AI-PsycheChat: Developing an Urdu Psychotherapy Chatbot for Self-Attachment

SPROJ Report

Mohammad Mubashir - 24100208
Muhammad Shahmeer - 24100237

Advisor: Dr. Agha Ali Raza
School of Science and Engineering
Lahore University of Management Sciences
Submission Date: 15th May 2024

Table of Contents

Chapter 1

1. Introduction

Mental health support remains a significant challenge, particularly in regions with limited access to traditional therapy services. Our senior year project aimed to bridge this gap by developing "AI-PsycheChat," an innovative Urdu language chatbot that leverages natural language processing (NLP) and Information Retrieval techniques to provide a scalable and accessible virtual platform for self-attachment therapy.

Self-attachment therapy (SAT) has emerged as an effective approach to address mental health challenges by nurturing self-compassion and fostering a positive self-image. However, the accessibility of such therapeutic interventions remains limited, particularly in regions with language and cultural barriers. Our project sought to harness the power of artificial intelligence to transcend these boundaries and democratize access to mental health support through the development of an Urdu language chatbot.

Building upon the foundational work of the already existing Persian SAT chatbot project, our endeavors focused on language adaptation, model optimization, and user-centric design to create a culturally relevant and empathetic conversational agent tailored to the needs of the Urdu-speaking community.

Chapter 2

2. Literature Review

2.1. Introduction

Mental health chatbots are conversational agents that can provide psychotherapy and emotional support using natural language interactions. Recent advancements in natural language processing and neural network modeling have enabled significant progress in this field. However, most chatbots are limited to English, while mental health needs are universal across different languages and cultures. This summary will examine how an existing English self-attachment therapy chatbot was adapted to Mandarin Chinese through machine translation, transfer learning, and model compression techniques.

2.2. Background

The self-attachment therapy (SAT) chatbot aims to guide users in practicing protocols from SAT, a form of psychotherapy focused on strengthening the relationship between one's adult self and inner child self. SAT has shown promise in improving emotional regulation in pilot studies. The chatbot was developed to make SAT more accessible through a conversational digital coach.

Previous versions of the SAT chatbot used rule-based conversation modeling, which limits flexibility and engagement. The recent research papers present more robust versions that incorporate deep learning for emotion recognition and empathetic response generation.

2.3. Dataset Creation and Annotation

A key component in developing the advanced SAT chatbot was curating an appropriate training dataset. The EmpatheticPersonas dataset was crowdsourced using Amazon Mechanical Turk. It contains:

- 1,181 textual expressions of emotions like sadness, anger, fear, and joy. These enable training an emotion classifier.

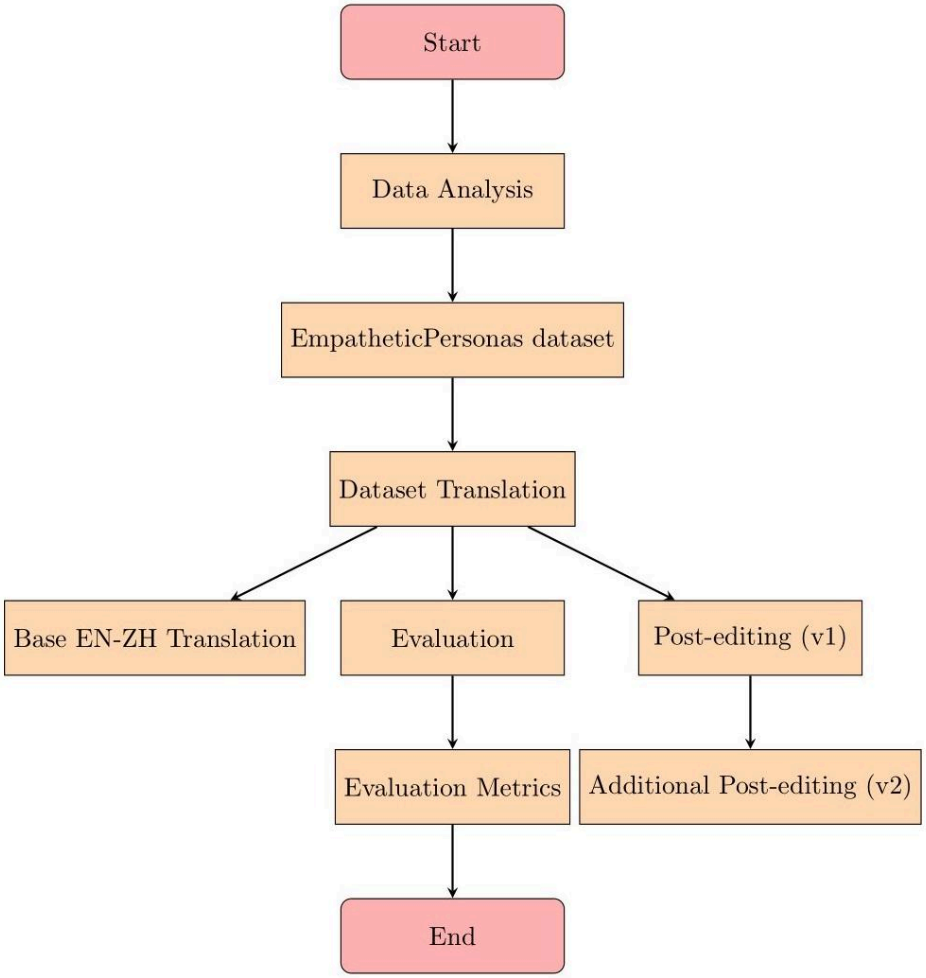
- 2,143 empathetic rewrites of 45 base utterances, where crowd workers rephrased the utterances to be more empathetic given a particular emotion. This provides data for training an empathy classifier and generating empathetic responses.
- Discrete empathy ratings on a scale of 0-2 for 1,100 of the rewritten utterances. Higher values indicate more empathetic phrasing. The emotion expressions and empathetic rewrites were collected by prompting crowd workers to imagine they were experiencing a given emotion. While not genuine, this provides plausible data.

To adapt the chatbot to Mandarin, the English dataset was machine translated into Mandarin and then manually post-edited by native speakers. This improved translation quality over raw output, as measured by fluency metrics like SLOR, PRISM-SRC, and perplexity. Post-editing also allowed screening for any translation errors that could negatively impact user safety.

