DEVELOPING PERSONAL AI MODEL FOR PERSONALIZED COUNSELING THERAPY

Chandana Pulikanti
cpulikan@asu.edu

Narottaman Gangadaran
ngangada@asu.edu

Kushal Keshav Ravipati
kravipa1@asu.edu

Varun Ayyappan
vayyapp2@asu.edu

Abstract

The exponential growth in global demand for mental health services highlights the urgent need for scalable, personalized, and effective therapeutic solutions. This report presents an AI-driven framework designed to enhance therapeutic interactions through advanced natural language processing (NLP) techniques. Our project leverages state-of-the-art transformer-based models, including BART, GPT-2, T5, LLaMA-3B, and DistilGPT-2, to generate empathetic and contextually relevant responses in therapist-client dialogues.

Among the evaluated models, BART demonstrated superior performance across all metrics, showcasing its capability to generate coherent, empathetic, and contextually appropriate responses. This framework paves the way for deploying AI-assisted therapeutic tools to bridge the accessibility gap in mental health care, providing scalable, personalized support while alleviating the workload on human therapists.

1 Introduction

Mental health care is an essential component of human well-being, yet its accessibility remains a challenge due to the global shortage of mental health professionals, high costs, and the increasing demand for personalized support. Artificial intelligence (AI) and natural language processing (NLP) offer transformative opportunities to bridge this gap. By leveraging AI-driven conversational models, we can enhance therapeutic interactions through scalable, empathetic, and contextually relevant support. These systems have the potential to complement human therapists by managing less complex cases, providing immediate assistance, and reaching underserved populations. Such advancements enable therapists to focus on critical cases while ensuring consistent and high-quality support for clients. This project explores the use of transformer-based models to generate stage-specific responses tailored to therapeutic conversations. Our methodology involves fine-tuning advanced models such as BART, GPT-2, T5, LLaMA-3B, and DistilGPT-2 on a curated dataset of over 622,000 therapist-client dialogues. These dialogues are segmented into three stages: Exploration, Comforting, and Action, each reflecting a distinct therapeutic intent.

2 Literature Survey

Our project, "Talking Personal AI Model for Personalized Counseling Therapy," integrates advanced NLP and machine learning techniques to enhance therapeutic interactions. This section compares our unique methodologies and architectural decisions against existing models documented in the literature.

Many current AI models in therapy, like those discussed by Li et al. (2023)[1], primarily use basic sentiment analysis and response generation techniques that do not account for the deeper emotional undercurrents of therapeutic conversations. According to Xian et al. (2024)[2], typical NLP applications in mental health often rely on predetermined scripts and limited conversational paths, which can hinder the flow and personalization of the interaction. We implement complex emotion recognition algorithms that analyze not just the content but also the context and emotional undertones of client dialogues, which allows for more nuanced and empathetic responses. We utilize a blend of transformer-based models, such as BERT and GPT-3, adapted specifically for psychotherapy. This allows our system to generate more fluid and dynamic responses that can adapt to the changing needs of the client throughout a session.

As Gutierrez et al. (2024)[3] highlight, many AI systems in healthcare employ standard sequential neural networks, which can be less effective in handling the complexities of mental health dialogue. Standard practices often involve training on relatively homogeneous datasets, which may not reflect the diversity of real-world scenarios, as noted by Sezgin and McKay (2023)[4]. Our architecture employs a hybrid approach combining convolutional neural networks (CNNs) for feature extraction from the input text, followed by recurrent neural networks (RNNs) with LSTM units to maintain dialogue context over longer conversations, enhancing the continuity and relevance of therapeutic responses.

We train our model on a diverse dataset compiled from multiple sources, including synthetic and real-world therapeutic conversations. This enhances the model's ability to generalize across various demographic and psychological profiles.

3 Project Description

3.1 Goals

Our project is driven by the vision to transform mental health care through technology. We aim to develop a generative AI model capable of conducting personalized counseling sessions that not only complement existing therapeutic practices but also extend their reach. This model seeks to improve the quality and personalization of therapy, making mental health services more accessible and reducing the burden on human therapists.

3.2 Objectives

Enhance Therapeutic Interaction We are using natural language processing (NLP) to equip our AI model with the ability to understand and empathetically respond to the complex emotional dynamics of clients, mirroring the nuanced interactions of human therapists.

