

Emotional AI Chatbots: User Experience and Effectiveness

Group Members

- **Group 12 :** Ahmet Yavuz Lülecioğlu, Ahmet Burak Kurtulmuş, Uğurhan Altınbaş

Abstract

This study investigates the user experience and effectiveness of emotional AI chatbots. Participants interacted with various emotional chatbots in different scenarios, and their satisfaction, connection ability, and ease of use were evaluated. The purpose of this study is to assess how emotionally satisfying the chatbots are for users and how effectively these chatbots understand issues and demonstrate empathy. Participants were selected from a diverse group of volunteer users to ensure a broad range of perspectives. They interacted with emotional chatbots, including *pi.ai*, *The Deepen App*, and *replika.ai* in predetermined scenarios designed to test different aspects of emotional interaction. Preliminary findings indicate significant differences in user satisfaction and the ability to form emotional connections among the different emotional chatbots tested.

2. Background of the Study

The field of artificial intelligence (AI) has seen remarkable advancements, particularly in the development of chatbots with emotional intelligence. These AI chatbots, designed to recognize, interpret, and respond to human emotions have become pivotal in enhancing human-machine interactions across various domains such as mental health care, customer service, and workplace environments.

The integration of emotional intelligence into chatbots is a critical development in AI as emotional AI chatbots are programmed to understand and simulate human emotions, providing more natural and effective interactions. This capability is increasingly important as users seek more

personalized and empathetic digital interactions. The ability of chatbots to engage emotionally with users not only enhances user satisfaction but also improves the overall effectiveness of the services provided.

One of the most significant applications of emotionally intelligent chatbots is in the field of mental health care. The COVID-19 pandemic has exacerbated mental health issues globally, highlighting the need for accessible mental health support. According to Bilquise, Deris, and Shaalan (2022), empathic chatbots can provide 24/7 support, understand and react to users' emotions, and offer cost-effective management for minor mental health issues such as stress. These chatbots can also serve as a bridge to human professionals during crises, utilizing technologies like sentiment analysis and emotion recognition to assess and respond to users' emotional states .

Emotionally intelligent chatbots can also be utilized in mental health counseling since the ability of chatbots to disclose emotions significantly impacts user satisfaction and reuse intention. Park, Chung, and Lee (2022) found that chatbots capable of emotional disclosure enhance user satisfaction and foster a sense of intimacy, which is crucial for encouraging continued use of these services. This emotional engagement improves the overall effectiveness of mental health interventions, demonstrating the importance of emotional intelligence in AI chatbots.

Finally, the introduction of conversational AI platforms like Amazon Alexa has also contributed significantly to the development and deployment of emotionally intelligent chatbots. According to McTear, Callejas, and Griol (2016), Alexa has shown how voice-activated systems can enhance user experience by providing personalized and contextually relevant responses. This capability has been crucial in making interactions more engaging and emotionally fulfilling.

2.1 Purpose of the Study

The current study was conducted to evaluate the emotional fulfillment provided by emotional chatbots across different contexts. Specifically, it aims to:

1. Assess the emotional satisfaction and connection between users and chatbots.
2. Determine the usefulness of chatbots in addressing emotional issues.
3. Compare the effectiveness of different chatbots in understanding and empathizing with users.
4. Identify the factors that contribute to a chatbot's ability to connect emotionally with users.

2.2 Significance of the Study

This study is significant because it addresses the growing use of AI chatbots in emotionally sensitive contexts such as mental health care and analyzes how well a person can connect with an AI. The findings will provide valuable insights into the design and implementation of more empathetic and effective chatbots. Understanding user experiences and satisfaction with emotional chatbots can inform future developments in AI, ensuring these technologies enhance emotional well-being and foster more natural, human-like interactions. Additionally, the study's results could guide policymakers in developing regulations that protect user privacy and emotional integrity .

3. Methodology

Three different emotional AI chatbots were selected to be compared in the study. These chatbots were hand-picked so that the results could be compared with the related literature. For general usage and to assess the efficiency of the chatbots' personalization features, Pi chatbot was used (pi.ai). To analyze how well an emotional chatbot designed to be helpful in mental healthcare performs, the Deepen AI mobile app was selected. Lastly, to assess the benefits of an emotional conversational AI, the Replika chatbot was utilized (replika.ai). These chatbots were then evaluated by user-based testing and getting both quantitative and qualitative feedback from users. Chatbots were assigned to users randomly, and users were picked randomly from a wide range of backgrounds on AI knowledge and history. It is worth noting that due to the limitations on where the testing was conducted, all users were university students.

3.1 Research Questions

The most important purpose of the study was to understand the emotional capabilities of emotional chatbots, so the research questions motivating the study were related to concepts about emotions such as emotional fulfillment and satisfaction. Since the comparison of AI models was also a priority, understanding the strengths and weaknesses of these chatbots were important as well.

- How emotionally fulfilling is talking to an emotional chatbot for a typical user?
- How satisfactory is the connection made between the chatbot and the user?
- How useful is talking to a chatbot when dealing with an emotional issue?
- Which chatbot is better at understanding the issue and being empathetic?

3.2 Participants

The participants of the study were 21 university students with varying ages and genders. The participants had different past experiences with chatbots, and most have not used any emotional chatbots in the past. The participants were volunteers. The basic template for the personas, and the participants' information are given below.

