International Journal of Computational Science and Engineering Research
ISSN: XXXX- XXXX(Online) , http://www.ijcser.com/
Regular Issue , Vol. 2, Issue. 2, 2025 , Pages: 11 - 17
Received: 12 January 2025 ; Accepted: 25 April 2025 ; Published: 12 May 2025
Research Paper , https://doi.org/10.XXX/XXXX.XXXX.XX.XXXX

Data Augmentation for Cognitive Behavioral Therapy : Harnessing ERNIE Language Models Through Artificial Intelligence

Kondreddygari Archana [1], Suram Indhra Sena Reddy [2], Shaik Meethaigar Jameer Basha [3], Shaik Karishma [4]

[1,2,3,4] Department of Department of Computer Science and Systems Engineering, Sree Vidyanikethan Engineering College, Tirupati, India ; geethikakotamsetty@gmail.com

Abstract: The effectiveness of Cognitive Behavioural Therapy (CBT) for overcoming the irrational thought patterns that give rise to mental disorders is a known fact but pinpointing the cognitive pathways accurately for personalized treatment is the key. The advent of social media has made it possible for people to express their negative emotions, in a way that they disclose their cognitive distortions, and in the case of severe forms, manifest suicidal tendencies. Nevertheless, the techniques developed to analyze these lines of cognitive pathway are lacking hence, psychotherapists will have to wait until the symptoms get worse before they are able to act on time and with the right tools and methods, online environments will become a reality where they are the first to intervene in the situation. Decision Tree and Random Forest are the primary Hierarchical Text Classification models. They help the system for separating inputs from the users into the different cognitive pathways and simultaneously for detecting the negative thought pattern with the aid of these models. To be more specific, for the extraction of negative sentiment and only of that kind, the system is built with BERT for sentiment analysis in social media data. This system goes beyond the available ones by covering not only the negative thoughts but also predicting some subcategories within the negative and other physical as well as mental health categories. This update allows the understanding of the problems and also gives a basis for early detection and treatment to the psychotherapists in the form of psychological and mental health issues intervention and therefore helps in the prediction of social-emotional patterns.

Keywords: Cognitive Behavioral Therapy (CBT), Data Augmentation, Sentiment Analysis, Acceptance and Commitment Therapy , Text Classification, Phobias.

1. Introduction

The cognitive process is the mental internal activity of cramming with many activities, ideas, feelings, etc. and by passing on the senses to the corresponding emotions. Cognitive Behavioral Therapy (CBT) is a structured and problem-solving direction of the psychology that resolves the mental health predicaments of an individual. CBT's main focus is on recognizing and changing the way thoughts, feelings, and behaviors relate and the development of good mental health. Such is the pace of its growth that the method has been impacted by so many developments within the fields of psychology and technology that it has become one of the best-used and most researched ways in the care of mental health Cognitive and behavioral health changes went through an amazing pace of development, yet that is influenced by such advancements like a chat bot in AI and synthetic data for the detection of emotion and many others.

Data augmentation in Cognitive Behavioral Therapy (CBT) focuses on expanding and diversifying datasets to capture nuanced therapeutic contexts, such as thoughts, emotions, and behaviors. Using advanced models like RoBERTa and PEGASUS, data augmentation rapidly generates varied and contextually relevant text, significantly enhancing the ability to recognize cognitive patterns, classify psychological states, and simulate personalized therapy interventions. Depression is a

form critical situation that users go threw in daily life because of situations at work.

The combination of real-time chatbots and virtual therapists has really improved how accessible CBT interventions are. These AI-driven chatbots can analyze conversations, pick up on signs of distress, and provide helpful therapeutic advice, effectively connecting professional therapy with self-help options. This is especially useful for those who can't access traditional therapy due to financial, geographical, or social challenges.

One significant use of deep learning in CBT is hierarchical text summarization, where AI models pull out key cognitive elements—like the activating event, belief, consequence, and disputation—from what users share. This allows AI-powered CBT systems to give structured feedback and personalized therapeutic suggestions based on a person's emotional state. By using explainable AI techniques, such as attention mechanisms found in transformer models, these systems can provide clear and understandable insights, making them more trustworthy for both clinical and self-help purposes.

