Pertanika J. Sci. & Technol. 34 (S3): 27 - 53 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

The Sky Emotional Chatbot Companion: Proactive AI Intervention Fostering Emotional Connectivity to the Mind

Baby Vadlana, Karnam Akhil*, Tamma Bhuwan Vamsi, and Tejasree Lokireddy

Department of Computer Science and Engineering, VNR Vignana Jyothi Institute of Engineering and Technology, Vignana Jyothi Nagar, Bachupally, 500118 Hyderabad, Telangana, India

ABSTRACT

In recent years, mental health issues, emotional instability and depression have increased exponentially, affecting millions of people of all ages. The increasing percentage of mental illness, integrating with emotional instability and suboptimal perception, has signified the need for innovative and quick solutions to support and enrich emotional well-being. As traditional therapeutic practices become increasingly demanded, with the rise in advancements of technology, specifically with the increased use and evolution of Artificial Intelligence (AI), it paves the way to promising intervention pathways. To address issues of emotional instability, loneliness, and subjectivity, this paper proposes and develops the SKY Emotional Chatbot, designed to support emotional health and promote mental well-being. This proposed framework deploys transformer models and implements the PaLM (Pathways Language Model) framework to enable the chatbot to engage in interactive conversations. The chatbot's conversations will be based on users' previous conversations with it. Training and testing were conducted to ensure the chatbot achieves high accuracy and effectiveness in its conversations. Various transformer models, including DistilBERT, DistilRoBERTa, BERT, RoBERTa, ALBERT, XLNet, DeBERTa, and Electra, were used, achieving 97% precision. The SKY Emotional Chatbot is not merely reactive; it can reference and build on previous conversations, ensuring continuity in dialogue and fostering deeper emotional bonding and resonance with users. By providing a consistent and empathetic virtual companion, the chatbot plays an important role in mitigating negative feelings of isolation and despair, thereby making a proactive contribution to mental health care. The proposed framework focuses on the growing need for creative, innovative mental health solutions, the technological advancements that enable such innovations, and the potential of the SKY Emotional Chatbot to foster emotional well-being and resilience among users.

Keywords: Depression, emotional well-being, SKY emotional chatbot, transformer models

ARTICLE INFO

Article history:
Received: 03 August 2026
Accepted: 06 October 2026
Published: 20 November 2026

DOI: https://doi.org/10.47836/pjst.34.S3.02

E-mail addresses:
baby_v@vnrvjiet.in (Baby Vadlana)
akhilresearch18@gmail.com (Karnam Akhil)
bhuwanvamsi@gmail.com (Tamma Bhuwan Vamsi)
tejasreelokireddy@gmail.com (Tejasree Lokireddy)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

The origin of depression means looking at a very complex interplay of biological factors. The main part of this is the disturbance of neurotransmitter systems, notably the monoamine systems that consist of hormones like serotonin, (Agrawal et al., 2025) norepinephrine and dopamine, but also involve the glutamate and GABA systems, according to Colita et al. (2024) and Fouad et al. (2024). Like DNA methylation and histone changes, these can strongly influence the expression of genes associated with depression. Moschny (2020) shed light on stress and external factors. Additionally, prolonged exposure to stress often results in observable morphological changes in brain regions involved in emotional regulation, such as the hippocampus, prefrontal cortex, and amygdala, leading to dendritic regression and alterations in synaptic connectivity (Ren et al., 2025). The neuroimmune system plays an important role because the pathophysiological schema of depression involves immunological dysregulation and inflammatory responses, highlighting the possible neuroimmune origins of disorders (Feng et al., 2025; Tarai et al., 2019). According to Moschny (2020) and Ren et al. (2025), there is a need for a comprehensive approach to treatment and scholarly research because these biological mechanisms are not discrete; rather, they are interconnected and involved in complex interactions that create a multifaceted network that aids both the initiation and continuation of depression.

The availability of emotional and practical support through community aid improves quality of life and increases self-esteem, thereby serving as an important buffer against the impact of social support on depressive symptoms (Alsabi et al., 2022). In the context of postpartum depression, social support, peer support, and community and culturally responsive intervention strategies have been shown to be effective in improving the symptoms, maternal and child health, and health outcomes (Norazman & Lee, 2024). Similarly, among the elderly, the use of technology with social networks has the potential to significantly reduce the social and physical aspects of depression, thereby improving the effectiveness of depression treatment (Lei et al., 2023). The way and quality of social groups are the key. For example, there are correlations between variables such as network size, density, and diversity of support levels and the perception of social support, which is key to people handling depression (Taylor et al., 2022). In clinics, although the effectiveness of social support is a bit lessened, it proves very useful, especially for new parents, because of the stigma and bias which can cause worse depression symptoms (Nakamura et al., 2020). All in all, these observations show why it is important to make and keep social ties and connections and to fight back stigma to prevent and cure depression at any age and in any part of the world.

Regarding the treatment of depressive symptoms in patients with chronic health conditions, Cognitive Behavioural Therapies, or CBTs, have proven to be a benchmark in treatment and have shown effectiveness at various stages. Traditional CBTs such as Beck's

Cognitive Therapy and Behavioural Activation are supported by a substantial body of evidence that demonstrates an individual's capacity to manage their depressive episodes and prevent them, but emerging techniques in the line of Mindfulness-Based Cognitive Therapy are also hinting at promising results, according to Fouad et al. (2024). Systematic review and meta-analysis of research from Mavranezouli et al. (2024) shows that CBTs bring moderate but distinct improvements in both depressive and anxiety levels in people with chronic diseases. What's more, a network meta-analysis, conducted by Li et al. (2024), found that more complex CBT methods, such as ultra-complex CBT, are adequate, albeit with reservations, for individuals with comorbid mental disorders, especially in persons between 18 and 58 years old. This stresses that although basic CBT is efficient, more involved interventions provide more and diverse benefits, and the need to follow up with consistent treatment programmes is paramount when addressing depression in patients with chronic illness.

Moreover, authors of multiple empirical studies also argue for the importance of EI to the analysis and assessment of depression. EI includes self-understanding, emotion regulation, and interpersonal competence in managing depressive states (Fan & Ma, 2024). For instance, a study of a female subject who is suffering from depression showed her self-awareness and mechanisms of coping and behavioural intervention, which would be essential to the successful treatment of her condition, such as social withdrawal and therapy, which are important to managing her disease. (Shen et al., 2024). Also, a study carried out in a medical college examined teachers with severe depressive symptoms and found a varying correlation between EI and depressive states and suggests that heightened EI can reduce depressive symptoms (Zeidner et al., 2012). Further support is found for the claim that individuals with high levels of EI are more adept at dealing with negative emotions such as anxiety and stress, often linked with depression (Volz et al., 2024). For older folks, EI was not equal; women had more feelings of care, and this was linked to sad signs, showing that being more aware can put one at risk of being sad (Bartlett et al., 2024). For teenagers, EI did not have a connection to sadness, but there was a link to being strong, which later helped mental health, and so making EI even better could help mental health get stronger because of stronger strength (Piro et al., 2024). According to the results of these tests, emotional intelligence (EI) is a useful component of new approaches to healing and improved mental health because it plays a significant role in preventing depression in many groups. While there isn't a direct correlation between emotional intelligence (EI) and depression in teenagers, EI plays a significant role in developing strong emotional control, which can improve mental health. Therefore, improving EI is a good way to improve mental health because of its strength (Piro et al., 2024).

For example, Anisha et al. (2024) cited an AI chatbot used in the diagnosis of depression as an evidence-based way to address persistent questions about common treatments and

the need to provide sufficiently skilled mental health professionals worldwide (Anisha et al., 2024; Zhong et al., 2024). AI chatbots provide tailored support and interaction that significantly improve access to mental health care and aid in the early detection and diagnosis of depressive symptoms (Manole et al., 2024). These technologies and digitised innovations are particularly invaluable at the primary care levels, where there is a need to intervene quickly. Chatbots offer round-the-clock assistance, helping users manage mental health concerns more consistently and effectively (Kang & Hong, 2025).

Nevertheless, the interdisciplinary nature of chatbot sentiment and emotion detection raises significant ethical concerns that lie beneath the surface. The biggest problem is the risk that the system might misinterpret or oversimplify users' emotional states due to biases in the training data. These biases can lead to incorrect evaluations and inappropriate system responses. So, in addition to thorough oversight of the work by mental health specialists and adherence to well-defined regulations, it is essential to ensure that the chatbot's outputs are clinically valid, safe, and ethically sound (Denecke & Gabarron, 2024; Guo et al., 2024). Despite these difficulties, the number of empirical studies supporting the idea of AI chatbots as a source of mental and emotional well-being, behavioural change, and therapy supplementation is increasing (Ferrario et al., 2024).

Additionally, even though sentiment and emotion analysis improve chatbot functionality, biased datasets and the misinterpretation of emotional cues continue to raise ethical concerns (Denecke & Gabarron, 2024). Future developments include improving natural language processing models and incorporating AI-enhanced, clinically informed decision-support systems. Improvements in traditional strategies and diagnostic accuracy, as well as strengthening orthodox approaches, can contribute to a more efficient and equitable mental health care system, as noted by Fan (2024).

