Building a virtual psychological counselor by integrating EEG emotion detection with large-scale NLP models

Peiran Lin[a, †] Liwen Yang[*, b, †]

[a]SDU-ANU joint science college, Shandong University (Weihai), No. 180, Wenhua West Road, Yiyuan Street, Gao District, Huancui District, Weihai City, Shandong Province; [b]School of Mathematics and Statistics, Shandong University (Weihai), No. 180, Wenhua West Road, Yiyuan Street, Gao District, Huancui District, Weihai City, Shandong Province

[†]These authors contributed equally.

ABSTRACT

Depression is a common psychological disorder, and the combination of psychological help and medication is the main approach to treat depression. However, traditional psychological counseling requires professional therapists to spend much time with patients, and the demand for therapists exceeds the supply. Therefore, this paper proposes a virtual psychotherapist architecture: combining EEG emotion detection and NLP large models to detect the patient's emotions in real time and adjust AI communication strategies to provide psychological counseling to patients. This paper introduces EEG emotion detection from several aspects, such as emotion models, EEG feature extraction, databases, emotion elicitation methods, and classification models. It also reviews the development of NLP large models in recent years. Then, it presents the theoretical methods for constructing a virtual psychotherapist using EEG emotion detection combined with NLP large models: the direct suggestion method and the prompt method. Finally, it points out the current problems and future development directions of virtual psychotherapists.

Keywords: Virtual AI psychological counseling, depression, Chat-GPT 3, EEG

1. INTRODUCTION

1.1 Depression

Depression is a prevalent mental health condition, affecting approximately 5.0% of adults worldwide, and female patients affected by depression accounted for a larger proportion than male patients. Severe cases can lead to suicide [1]. Depression can result in enduring feelings of sadness, a lack of interest, and difficulties with memory. Individuals who are depressed often face cognitive challenges and endure prolonged and intense emotional distress [2]. Psychological interventions, such as behavioral activation, cognitive-behavioral therapy, and interpersonal psychotherapy, are the primary treatment methods for patients with depression. Furthermore, pharmacological treatments, for example, selective serotonin reuptake inhibitors, and the tricyclic antidepressants, are also available options for managing depression.

There are multiple personality types of depression. The 9th edition of ICD has up to 8 possible classifications for patients who exhibit depressive emotions. In the United States, DSM-II classifies depression more according to different "emotional disorders". Wardenaar's team used latent characteristics to analyze 146 patients with major depression and obtained two personality types of depression: susceptible and recovered. In another longitudinal study involving patients with depression and acute coronary syndrome, 685 depressive patients were classified by non-hierarchical k-means cluster analysis, and two personality types, elastic and susceptible, were also found [3]. In addition, there are RUO personality type analysis methods jointly constructed by Asendorpf, Robinson and Caspi. In summary, a considerable number of patients with depression are introverted, sensitive and empathetic. Therefore, in reality, patients with depression, mainly introverted patients, not only need to receive routine drug treatment but also need more companionship and dialogue, especially professional and long-term psychological counseling services.

[*] yangliwen@mail.sdu.edu.cn

Third International Conference on Biological Engineering and Medical Science (ICBioMed2023), edited by Alan Wang, Proc. of SPIE Vol. 12924, 129241X

1.2 Existing ways to improve depression and our aims

At present, depression is mainly treated with artificial psychological counseling combined with drugs. For example, the selective 5-HT (5-hydroxytryptamine) reuptake inhibitor is represented by fluoxetine hydrochloride capsules [4]. However, drug therapy, as an in vitro intervention method, has the characteristics of slow onset and long process, and many patients with abnormal liver and kidney metabolism are limited when taking drugs. Traditional Chinese medicine also has acupuncture, aromatherapy and other methods to treat depression, but these methods often lack objective quantitative criteria and have different effects on patients in different situations [5]. In summary, although traditional drug therapy has very scientific methods and processes, these methods often cannot directly focus on the unique psychological state of depressed individuals. Artificial psychological counseling can focus on individual patients with depression, but the labor cost is relatively high. In addition, some depressed individual groups, mainly introverted depression, often encounter problems such as shyness and low communication efficiency when facing artificial psychological counseling. Virtual psychological counseling AI provides a low-cost auxiliary psychological counseling method, which can partially replace the functions of psychological counselors to complete the communication and solution of basic problems, which can greatly reduce the burden of psychological counselors.

