Grenze International Journal of Engineering and Technology, Jan Issue

Augmented Humanoid Counselor for Mental Health using LSTM Neural Network

Sonali Sharma[1], Vinayak Unde[2], Sakshi Jadhav[3], Dev Prajapati[4] and Deepshikha Chaturvedi[5]
[1-5]Department of Computer Engineering, Shah and Anchor Kutchhi Engineering College, Mumbai, India
Email: {sonali.sharma_19, vinayak.unde_19, sakshi.jadhav_19, dev.prajapati_19, deepshikha.chaturvedi}@sakec.ac.in

***Abstract*— In every period of life, from childhood and adolescence to maturity, mental health is vital. Good mental health allows individuals to enhance their overall quality of life. However, due to a lack of willingness, many people neither share their mental health issues with their close ones nor seek professional help. To address this issue, we have developed a lifelike counselor who can offer effective and compassionate mental health conversations by uniting natural language processing and augmented reality. The Augmented Humanoid Counselor, a virtual counselor equipped with a conversational model, is designed to give a human touch by assisting individuals to manage their mental health. We have designed various neural network-based conversational models using Long Short-Term Memory (LSTM), Bi-directional LSTM, and LSTM with an attention layer. The LSTM model, enhanced with an attention layer, enables itself to concentrate on the input sequence's most instructive segments while disregarding the less important ones, and thus has demonstrated its ability to provide more interactive, relatable responses compared to other models. This conversational model is used by our Augmented Humanoid Counselor to respond to the user in a sympathetic way to give them support for their mental health and make them feel better.**

***Index Terms*— augmented reality, natural language processing, neural network, deep learning, humanoid, counselor, mental health, LSTM, BiLSTM and attention layer.**

I. INTRODUCTION

Mental health is one of the most crucial components of both individual and community well-being. It can affect not only one's way of living life but also one's connections and physical health. However, this also works vice versa. Interpersonal relationships, medical issues, and aspects of people's lives can all affect mental health. A person's ability to enjoy life, lead a healthy lifestyle, and be more productive in both their personal and professional lives can be maintained by taking good care of their mental health. Recognizing the significance of prioritizing mental health and taking actions to maintain and improve it is essential.

In this research work, we have created an augmented humanoid counselor with the intention of offering a resolution to people facing mental health issues. An augmented humanoid counselor is a virtual human that acts as a mental health counselor in augmented reality, an accessible medium that offers a seamless experience. It offers an environment for people to share their feelings, emotions, and thoughts and receive moral support whenever they require it. In addition to providing a safe and non-judgmental space to talk about mental health, augmented reality has been incorporated into our application to enhance the user's experience and understanding of their mental health and to create an immersive and interactive experience for users.

Our augmented reality and natural language processing-based application using the LSTM neural network is

Grenze ID: 01.GIJET.10.1.340_1

developed to provide a unique and engaging platform for people to improve their mental health, gain insights into their own mental health, and develop skills to manage their mental health challenges. As compared to traditional face-to-face counseling or even chatbots, people will feel less stigmatized by seeking help from an augmented counselor, which could encourage more people to get mental health care; hence, we believe that our augmented humanoid counselor can make a significant contribution to mental health improvement.

II. Related Work

With so many individuals suffering from different mental health issues, the mental health crisis has grown to be a global concern. The aim of this literature survey is to examine the existing solutions available in the sphere of mental health research.

A chatbot application that uses natural language processing to comprehend the user's state and sensitively observe ongoing emotional changes in the user has been introduced using emotion recognition techniques to generate personalized responses in the paper [1]. According to the paper [7], many factors contribute to the high rate of mental health problems among students; hence, a chatbot has been developed using emotion recognition, especially designed for students, to promote their mental health and resolve the students' mental health issues at the right time.

The paper [9] proposes that conversational systems based on artificial intelligence can be effective in suicide prevention and cognitive-behavioral-therapy (CBT) if there are applications that can identify suicide intent on various platforms, such as social media platforms, browsers, etc. Work has been done to detect and prevent suicide using deep learning algorithms.

Deep learning and machine learning techniques have proven to be the most effective approaches for depression detection. A convolutional neural network has been used on audio data and the transformer model BERT on textual data to detect whether a person has depression or not [11].