Problem Statement

Nowadays, there is enormous growth in challenges faced due to depression and emotional instability. Treating individuals with these symptoms of loneliness and emotional inexpressiveness, which are issues of subjectivity, has become a predominant task for better and healthier living. Existing models face many challenges in providing an optimal solution. To achieve this, we deploy an AI-based Chatbot and propose SKY Chatbot as an emotional support companion for continuous conversations.

Research Questions (RQ)

RQ1 : How do we assess the performance of emotions and sentiments associated with subjectivity concerns?

RQ2 : How do you tackle the challenge of Ethical Concerns while dealing with personalised intervention?

RQ3 : How do you deal with the unstructured data in CHATBOTS since the conversation is based on Text analytics?

RQ4 : How can data from previous conversations be mathematically modelled to enhance continuity and contextual awareness in dialogues?

Related Works

Denecke and Gabarron (2024) proposed chatbots for depression that use AI to assess users' emotional states and sentiments, aiming to offer empathetic responses and suitable recommendations. The ethical challenges to this integration of sentiment and emotion analysis are the possibility of misinterpreting emotional expressions among individuals experiencing depression (Yang et al., 2019), the probability of biases associated with training data, and the potential to misunderstand emotions and emotional representations among depressive individuals. The underlying research challenge, highlighted, is whether these chatbots offer the right combination of sentiment and emotion analysis to make a richer understanding of sentiment and emotion. There is ongoing demand for further research into the ethical complications of biased training data and perceptions concerning sentiment analysis, and a lack of understanding of how such issues affect the effectiveness and safety of chatbots and interventions.

Additionally, prolonged exposure to stress frequently results in observable morphological changes in brain regions involved in emotional regulation, such as the hippocampus, prefrontal cortex, and amygdala, leading to dendritic regression and alterations in synaptic connectivity (Ren et al., 2025). The neuroimmune system plays an important role because the pathophysiological schema of depression involves immunological dysregulation and inflammatory responses, highlighting the possible neuroimmune origins of disorders (Feng et al., 2025; Tarai et al., 2019). According to Moschny (2020) and Ren et al. (2025), there is a need for a comprehensive approach to treatment and scholarly research because these biological mechanisms are not discrete; rather, they are interconnected and involved in complex interactions that create a multifaceted network that aids both the initiation and continuation of depression.

Khan and Bokhari (2024) shared AI driven chatbots that offer 24 hour support, tailored contact, and early signs of disease in general practice however, there are still moral issues to consider evaluate the accuracy of the data privacy and healthcare system integration to succeed, the interest of those involved, the need for extensive training of care workers, and the sensitive ethical issues that could arise if care bots were to identify mental health issues even if the person was unaware of what the bot thought now the author informs us of a research gap that would lead to reliable ways to find out if AI chatbots could cut the rates of depression and protect the data of patients we need a better look at how AI chatbots could join current health systems.

Voon et al. (2026) demonstrated the development of a human-centred chatbot that uses the Depression Anxiety and Stress Scale-21 (DASS-21) to support the prevention of mental health disorders. The chatbot was developed with input from mental health professionals and tested by practitioners and end users. According to the findings, the chatbot is a great tool for preventing mental illness. He et al. (2020) Certain aspects of the application, like the amount of information on designated pages and the application interface, were rated "Average" by a subset of college students. Because there may be knowledge-based issues with the data and the chatbot's general features, this rating indicates future directions for improvement.

Chen et al. (2024) proposed contemporary chatbots that focus on depression. They tend to be general and less personal, merely providing general advice or expressions of emotion. PsyChatbot aims to overcome such limitations by utilising a counselling dialogue framework based on CBT to help users reflect on their negative perceptions through retrieval-based question-and-answer algorithms. Most of the current psychological counselling systems for depressed people support only English speakers, which has left out a large population of depressed Chinese individuals who need improved responsiveness in mental health support. Therefore, PsyChatbot targets the Chinese population.

Heinz et al. (2024) introduced a novel therapy chatbot named Therabot, driven by AI, designed to address mental health issues. In a controlled RCT involving 210 subjects, Therabot showed decreases in symptoms of major depressive disorder (MDD). Users reported engagement and a robust therapeutic rapport, like sessions with human therapists, suggesting that AI chatbots could serve as a viable treatment option for depression. The study primarily examined Therabot's performance in MDD, Generalised Anxiety Disorder (GAD), and Chronic Health-Related Fatigue Disorder (CHR-FED) but did not assess its applicability to other psychological conditions.

Lee and Lee (2024) studied the use of chatbot-based interpretation bias modification (CBM-I) programs to increase cognitive biases associated with depression in early adulthood. Their findings showed that those receiving CBM-I treatment experienced considerably fewer depressive symptoms and better cognitive flexibility and attention control than the control group did. The results of this study, which covers the development of digital interventions for depression and the promotion of mental well-being, are very promising. But there is only evidence that the chatbot-driven CBM-I is working, and it is still not being used to determine the effect this change has on existing conditions.

Dharrao and Gite (2024) created the TherapyBot. TherapyBot is a transformer-based model trained for mental health conversations. TherapyBot offers a safe environment for those who are unwilling to talk about their struggles. TherapyBot has been trained on an open-domain and therapist-client model, enabling it to generate empathetic responses. TherapyBot has been shown to address mental health issues with a loss of 0.29 and a

perplexity of 1.34. The study also offers suggestions for future work. Future work involves developing larger models to achieve better outcomes, applying reinforcement learning for adaptive responses, and implementing generative adversarial networks (GANs) for chatbots. The authors also suggest implementing emotion recognition.

METHODOLOGY

Proposed Architecture

In this research, SKY* Chatbot, an advanced conversational AI system, provided emotional support to individuals facing emotional difficulties. SKY is built on the Pathways Language Model (PaLM) Architecture and can be easily accessed via a web interface. A second feature is the integration of emotion recognition with personalised content recommendation. Based on the emotional state identified from the user's text, the Interests Fetching API can suggest content such as music, videos, and blogs. The intention is not only to generate a conversational reply but also to provide additional content that may help the user remain engaged during the interaction.The chatbot uses eight transformer-based models for training and testing: BERT, RoBERTa, ALBERT, XLNet, DeBERTa, DistilBERT, DistilRoBERTa, and Electra, providing accurate, high-performance comprehension of user input and responses. Thus, SKY's contribution is not based on a single language model but on integrating emotion identification, conversational history, personalised recommendations, and a conversational interface into a single framework.

As Figure 1 shows, SKY has an Interests Fetching Application Programming Interface (API) that gives personalised recommendations for content, including music, videos, and blogs, based on the user's emotional state and interests. The motive of this feature is to uplift and engage the user during their interaction (Reis, 2024). The database stores user credentials, emotional states, previous responses, and preferences, which helps keep the user experience personalised and helps the chatbot understand the user better.

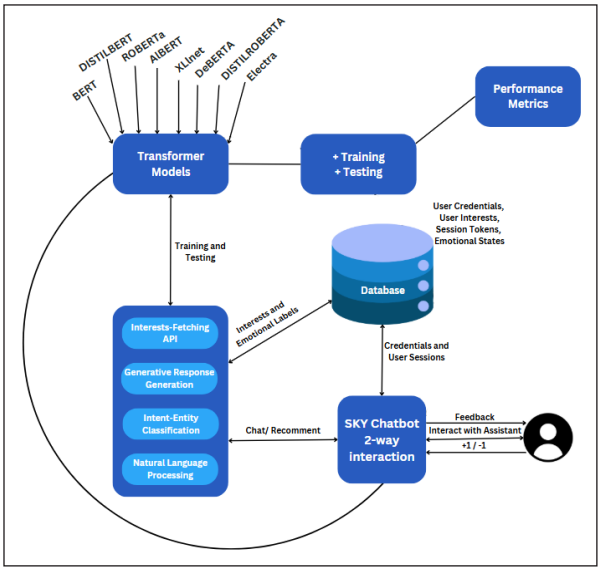


Figure 1. Architecture of the SKY chatbot

Motivation and Hypothesis of SKY Chatbot

The name SKY originates from the authors' hypothesis that the sky, with its beautiful sunrises, magical twilights, and countless stars, projects peace and wonder to anyone who looks at it. The goal is to make users feel the same comfort when interacting with the SKY Chatbot as they do when looking at the night or morning sky.

Scenario 1, as shown in Figure 2, depicts a person who begins with sadness and loneliness, burdened by emotional weight. He resorts to talking to the SKY Chatbot. Throughout, the mood shifts from sadness to happiness as you interact with SKY. He walks outside, standing under a clear, starry sky, where the stars bring hope and a sense of comfort and peace.

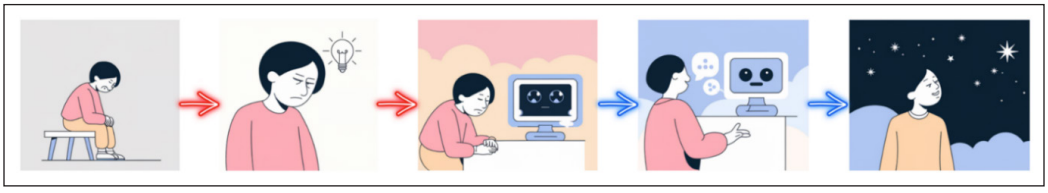


Figure 2. Scenario 1: A person using SKY Chatbot and feeling better

PaLM Model

A single, compelling model applies to several tasks and domains. Creating a new Pathways system to manage accelerator distributed computing was a big step in the right direction. Using the Pathways system, training a single model across several Tensor Processing Unit (TPU) v4 Pods was successful. PaLM is a Transformer model with just dense decoders and 540 billion parameters. Researchers evaluated PaLM using hundreds of language production and comprehension tasks and found that it often outperforms state-of-the-art few-shot performance.

