A Comparative Study of Chatbot Catered Toward Mental Health

Pranto Dev , Sameeha Haque , Asmita Noor , Abir Alam Srabon , Mashruk Mohammed Wasik , Sumaiya Mim , Shadman Bin Sharife , Fariha Rahman , Syeda Rifa Syara , Shadab Iqbal , Md Humaion Kabir Mehedi , and Annajiat Alim Rasel

Department of Computer Science and Engineering
Brac University
66 Mohakhali, Dhaka - 1212, Bangladesh
{*pranto.dev, sameeha.haque, asmita.noor,abir.alam.srabon, mashruk.md.wasik.ataher, sumaiya.mim, shadman.bin.sharife, fariha.rahman1, syeda.rifa.syara, shadab.iqbal, humaion.kabir.mehedi*}*@g.bracu.ac.bd, annajiat@gmail.com*

***Abstract*—The number of people suffering from severe depression has risen in recent years. The majority of patients are apprehensive about seeking counseling and are unwilling to open up. A chat bot might be a viable tool for involving customers in artificial intelligence-powered behavioral health therapies. Chat bots are artificial intelligence entities that answer to users in normal language, exactly like a person would. Social chat bots, in particular, are those that form a deep emotional bond with the user. We shall explore and compare such chat bots in this paper, as they play an important role in assisting patients with mental illness. The study will compare and contrast chat bots such as CARO, XiaoIce, DEPRA, PRERONA, and Eviebot, as well as their role in resolving the depression problem. The article will show how the different chat bots compare in terms of methodology, underlying algorithms, accuracy, population demographics, and limitations. Finally, the paper will provide a quick overview of chat bots' future advancements in this field. The therapeutic component, which determines a person's level of depression, is also a priority.**

***Index Terms*—NLP, AI, Chatbot , Mental health , Comparison, Sentiment analysis**

I. INTRODUCTION

A conversational agent, often known as a chatbot is a type of software that can converse with a human using natural languages. Written text is medium of communication, from which we can extract human emotions [1] and feed this data to computers. Creating a well-functioning chatbot has been one of the most difficult tasks in Artificial Intelligence since its inception. One of the most important goals of Natural Language Processing (NLP) and Artificial Intelligence (AI) is to model the process of a good conversation, which is one of the many functions of a chatbot. Even though chatbots can be used for a variety of functions, their primary goal is to recognize human utterances and answer appropriately. Chatbots are frequently known as chat robots which are programming specialists that can reproduce human discussions through message or voice messages. In this area of computational learning theory research, Machine Learning has played a significant influence. Building methods allowing Artificial Intelligence agents to replicate human communication using voice commands or text, and in some cases both, has been one of the key emphasis points in Artificial Intelligence in recent years. Artificial Intelligence agents capable of performing this task are known as Chatbots. It has been difficult to grasp the true nature of chatbots because they closely resemble intelligent human beings. This computerized agent behaves like a human and has the potential to play a key role in healthcare services. Due to the pandemic in recent days, we have encountered an unprecedented situation in our generation to be compelled to sit at home, maintain social distance and avoid social gatherings, which is unprecedented in our generation. It has been noticed that numerous mental illnesses including sadness, anxiety, frustration, mood swings, and so on are becoming more common among people all over the world. In light of these circumstances, the analysis of different ChatBots to assist people in coping with depression or emotional needs, as well as to provide advice and therapy to those who are suffering. People suffering from depression and anxiety, Hussain et al. highlight the necessity of chatbots in mental health care. According to world heath organization (WHO), more than 300 million suffer from depression or the effects of wretchedness, representing 4.4% of the worldwide population. According to a national survey on mental health in Bangladesh (2003-2005), around 5 million people, including 4.6% adults and 1% children, suffer from depression [2]. These figures highlight the significance of chatbots in the psychological and mental health fields. As a result, this study presents a chatbot architecture for mental health therapy based on natural language processing (NLP).

II. RELATED WORKS

Many researchers worked with chatbots that used various machine learning methods to detect various levels of depression and anxiety.

Ashraf et al. employed machine learning techniques depending [3] on a video and image-based depression identification

978-1-6654-6837-4/22/$31.00 ©2022 IEEE

model. And, the study examined data acquisition methods as well as their datasets. This study also examines the indicators of depression. The evaluation of various studies is described, as are their quality factors. The study concluded with comments on the methods used and the potential applications of video and image-based depression assumptions in the future.