The study of the paper [12] adds to the body of knowledge by outlining specific, well-organized areas for future research in the field of artificial intelligence and mental health and describing the framework for mental health chatbots in public organizations. The papers [13-14] show that the use of DialoGPT, an advanced transformer model, makes the chatbot more accurate in handling conversations related to mental health issues as the generated responses appear to be more real.

Interactive digital representations of real humans, known as "virtual humans" are capable of recognizing actual humans and giving appropriate verbal and behavioral responses [2]. Along with human-to-virtual human interactions, human-to-virtual animals' interactions are the potential benefit of using augmented and virtual reality [3].

Virtual humans have the capability to act as breathing coaches [8]. Multiple factors are considered while establishing trust between the user and intelligent adaptive agents, such as augmented and virtual agents, which include expression and lip synchronization with the speech and behavior of the agents [4].

The facial expression of a virtual and augmented human should be dynamic to make it realistic [10]. A spoken interaction between two avatars in augmented reality has been presented in the paper [5]. The paper [6] presents a virtual assistant for mental health using a Naive Bayes classifier and neural network to analyze sentiment.

LSTM, a type of recurrent neural network, excels at preserving and utilizing long-term information. Its ability to capture and process sequential patterns makes it a popular choice for various tasks, such as natural language processing [15]. Bidirectional LSTM (BiLSTM) has the capability to capture contextual information from past and future contexts simultaneously. BiLSTM has proven to be effective in tasks where understanding the entire sequence, including its surrounding context, is crucial [16].

Chatbots for mental health cannot offer real human empathy or emotional support as they don't deliver realistic experiences. Users can become completely disengaged from mental health support if they experience unsatisfactory interactions with chatbots. The use of augmented reality and natural language processing makes mental health support more motivational and engaging due to its ability to construct immersive and interactive environments. Compared to text-based chatbots, users can have a more human-like experience with an augmented counselor because they may see and interact with a relatable or humanoid counselor.

III. Problem Statement

A lot of individuals struggle with mental health conditions like stress, anxiety, and depression. However, people do not prioritize their mental health when they are in bad mental health. Many people are unwilling to share it with their loved ones or seek expert assistance. This could have serious consequences that lower one's productivity.

Therefore, in order to address this problem, we have created an augmented reality and natural language processing-based application that can assist those who are struggling with mental health issues through friendly and empathetic conversations.

IV. PROPOSED METHODOLOGY

A. System Architecture

The system has been designed by uniting augmented reality and natural language processing, as shown in Fig. 1. Augmented reality is used to create a virtual humanoid counselor that appears to be a human counselor. Natural Language Processing is used to develop a conversation model that is used by our Augmented Humanoid Counselor to give responses to the user. The user needs to give the camera and microphone access to start with the application. Location tracking and flat surface detection are used to render the AR human at the appropriate place. The user's question or statement is taken as input text. The input text is given to the conversational model; the conversational model in the backend sends appropriate responses to the user, which are then delivered as speech. In this way, the user gets the feeling that our augmented counselor is talking with them.

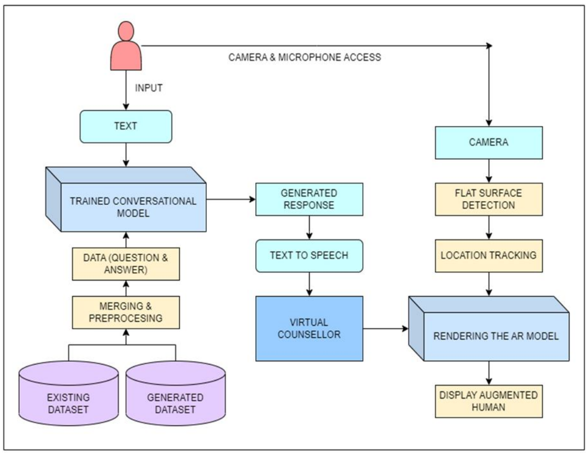

Figure 1. Proposed System Architecture

B. Questionnaire

With the help of counselors, a questionnaire has been generated. It is a type of evaluation technique used to learn more about a person's mental health and wellness. A number of questions or statements about many areas of mental health, such as mood, emotions, behavior, and cognition, are included in it. A questionnaire is designed to assist in better understanding a patient's mental health condition and spotting any possible problems. The questions are related to symptoms and behaviors in accordance with mental health issues. These questions are presented to the users one at a time, their responses are recorded, and a score is created based on their responses. Using a mathematical model, the score assesses whether or not a person requires support with mental health counseling.