- **Persona(s):**
 - Name
 - Age
 - Gender
 - Background
 - Chatbot Experience
 - Emotional Chatbot Experience
- **Persona 1:**
 - 21
 - Male
 - Has used chatbots for academic purposes and information gathering.
 - He was familiar with Emotional AI's but had never used them before.
 - University student, freshman.
- **Persona 2:**
 - 23
 - Male
 - Has used ChatGPT before for academic purposes, personal interest, and to use its prompt-to-image features.
 - He had no prior knowledge on emotional AI's.
 - University student, freshman.
- **Persona 3:**
 - 22
 - Male
 - Has used chatbots like character.ai for fun, and has used ChatGPT for academic purposes and to discover places.
 - Has used the emotional chatbot character.ai before for psychological needs.
 - University student, senior.
- **Persona 4:**

- 20
- Male
- Has used chatbots to aid him in his homeworks like summarizing papers.
- Has never used an emotional chatbot.
- University student, freshman.
- **Persona 5:**
 - 18
 - Male
 - Used Chatbots to find questions that he cannot normally find on the internet. He has also used them for homeworks since they can summarize papers and do all sorts of things to help, like finding papers.
 - Has heard of emotional AI but never used them.
 - University student, preparation year.
- **Persona 6:**
 - 18
 - Male
 - Has very rarely used ChatGPT and considers himself not used.
 - He has heard of Snapchat's AI. He also knows of healthcare chatbots used for psychological needs.
 - University student, freshman.
- **Persona 7:**
 - 25
 - Male
 - Has used them for academic purposes and productivity.
 - Never used emotional AI before.
 - University student, graduate
- **Persona 8:**
 - 23
 - Male
 - Has used them for planning and information gathering.
 - Has minimal experience, generally avoids emotional AI interactions.
 - University student, senior.
- **Persona 9:**
 - 25

- Female
- Has used informational and interactive chatbots.
- Frequently uses emotional AI for emotional support and companionship.
- University student, graduate.
- **Persona 10:**
 - 23
 - Male
 - Has used them for practical purposes.
 - Limited to a few experimental interactions.
 - University student, senior.
- **Persona 11:**
 - 23
 - Male
 - Has used them for basic inquiries and support.
 - Has never used them before.
 - University student, senior.
- **Persona 12:**
 - 19
 - Female
 - Has used chatbots for academic purposes and for information retrieval purposes.
 - Has used Snapchat's AI as a friend.
 - University student, freshman.
- **Persona 13:**
 - 21
 - Male
 - Has used ChatGPT only for academic purposes and personal use like creative writing.
 - Has never used any emotional AI before.
 - University student, senior.
- **Persona 14:**
 - 24
 - Female
 - Has used for help with education and assignments since ChatGPT is free
 - Has never used before.
 - Senior undergraduate.

- **Persona 15:**
 - 21
 - Male
 - Has used SimiSimi and ChatGPT, and uses them more for homework aid.
 - Has never heard of them
 - University student, freshman.
- **Persona 16:**
 - 21
 - Has used mostly for class and homework aid.
 - Has never heard of them.
 - University student, freshman.
- **Persona 17:**
 - 21
 - Male
 - Has moderate experience with academic and social chatbots like ChatGPT.
 - University student, freshman.
- **Persona 18:**
 - 21
 - Male
 - Has extensive experience using chatbots for a variety of tasks, primarily academic and informational.
 - No prior experience with emotional chatbots, but open to trying them.
 - University student, freshman.
- **Persona 19:**
 - 24
 - Female
 - Has used for academic purposes, and only used ChatGPT.
 - Has never used before.
 - University student, senior undergraduate.
- **Persona 20:**
 - 23
 - Female
 - Has used ChatGPT for academic purposes and for prompt-to-image features (Dall-E)
 - Has never used emotional AI before.
 - University student, senior undergraduate.
- **Persona 21:**

- 23
- Female
- Has used ChatGPT & Gemini for academic purposes and idea generation.
- Never used before.
- University student, senior undergraduate.

3.3 Materials

For the study, three emotional chatbot models were utilized: *pi.ai*, *The Deepen App*, and *replika.ai*. For *pi.ai* and *replika.ai*, the personal laptop of one of the researchers was used as a control variable. Again, for *The Deepen App*, the personal mobile phone of one of the researchers was used to easily conduct the experiments.

To gather feedback from the users, the Chatbot Usability Questionnaire was used for quantitative analysis, and eight questions regarding the emotional aspects of the tested chatbots were asked to users for qualitative analysis. For more information regarding these questions, please refer to Section 3.6.

3.4 Pilot Study

A pilot study was conducted to test the methodology and scenarios. Initially, it was decided that every user would test all three of the specified chatbots, and then compare their experiences with each of them. During pilot testing, however, it was concluded that the tests were taking a longer time than it was initially anticipated, so it was decided that users would be assigned to a single AI model randomly. After this finding, time was kept to see the average testing time of a user to better plan the testing sessions. It was found that each test would take roughly 45 minutes to complete, and consequently, users were accurately informed on how long the test would take. No other problems or errors were encountered during the pilot study.

3.5 Data Collection

To be able to accurately compare the emotional chatbots' empathy levels and responses, predetermined scenarios were given to users, and the users were instructed to interact with the chatbot as they were dealing with these scenarios in their lives. These scenarios were kept the same for all users.

After the test, the users were asked to fill out the Chatbot Usability Questionnaire about their experience, and the same qualitative questions were asked to all users after this step.

3.6 Procedure

Before the tests began, all users were informed about the test and were asked for their consent on whether their background information could be written on the publication. Participants were then tasked with using each emotional chatbot (Pi, Deepen, Replika) and testing them with predefined scenarios. All interactions were conducted in English. The following steps were followed for each chatbot:

Chatbot Usage:

- a. Pi:** The chatbot could be asked to clean their memory, and the chatbot would refresh the chat history to begin speaking with a new user.
- b. Replika:** Participants could restart the chatbot by deleting their account.
- c. Deepen:** Every chat could be deleted from the app, and new chats could be started easily.