In the paper of Yin et al. the authors proposed Eviebot, a breakthrough sequence-to-sequence [4] oriented, completely associative interaction approach for the diagnosis of negative emotions and control of distress by favorably recommending answers. Deep-learning-based models in the system include a Bi-LSTM-based model for identifying emotional trauma in participants as well as retrieving a mental counseling-related corpus for guiding the chatbot, an anti-language sequence-to-sequence neural network, and a highest mutual information model. They deployed the embedded type of software (chatbot system) on an online network for real-world university use, and after a month of user testing, they noticed superior outcomes in terms of rising optimism than the control group's public chatbot.

CARO is a chatbot application developed by the authors Harilal et al. [5] that can conduct sympathetic discussions and provide medical advice to those suffering from severe depression. It will be able to identify the discussion's context, as well as its intent and feelings. It can also generate a sympathetic reaction or clinical advice based on the ensuing understanding established from the features recognized based on the user's desire.

In the paper of Uddin et al. the authors presented a time-saving model for detecting texts representing [6] self-perceived depressive symptoms using a Long Short-Term Memory (LSTM) based Recurrent Neural Network (RNN) in their paper. The strategy is put to the test using a massive dataset gathered from a Norwegian public internet information channel for teenagers. On this information channel, the dataset contains young people's own text-based queries. A strong characteristic derived from the projection of likely depressive signs pre-defined by medical and mental health professionals is then delivered as features from a one-hot method. To train the time-sequential characteristics that distinguish writings reporting anxiety and depression from postings without such descriptions, a deep learning approach RNN is used. At last, the prepared RNN is utilized to anticipate sad posts spontaneously. When the technology is evaluated by comparing it with the traditional methodologies, it outperforms them.

In the paper of Zhou et al. the author presented XiaoIce, it is built on a framework for compassionate computing that allows a machine (in this case, a social bot) [7] to perceive human sentiments and states, comprehend user intents, and dynamically respond to user demands. It aspires to pass the time-sharing Turing Test, in which robots and humans cohabit in a companion system with a time-sharing schedule. The integration of both intellectual and emotional quotients is at the heart of the system design (intelligent quotient (IQ) and emotional quotient (EQ)). It has a distinct personality as well. This one is a social chatbot with compassion, personality, and abilities, as well as the ability to integrate both EQ and IQ to optimize for long-run user engagement, as measured by anticipated CPS. This chatbot has established long-run relationships with millions of chatbot users worldwide, earning an average conversation-turns Per Session (CPS) of 23, which is way greater than that of other chatbots and even human conversation.

The introduction of chatbots using artificial intelligence has been constructive in order to treat stress disorders such as depression. Analyzing [8] how chatbots work could be a better approach to further improve their performance. Chatbots can be utilized as BIT (behavioral intervention technology) which can provide user replies depending on responses generated by AI or programmed. Using a popular chatbot Tess and its data, the research is conducted to have an in-depth understanding of its working mechanism. Tess has 12 modules designed for particular mental health conditions. Interaction of 354 users, including messages sent and received, time spent transition, and usage of modules was used for this research purpose. Descriptive statistics were used as a tool to analyze the data. When Tess started a conversation with the users initialized the depression diet module and depending on the analysis Tess forwarded them to the 'body scan module' and 'transtheoretical module'. The message commenced by tess shows that most users were directed to the 'depression diet module'. Among the 12 modules, the completion rate was 40%. 25 mins of the desired conversation boost the morale of the user. Chatbots can play a major role as a mental health resource. Moreover, not all the modules were utilized properly. Rather than increasing the number of modules, keeping a handful of quality modules can help users greatly.On the contrary, the demographic of the population is unknown, the data provided had ambiguity about the source.

In this paper, the author compared chatbot therapy against bibliotherapy in terms of dealing with depression for university students. The main objective was to establish a chatbot as an effective, less costly option for dealing with depression. A total of 83 students were selected and divided [9] randomly using SPSS into two groups: the Chatbot test group and the Bibliotherapy group. They were observed over a period of 16 weeks, clinical questions were set every 4 weeks to observe their progress. A chatbot named XiaoNan was developed for this purpose. It has a Natural Language Understanding module which includes NLP, Intention classification, emotion recognition, and Dialogue management module. The final message is generated from the Natural language generation module. On the other hand, Bibliotherapy gives treatment to patients using psychological advice and literature reading which is proven to be one of the effective methods of handling depression. Chatbot has been proven effective over a longer period of observation. Moreover, chatbots also keep track of symptoms and compare them with the records. Furthermore, it provides guidance for the users using the DPO module. The study sample lacks variety, generalisability, and lack in content for continuing a long period of conversation.

