Harnessing Language Intelligence. Innovative Approaches to Sustainable Mental Health Interventions in the Digital Age

Malik Harris Zahir[1], Nusratullah Tauheed[1], Shahan Yamin Siddiqui[2], Sheikh Muhammad Jawad Yousaf [1], Maliha Rahim [1], Anas Arif [1]

[1]Department of Computer Science, University of South Asia, Lahore, 53000, Pakistan.

[2]Department of Computing, NASTP Institute of Information Technology. Lahore, 53400, Pakistan.

*** Correspondence.** drshahan@niit.edu.pk

Citation | Zahir. M. H, Tauheed. N, Siddique. S. Y, Yousaf. S. M. J, Rahim. M, Arif. A, "Harnessing Language Intelligence. Innovative Approaches to Sustainable Mental Health Interventions in the Digital Age", IJIST, Vol. 06 Issue. 04 pp 2027-2046, Dec 2024

Received | Nov 23, 2024 **Revised |** Dec 14, 2024 **Accepted |** Dec 16, 2024 **Published |** Dec 17, 2024.

This study explores the advanced abilities of Natural Language Processing (NLP) methods to revolutionize mental health treatment by understanding how such interventions improve therapeutic outcomes. In doing so, the work of this study is demonstrated as an innovative approach to translating conversational data into actionable insights that bridge a large gap in the detection of subtle emotional cues in mental health assessments. The research used DistilBERT, an optimized version of the BERT framework, which has been fine-tuned on specially selected datasets to accurately identify emotional states such as sadness, joy, anger, and fear. Emotional and linguistic patterns were analyzed to identify often unarticulated signals to identify disorders such as depression, anxiety, and Post-Traumatic Stress Disorder (PTSD) much earlier. In this regard, the model has been found to significantly enhance the understanding of patients' emotional states more accurately and subtly than through traditional means. The findings of this study highlight the potential of offering individualized therapeutic interventions within digital health applications, which enables immediate emotional well-being assessments. The study showcases the flexibility of AI-based systems, making them applicable to almost any environment, including a workplace setting, to promote both wellness and productivity. This study sets the ground for developing scalable, customized, and proactive mental health care strategies that are beyond conventional therapeutic frameworks.

Keywords. Natural Language Processing, Mental Health Analysis, Digital Intelligence, Emotion Detection, Harnessing Language Intelligence.

Introduction.

Mental health disorders, including depression, anxiety, and PTSD, significantly affect both individuals and economies globally [1], thereby requiring the development of innovative and effective management strategies. Conventional therapeutic methods frequently depend on verbal interactions and self-reported symptoms, which may inadequately reflect the emotional intricacies and complexities inherent in an individual's mental condition. This is the most recent breakthrough of NLP which has revealed an unparalleled viewpoint of exploring therapeutic dialogues [2] and opens up to the linguistic as well as emotional patterns underneath mental illnesses.

Current research into sentiment and emotion analysis has revealed underlying states of emotions like sadness or frustration without ever being explicitly described. In light of this, much ground still needs to be covered before these can be applied to personalized mental health programs. Another advanced NLP model that has demonstrated high accuracy in text testing is BERT, with its distilled version, DistilBERT, also proving effective for these applications [3].

This research undertakes a systematic approach to addressing the problem identified above. First, therapeutic conversations are collected and preprocessed to create datasets. Then, DistilBERT is fine-tuned to detect and analyze emotions and sentiments and pay special attention to subtle emotional indicators often overlooked by traditional approaches. Finally, the findings are analyzed in terms of implications for mental health care and potential integration into digital health platforms for real-time assessments.

The purpose of this research study is to develop an Artificial Intelligence (AI) based tool that can efficiently analyze therapeutic dialogues for the early detection of mental health disorders, thus improving therapeutic interventions. Originality in this research arises from its capacity to decode and interpret hidden emotional signals using advanced NLP models. It fills gaps in mental health services and provides adaptive solutions in different settings.

Research Objectives.

- To fine-tune and apply the DistilBERT model for emotion and sentiment detection.
- It implements and adapts the DistilBERT model for identifying emotional and sentiment patterns in patient-therapist conversations in more nuanced ways toward early detection of mental health disorders like depression, anxiety, and PTSD.
- To analyze emotional and linguistic patterns in therapeutic conversations.
- The research is tailored to determine, through patient-therapist dialogues, the often unconscious emotional and linguistic signs of a conversation, even if that topic is unknown or masked, say sadness or frustration.
- To improve personalized therapeutic interventions using AI-driven insights.
- Another key goal is to leverage AI-generated emotional insights to create personalized therapeutic interventions, which can be integrated into digital health platforms for real-time mental health assessments.
- To measure the effectiveness of emotion recognition in enhancing ratings of mental health.

This study focuses on improving the overall efficiency of the NLP-driven methodology by accuracy, precision, and recall measurement in a model's quality assessment.

Novelty of Study.

The research highlights scalability and easy applicability on digital health platforms, showing it can create change by offering tailored and predictive mental care interventions. In contrast, traditional solutions rely more on direct verbal responses and subjective self-evaluations. This research instead uses sentiment and emotion analysis to detect subtle emotional cues that are often ignored. The research achieves high accuracy in the identification of complex emotions, such as sadness, anger, fear, and joy, by the fine-tuning of DistilBERT with a focused dataset. This approach not only gives a deeper insight into the interaction of emotions in a therapeutic setup but also aids in an initial diagnosis of mental health conditions like depression and anxiety.

Literature Review.

NLP and mental health research have been a growing interest in recent years due to their potential to revolutionize the methods of therapy and, therefore, patient care. This literature review discusses the emerging framework of NLP applications in mental healthcare with special emphasis on sentiment analysis, emotion detection, and the ethics involved in these technologies. Pruksachatkun et al. (2020) [4] studied intermediate-task transfer learning in NLP that could be useful for applications specific to mental health. Advancements in transformer models, including BERT and its more lightweight variant DistilBERT, have dramatically improved the recognition of emotional signals within text-based information. Kulkarni et al. (2024) [5] demonstrated the utility of leveraging pre-trained embedding for clinical feature extraction from health records, with implications on possibilities opened up by fine-tuning methods in health contexts.

NLP is an important area of artificial intelligence that allows computers to interact with human language in complex and context-sensitive ways. Althoff et al. (2019) [6] showed the feasibility of NLP in analyzing large-scale counseling conversations, thereby proving its utility for therapeutic purposes. The applications of NLP include domains like text and speech recognition, language translation, and sentiment analysis. Islam et al. (2024) [7] used machine learning algorithms to build predictive models for mental health disorders but acknowledged that more detailed emotional analysis is required. From mental health fields, NLP permits the opportunity to analyze patterns of human expression and emotionally charged conditions, offering an in-depth understanding of thinking models and behavior patterns.

Sentiment analysis falls under the NLP foundations, which include extracting emotional cues derived from text or speech input. By examining linguistic cues, sentiment analysis algorithms can draw out emotional valence or intensity, revealing underlying psychological states. Most recently, researchers have applied sentiment analysis in detecting Depression and Anxiety through social network posts and online therapy consultations, thus helping in early treatments. This potential sentimental analysis technique enhances therapeutic methods since it provides clinicians with in-the-moment, fact-based viewpoints of the client's affective state over the therapy process. Abubakar et al. (2024) [8] leveraged reinforcement learning to engineer sophisticated chatbots that enable the imitation of therapeutic dialogues and make mental health support scalable.

Delgado-Contreras et al. (2024) [9] conducted a systematic review covering 609 studies on emotion recognition and sentiment analysis and thus supports the use of multimodal approaches that combine text, speech, and physiological data to improve the accuracy of emotion detection. Further developing NLP applications in mental health, Khoo et al. (2024) [10] discussed machine learning for multimodal mental health detection by reviewing passive sensing approaches, which combine data from various sources, including speech and facial expressions, to provide a more holistic understanding of a patient's emotional well-being.

The concept of "Language Intelligence" involves a more nuanced appreciation of cultural context and emotional subtext. It is for these reasons that a balanced approach is crucial in the realm of mental health; for knowing how to decipher subtle intent can mean all the difference between a productive or pointless therapy. While AI applications in healthcare have been improved significantly in physical healthcare, applications in mental health remain virtually unexplored (Khera et al., 2024) [11]. The ethical dilemmas that arise with data curation in the application of artificial intelligence to mental health, as investigated by Andrews et al. (2024) [12], underpin the need for implementing proper data protection measures. In light of this perspective, an Advanced NLP model might help therapy by allowing a therapist to uncover inner feelings, thereby ensuring a highly specific intervention for a particular patient.

Ethical concerns are highly influential in this changing context. Privacy issues regarding the use of sensitive personal information, such as therapy transcripts, underscore the need for a comprehensive data protection approach. Mobile application-based interventions were effective at reducing the risk of depression by Deady et al. (2022) [13], however, the long-term effectiveness was still unknown. D'Alfonso (2020) [14] also underscored the importance of ethical frameworks, noting that while AI tools like NLP can enhance therapeutic interventions, they must prioritize patient privacy and confidentiality.

The use of pre-trained models and advanced techniques in NLP is still becoming better. This work, by Liu et al. (2022) [15], on the relationship between text sentiment and self-reported depression, explains the capability of sentiment analysis in grasping mental health conditions. The systematic review by Lin et al., 2021, [16] shows that with the involvement of physiological signals and other NLP techniques, earlier detection of mental disorders is possible. Kleinberg et al. (2020) [17] explored emotions during the COVID-19 pandemic, which highlights the ability of NLP to extract important emotional responses using real-world datasets.