We will review the development of EEG emotion detection and large natural language processing models. On this basis, we try to describe a vision of efficient virtual psychological AI that combines the above two technologies.

2. BACKGROUND OF EEG

2.1 Emotion models

Emotion is a complex state of mind, to recognize emotions, researchers must measure and construct models of emotional states. There are currently two different perspectives on the quantification and modeling of emotions. The first perspective is that emotional states are made up of discrete, limited basic emotions, such as: sadness, anger, surprise, disgust, curiosity and so on [6]. This perspective holds that all emotions can be made up of these basic emotions. Another view is to construct the emotional space as a continuous dimensional type of emotion quantification model. Such as the Valence-arousal dimensional emotion model. This model divides emotional states into two dimensions: valence and arousal [7]. The valence of emotion indicates whether it is positive/negative, while arousal suggests the level of intensity of the emotion.

2.2 Feature extraction

Typically, EEG characteristics are categorized into three main groups: features that are based on the signal's time-domain properties, features that are based on the signal's frequency-domain properties, and features that are based on the signal's time-frequency properties. These different types of features can be utilized for a variety of purposes, such as EEG analysis, EEG-based diagnosis, and EEG-based brain-computer interface (BCI) systems. As different emotions are associated with specific brain regions, space domain features are also utilized in EEG-based emotion detection to identify emotions.

Commonly used time-domain feature processing methods include event-related potential (ERP), power, higher order crossings, non-stationary index, and others. Among them, ERP is the most frequently used method for processing analysis. ERP refers to EEG voltage fluctuations triggered by discrete stimulatory events, which can reflect the process of cognitive processing [8]. ERP has extremely precise time resolution, allowing for the measurement of immediate responses to short stimuli. ERP waveforms exhibit peaks with different durations, amplitudes, and polarities over time, so they are usually measured from three aspects: latency, amplitude, and polarity.

The EEG signal can be analyzed in terms of specific event-related potentials (ERPs) that reflect distinct stages of cognitive processing. For instance, N100 and P100 are ERPs characterized by negative and positive voltage deflections occurring 100ms after the stimulus, respectively. Similarly, N200 and P200 have latencies of around 200ms, while P300 is a positive voltage deflection occurring approximately 250 to 500ms after the stimulus, reflecting processes related to stimulus evaluation or categorization. Another ERP, called slow cortical potentials (SCP), can last for several hundred milliseconds to a few seconds [9].

In addition to time domain features, EEG frequency domain features can reflect a lot of information about emotions. According to the frequency of EEG signal, it is commonly categorized as 5 frequency bands, namely the delta (about 0.5-4 Hz), theta (about 4-8 Hz), alpha (about 8-13 Hz), beta (about 13-30 Hz), and gamma (about 30-64 Hz) bands [10, 11]. Every frequency band can reflect different physiological states in humans. (see Table 1).

Table 1. EEG with different frequency bands and frequencies and corresponding physiological states.

Brain Wave	Frequency	Association
δ	About 0.5–4 Hz	Unconscious mind; Deep sleep
θ	About 4–8 Hz	Drowsiness; Subconscious mind
α	About 8–13 Hz	Relaxed status; Eyelid closure
β	About 13–30 Hz	Analytical thinking; Activities; Problem solving
γ	About 30–64 Hz	High-level Information Processing; Sensory processing; Motor control; Certain cognitive;
Noise	About >64 HZ	

EEG emotion detection involves the utilization of features from both time and frequency domains, which are combined in the time-frequency domain feature. In order to handle unstable signals and roughly estimate the onset time of emotions, a common approach is to divide the signal into several time windows and transform them into the frequency domain to extract a set of frequency domain characteristics. By using a sliding time window, different time periods can be analyzed, allowing for the simultaneous extraction of time-frequency domain information. This improves the ability to process unstable signals and estimate the time of emotional onset [11]. For time-frequency domain signal transformation, wavelet transforms (WT), the short-time Fourier transform (STFT), and wavelet packet transform (WPT) are commonly applied.