In addition to analyzing text, AI-driven CBT systems also aim to predict behavioral patterns based on identified emotional states. Machine learning models trained on psychological data can determine if someone is likely to engage in positive behaviors or unhealthy coping strategies. This predictive ability allows for real-time interventions, where the AI can recommend relaxation techniques, mindfulness exercises, or motivational affirmations tailored to the user's thought processes. Moreover, methods like job description parsing and personality trait analysis have been investigated to better understand behavioral tendencies in workplace and social settings, further enhancing AI-based CBT applications.

2. Literature Survey

Ghanadian et al. highlighted the importance of generating data for socially anxious individuals to improve the discovery of suicidal creativity using large language models (LLMs). Their research emphasizes the need for effective data generation methods to address the lack of data in mental health research, which can lead to more accurate and compassionate AI-driven interventions [1].

Omarov et al. merged Cognitive Behavioral Therapy (CBT) with AI and machine learning, resulting in an AI-powered mobile psychologist chatbot. This study demonstrated how a simple chatbot can serve as a scalable, accessible, and more personalized tool for delivering CBT to communities in need of such support [2].

Mittal et al. provided a comprehensive overview of how artificial intelligence is transforming mental healthcare, particularly in the delivery of CBT. They discussed the integration of natural language processing (NLP) and machine learning for tasks like emotion detection, therapy delivery, and mental health monitoring [3]. Gupta et al. introduced a chatbot designed for mental health that utilizes NLP techniques. This research focused on the chatbot's ability to provide timely, contextually relevant responses, suggesting that AI could streamline traditional CBT methods [4].

Rani et al. created a mental health bot that integrates CBT principles with remote health monitoring capabilities. Their work illustrates how AI and CBT can work together to offer therapeutic interventions and continuous health tracking, helping users maintain a healthy lifestyle [5].

Kwak et al. proposed a video event recognition framework based on scripts, highlighting its potential applications in mental health. Their research focused on constraint flow methods to understand user behavior and context, effectively enhancing the CBT system [6].

Doan et al. have implemented a highly effective fine-tuning method to enhance large language models for Vietnamese chatbots. This outcome underscores the importance of localization, which is essential for making AI-driven cognitive behavioral therapy (CBT) tools accessible to a wider audience [7].

Hu et al. introduced Low-Rank Adaptation, a parameter-efficient fine-tuning system for large language models, which has been utilized in CBT systems aimed at deployment in resource-limited areas for AI-supported mental health conditions [8].

Chavan et al. expanded on the LoRA framework for parameter-efficient fine-tuning, providing insights into optimizing AI models for scalable and customizable CBT delivery [9]. He et

al. examined the current applications of large language models in mental health and proposed a generalist AI framework. The authors focused on the adaptability of LLMs to tackle various mental health issues, including the delivery of CBT [10]. Chesney et al. conducted a meta-review of the risks associated with all-cause and suicide mortality in individuals with mental health disorders. Their findings highlight the urgent need to develop AI-driven CBT systems to address the mental health crisis and provide timely interventions [11].

3. Existing System

Detecting signs of depression in text is a fascinating process that the system handles automatically. It kicks off by gathering raw text data from various sources, including social media posts, medical records, patient journals, and survey responses. The richness and diversity of this data play a vital role in ensuring accurate analysis. Once the data is collected, it goes through a cleaning process where unnecessary characters like punctuation, numbers, and special symbols are stripped away.

This step also addresses any missing information and standardizes the text by converting everything to lowercase, setting the stage for deeper analysis. Next up is tokenization and lemmatization. Tokenization breaks the text down into individual words or tokens, which allows algorithms to examine the text at a granular level.

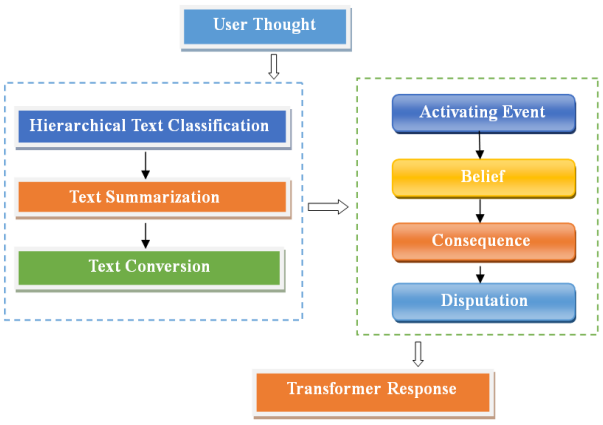

Fig 1. Internal Operational Working Mechanism

Lemmatization then simplifies words to their base forms; for example, "running," "runs," and "ran" all become "run." This standardization boosts the accuracy of the analysis. The cleaned-up text is then fed into a deep learning model known as BERT (Bidirectional Encoder Representations from Transformers) (Fig. 1). BERT creates contextualized word embeddings by taking into account the context surrounding each word, capturing nuanced meanings and relationships. This understanding is crucial for spotting emotional cues and language patterns associated with depression. In the end, the BERT model classifies the emotion in the text, helping to determine if it shows signs of depression.