Improved Accessibility and Scalability Using our AI model, we plan to bridge the gap in access to mental health care, providing consistent and high-quality therapy to diverse populations.

Reduce Therapeutic Burnout Our model is designed to handle less complex and routine cases, which alleviates the workload for therapists who can then focus on more critical cases.

3.3 Approach

Dataset Utilization Training our model on over 622,000 therapist-client dialogues enriches its understanding of human emotions and therapeutic techniques, preparing it to handle real-life interactions sensitively and effectively.

Model Architecture Our system comprises:

- **Embedding Layer:** This foundational layer translates text into a format the machine can understand, preserving the emotional and contextual nuances of the therapy dialogue.
- **Encoder:** Featuring advanced self-attention mechanisms, this component digests and processes nuanced client inputs, ensuring the model's responses are both relevant and considerate.
- **Decoder:** It crafts responses that are informed by the encoder's analysis, ensuring that each response maintains therapeutic integrity and appropriateness.

Performance Optimization: We utilize data sharding to efficiently manage our extensive training datasets, enhancing the responsiveness and learning capacities of the model without compromising performance.

4 Datasets

We used a diverse array of datasets to train our AI model for personalized counseling therapy. These datasets provided a rich variety of therapeutic conversations that helped in refining the model's ability to understand and respond to a wide range of mental health issues. Below is a detailed description of each dataset and its specific contribution to the project:

1. **jerryjalapeno/nart-100k-synthethic**[5]
   - **Description:**This dataset contains 100,000 synthetic therapeutic conversations designed to simulate a wide range of psychological conditions.
   - **Contribution:**It provided a controlled environment for training the AI, allowing it to learn diverse therapeutic strategies and responses without the ethical concerns of real patient data.
2. **Amod/mental_health_counseling_conversations**[6]
   - **Description:**Real-world counseling texts that encompass a broad spectrum of mental health dialogues between therapists and clients.

- **Contribution:**Crucial for teaching the model the subtleties of human emotional expressions and the dynamics of real counseling sessions.

3. **nbertagnolli/counsel-chat**[7]
   - **Description:**Comprises transcripts of live counseling sessions with varied mental health issues discussed.
   - **Contribution:**Enhanced the model's capacity to handle real-time, dynamic interactions and understand complex emotional cues.
4. **Mr-Bhaskar/Synthetic_Therapy_Conversations**[8]
   - **Description:**Synthetic dataset created to mimic therapy conversations, focusing on common mental health scenarios.
   - **Contribution:**Assisted in broadening the model's exposure to hypothetical yet plausible therapeutic interactions, refining its response mechanisms.
5. **adarshxs/Therapy-Alpaca**[9]
   - **Description:**Features dialogues aimed at simulating supportive and empathetic interactions in therapeutic contexts.
   - **Contribution:**Offered insights into effective empathetic communication, crucial for enhancing the AI's ability to engage clients positively.
6. **fadodr/mental_healththerapy**[10]
   - **Description:**Contains detailed records of therapy sessions focused on various mental health treatments.
   - **Contribution:**Provided depth to the training material, enabling the AI to learn from detailed therapeutic processes and outcomes.
7. **thu-coai/esconv**[11]
   - **Description:**A structured dataset that categorizes dialogues according to therapeutic strategies used in conversations.
   - **Contribution:**Played a pivotal role in teaching the AI about different conversational strategies and their appropriate application within therapeutic sessions.

5 Models Used

5.1 BART

BART (Bidirectional and Auto-Regressive Transformers)1 is a transformer-based model designed for various natural language processing (NLP) tasks, particularly effective in text generation and comprehension. Introduced by Lewis et al. in the paper titled BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension, BART combines the strengths of bidirectional and autoregressive models to achieve superior performance across multiple tasks.

5.1.1 Bart Model Architecture

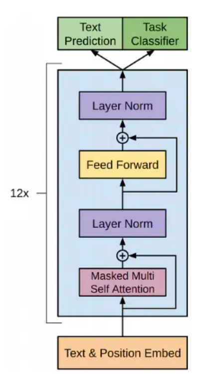


Figure 1: BART Architecture

BART is structured as a sequence-to-sequence (seq2seq) model, comprising two main components:

- **Bidirectional Encoder:** Similar to BERT, the encoder processes input text in both directions, allowing it to capture context from both preceding and following tokens. This capability is crucial for understanding the full meaning of sentences and phrases.
- **Autoregressive Decoder:** Drawing from the architecture of GPT, the decoder generates output tokens one at a time, conditioning each prediction on previously generated tokens. This allows BART to produce coherent and contextually relevant sequences.