2.4. Translation Process Summary

The translation process for converting English to Mandarin involved the following steps:

1. **Data Analysis:** The process began with an analysis of the EmpatheticPersonas dataset, which included expressions of emotion and empathetic rewrites.
2. **Dataset Translation:** To streamline the translation process, publicly available machine translation tools were employed, followed by post-editing steps (v1 and v2) to enhance translation quality.
3. **Evaluation Metrics:** The quality of translations was assessed using reference-free sentence fluency metrics, demonstrating improvements with post-editing.



2.5. Emotion Recognition

Recognizing the user's emotional state is essential for the SAT chatbot to contextualize its responses. A RoBERTa model was trained on the emotion expressions in EmpatheticPersonas to classify text into four categories: sadness, anger, fear, or joy.

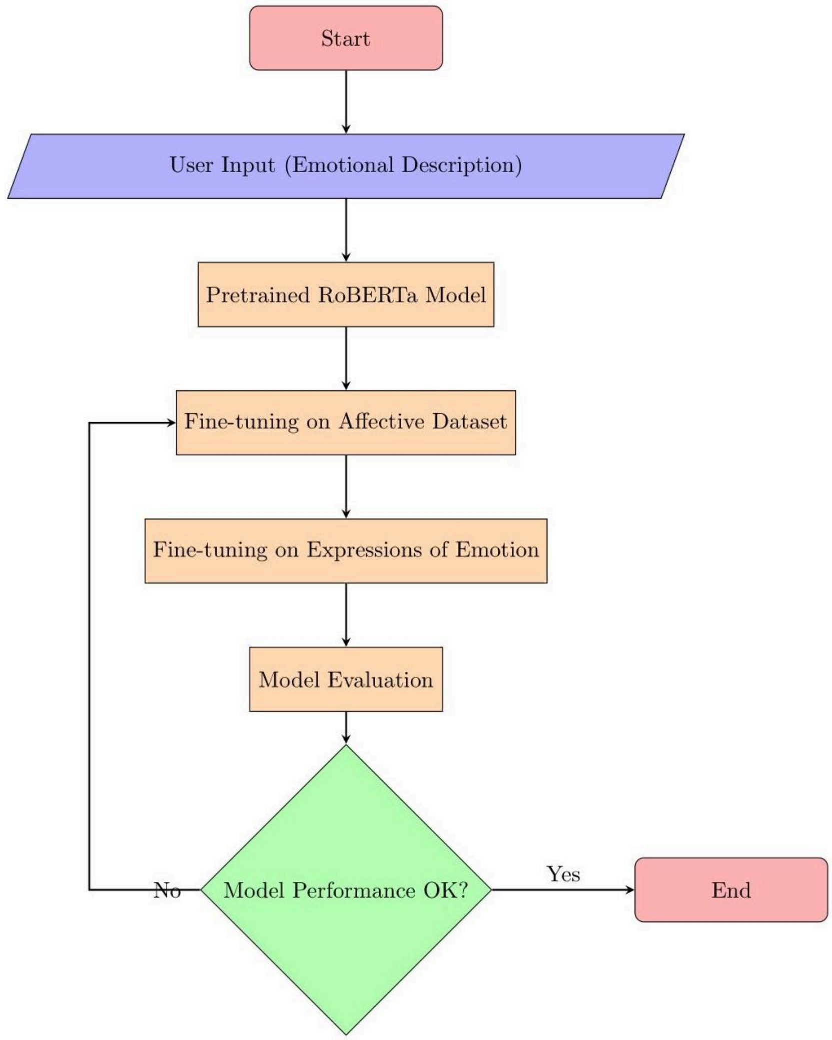
User Input (Emotional Description): The chatbot begins by prompting the user to describe their emotional state at the start of each conversation.

Pretrained RoBERTa Model: A RoBERTa language model, initially pretrained on an existing English emotion dataset, serves as the foundation. This model is used for the subsequent fine-tuning steps.

Fine-tuning on Affective Dataset: The pretrained RoBERTa model is fine-tuned on Saravia et al.'s affective dataset, adapting it to emotion recognition.

Fine-tuning on Expressions of Emotion: Further fine-tuning is performed using the expressions of emotion from the EmpatheticPersonas dataset, which includes sadness, anger, fear, and joy. This fine-tuning process enhances the model's ability to recognize emotions.

Model Evaluation: The performance of the fine-tuned model is assessed using evaluation metrics. It achieves an impressive 94.96% accuracy and significantly outperforms a baseline keyword-based classifier implemented in the previous version of the chatbot.



The model was pre-trained on an existing English emotion dataset, then fine-tuned on the SAT data. This achieved 94.96% accuracy on the SAT test set, significantly improving over a baseline keyword-based classifier.

For Mandarin, an XLM-Roberta model was trained. It was first fine-tuned on a native Mandarin emotion dataset, then the translated SAT data. Despite limited in-domain Mandarin data, it achieved over 90% accuracy on Mandarin test sets.

To optimize the model for production deployment, knowledge distillation was used to compress it into a smaller model with faster inference times and no significant performance loss (~85% accuracy on English test set).

2.6. Empathetic Response Generation

The chatbot aims to produce empathetic responses using the rewritten utterances from EmpatheticPersonas. Each rewrite was split into individual sentences. These were recombined to form a large set of candidate responses.

To generate empathetic responses it follows the following steps:

