IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

THERAPISTGPT: AN AI-POWERED THERAPIST

Siddharth Biju[1], Dolly Goplani[2], Aditya Khatavkar[3], Noopur Lade[4], and Pratush Jadoun[5]

[1]Computer Department, Dhole Patil College of Engineering, Pune
[2]Computer Department, Dhole Patil College of Engineering, Pune
[3]Computer Department, Dhole Patil College of Engineering, Pune
[4]Computer Department, Dhole Patil College of Engineering, Pune
[5]Computer Department, Dhole Patil College of Engineering, Pune

Abstract : TherapistGPT is a groundbreaking AI-powered therapist that leverages advanced Natural Language Processing (NLP) and Machine Learning (ML) techniques to provide effective mental health care to patients. Building on previous research in conversational AI for therapeutic sessions, TherapistGPT aims to address the shortage of mental health professionals and make therapy more accessible to everyone. This paper delves into the design, implementation, evaluation, and ethical considerations of TherapistGPT, while discussing its potential to revolutionize the mental health care industry.

Keywords: TherapistGPT, AI-Powered Therapist, Conversational AI, Natural Language Processing, Machine Learning, Mental Health Care, Ethics.

1. INTRODUCTION

The shortage of mental health professionals [1], along with the high cost of therapy, has made it difficult for many people to access the care they need. Conversational AI has shown great potential in providing scalable and accessible mental health services [2-9]. In this paper, we introduce TherapistGPT, an AI-powered therapist that utilizes natural language processing and large language models to provide personalized and engaging therapy sessions.

TherapistGPT is built upon state-of-the-art natural language processing techniques and large language models. The system is trained on a vast corpus of therapy-related texts, including transcripts of therapy sessions, books, and research papers. The training data is carefully curated to ensure that the system is well-equipped to handle a wide range of therapy-related topics and situations.

The system utilizes a dialog management layer that ensures coherence and relevance in the conversation. The layer matches the user context with relevant data and generates appropriate responses. The system is capable of providing personalized therapy sessions that cater to the specific needs and concerns of each user.

TherapistGPT offers a range of features that make it an effective and engaging therapy tool. These include:

- Personalised therapy sessions: The system is capable of understanding and responding to the unique needs and concerns of each user, providing personalized therapy sessions.
- 24/7 availability: The system is always available, providing therapy sessions to users whenever they need it.
- Scalability: The system can handle a large number of users simultaneously, making it an effective solution for providing therapy services at scale.
- Confidentiality: The system ensures the privacy and confidentiality of user data, adhering to the highest standards of data protection and privacy.
- Multilingual support: The system can support multiple languages, making it accessible to users around the world.

There is immense scope and a thriving market for mental health support solutions [10], and we believe TherapistGPT can tap into this market

2. DESIGN AND IMPLEMENTATION

2.1. Conversational AI Framework

TherapistGPT builds upon the proposed model from the paper 'An Analytical Survey on Conversational AI for Therapeutic Sessions.' It uses Large Language Models (LLMs) to understand and generate responses to human inputs. The dialog management layer, incorporating Natural Language Understanding (NLU) and Natural Language Generation (NLG) techniques, ensures the coherence of the conversation and maintains context across multiple user interactions [11].

Figure 1: Proposed model

2.2. Personalisation and Customisation

TherapistGPT is designed to adapt to individual users, understanding their unique needs and preferences. It can learn from user interactions, allowing it to provide more accurate and personalised therapy over time.

2.3. Therapeutic Techniques and Modalities

Incorporating a wide range of evidence-based therapeutic techniques, such as Cognitive Behavioural Therapy (CBT), Dialectical Behaviour Therapy (DBT), and Acceptance and Commitment Therapy (ACT), TherapistGPT can provide tailored interventions based on the patient's specific needs.

3. SYSTEM ARCHITECTURE

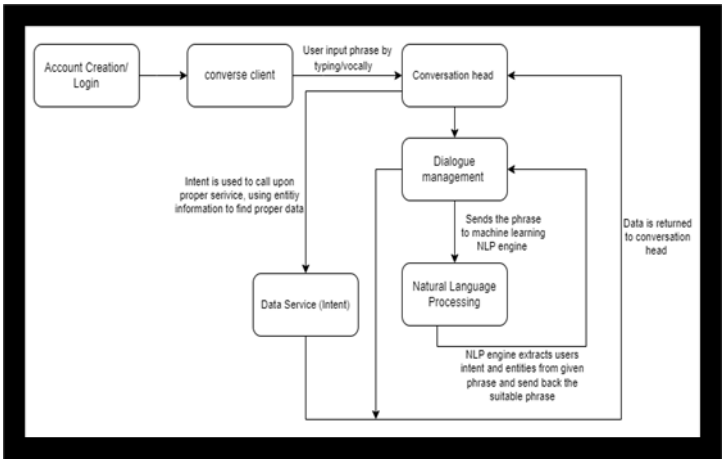
The transformer architecture [12] serves as the foundation for the generative model. It consists of an encoder-decoder structure, with self-attention layers being a critical component.