4. Research Work

The Proposed System chatbot is designed to engage in conversations that feel natural and empathetic. Initially, the chatbot receives the user's message as Input Text. This message then enters a "Data Preprocessing" phase. During this phase, the chatbot tidies up the message by removing unnecessary elements like extra symbols and correcting any missing words. It also ensures that all the text is formatted consistently for smoother processing. After this cleanup, the message moves on to Emotion Classification. This is where powerful AI models like BERT and DistilBERT come into play, helping to analyze the emotional content effectively.

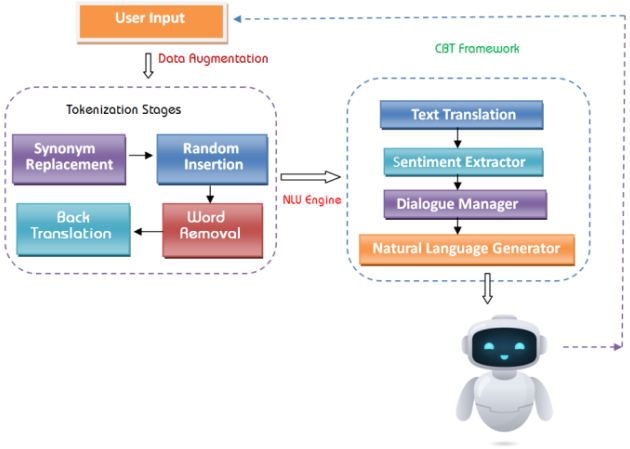

Fig 2. Proposed Architecture Working Flow

The emotions identified play a crucial role in guiding the Response Generator. This component utilizes a sophisticated AI model known as XLNET, which excels at grasping and producing text that feels human-like. XLNET takes into account the user's emotions and the context of the entire conversation, crafting responses that are not only informative but also empathetic and relevant to the user's feelings (see Fig. 2). Ultimately, the Bot Response is presented to the user, wrapping up the conversation. This system combines cutting-edge AI techniques with advanced language understanding models to create a

chatbot that can interpret and respond to user emotions in a way that feels natural and engaging, much like a real human interaction. A. ALGORITHM The AI bot generates responses based on the user's input text. First, a dataset is embedded into directories to carry out data preprocessing techniques. These techniques include synonym replacement and text removal, which help the AI bot easily identify emotions and generate helpful suggestions. This process aids in categorizing the user's sentiment based on the input text, allowing the system to classify emotions as positive or negative using BERT and RoBERTa models. Then, based on the user's intent, a response is crafted to offer healthy suggestions that steer users away from negative thoughts.

Step 1: The user begins by entering a text message T, which consists of a sequence of words w, where wi represents the i-th word in the sentence.

$T = \{w1, w2, \ldots, wn\}$

Step 2: For each word wi in T, perform part-of-speech (POS) tagging, where pi is the part-of-speech tag for wi. Retrieve a set of synonyms S(wi) from a database, where sj are the synonyms of wi that correspond to the same part of speech pi.

$POS(wi) = Pi\ S(wi) = \{s1, s2, \ldots, sm\}$

Step 3: The BERT model then classifies sentiment based on the vocabulary replacement, while DistilBERT is used for categorizing sentiment into positive, negative, or neutral. The sequence of words S(wi) is passed to the BERT model for text classification, embedding si into ei, and processing is completed with BERT.