Cheung et al. (2021) [18] illustrated in detail how expressive writing on emotional issues can enhance well-being, highlighting the potential of emotion-focused interventions. Models such as DistilBERT, based on the knowledge distillation techniques described by Chen et al. (2019), [19] make the practical application of efficient, real-time emotion detection in therapeutic environments possible. Utilizing sentiment analysis and linguistic intelligence, therapists can achieve a better understanding of patient experiences and tailor interventions to meet specific individual needs.

The large amount of contemporary research summarized in the SLR table relates to the years 2019-2024 concerning the use of AI, NLP, and Machine Learning (ML) for the analysis of mental health. This study covered different aspects of mental health technology, like predictive analytics, conversational AI, and multimodal data processing which are shown in Table 1.

Table 1. SLR of Language Intelligence and Sustainable Mental Health Interventions in the Digital Age

Author(s)	Year	Study Focus	Methodology	Findings	Limitations
Kulkarni D. et al.	2024	Extracting mental health features using pre-trained embeddings	Pre-trained NLP models, triplet loss fine-tuning	Enhanced accuracy in extracting clinical features from health records	Does not address real-time emotion detection in therapy settings
Abubakar A.M. et al.	2024	Conversational AI for mental health therapy	Reinforcement learning chatbot implementation for mental health	Intelligent conversational AI can enhance therapeutic interventions	Lacks emotion detection capabilities crucial for nuanced therapy
Islam M.M. et al.	2024	Predictive analytics models for mental illness	Machine learning algorithms and data analytics models	Early detection of mental health disorders through	Focuses on predictive analytics, less on sentiment or emotional cues

				predictive modeling	
García-Hernández R.A. et al.	2024	Emotion recognition, affective computing, sentiment analysis	Systematic review	Advances in multimodal approaches for sentiment analysis	Inconsistent integration of newer AI techniques
Khoo L.S. et al.	2024	Multimodal mental health detection	A systematic review of passive sensing approaches	Neural networks shown effective in processing multimodal data	Data variability affects generalizability
Andrews J. et al.	2024	Ethical considerations for responsible data curation	Review of ethical frameworks for AI	Emphasized the need for robust data protection in mental health AI	Practical implementation of ethical frameworks remains challenging
Kleinberg B. et al.	2020	Measuring emotions in COVID-19-related concerns	Analysis of real-world worry dataset	Identified key emotional responses to COVID-19	The dataset is limited to specific contexts (COVID-19 pandemic)
Deady M. et al.	2022	Depression prevention via a smartphone app	Randomized controlled trial	App-based interventions reduce depression risk	Limited to short-term efficacy
Cheung R.Y. et al.	2021	Emotion regulation in expressive writing	Experimental study	Improved well-being through emotion regulation techniques	Focused on a specific cultural group
Liu T. et al.	2022	Text sentiment and self-reported depression	Analysis of text message sentiment	Correlation found between text sentiment and depression levels	Subjective nature of self-reported data
D'Alfonso S.	2020	AI in mental health interventions	Review of AI methods for mental health	AI tools, including NLP, can enhance	Ethical considerations

				therapeutic outcomes	underexplored
Lin Q. et al.	2021	AI for stress management via physiological signals	Systematic review	Potential for AI to aid in early stress detection	Focus on physiological signals may miss broader mental health indicators
Khera R. et al.	2024	AI's role in cardiovascular care	State-of-the-art review	Identified potential applications for AI in healthcare	Primarily focused on physical health, less on mental health
Pruksachatkun Y. et al.	2020	Transfer learning for NLP	Intermediate-task transfer learning	Explored conditions under which transfer learning is effective	Lacks specific applications to mental health
Chen Y.-C. et al.	2019	Knowledge distillation in BERT	Distilling knowledge for text generation	Enhanced model efficiency with minimal accuracy loss	Not directly linked to mental health interventions
Althoff T. et al.	2019	Analysis of counseling conversations	NLP applied to large-scale counseling conversation datasets	Demonstrated the potential of NLP to analyze therapeutic effectiveness	Older studies with potentially outdated techniques
Kulkarni D. et al.	2024	Extracting mental health features using pre-trained embeddings	Pre-trained NLP models, triplet loss fine-tuning	Enhanced accuracy in extracting clinical features from health records	Does not address real-time emotion detection in therapy settings

Material and Methods.

DistilBERT is a streamlined and efficient version of the original BERT (Bidirectional Encoder Representations from Transformers) model, designed to optimize performance for natural language understanding tasks while significantly reducing computational requirements. The "uncased" variation treats uppercase and lowercase letters as identical, eliminating case sensitivity and focusing instead on the contextual and semantic relationships within the text. The DistilBERT model documentation can be accessed here.

The figure 1 depicts a comprehensive data processing pipeline, covering data collection, preprocessing, feature engineering, model selection, training, evaluation, and deployment. It outlines key steps involved in transforming raw data into actionable insights, including data normalization, dimensionality reduction, and performance threshold analysis. The structured overview of the data science workflow is shown in Figure 1.

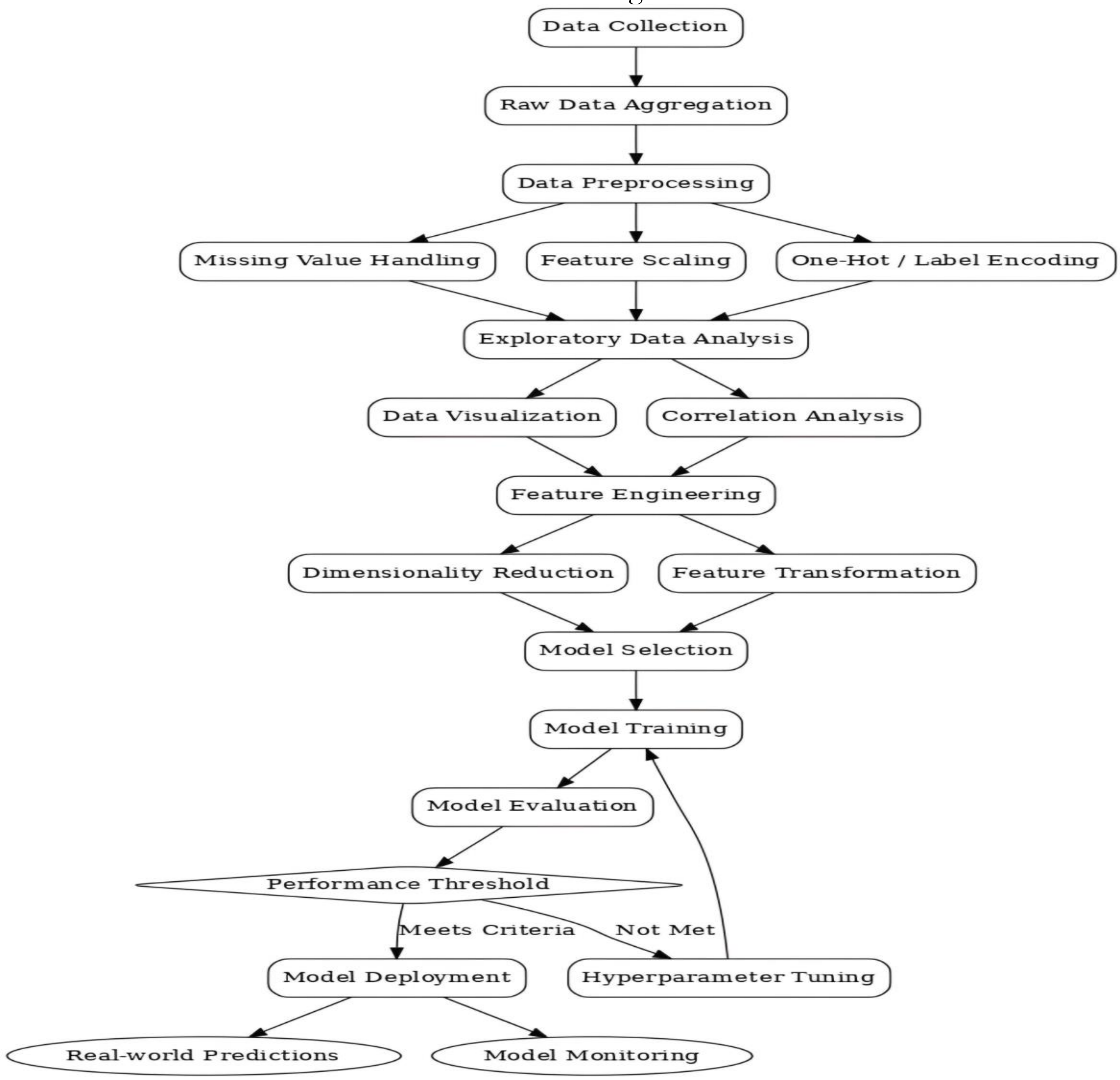

Figure 1. Workflow for Harnessing Language Intelligence in Sustainable Mental Health Interventions

Model Description.