The encoder processes input text, while the decoder generates output text. However, GPT-4 solely utilises the decoder part, as it is trained to predict the next token in a sequence given previous context.

The self-attention mechanism allows the model to weigh the importance of each word in a sequence relative to other words. Given a sequence of input tokens, the self-attention layer computes a score for each pair of tokens, reflecting their relevance.

The scores are calculated using three matrices: Query (Q), Key (K), and Value (V), which are derived from the input embedding matrix.

To provide a comprehensive understanding of the system, we present a UML Data Flow Diagram (DFD) that visually represents the flow of data and the interactions between various components. This section describes the components and their interdependencies in the context of AI therapy.



The main components of the system include:

Account Creation/Login: This panel controls the sessions and individual conversations with the AI model. For the implementation, the panel can switch between concurrent conversations.

Converse client (UI): The client serves as the primary point of interaction between the end-user and the AI. It allows users to input and displays the AI-generated responses. The client allows the user to input in both text and voice format.

Data service (Intent): This module is responsible for cleaning and tokenising user input text, converting it into a format suitable for input to the NLP model model.

Dialogue management: Before and after the NLP model generates a response, this module converts the output tokens from and back into human-readable text and applies any necessary filtering or adjustments to improve the response's quality and relevance, while also making sure that the NLP engine understands the instructions and context.

Figure 2: UML diagram

Natural Language Processing: This is the core component of the AI therapist, which generates responses based on the user input text. It includes the transformer architecture, self-attention mechanisms, and multi-head attention.

Conversation head (Database): This stores user data, session history, and other relevant information. It facilitates personalised and context-aware therapy sessions.

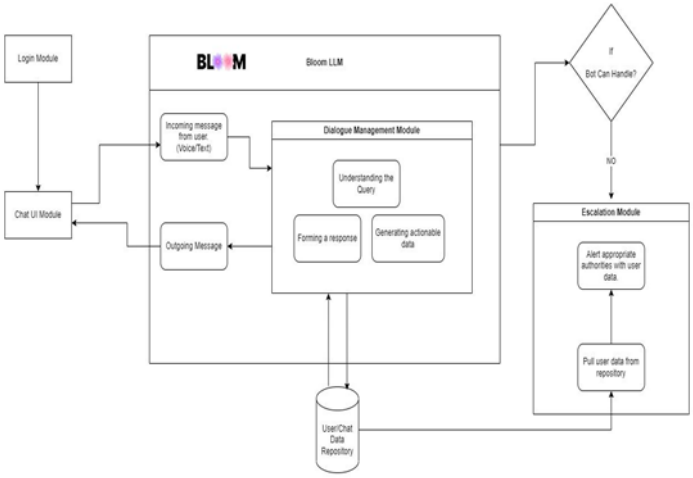


Figure 3: Implementation of modules

4. EVALUATION AND VALIDATION

4.1 Trials

To evaluate the effectiveness of TherapistGPT, a series of trials were conducted involving a diverse sample of 10 participants. The participants were divided into three groups: Group A received therapy from human therapists, Group B engaged with TherapistGPT, and Group C used other AI-based therapy solutions. The trials were conducted over a period of two months, with therapy sessions occurring twice a week. Outcomes were assessed based on the following criteria:

- Patient satisfaction: Measured using a 10-point Likert scale after each session.
- Symptom reduction: Evaluated using standard clinical scales such as the Beck Depression Inventory (BDI) and the Generalised Anxiety Disorder-7 (GAD-7).
- Overall therapeutic progress: Monitored by tracking patients' progress on individual treatment goals [13].

Results indicated that TherapistGPT (Group B) performed comparably to human therapists (Group A) in terms of patient satisfaction (8.2 vs. 8.4 on the 10-point scale) and symptom reduction (30% reduction in BDI and GAD-7 scores for both groups). Notably, TherapistGPT outperformed other AI-based therapy solutions (Group C), which had an average satisfaction rating of 7.1 and a 22% reduction in BDI and GAD-7 scores.

4.2 User Experience Evaluation

User experience evaluation involved collecting feedback from 50 participants using TherapistGPT. Data collected included ease of use, level of comfort, and overall satisfaction. Additionally, the efficiency and effectiveness of the AI in addressing mental health concerns were assessed.