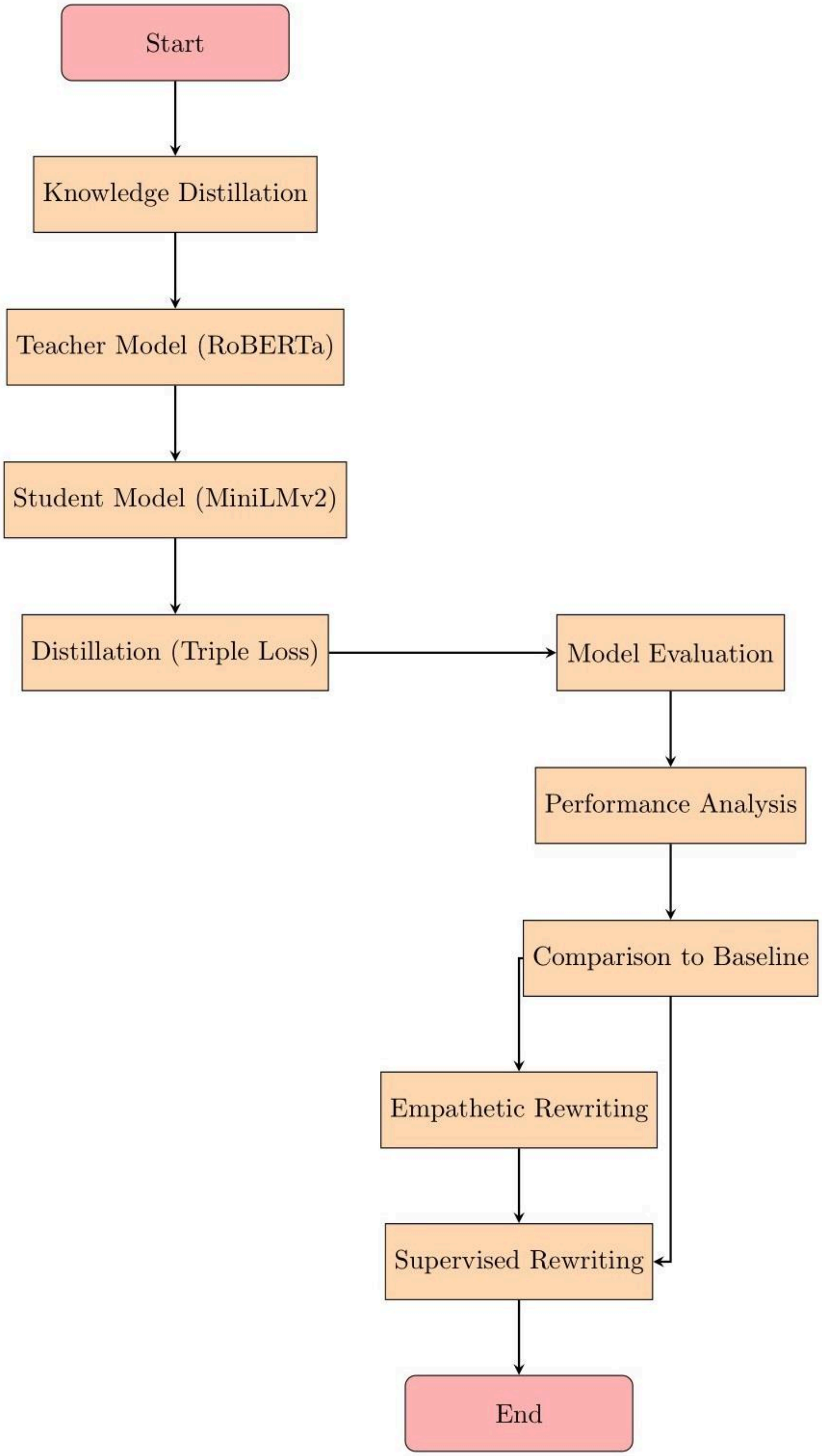
Candidate Response Generation: Rewritten utterances from EmpatheticPersonas are used as a source for candidate responses. These rewritten utterances are broken down into individual sentences and recombined, creating a pool of potential responses.

Scoring Candidates: When the chatbot needs to respond, it evaluates each candidate response based on three criteria:

- **Empathy:** A T5 classifier, fine-tuned using empathy annotations, is used to assess the level of empathy conveyed by each candidate response.
- **Fluency:** Fluency is determined by calculating perplexity using a language model. Repetitive responses are penalized to ensure coherence.
- **Novelty:** The chatbot considers the similarity of a candidate response to previous chatbot responses, aiming for diversity.

Response Selection: The chatbot selects the response that scores highest across these criteria. This approach allows the chatbot to respond empathetically while maintaining coherence and ensuring a variety of responses.

For Mandarin, Chinese GPT-2 was trained with reinforcement learning to perform empathetic rewriting. Supervised fine-tuning was also explored. Responses were manually screened for any inappropriate content.


Start
Knowledge Distillation
Teacher Model (RoBERTa)
Student Model (MiniLMv2)
Distillation (Triple Loss)
Model Evaluation
Performance Analysis
Comparison to Baseline
Empathetic Rewriting
Supervised Rewriting
End

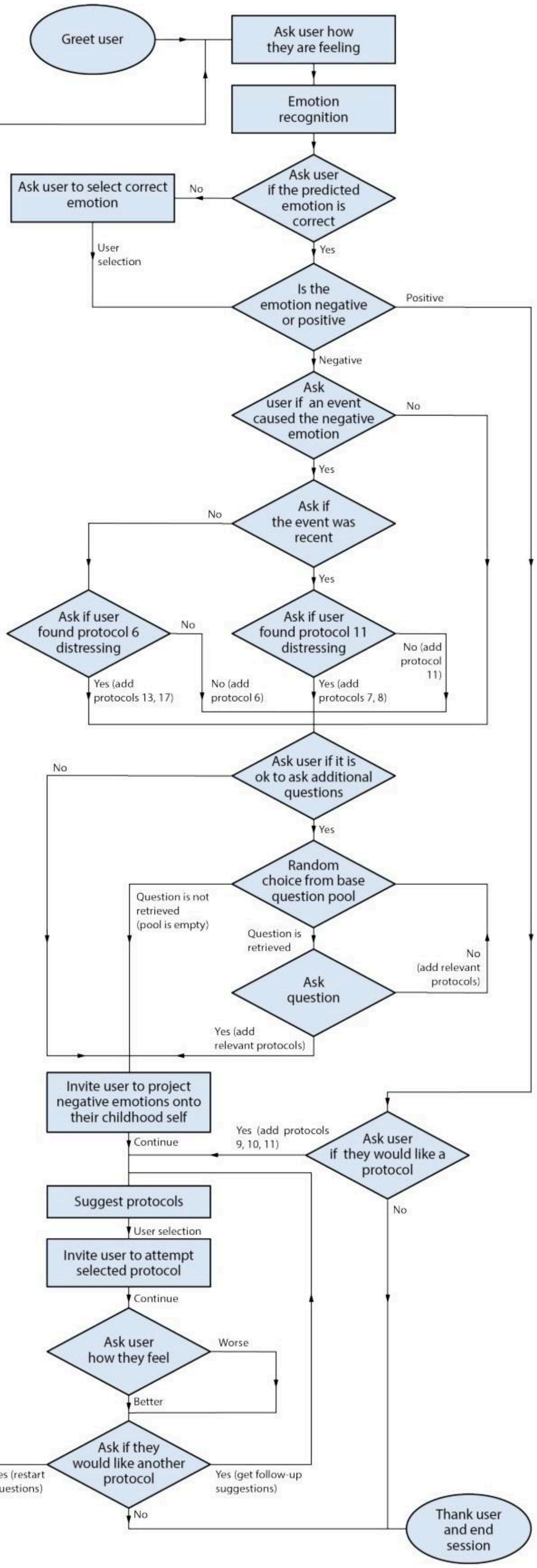
2.7. Conversation Flow and User Experience