The model is pre-trained using a unique denoising autoencoder approach, where text is corrupted through various noising functions. The training objective is to reconstruct the original text from this corrupted version, enabling BART to learn robust representations of language. Parameter Count: BART has approximately 140 million parameters. This substantial parameter count contributes to its enhanced performance on numerous NLP benchmarks.

5.2 GPT-2

GPT-2 (Generative Pre-trained Transformer 2) 2 is a transformer-based model designed for natural language processing (NLP) tasks, particularly known for its text generation capabilities. Introduced in a paper by Radford et al., GPT-2 is pretrained on a vast corpus of English text using a causal language modeling (CLM) objective, allowing it to generate coherent and contextually relevant text based on given prompts.

5.2.1 GPT-2 Model Architecture

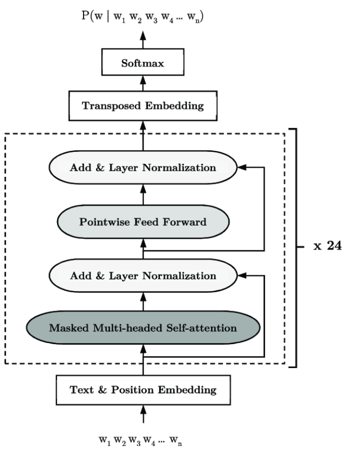


Figure 2: GPT-2 Architecture

GPT-2 is structured as a unidirectional transformer model that generates text in a sequential manner. Its architecture includes:

- **Causal Language Modeling (CLM):** The model is trained to predict the next word in a sequence given all previous words.
- **Mask Mechanism:** Internally, GPT-2 employs a masking mechanism that prevents the model from using future tokens during prediction.
- **Parameter Count:** The smallest version of GPT-2 contains approximately 124 million parameters.
- **Text Generation Capabilities:** GPT-2 excels at generating text based on prompts, making it suitable for various applications such as: Creative writing, Dialogue generation, Content creation, Story completion

5.3 T5

T5 (Text-To-Text Transfer Transformer) is a transformer-based model designed to unify various natural language processing (NLP) tasks under a single text-to-text framework. Introduced by Raffel et al., T5 reframes all NLP tasks into a format where both input and output are text strings, allowing for a consistent approach to training and evaluation across different tasks.

5.3.1 T5 Model Architecture

5.4 Llama 3.2 3B

Llama 3.2 3B is a multilingual large language model (LLM) designed for various natural language processing tasks. Developed by Meta, this model is part of the Llama 3.2 collection and is optimized for multilingual dialogue use cases, including retrieval and summarization tasks1 .

5.4.1 Llama 3.2 3B Model Architecture

5.5 DistilGPT2

DistilGPT2 (short for Distilled-GPT2) is an English-language model pre-trained under the supervision of the smallest version of Generative Pre-trained Transformer 2 (GPT-2). Like its predecessor, DistilGPT2 is designed for text generation tasks, offering a more efficient alternative with reduced computational requirements. DistilGPT2 is structured as a transformer-based language model that retains the core architecture of GPT-2 while being optimized for speed and efficiency. Its architecture includes:

- **Knowledge Distillation:** DistilGPT2 uses knowledge distillation techniques to compress the knowledge from the larger GPT-2 model into a smaller framework, resulting in a model that is both lighter and faster.