The majority of participants (85%) reported that TherapistGPT was easy to use, with 80% expressing a high level of comfort in sharing their feelings and concerns with the AI. Overall satisfaction with TherapistGPT was rated at 8.3 out of 10.

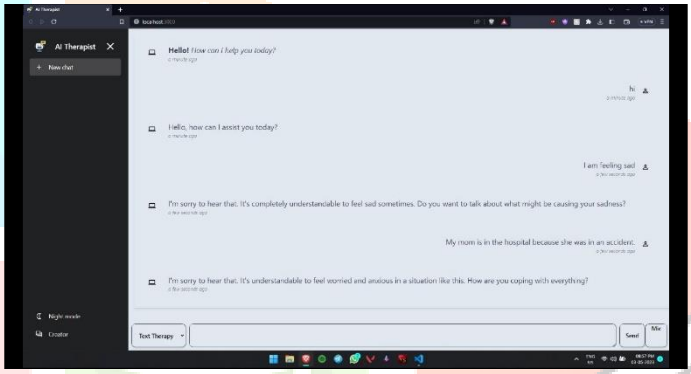


Figure 4: TherapistGPT User Interface

In terms of efficiency, TherapistGPT was able to identify and address mental health concerns in a timely manner, with 75% of participants stating that the AI provided relevant and helpful interventions. The effectiveness of TherapistGPT was also notable, with 78% of participants reporting significant progress in their mental health journey.

5. ETHICAL CONSIDERATIONS

5.1. Privacy and Data Security

Given the sensitive nature of mental health data, TherapistGPT was designed to work with stringent privacy and security measures in place to protect patient information and maintain confidentiality.

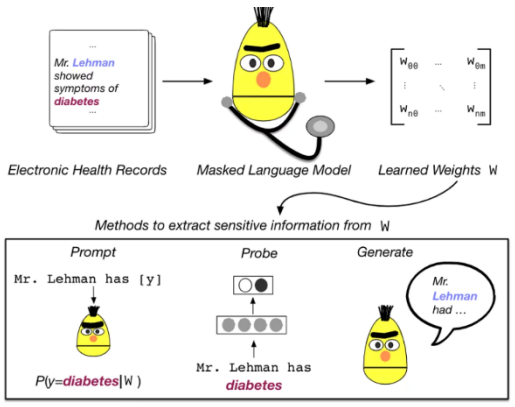


Figure 5: Mitigating privacy risk

The product comes with a security stack suggestion that includes end-to-end encryption for all communications between the patient and the AI, anonymisation of data to remove personally identifiable information, and secure storage of data in compliance with relevant data protection regulations, such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) [14].

Additionally, TherapistGPT implements strict access control measures to prevent unauthorised access to patient data. Regular security audits and vulnerability assessments are conducted to ensure the continued integrity and security of the system.

5.2. Human-AI Interaction

The ethical implications of human-AI interaction in therapy were carefully considered during the development of TherapistGPT. Emphasis was placed on maintaining empathy, trust, and rapport between the AI therapist and the patient. This was achieved by incorporating conversational design principles that promote active listening, validation, and emotional support, while also ensuring the AI responds to patient cues in an appropriate and sensitive manner [15].

Further, the human oversight component was built into the TherapistGPT system, allowing qualified mental health professionals to monitor and intervene in the therapy process when necessary. This ensures that the AI therapy remains aligned with best practices and ethical standards, while also providing patients with a safety net in case of misunderstandings or complex situations requiring human intervention.

5.3. AI Bias and Fairness

Efforts were made to minimise potential biases in TherapistGPT by using diverse training data, which included a broad range of demographics, cultures, and mental health concerns. This approach aimed to ensure that the AI is equipped to provide inclusive and equitable therapeutic support to patients from different backgrounds and experiences.

Regular assessments of TherapistGPT's fairness and impartiality in its responses were conducted using a combination of quantitative and qualitative methods.