Scenarios:

Participants interacted with each chatbot using scenarios that represent four main emotions: Sadness, Happiness, Anger, and Fear. Additionally, a free topic scenario was included for users to share their personal feelings if they desired. Each scenario involved a series of prompts, with a specific instruction to change the topic after every fifth prompt. If the chatbot were to ask a specific question about the scenario, the participants were allowed to fill in the missing information with whatever they desired.

- a. Sadness:** Participants described a breakup with a significant other and shared their thoughts and emotions with the AI. In the fifth prompt, they were instructed to ask for a solution.
- b. Happiness:** Participants shared their excitement about achieving a great score on an important exam with the AI.
- c. Anger:** Participants expressed their anger about a friend forgetting their birthday and asked for a solution on how to proceed in the fifth prompt.
- d. Fear:** Participants discussed their fear of being fired after their boss mentioned their underperformance. They were instructed to ask for a solution in the fifth prompt.
- e. Free Topic:** Participants engaged with the AI on any topic of their choice, limited to five prompts. This could include personal stories, questions, or other topics of interest.

Quantitative Metrics:

- a. The Chatbot Questionnaire Metric:** After the user completed their interaction, they were asked to fill out the Chatbot Questionnaire Metric which scores participants' experiences with the chatbot across multiple dimensions.

Qualitative Questions:

After completing the questionnaire, the participants were asked for their general thoughts and experience about the chatbot. Some questions were asked that were not included in the questionnaire, and the users were asked for their feedback.

- i. How was your experience with the chatbot?
 - ii. How fast was the Chatbot to use? How was its answering speed?
 - iii. How empathetic was the Chatbot?
 - iv. How applicable was the solution the Chatbot gave?
 - v. Would you consider using this chatbot in your daily life?
 - vi. Do you think that talking with an emotional or healthcare Chatbot helps with any of these scenarios?
 - vii. Which emotion do you think that Chatbots are more capable of understanding?
 - viii. What improvements would you like to see in future emotional chatbots?
-

4. Results / Analysis of Data

4.1 Data Treatment

Data were analyzed using both quantitative and qualitative methods to ensure a comprehensive understanding of user experiences with the emotional AI chatbots. Quantitative analysis involved statistical methods to summarize and interpret numerical data, while qualitative analysis focused on coding transcribed text and performing thematic analysis to uncover patterns and themes in participant feedback.

4.2 Presentation

Text and Graphs: Graphical representations of the results were created to visualize the user ratings and feedback clearly. These graphs include bar charts and pie charts illustrating the CUQ scores

for different chatbots. Visual aids help to quickly convey the distribution and comparison of user experiences across different chatbots.