$Hi = BERT(ei)$

Step 4: At this point, the AI Bot starts crafting responses based on how the input text is classified. These responses come from a dataset filled with various intents and replies, and the conversation keeps flowing by offering suggestions that align with the user's emotions. The formula

$$P(yi|Tinput,y1,\ldots yi-1)=Softmax(W.hi+b)$$

helps in this process. The possible responses are represented as $Tresponse=\{y1,y2,y3,\ldots,ym\}$. B. Implementation To create a model that generates responses based on what users input, we first need a dataset. We kick things off by setting up an AI framework using online platforms like RASA, which helps us design a bot that can effectively handle a question-and-answer format, similar to a Cognitive Behavioral Therapy (CBT) model. When users interact with the bot, it checks the dataset for word matches, including synonyms. This allows the bot to gauge the user's emotions based on their input and provide suggestions that resonate with what they're feeling. This approach enhances the interaction between the user and the AI bot, ensuring that all emotions are acknowledged and useful suggestions are offered. Platforms like Twitter and Instagram serve as rich sources for capturing user thoughts, emotions, and feelings. Social Media Sentiment Analysis Dataset: This dataset is incredibly valuable as it dives into the emotions and sentiments expressed across various social media platforms like Twitter, Facebook, and Instagram. It plays a crucial role in tasks such as sentiment analysis, opinion mining, and emotion detection through Natural Language Processing (NLP). The dataset is filled with a variety of posts and comments from users, categorized into sentiment labels like positive, negative, and neutral, making it a fantastic resource for supervised learning aimed at sentiment classification. Typically, the dataset includes several key attributes, with the content of the post being the most significant, as it serves as the primary data source for sentiment analysis. Additionally, it captures the timestamp of each post, providing valuable insights into how sentiments evolve over time.

The Twitter and Reddit Sentiment Analysis Dataset is a valuable tool for diving into how users express their feelings on two of the biggest social media platforms: Twitter and Reddit. This dataset gathers posts from these sites, categorizing them into various sentiment types like positive, negative, and neutral. This makes it incredibly useful for tasks related to sentiment analysis, opinion mining, and emotion detection. With this dataset, you can build machine learning models that uncover the sentiments behind social media text, which can help in monitoring user intentions, tracking social trends, and even spotting potential mental health concerns based on online behavior. Typically, the dataset includes several key features, starting with the actual text of the posts, which is the main

data for sentiment classification. It also includes the timestamp for each post, allowing for the analysis of sentiment trends over time and the identification of shifts in emotions, reactions, or opinions. The author attribute reveals who created the post, which is crucial for examining individual or demographic differences in sentiment. Additionally, the platform attribute shows whether the post is from Twitter or Reddit, enabling a comparison of sentiment patterns across these platforms, as each has its own unique user culture and content style.

5. Result Analysis

In this study, we dive into datasets that center around sentiment analysis related to mental health, specifically focusing on detecting signs of suicide and depression. We utilized two publicly available datasets from Kaggle, both of which offer a wealth of labeled textual data that's perfect for natural language processing (NLP) tasks in the mental health arena. The first dataset we explored is the Suicide Watch dataset. It features posts created by users on online forums where they share their emotional experiences. This dataset includes various fields such as statement, status, class, and text. The text field captures the main message from the user, while the class field indicates whether the post is categorized as suicidal, depressed, or normal. This structured approach helps us gain insights into the psychological state of the users, making it an invaluable resource for developing and assessing models aimed at detecting suicide and depression. The second dataset we examined is the Sentiment Analysis for Mental Health. This one zeroes in on sentiment classification within discussions about mental health. It also contains fields like text, statement, status, and class. The class label typically reflects the sentiment category—be it positive, negative, or neutral—making it ideal for training sentiment analysis models that can pick up on the emotional tone in user messages. Both datasets offer crucial features for our classification tasks, especially the text and class columns, which are foundational for supervised learning. They closely resemble real-world conversations about mental health online, aiding in the training of models that can recognize emotional distress and provide early alerts in digital spaces. Their structured nature and accessibility make them perfect for research and experimentation in AI-driven cognitive and emotional analysis systems