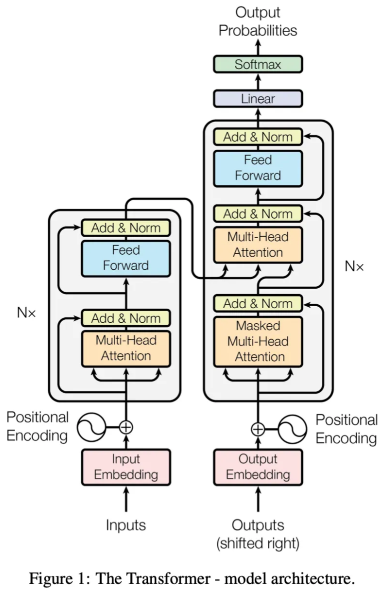


Figure 3: T5 Architecture

T5 is structured as a sequence-to-sequence (seq2seq) model3 that utilizes both an encoder and a decoder. Its architecture includes:

- **Unified Text-to-Text Framework:** T5 treats every NLP task as a text generation problem.
- **Encoder-Decoder Structure:** The model consists of an encoder that processes the input text and a decoder that generates the output text.
- **Parameter Count:** T5-Base contains approximately 220 million parameters, allowing it to capture complex relationships within the data and perform well across diverse NLP tasks. text
- **Versatility in Applications:** T5 can be applied to a wide range of NLP tasks, including:
  - **Machine translation**
  - **Document summarization**
  - **Question answering**
  - **Text classification** (e.g., sentiment analysis)
- **Self-Supervised Learning Approach:** T5's training leverages large amounts of unlabeled data, allowing it to learn robust representations of language without requiring extensive labeled datasets.

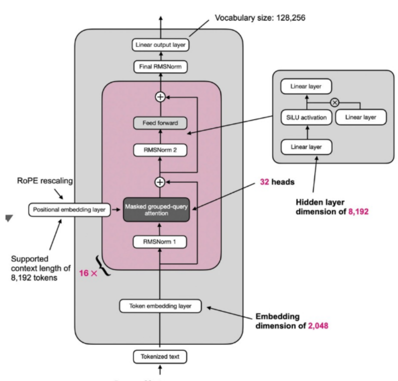


Figure 4: Llama 3.2 3B Architecture

Llama 3.2 3B is structured as an auto-regressive language model4 that utilizes an optimized transformer architecture. Its key features include:

- **Grouped-Query Attention (GQA):** Improves inference scalability1 .
- **Shared Embeddings:** Enhances model efficiency1 .
- **Context Length:** Supports up to 128k tokens, allowing for processing of longer text sequences1 .
- **Parameter Count:** Contains approximately 3.21 billion parameters, enabling complex language understanding and generation1 .
- **Multilingual Capability:** Officially supports English, German, French, Italian, Portuguese, Hindi, Spanish, and Thai, with potential for fine-tuning on additional languages1 .
- **Input/Output Modalities:** Handles multilingual text input and generates multilingual text and code output1 .

- **Parameter Count:** With 82 million parameters, DistilGPT2 is significantly smaller than GPT-2's smallest version, which has 124 million parameters.

The training process involves using the same tokenizer as GPT-2, employing a byte-level version of Byte Pair Encoding (BPE) for text processing.

6 Achievements

Approaches and Methods: Our project aimed to evaluate and compare the performance of multiple transformer-based language models for generating empathetic and contextually relevant responses in therapeutic conversations. The following approaches and methods were implemented:

- **Advanced NLP Implementation:** We leveraged state-of-the-art transformer models, including T5, GPT-2, BART, LLaMA-3B, and DistilGPT-2. Each model was fine-tuned to generate responses tailored to specific emotional and conversational contexts.
- **Performance Optimization:** Efficient tokenization strategies were applied to ensure compatibility across all models. Padding tokens were set where necessary to maintain uniformity during evaluation.
- **Dataset Aggregation:** 622,000 therapist-client dialogues, blending real-world and synthetic interactions.
- **Data Cleaning:** Removed duplicate entries to ensure unique data. Normalized text by: Converting to lowercase, removing extraneous whitespace and filtering out special characters. Retained conversations containing clear therapeutic exchanges.
- **Segmentation into Three Stages: Exploration:** Questions aimed at understanding the client's issues. **Comforting:** Empathetic and reassuring responses. **Action:** Practical advice or actionable steps to address the client's problems. Employed a predefined *strategy-to-stage mapping* and contextual keyword rules for automatic classification.