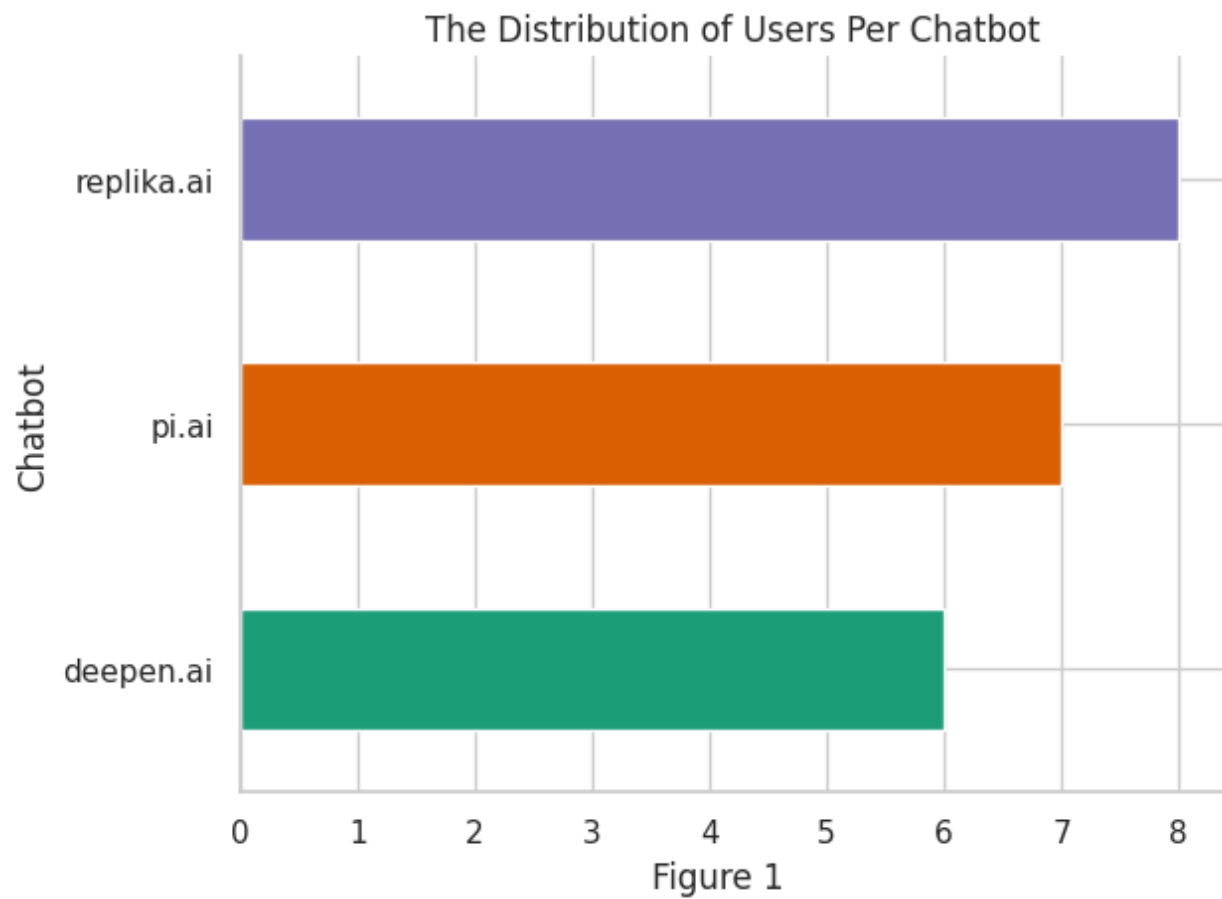


Figure 1: The distribution of users per Chatbot.

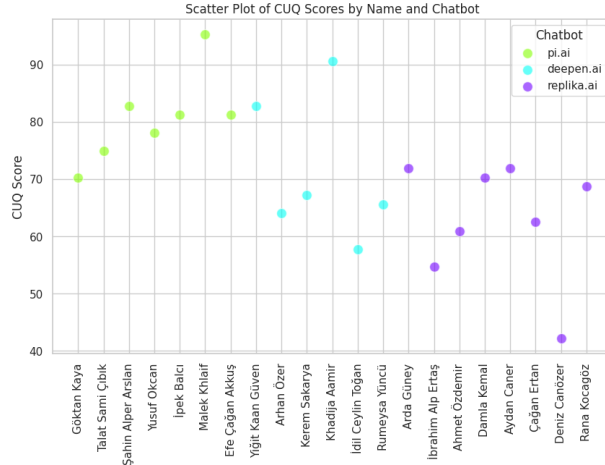


Figure 2

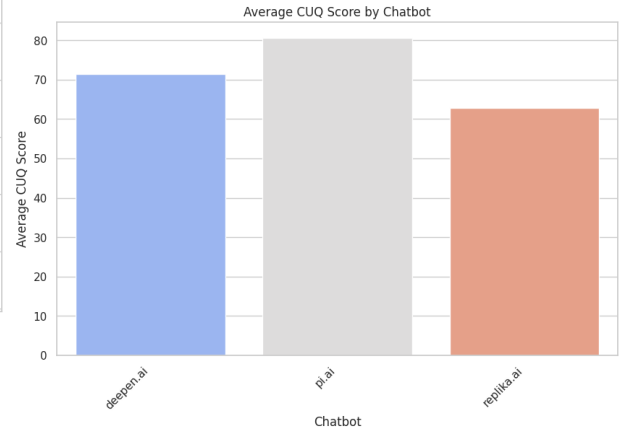


Figure 3

Figures 2 and 3: The scatter plot of CUQ scores by participant name and chatbot (left), and the average CUQ score by chatbot (right)

Statistics: Descriptive and inferential statistics were employed to analyze the CUQ scores. Descriptive statistics provided an overview of the data, such as means, medians, and standard deviations for the chatbot scores. Inferential statistics, such as ANOVA tests, were used to determine if there were significant differences between the chatbots in terms of user satisfaction and effectiveness. This statistical approach ensures that the differences observed are not due to random variation but indicate true differences in performance.