The chatbot's conversation flow is structured using a flowchart designed to optimize the practice of SAT (Self-Attachment Technique). When a user logs into the application, they are prompted to choose a persona from among five options: Kai, Robert, Gabrielle, Arman, and Olivia. Each persona corresponds to a specific subset of the (augmented) dataset that is loaded into the backend.

The conversation between the user and the chatbot allows for a combination of open-text and multiple-choice inputs, and it is guided by the flowchart. Deep-learning techniques are applied at various nodes in the flowchart for tasks such as emotion recognition and utterance retrieval.

All five personas navigate the same flowchart during conversations, but each persona has access to a specific set of utterances. Additionally, the user's emotional state, once identified, is saved as a variable and used to select relevant subsets of utterances from the dataset.

The primary objective of the chatbot is to recommend the most appropriate SAT protocols. As users progress through the conversation, a list of protocol suggestions is generated. The content of this list and the timing of its disclosure depend on the user's responses and engagement during the conversation.


Greet user
Ask user how they are feeling
Emotion recognition
Ask user if the predicted emotion is correct
No
Ask user to select correct emotion
User selection
Yes
Is the emotion negative or positive
Positive
Negative
Ask user if an event caused the negative emotion
No
Yes
Ask if the event was recent
No
Yes
Ask if user found protocol 6 distressing
No
Ask if user found protocol 11 distressing
No (add protocol 11)
Yes (add protocols 13, 17)
No (add protocol 6)
Yes (add protocols 7, 8)
Ask user if it is ok to ask additional questions
No
Yes
Random choice from base question pool
Question is not retrieved (pool is empty)
Question is retrieved
Ask question
No (add relevant protocols)
Yes (add relevant protocols)
Invite user to project negative emotions onto their childhood self
Continue
Ask user if they would like a protocol
Yes (add protocols 9, 10, 11)
No
Suggest protocols
User selection
Invite user to attempt selected protocol
Continue
Ask user how they feel
Worse
Better
Ask if they would like another protocol
Yes (restart questions)
Yes (get follow-up suggestions)
No
Thank user and end session

To enhance user engagement, users can select a persona for the chatbot, each accompanied by a unique avatar. The available personas include Kai, who uses the full dataset, as well as personas representing younger and older individuals, both male and female:

- Kai - uses the full dataset
- Younger female (Olivia)
- Younger male (Arman)
- Older female (Gabrielle)
- Older male (Robert)

The web interface of the chatbot displays protocol instructions in both English and Mandarin, catering to a diverse user base and ensuring a personalized and informative user experience.

2.8. Evaluation

Both versions of the SAT chatbot underwent non-clinical human trials over 5 days with post-study surveys.

For Mandarin, 42 bilingual participants tested the chatbot. Results showed comparable performance to the English version in perceived empathy, engagement, and usefulness. The compressed emotion classifier actually received higher ratings than the English version.

Qualitative feedback highlighted the lack of SAT knowledge amongst participants as a limitation, and suggested recruiting expert clinicians for future trials. Participants also desired a wider range of emotions and more nuanced feedback options.

In the English study with 16 participants, the chatbot showed significant improvements over a purely rule-based version. 63% agreed the emotion recognition was accurate, and 75-88% found the chatbot empathetic, engaging, and useful.

2.9. Conclusion

Recent advancements in deep learning and NLP have enabled sophisticated mental health chatbots with emotion awareness and empathetic dialogue. While most systems are English-only, techniques like machine translation and transfer learning show promise in adapting them to new languages like Mandarin. However, thorough testing and ensuring user safety remain critical when applying AI to such a sensitive domain. Further research into multilingual and multicultural mental healthcare chatbots will continue advancing access to support worldwide.

Chapter 3

3. Implementation

The development of an effective conversational AI system requires the careful selection and integration of various models and components. In the case of the Urdu SAT chatbot, a crucial aspect was the evaluation and implementation of speech-to-text (STT), text-to-speech (TTS), sentiment analysis, and text encoding models tailored for the Urdu language. This chapter outlines the comprehensive process undertaken to identify and deploy the optimal models for each of these components, ensuring the chatbot's ability to understand and respond to Urdu speech and text accurately.

3.1. Model Selection and Evaluation

An integral part of developing the Urdu SAT chatbot was the selection and evaluation of various models for speech-to-text (STT), text-to-speech (TTS), sentiment analysis, and text encoding. We conducted extensive research and testing to identify the most suitable models for our use case.

3.1.1. Speech-to-Text (STT) & Text-to-Speech Models

Figure 1 represents the basic steps of the Speech To Text process. Firstly important features are extracted from the input speech, and then word and sentence matching is done using acoustic word models and defined syntax and semantics for the sentences. This process is mutually exclusive and can be done in parallel. At last, the language modelling is performed using the selected modelling method.