C. Conversational Model

The objective is to develop a model capable of comprehending human language, understanding its meaning, and generating appropriate responses that make sense in the context of the discussion.

1. *Data Collection***:** The dataset consists of dialogues in the form of questions and answers. We have used the empathic dialogue dataset, which comprises open-world dialogues. We extracted questions and answers from the dialogue, which accounts for more than 64,000 open-world questions and answers. Additionally, various dialogue interactions between counselors and mental health patients were used. The dialogues were converted into questions and answers, which accounts for more than 20,000 question and answer creations. Furthermore, we incorporated a dataset from various mental health organizations FAQs, that encompasses frequently asked questions about various mental health issues; this accounts for more than 2500 question and answer creations. All these were merged, which led to the formation of a huge database that accounts for more than 2,50,000 questions and answers.
2. *Data Pre-processing***:** All the data gathered is presented in question-and-answer format. The collected data is processed to eliminate noise and clean up the text.

3. *Tokenization:* In the case of LSTM models, tokenization, a crucial step in natural language processing (NLP), is especially significant. Tokenization is used in LSTM models to turn unprocessed text data into numerical vectors that can be given to the model.
4. *Model Building:* By specifying architecture and configuring hyperparameters, the LSTM, Bi-LSTM, and LSTM with Attention Layer models are developed. The model is built to accept input sequences and produce predicted word or phrase sequences.
5. *Model Training***:** Prepared data is utilized to train the LSTM models. To reduce the loss function, the model's parameters are modified as we fed the input sequences into it.
6. *Hyper parameters Tuning***:** At each stage of model training, the rate of learning, batch size, and epoch numbers are adjusted to enhance performance.

D. Model Architecture

*Seq2seq Encoder Decoder Model based on LSTM***:** The encoder block of our model incorporates LSTM cells to sequentially process each token in the input sequence, as shown in Fig. 2. Its objective is to extract all the relevant data from the input sequence and condense it into a fixed-length vector known as a "context vector." Once the encoder has processed all the tokens, it transfers this context vector to the decoder.

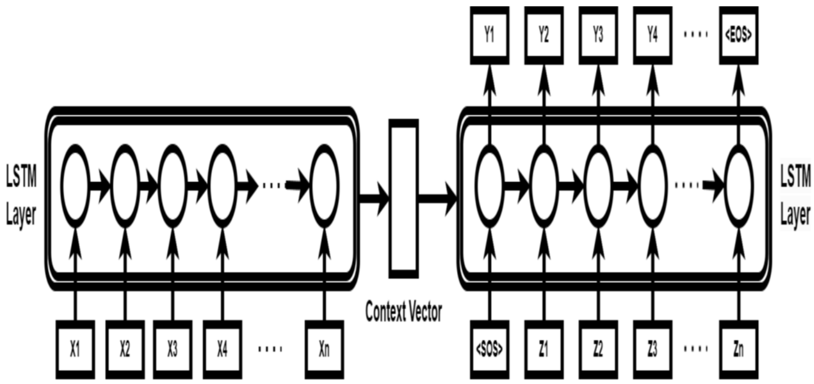

Figure 2. Seq2seq Encoder Decoder Model based on LSTM

The context vector is specifically designed to encompass the complete meaning of the input sequence, enabling the decoder to generate accurate predictions when constructing the target sequence. In contrast, the decoder consists of LSTM cells as well. It initially receives the context vector generated by the encoder as input and proceeds to predict the tar-get sequence token-by-token. The process is as depicted below:

Step 1: At the initial time step, the decoder takes a special symbol "<SOS>" as input to signify the beginning of the output sequence. Using this input along with the context vector, the decoder generates the output for the first time.