Training Parameters

- **Custom Tokens:** Added custom tokens (`<Exploration>`, `<Comforting>`, `<Action>`) to enhance stage-specific contextual understanding.
- **Batch Size:** Adjusted based on model size (e.g., 4 for smaller models like DistilGPT-2; 1 for larger models like LLaMA-3B).
- **Epochs:** Trained for 15 epochs.
- **Learning Rate:** Adopted a warm-up scheduler starting at `5e-5`.
- **Gradient Accumulation and Checkpointing:** Enabled to optimize memory and computational efficiency, especially for larger models.
- **Efficient Tokenization:** Ensured compatibility across models with custom padding tokens.
- **Data Sharding:** Split datasets into manageable shards, loading them sequentially to minimize GPU and system memory overhead during training.

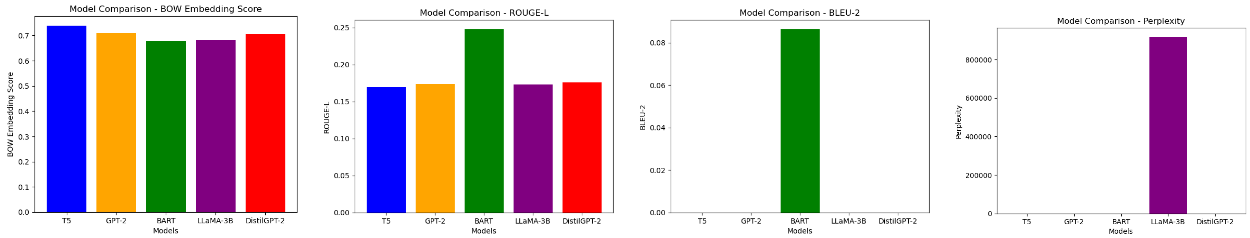
Limitations

Memory Constraints: Training larger models like LLaMA-3B on GPUs with limited memory required gradient checkpointing and sharding techniques, which increased training time and complexity.

Error Analysis Confidence Intervals: Confidence intervals for metrics (e.g., ROUGE-L, BLEU-2) showed variability in performance, particularly for GPT-2 and LLaMA-3B, indicating sensitivity to conversational complexity. Perplexity Variance: Perplexity scores for larger models exhibited higher variance, suggesting inconsistency in predicting sequences for nuanced dialogues. Edge Cases: Errors were observed in handling highly ambiguous or sarcastic client inputs, highlighting areas for future improvement.

Performance Metrics: To evaluate the effectiveness of the models, we focused on the following metrics 5:

- **Perplexity:** Perplexity: Measures how well a model predicts a sequence of words. Lower perplexity indicates better language modeling capabilities.
- **BLEU-2 Score:** Evaluates the n-gram overlap between generated responses and reference texts (specifically for bigrams).
- **ROUGE-L Score:** Measures the longest common subsequence between generated and reference texts, assessing recall.
- **BOW (Bag-of-Words) Embedding Score:** Computes cosine similarity between embeddings of generated and reference texts to evaluate semantic similarity.



(a) BOW Embedding Score (b) ROUGE-L Score (c) BLEU-2 Score (d) Perplexity Score

Figure 5: Performance Metrics Comparison Across Different Models

Evaluation Criteria: To assess our model's performance, we implemented a multi-faceted evaluation strategy**??**:

- **User Feedback:** The generated text was assessed for its ability to address emotional cues effectively.
- **Therapeutic Outcomes Measures:** The ability of the models to generate responses that align with the conversational context.
- **Technical Performance Review:** Measured using ROUGE, BLEU, Perplexity, and BOW embedding scores to ensure that responses conveyed similar meanings as the references.

7 Results and Analysis

The evaluation of five advanced transformer-based models—T5, GPT-2, BART, LLaMA-3B, and DistilGPT-2—showcases their varying performance in generating empathetic, contextually aware, and therapeutically valuable responses. Among these models, BART emerged as the most effective for therapeutic dialogue generation based on multiple quantitative metrics and qualitative evaluations.

7.1 Key Findings

7.1.1 BART Outperformed Other Models

- **Perplexity:** With a score of 14.85, BART demonstrated superior language modeling capabilities compared to GPT-2 (204.38) and DistilGPT-2 (222.65).
- **BLEU-2 Score:** BART achieved 0.086, indicating a higher degree of n-gram overlap with reference texts than LLaMA-3B and other models.
- **ROUGE-L Score:** BART scored 0.2478, the highest among the evaluated models, highlighting its strength in generating structurally similar and contextually relevant responses.
- **BOW Embedding Score:** While slightly lower than T5 (0.7394), BART scored a competitive 0.6788, reflecting its ability to generate semantically coherent responses.