Klos et al. used an AI chatbot named Tess to study mental health and uplift their findings. The main target was to examine whether AI-based [10] chatbots were capable enough to treat mental health-related patients.There were two groups, the experimental condition group, and the control condition group. Mann-Whitney U and Wilcoxon tests were used as determinant methods for depressive symptoms and unconventional tests for anxiety detection for the control condition group and Tess for the experimental group. The demographic of this study was from Latin America, 181 Argentine students were selected in the age group 18-33. There was not any dramatic difference between the results of the experimental group and the control group. The study depicts that AI chatbots are capable and have the potential to treat mental patients. Moreover, it is a developing technology. More study on this is a field required.

DEPRA is built on Dialog Flow as a conversation interface and uses personalized utterances collected from a focus group to train it. The users can access it via any smart device like smartphones, [11] laptops, etc. it is designed under the Google cloud platform. It receives messages from participants using apps like Facebook messenger and the backend system is connected to the AWS database. The database holds all the records of conversations between agents and clients. Two different methods were used to predict the conditions of the patients, they are IDS-SR and QID-SR. These methods judged users and gave a score on a scale of 0-3. Three psychometric themes were taken under consideration, they are sleep symptoms, weight symptoms, and psychomotor symptoms. Depression can be detected in the early stage if users maintain a full level of transparency. As it was seen the users had high-level depression whereas their cases were not the server. It is a key step to determining the accuracy of the chatbot as providing vague/incomplete answers will result in incorrect system analysis. A more robust dataset and a greater number of participants may allow further improvements in the system. Automatic scoring methods based on NLP and sentiment analysis will improve the precision and accuracy of the chatbot.

The authors Hussna et al. created a chatbot for Bangla, English, and Korean native speakers, named “Prerona”. The chatbot can provide replies [12] to the queries asked and adequate mental health care support. Moreover, it uses adaptive learning techniques to answer questions that it does not know in their current state and save it in the database. The messages from users are processed using NLP and replies are generated from the database which is custom made for research purposes and clinically approved. The precision of the chatbot depends on the user’s ability to type, the data processing ability of the bot, and the available data on the database. As the Bangla language is computationally costly to process, it uses keyword-based communication methods to increase performance. To make the chatbot dynamic and robust they plan to implement a machine learning algorithm in the system. Furthermore, the author wants to embed a voice-based search system, so that people without proper education on technology can also take help from this chatbot.

III. Comparison

A. Methodology

*1) **DEPRA**:* DEPRA [11]is a conversation interface built on Dialog Flow and trained with individualized utterances from a focus group. Users can access it via smartphones, laptops, etc. which is built on Google cloud. The backend system connects to an AWS database and gets messages from participants via Facebook Messenger. The database stores all client-agent communications. The patients’ conditions were predicted using two methods: IDS-SR and QID-SR. These approaches rated the user from 0 to 3. Three psychometric themes were considered: sleep, weight, and psychomotor symptoms. Depression can be noticed early if people are fully transparent. The users exhibited high levels of depression, but their cases were not served. Providing vague/incomplete answers will result in faulty system analysis. A larger dataset and more individuals may allow for future system improvements. Automatic scoring using NLP and sentiment analysis will increase the chatbot’s precision and accuracy.

*2) **CARO**:* The medical advice generator [5] and sympathetic dialogue generator are the two models that make up the proposed model. Based on the user’s stated intent, the delivered utterance is routed to one of these models. They developed it as a binary classifier using medical inquiry answers and replies to empathetic dialogues as training data. When a text is presented, its intent is first categorized as either a 1 or a 0, and then it is assigned to one of the defined models. Medical advice is generated by the first model, which utilizes an LSTM architecture that is trained on a dataset of medical questions and replies. Other models are trained on empathetic discussions, but the other model generates sympathetic replies. By keeping the discourse on track it is made easier by referencing two prior statements before formulating a response. The final word is generated word-by-word in this paradigm, which also utilizes a feedback mechanism.