Step 2: During the second time step, the previous output becomes the input for the ongoing step. The output produced in this time step represents the second word in the target sequence.

This pattern continues for subsequent time steps, where the output of each step is used as input for the next step. The process persists until the decoder generates the special symbol "<EOS>", which indicates the end of the output sequence.

Seq2seq Encoder Decoder Model based on Bi- LSTM: The encoder block in our model consists of three Bi-LSTM layers as shown in the Fig. 3. The input sequence in the first layer is processed in both forward and backward directions, which captures data from both future and past time steps. This layer aims to encapsulate all the relevant in-formation and passes its output data as input to the second Bi-LSTM layer. Similarly, the second layer output serves as input to the third Bi-LSTM layer. This sequential processing enables the encoder to have access to data from earlier and future time steps at the current time step. Ultimately, the last output of Bi-LSTM layer serves as the context vector representation for the input sequence, which is then passed on to the decoder. The context vector is specifically designed to grab the entire understanding of the input sequence and offers a more accurate and comprehensive representation compared to the one generated by a standard LSTM. The decoder component consists of LSTM cells. It takes the context vector generated by the encoder as its initial input and proceeds to predict the target sequence token by token. The process follows the same framework as the decoder block of the LSTM-based sequence-to-sequence (seq2seq) model consisting of encoder-decoder model. The output sequence generated by this architecture is more precise and well-structured compared to the output produced by a standard LSTM model.

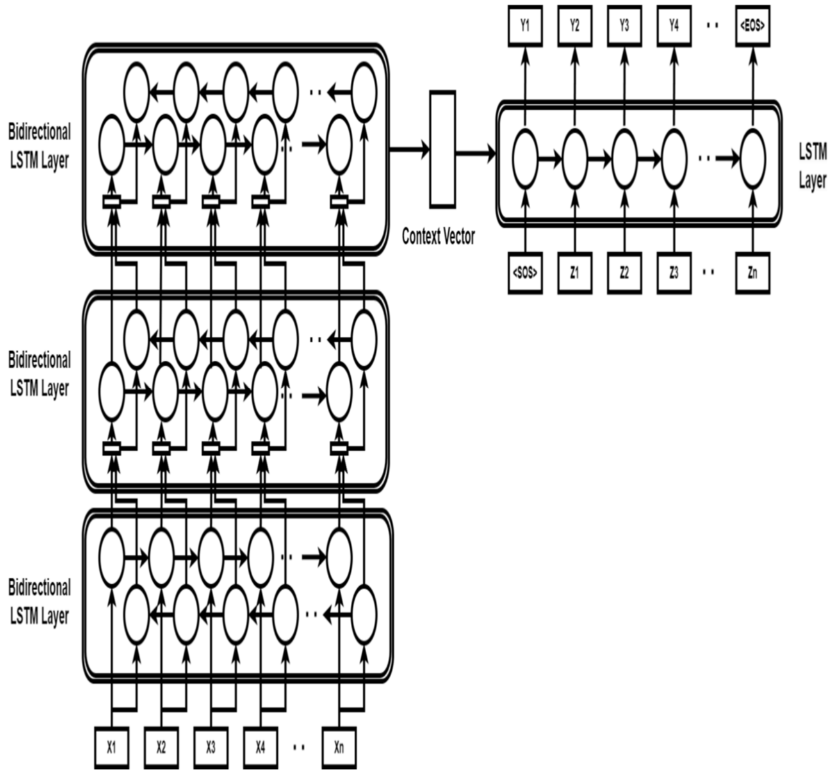

Figure 3. Seq2seq Encoder Decoder Model based on Bi- LSTM

Seq2seq Encoder Decoder Model consisting of LSTM and Attention Mechanism: The LSTM and Bi-LSTM-based model described earlier faces a significant limitation when working with long sequences. As the model processes more tokens in the input sequence, important information from the initial tokens tends to be diluted or lost. This is due to the context vector, which is responsible for encoding the input sequence's information to a fixed-length vector. However, this vector's limited capacity makes it challenging to handle longer sentences effectively. The context vector tends to prioritize recent important tokens over earlier ones, leading to a loss of information. To overcome this challenge, an attention mechanism was introduced.