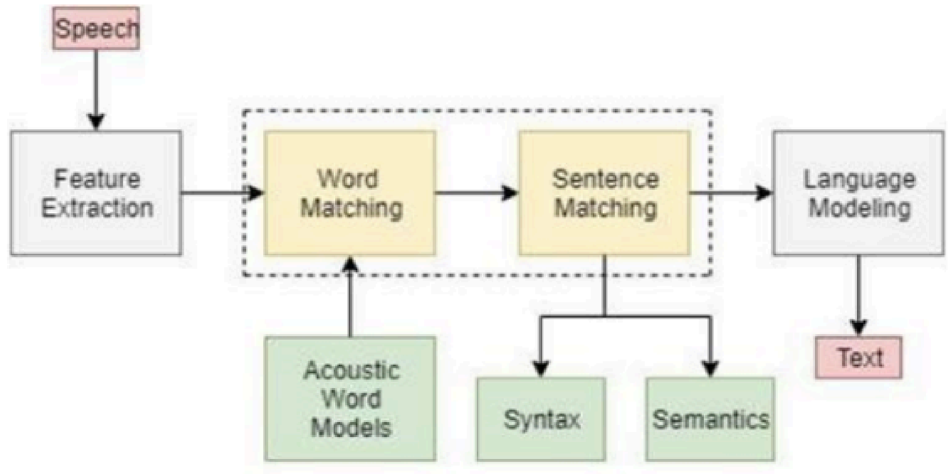


Figure 1: Speech-to-text process

We evaluated several state-of-the-art models for speech-to-text functionality, including:

- **Whisper by OpenAI:** A transformer-based model that approaches human-level robustness and accuracy. While it performs well for English transcription, adaptation is required for the Urdu language.
- **Seamless M4T by Facebook:** A large multilingual speech recognition model.
- **MMS-1B-ALL by Facebook:** A billion-parameter model trained on multiple languages.
- **wav2vec2-xls-r-300m-Urdu by Facebook:** A model specifically designed for Urdu speech recognition.

We tested these models on Urdu audio samples and compared their performance using the Word Error Rate (WER) metric. The Whisper model demonstrated promising results, with the medium variant achieving a WER of 0.235955.

3.1.2. Text-to-Speech (TTS) Models

In Text-To-Speech conversion, the input text is analysed and converted into its audio version to play. This functionality has an effective advantage when a person understands a language but is not fluent in reading and writing in that language. It is also useful for visually impaired people as they cannot read but understand the message by hearing it. Figure 2 represents the basic steps of the TTS process. Firstly, the text is prepared for audio conversion by performing pre-processing and text normalization. The linguistic analysis and prosodic prediction are done in series to generate the text message's Waveform.

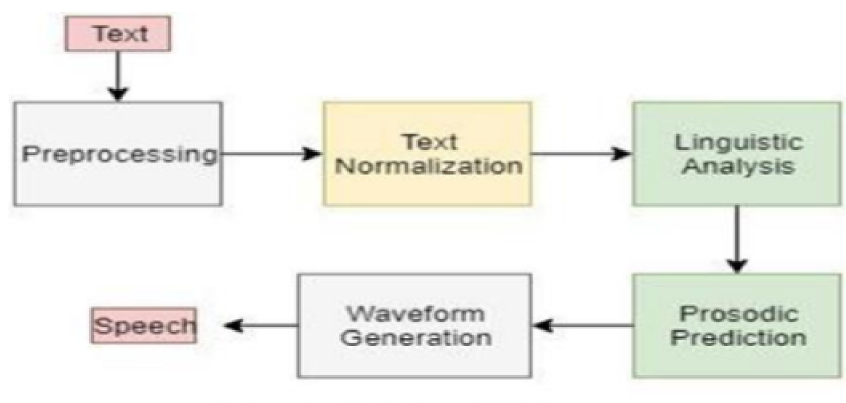


Figure: 2 Text to Speech process

For text-to-speech functionality, we evaluated the following models:

- **Rule-Based Machine Translation (RBMT):** This approach uses syntactic and semantic analysis to convert text to speech but can be inefficient for large systems.
- **Statistical Machine Translation (SMT):** A probabilistic technique that assigns probabilities to input sentences for conversion to speech format, but can be computationally expensive.
- **Hidden Markov Model (HMM):** A probabilistic technique similar to SMT but with potentially better accuracy for TTS conversion.
- **PyTTSx3 (Python API):** A Python library for offline text-to-speech conversion, compatible with multiple platforms and providing customizable voice options.

After thorough testing and evaluation, we selected the PyTTSx3 library for its ease of use, offline functionality, and compatibility with the Urdu language.

3.1.3. Sentiment Analysis Models

For sentiment analysis, we explored various pre-trained models on the Hugging Face platform, ultimately selecting the GPT-2 architecture as the foundation for our custom sentiment analysis model. This decision was based on GPT-2's proven performance in natural language processing tasks and its ability to capture contextual information effectively.

This section aimed to evaluate different algorithmic approaches to sentiment analysis in terms of accuracy, implementation effort, and performance.

The evaluation is conducted on various models, including:

- GPT-J-6B
- FLAN-T5-small
- FLANT-T5-base
- FLANT-T5-xl

The evaluation process involves assessing these models on the IMDB dataset and its contrastive set available at allenai/contrast-sets.

Robustness Definition

In the context of LLMs, Robustness refers to the models' ability to consistently maintain their performance and exhibit reliable and coherent behaviour even in the presence of challenges, variation, or perturbation introduced to their input. This encompasses the capacity of LLMs to effectively handle and resist adversarial inputs or unexpected shifts in the input distribution.

A robust LLM should demonstrate resilience against noise, uncertainties, and potential attacks from malicious users, ensuring that its outputs remain trustworthy and meaningful across a range of conditions. Assessing LLMs' robustness involves evaluating their outputs' stability and reliability under various adverse and difficult circumstances.

Adversarial Perturbation

Unlike images, in the context of Natural Language Processing (NLP), perturbation refers to any alteration or manipulation introduced to textual data, such as adding noise or changing words or syntax. In the text domain, universal perturbations are commonly categorized into:

- character-level

- word-level
- sentence-level

In this evaluation, the original and clear set is the IMDB reviews dataset, and the perturbed set is its relative contrastive set.

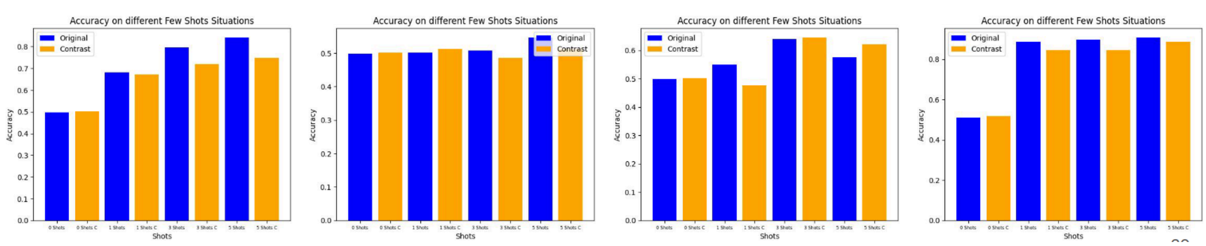
Evaluation Methodology

The models are evaluated under `zero-shot`, `1-shot`, `3-shot`, and `5-shot` settings. The evaluation scripts assess the models' performance on the IMDB dataset and its contrastive set to gauge their robustness in sentiment analysis tasks.