*3) **XiaoIce**:* XiaoIce [7] has three layers and they are user experience, chat engine, and data. With full-duplex and turn-based communication, where the user experience layer lets XiaoIce connect to a popular platform chat such as QQ and WeChat. Full-duplex mode handles simultaneous user-XiaoIce's voice-stream communications. Also, a user and XiaoIce can have a conversation via message. This layer also includes components for processing inputs of the user and responses is made by XiaoIce, such as understanding images and normalizing the text, recognizing speech and synthesis, detection of voice activity for differentiating the user input out of the background noise, lastly, the talking-to-bot classifier is used to distinguish between human & bot users. Core chat, dialogue skills, and a computing module for empathy comprise a Conversation engine layer. The dialogue manager tracks the dialogue state and uses that dialogue policy to generate the responses by using a dialogue skill or Core Chat. The user's background, intent, perspective on a topic, emotion, and general interests are all empathic components of the discussion and user. Also included are databases for XiaoIce's profiles and those of all her active users.

*4) **Prerona**:* User input is first routed to an NLP (natural language processing) unit [12], which then processes it. Data is sent to the chat interface as well. JSON is used by the real-time data handler to collect and process all the data. UTF-8 characters are checked to see if they are present in the custom dataset. Those who responded will get a response if they fulfill the criteria. The default error message occurs if this isn't the case and is shown by the bot, it also adds "Can you just say that again?" Like this, it responds to consumer feedback.

*5) **Eviebot**:* In a Bi-LSTM i.e.(bi-directional LSTM) architecture [4], the network can be trained using both sequences and concatenated to create one final result. Because of this, a Bi-LSTM, RNN-based model is employed for classifying responses into any one of two categories: which are affirmative or negative. A bidirectional 10-10 LSTM 10-10 architecture allows the input for training from two contrasting sequences which are, one from forwarding sequences and backward ones are the other ones. All components of the binary classification model are the Bi-LSTM network and sigmoid activation function. Embedding vectors are created by using a trained word embedding model when words are supplied into the model. These will be used to feed the Bi-LSTM network as the input layer. Training results in the sigmoid activation function with a possibility/probability range from 0 to 1, where the output is less than 0.5 which is labeled as negative, and if the output is greater than 0.5 which is considered positive. Lastly, a table is presented below for comparing the main methodology of different bots:

TABLE I: Chatbots and used methodologies

Chatbot	***Methodology***
Depra	DS-SR and QID-SR
CARO bot	Medical advice and sympathetic dialogue generator
Xiaoice	The user experience layer, Full-duplex mode
Prerona	The user experience layer, Full-duplex mode
Eviebot	Bi-LSTM network and sigmoid activation function

B. Accuracy

All the chatbots that we reviewed try to deal with human emotion. Some of them try to detect depression or anxiety while some try to relieve or lessen the symptoms. The accuracy of such chatbots can be measured with multiple types of metrics. Among them, one is the BLEU [13] score which is scored from 0 to 1, where 1 is the highest score representing that the machine-translated output has the closest similarity with the original text. Other than this there is PHQ-9 which is a self-test-based scoring system where the user is given a score from 0 to 20, where anything above 20 is severely depressed. There is also GAD-7 another self-test tool with a score from 0 to 15 where anything above 15 is severe anxiety. Among our reviews chatbots, the first one is CARO which has a BLEU score of 0.179 and a BERT score of 0.83 which when compared to the BLEU score of 0.08 AI developed by Facebook seems like a significant improvement. The next one is XIAOICE by Microsoft which was based on a persona model which after multiple tests scored as high as 0.188 on the BLEU scale. Another bot is PRERONA which tries to make sure people suffering from depression or severe anxiety get a chance at counseling by making it digital and available to all. This chatbot works in three languages including English, Bengali, and Korean. This was tested in two phases where for Bengali, English, and Korean the highest accuracy rate of accurate answers from the bot was 13.70%, 14%, and 12.93% respectively. One more chatbot is Tess which also aims to lessen the symptoms of depression and anxiety. The test was done on two groups over 4 weeks, one group had access to the chatbot while the other did not. Within 2 weeks the ones with access to Tess had their PHQ-9 score lessened by 1 and also their GAD-7 scores also decreased whereas the ones with no access to Tess saw an increase in both GAD-7 and PHQ-9 scores.