PaLM demonstrates the first large-scale implementation of the Pathways system to expand training to 6144 chips, the most extensive TPU-based system configuration 31 used for training to date. The training is scaled via data parallelism at the Pod level, with traditional data and model parallelism within each Cloud TPU v4 Pod. This knowledge represents a significant increase in scale over most previous LLMs, which were either trained on a single TPU v3 Pod (GLaM, LaMDA, and more), used multiple TPU v3 Pods (Gopher) with a maximum scale of 4096 TPU v3 chips, or scaled to 2240 A100 graphics processing units (GPUs) across GPU clusters using pipeline parallelism (Megatron-Turing NLG). A mix of English and multilingual datasets, including GitHub code, novels, Wikipedia, discussions, and high-quality online articles, was used to train PaLM.

Transformer Models

Transformers are a revolutionary model architecture in NLP that operate independently using the self-attention mechanism introduced by Vaswani et al. (2017). Thus, they can process input data in parallel rather than sequentially (Zandieh et al., 2024). This design allows these architectures to capture long-range dependencies and relationships within the data better than any traditional recurrent neural network (RNN) or convolutional neural network (CNN) ever could. Achieving the highest scores across various tasks, particularly in machine translation, Transformers have loomed large, with reduced training time and computational costs and improved translation quality. To evaluate the SKY Chatbot implemented using the PaLM model, eight transformer models were widely used: BERT, ALBERT, RoBERTa, DeBERTa, DistilBERT, Electra, XLNet, and DistilRoBERTa.

Algorithm for Analysing SKY Chatbot Accuracy: A Mathematical Approach

To evaluate the accuracy of SKY Chatbot, each transformer model must follow the algorithm to obtain the best results:

Step I. Data Collection and Pre-Processing

- Collect conversation data*: D = {(u_i, c_i, t_i)},* where *u_i* is user input, *c_i* is chatbot response, and *t_i* is timestamp.
- Text Cleaning and tokenising data: *D′ = {((u_i)_t, (c_i)_t)}*, where *_t* denotes tokenisation.
- Split data: *D′ = {D_train, D_val, D_test}*.

Step II. Model Training and Evaluation

- Define a loss function *L* (e.g., cross-entropy for classification tasks).
- Initialise model parameters θ.
- For epoch in epochs:
- For *(u, c)* in *D_train*:
- Compute prediction $\hat{y}$ using the model with parameters θ.
- Calculate loss: loss = *L(ŷ, c)*.
- Update parameters: $\theta = \theta - \eta * \nabla_\theta$ loss, where η is learning rate.
- Evaluate on *D_val* and adjust hyperparameters if necessary.

Step III. Integrating Conversation History

a) Feature Engineering:

- Define a function *H* to extract historical features:
- *H(u_i, c_i, t_i) = (u_i-N, ..., u_i-1, c_i-N, ..., c_i-1)*, where *N* is the history window size.
- Augment training data: *D_train_hist = {(H(u_i, c_i, t_i), (u_i)_t, (c_i)_t)}*.

b) Model Adaptation:

- To modify the model for the functioning of the accept historical features as input: $\hat{y} = f(\theta, (u_i)_t, (c_i)_t, H(u_i, c_i, t_i))$.
- Re-train the model on *D_train_hist*.

Step IV. Iterative Refinement

a) Continuous Monitoring Evaluation:

- Periodically assess the model on various data sources: *D_test_new*.
- Calculate metrics: $M(D_test_new, \hat{y})$.

b) Model Enhancement:

- If $M(D_test_new, \hat{y})$ falls below a threshold:
- Collect more data: *D_new*.
- Re-train the model on $D_train \cup D_new$.
- Adjust hyperparameters or model architecture.

Reproducibility and Experimental Procedure

The experimental procedure consists of four main stages: data preparation, model training, conversation-history integration, and performance evaluation.

First, conversational data are collected as user input, chatbot responses, and corresponding interaction information. The text is cleaned and tokenised before being divided into training, validation, and test subsets.

During model training, the selected transformer architecture processes the textual input and generates the predicted class. Cross-entropy loss is used for classification, and the model parameters are updated iteratively during training. Performance on the validation set is monitored to assess the model's behaviour before final evaluation. Conversation history is incorporated by considering a window of previous user inputs and chatbot responses. These historical interactions provide contextual information that is combined with the current interaction. This enables the system to preserve short-term conversational continuity. After training, the models are evaluated using accuracy, precision, recall, and F1-score. Confusion matrices and emotion-wise results are also examined to identify differences in performance among emotion categories.

The reported implementation was executed on a system equipped with an NVIDIA GeForce RTX 4080 GPU and 32 GB RAM. Reporting the preprocessing procedure, data split, model configuration, evaluation measures, and computing environment improves the transparency of the experimental setup and supports the study's reproducibility.

RESULTS AND DISCUSSIONS

The whole SKY Chatbot has been implemented as a system with hardware requirements that include a GeForce RTX 4080 GPU and 32 GB of RAM. The SKY Chatbot integrates with a web interface, making it easily accessible to everyone.

The SKY Chatbot and User Interaction

Figure 3 shows the home page of the SKY Chatbot. Figures 3 and 4 show the interface and user interaction. SKY initially welcomes the users with a greeting message, after which the conversation begins. Users can chat with the chatbot about their emotional weights, and SKY will help them lift them off using its trained ability to understand emotions.

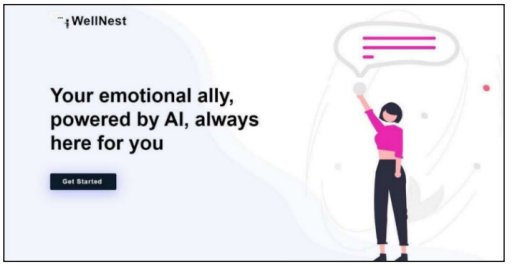


Figure 3. Home page of the SKY Chatbot

Users can chat with the chatbot about their emotional weights, and SKY will help will help them lift them off using its trained ability to understand emotions as shown in Figure 5.

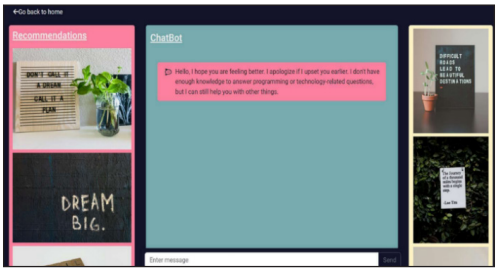


Figure 4. SKY Chatbot page

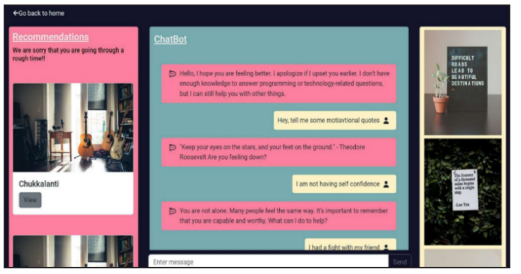


Figure 5. SKY Chatbot responses

Accuracy, F1-Score, Precision, and Recall Evaluation Metrics of All Models

Table 1 represents the metrics: the Accuracies, Precisions, Recalls, and F1-Scores, of all eight Transformer models and the number of epochs each model went through.

DeBERTa clearly outshines all other models, delivering the highest performance across accuracy (0.937), F1-score (0.935), precision (0.9384), and recall (0.937). BERT and ALBERT trail closely, with BERT slightly better at precision (0.9335) but ALBERT showing a more balanced F1-score (0.9323). In contrast, lighter models like DistilBERT and Electra offer efficiency but fall short in performance, with accuracy, F1-scores, precision, and recall all hovering around 0.926 to 0.928. This comparison highlights that while smaller models are resource-efficient, DeBERTa's robust precision and recall make it the clear winner.

Table 1
Summary of statistics of all models

Model	Epochs	Accuracy	Precision	Recall	F1-Score
DistilBERT	6	0.9275	0.926967704	0.9275	0.926177239
DistilRoBERTa	7	0.9305	0.930206809	0.9305	0.928776377
BERT	5	0.933	0.933546594	0.933	0.932790438
RoBERTa	9	0.929	0.929603594	0.929	0.927191908
ALBERT	10	0.9325	0.933316551	0.9325	0.932296927
XLNet	7	0.93	0.932712232	0.93	0.93008278
DeBERTa	8	0.937	0.938377823	0.937	0.934958303
Electra	7	0.9265	0.927955736	0.9265	0.924947401

F1-Score, Precision, and Recall Across Classes

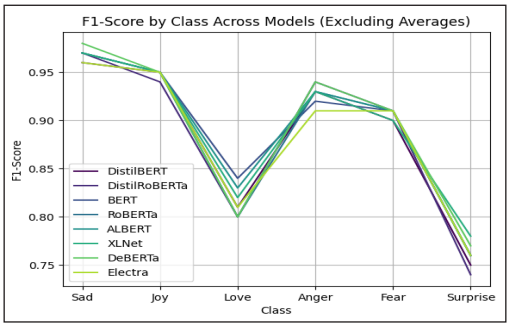


Figure 6. F1-Score per emotion

As shown in Figures 6, 7 and 8, DeBERTa and ALBERT regularly top the list in emotion classification for emotions such as Joy, Fear and Surprise, achieving strong F1-scores, precision and recall. Joy, which is the simplest to detect due to its linguistic signals, achieves the best performance across all models, whereas Love falls short, being the toughest emotion to classify correctly. DeBERTa stands out notably in precision for Fear, minimising positives, and in recall, for Surprise, successfully identifying even subtle occurrences. While DistilBERT and Electra struggle with precision and recall for emotions like Love and Fear, DeBERTa's ability to generalise well, even with imbalanced data, makes it the most reliable model overall.