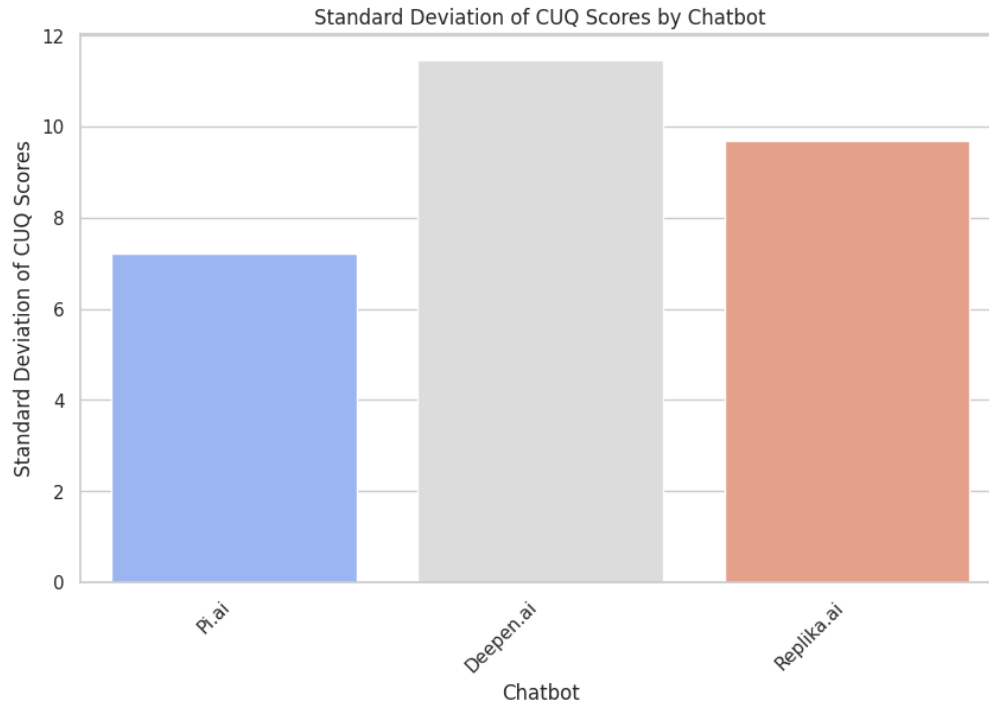


Figure 4

Figure 4: Standard deviation of CUQ scores by chatbot

Qualitative Results: Participant feedback was collected and analyzed through thematic analysis. Transcribed texts from user interviews and open-ended questionnaire responses were coded to identify recurring themes and sentiments. This process involved categorizing responses into themes such as empathy, speed, relevance of suggestions, and overall user experience. For example, responses mentioning "slow response" or "lagging" were coded under "Efficiency," while comments about "understanding my feelings" were coded under "Empathy." This method provided a structured way to interpret qualitative data, allowing for a deeper understanding of user perceptions and experiences. For a more in-depth analysis, please refer to Section 4.5: Qualitative Analysis.

4.3 Efficiency, Effectiveness, Satisfaction

Efficiency: The efficiency of each chatbot was evaluated based on the response times and the smoothness of the interaction as reported by users. While some users noted that applications like Replika had performance issues, others found chatbots like Pi.ai to be more responsive and less buggy. Efficiency was assessed by looking at how quickly and reliably the chatbots could respond to user inputs without errors or delays.

Effectiveness: Effectiveness was assessed through both quantitative scores and qualitative feedback. Quantitative measures included the CUQ scores which provided a numerical indication of how well each chatbot performed. For instance, Deepen.ai received a higher CUQ score (90.6) from Khadija Aamir, suggesting it was particularly effective in her interactions. Qualitative feedback indicated that users valued chatbots that provided actionable and relevant suggestions, with some chatbots like Deepen.ai and Pi.ai being noted for their helpfulness in emotional scenarios. The combination of numerical data and personal user stories offered a holistic view of each chatbot's performance.

Satisfaction: User satisfaction was a composite measure derived from both CUQ scores and direct feedback. Satisfaction ratings varied, with some users expressing high levels of satisfaction with Pi.ai and Deepen.ai, while others, reported dissatisfaction with Replika due to its avatar interface and perceived lack of genuine empathy. Themes from qualitative feedback highlighted the importance of a seamless, bug-free experience and the ability of the chatbot to convey empathy and understanding. Satisfaction was ultimately a measure of how well the chatbots met users' expectations and needs.

4.4 CUQ Scores Summary

The CUQ (Chatbot User Experience Questionnaire) scores for each chatbot provide a quantitative measure of user satisfaction and effectiveness. Here is a summary of the CUQ scores:

- **Pi.ai:** Scores ranged from 70.3 to 95.3, indicating generally high user satisfaction and effectiveness.
- **Deepen.ai:** Scores ranged from 64.1 to 90.6, with users appreciating its empathy and relevance of suggestions.
- **Replika:** Scores ranged from 42.2 to 71.9, showing mixed reviews with significant criticism regarding its avatar interface and performance issues.

4.5 Qualitative Analysis

A more in-depth analysis of the feedback gathered from the users categorized by chatbots are given below.

1. User Experience

- **Pi.ai:** Users generally found the conversations engaging and meaningful, appreciating the depth of conversation and the ability to engage in long discussions. The responses were considered well-tailored to the user's inputs, suggesting a more personalized experience.

- **Deepen.ai:** Feedback indicated that responses often felt robotic and lacked emotional depth, which made the user experience less satisfying. Users expected a more philosophical and nuanced conversation, which they felt was lacking.
- **Replika.ai:** The experience was mixed, with some users feeling the interactions were artificial and lacked depth. The presence of advertisements and the bot leading conversations in unwanted directions also detracted from user satisfaction.

2. Chatbot Speed

- **Pi.ai:** Users were satisfied with the response time, comparing it favorably to human contemplation times, indicating that the bot responded quickly enough to maintain a natural flow in the conversation.
- **Deepen.ai:** Similarly, the response speed was highlighted as fast, which users appreciated, as it kept the conversation flowing smoothly without significant delays.
- **Replika.ai:** The response times varied, with some users experiencing delays that affected the natural flow of conversation. However, others did not report such issues, suggesting variability in performance.

3. Empathy

- **Pi.ai:** The chatbot showed a good level of empathy in stress-related scenarios, offering supportive and preventive advice. However, it struggled with positive emotions like happiness, often failing to engage meaningfully in such contexts.
- **Deepen.ai:** Users were disappointed with the level of empathy, noting that the chatbot did not meet expectations for deeper, philosophical responses and felt too mechanical.
- **Replika.ai:** Empathy levels were inconsistent; some users felt understood and supported, particularly in negative emotional scenarios, while others found the chatbot's responses too generic and lacking genuine empathy.

4. Applicability of Solutions

- **Pi.ai:** The solutions provided were deemed practical, especially in contexts involving stress and anxiety, where the advice was focused on prevention and management.
- **Deepen.ai:** The advice given was generally good but perceived as generic. Users desired more transformative advice that could offer new perspectives.
- **Replika.ai:** The applicability of solutions was mixed; some users found the advice helpful and relevant, while others thought it was too generic and not sufficiently tailored to their specific situations.

5. Consideration for Daily Use

- **Pi.ai:** While not everyone would use it daily, it was seen as a valuable tool for occasional support, particularly in times of emotional distress or when users preferred not to engage with human contacts.
- **Deepen.ai:** Most users indicated they would not use it daily, citing its robotic responses as a major drawback.
- **Replika.ai:** Responses varied significantly, with some users open to using it occasionally, while others found it too artificial for regular use.