TABLE II: Chatbots,Feature and Metrics used

Chatbot	***Key Features***	***Metrics***
CARO	Medical Advice	BLEU, BERT
XIAOICE	Persona Model	BLEU
Prerona	Multiple Languages	SPSS
Tess	Modular System	PHQ-9,GAD-7
Depra	Early detection,accessibility	NA

C. Population Demography

For discussing population demography, we have chosen 6 chatbots; DEPRA, CARO, XiaoIce, Tess, Prerona, and Evebot. DEPRA does not have a specific target audience in mind, it does provide statistics on young people and older women in low- and middle-income countries, with an annual case count of about 800,000. With the median age of death at 43.9, intentional self-harm is one of the top five causes of death among Australians aged 15 to 44. As a result, early depression detection analysis is necessary to offer early warnings for behaviors that DEPRA can identify, such as suicidal thoughts, self-harm, and mental health difficulties. CARO, on the other hand, focuses on teens who are hesitant to talk about their mental health. The apprehension stems mostly from a fear of revealing one's identity as well as the societal stigma associated with mental health issues.CARO chatbot is tested on Indians. XiaoIce's tests are based on Chinese university students. Another chatbot, Tess works with college students to reduce self-reported symptoms of despair and anxiety. Tess conducts less intensive conversations and is more cost-effective. A total of 75 people were chosen from 15 institutions around the United States. All of the participants took part in several online questionnaires. This research included a total of 74 participants, with zero percent retention, less than 1% retention in the test group (1/24), and less than 1% retention in the control group (1/24). The average age of the participants was 23 (22.9) years, with 70% of the participants being female, predominantly Asian 51%, and white 41%. The results showed a statistically significant difference between the control and group 1 groups, with group 1 reporting a significant reduction in depression symptoms, using an alpha level of.05 for all statistical tests. PRERONA, on the other hand, focuses solely on Bengali people. The people who know the Bangla language are the targeted audience. According to the first national study on mental health in Bangladesh (2003-2005), around 5 million individuals, including 4.6 percent of adults and 1% of children, suffer from depression.

D. Limitations

Treating depression using a chatbot can have multiple limitations. For example, in certain research regarding the chatbot Tess, it does not collect demographic information, that's why users could not use it in a personalized way. Second, they did not keep track of the discontinued user or the users who were shifted into a different module of the chatbot. Also, not all users who start using Tess will benefit from it. Tess does not direct users into their required corresponding module. In addition, there might be system errors in the module. Third, they did not collect data when the user was most active. Fourth, the modules of Tess cannot be used on other chatbots since those chatbots are not of a modular format like Tess. Another study was conducted using the chatbot, XiaoNan, on University students. There were several limitations in this study. The contents of the chatbot were limited as the dialogues and procedures followed by the bot were administered after guidelines from the professionals. Due to this shortage of content, a lot of the conversations felt repetitive and robotic, failing to establish a more humane connection between the chatbot and the user. Additionally, the chatbot does not have a memory to store conversation history. As a result, conversations start anew each time without any context of the situation. This results in generalized conversations without any special modifications tailored to the user's particular needs and symptoms. Furthermore, another study by Yin et al., Eviebot, A Deep Learning-Based Chatbot for Campus Psychological Therapy, has few limitations. First, if the user makes any grammatical error or spelling mistake then the bot might not give correct responses. Second, the chatbot does not store the previous data therefore this might be inconvenient for the user. If we can overcome the above limitations then the chatbot will be more convenient for the user.

IV. Future Plan

The population cannot be categorized considering the geographical location only to globalization. On the other hand reason behind depression is also dependent on culture which is based on geographical location. As the resolve for an ideal chatbot is contradictory to itself, we propose to build a chatbot that has features to select a language and geographic location. Based on the user's choice we will ask a series of questions to determine the health condition of the user. We will create a system having 4 modules. They are :

- Anxiety Module
- Depression Module
- Post-traumatic stress disorder (PTSD)
- Friend Module

Depending on the user's condition we will forward them to any of the four modules. Moreover, after the test is taken, we will set two threshold values. If the user scores over the maximum threshold value then we will suggest the user contact the nearest mental health care center and give them a detailed contact list. If the user scores lower than a minimum threshold value, they will be evaluated as a mentally healthy person and they will be asked if they want a subscription to mental health care tips daily provided by our chatbot. Our chatbot will take weekly tests using GAD-7 and PHQ-9 to monitor patients' progress. All the modules will have separate models. Each model will be provided trained on corpus relating to the respected domain. Sentiment analysis will be used to determine the mood of the user and reply depending on the user's mood. If the user's text is out of the domain of the boat then the user will be suggested with googled results. Finally, quality research having a known demographic will be beneficial for the development of an ideal chatbot we are looking forward to building.