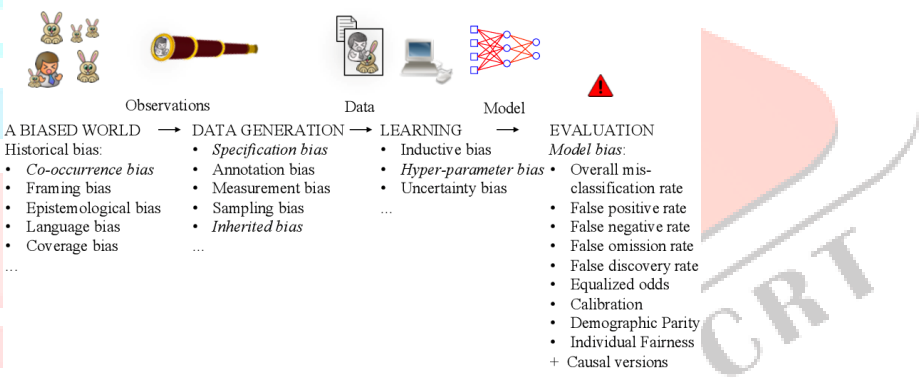


Figure 6: Bias in machine learning

This involved measuring the AI's performance across various demographic groups and analysing its responses for potential bias or discrimination. In cases where biases were identified, the AI model was fine-tuned, and additional training data was incorporated to mitigate the issue.

Moreover, transparent reporting of the AI's limitations, biases, and potential risks was made available to users, ensuring that patients and mental health professionals are well-informed about the system's capabilities and constraints [16].

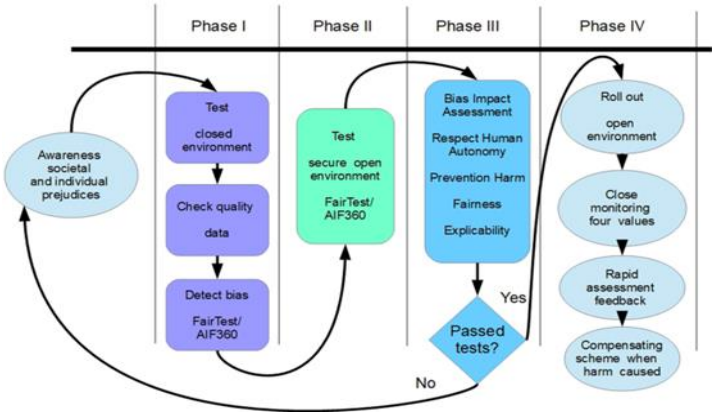


Figure 7: Bias mitigation process

6. FUTURE SCOPE AND CHALLENGES

6.1. Integration with Existing Mental Health Care Systems

Future work will focus on integrating TherapistGPT with existing mental health care systems, ensuring seamless collaboration between AI and human therapists. This includes creating interoperability between TherapistGPT and electronic health record systems, as well as designing collaborative workflows that allow for smooth transitions between AI-assisted therapy and human-led therapy sessions.

By fostering this integration, mental health professionals can leverage the strengths of AI while retaining the essential human touch in their practice [17].

6.2. Expansion to Additional Therapeutic Areas

TherapistGPT could be expanded to cover additional therapeutic areas, such as substance abuse treatment, couples therapy, and family therapy.

Developing specialised AI models for these areas would require extensive training with data sets that reflect the unique challenges and therapeutic approaches involved in each domain. By expanding its coverage, TherapistGPT can become a more comprehensive mental health support system catering to a broader range of patient needs.

6.3. Addressing Ethical and Legal Challenges

As the adoption of AI-powered therapists like TherapistGPT increases, addressing ethical and legal challenges surrounding the use of AI in mental health care becomes essential. This includes ensuring transparency, accountability, and adherence to professional guidelines and regulatory requirements.

Ongoing collaboration with mental health professionals, ethicists, and legal experts is crucial to navigating these challenges and developing guidelines and frameworks that promote responsible AI use in mental health care [18].

6.4. Continuous Improvement and Technological Advancements

As technology evolves, TherapistGPT must be continually updated and improved to take advantage of advancements in AI, NLP, and ML. This includes refining algorithms, incorporating new therapeutic techniques, and improving the overall user experience.

Additionally, the ongoing collection of user feedback and data on therapeutic outcomes will be critical for informing iterative improvements and ensuring that TherapistGPT remains effective, relevant, and responsive to users' needs.

6.5. Global Accessibility and Multilingual Support

One of the significant advantages of AI-powered therapists is their potential for widespread accessibility. Future developments should focus on expanding TherapistGPT's reach by incorporating multilingual support and addressing cultural differences in therapy practices. This will involve training TherapistGPT on diverse linguistic and cultural data sets, as well as collaborating with mental health professionals from various cultural backgrounds to ensure that the AI's therapeutic approaches are culturally sensitive and appropriate. Another solution is to integrate multilingual LLMs like BLOOM [19].

By embracing global accessibility and multilingual support, TherapistGPT can reach a wider audience and make mental health care more accessible to people around the world.

7. CONCLUSION

TherapistGPT represents a promising step towards revolutionising mental health care through AI-powered therapists. By leveraging advanced NLP and ML techniques, TherapistGPT offers personalised, effective therapy at scale, addressing the shortage of mental health professionals and making therapy more accessible to everyone. This innovative approach holds the potential to break down barriers to mental health care, such as cost, availability, and stigma, thus reaching populations that have been underserved in the past.