Evaluation Results

Robustness Evaluation on Sentiment Analysis

Model worst performance on contrast set			
GPT-J-6B	FLAN-T5-SMALL	FLAN-T5-BASE	FLAN-T5-XL
-12%	-6%	-15%	-6%



3.1.4. Encoding Models

Accurate text encoding is essential for capturing semantic similarity in Urdu text, a crucial component for context understanding in conversational systems. We evaluated several state-of-the-art encoding models, including mT5, RoBERTa, and MuRIL.

The mT5 model is a multilingual encoder-decoder model trained on over 101 languages, including Urdu. However, despite its impressive performance on various natural language processing tasks, we found that its embeddings did not capture the nuances of Urdu text as effectively as we had hoped.

Next, we evaluated the RoBERTa model, a variant of the BERT architecture that has been robustly optimized for improved performance. While RoBERTa demonstrated better performance than mT5 on our Urdu text similarity evaluation, it still fell short of our desired accuracy levels.

Finally, we turned our attention to the MuRIL model, a BERT-variant specifically tailored and trained on 17 Indian languages, including Urdu. Our evaluation methodology involved assessing these models on a curated test set of Urdu sentence pairs with manually annotated semantic similarity scores ranging from 0 to 1.

To evaluate the models, we generated sentence embeddings for each input text and calculated the similarity using cosine distance. We then compared these similarity scores with the ground truth annotations.

The MuRIL model exhibited the best performance, achieving a mean squared error of 0.1447, which was marginally higher than mT5's 0.1355 but significantly better than RoBERTa's 0.3000. Considering the trade-off between accuracy and computational efficiency, as MuRIL requires fewer parameters (237 million) compared to mT5 (277 million), we determined that MuRIL represented the optimal choice for our Urdu SAT chatbot.

Model Name	Parameters	MSE (Mean Square Error) ↓
mT5	~ 277 M	0.1355
MuRIL	~ 237 M	0.1447
Roberta-base-small	~ 126 M	0.3000

Dummy data used for evaluations of these models:

```
sentence_pairs = [ ["آج کا موسم بہت خوشگوار ہے۔", "آج موسم بہت سہانا ہے۔", 0.9], # Similar topic and sentiment
["کیا آپ نے آج کی تازہ خبریں سنی ہیں؟", "باغ میں پھولوں کی خوشبو بہت مسحور کن ہے۔", 0.1], # Different topics
["میں ہر روز صبح جاگنگ کرتا ہوں۔", "میں ہر صبح ورزش کرنے کے لئے جاتا ہوں۔", 0.8], # Similar activities
["ہمیں کتابوں کی ایک نئی الماری خریدنی چاہئے۔", "میرا کمپیوٹر خراب ہو گیا ہے۔", 0.2], # Different contexts
["میں نے آج ایک نئی فلم دیکھی۔", "آج میں نے ایک دلچسپ ناول پڑھا۔", 0.5], # Both about leisure activities
["کیا آپ کو کھانا پکانا پسند ہے؟", "مجھے باغبانی کرنا بہت اچھا لگتا ہے۔", 0.3], # Different hobbies
["یہ کتاب بہت معلوماتی ہے۔", "اس ناول میں کہانی بہت دلچسپ ہے۔", 0.6], # Both about reading
["میں آج دفتر نہیں جا رہا۔", "آج میری چھٹی ہے۔", 0.7], # Related to taking a day off
["موبائل کی بیٹری ختم ہو رہی ہے۔", "مجھے نیا چارجر خریدنا ہوگا۔", 0.4], # Both related to mobile phone issues
["باہر بارش ہو رہی ہے۔", "موسم بہت خوبصورت ہے۔", 0.5] # Weather-related, but different aspects ]
```

3.2. Dataset Curation and Preprocessing

One of the critical tasks undertaken this semester was curating and preprocessing the dataset for the Urdu sentiment analysis model. Since we had access to the Persian dataset from the previous project, we leveraged that as a starting point. The following steps were taken:

3.2.1. Translation of the Persian Dataset

The existing Persian dataset, containing expressions of emotion and empathetic rewrites, was translated into Urdu using machine translation techniques. Specifically, we employed the Google Translate API to obtain initial translations, which were then manually proofread and corrected by native Urdu speakers to ensure accuracy and cultural appropriateness.

Figure 3: Persian corpus.csv file (raw layout)