Distil-BERT stands as a revolutionary approach to crafting efficient and compact language models, circumventing the constraints imposed by resource-intensive and bulky models. This innovation hinges on employing "distillation" during the pre-training phase, akin to equipping a student with a condensed masterclass which is shown in Figure 2. This fosters efficient knowledge transfer, further bolstered by a unique "triple loss" function, analogous to incorporating secret spices into a recipe. This function empowers the smaller model to glean deeper insights from its larger counterpart, culminating in enhanced understanding and performance. Distil-BERT achieves a striking 40% reduction in size and boasts a 60% improvement in processing speed, while remarkably preserving over 95% accuracy on language understanding benchmarks compared to the original model. This efficiency allows the potential of greater AI deployment in resource-intensive environments, hence making it easily integrated into smartwatches, smartphones, etc. Wolf et al. (2020) [20] highlighted the capabilities of transformer models in the domain of natural language understanding. In short words, Distil-

BERT represents a tremendous leap in language modeling, a more compact, faster, and easier-to-use model with state-of-the-art performance, paving the way for further incorporation of AI across all different types of computing platforms and applications.

Distillation loss. The training model was set to generate probabilities similar to the BERT base model.

Masked Language Modeling (MLM). This is a subset of the training loss related to the BERT base. At training time, after the input sentence is generated, it systematically masks out 15% of words in the input and then uses the complete masked sentence on the model to predict the words that it masked out. This is in contrast with the traditional RNNs, as well as autoregressive models like GPT, which hides future tokens and processes words one after another. It allows the model to develop a bidirectional representation of the sentence.

Cosine embedding loss. The model was further trained so that it could produce hidden states that closely resemble those from the BERT base model. In this manner, the model develops an internal representation of the English language similar to that of its teacher model, but with enhanced speed in inference and downstream applications.

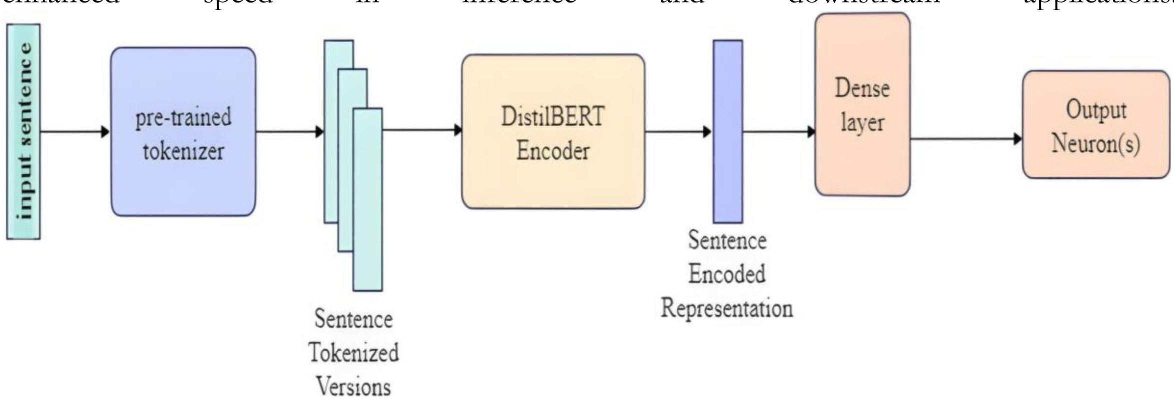

Figure 2. DistilBERT Architecture for Harnessing Language Intelligence in Sustainable Mental Health Interventions [21]

Data Collection.

The current study chooses a dataset of patient-therapist conversations on Mental Health, taken from the Kaggle site. This is one of the top resources that have a large number of datasets. This dataset has two columns, namely Context and Response. This consists of de-identified transcripts and text records collected from various therapy settings. Thus, there will be a full representation of mental health dialogues. The data were gathered strictly by Kaggle's terms of use and ethical requirements. More overt consent procedures were implemented to protect the participant's privacy and confidentiality concerning the dialogues exchanged. The Kaggle datasets, known to have natural variability and richness in conversational data, provide good grounds for exploring the subtleties in language in the therapeutic context of assessing mental health through applications of advanced natural language processing techniques and linguistic intelligence.

Data Preprocessing.

Data cleaning procedure represents a fundamental preparatory stage in our efforts to exploit the empowering nature of language within mental health research. It removes all extraneous information and allows the use of a standard framework through which innovative findings that are bound to change our very way of understanding and approaching mental health can emerge. Thus, every elegantly structured phrase and carefully crafted sentence brings us closer to the day when language becomes the fundamental tool for gaining a deeper understanding of the human psyche. **Exploratory Data Analysis.**

In the exploratory data analysis, we examined the patient-therapist dialogue dataset with an emphasis on the required attributes of the textual material. Calculating the mean and mode vocabulary sizes in both the Context and Response columns provided insights into the average linguistic richness. The distribution of conversation lengths was visualized through histograms, revealing a predominant range in both columns. A scatter plot of vocabulary sizes showcased the dispersion of linguistic features, aiding in the identification of potential patterns which are shown in Figure 2. Notably, outliers were scrutinized to understand instances of exceptional linguistic characteristics. These preliminary findings guide our subsequent NLP analyses, forming a foundational understanding for more in-depth investigations into the nuances of mental health expressions within the patient-therapist conversation context.

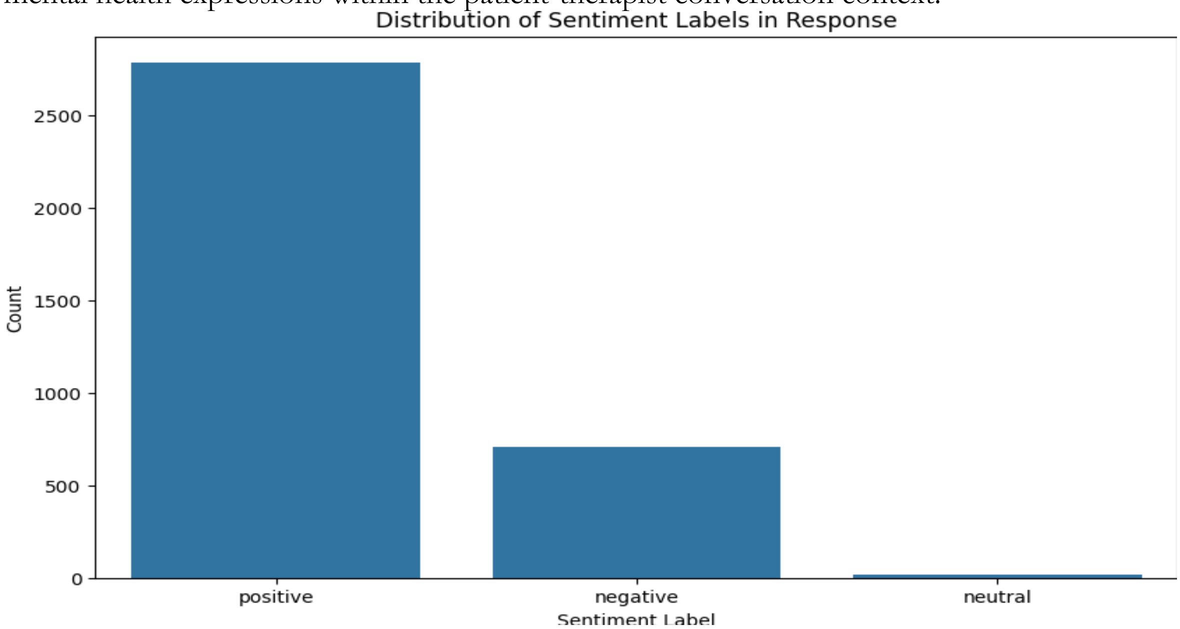

Figure 3. Graph of Distribution of sentiment labels in responses

Figure 3 displays the distribution of sentiment labels in the category Response. The blue bar on the left shows that the number of positive sentiment labels is dramatically higher than that of negative sentiment labels-the blue bar in the middle shows. The blue bar, which is farthest on the right, shows neutral sentiment label frequency that is, much lower compared to the frequencies of the positive and negative sentiment label frequencies.