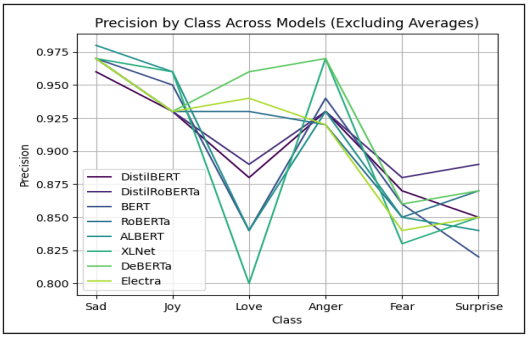


Figure 7. Precision per emotion

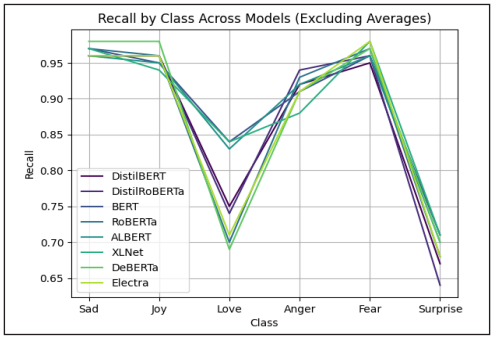


Figure 8. Recall per emotion

Macro and Weighted F1-Score, Precision, and Recall Across All Models

Macro Average F1-Score, Precision, and Recall metrics measure the average performance across all emotional classifications. Each category is treated as equally important, regardless of its frequency in the dataset. Weighted Average F1-Score, Precision, and Recall metrics calculate the mean performance across different classifications while adjusting for the frequency with which instances appear in each category. This method addresses class imbalances by prioritising classes that are more frequently represented.

Figures 9, 10, and 11 show that the transformer model with the best macro scores is XLNet. The transformer model with the best weighted scores is DeBERTa.

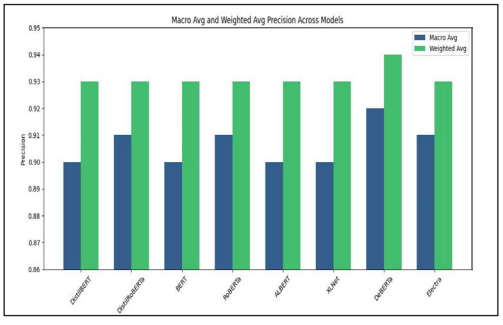


Figure 9. Macro average and weighted average Precision for transformer models

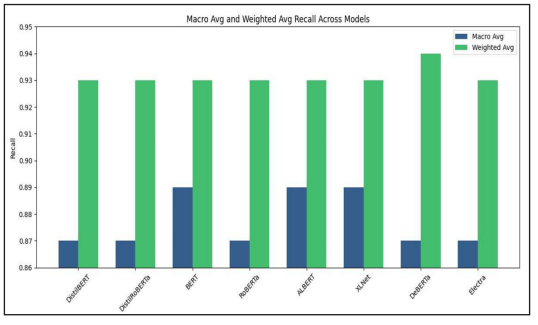


Figure 10. Macro average and weighted average Recall scores for transformer models

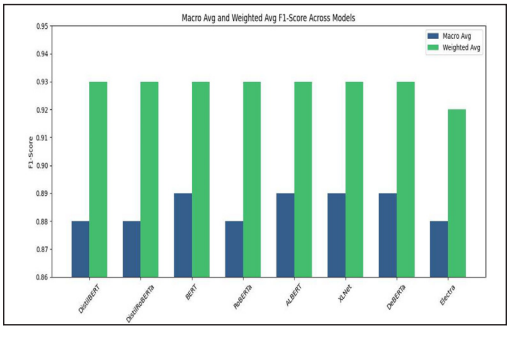


Figure 11. Macro average and weighted average F1-Scores for transformer models

BERT

Based on the information from Devlin et al. (2019), BERT's power is attributed to the bidirectional design that enables it to get the context not only from the left side but also from the right, whereas the traditional models are unidirectional. The incorporation of MLM during its pretraining phase enables it to develop a deep and quite complex understanding of the language, which is the main reason it is very effective at differentiating the emotions of Joy and Sadness. Nevertheless, as a matter of fact, BERT, as a machine learning model in

Figure 12, is less able to identify the emotions of Love and Surprise, which is probably due to the subtle and overlapping nature of these linguistic cues. Here, in Figure 13, it is evident that the entire procedure of its training is heavy on the resources and takes a longer time to converge as compared to the likes of the lightweight model DistilBERT, though the outcome in terms of high accuracy and strong F1-scores in most categories is what really counts. Slight overfitting can be noticed through the validation loss, which is somewhat up and down.

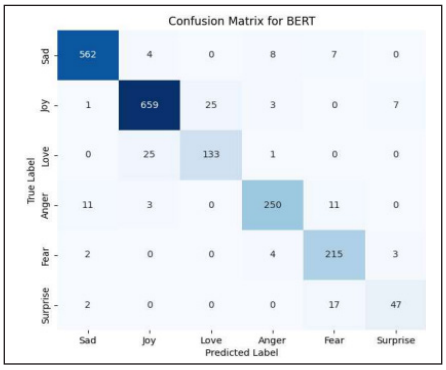


Figure 12. Confusion matrix for BERT

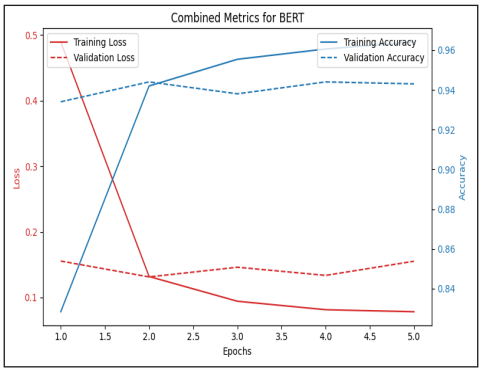


Figure 13. Epochs metrics of BERT

ALBERT

As per Lan et al. (2019), ALBERT stands out for its efficient architecture. As shown in Figure 14, ALBERT excels at classifying emotions such as Joy and Sadness, achieving high precision and recall. Its training process is efficient, with rapid improvements in accuracy and early loss stabilisation, indicating strong generalisation capabilities, as demonstrated in Figure 15. Although ALBERT struggles with classifying more ambiguous emotions like Love, its weighted and macro average F1-scores are among the highest, reflecting consistent performance across emotional categories.

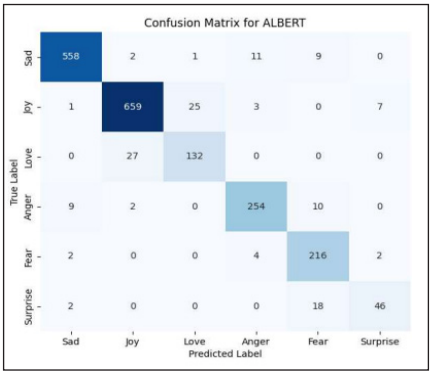


Figure 14. Confusion matrix for ALBERT

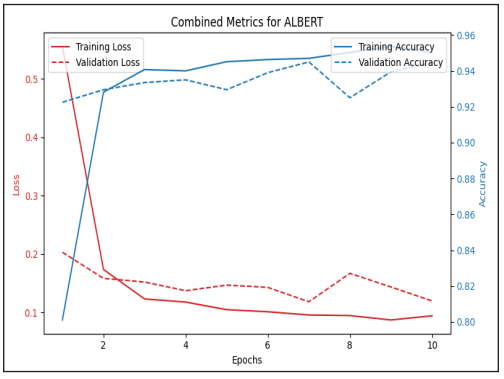


Figure 15. Epochs metrics of ALBERT

RoBERTa

According to Liu et al. (2019), RoBERTa enhances the BERT framework through refining training techniques. Expanding datasets demonstrate that BERT's initial training was suboptimal and could be outperformed by better training approaches. By removing limitations such as Next Sentence Prediction (NSP) and augmenting the training data, RoBERTa attains top-tier performance on benchmarks including GLUE, RACE and SQuAD. Although it performs admirably in classifying emotions such as Joy, Fear, and Surprise, it has difficulty with Love, showing 48 misclassifications, as depicted in Figure 16. RoBERTa demonstrates efficient training by rapidly attaining peak validation accuracy, yet it exhibits overfitting after the fifth epoch, as shown in Figure 17. Its F1-scores remain very competitive, surpassing BERT in categories, especially in accurately identifying complex emotions like Fear, with reduced false positives. Despite occasional inconsistencies in validation loss, RoBERTa demonstrates robust generalisation across diverse emotional categories, effectively balancing accuracy and efficiency.