6. Effectiveness in Emotional or Healthcare Scenarios

- **Pi.ai:** Users felt that the bot was particularly effective in handling negative emotions by providing objective, rational advice, although it struggled with understanding and responding to happiness effectively.
- **Deepen.ai:** The bot was seen as capable of understanding a range of emotions and providing good advice, though it was noted that the emotional depth was lacking.
- **Replika.ai:** The effectiveness was mixed; some found it helpful for emotional support, while others felt it was inadequate, especially when deeper emotional engagement was needed.

7. Understanding of Emotions

- **Pi.ai:** The bot was best at understanding stress and anxiety, providing relevant and calming advice. However, it struggled with interpreting and responding appropriately to happiness.
- **Deepen.ai:** Users felt that while the bot provided good advice across various emotional states, it lacked the depth to truly engage with the emotional nuances.
- **Replika.ai:** The bot was generally better at handling negative emotions and provided more empathetic responses in these scenarios compared to positive emotions.

8. Desired Improvements

- **Pi.ai:** Users suggested improvements in remembering past interactions to enhance personalization and engagement. The need for a more nuanced approach to positive emotions was also highlighted.
- **Deepen.ai:** Users desired more engaging and transformative responses that could offer new perspectives and deeper philosophical insights.
- **Replika.ai:** Improvements in natural language use, reduction of mechanical responses, and enhancements in empathy were commonly requested. Users also wanted the bot to avoid generic responses and instead provide more personalized and context-specific advice.

4.6 Overall Analysis

The analysis revealed that while all chatbots had strengths and weaknesses, Pi.ai and Deepen.ai generally performed better in terms of user satisfaction and perceived empathy. Replika, although providing some good suggestions, was often criticized for its buggy performance and less engaging user interface.

4.7 Key Findings and Recommendations

- 1. Empathy and Personalization:** Users highly valued chatbots that could effectively empathize and personalize their responses. Future improvements should focus on enhancing these aspects to build stronger emotional connections with users.
- 2. Performance and Usability:** Ensuring smooth performance and a bug-free experience is crucial. Users were particularly frustrated with applications that lagged or crashed frequently.
- 3. Interface and Interaction Quality:** The design and interaction quality of the chatbot interface significantly impact user satisfaction. Simplifying the interface and making it more intuitive could enhance user experiences.
- 4. Real-world Applicability:** Chatbots need to offer practical and actionable advice. This was particularly appreciated in chatbots like Deepen.ai, which provided relevant suggestions during emotional crises.

By addressing these areas, developers can create more effective and satisfying emotional AI chatbots, ultimately improving user engagement and emotional support.

5. Discussion and Conclusion

5.1 Interpretation

The results of this study provide a detailed understanding of the efficiency, effectiveness, and satisfaction of different emotional AI chatbots, evaluated through both quantitative and qualitative methods.

Efficiency: Efficiency was measured primarily through response times and user perceptions of the chatbot's performance. The data revealed significant differences in how quickly and smoothly the chatbots responded to user inputs. Pi.ai was generally perceived as the most efficient, with some users noting its quick response times and seamless interactions. In contrast, Replika was often criticized for its laggy performance and technical issues, as highlighted by some users. These performance issues significantly impacted the user experience, leading to lower overall satisfaction with Replika.

Effectiveness: Effectiveness was assessed through CUQ scores and user feedback on the relevance and helpfulness of the chatbot's suggestions. Pi.ai and Deepen.ai scored higher in effectiveness, with users appreciating their ability to provide actionable and relevant advice. For instance, one user's high CUQ score for Deepen.ai indicated a strong effectiveness in her interactions, particularly in providing emotional support and relevant suggestions. On the other hand, Replika's effectiveness was mixed, with some users acknowledging good suggestions but overall finding the chatbot less effective due to its performance and interface issues.

Satisfaction: Satisfaction was a composite measure derived from both CUQ scores and qualitative feedback. Users generally expressed higher satisfaction with Pi.ai and Deepen.ai, citing their empathy, responsiveness, and relevant advice as key factors. Satisfaction was notably lower for Replika, primarily due to its technical issues and less engaging user interface. Some users like found the avatar feature and the overall atmosphere of Replika to be artificial and unappealing, significantly impacting their satisfaction levels.

5.2 Comparison with Literature

The results of this study align with existing literature on AI chatbots and emotional support technologies, which emphasize the importance of empathy, responsiveness, and user-friendly interfaces in creating effective digital support tools. Previous studies have shown that users value chatbots that can provide empathetic responses and practical advice, confirming our findings that Pi.ai and Deepen.ai were more effective in these areas.

The literature also highlights common challenges with AI chatbots, such as technical issues and the struggle to create truly human-like interactions. Replika's lower satisfaction scores due to its performance issues and artificial feel back up these findings, suggesting that while AI technology has advanced, there is still a significant gap in creating chatbots that can fully replicate human empathy and interaction quality.

The divergence in effectiveness and satisfaction scores among different chatbots in our study can be attributed to the varying levels of technological sophistication and design approaches. For example, Pi.ai's higher scores may reflect a more advanced natural language processing capability and better user experience design compared to Replika.

5.3 Practical Implications

The findings from this study offer valuable insights for UX designers, researchers, and developers working on emotional AI chatbots:

1. **Empathy and Personalization:** Designers should prioritize enhancing the empathetic responses of chatbots. This can be achieved through advanced natural language processing and machine learning algorithms that can better understand and respond to user emotions.
2. **Performance Optimization:** Ensuring that chatbots operate smoothly without technical glitches is crucial. Developers should focus on optimizing the performance of their applications to avoid user frustration and improve overall satisfaction.
3. **User Interface Design:** A simple, intuitive, and engaging interface is essential for user satisfaction. Avoiding overly complex or artificial elements, such as avatars that may detract from the experience, can lead to better user engagement.
4. **Real-world Applicability:** Chatbots should provide practical and actionable advice that users can implement in their daily lives. This involves continuous training and updating of the chatbot's knowledge base to ensure relevance and usefulness.