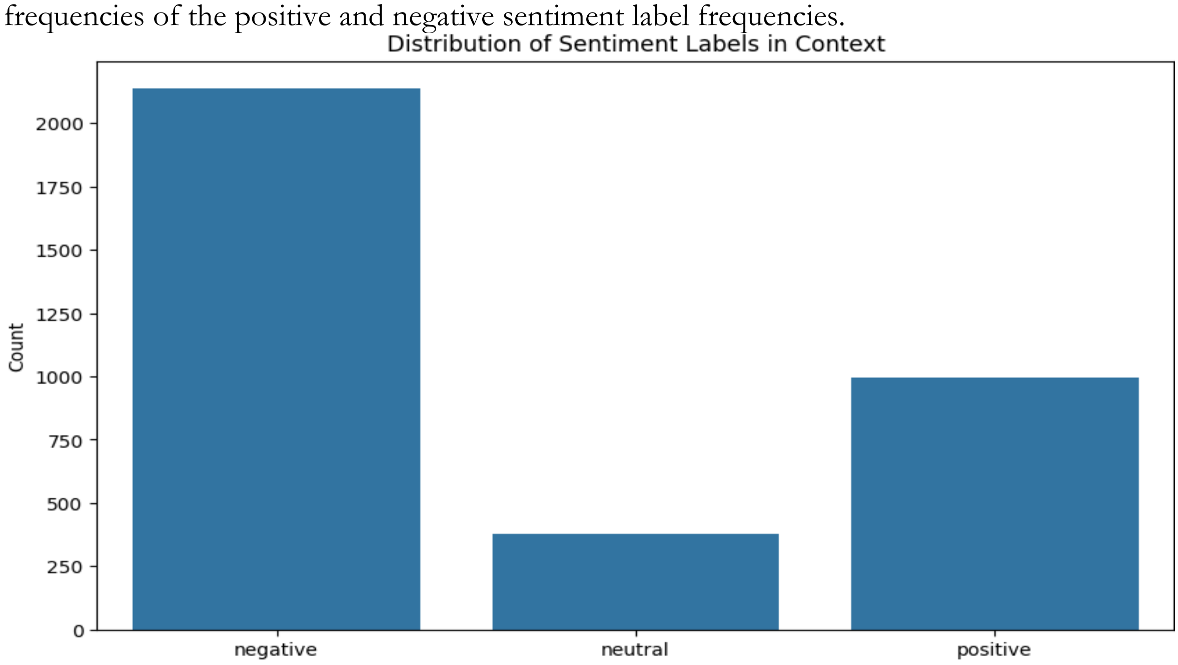

Figure 4. Graph of Distribution of sentiment label in Context

Figure 4 represents the frequency of sentiment labels found in the Context category. The left blue bar represents the number of negative sentiment labels, which is noticeably higher than the number of neutral sentiment labels shown by the blue bar in the middle. In contrast, the number of positive sentiment labels, indicated by the blue bar on the far right, is lower than both the negative and neutral sentiment labels.

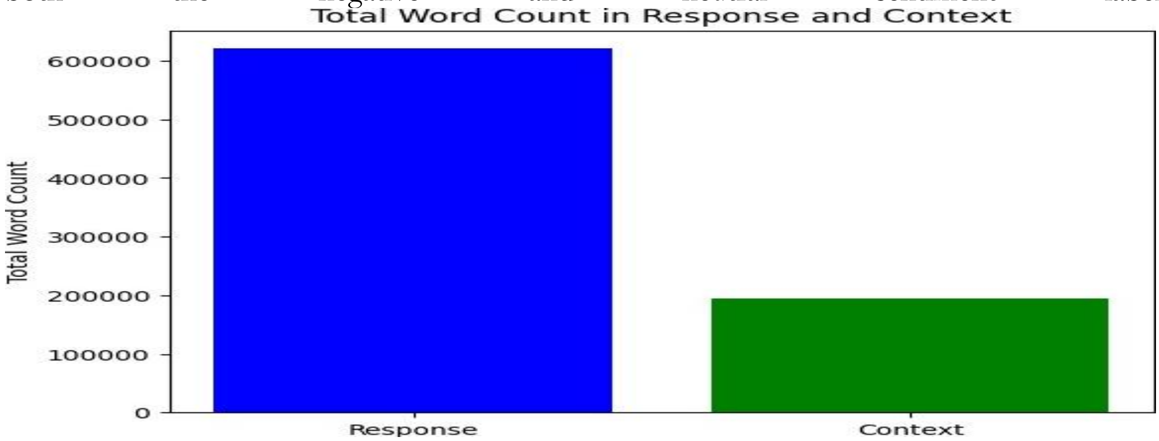

Figure 5. Graph of Word count in response and context

The graph (Figure 5) is divided into two categories. Response and Context which are shown in Figure 5. The left-hand side of the blue bar is much higher because it is the total number of words about Response, while the small green one on the right represents the total number of words for the Context.

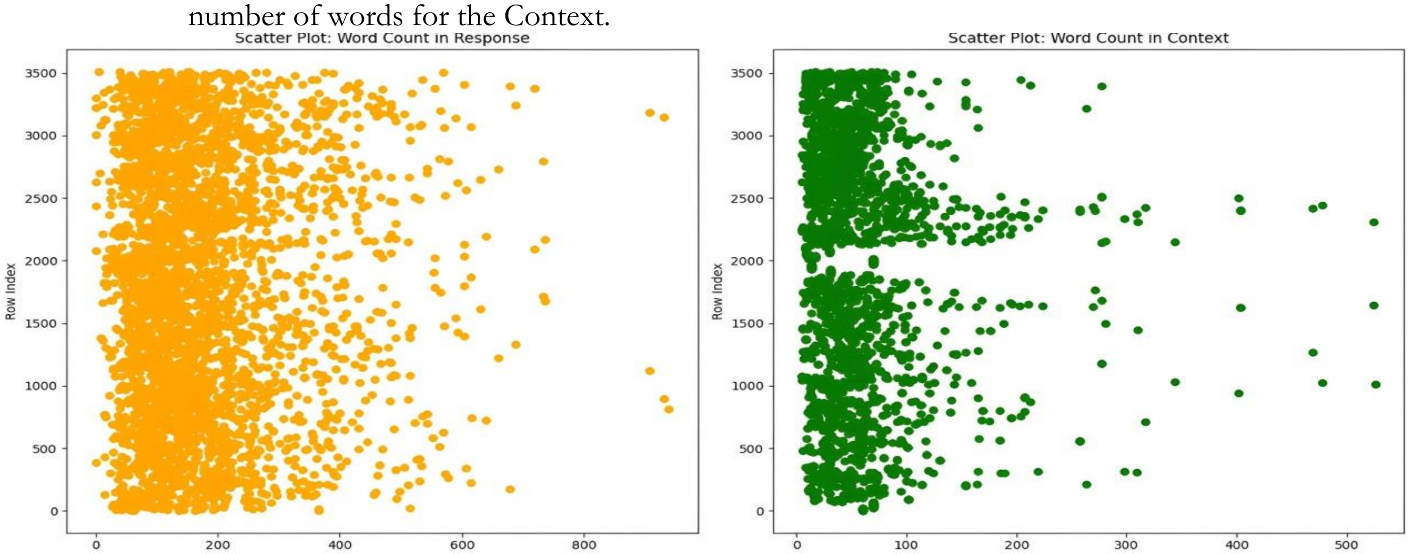

Figure 6. Scatter plot of Word count in response and context

1. The scatter plots compare the word counts of items in the dataset.
2. The left-hand plot is a distribution of word counts in the response. Most rows are less than 200 words, so the responses were overall shorter.
3. The proper plot displays the count in terms of words. More rows are less than 100 words, indicating short text in the context which is shown in Figure 6.
4. These findings depict how brief responses and contexts in the dataset have been.

Training of Model.

The distilBERT model has been utilized and further fine-tuned using the emotion dataset from Hugging Face. This fine-tuning allows the model to conduct wide-ranging sentiment analysis, yielding an output having the emotion score along with the related labels as sad, happy, love, joy, anger, fear, and surprise. Hyper-parameters used for training were.

1. Batch size = 64

2. Learning rate = 2e-5
3. Weight decay = 0.1
4. Epoch = 8

The training results are presented in the table 2.

Table 2. Training Results

Epoch	Training Loss	Validation Loss	Accuracy	F1-score
1	No log	0.199009	0.931500	0.931880
2	0.117600	0.176432	0.936000	0.935400
3	0.117600	0.167419	0.939500	0.939280
4	0.075900	0.173544	0.935500	0.934821
5	0.075900	0.194095	0.931000	0.931460
6	0.043900	0.219983	0.932000	0.932384
7	0.043900	0.230020	0.934500	0.934556
8	0.023900	0.235640	0.937000	0.936711

Table 2 describes the training statistics of the model per epoch. The model was trained within 8 epochs. After completion of the model training, we evaluated to measure accuracy, precision, recall, and F1 score [22] based on ground truth and model prediction.

Mathematical Formulas.

Metrics such as precision, recall, and F1-score provide a very rich view of model performance (Powers, 2020) [22].

F1 Score. It balances between precision and recall. The formula is.

$$F1 = 2\times((Precision\times Recall))/((Precision+Recall))$$

Precision. The number of true positive predictions out of the total number of predicted positives.

$$Precision = (True\ Positives)/\ ((True\ Positives+False\ Positives))$$

Recall. The ratio of all correctly predicted positive observations to all observed in the actual class.

$$Recall = (True\ Positives)/\ ((True\ Positives+False\ Negatives))$$

Table 3. Performance of the Proposed Model

Accuracy.0.9395				
Classification Report	**Precision**	**Recall**	**F1-score**	**Support**
Sadness	0.95	0.98	0.96	550
Joy	0.95	0.96	0.95	704
Love	0.90	0.85	0.88	178
Anger	0.97	0.93	0.95	275
Fear	0.92	0.87	0.89	212
surprise	0.83	0.90	0.86	81
Accuracy			0.94	2000
Macro avg	0.92	0.92	0.92	2000
Weighted avg	0.94	0.94	0.94	2000

In Table 3, the classification report of the model with some significant performance indicators such as precision, recall, F1-score, and accuracy of all the classes, thus giving an accurate summary of the model's ability to make precise predictions.

Confusion Matrix.

Table 4. Confusion Matrix of the Proposed Model

Sadness	538	1	1	6	4	0
Joy	6	675	16	0	4	3
Love	1	25	152	0	0	0
anger	9	3	0	256	7	0

Fear	12	1	0	2	185	12
surprise	1	5	0	0	2	73

Table 4 provides the confusion matrix generated by the model using test data, representing classification performance in the form of true positives, false positives, true negatives, and false negatives for each class. This provided rich information about model accuracy patterns as well as its error patterns.