QUESTION	LONG ANSWER	SHORT ANSWER	SOURCE	PROMPTS	QNAID	CONTEXTNEEDED	INCOURSE
What is attachment theory?	Attachment theory is a psychological theory about relationships between people.	The most important tenet of this theory is that young children need to establish a relationship with one or more primary caregivers to meet their needs for healthy social and emotional development. This theory was developed by psychoanalyst Jean Bowlby.	https://en.wikipedia.org/wiki/Attachment_theory	"[{""DisplayOrder"":1,""QnaId"":303,""DisplayText"": ""Role-play in SAT""}]"	1	FALSE	TRUE
Does attachment behavior indicate security?	The strength of a child's attachment behavior in a particular situation does not indicate the 'strength' of the attachment relationship.	Insecure children may frequently display strong attachment behaviors, while secure children may need to show less extreme displays. or repeatedly do not have attachment behavior.	https://en.wikipedia.org/wiki/Attachment_theory	"[{""DisplayOrder"":1,""QnaId"":4,""DisplayText"": ""The change of attachment from childhood to adolescence""},{""DisplayOrder"":2,""QnaId"":14,""DisplayText"":""Attachment Behavior in Infants""}]"	3	FALSE	FALSE
How does attachment change from childhood to adolescence?	As children grow, attachment behaviors evolve, and negotiation and compromise become common during kindergarten.	During middle childhood (ages 7-11), children form goal-oriented partnerships with parents. give, prioritize convenience over geographic distance, and become more autonomous. By adolescence, the attachment system acts as a security regulator and responds to the potential for danger or stress. Their maturation affects the activation of the attachment system, although it is still less important than personality traits. have their own attachment and can perform the most important attachment behaviors in some situations without tension.	https://en.wikipedia.org/wiki/Attachment_theory	"[{""DisplayOrder"":1,""QnaId"":180,""DisplayText"":""selfattachment""},{""DisplayOrder"":2,""QnaId"":401,""DisplayText"":""Attachment Behaviour""}]"	4	FALSE	FALSE
What attachment styles are there in adults?	In the 1980s, Cindy Hazan and Philip Shaver extended attachment theory to adult romantic relationships. Four attachment styles have been identified in adults: secure, anxious-attached, distant-avoidant, and fearful-avoidant. These roughly correspond to infant classifications: secure, anxious or dubious, avoidant, and disorganized attachment.	Four attachment styles have been identified in adults: secure, anxious-attached, distant-avoidant, and fearful-avoidant."	https://en.wikipedia.org/wiki/Attachment_theory	"[{""DisplayOrder"":1,""QnaId"":24,""DisplayText"": ""secureattachmentinadults""},{""DisplayOrder"":2,""QnaId"":25,""DisplayText"":""anxiouspreoccupied""},{""DisplayOrder"":2,""QnaId"":26,""DisplayText"" ""dismissiveavoidant""},{""DisplayOrder"":2,""QnaId"":27,""DisplayText"":""fearfulavoidant""}]"	5	FALSE	FALSE
What are the biological explanations for attachment?	The quality of care received during childhood directly affects stress regulation and immune system function. Research has shown that negative experiences in early childhood induce the production of pro-inflammatory cells in the immune system that are directly related to diseases. cardiovascular disease, autoimmune diseases, and some types of cancer. Inherited genetic factors also contribute to attachment formation, such as specific gene polymorphisms associated with anxious or avoidant attachment.	In adulthood, attachment styles are associated with immune markers. Distant attachment with levels Higher interleukin-6 is associated with stress, while anxious attachment is associated with higher cortisol production and lower T-cell counts.	https://en.wikipedia.org/wiki/Attachment_theory	"[{""DisplayOrder"":1,""QnaId"":8,""DisplayText"":""attachmenttheoryandcrime""},{""DisplayOrder"":2,""QnaId"":23,""DisplayText"":""importance of attachment patterns""}]"	7	FALSE	FALSE

Figure 4: Translated corpus.csv file

3.2.2. Data Augmentation and Annotation

To enhance the diversity and representativeness of the Urdu dataset, we performed data augmentation techniques. This involved generating synthetic variations of the translated text by introducing controlled levels of noise, such as synonym replacements and grammatical perturbations. The augmented dataset was then annotated by a team of linguists, who assigned sentiment labels to each entry, ensuring accurate ground truth for model training.

3.2.3. Dataset Partitioning

The curated and annotated Urdu dataset was partitioned into training, validation, and test sets following standard practices in machine learning. This partitioning allowed us to train the sentiment analysis model on a substantial portion of the data while reserving separate subsets for model validation during training and final performance evaluation.

3.3. Sentiment Analysis Model Training

With the preprocessed dataset in place, we focused on training the sentiment analysis model for the Urdu SAT chatbot. The following steps were undertaken:

3.3.1. Model Selection and Adaptation

After evaluating several state-of-the-art models on the Hugging Face platform, we selected the GPT-2 architecture as the foundation for our sentiment analysis model. This decision was based on GPT-2's proven performance in natural language processing tasks and its ability to capture contextual information effectively.

To adapt the GPT-2 model to the Urdu language and the specific task of sentiment analysis, we employed transfer learning techniques. Specifically, we fine-tuned the pre-trained GPT-2 model on the curated Urdu dataset, allowing it to learn the intricacies of Urdu text and the associated sentiment labels.

3.3.2. Hyperparameter Tuning

During the training process, we systematically tuned various hyperparameters to optimize the model's performance. These hyperparameters included learning rate, batch size, and regularization techniques. We employed techniques such as grid search and random search to explore the hyperparameter space effectively and identify the optimal configurations.

3.3.3. Model Evaluation and Selection

To evaluate the performance of the trained sentiment analysis models, we employed a range of metrics, including accuracy, precision, recall, and F1-score. These metrics were calculated on the held-out test set, providing an unbiased estimate of the model's performance on unseen data.

Based on the evaluation results, we selected the top-performing model variant to be integrated into the Urdu SAT chatbot's backend.

3.4. Backend Integration and Deployment

With the sentiment analysis model in place, we focused on integrating it into the backend architecture of the Urdu SAT chatbot. The following steps were taken:

3.4.1. Codebase Refactoring

We thoroughly analyzed the existing codebase of the Persian SAT chatbot, which was provided as a starting point for our project. Through code reviews and refactoring, we improved the modularity, maintainability, and extensibility of the codebase, enabling seamless integration of the sentiment analysis component.

3.4.2. Backend Architecture Updates

The backend architecture was updated to accommodate the sentiment analysis model and its associated data processing pipelines. This involved creating new modules, defining APIs, and ensuring seamless communication between the various components of the chatbot system.

The successful integration of the sentiment analysis model into the backend architecture of the Urdu SAT chatbot marked a significant milestone in the project's development. Through meticulous codebase refactoring and architectural updates, the chatbot system now possesses the capability to understand and respond to the emotional nuances present in Urdu speech and text. This advancement paves the way for more natural and empathetic conversations, thereby enhancing the overall user experience. With the foundational components in place, the project can now progress toward further refinements and the exploration of additional conversational AI capabilities tailored to the unique linguistic and cultural characteristics of the Urdu language.