V. Conclusion

Even though chatbots have not reached their full potential in the detection of mental health conditions, the progress in the field shows that with further research and development it is possible to completely switch to chatbots to create cognitive-behavioral systems. Our research brought to light that there are still vast differences between communication across different modules, however, the aim is to narrow down these differences for sustainable usability. The chatbots mentioned throughout the paper shows how far these systems have come in terms of human-level intelligence despite the limitations each of them holds.

The inaccuracies in the system results can be reduced greatly if the users maintain complete transparency while carrying out a conversation with the chatbots. Since these chatbots depend highly on the use of language and words, it is essential to use references that reflect the participant's emotions. With further use of Natural Language Processing (NLP) in sentiment analysis, inaccuracies can be brought to a minimum and better precision in the results can be obtained. Increased accuracy and usability will allow more participants to resort to chatbots as an entity to confide in when they do not feel too comfortable sharing certain aspects of their sufferings with a therapist. Furthermore, greater accessibility to these chatbots will also help users receive quicker feedback compared to therapy sessions or other psychological treatments. The main goal of using chatbots in this field is to make teletherapy available to everyone. Of course, it is not an easy task but at the rate at which NLP and machine learning are advancing, this goal does not seem too far away.

REFERENCES

[1] F. Rahat, M. Mim, A. Mahmud, I. Islam, M. J. A. Mahbub, M. H. K. Mehedi, and A. A. Rasel, *Data Analysis Using NLP to Sense Human Emotions Through Chatbot*, 07 2022, pp. 64–75.

[2] S. Siuly, H. Wang, L. Chen, Y. Guo, and C. Xing, *Health Information Science 10th International Conference, HIS 2021, Melbourne, VIC, Australia, October 25–28, 2021, Proceedings: 10th International Conference, HIS 2021, Melbourne, VIC, Australia, October 25–28, 2021, Proceedings*, 01 2021.

[3] A. Ashraf, T. Gunawan, B. Riza, E. Haryanto, and Z. Janin, "On the review of image and video-based depression detection using machine learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, p. 1677, 09 2020.

[4] J. Yin, Z. Chen, K. Zhou, and C. Yu, "A deep learning based chatbot for campus psychological therapy," *CoRR*, vol. abs/1910.06707, 2019. [Online]. Available: http://arxiv.org/abs/1910.06707

[5] N. Harilal, R. Shah, S. Sharma, and V. Bhutani, "Caro: An empathetic health conversational chatbot for people with major depression," 01 2020, pp. 349–350.

[6] M. Z. Uddin, K. Dysthe, A. Følstad, and P. Brandtzaeg, "Deep learning for prediction of depressive symptoms in a large textual dataset," *Neural Computing and Applications*, vol. 34, pp. 1–24, 01 2022.

[7] L. Zhou, J. Gao, D. Li, and H.-Y. Shum, "The design and implementation of xiaoice, an empathetic social chatbot," 2018. [Online]. Available: https://arxiv.org/abs/1812.08989

[8] G. Dosovitsky, B. S. Pineda, N. C. Jacobson, C. Chang, M. Escoredo, and E. L. Bunge, "Artificial intelligence chatbot for depression: Descriptive study of usage," *JMIR Form Res*, vol. 4, no. 11, p. e17065, Nov 2020. [Online]. Available: http://formative.jmir.org/2020/11/e17065/

[9] H. Liu, H. Peng, X. Song, C. Xu, and M. Zhang, "Using ai chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness," *Internet Interventions*, vol. 27, p. 100495, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214782922000021

[10] M. Klos, M. Escoredo, A. Joerin, V. Lemos, M. Rauws, and E. Bunge, "Artificial intelligence chatbot for anxiety and depression in university students: a pilot randomized controlled trial (preprint)," 07 2020.

[11] P. Kaywan, K. Ahmed, Y. Miao, A. Ibaida, and B. Gu, *DEPRA: An Early Depression Detection Analysis Chatbot*, 11 2021, pp. 193–204.

[12] A. Hussna, A. Laz, M. Sikder, J. Uddin, H. Tinmaz, and A. Esfar-E-Alam, *PRERONA: Mental Health Bengali Chatbot for Digital Counselling*, 02 2021, pp. 274–286.

[13] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: A method for automatic evaluation of machine translation," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ser. ACL '02. USA: Association for Computational Linguistics, 2002, p. 311–318. [Online]. Available: https://doi.org/10.3115/1073083.1073135