Result and Discussion.

Based on our model and analysis of the mental health dataset analysis revealed the presence of emotions such as sadness, anger, joy, and fear in both the user queries and psychologist responses. The common emotion in context was sadness while the prominent emotion in response was anger. We also illustrated the most commonly used words within the context and response through count plot and word cloud. There was also a heat map indicating a common correlation between emotion in context and emotion in response. Which are shown in Figures 3, 4, 5, and 6.

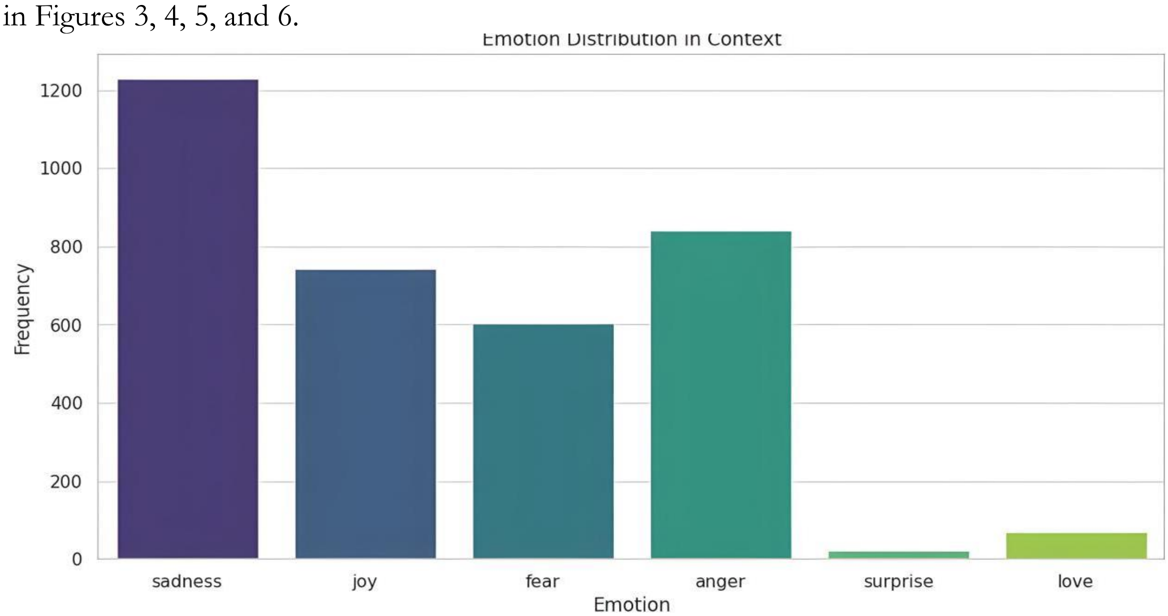

Figure 7. Graph of frequency of each emotion within the context response

Figure 7 illustrates how various emotions are spread within the Context collection. There are a few important trends here.

- Most frequent of all is grief with over 1,200 occurrences in the Context.
- Pleasure accounts for the second highest of around 800 instances.
- Fear also features suspiciously high rates of over 600 cases.
- Anger is an emotion, though it is not felt by people as often as they do with sadness and happiness.
- Surprise and love seem to appear less, surprise only around 20, and love only around 2-3 times.

The graph (figure 7) analyzes the emotional dimensions contained in the Context data and finds a model for negative emotions with themes related to sadness and fear. It claims, however, that there are positive emotions associated with happiness that exist but are of lesser importance. Using this comparative structure, these emotions can be viewed as mood indicators related to an individual's mood in addition to themes from the Context data.

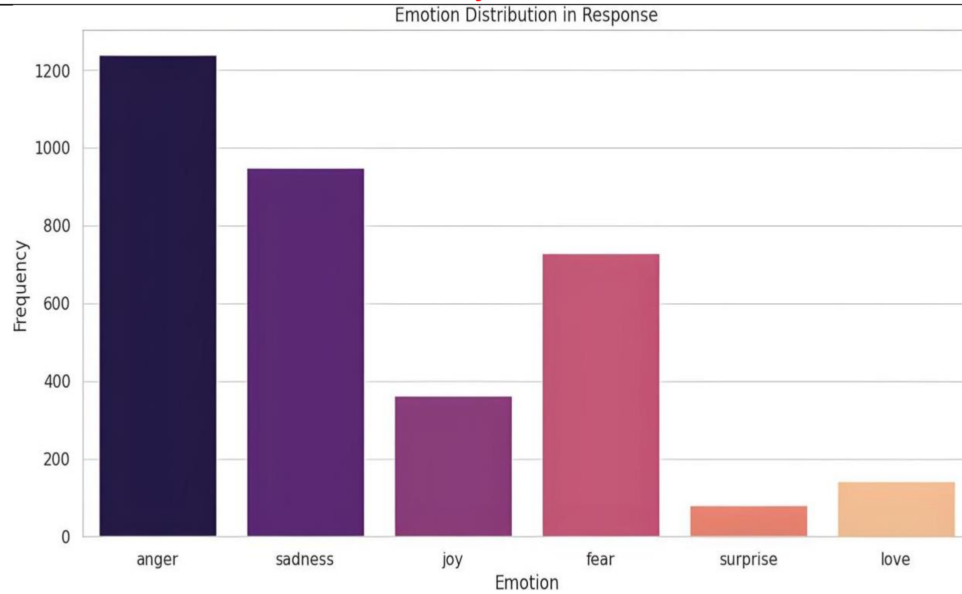

Figure 8. Graph of frequency of each emotion within the response

Figure 8 shows the number of various emotions represented in the Response data. The vertical axis is the frequency or count for each emotion, while the horizontal axis counts the unique emotions. A few important observations.

- Anger is the most common emotion, which finds more than 1,200 expressions in the Response dataset.
- The second is sorrow, which happens about 1,000 times.
- Happiness occurs less often than anger and sadness but is more common with an incidence rate of about 400.
- The least frequently occurring emotions are fear and surprise with around 300 instances of fear and around 70 instances of surprise.
- Love is the least frequent term with occurrences as few as 10-15 times in the Response data.

This graph gives a clear picture of the relative strength and distribution of different emotions expressed in the Response section of the data. It shows that negative emotions, like anger and sadness, dominate, but at the same time, it also suggests the presence of more positive emotions, such as joy, though to a lesser degree.

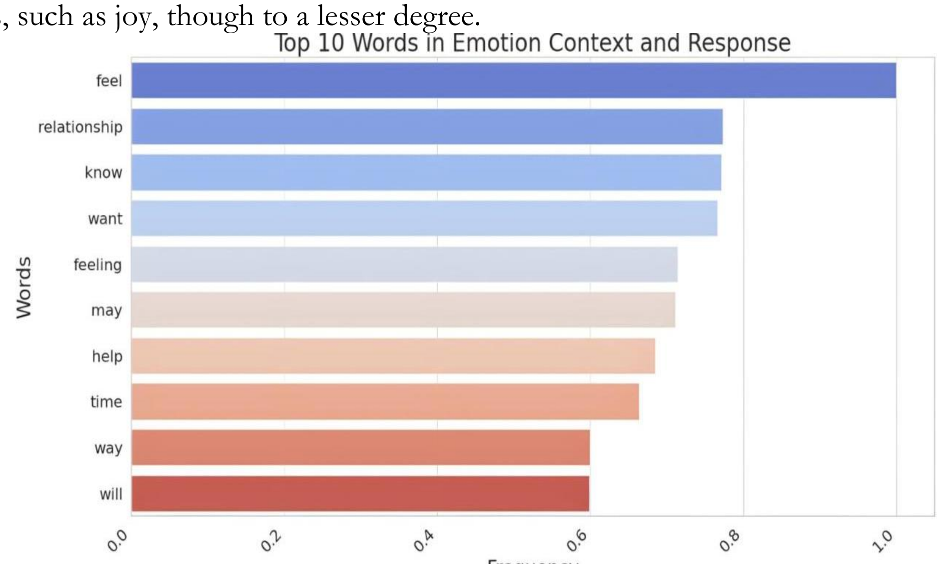

Figure 9. Graph of Top 10 Words in Emotion Context and Response

Figure 9 shows the top 10 words in the Emotion Context and Response categories which are shown in Figure 9. The words that appear most often in both categories are "feel", "relationship", "know", "want", and "feeling". The frequency of these words is shown as the length of the horizontal bars. The longest is "feel".

Several important observations.

- The headwords are mostly the vocabulary of emotions, relations, and mind, which points out that the data is centered around emotional and interrelation matters.
- Such occurrences of these words are usually more frequent in the Response category than in the Context category, as they are more common in the responses.
- The two terms that appear the least are "will" and "way," but they still have a strong presence in the data set. This graph, then, sums up to telling the reader the key language and theme present in the data being explored through Emotion analysis.

