107457, 2023.

[7] P. Cuijpers, C. Miguel, M. Harrer, C. Y. Plessen, M. Ciharova, D. Ebert, and E. Karyotaki, “Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including 409 trials with 52,702 patients,” *World Psychiatry*, vol. 22, no. 1, pp. 105–115, 2023.

[8] J. Song, B. Litvin, R. Allred, S. Chen, T. D. Hull, and P. A. Areán, “Comparing message-based psychotherapy to once-weekly, video-based psychotherapy for moderate depression: Randomized controlled trial,” *Journal of Medical Internet Research*, vol. 25, p. e46052, 2023.

[9] T. Kannampallil, O. A. Ajilore, N. Lv, J. M. Smyth, N. E. Wittels, C. R. Ronneberg, V. Kumar, L. Xiao, S. Dosala, A. Barve *et al.*, “Effects of a virtual voice-based coach delivering problem-solving treatment on emotional distress and brain function: a pilot rct in depression and anxiety,” *Translational Psychiatry*, vol. 13, no. 1, p. 166, 2023.

[10] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, “Language models are few-shot learners,” *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[11] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu *et al.*, “Summary of chatgpt/gpt-4 research and perspective towards the future of large language models,” *arXiv preprint arXiv:2304.01852*, 2023.

[12] M. Maaz, H. Rasheed, S. Khan, and F. S. Khan, “Video-chatgpt: Towards detailed video understanding via large vision and language models,” *arXiv preprint arXiv:2306.05424*, 2023.

[13] B. Lamichhane, “Evaluation of chatgpt for nlp-based mental health applications,” *arXiv preprint arXiv:2303.15727*, 2023.

[14] M. M. Amin, E. Cambria, and B. W. Schuller, “Will affective computing emerge from foundation models and general ai? a first evaluation on chatgpt,” *IEEE Intelligent Systems*, vol. 38, p. 2, 2023.

[15] K. Yang, S. Ji, T. Zhang, Q. Xie, and S. Ananiadou, “On the evaluations of chatgpt and emotion-enhanced prompting for mental health analysis,” *arXiv preprint arXiv:2304.03347*, 2023.

[16] W. Qin, Z. Chen, L. Wang, Y. Lan, W. Ren, and R. Hong, “Read, diagnose and chat: Towards explainable and interactive llms-augmented depression detection in social media,” *arXiv preprint arXiv:2305.05138*, 2023.

[17] I. Pirina and Ç. Çöltekin, “Identifying depression on reddit: The effect of training data,” in *Proceedings of the 2018 EMNLP workshop SMM4H: the 3rd social media mining for health applications workshop & shared task*, 2018, pp. 9–12.

[18] G. Coppersmith, M. Dredze, C. Harman, K. Hollingshead, and M. Mitchell, “Clpsych 2015 shared task: Depression and ptsd on twitter,” in *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, 2015, pp. 31–39.

[19] T. Gui, L. Zhu, Q. Zhang, M. Peng, X. Zhou, K. Ding, and Z. Chen, “Cooperative multimodal approach to depression detection in twitter,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 110–117.

[20] Y. Wang, Z. Wang, C. Li, Y. Zhang, and H. Wang, “A multitask deep learning approach for user depression detection on sina weibo,” *arXiv preprint arXiv:2008.11708*, 2020.

[21] M. Garg, C. Saxena, S. Saha, V. Krishnan, R. Joshi, and V. Mago, “Cams: An annotated corpus for causal analysis of mental health issues in social media posts,” in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, 2022, pp. 6387–6396.

[22] V. Desu, N. Komati, S. Lingamaneni, and F. Shaik, “Suicide and depression detection in social media forums,” in *Smart Intelligent Computing and Applications, Volume 2: Proceedings of Fifth International Conference on Smart Computing and Informatics (SCI 2021)*. Springer, 2022, pp. 263–270.

[23] M. Yang, C. Cai, and B. Hu, “Clustering based on eye tracking data for depression recognition,” *IEEE Transactions on Cognitive and Developmental Systems*, pp. 1–1, 2022.

[24] H. Lu, S. Xu, X. Hu, E. Ngai, Y. Guo, W. Wang, and B. Hu, “Postgraduate student depression assessment by multimedia gait analysis,” *IEEE MultiMedia*, vol. 29, no. 2, pp. 56–65, 2022.

[25] Z. Liu, H. Yu, G. Li, Q. Chen, Z. Ding, L. Feng, Z. Yao, and B. Hu, “Ensemble learning with speaker embeddings in multiple speech task stimuli for depression detection,” *Frontiers in Neuroscience*, vol. 17, p. 1141621, 2023.

[26] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, “Robust speech recognition via large-scale weak supervision,” in *International Conference on Machine Learning*. PMLR, 2023, pp. 28 492–28 518.

[27] Y. Yang, C. Fairbairn, and J. F. Cohn, “Detecting depression severity from vocal prosody,” *IEEE transactions on affective computing*, vol. 4, no. 2, pp. 142–150, 2012.