The attention mechanism addresses the issue by allowing model to moderately focus on input sequence for generating the output sequence. It takes into account the outputs of the LSTM cells in the last encoder layer at each time step. There are two types of attention mechanisms: global attention, which considers all hid-den states/outputs of encoder layer at different time steps, and local attention, which focuses only on the most recent hidden states/outputs of the encoder layer.

In our model, we have implemented a global attention mechanism to ensure that attention is given to the entire input sequence by considering all time step out-puts as shown in the Figure. By incorporating the input sequence, context vector, and attention mechanism, we aim to generate more accurate, contextually relevant, and well-structured responses. Moving on to the decoder block, it consists of LSTM cells. The initial input to decoder includes the context vector generated by encoder, along with the information from the attention layer. The decoder then predicts the target sequence based on token using the following process: At every particular time step, the decoder takes a special symbol "<SOS>" as input, and mark the start of the output sequence. Utilizing current input, the con-text vector, and the attention layer information, the decoder generates the output for the ongoing time step. The result from the previous time step and the attention layer data are also considered at each time step. This process functions iteratively, with each step result tends to become the input for the next step. The iteration persists until the decoder generates the special symbol "<EOS>", signifying the end of the output sequence.

Therefore, we have employed a Seq2seq Encoder-Decoder model consisting of LSTM and Attention Mechanism to address the challenge of generating more contextually accurate and well-framed responses for user queries or inputs.

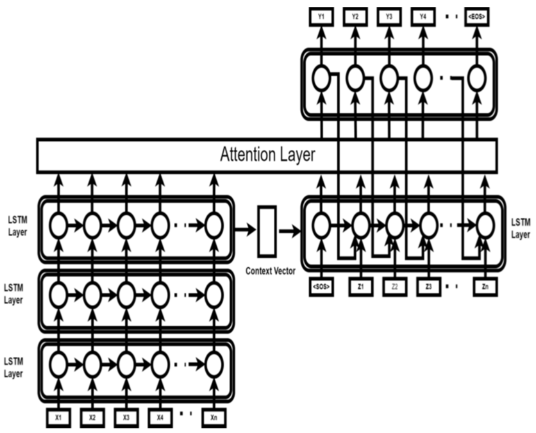

Figure 4. Seq2seq Encoder Decoder Model consisting of LSTM and Attention Mechanism

E. Augmented Humanoid Counselor

Augmented reality technology combines real-time computer-generated graphics and objects with one's view of the natural world. The purpose of augmented reality (AR) is to overlay a virtual humanoid counselor onto the physical environment.

i. *User Interface*: In augmented reality (AR), user interfaces are created to give users a way to interact with virtual items and information superimposed over the real-world environment. In our application, users can interact through text.

ii. *Camera and Microphone Access*: To provide users with more immersive experiences, access to the camera and microphone is a crucial element in AR applications. Camera access is needed to detect the real world. Microphone access is used by the augmented counselor to speak.

iii. *Flat Surface Detection*: Floor is detected by a plane detection mechanism to position virtual items and link them to actual surfaces. This is accomplished by looking at the camera feed and locating regions of the image that seem to be planar surfaces.

iv. *Location Tracking*: It makes it possible to attach virtual objects to specific locations in the real world, resulting in a more realistic AR experience. The location of the floor surface is captured with respect to the user for placing the virtual counselor in the physical world.

v. *Rendering AR Human*: The process of showing computer-generated objects in the physical-world environment is known as rendering in augmented reality (AR). In order to show AR humans accurately in the user's field of view, our application uses real-time rendering. To avoid lag or other performance problems, the virtual counselor is rendered swiftly and effectively.

vi. *Text to Speech*: Text to speech plugin is added to convert the text into speech and is synchronized with the AR model.

Figure 5. Augmented Humanoid Counselor

V. RESULT AND ANALYSIS

Our conversational model was developed using Long Short-Term Memory (LSTM), which excels at preserving and utilizing long-term information; Bi-Directional LSTM, which captures contextual information from past and future contexts simultaneously; and LSTM with Attention Layer architectural models.