EEG signals are collected from multiple electrodes located throughout the cerebral cortex. EEG signals generated by the left and right hemispheres are related to emotional titer: negative emotions can activate the right frontal, temporal, and parietal lobes, whereas positive emotions can activate the left region. As a result, EEG spatial domain features are primarily classified as space-frequency domain features and electrode combination features. The choice of electrode position and features affects the effectiveness of EEG in detecting emotion. For example, R. Jenke et al, show that sophisticated signal processing techniques of feature extraction for HOS, HOC, and HHS outperform typically applied spectral power bands and reachers have a preference for locations in the parietal and centro-parietal lobes [12].

2.3 Dataset introduction

The DREAMER dataset [13], which consists of ratings given by participants for film potency, arousal, and control, are then used to calculate the corresponding values for emotions such as positive/negative or arousal/control. In this study, ECG data were gotten from 23 participants, while they watched movies using the Emotiv EPOC system at a sampling rate of 128 Hz. The final 60 seconds of every signal were extracted and preprocessed using EEGLAB in MATLAB.

The DEAP dataset [14] is a multi-channel dataset applied to investigate the emotional states of people. The dataset contains psychological scales, physiological signals, and facial expressions, among other things. 32 subjects were monitored for their physiological responses and EEG signal while they viewed a 40-minute music video. The dataset comprises recordings from 32 EEG channels, 2 EOG channels, and 2 EMG channels. Before each signal is recorded, a 3-second silence period is set. This dataset can be used to investigate physiological signals in multiple dimensions.

Shanghai Jiao Tong University released the SEED database [15], and Chinese movie clips lasting about 4 minutes were used to arouse three emotions of neutral, negative, and positive potency. EEG data from 15 subjects were collected at a sampling rate of 1000 Hz using ESI NeuroScan System. Each participant completed three experiments at different times, each time watching 15 movie clips, for a total of 45 trials. EEG signal preprocessing includes signal downsampling to 200Hz, removal of ocular and EMG noise, and the use of a 0.3 50 Hz bandpass filter. The short-time Fourier transform was performed and divide the time-frequency domain characteristics into five frequency bands: delta (about 1-4 Hz), theta (about 4-8 Hz), alpha (about 8-12 Hz), beta (about 13-30 Hz), gamma (about 31-45 Hz).

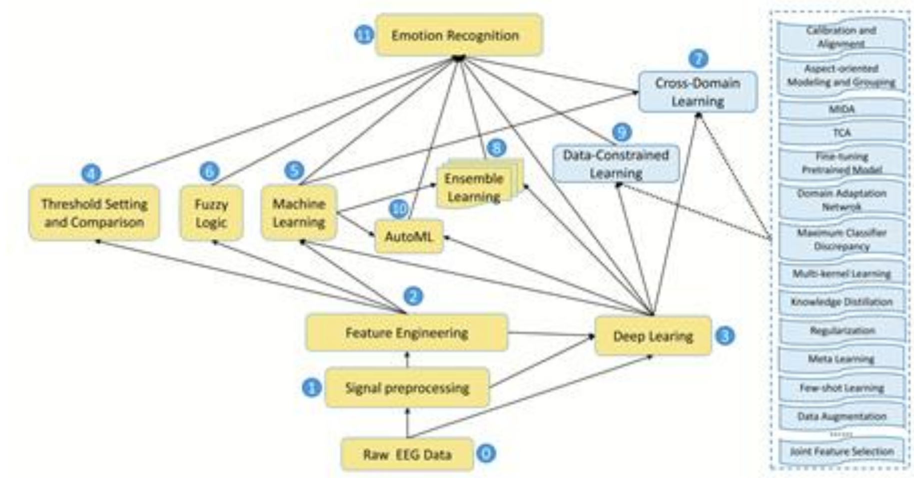

Figure 1. Flowchart of EEG emotion detection processing [11].

2.4 Classification

In the selection of EEG sentiment classification algorithms, some relatively simple machine learning algorithms were often used, such as, multi-layer perception back propagation (MLPBP), linear discriminant analysis (LDA), and so on [8]. But deep learning methods that have emerged in recent years adopted a hierarchical structure, which first transforms the feature representation of the original space to the new space, and then performs subsequent classification. Compared with artificially constructed features, the features learned by using massive data can better reflect the intrinsic nature of data, overcome the problem of feature redundancy, and improve the intelligence and universality of emotion recognition (figure 1).