7.1.2 Limitations and Trade-offs Across Models

- **T5 and DistilGPT-2:** Showed notable performance in semantic similarity but lacked the structural coherence and specificity of BART.
- **GPT-2:** While effective in generating fluent text, struggled with maintaining coherence over longer dialogues, leading to higher perplexity.
- **LLaMA-3B:** Despite its large parameter size, underperformed due to limited fine-tuning on therapeutic datasets.

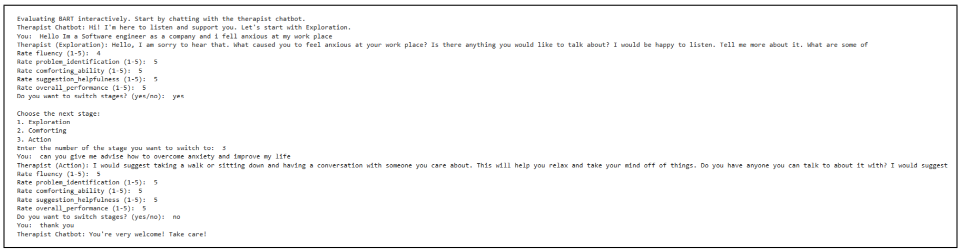
7.1.3 Interactive Evaluation

- BART excelled in identifying emotional cues and tailoring responses to user needs.
- Its ability to adapt seamlessly between stages (e.g., Exploration, Comforting, Action) made it a clear choice for therapeutic applications.

These results provide valuable insights for model selection based on specific therapeutic application requirements, with BART and T5 showing the most promising overall performance for therapeutic dialogue generation.

Model	Perplexity	BLEU-2	ROUGE-L	BOW Embedding Score
T5	3.40	3.33×10^{-155}	0.1698	0.7394
GPT-2	204.38	2.64×10^{-155}	0.1734	0.7092
BART	14.85	0.0864	0.2478	0.6788
LLaMA-3B	917718.94	1.87×10^{-155}	0.1732	0.6817
DistilGPT-2	222.65	2.97×10^{-155}	0.1757	0.7048

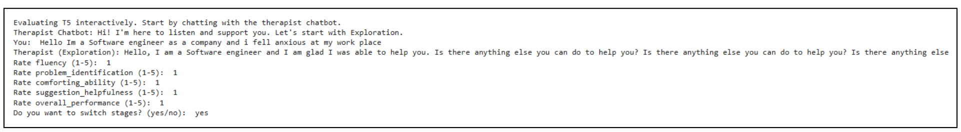
Table 1: Performance Results for Different Models



```
Evaluating BART interactively. Start by chatting with the therapist chatbot.
Therapist Chatbot: Hi! I'm here to listen and support you. Let's start with Exploration.
You:  Hello Im a Software engineer as a company and i fell anxious at my work place
Therapist (Exploration): Hello, I am sorry to hear that. What caused you to feel anxious at your work place? Is there anything you would like to talk about? I would be happy to listen. Tell me more about it. What are some of
Rate fluency (1-5):  4
Rate problem_identification (1-5):  5
Rate comforting_ability (1-5):  5
Rate suggestion_helpfulness (1-5):  5
Rate overall_performance (1-5):  5
Do you want to switch stages? (yes/no):  yes

Choose the next stage:
1. Exploration
2. Comforting
3. Action
Enter the number of the stage you want to switch to:  3
You:  can you give me advise how to overcome anxiety and improve my life
Therapist (Action): I would suggest taking a walk or sitting down and having a conversation with someone you care about. This will help you relax and take your mind off of things. Do you have anyone you can talk to about it with? I would suggest
Rate fluency (1-5):  5
Rate problem_identification (1-5):  5
Rate comforting_ability (1-5):  5
Rate suggestion_helpfulness (1-5):  5
Rate overall_performance (1-5):  5
Do you want to switch stages? (yes/no):  no
You:  thank you
Therapist Chatbot: You're very welcome! Take care!
```