The LSTM model's accuracy has shown notable progress over the course of training. Fig. 6 shows that starting at 88% accuracy after 300 epochs, it increased to 97% after 400 epochs. These results demonstrate that the model has steadily improved its ability to make predictions accurately and has developed a strong understanding of the hidden patterns in the training data.

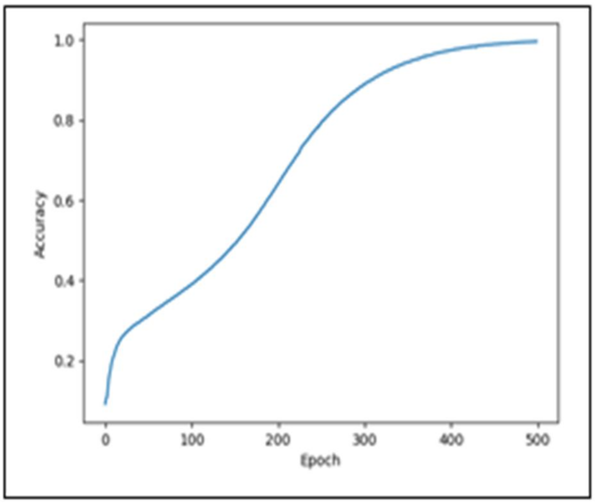

Figure 6. LSTM based Seq2seq Encoder Decoder Model’s Accuracy

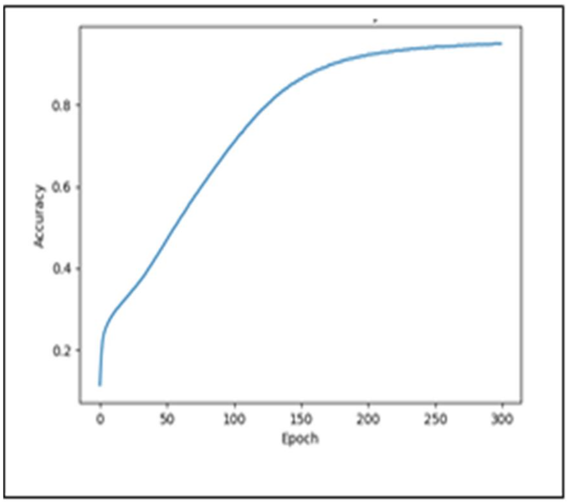

Figure 7. BiLSTM Seq2Seq Encoder Decoder Model’s Accuracy

The bidirectional LSTM model, trained for 300 epochs, has demonstrated impressive performance with an accuracy of 95%, as shown in Fig. 7. The BiLSTM architecture permits the model to reflect dependencies in both past and future contexts by incorporating both forward and backward information flow. This enables the model to effectively understand persistent dependencies and relationships within data that are arranged sequentially. The achieved accuracy of 95% indicates that the BiLSTM model has successfully captured important contexts from the training data.

Fig. 8. The LSTM with Attention model, trained for 75 epochs, has demonstrated remarkable performance with an accuracy of 97%. The LSTM architecture with an attention mechanism enables the model to concentrate only on the input sequence's most instructive segments while disregarding the less important ones. Overall, the LSTM with Attention model's high accuracy and ability to selectively attend to informative input features make it a promising approach for sequence prediction tasks.

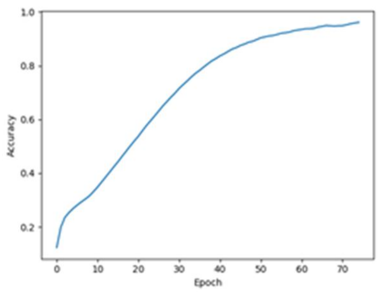

Figure 8. Accuracy of Seq2Seq Encoder Decoder LSTM Model with Attention Layer

The accuracy of each conversational model at a given number of epochs is displayed in the above table. As the model is modified from LSTM to Bi-LSTM and Bi-LSTM to LSTM with Attention Layer, the number of epochs required to obtain the desired accuracy is reduced. Hence, the AR model is integrated with the Seq2Seq Encoder Decoder LSTM Model with Attention Layer, which gives the best result in friendly conversations.