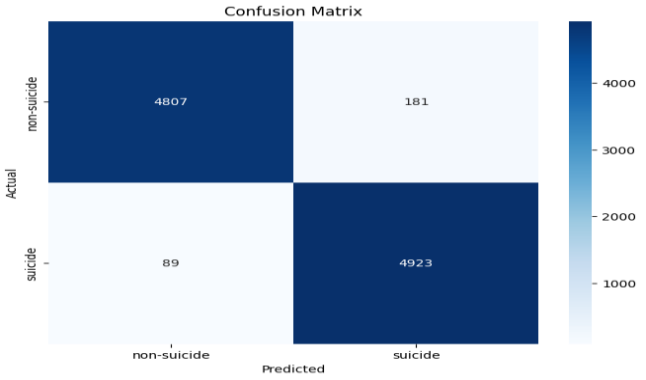

Figure 3. Confusion Matrix of BERT Model for Suicide Detection

The data shows that the model successfully identified 4,807 cases where individuals were not suicidal and 4,923 cases where they were. However, it did make some errors, misclassifying 181 non-suicidal texts as suicidal (these are known as false positives) and 89 suicidal texts as non-suicidal (false negatives). These findings suggest that the BERT model is quite effective at telling the difference between the two groups, demonstrating a strong ability for accurate classification. The relatively small number of misclassifications further emphasizes the model's strength and dependability in spotting mental health risks based on what users write. This level of performance is crucial for real-time applications, where quickly and accurately identifying suicidal thoughts can make a significant difference in providing timely help and support.

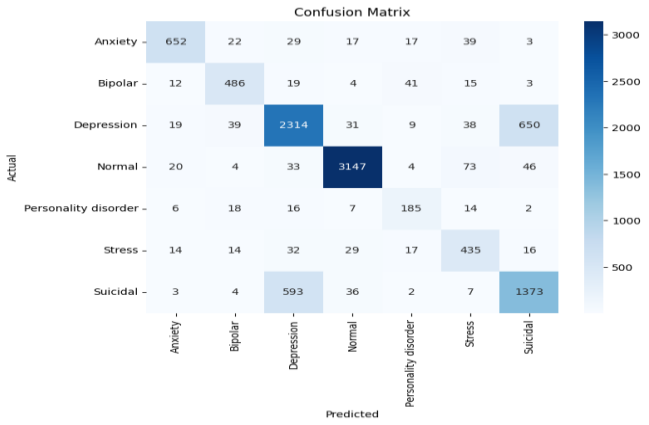

Figure 4.Confusion Matrix of BERT Model for Mental Health Disorders

The confusion matrix provides a clear picture of how well the model is performing across seven different classes, highlighting both its strengths and areas that need some work. For instance, it accurately classified 652, 486, 2314, 3147, 185, 435, and 1373 instances in their respective classes, showing impressive predictive abilities, especially in the third and fourth

categories where the accuracy is notably high. On the flip side, the matrix also points out some challenges, like the significant number of misclassifications between the third and seventh classes, with 650 and 593 errors respectively. This suggests that the model struggles with semantic or contextual overlaps that it can't quite sort out. We also see some moderate misclassifications between nearby categories, such as classes one and two, and classes five and six. Despite these hurdles, the model generally performs well across various categories, which is encouraging for tackling complex classification tasks in multi-class settings. With a bit more fine-tuning and the addition of domain-specific knowledge, we could really boost its ability to distinguish between classes, especially in situations where there's a lot of contextual ambiguity.

6. Conclusion

In this project focused on cognitive behavioral therapy (CBT), we developed a framework that leverages machine learning and natural language processing to analyze text created by users. The goal is to classify this text into important mental health categories like Depression, Anxiety, Stress, and Normal. This system aims to facilitate early detection and a better understanding of cognitive patterns by automatically analyzing textual data. We utilized models such as Decision Tree and Random Forest for clear and ensemble-based classification, alongside BERT, a robust transformer-based deep learning model that excels in context-aware text analysis. Each model was trained and tested using real user inputs, achieving impressive accuracy in identifying psychological states. The BERT model really stood out, showcasing its ability to grasp the semantic and contextual nuances of user text, making it particularly effective for mental health classification tasks. It picks up on subtle language cues that might reveal deeper emotional or psychological issues. On the other hand, the Decision Tree model provides straightforward, rule-based outputs that clarify the reasoning behind its classifications, while Random Forest enhances reliability by merging multiple decision trees to minimize variance and prevent overfitting. By blending traditional machine learning with deep learning, we achieved both accuracy and clarity in the system's results. Looking ahead, we can expand the system to offer personalized CBT recommendations based on the user's emotional state. These could range from motivational messages and coping strategies to referrals for professional help. Future upgrades might also include integrating speech and facial expression analysis, supporting multiple languages, and employing more advanced models like RoBERTa or DistilBERT to boost processing speed and accuracy. Additionally, tracking user sentiment over time could help visualize mental health trends, and enabling chatbot interactions could transform this project into a more engaging and supportive experience.