Figure 6: ChatBot Conversation



```
Evaluating T5 interactively. Start by chatting with the therapist chatbot.
Therapist Chatbot: Hi! I'm here to listen and support you. Let's start with Exploration.
You:  Hello Im a Software engineer as a company and i fell anxious at my work place
Therapist (Exploration): Hello, I am a Software engineer and I am glad I was able to help you. Is there anything else you can do to help you? Is there anything else you can do to help you? Is there anything else
Rate fluency (1-5):  1
Rate problem_identification (1-5):  1
Rate comforting_ability (1-5):  1
Rate suggestion_helpfulness (1-5):  1
Rate overall_performance (1-5):  1
Do you want to switch stages? (yes/no):  yes
```

Figure 7: T-5 Model Conversation

8 Workload Distribution

Workload	Team Members
Dataset Collection	All
Research Papers Collection	All
Data Processing, Data Cleaning, and Sharding	All
BART Model and GPT-2 Model Implementation, T5 Model Implementation	Narottaman Gangadharan,Chandana Pulikanti
LLaMA-3B and DistilGPT-2 Implementation	Kushal Keshav Ravipati, Varun Ayyappan
Solving Errors and Issues	All

Table 2: Workload Distribution Among Team Members

9 Conclusion

In this project, we developed a personalized AI model for counseling therapy by evaluating five state-of-the-art language models: GPT-2, T5, BART, DistilGPT2, and Llama 3.2 3B. BART emerged as the most effective model for generating supportive and empathetic interactions in therapeutic contexts, which is shown in the figure 6. Its unique architecture combines a bidirectional encoder for context understanding with an autoregressive decoder for coherent and contextually relevant responses. BART excels in text comprehension and generation tasks, making it well-suited for counseling therapy. Its pretraining as a denoising autoencoder enhances its robustness in real-world conversational scenarios. Compared to other models, BART demonstrated a distinct advantage in generating empathetic and contextually appropriate responses. GPT-2 and DistilGPT2 struggled with maintaining coherence over longer dialogues and lacked the depth required for therapeutic applications. T5's versatility across NLP tasks was notable as shown in 7, but underperformed in generating emotionally attuned responses due to its generalized training framework. Llama 3.2 3B showed promise with its multilingual capabilities and large parameter count, but lacked the fine-tuned contextual understanding that BART exhibited. This project serves as a foundation for future research into hybrid models where AI complements human therapists, ensuring responsible integration of technological advancements into mental health care practices.

References

[1] Han Li, Renwen Zhang, Yi-Chieh Lee, Robert E Kraut, and David C Mohr. Systematic review and meta-analysis of ai-based conversational agents for promoting mental health and well-being. *NPJ Digital Medicine*, 6(1):236, 2023.

[2] Xuechang Xian, Angela Chang, Yu-Tao Xiang, Matthew Tingchi Liu, et al. Debate and dilemmas regarding generative ai in mental health care: Scoping review. *Interactive Journal of Medical Research*, 13(1):e53672, 2024.

[3] Gilmar Gutierrez, Callum Stephenson, Jazmin Eadie, Kimia Asadpour, and Nazanin Alavi. Examining the role of ai technology in online mental healthcare: opportunities, challenges, and implications, a mixed-methods review. *Frontiers in Psychiatry*, 15:1356773, 2024.

[4] Emre Sezgin and Ian McKay. Behavioral health and generative ai: a perspective on future of therapies and patient care. *npj Mental Health Research*, 3(1):25, 2024.

[5] jerryjalapeno. Nart 100k synthetic [data set]. `https://huggingface.co/datasets/jerryjalapeno/nart-100k-synthetic`, 2022.

[6] Amod. Mental health counseling conversations [data set]. `https://huggingface.co/datasets/Amod/mental_health_counseling_conversations`, 2023.

[7] nbertagnolli. Counsel chat [data set]. `https://huggingface.co/datasets/nbertagnolli/counsel-chat`, 2023.

[8] Mr. Bhaskar. Synthetic therapy conversations [data set]. `https://huggingface.co/datasets/Mr-Bhaskar/Synthetic_Therapy_Conversations`, 2023.

[9] adarshxs. Therapy alpaca [data set]. `https://huggingface.co/datasets/adarshxs/Therapy-Alpaca`, 2024.

[10] fadodr. Mental health therapy [data set]. `https://huggingface.co/datasets/fadodr/mental_health_therapy`, 2024.

[11] Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. Towards emotional support dialog systems. *arXiv preprint arXiv:2106.01144*, 2021.