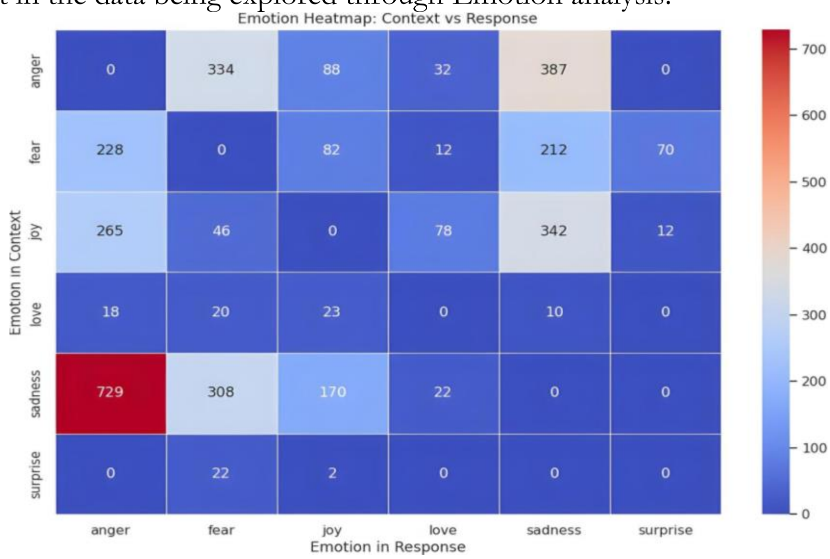

Figure 10. Heat map of emotion label within the dataset

This "Emotion Heatmap" shown in Figure 10, the intensity levels of different emotions expressed in the Context and Response datasets, comparing them against each other. Here is a concise overview of what the heatmap presents.

The vertical axis is comprised of enumerations of various categories, including anger, fear, joy, love, sadness, and surprise. The horizontal axis shows the information about Context (left) and Response (right).

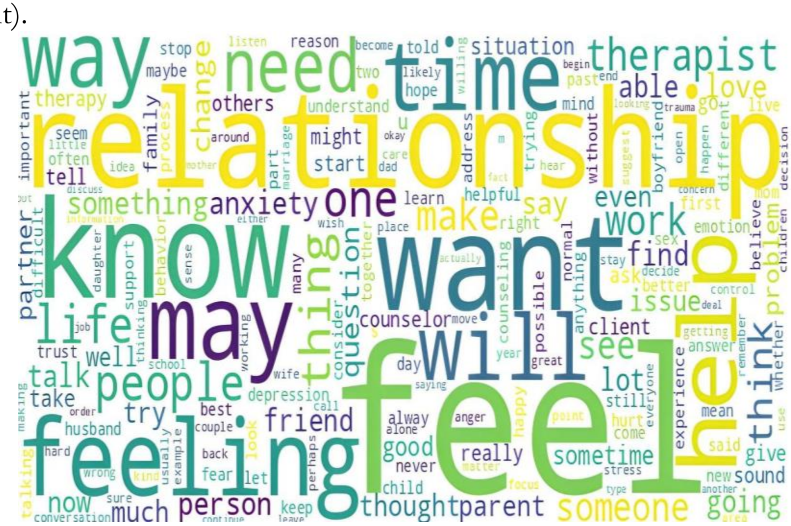

Figure 11. Representation of Therapy-Related Vocabulary and Emotional Factors in the Dataset.

The image is a representation of the very wide vocabulary and phrases that include therapy-related data in the dataset, mental health factors, emotions, and interactions between people which are shown in Figure 11. It does seem to be a picture that is full of movement and energy as it puts forward common themes and problems that could come up during therapy.

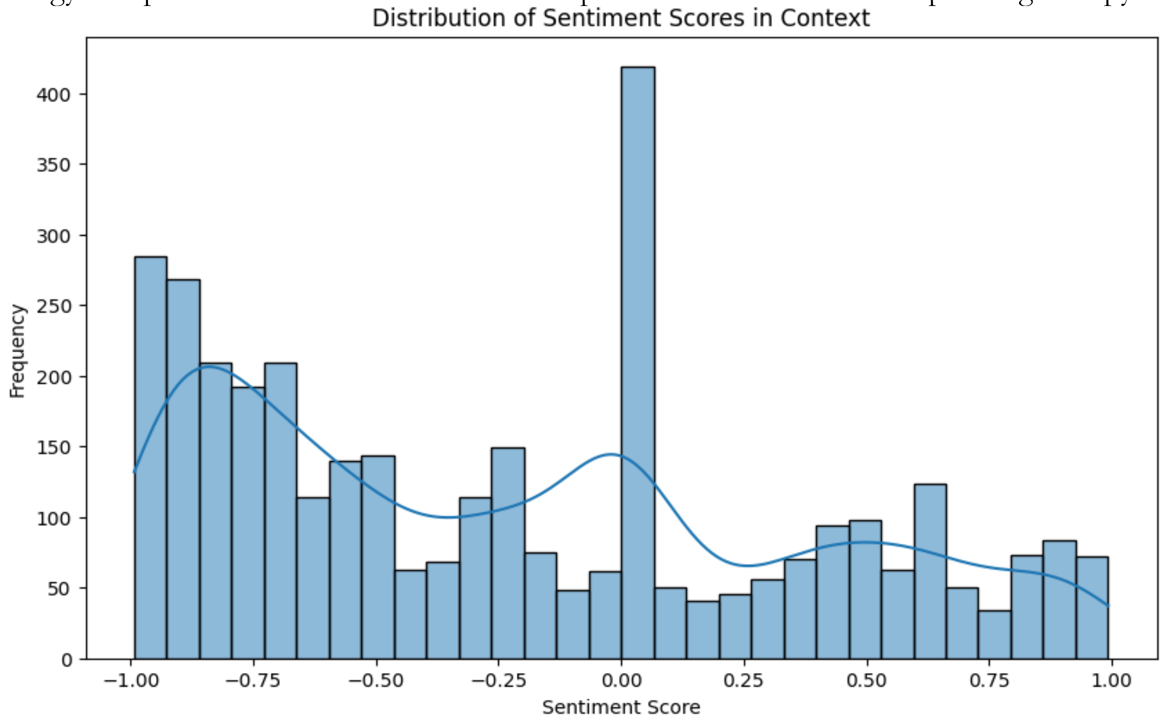

Figure 12. Graph of Distribution of Sentiment Score in Context

Figure 12 depicts the distribution of sentiment scores in the Context category. The data points spread over a range of sentiment scores with the highest density around the neutral score of 0.00. Scores spread out from strongly negative at about -1.00 to strongly positive at about 1.00, showing the diverse range of sentiments expressed in the Context.

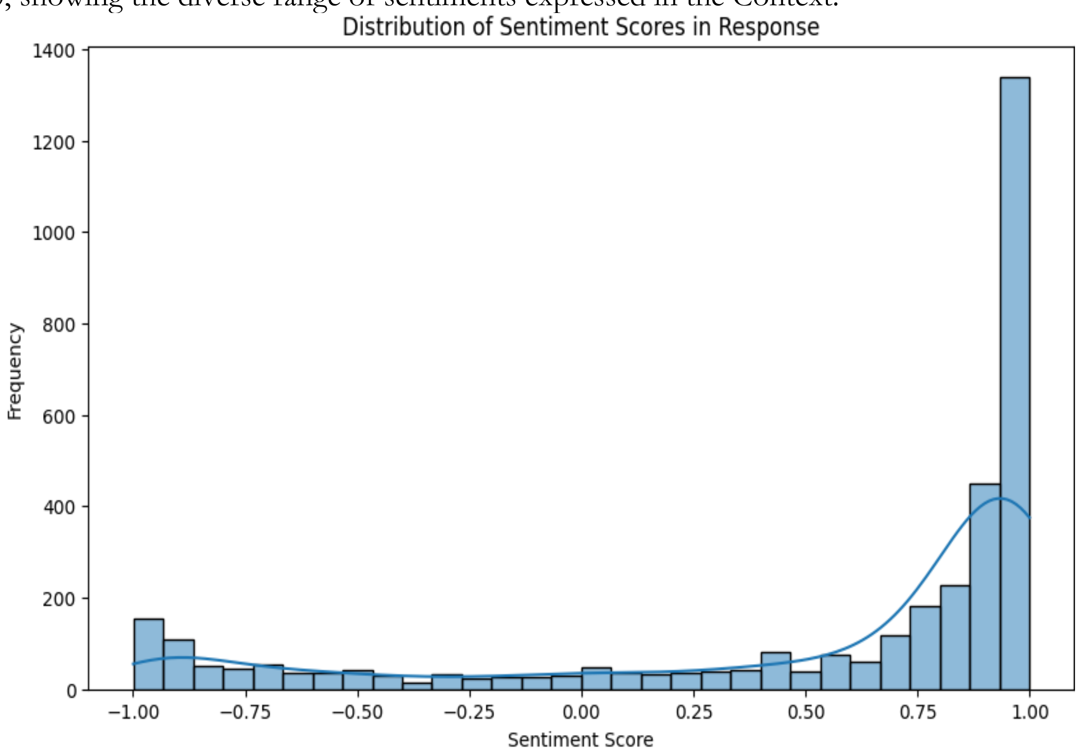

Figure 13. Graph of Distribution of Sentiment Score in Response.

Figure 13 reveals the distribution of sentiment scores categorized in Response. The data spread over the range of sentiment scores from the strongly negative value close to 1.00 up to the strongly positive value approaching the value of 1.00. The peak frequency is located near the neutral score, namely 0.00. Nevertheless, a large quantity of positive sentiment scores larger than 0.50 demonstrates a diversified range of emotions expressed within the Response data.

We used Streamlit to create a demonstration application that would display the predictive functionalities of the model and its outcomes in a more user-friendly manner which are shown in Figure 14.