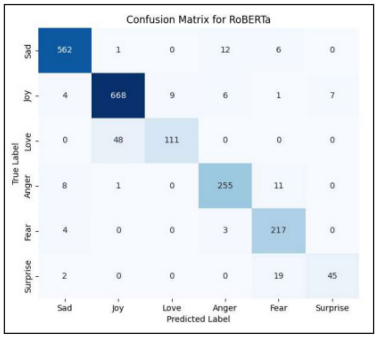


Figure 16. Confusion matrix for RoBERTa

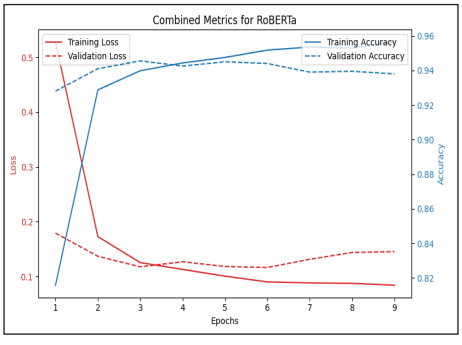


Figure 17. Epochs metrics of RoBERTa

DeBERTa

As He et al. (2020) noted, DeBERTa improves on earlier models such as BERT and RoBERTa with two main innovations: a disentangled attention mechanism and a better masked decoder. By using two separate vectors for each token to distinguish semantic meaning from positional information, DeBERTa enhances attention calculations by accounting for both content and location. Its improved mask decoder includes absolute positions, which helps with masked token predictions during pretraining. With the highest macro and weighted-average F1-Scores, DeBERTa sets a standard for emotion classification, demonstrating impressive precision and recall, particularly for more difficult emotions such as Fear and Surprise. With the highest macro and weighted - average F1-Scores as shown in Figure 18, DeBERTa sets a standard for emotion classification as shown in Figure 19, demonstrating impressive precision and recall, particularly for more difficult emotions such as fear and surprise.

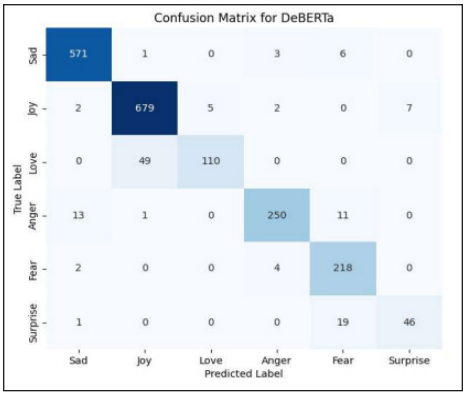


Figure 18. Confusion matrix for DeBERTa

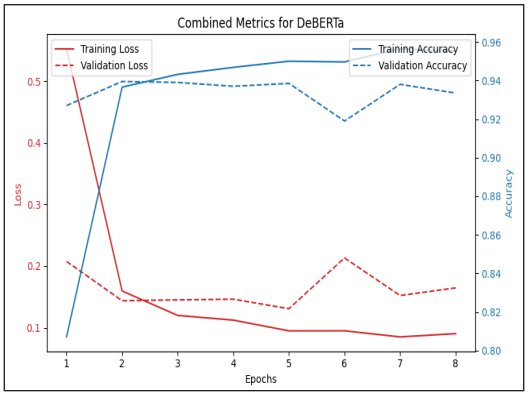


Figure 19. Epochs metrics of DeBERTa

DistilBERT

As Figure 20 illustrates, DistilBERT performs exceptionally well in classifying Joy (668 true positives). It is a very effective option for computationally limited tasks because, as shown in Figure 21, its training process is faster, with a sharp decline in training loss and a quicker plateau in validation accuracy. DistilBERT achieves impressive F1-scores and performs well across most models despite its smaller size. However, it is not as effective as larger models such as DeBERTa, particularly in handling emotional categories like Love and Fear. DistilBERT's recall for uncommon emotions like surprise is lower, underscoring its limitations with underrepresented emotions, but it maintains a relatively high precision, minimising false positives in Joy and Sadness.

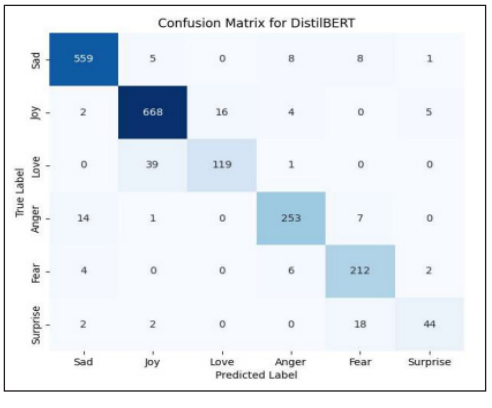


Figure 20. Confusion matrix for DistilBERT

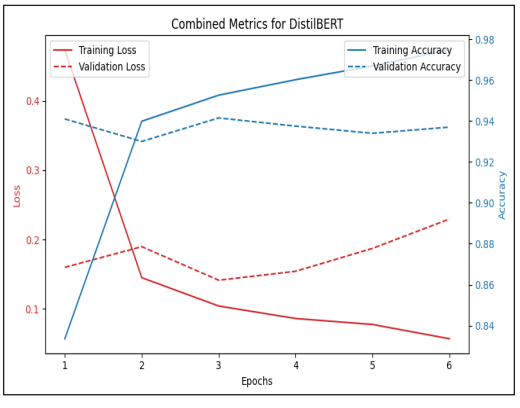


Figure 21. Epochs metrics of DistilBERT

Electra

According to Clark et al. (2020), ELECTRA changes the game by introducing a novel transformer architecture that uses both a generator and a discriminator, thereby forming a dual-transformer model for pretraining. In a masked language modelling (MLM) setting, the generator replaces tokens in the sequence, whereas the discriminator identifies which tokens have been replaced; thus, it performs a replacement-token detection task. The new method is more productive in training than the standard MLM because training proceeds more quickly. As a matter of fact, Figure 22 reveals that ELECTRA is very good at Joy prediction (668 true positives), but it confuses Love (44 misclassifications) and occasionally confuses Surprise (45 false positives). The model learning fast during the initial training epochs is depicted in Figure 23, where high training accuracy was achieved. However, the model exhibits poor generalisation, as evidenced by a high validation loss and an early plateau in validation accuracy. Despite the model's training technique being quite beneficial, ELECTRA's macro-average F1 scores are lower than those of other models. It is strong in Joy and Sadness but weak in rare emotions, and its low precision in misclassifying surprise and love is its major problem. Its recall is good but not as strong as the other models, suggesting it lacks the ability to detect subtle emotional cues.

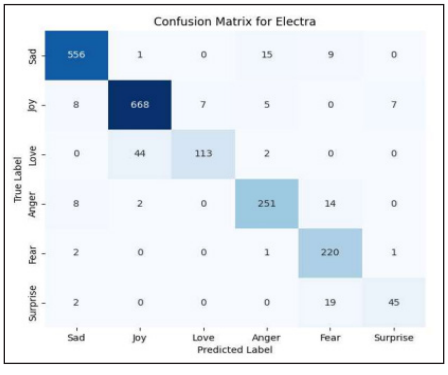


Figure 22. Confusion matrix for Electra

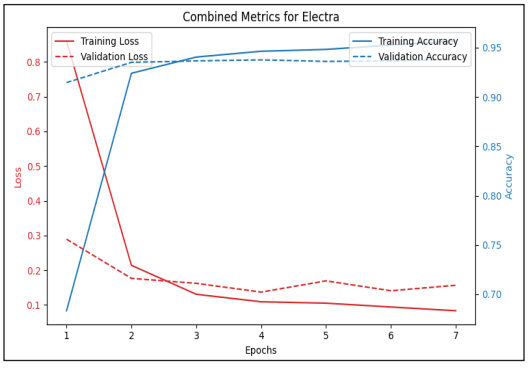


Figure 23. Epochs metrics of Electra

DistilRoBERTa

As discussed by Sanh et al. (2019) DistilRoBERTa, a compressed variant of the RoBERTa model, achieves a smaller footprint and lower complexity via knowledge distillation while preserving robust performance across several natural language understanding tasks, particularly in real-time scenarios such as text classification, sentiment detection, and content moderation. Its confusion matrix in Figure 24 highlights accuracy for Joy (669 true positives) and a few errors, for Sadness (6 misclassifications) as shown in Figure 25.