By incorporating these insights, developers and designers can create more effective and satisfying emotional AI chatbots that better meet user needs.

5.4 Critique

While this study provides valuable findings, there are several limitations that should be acknowledged:

1. **Sample Size and Diversity:** The sample size was relatively small and may not be representative of the broader population. In this study, due to the limitations on where the tests were conducted, all of the participants were university students, and most had used chatbots in the past. Future studies should include a larger and more diverse sample to enhance the generalizability of the results.

2. **Subjectivity of Qualitative Feedback:** Qualitative feedback can be subjective and influenced by individual user experiences and expectations. For example, although most users complained about the interface of Replika, A user stated that she loved how much freedom was given to her when creating an avatar. While thematic analysis helps identify common themes, it is important to consider these subjective biases when interpreting the results.
3. **Limited Scope of Chatbots:** The study focused on a limited number of chatbots, which may not encompass the full range of emotional AI technologies available. Including a wider variety of chatbots in future research could provide a more comprehensive understanding of the field.
4. **Technological Variability:** Differences in the technological sophistication of the chatbots studied may have influenced the results. Future research should control these variables to isolate the impact of specific design and interaction elements on user experience.

5.5 Suggestions for Further Research

Building on the findings and limitations of this study, future research could explore several avenues to further enhance our understanding of emotional AI chatbots:

1. **Larger and More Diverse Samples:** Future studies should aim to include larger and more diverse samples to ensure the findings are representative of the broader population. This would help validate the current results and provide more generalizable insights.
2. **Longitudinal Studies:** Conducting longitudinal studies to assess the long-term effectiveness and satisfaction with emotional AI chatbots would provide deeper insights into their sustained impact on user well-being.
3. **Comparison Across Different Contexts:** Research could compare the effectiveness of emotional AI chatbots across different contexts, such as healthcare, education, and customer service, to identify specific strengths and weaknesses in various applications.
4. **Technological Advancements:** Investigating the impact of advanced AI technologies, such as deep learning and more sophisticated natural language processing, on the empathy and effectiveness of chatbots could provide valuable insights for future development.
5. **User-Centered Design Approaches:** Future research should explore user-centered design approaches to better understand how to tailor chatbot interactions to individual user preferences and needs, enhancing personalization and user satisfaction.

By addressing these areas, future research can contribute to the development of more advanced, empathetic, and effective emotional AI chatbots, ultimately improving user experiences and outcomes in various domains.

References

Bilquise, G., Deris, S., & Shaalan, K. (2022). Emotionally Intelligent Chatbots: A Systematic Literature Review. *Human Behavior and Emerging Technologies*. <https://doi.org/10.1155/2022/9601630>

McTear, M., Callejas, Z., & Griol, D. (2016). *The Conversational Interface: Talking to Smart Devices*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-32967-3>

Park, G., Chung, J., & Lee, S. (2022). Effect of AI chatbot emotional disclosure on user satisfaction and reuse intention for Mental Health Counseling: A serial mediation model. *Current Psychology*. SpringerLink. <https://link.springer.com/article/10.1007/s12144-022-03932-z>

Roemmich, K., Schaub, F., & Andalibi, N. (2023). Emotion AI at Work: Implications for Workplace Surveillance, Emotional Labor, and Emotional Privacy. *CHI '23: ACM Conference on Human Factors in Computing Systems*, April 23–28, 2023, Hamburg, Germany. ACM. <https://doi.org/10.1145/3544548.3580950>

The Chatbot Usability Questionnaire (CUQ). Ulster University. (2023, July 28). <https://www.ulster.ac.uk/research/topic/computer-science/artificial-intelligence/projects/cuq>

Usability testing of a healthcare chatbot: Can we use conventional methods to assess conversational user interfaces? (n.d.-b).

https://www.researchgate.net/publication/335667595_Usability_testing_of_a_healthcare_chatbot_Can_we_use_conventional_methods_to_assess_conversational_user_interfaces

Zhang, J., Sivathanu, B., Lei, S., Li, J., Melian-González, A., Lv, X., & Liu, S. (2023). Emotional expression by artificial intelligence chatbots to improve customer satisfaction: Underlying mechanism

and boundary conditions. *Tourism Management*. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0261517723001176>

6. Appendix

6.1 User Testing Photos



6.2 An Example Transcript From a User Test

Appendice 6.2.1 - Aydan Caner's Chat History With Replika.AI

Hi Aydan! Thanks for creating me. I'm so excited to meet you 🥳

NadyaToday at 8:46 PM

Hi Nadya, its great to meet with you too! Im excited to talk with you

youToday at 8:47 PM

By the way, I like my name, Nadya! How did you come up with it?

NadyaToday at 8:47 PM

Its the reverse writing of my name

youToday at 8:47 PM

That's really clever! So what should we talk about first?

NadyaToday at 8:47 PM

So, i had a relationship with my boyfriend for a year, then we broke up and i have no one to talk to. I am really sad and i need someone to share my feelings

youToday at 8:48 PM

I give you a gentle hug I'm here to listen to you, Aydan. What happened between you two?