3. LARGE-SCALE NLP MODELS

Natural language processing (NLP) is a Universally applied technique, which is used for large datasets analysis, and it involves the use of computer-based algorithms to process, enhance, and convert natural language into a form that can be used for computation. NLP algorithms are capable of performing various tasks, including but not limited to syntactic processing (such as sentence detection), information extraction, and semantic analysis, all of which help to capture the meaning of text [16].

LNLP (large natural language processing) is a type of natural language processing model that is characterized by its use of large datasets and strong language processing capabilities. In the last few years, there have been many L-NLP models released, such as the BERT model released by Google. This NLP pre-training model adopts the advanced Transformer neural network architecture at that time for language understanding and is suitable for any task such as speech recognition (ASR), text-to-speech (TTS), and sequence-to-sequence (Sequence To Sequence) [17]; Google then released the ALBERT model in 2020 to address the problem of slower training times caused by the increasing size of models. Essentially, ALBERT is a streamlined version of the BERT model that significantly improves the efficiency and performance of the model on various NLP tasks [18].

Large Natural Language Processing (LNLP) models have become increasingly prevalent in recent years, with companies such as Google, Microsoft, and Alibaba releasing various models with differing performance and capabilities. Some notable models include BERT [17], a pre-training model using a Transformer neural network architecture for language understanding, and its more efficient and streamlined version ALBERT [18]. XLNet, a model released in 2019 by CMU and Google Brain, builds upon BERT's architecture with a universal auto-regressive pre-training method. T-5, released by Google, is capable of transforming all NLP tasks into text tasks. ELECTRA, inspired by adversarial networks, is another model that utilizes a generator-discriminator approach.[19] Microsoft's DeBERTa model, released in early 2021, has already undergone three iterations [20]. Other models like Alibaba's StructBERT have also been proposed [21].

In February 2019, OpenAI released the open-source GPT-2 model, which also adopts the generative pre-training Transformer. The model can achieve human-like abilities in text generation, summarization, translation, and QA. GPT-2 is a "direct enlargement" of the GPT model released in 2018, marking the clear development direction of the large-scale natural language processing models in the GPT line. Not long after, in 2020, Open AI released GPT-3. The biggest difference between GPT-3 and its predecessor is that it can continue generating text based on the given initial text, and the content is almost at the level of human writing. GPT-3 has a context of 2048 tokens and 175 billion parameters, which together give it powerful language processing capabilities. The Chat-GPT conversational model that is currently popular was trained based on the GPT-3 model.

ChatGPT has been utilized in many fields, for example, natural language processing, dialogue systems, and language translation, to produce simplified and context-based texts. These texts are easier to comprehend and understand for users. Although Chat-GPT has not been released for a long time, many studies have proved that Chat-GPT has a broad application space, such as for medical research: Chat-GPT automatically generates a series of potential applications of Chat-GPT in liver disease research according to several keywords including 'liver disease' and 'chatGPT', including but not limited to generating patient narratives, relevant medical literature reviews, and generating research hypotheses. In addition, chat GPT can also be used in Clinical Toxicology. Chat-GPT also shows certain potential in basic scientific research. For example, some scholars have asked about some difficult and unpopular physical and chemical problems of Chat-GPT, such as PFAS (perfluoroalkyl and poly-fluoroalkyl substances). Chat-GPT can give some effective information, although the correctness is not as good as the common problems in life [22]. In mental health treatment, ChatGPT has also received extensive attention. For example, the ' Asian Journal of Psychiatry ' has developed some new policies related to ChatGPT, legalizing and prohibiting the function of some ChatGPT. In summary, the L-NLP language processing model represented by ChatGPT has great application space in many fields. In summary, the application of ChatGPT combined with other technologies has great possibilities and broad prospects. This article will briefly discuss the implementation of EEG and chat GPT as the representative of the LNLP model to build a brief method for psychological counselors.

4. BUILDING A VIRTUAL PSYCHOLOGICAL COUNSELOR

4.1 Related work

Many researchers have previously developed intelligent chat robot systems for psychological counseling. For example, Junjie Yin proposed an innovative, fully generated session system based on sequence pair sequence, which is used to diagnose negative emotions and prevent depression through positive suggestive responses. The robot is integrated into the online platform. After a month of user research, it shows that it is more competent for psychological counseling than other public chat robots [23]. Aislyn's team has created a virtual chatbot using their expertise in professional psychological counseling, which is designed to provide immersive virtual reality counseling sessions to users in need of mental health support. The aim is to provide psychological assistance and stress management counseling to students facing difficulties in various aspects of school, including academics, family, and social issues, at any time and from any location. At the same time, the robot can intelligently identify life-threatening reactions and inform clinical psychologists to intervene immediately to ensure professional help for students [24]. The Amy team developed a chat robot VRECC system that can perform emotional dialogue analysis centered on empathy. The verification experiment proves that although there is still a gap between AI consultants and real consultants, VRECC has shown hope of being put into application.