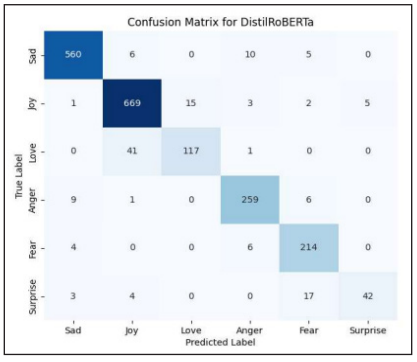


Figure 24. Confusion matrix for DistilRoBERTa

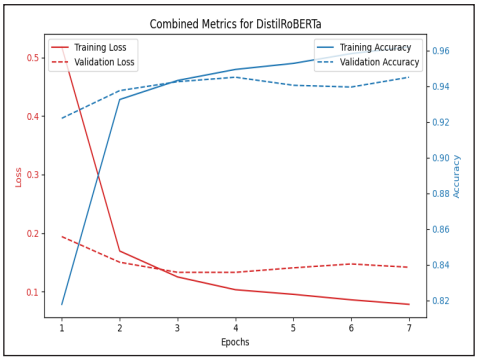


Figure 25. Epochs metrics of DistilRoBERTa

XLNet

By adding a permutation-based autoregressive pretraining technique, XLNet overcomes some of BERT's shortcomings by learning bidirectional context and improving its modelling of token dependencies. As seen in Figure 26, XLNet struggles to correctly classify Love and occasionally misclassifies Fear and Surprise, but it is excellent at classifying Sadness and Joy. Figure 27 shows that XLNet is efficient during training, but there is some instability in validation loss and accuracy, indicating potential difficulties in generalising to new data. Despite these variations, it performs better than BERT and ALBERT. Even though XLNet outperforms its predecessors by a small margin, further tweaking may be necessary to maximise its performance, especially for less common emotional categories such as Surprise and Love.

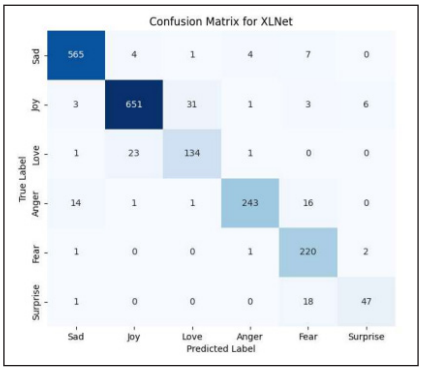


Figure 26. Confusion matrix for XLNet

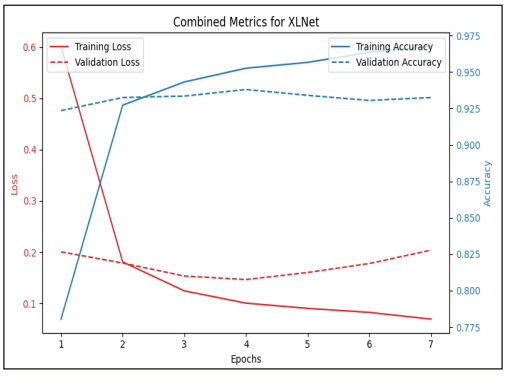


Figure 27. Epochs metrics of XLNet

Comparison of Metrics Across Each Class

Figures 28, 29, 30, 31, 32 and 33 compare Precision, Recall, and F1-Scores for all eight transformer models of the emotional categories Anger, Fear, Joy, Love, Surprise, and Sadness, respectively.

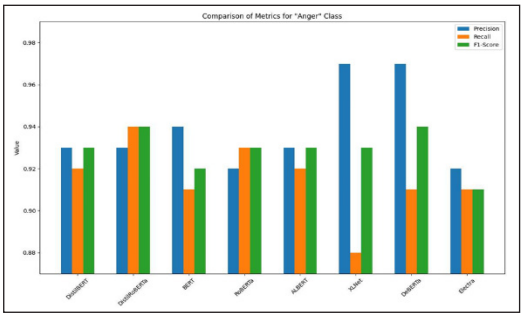


Figure 28. Scores for Anger emotion

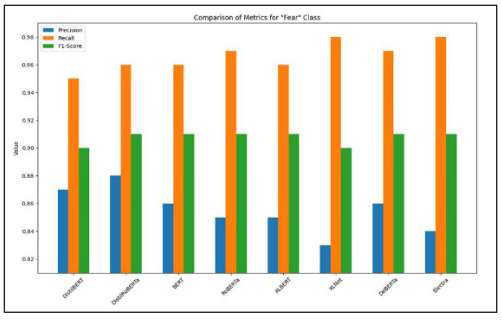


Figure 29. Scores for Fear emotion

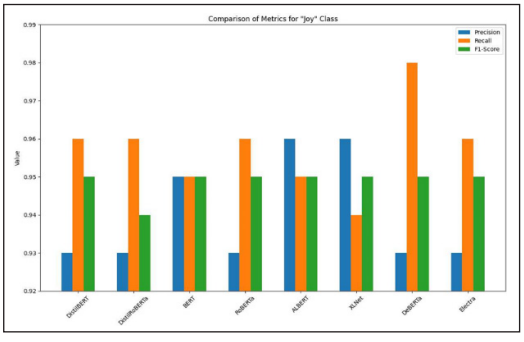


Figure 30. Scores for Joy emotion

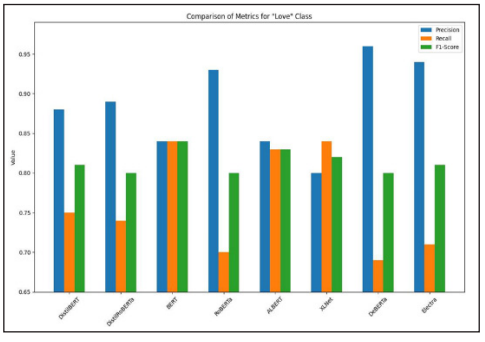


Figure 31. Scores for Love emotion

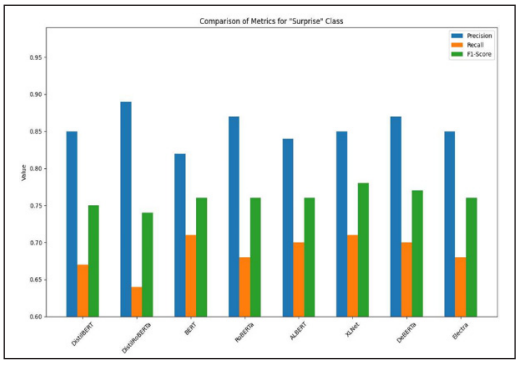


Figure 32. Scores for Surprise emotion

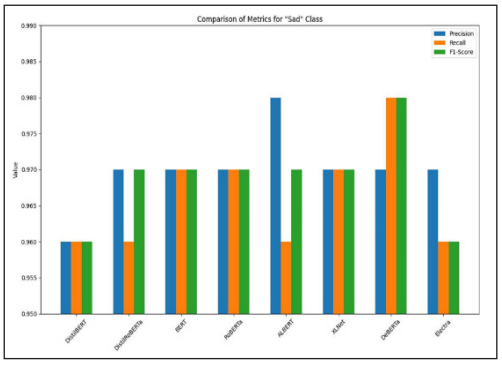


Figure 33. Scores for Sad emotion

DistilBERT and DistilRoBERTa both performed well, especially in recognising Sad and Fear emotions, with DistilRoBERTa being a bit more consistent overall. BERT did great with Joy and Surprise, while RoBERTa was slightly better at Joy but struggled with Sad and Fear. ALBERT was decent across the board but had trouble with Anger and Surprise. XLNet really stood out for Anger, Fear, and Surprise, outperforming the others in those areas. DeBERTa was the most reliable overall, especially strong in Love and Fear, with top precision for Love and Surprise—though it missed some instances in those categories.

Electra had the weakest performance, particularly with Sad and Anger, but did okay with Fear and Surprise. Overall, DeBERTa and XLNet were the most consistent and effective models, while Electra fell behind.

DISCUSSION

Formally, depression and mental issues connected to the mind are discussed. The emergence of chatbots in the current scenario is illustrated. The SKY Chatbot was developed using the PALM model to enable continuous, chat-based interactions that build on previous conversations. To evaluate the model (chatbot), various AI-integrated Deep Learning transformers were used for performance analysis.

The performance metrics, including Accuracy, F1-score, Precision, and Recall, of multiple models are illustrated in the paper. DeBERTa demonstrates the model's supremacy on text classification tasks. Accuracy was achieved by models such as ALBERT and BERT. Nevertheless, DeBERTa consistently outperformed every other model on every metric. This finding suggests that DeBERTa might be the ideal model in circumstances where obtaining a higher percentage of true positives is crucial. Even though they perform relatively poorly, lightweight models like DistilBERT and Electra can be used in situations where the main goal is efficient resource utilisation and minor trade-offs on accuracy and other metrics are acceptable. It becomes clear how model size, training time, and performance are balanced. While lightweight models are better suited for tasks where resource utilisation is the top priority, more comprehensive models, such as ALBERT and DeBERTa, are better at generalisation.

The models under consideration in this study mostly show good results, with Joy being the easiest emotion to predict by all the platforms. Sadness, similarly, shows good results; for example, DeBERTa, BERT, and ALBERT correctly identify most Sadness instances. On the other hand, Love is the most difficult category for the models to identify, as many models mislabel a high percentage of actual Love instances as other emotions, e.g., joy or Anger. It is worth mentioning that, in general, Fear and Surprise have fewer classification issues across models, suggesting that these models are better at differentiating these emotions from others. While all models perform well, DeBERTa stands out a bit more in terms of its ability to classify all emotional categories with a relatively lower number of errors, especially in the more complex category of Surprise.

In conclusion, DeBERTa and ALBERT are recommended for large-scale emotion classification tasks, given their ability to achieve high cross-performance across all metrics. In the future, where computing resources are limited, DistilBERT and DistilRoBERTa can strike a helpful trade-off between computational speed and accuracy. The results indicate that DeBERTa achieved the best overall performance among all models, especially in the Love and Fear categories. At the same time, XLNet ranked relatively high in distinct

dimensions, particularly in terms of Anger. The best models in the Surprise and Joy categories were BERT and DistilBERT. However, the ones that worked are more heavily covered in categories such as Electra and have been quite limited, thereby hinting that higher-performing models, such as DeBERTa and XLNet, might have a penchant for the finer classification in sentiment tasks.