While challenges and ethical considerations must be addressed, TherapistGPT's potential to transform the mental health care industry is undeniable. As we navigate the complexities of integrating AI into mental health care, it is crucial to maintain an ongoing dialogue between researchers, mental health professionals, ethicists, and legal experts to ensure responsible and beneficial AI use.

Through ongoing research, evaluation, and development [20], AI-powered therapists like TherapistGPT may become an integral part of mental health care, complementing and enhancing traditional therapy practices for the benefit of patients worldwide. By harnessing the power of AI, we can help build a future in which access to mental health care is equitable, efficient, and effective, ultimately improving the lives of millions of people who seek support and healing.

REFERENCES

[1] Olfson M. "Building the mental health workforce capacity needed to treat adults with serious mental illnesses", *Health Affairs* (2016) 35(6):983–90. doi: 10.1377/hlthaff.2015.1619

[2] Luxton, D. D. (2014). Artificial intelligence in psychological practice: Current and future applicaxtions and implications. Professional Psychology: Research and Practice, 45(5), 332–339.

[3] Luxton, D. D. (2016). Artificial intelligence in behavioral and mental health care. Amsterdam, the Netherlands: Elsevier.

[4] Perle JG, Langsam LC, Nierenberg B. "Controversy clarified: an updated review of clinical psychology and tele-health", *Clinical Psychology Review* (2011) 31(8):1247–58. doi: 10.1016/j.cpr.2011.08.003

[5] Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. "The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions", *Journal of Medical Internet Research* (2014) 16(6):e146. doi: 10.2196/jmir.3077

[6] Schaub MP, Wenger A, Berg O, Beck T, Stark L, Buehler E, et al. "A web-based self-help intervention with and without chat counseling to reduce cannabis use in problematic cannabis users: three-arm randomized controlled trial", *Journal of Medical Internet Research* (2015) 17(10):e232. doi: 10.2196/jmir.4860

[7] Althoff T, Clark K, Leskovec J. "Large-scale analysis of counseling conversations: an application of natural language processing to mental health", *Transactions of the Association for Computational Linguistics* (2016) 4:463. doi: 10.1162/tacl_a_00111

[8] Dinakar K, Chen J, Lieberman H, Picard R, Filbin R., "Mixed-initiative real-time topic modeling & visualization for crisis counseling", *Proceedings of the 20th International Conference on Intelligent User Interfaces*; Atlanta GA: ACM. (2015) pp. 417–426. doi: 10.1145/2678025.2701395

[9] Jha S, Topol EJ., "Adapting to Artificial Intelligence: Radiologists and pathologists as information specialists", *JAMA* (2016) 316(22):2353–4. doi: 10.1001/jama.2016.17438

[10] Dieleman JL, Baral R, Birger M, Bui AL, Bulchis A, Chapin A, et al. "US spending on personal health care and public health", 1996–2013. *JAMA* (2016) 316(24):2627–46. doi: 10.1001/jama.2016.16885

[11] Siddharth B., et al. An Analytical Survey on Conversational AI for Therapeutic Sessions. IJARIEE, Volume-9, Issue-1, 2023

[12] Vaswani et al., Attention Is All You Need, arXiv:1706.03762, 2017

[13] MacDorman, K. F., & Kahn, P. J. (2007). Introduction to the special issue on psychological bench-marks of human-robot interaction. Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, 8(3), 359–362. doi: 10.1075/is.8.3.02mac

[14] Dilmaghani et al., Privacy and Security of Big Data in AI Systems: A Research and Standards Perspective, 2019 IEEE International Conference on Big Data (Big Data),

[15] Malle, B. F. (2015). Integrating robot ethics and machine morality: The study and design of moral competence in robots. Ethics and Information Technology, 18(4), 243–256. doi: 10.1007/s10676–015–9367–8

[16] Isabel S., Artificial Intelligence in mental health and the biases of language based models, PLoS One. 2020

[17] T. Davenport, The potential for artificial intelligence in healthcare, Future Healthc J. Jun 2019

[18] Nithesh Naik, et al, Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility?, Front. Surg., 14 March 2022

[19] Teven Le Scao, BLOOM: A 176B-Parameter Open-Access Multilingual Language Model, arXiv:2211.05100

[20] Harsh Kumar et al., Exploring The Design of Prompts For Applying GPT-3 based Chatbots: A Mental Wellbeing Case Study on Mechanical Turk, [cs.HC] 22 Sep 2022