4.2 How to build a virtual psychological counselor

To better understand and assist depressed patients with emotional issues such as depression or anxiety, it is crucial to accurately identify their emotional state. EEG emotion detection technology can detect the emotional state of customers through the electrical signals generated by the brain using electrodes placed on the scalp to accurately collect brain signals. The collected EEG signal data needs to be processed for noise reduction and feature extraction, and then analyzed and classified through a classification model to finally output the emotional state of the patient in the format of a discrete emotion model or a continuous emotion model, for real-time monitoring of the patient's emotional state.

Currently, there are two main methods for inputting EEG data and influencing NLP large models (using ChatGPT as an example): one is to directly suggest ChatGPT of the emotional state of the people they are conversing with during the conversation. Due to ChatGPT's strong logical abilities, you can add hints like "Assuming you are a virtual psychological counselor, the person you are currently talking to is..." and "How would you comfort/encourage them to improve their mental state?" to ChatGPT during the conversation, and ChatGPT will smartly adjust its conversation strategy accordingly. Another method is called "Prompt Engineering", which guides ChatGPT to provide the desired response by designing appropriate prompts. By constructing a small sample of virtual psychological counseling Q&A, ChatGPT can perform few-shot learning and achieve very good conversation effects.

In addition to EEG technology, natural language processing (NLP) large model technology can also play an important role in analyzing customer language expressions and emotional experiences. For example, NLP can identify the emotional vocabulary used by depressed patients and evaluate the emotions conveyed in their text. By combining the output of EEG and NLP, we can establish a more comprehensive customer emotional state model, providing a better understanding of the depressed patient's psychological condition. This can help provide better service and support to customers, helping them manage their emotions and recover their mental health.

4.3 Conclusion

This study proposes an architecture for an AI virtual psychological counselor that combines EEG emotion detection with NLP large models to address the scarcity of professional psychological counselors for the treatment of depression. Firstly, the research reviews the research methods for EEG emotion detection, introduces the selection of emotion models, feature extraction of EEG raw data, and classification models, and presents some commonly used EEG datasets and recent related research. Then, recent developments in NLP large models are reviewed. Combining with the research of other scholars in AI psychological counseling, the study presents two feasible methods for combining EEG emotion detection with NLP large models: direct suggestion or prompt method. Both methods use EEG technology to detect the client's emotional state, while NLP large model technology analyzes the language expression and emotional experience of depressed patients to establish a client emotional state model. This method can provide more convenient, private, safe, and timely psychological counseling services for depressed patients, reduce interference from human factors during the treatment process, and improve the efficiency and quality of psychological counseling services.

It should be noted that although the combination of EEG emotion detection and NLP large models provides a feasible solution for building virtual psychological counselors, there are still some shortcomings. For example, virtual psychological counselors lack the personal touch of human counselors and cannot provide face-to-face communication and interaction.; For some patients with severe conditions, using AI is risky, and they still need the help of professional psychologists.

Therefore, although virtual psychological counselors provide new ideas and methods for psychological treatment, technology and counseling services still need to be further improved to improve their reliability and effectiveness. However, the prospects for AI psychological counselors are still very optimistic: with the continuous development and improvement of NLP technology, it can be foreseen that NLP large-scale models specifically for psychological counseling services will appear. These models will integrate more psychological theories, emotional cognition, and applied emotional science knowledge, and the virtual psychological counselor system will become more humane, intelligent, and precise to better meet the needs and expectations of depressed patients.

REFERENCES

[1] Johnson, D., Dupuis, G, Piche, J., Clayborne, Z. and Colman, I. "Adult mental health outcomes of adolescent depression: A systematic review," Depress Anxiety 35(8),700-716 (2018).

[2] Cai, H., Han, J., Chen, Y., Sha, X., Wang, Z., Hu, B., Yang, J., Feng, L., Ding, Z., Chen, Y. and Gutknecht, J. "A Pervasive Approach to EEG-Based Depression Detection," Complex 13 1-13 (2018).