DeBERTa and ALBERT are recommended for large-scale emotion classification tasks, given their ability to achieve high cross-performance across all metrics. In the future, where computing resources are limited, DistilBERT and DistilRoBERTa can strike a helpful trade-off between computational speed and accuracy.

The following is the list of solutions for research Questions (RQ) addressed in the paper.

RQ1 : The mention of using various transformer models suggests that the SKY Emotional Chatbot is equipped to analyse emotions and sentiments effectively. The high precision (97%) indicates that the chatbot can accurately assess emotional states, addressing this research question.

RQ2 : Ethical concerns can be addressed within personalised interventions. This requires more information on privacy, data usage, and the ethical guidelines used in the chatbot's design. It is a challenging task.

RQ3 : Using transformer models like BERT and its variants shows a strong ability to process unstructured text data. These models are built to handle linguistic nuances, enabling the chatbot to effectively manage unstructured conversational data.

RQ4 : The proposed chatbot can refer to previous conversations, indicating that methods are in place to improve continuity and contextual awareness. This suggests a mathematical or algorithmic approach to modelling past interactions for coherent dialogue.

FUTURE SCOPE

Several interesting research sources on sentiment and emotion classification are currently being investigated using transformer architectures. Given some of these transformer models fine-tuned on specialised datasets, an interesting direction of research is to examine the effects of fine-tuning. A larger extent of how these models can generalise across a wider spectrum of emotional states is to make the model appropriate to use on individual emotion datasets or to include additional emotions. Such fine-tuning may improve performance in specific emotion categories that are harder to determine and would provide much more insight into the model’s capabilities. This could help improve performance in contexts such as customer feedback, social media interactions, or customer service experiences, where contextual details and subtle tone are important.

Text, audio, and video are all included in multimodal data integration, which opens new possibilities for chatbots to develop emotional intelligence. To gain a deeper understanding of the user's emotional state, the chatbot can employ multimodal deep learning models and feature fusion. For example, the system can identify subtle emotional cues from text, facial expressions, and vocal tone to provide more accurate and sympathetic responses. These enhanced perceptions will not only make it easier for the chatbot to recognise emotions but also enable it to adapt in complex scenarios where a single input might be ambiguous or insufficient in context.

Incorporating novel techniques in mathematical modelling, specifically aimed at tensor-based representations, could significantly improve model performance. Certain tensor methods are effective at capturing interactions among multiple representations, while the modal attention mechanism can focus on the most pertinent information for context-based emotional recognition. This technological advancement could enable chatbots to respond to users consistently and accurately. In addition, dynamic attention to different modalities will enable SKY to function in various environments as an articulate emotional support system.

The ethical problems with the subject's applications become increasingly important as systems improve at identifying emotions. Better, more accurate models may make it easier for people to exploit weaknesses or manipulate users, which calls for close supervision. Maintaining trust and avoiding overly intrusive emotional interventions present ethical challenges. The systems should incorporate precise consent mechanisms to inform users that their emotional data is being used, given these ethical concerns. To maintain the emotional interventions' non-intrusiveness and respect, oversight mechanisms should ensure transparency in their execution.

Finally, exploring ways to minimise problems such as overfitting and underfitting in low-data environments is an important area for improvement. Through ways such as knowledge distillation, adaptive loss functions, and pre-training strategies, we can better generalise emotional categories. This allows smaller models, such as DistilBERT, to significantly improve efficiency and accuracy, creating more responsive and robust emotional recognition models for real-time applications.

Therefore, future research should focus on fine-tuning these transformer-based models. They offer great opportunities to integrate multimodal data, adopt hybrid approaches, and apply specific methods to address various discrepancies. As emotion recognition continues to develop, these efforts could lead to more reliable, scalable, and practical research applications.

CONCLUSION

These findings demonstrate a thorough examination of transformer-based models using a range of performance metrics. With consistent scores in Accuracy, F1-Score, Precision, and

Recall, these results demonstrate that DeBERTa is the most adaptable model. However, models like DistilBERT or Electra can be considered depending on task requirements (e.g., resource constraints or the need to achieve better results).

DeBERTa demonstrated superiority across all emotions and confusion-matrix metrics, even though all models performed well. However, Love remains a major challenge for all models, and future work should focus on methods to improve classification accuracy for these emotionally complex categories.

Emotion classification can be done well by transformer-based models such as DeBERTa and ALBERT, which are able to show very high performance in complicated categories like Fear and Surprise. Nevertheless, issues remain, particularly in the classification of Love, a category in which all models have struggled to achieve higher F1-Score, Precision, and Recall scores. Although complex models like DeBERTa and ALBERT have generally yielded better results, lighter models such as DistilBERT and Electra have incurred efficiency-accuracy trade-offs, making them suitable for situations where considerable resource consumption is required.

Studying just about each of the eight different transformer-based models for emotion classification, this research has essentially figured out their comparison criteria to be F1-Score, Precision, and Recall. The heatmap representations of these statistics are quite informative, as they show how the different models align with the various emotion classes in terms of their strengths and weaknesses. DeBERTa and ALBERT, as usual, are the top two performers, whereas DistilBERT and DistilRoBERTa can be considered good choices for the task of efficient performance to some extent.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to VNR Vignana Jyothi Institute of Engineering and Technology for providing a supportive academic environment. The authors also extend their appreciation to everyone who contributed guidance, encouragement, and support throughout the development and completion of this research.

REFERENCES

Agrawal, R., Monteiro, K., Narasimhamurti, N. S., Sharma, S., Suryawanshi, A., Bariya, A., Narvekar, S., Bagnoli, L., Saxena, M., Magoun, L., Parsekar, S. S., Pozuelo, J. R., Lesh, N., Sood, M., Sharma, T., Yadav, H., Bhan, A., Nadkarni, A., & Patel, V. (2025). A conversational agent (PracticePal) to support the delivery of a brief behavioral activation treatment for depression in rural India: Development and pilot-testing study. *JMIR Formative Research, 9*, Article e73563. https://doi.org/10.2196/73563

Alsabi, R. N. S., Zaimi, A. F., Sivalingam, T., Ishak, N. N., Alimuddin, A. S., Dasrilsyah, R. A., Basri, N. I., & Jamil, A. A. M. (2022). Improving knowledge, attitudes and beliefs: A cross-sectional study of postpartum depression awareness among social support networks during COVID-19 pandemic in Malaysia. *BMC*

Women's Health, 22(1), Article 221. https://doi.org/10.1186/s12905-022-01795-x

Anisha, S. A., Sen, A., & Bain, C. (2024). Evaluating the potential and pitfalls of AI-powered conversational agents as humanlike virtual health carers in the remote management of noncommunicable diseases: Scoping review. *Journal of Medical Internet Research, 26*, Article e56114. https://doi.org/10.2196/56114

Bartlett, C., Littlewood, A., McCool, R., Mak, C., & Chen, S. Y. (2024). EE546 postpartum depression: Systematic literature reviews of economic evaluations of pharmacological and non-pharmacological treatments, utilities, costs, and healthcare resource use [Conference abstract]. *Value in Health, 27*(12), S163. https://doi.org/10.1016/j.jval.2024.10.826

Chen, T., Shen, Y., Chen, X., & Zhang, L. (2024). PsyChatbot: A psychological counseling agent towards depressed Chinese population based on cognitive behavioural therapy. *ACM Transactions on Asian and Low-Resource Language Information Processing*. Advance online publication. https://doi.org/10.1145/3676962

Clark, K., Luong, M.-T., Le, Q. V., & Manning, C. D. (2020). *ELECTRA: Pre-training text encoders as discriminators rather than generators* [Preprint]. arXiv. https://arxiv.org/abs/2003.10555v1

Colita, C.-I., Hermann, D. M., Filfan, M., Colita, D., Doepnner, T. R., Tica, O., Glavan, D., & Popa-Wagner, A. (2024). Optimizing chronotherapy in psychiatric care: The impact of circadian rhythms on medication timing and efficacy. *Clocks & Sleep, 6*(4), 635-655. https://doi.org/10.3390/clockssleep6040043

Denecke, K., & Gabarron, E. (2024). The ethical aspects of integrating sentiment and emotion analysis in chatbots for depression intervention. *Frontiers in Psychiatry, 15*, Article 1462083. https://doi.org/10.3389/fpsyt.2024.1462083

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171-4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Dharrao, D., & Gite, S. (2024). TherapyBot: A chatbot for mental well-being using transformers. *International Journal of Advances in Applied Sciences, 13*(1), 1-12. https://doi.org/10.11591/ijaas.v13.i1.pp1-12

Fan, H., & Ma, S. (2024). Editorial: Effects of exercise intervention on emotional health and emotional processing: Evidence related to psychology and physiology. *Frontiers in Psychology, 15*, Article 1474722. https://doi.org/10.3389/fpsyg.2024.1474722

Fan, J. (2024). *AI psychotherapists and positive psychology: A systematic literature review of current literature and future directions* [Master's thesis, Lappeenranta–Lahti University of Technology LUT]. LUTPub. https://lutpub.lut.fi/handle/10024/168817