TABLE I. ACCURACY TABLE

Model	Number of Epochs	Accuracy
LSTM	400	97.2%
Bi-LSTM	300	95%
LSTM with attention layer	75	97%

VI. CONCLUSIONS

In this research project, we have successfully developed an interactive and immersive experience for individuals dealing with mental health issues using natural language processing and augmented reality technology. We collected data from various sources, performed data cleaning and pre-processing, and trained various neural network conversational models, which included an LSTM model, a BiLSTM model, and an LSTM with Attention Layer model for text generation. Our findings indicated that LSTM with Attention Layer gives the desired accurate results for conducting mental health-related conversations, as it has displayed remarkable performance after being trained for 75 epochs with an accuracy of 97%. The model was integrated with an augmented human designed to act as a counselor for users, and text and voice interaction were enabled using text-to-speech modules. The application was deployed on AR-supported mobile devices for easy accessibility. Overall, the project demonstrated the potential of NLP and AR technology in addressing mental health issues and providing individuals with a supportive and interactive platform to improve their mental wellbeing.

REFERENCES

[1] Dongkeon, LeeKyo-Joong, OhHo-Jin Choi, "The chatbot feels you - a counseling service using emotional response generation", IEEE International Conference on Big Data and Smart Computing (BigComp), April 2017

[2] Arno Hartholt, Ed Fast Adam, Reilly Wendy Whitcup Matt, Liewer Sharon Mozgai, "Ubiquitous Virtual Humans: A Multi-platform Framework for Embodied AI Agents in XR", IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), December 2019

[3] Nahal Norouzi, "Augmented Reality Animals: Are They Our Future Companions", 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), March 2020

[4] Ningyuan Sun, Jean Botev, "Intelligent Adaptive Agents and Trust in Virtual and Augmented Reality", IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), November 2020

[5] Shinji Miyake, Akinori Ito, "A spoken dialogue system using virtual conversational agent with augmented reality", Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference, December 2020

[6] Sreevidya Iyer, Dhanashree Shetty, Purva Badgujar, Anjali Nambiar, Kajal Jewani, "A Proposal for Virtual Mental Health Assistant", 7th International Conference on Advanced Computing and Communication Systems (ICACCS), March 2021

[7] Varshaa Dhanasekar,Yenugu Preethi,Vishal S, Praveen Joe I R, Booms Poolan M, "A Chatbot to promote Students Mental Health through Emotion Recognition", Third International Conference on Inventive Research in Computing Applications (ICIRCA), September 2021

[8] Sanobar Dar, Aniko Ekart, Ulysses Bernardet, "The Virtual Breathing Coach", IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), March 2022

[9] Chandra Mohan Dharmapuri , Ashutosh Agarwal, Faizan Anwer, Jigyasa Mahor, "AI Chatbot: Application in Psychiatric Treatment and Suicide Prevention", International Mobile and Embedded Technology Conference (MECON), March 2022

[10] Shu Wei, "[DC] Dynamic facial expressions on virtual humans for virtual reality (VR) mental health therapy", IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), March 2022

[11] Shreyas Md, Hemant Satish, K S Koulini, Aleeza Inamdar, U Ananthanagu, "A Radical Approach To Depression Detection", IEEE 7th International conference for Convergence in Technology (I2CT), April 2022

[12] Nadja Damij, Suman Bhattacharya, "The Role of AI Chatbots in Mental Health Related Public Services in a (Post)Pandemic World: A Review and Future Research Agenda", IEEE Technology and Engineering Management Conference (TEMSCON EUROPE), April 2022

[13] Reuben Crasto, Lance Dias, Dominic Mira, Deepali Kayande, "CareBot: A Mental Health ChatBot", 2nd International Conference for Emerging Technology (INCET), May 2021

[14] Sushruth P Konapur, Tejas Krishna, Varun G, Uma R, Sarojadevi H, "Design of a Chatbot for People Under Distress using Transformer Model", 2nd Global Conference for Advancement in Technology (GCAT), October 2022

[15] Sepp Hochreiter, Jurgen Schmidhuber, "Long Short- Term Memory", Neural Computation, Decemeber 1997

[16] Oren Melamud, Jacob Goldberger, Ido Dagan, "Learning Generic Context Embedding with Bidirectional LSTM", Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, 2016