Future Scope

The future of Cognitive Behavioral Therapy (CBT) is incredibly promising, with the potential to revolutionize mental health care by making it more accessible, personalized, and adaptable. Cutting-edge AI technologies, like natural language understanding (NLU) and sentiment analysis, are paving the way for the evolution of CBT models. These advanced tools can not only help identify emotions from text but can also be applied to tackle other cognitive challenges, such as sleep disorders and phobias. By gathering data from users, we can create models that address these issues without needing direct intervention from therapists. Given the current shortage of mental health professionals, effective identification and treatment of phobias can be quite challenging. However, a well-trained model could pinpoint these problems and offer valuable treatment suggestions to help individuals overcome cognitive difficulties. Additionally, these advancements could be integrated into wearable devices, aiding in stress management through sensor technology. All of these exciting developments are on the horizon, thanks to the latest AI and NLP models.

References

[1] World Health Organization, “Depressive disorder (depression),” World Health Organization, 2023.

[2] Y. Huang, Y. Wang, H. Wang, Z. Liu, X. Yu, J. Yan, Y. Yu, C. Kou, X. Xu, J. Lu et al., “Prevalence of mental disorders in china: a cross sectional epidemiological study,” The Lancet Psychiatry, vol. 6, no. 3, pp. 211–224, 2019.

[3] J. Robinson, G. Cox, E. Bailey, S. Hetrick, M. Rodrigues, S. Fisher, and H. Herrman, “Social media and suicide prevention: a systematic review,” Early intervention in psychiatry, vol. 10, no. 2, pp. 103–121, 2016.

[4] 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.

[5] Ellis and W. Dryden, The practice of rational emotive behavior therapy. Springer publishing company, 2007.

[6] D. William, S. Achmad, D. Suhartono and A. P. Gema, "Leveraging BERT with Extractive Summarization for Depression Detection on Social Media," 2022 International Seminar on Intelligent Technology Authorized licensed use limited to: Indian Institute of Technology.

[7] Andrews, G., Anderson, T. M., Slade, T., & Sunderland, M. (2008). Classification of anxiety and depressive disorders: problems and solutions. Depression and anxiety, 25(4), 274-281.

[8] K. Rani, H. Vishnoi and M. Mishra, "A Mental Health Chatbot Delivering Cognitive Behavior Therapy and Remote Health Monitoring Using NLP And AI," 2023 International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2023, pp. 313- 317, doi: 10.1109/ICDT57929.2023.10150665.

[9] M. Aragon, A. P. L. Monroy, L. Gonzalez, D. E. Losada, and M. Montes,"DisorBERT: A Double Domain Adaptation Model for Detecting Signsof Mental Disorders in Social Media," in Proceedings of the 61st AnnualMeeting of the Association for Computational Linguistics (Volume 1:Long Papers), 2023, pp. 15 305–15 318.

[10] T. He, G. Fu, Y. Yu, F. Wang, J. Li, Q. Zhao, C. Song, H. Qi, D. Luo,H. Zou et al., "Towards a psychological generalist ai: A survey ofcurrent applications of large language models and future prospects,"arXiv preprint arXiv:2312.04578, 2023.

[11] H. Qi, Q. Zhao, C. Song, W. Zhai, D. Luo, S. Liu, Y. J. Yu, F. Wang,H. Zou, B. X. Yang et al., "Evaluating the efficacy of supervised learningvs large language models for identifying cognitive distortions and suicidal risks in Chinese social media," arXiv preprint arXiv:2309.03564,2023.

[12] Y. Sun, S. Wang, S. Feng, S. Ding, C. Pang, J. Shang, J. Liu, X. Chen,Y. Zhao, Y. Lu et al., "ERNIE 3.0: Large-scale knowledge enhancedpre-training for language understanding and generation," arXiv preprintarXiv:2107.02137, 2021.

[13] Meghrajani, V.R., Marathe, M., Sharma, R., Potdukhe, A., Wanjari, M.B., Taksande, A.B., Meghrajani Jr, V.R. and Wanjari, M., " A Comprehensive Analysis of Mental Health Problems in India and the Role of Mental Asylums", Cureus, vol. 15, no. 7, 2023.

[14] Parviainen, J. Rantala, J., " Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care", Medicine, Health Care and Philosophy, vol.25, no.1, pp.61-71, 2022.of Applied Studies, August, 2013, DOI:10.3886/ICPSR30122.v2