Feng, X., Ma, X., Li, J., Zhou, Q., Liu, Y., Song, J., Liu, J., Situ, Q., Wang, L., Zhang, J., & Lin, F. (2025). Inflammatory pathogenesis of post-stroke depression. *Aging and Disease, 16*(1), 209-238. https://doi.org/10.14336/ad.2024.0203

Ferrario, A., Sedlakova, J., & Trachsel, M. (2024). The role of humanization and robustness of large language models in conversational artificial intelligence for individuals with depression: A critical analysis. *JMIR*

Mental Health, 11, Article e56569. https://doi.org/10.2196/56569

Fouad, M. A., Tadros, M. G., & Michel, H. E. (2024). A comprehensive review of the pathophysiology of depression. *Archives of Pharmaceutical Sciences Ain Shams University, 8*(1), 122-132. https://doi.org/10.21608/aps.2024.271710.1160

Guo, Z., Lai, A., Deng, Z., & Li, K. (2024). Evaluating the feasibility and acceptability of a GPT-Based chatbot for depression screening: A mixed-methods study. In X. Xie, G. Powathil, I. Styles, & M. Ceccarelli (Eds.), *Artificial intelligence in healthcare* (Lecture Notes in Computer Science, Vol. 14975, pp. 249-263). Springer. https://doi.org/10.1007/978-3-031-67278-1_20

He, P., Liu, X., Gao, J., & Chen, W. (2020). *DeBERTa: Decoding-enhanced BERT with disentangled attention* [Preprint]. arXiv. https://arxiv.org/abs/2006.03654v1

Heinz, M. V., Mackin, D., Trudeau, B., Bhattacharya, S., Wang, Y., Banta, H. A., Jewett, A. D., Salzhauer, A., Griffin, T., & Jacobson, N. C. (2024). *Evaluating Therabot: A randomized control trial investigating the feasibility and effectiveness of a generative AI therapy chatbot for depression, anxiety, and eating disorder symptom treatment* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/pjqmr

Kang, B., & Hong, M. (2025). Development and evaluation of a mental health chatbot using ChatGPT 4.0: Mixed methods user experience study with Korean users. *JMIR Medical Informatics, 13*, Article e63538. https://doi.org/10.2196/63538

Khan, H., & Bokhari, S. F. H. (2024). Integrating artificial intelligence (AI) chatbots for depression management: A new frontier in primary care. *Cureus, 16*(8), Article e66857. https://doi.org/10.7759/cureus.66857

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). *ALBERT: A lite BERT for self-supervised learning of language representations* [Preprint]. arXiv. https://arxiv.org/abs/1909.11942v1

Lee, J., & Lee, D. (2024). The effects of a chatbot-based interpretation bias modification on early adulthood depression [Conference abstract]. *European Psychiatry, 67*(S1), S532. https://doi.org/10.1192/j.eurpsy.2024.1105

Lei, X., Wu, H., Deng, Z., & Ye, Q. (2023). Self-disclosure, social support and postpartum depressive mood in online social networks: A social penetration theory perspective. *Information Technology & People, 36*(1), 433-453. https://doi.org/10.1108/itp-12-2020-0825

Li, J., Li, Y., Mei, X., He, H., Yu, L., Ho, G. W. K., & Chan, E. A. (2024). Effectiveness of cognitive-based interventions on psychological distress in adolescents with physical disabilities: A systematic review and meta-analysis. *International Journal of Mental Health Nursing, 33*(6), 1729-1742. https://doi.org/10.1111/inm.13391

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A robustly optimized BERT pretraining approach* [Preprint]. arXiv. https://arxiv.org/abs/1907.11692v1

Manole, A., Cârciumaru, R., Brînzaș, R., & Manole, F. (2024). Harnessing AI in anxiety management: A chatbot-based intervention for personalized mental health support. *Information, 15*(12), Article 768. https://doi.org/10.3390/info15120768

Mavranezouli, I., Megnin-Viggars, O., Pedder, H., Welton, N. J., Dias, S., Watkins, E., Nixon, N., Daly, C.

H., Keeney, E., Eadon, H., Caldwell, D. M., O'Donoghue, K. J.-M., Stockton, S., Arnold, S., Thomas, J., Kapur, N., & Pilling, S. (2024). A systematic review and network meta-analysis of psychological, psychosocial, pharmacological, physical and combined treatments for adults with a new episode of depression. *eClinicalMedicine, 75*, Article 102780. https://doi.org/10.1016/j.eclinm.2024.102780

Moschny, N. K. (2020). *Epigenetics in psychiatry: Characterization of the immune system and its methylome in depressed patients receiving electroconvulsive therapy* [Doctoral dissertation, University of Veterinary Medicine Hannover]. https://nbn-resolving.org/urn:nbn:de:gbv:95-114539

Nakamura, A., Sutter-Dallay, A.-L., El-Khoury Lesueur, F., Thierry, X., Gressier, F., Melchior, M., & van der Waerden, J. (2020). Informal and formal social support during pregnancy and joint maternal and paternal postnatal depression: Data from the French representative ELFE cohort study. *International Journal of Social Psychiatry, 66*(5), 431-441. https://doi.org/10.1177/0020764020911409

Norazman, C. W., & Lee, L. K. (2024). The influence of social support in the prevention and treatment of postpartum depression: An intervention-based narrative review. *Women's Health, 20*, Article 17455057241275587. https://doi.org/10.1177/17455057241275587

Piro, A., Curci, A., & Mangiulli, I. (2024). Detection of cognitive-emotional feedback in interviewer-interviewee interaction [Conference abstract]. *Preliminary Reports and Negative Results in Life Science and Humanities, 1*(2, Suppl.), 170. https://cab.unime.it/journals/index.php/PRNR/article/viewFile/4892/pdf

Reis, P. M. (2024). *Systematic review on psychotherapy in the context of psychedelic therapy* [Master's thesis, University of Porto]. https://doi.org/10.34626/bj3a-rw68

Ren, S., Guo, Z., Zhang, J., He, Y., Sun, Z., & Yang, J. (2025). Transcriptomic alterations in the hippocampus and prefrontal cortex of rats with chronic unpredictable stress induced by low-intensity pulsed ultrasound. *Molecular Neurobiology, 62*(5), 5917-5929. https://doi.org/10.1007/s12035-024-04656-w

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter* [Preprint]. arXiv. https://arxiv.org/abs/1910.01108v1

Shen, Y., Zou, X., & Hu, G. (2024). Editorial: Digital technology in the management and prevention of diabetes. *Frontiers in Endocrinology, 15*, Article 1511236. https://doi.org/10.3389/fendo.2024.1511236

Tarai, S., Mukherjee, R., Gupta, S., Rizvanov, A. A., Palotás, A., Chandrasekhar Pammi, V. S., & Bit, A. (2019). Influence of pharmacological and epigenetic factors to suppress neurotrophic factors and enhance neural plasticity in stress and mood disorders. *Cognitive Neurodynamics, 13*(3), 219-237. https://doi.org/10.1007/s11571-019-09522-3

Taylor, B. L., Nath, S., Sokolova, A. Y., Lewis, G., Howard, L. M., Johnson, S., & Sweeney, A. (2022). The relationship between social support in pregnancy and postnatal depression. *Social Psychiatry and Psychiatric Epidemiology, 57*(7), 1435-1444. https://doi.org/10.1007/s00127-022-02269-z

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30, pp. 5998-6008). Curran Associates. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Volz, F., Wolf, K., Fung, C., Carroll, I., Lahmann, C., Lützen, N., Urbach, H., Klingler, J.-H., Beck, J., & El Rahal, A. (2024). Impact of spinal CSF leaks on quality of life and mental health and long-term reversal by surgical closure. *Neurology: Clinical Practice, 14*(2), Article e200272. https://doi.org/10.1212/cpj.0000000000200272

Voon, V. L. W., Tan, P. P., & Sim, C. W. (2026). Towards human-centric mental health chatbot design: Leveraging on DASS-21. *Journal of Advanced Research in Applied Sciences and Engineering Technology, 63*(1), 37-51. https://doi.org/10.37934/araset.63.1.3751

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2019). *XLNet: Generalized autoregressive pretraining for language understanding* [Preprint]. arXiv. https://arxiv.org/abs/1906.08237v1

Zandieh, S., Abdollahzadeh, S. M., Sadeghirad, B., Wang, L., McCabe, R. E., Yao, L., Inness, B. E., Pathak, A., Couban, R. J., Crandon, H., Torabiardakani, K., Bieling, P., & Busse, J. W. (2024). Therapist-guided remote versus in-person cognitive behavioural therapy: A systematic review and meta-analysis of randomized controlled trials. *Canadian Medical Association Journal, 196*(10), E327-E340. https://doi.org/10.1503/cmaj.230274

Zeidner, M., Matthews, G., & Roberts, R. D. (2012). The emotional intelligence, health, and well-being nexus: What have we learned and what have we missed? *Applied Psychology: Health and Well-Being, 4*(1), 1-30. https://doi.org/10.1111/j.1758-0854.2011.01062.x

Zhong, W., Luo, J., & Zhang, H. (2024). The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: A systematic review and meta-analysis. *Journal of Affective Disorders, 356*, 459-469. https://doi.org/10.1016/j.jad.2024.04.057