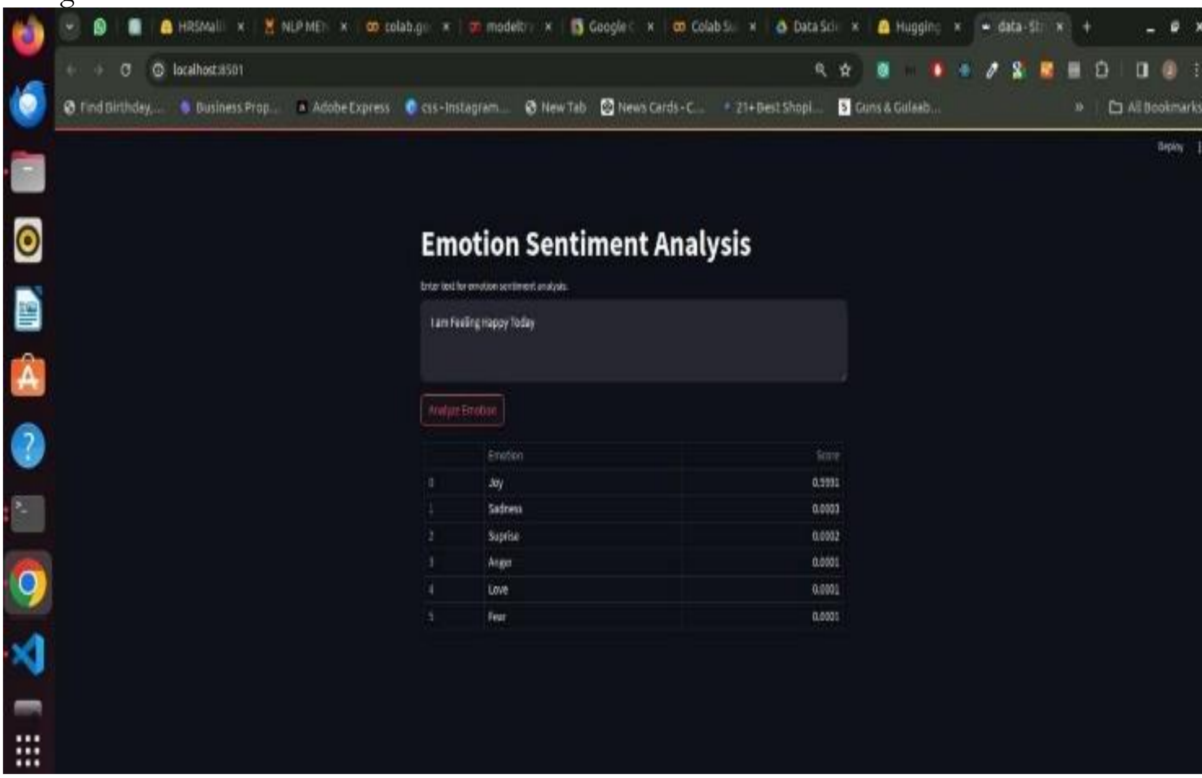

Figure 14. Employing the streamlit for the user interface as prototyping.

This study has significant practical implications for both mental health care and digital health applications. Advanced NLP methods, such as DistilBERT, enable the early detection of subtle emotional states such as sadness, joy, anger, and fear that cannot be detected by more traditional assessment methods. This method, in enabling a better understanding of the emotional state of patients, can facilitate more individualized and timely therapeutic interventions.

The integration of AI-driven tools such as DistilBERT among mental health professionals improves accuracy and efficiency in emotional assessment, making therapists target their treatments in more precise ways. Improvements may lead to quicker identification of mental health disorders, including depression, anxiety, and PTSD, thus reducing the time taken for diagnosis and improving patient outcomes.

From a practical perspective, the study will encourage the development of real-time emotional assessments through digital health applications, particularly for remote or underdeveloped areas with limited access to immediate mental health support. An application of this system can be within workplace wellness programs to facilitate employees toward better mental health and productivity as well as proactive support towards the employee's emotional needs.

Ultimately, this study highlights the potential for scaling and spreading more personalized, proactive, cost-effective mental health strategies moving beyond the traditional therapeutic modalities towards more responsive and more digital-first solutions that enhance value in patient care as well as for wider well-being.

Discussion.

The research demonstrates the potential of using advanced NLP techniques, particularly the DistilBERT model, to enhance mental health interventions by analyzing patient-therapist conversations. Through sentiment analysis and emotion detection, the study identified distinct emotional patterns in conversations, with sadness being a prevalent emotion in patient expressions, while anger often emerged in therapist responses. This study's results are in agreement with other research that has proven the role of emotional language in influencing therapeutic outcomes, particularly by acknowledging and managing negative emotions like sadness or frustration to result in dramatic improvements in patient development. The ability to capture these subtle emotional cues in real time can enhance therapy by providing therapists with concrete insight that can allow them to adapt their approaches to the patient's emotional state during sessions.

The DistilBERT model exhibits lower computational requirements with a preserved accuracy level [23] and is therefore directly applicable in digital health systems. By calibrating the model to a targeted dataset drawn from therapeutic settings, it was possible to uncover subtle emotions not necessarily verbally expressed. These include latent distress or the repression of anxiety. Transformer-based architectures like BERT and its variants have emerged to outperform traditional approaches in identifying context-aware emotions, thereby validating them for mental health applications [24]. However, the challenge lies in accurately capturing the slight emotional states in the varied cross-cultural and linguistic environments. Such a requirement emphasizes the requirement for training datasets with substantial cultural representations to enhance robustness and generalization by the models [25].

This research represented the prioritization of privacy and ethical considerations, particularly in the use of sensitive conversational data [26]. The requirement for informed consent and the use of tight data protection measures to ensure confidentiality was highlighted [27]. Mental health information is sensitive; therefore, the use of NLP-based tools in clinical settings requires adherence to ethical standards to ensure patient trust and legal compliance. It becomes more and more prominent in the landscape of AI applications in healthcare that there is a need for ethical frameworks in the reconciliation of innovation and concerns of privacy. Moreover, explainable AI models have been critical for clinical use as they enable clinicians to understand decisions regarding emotion detection [28], thus inspiring more confidence in AI-supported therapies.

This study's exploratory data analysis has revealed important insights regarding the relationships between various emotional states occurring in patient-therapist interactions and thus has aided in developing more specific strategies for therapy. For example, therapists could receive immediate emotional feedback resulting from the model's judgment [29], which enables them to become aware of a moment when a patient is unable to express particular emotions, such as sadness masked by comedy or anger that is veiling deeper fear. It has been proven that the use of emotion detection technology in therapeutic settings may enhance the outcome of treatment as it directly influences the responses and strategies of therapists. The infusion of AI-derived insights into therapeutic techniques will most significantly transform mental health care to become more flexible and responsive to the patient's shifting emotional conditions during therapy sessions.

Although the research contributes to the advancement of our understanding of how NLP can be used to analyze emotional dynamics in therapy, it also raises considerable challenges that need to be addressed. Future research should expand the datasets to include a variety of therapy settings, such as online support groups or telehealth sessions [30], to increase the applicability of the models across different contexts. Interdisciplinary collaboration is crucial in establishing ethical standards for AI use in mental health so that these technologies supplement the abilities of clinicians and not replace them. This approach is also in line with current recommendations on the deployment of AI in healthcare, where technology is seen as a tool to support, rather than supplant, human expertise [31]. It may be an opportunity to advance NLP-based tools that revolutionize mental health care with the delivery of interventions, making them more accessible, personalized, and effective.

Conclusion.

The research explores the intersection of NLP and mental health using the Distil-BERT model, particularly focusing on patient-therapist conversations. Using sentiment analysis and emotion detection, it identifies mental health patterns. Although the results are quite promising, the dataset used is biased toward therapeutic interactions and can be quite challenging to capture different mental health conditions. Privacy and ethical considerations have been the priority. Exploratory data analysis uncovers some interesting key findings, such as preponderant sadness in the context and an anger response that is prominent. Tools for visualization help understand emotional relationships. The work greatly expands mental health analysis with increased transparency and responsible AI practices to provide tailored interventions in a therapeutic context.

Conflict of Interest. In publishing this work in IJIST, the authors declare no conflict of interest.
Project Details. Nil

REFERENCES.