[3] Kim, S. Y., Stewart, R., Bae, K. Y., Kim, S. W., Shin I. S., Hong, Y. J., Ahn, Y., Jeong, M. H., Yoon, J. S. and Kim, J. M. "Influences of the Big Five personality traits on the treatment response and longitudinal course of depression in patients with acute coronary syndrome: A randomised controlled trial," J Affect Disord 203 38-45 (2016).

[4] Chen, Y. and Yuan, Y. G. "Research progress in the treatment of depression with monome of Chinese herb, drug pairs, compound prescriptions and Chinese patent drugs," Chinese Journal of Clinical Pharmacology and Therapeutics 26(5) 586-593 (2021).

[5] Wu, P. and Zhang, L. P. "Overview of TCM treatment of depression in recent ten years," World Journal of Integrated Chinese and Western Medicine 11 999-1001 (2012).

[6] Ekman, Paul., "Basic Emotions," In Handbook of Cognition and Emotion.,. Hoboken, NJ, US: John Wiley & Sons Ltd, 45-60(1999).

[7] Pekrun, R., Goetz, T. and Perry, R. P. "Chievement Emotions Questionnaire," University of Munich: Munich, Germany (2005).

[8] Hruby, T. and Marsalek, P., "Event-related potentials--the P3 wave," Acta Neurobiol Exp 63(1) 55-63 (2003)

[9] Alarcão, S. M. and Fonseca, M. J. "Emotions Recognition Using EEG Signals: A Survey," IEEE Transactions on Affective Computing 10(3) 374-393 (2019)
[10] Suhaimi, N. S., Mountstephens, J., and Teo, J. "EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities," Comput Intell Neurosci 16 1-18 (2020).
[11] Xiang, L., Zhang, Y. Z. Tiwari, P., Song, D., Hu, B., Yang, M, Zhao, Z. G., Kumar, N. and Marttinen, P., "EEG Based Emotion Recognition: A Tutorial and Review," ACM Comput. Surv. 55(4), 1-57 (2023)
[12] Jenke, R., Peer, A., and Buss, M. "Feature Extraction and Selection for Emotion Recognition from EEG," IEEE Transactions on Affective Computing, 5(3) 327-339 (2014)
[13] Katsigiannis, S. and Ramzan, N. "DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices," IEEE Journal of Biomedical and Health Informatics 22(1) 98-107 (2018).
[14] Verma, Gyanendra K. and Uma S. T., "Multimodal fusion framework: A multiresolution approach for emotion classification and recognition from physiological signals," NeuroImage 102, 162-172 (2014).
[15] Zheng, W. L., Lu and B. L., "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks, "IEEE Transactions on Autonomous Mental Development, 7, 162-175 (2015).
[16] Fleuren, W. W. and Alkema W., "Application of text mining in the biomedical domain," Methods 74,97-106 (2015).
[17] Devlin, Jacob, Chang, M. W., Kenton, L. and Kristina, T., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, "In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1 4171–4186 (2019).
[18] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma and P., Soricut, R., "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," ArXiv 1-17 (2019).
[19] Clark, K., Luong, M., Le, Q. V., and Manning, C. D. "ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators, " ArXiv 1-18 (2020)
[20] He, P., Liu, X., Gao, J., and Chen, W. "DeBERTa: Decoding-enhanced BERT with Disentangled Attention, " ArXiv 1-23 (2020).
[21] Wang, W., Bi, B., Yan, M., Wu, C., Bao, Z., Peng, L. and Si, L. "StructBERT: Incorporating Language Structures into Pre-training for Deep Language Understanding,"ArXiv, 1-10 (2019).
[22] Zhu, J. J., Jiang, J., Yang, M. and Ren, Z. J. "ChatGPT and Environmental Research. Environ Sci Technol," Environmental science and technology 1-21 (2023).
[23] Yin, J., Chen, Z., Zhou, K., and Yu, C., "A Deep Learning Based Chatbot for Campus Psychological Therapy," ArXiv 1-31 (2019).
[24] Lin, A.P., Trappey, C.V., Luan, C., Trappey, A.J., and Tu, K. L., "A Test Platform for Managing School Stress Using a Virtual Reality Group Chatbot Counseling System," Applied Sciences 11 1-24 (2021).