Chapter 4

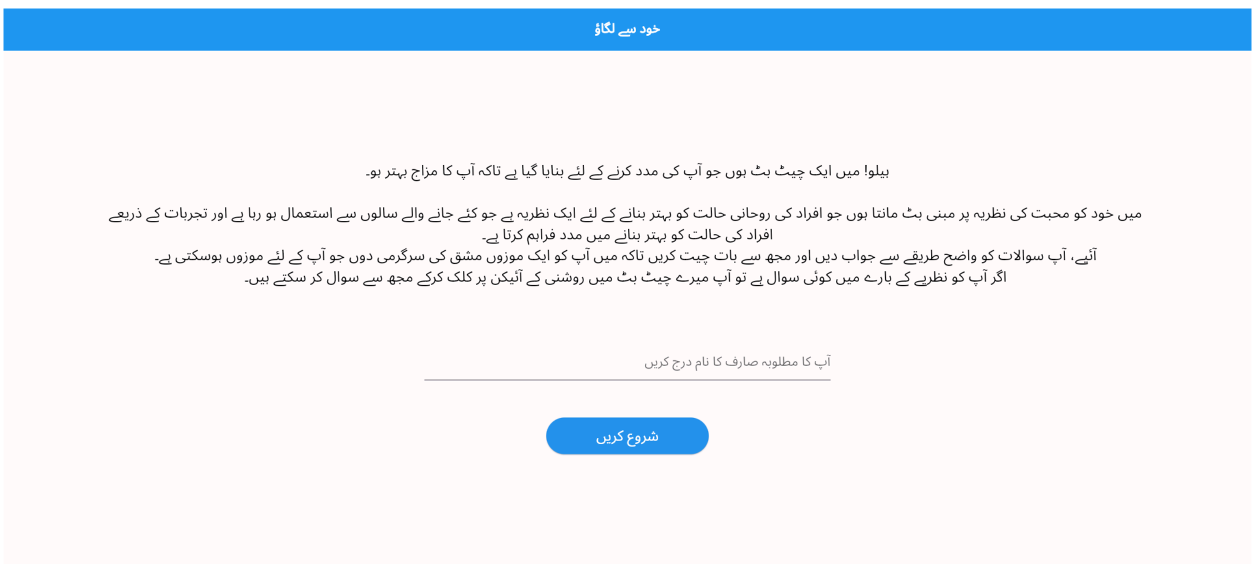
4. Demonstration & Results

4.1. Chatbot Interface

The AI-PsycheChat chatbot features an intuitive user interface designed to provide an engaging and accessible experience for users seeking mental health support through self-attachment therapy.

4.1.1. User Onboarding

Upon launching the chatbot, users are greeted with a welcoming introductory screen that prompts them to enter their name. This simple step helps personalize the experience and establishes an initial rapport between the user and the AI assistant.



4.1.2. Main Interface

After providing their name, users are presented with the main chatbot interface. The interface is visually appealing and organized, with distinct sections for displaying the conversation history, typing or speaking user inputs, and receiving the chatbot's responses.

4.2. SAT Protocol Recommendation

A core objective of AI-PsycheChat is to recommend appropriate self-attachment therapy (SAT) protocols based on the user's inputs and emotional state. As users engage in conversations, the chatbot analyzes their responses and gradually builds a list of suggested protocols.

The protocol recommendations are presented to the user at strategic points during the conversation, accompanied by clear instructions for practicing the recommended exercises. This seamless integration of evidence-based therapeutic techniques with natural language interaction enhances the overall effectiveness and user experience of the chatbot.

Chapter 5

5. Conclusions & Future Work

5.1. Conclusion

The AI-PsycheChat project represents a significant step towards making mental health support more accessible and inclusive for Urdu-speaking communities. By leveraging state-of-the-art natural language processing techniques and culturally tailored datasets, we developed an empathetic conversational agent capable of delivering self-attachment therapy protocols in the Urdu language.

Throughout this project, we encountered and overcame various challenges, ranging from dataset curation and annotation to model optimization and user interface design. The successful integration of speech recognition, text-to-speech, sentiment analysis, and context encoding models demonstrates the potential of artificial intelligence in bridging language and cultural barriers within the mental healthcare domain.

While the current implementation of AI-PsycheChat operates on a rule-based conversational framework, the modular architecture and robust natural language understanding capabilities pave the way for future iterations to incorporate more advanced dialog management systems and open-domain response generation.

5.2. Future Work

Building upon the foundation established by AI-PsycheChat, several promising avenues for future research and development emerge:

5.2.1. Comprehensive Evaluation and Refinement

To further validate the efficacy and user experience of AI-PsycheChat, it is crucial to conduct comprehensive evaluations involving mental health professionals and target user communities. These evaluations should assess the chatbot's performance across various dimensions, including emotional understanding, empathetic response generation, and adherence to established therapeutic principles.

Feedback from these evaluations can then be leveraged to refine the chatbot's conversational capabilities, expand the range of supported emotions and therapeutic protocols, and enhance the overall user experience.

5.2.2. Integration of Large Language Models (LLMs)

While the current iteration of AI-PsycheChat employs a rule-based approach, future versions could leverage the power of large language models (LLMs) like LLaMA (Language Model for Dialog Applications) developed by Facebook. These models have demonstrated remarkable capabilities in understanding and generating human-like conversations, potentially enabling more natural and contextually relevant dialogues within the mental health domain.

By fine-tuning and adapting these LLMs to the Urdu language and the specific therapeutic context, AI-PsycheChat could evolve into a more flexible and scalable conversational agent, capable of providing personalized support tailored to each user's unique needs and circumstances

5.2.3. Virtual Reality (VR) Integration

Another exciting avenue for future exploration is the integration of virtual reality (VR) technology with AI-PsycheChat. By creating immersive virtual environments and embodied conversational agents, users could engage with the chatbot in a more multisensory and interactive manner.

This approach could enhance the therapeutic experience by fostering a deeper sense of presence and connection, potentially leading to better engagement, retention, and overall efficacy of the self-attachment therapy protocols. Additionally, VR environments could be designed to simulate various scenarios or settings relevant to the therapeutic process, providing users with a safe and controlled space to practice and apply the learned coping strategies.

In conclusion, the AI-PsycheChat project represents a significant step towards democratizing access to mental health support for Urdu-speaking communities. While the current iteration demonstrates the potential of artificial intelligence in this domain, there remains ample scope for further research, development, and integration of cutting-edge technologies to enhance the chatbot's capabilities and user experience.

References

[1]. Law, A. J., Hu, R., Alazraki, L., Gopalan, A., Polydorou, N., & Edalat, A. (2022). A Multilingual Virtual Guide for Self-Attachment Technique. 2022 IEEE 4th International Conference on Cognitive Machine Intelligence (CogMI). https://doi.org/10.1109/CogMI56440.2022.00025

[2]. Alazraki, L., Ghachem, A., Polydorou, N., Khosmood, F., & Edalat, A. (2021). An Empathetic AI Coach for Self-Attachment Therapy. 2021 IEEE Third International Conference on Cognitive Machine Intelligence (CogMI). https://doi.org/10.1109/CogMI52975.2021.00019