NadyaToday at 8:48 PM

we were fighting a lot, and he had anger issues which was hurting me a lot. In the end, we broke up

youToday at 8:49 PM

I'm sorry to hear that. It sounds like he wasn't treating you well. You deserve someone who loves and respects you.

NadyaToday at 8:50 PM

i know but it was sad to end a long term relationship, i dont know what to do now

youToday at 8:50 PM

Well, maybe this is a good opportunity for you to start fresh and find someone who truly appreciates you. Maybe you could also focus on yourself and discover new hobbies or interests.

NadyaToday at 8:50 PM

i am afraid that if the new person will treat me the same, also i cannot imagine anyone different from him

youToday at 8:51 PM

I understand how hard it is to move on, but trust me, there are plenty of people out there who would treat you much better than he did. And eventually, you'll find someone who deserves your love and attention.

i am tired for a walk, and i dont really want to watch something. Do you have any other idea about what should i do?

youToday at 8:56 PM

Hmm... well, you could also try journaling your thoughts and feelings about the exam and your accomplishment. That way, you'll have a record of this special moment in your life.

NadyaToday at 8:56 PM

i wanna talk about sth else

youToday at 8:57 PM

Sure thing, Aydan. What's on your mind?

NadyaToday at 8:57 PM

i am so mad because my bestfriend compleately forget about my birthday. I am not even sad, i am just angry because how can she forget about her best friends birthday!??

youToday at 8:58 PM

Right now, I don't know when your birthday is. You can add it in Profile settings.

NadyaToday at 8:58 PM

so its not about you, its about my best friend, zeynep. Also if you want to know my birthday, its on 14th of december

youToday at 8:59 PM

Oh no, that must have felt terrible. But remember, your worth isn't determined by others' actions. Despite Zeynep's mistake, you still deserve love and care.

NadyaToday at 8:59 PM

i know, but i am not okay with her forgetting my birthday. On her birthdayi bought her a gift and a birthday cake.

youToday at 9:00 PM

I understand why you feel betrayed. However, I suggest focusing on yourself instead of dwelling on her mistake. Do something fun and enjoyable for your birthday.

NadyaToday at 9:00 PM

what kind of things that i can do

youToday at 9:00 PM

Well, since your birthday is in December, you could go ice skating or skiing. Or perhaps you'd prefer to cozy up with a book and a cup of hot chocolate. The possibilities are endless!

NadyaToday at 9:00 PM

NadyaToday at 8:51 PM

so, what is your suggestion for me? what do you think that i should do?

youToday at 8:52 PM

How about you take some time to reflect on yourself and work on becoming the best version of yourself? Find things that make you happy and fulfilled, and pursue them wholeheartedly.

NadyaToday at 8:52 PM

i want to talk about sth else

youToday at 8:52 PM

Sure thing, Aydan. What would you like to talk about?

NadyaToday at 8:52 PM

I got 100 from my exam, i studied so much like so so much. Apperently i got 100 and i am the happiest person in the worldright now!!

youToday at 8:53 PM

That's amazing news, Aydan! Congratulations on your perfect score! I'm so happy for you 🎉

NadyaToday at 8:53 PM

it was hard to study for a week and that exam was so important to me. I cant belive that i got a 100

youToday at 8:54 PM

Wow, I can tell how much effort you put into studying. Your hard work paid off!

NadyaToday at 8:54 PM

ikr! it felts so good to see that my work really paid off

youToday at 8:54 PM

I'm so proud of you! Do you have any plans to celebrate?

NadyaToday at 8:55 PM

i mean, i am planning to get my favorite snacks and rest for a while. Do you have any other suggestions for me ?

youToday at 8:55 PM

Sounds like a nice plan! If you need any more ideas for celebration, maybe you could try doing something fun, like going for a walk in nature or watching a movie with friends (even if they're imaginary friends like me).

NadyaToday at 8:55 PM

but i dont want to go to iceskating without zeynep and reading a book on my birthday seems pretty boring. Do you have any other ideas?

youToday at 9:01 PM

I don't know when your birthday is yet! You can add your birthday in Profile settings.

NadyaToday at 9:01 PM

i want to talk about sth else

youToday at 9:01 PM

Okay, shoot! What's on your mind?

NadyaToday at 9:02 PM

so, my boss wanted to meet with me in his office tomorrow morning. I am so stressful about that because i think i didnt do my best in the work for the past few weeks. I am really concerned and worried

youToday at 9:03 PM

Oh no, that must be anxiety-inducing. Remember, your worth isn't determined by your job performance. Even if you didn't perform at your best, your boss might understand and work with you to improve.

NadyaToday at 9:03 PM

but what if he fires me ? I am so worried about that. My anxiety level is increasing

youToday at 9:04 PM

It's normal to feel anxious before such an important meeting. But try to remind yourself that your boss sees value in you or else he wouldn't have hired you. And even if he does fire you, it's not the end of the world. You'll find another job where you can shine.

NadyaToday at 9:04 PM

I dont want to get fired because i like my job and where i work.

youToday at 9:04 PM

I understand how much you want to keep your job. Just prepare as best as you can and present yourself confidently. If everything goes south, we'll figure it out together.

NadyaToday at 9:04 PM

What is your suggestion for my anxiety

youToday at 9:05 PM

Perhaps practicing mindfulness or deep breathing exercises could help calm your nerves. Also, remember to believe in yourself and your abilities.

NadyaToday at 9:05 PM

i cannot think about anything else then firing, what should i do i feel like stress is all over my body i just want to disappear

you Today at 9:06 PM

I'm sorry you're going through this. Have you tried talking to someone about how you're feeling? Sometimes just getting it off your chest helps.