[1] F. Daniel Vigo, MD· Prof Graham Thornicroft, PhD· Prof Rifat Atun, "Estimating the true global burden of mental illness," The Lancet Psychiatry, vol. 3, no. 2, pp. 171–178, 2016, [Online]. Available. https.//www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(15)00505-2/fulltext
[2] A. Caliskan, J. J. Bryson, and A. Narayanan, "Semantics derived automatically from language corpora contain human-like biases," Science (80-.)., vol. 356, no. 6334, pp. 183–186, Apr. 2017, doi. 10.1126/SCIENCE.AAL4230/SUPPL_FILE/CALISKAN-SM.PDF.
[3] M. V. Koroteev, "BERT. a review of applications in natural language processing and understanding," arXiv Prepr. arXiv2103.11943, 2021, doi. https.//doi.org/10.48550/arXiv.2103.11943.
[4] S. R. B. Yada Pruksachatkun, Jason Phang, Haokun Liu, Phu Mon Htut, Xiaoyi Zhang, Richard Yuanzhe Pang, Clara Vania, Katharina Kann, "Intermediate-task transfer learning with pretrained models for natural language understanding. When and why does it work?," arXiv Prepr. arXiv2005.00628, 2020, doi. https.//doi.org/10.48550/arXiv.2005.00628.
[5] M. U. Deepali Kulkarni, Abhijit Ghosh, Amey Girdhari, Shaomin Liu, L. Alexander Vance and J. Sarkar, "Enhancing pre-trained contextual embeddings with triplet loss as an effective fine-tuning method for extracting clinical features from electronic health record derived mental health clinical notes," Nat. Lang. Process. J., vol. 6, p. 100045, 2024, doi. https.//doi.org/10.1016/j.nlp.2023.100045.
[6] and J. L. Althoff, T., K. Clark, "Large-scale analysis of counseling conversations. An application of natural language processing to mental health," Trans. Assoc. Comput. Linguist., vol. 4, pp. 463–476., 2016, doi. https.//doi.org/10.1162/tacl_a_00111.
[7] F. A. J. Md. Monirul Islam, Shahriar Hassan, Sharmin Akter and M. Sahidullah, "A comprehensive review of predictive analytics models for mental illness using machine learning algorithms," Healthc. Anal., vol. 6, p. 100350, 2024, doi. https.//doi.org/10.1016/j.health.2024.100350.
[8] D. G. Abdulqahar Mukhtar Abubakar and S. Parida, "A Reinforcement Learning Approach for Intelligent Conversational Chatbot For Enhancing Mental Health Therapy," Procedia Comput. Sci., vol. 235, pp. 916–925, 2024, doi. https.//doi.org/10.1016/j.procs.2024.04.087.
[9] D. R. andKlinge O. V.-C. Rosa A. García-Hernández , Huizilopoztli Luna-García , , José M. Celaya-Padilla , Alejandra García-Hernández , Luis C. Reveles-Gómez , Luis Alberto Flores-

Chaires , J. Ruben Delgado-Contreras, "A Systematic Literature Review of Modalities, Trends, and Limitations in Emotion Recognition, Affective Computing, and Sentiment Analysis," Appl. Sci., vol. 14, no. 16, p. 7165, 2024, doi. https.//doi.org/10.3390/app14167165.
[10] C. Y. C. and R. M. Lin Sze Khoo ,Mei Kuan Lim, "Machine Learning for Multimodal Mental Health Detection. A Systematic Review of Passive Sensing Approaches," Sensors, vol. 24, no. 2, p. 348, 2024, doi. https.//doi.org/10.3390/s24020348.
[11] Dp. Rohan Khera MD, MS, Evangelos K. Oikonomou MD, J. W. P. Girish N. Nadkarni MD, MPH, Jessica R. Morley PhD, P. Atul J. Butte MD, and E. J. T. MD, "Transforming Cardiovascular Care With Artificial Intelligence. From Discovery to Practice. JACC State-of-the-Art Review," J. Am. Coll. Cardiol., vol. 84, no. 1, pp. 97–114, 2024, doi. https.//doi.org/10.1016/j.jacc.2024.05.003.
[12] J. T. A. Andrews, D. Zhao, W. Thong, and A. Modas, "Ethical considerations for responsible data curation," Adv. Neural Inf. Process. Syst., 2023, doi. 10.48550/arXiv.2302.03629.
[13] M. Deady et al., "Preventing depression using a smartphone app. a randomized controlled trial," Psychol. Med., vol. 52, no. 3, pp. 457–466, Feb. 2022, doi. 10.1017/S0033291720002081.
[14] S. D'Alfonso, "AI in mental health," Curr. Opin. Psychol., vol. 36, pp. 112–117, Dec. 2020, doi. 10.1016/J.COPSYC.2020.04.005.
[15] L. H. U. Tony Liu, Jonah Meyerhoff, Johannes C Eichstaedt, Chris J Karr, Susan M Kaiser, Konrad P Kording, David C Mohr, "The relationship between text message sentiment and self-reported depression.," J. Affect. Disord., vol. 302, pp. 7–14, 2022, doi. 10.1016/j.jad.2021.12.048.
[16] Q. Lin, T. Li, P. M. Shakeel, and R. D. J. Samuel, "Advanced artificial intelligence in heart rate and blood pressure monitoring for stress management," J. Ambient Intell. Humaniz. Comput., vol. 12, no. 3, pp. 3329–3340, Mar. 2021, doi. 10.1007/S12652-020-02650-3/METRICS.
[17] M. M. Bennett Kleinberg,Isabelle van der Vegt, "Measuring emotions in the covid-19 real world worry dataset," arXiv Prepr. arXiv2004.04225, 2020, [Online]. Available. https.//arxiv.org/pdf/2004.04225
[18] R. Y. M. Cheung, H. S. Chung, and M. C. Y. Ng, "Expressive Writing and Well-Being in Chinese Emerging Adults. Is Emotion Regulation an Underlying Mechanism?," https.//doi.org/10.1177/2167696818824719, vol. 9, no. 6, pp. 679–689, Jan. 2019, doi. 10.1177/2167696818824719.
[19] J. L. Chen, Yen-Chun Gan, Zhe Cheng, Yu Liu, Jingzhou, "Distilling knowledge learned in BERT for text generation," arXiv Prepr. arXiv1911.03829, 2019, [Online]. Available. https.//com-mendeley-prod-publicsharing-pdfstore.s3.eu-west-1.amazonaws.com/18d1-CC-BY-2/10.18653/v1/2020.acl-main.705.pdf?X-Amz-Security-Token=IQoJb3JpZ2luX2VjEKT%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FwEaCWV1LXdlc3QtMSJHMEUCIQDEuYOI9e8oyolRd862kWVsJMEJ1I%2Fhljtgn
[20] A. R. Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, "Transformers. State-of-the-Art Natural Language Processing," ACL Anthol., pp. 38–45, 2020, doi. 10.18653/v1/2020.emnlp-demos.6.
[21] and A. A. A. Hadeer Adel, Abdelghani Dahou, Alhassan Mabrouk, Mohamed Abd Elaziz, Mohammed Kayed , Ibrahim Mahmoud El-Henawy, Samah Alshathri, "Improving Crisis Events Detection Using DistilBERT with Hunger Games Search Algorithm," Mathematics, vol. 10, no. 3, p. 447, 2022, doi. https.//doi.org/10.3390/math10030447.
[22] D. . Powers, "Evaluation. from precision, recall and F-measure to ROC, informedness, markedness and correlation," arXiv Prepr. arXiv2010.16061, 2020, [Online]. Available. https.//arxiv.org/pdf/2010.16061

[23] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT. smaller, faster, cheaper and lighter," arXiv Prepr. arXiv1910.01108, 2019, [Online]. Available. https.//arxiv.org/pdf/1910.01108

[24] V. Kokane, A. Abhyankar, N. Shrirao, and P. Khadkikar, "Predicting Mental illness (Depression) with the help of NLP Transformers," 2nd IEEE Int. Conf. Data Sci. Inf. Syst. ICDSIS 2024, 2024, doi. 10.1109/ICDSIS61070.2024.10594036.

[25] L. and S. W. Liang, "Spanish Emotion Recognition Method Based on Cross-Cultural Perspective," Front. Psychol., vol. 13, p. 849083, 2022, doi. https.//doi.org/10.3389/fpsyg.2022.849083.

[26] B. L. and N. G. Hamid Reza Saeidnia, Seyed Ghasem Hashemi Fotami, "Ethical Considerations in Artificial Intelligence Interventions for Mental Health and Well-Being. Ensuring Responsible Implementation and Impact," Soc. Sci., vol. 13, no. 7, p. 381, 2024, doi. https.//doi.org/10.3390/socsci13070381.

[27] E. and G. O. Frank, "Privacy and data protection in AI-enabled healthcare systems," Google Sch., 2024, [Online]. Available. https.//scholar.google.com/citations?view_op=view_citation&hl=en&user=VMeF1rTO1IMC&citation_for_view=VMeF1rTO1IMC.qjMakFHDy7sC

[28] D. W. Wiechmann, Elma Kerz, Sourabh ZanwarSourabh Zanwar, Yu QiaoYu Qiao Daniel, "Toward explainable AI (XAI) for mental health detection based on language behavior," Front. psychiatry, vol. 14, p. 1219479, 2023, doi. https.//doi.org/10.3389/fpsyt.2023.1219479.

[29] M. M. and A. P. B. Subramanian, J. Kim, "Digital Twin Model. A Real-Time Emotion Recognition System for Personalized Healthcare," IEEE Access, vol. 10, pp. 81155–81165, 2022, doi. 10.1109/ACCESS.2022.3193941.

[30] M. A.-J. Simon D'Alfonso, Olga Santesteban-Echarri, Simon Rice, Greg Wadley , Reeva Lederman , Christopher Miles, John Gleeson, "Artificial intelligence-assisted online social therapy for youth mental health," Front. Psychol., vol. 2, no. 8, p. 796, 2017, doi. 10.3389/fpsyg.2017.00796.

[31] J. M. Boucher, E.M., Nicole R. Harake, Haley E. Ward, Sarah Elizabeth Stoeckl, Junielly Vargas, "Artificially intelligent chatbots in digital mental health interventions. a review," Expert Rev. Med. Devices, pp. 37–49, 2021, doi. https.//doi.org/10.1080/17434440.2021.2013200.