Received October 29, 2024, accepted May 15, 2025, date of publication June 17, 2025.

Case Study

Case Study: Augmented Reality Enabled Mental Health Chatbot

Subbaraj Pravin Kumar*, Akash Kumar, and Anusha Amba Prasanna

Biomedical Engineering, Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu, India.

* Corresponding Author Email: pravinkumars@ssn.edu.in

ABSTRACT

Background and Objective: In recent years, there has been a growing demand for mental health support. This has led to a focus on providing personalized and continuous care. However, traditional mental health systems often have long wait times and limited support for engagements beyond clinical hours. The goal of this project is to create ARden, a digital companion using augmented reality, to help improve mental health for those in need.

Material and Method: This study aims to fine-tune a Large Language Model with domain-specific knowledge, ensuring a personalized and intelligent companion—ARden. The chatbot is integrated with the AR companion using an Application Program Interface (API). The mixed reality companion is accessible via a mobile application, making care available without the additional hardware costs associated with head-mounted displays.

Results: The development of ARden has introduced new possibilities for personalized and interactive mental health support. Early feedback suggests that the chatbot may help improve user engagement and satisfaction, supported by encouraging retention metrics. By combining augmented reality, large language models, and a character-based interface, ARden offers an approach that could contribute positively to mental health support.

Conclusion: ARden aims to help users with emotional regulation during long wait times between mental health interventions, overcome communication barriers, and provide exercises and suggestions to improve mental health wellbeing. This approach offers a promising solution to existing mental health challenges and holds potential for further improvement and scalability.

Keywords—*Augmented reality, Large language model, Mood score.*

INTRODUCTION

Mental health problems pose a critical public health burden, yet traditional solutions remain insufficient to address the growing demand. The advent of social media and increased isolation among the younger population have exacerbated mental health challenges for these individuals. According to NextStep Solutions, a leading provider of behavioral health software, 29% of the U.S. population experiences at least one form of mental illness.[1]

Along with the factors that have led to an increase in mental health issues, there is also a severe shortage of psychiatrists, with only nine per 100,000 people. This deficit contributes to negative mental health outcomes, including an increased risk of suicide.[1]

At present, individuals seek help from their family, social circles, the internet, and mental health professionals to overcome these issues. However, several barriers limit the effectiveness of current mental health systems, including social stigma and a shortage of mental health professionals.

Currently, most of the mental health help that individuals receive is in the form of cognitive behavioral therapy and medication. Cognitive behavioral therapy (CBT) is based on the idea that professionals are trained to help individuals overcome the issues they are facing and enable them to tackle problems on their own by using structured systems. However, due to understaffing, the current system struggles with increased costs, long wait times, and other challenges. Despite its shortcomings, seeking professional help remains the best option for those facing mental health challenges.

Recent studies have explored the potential of chatbots to improve mental health outcomes. Denecke et al.[1] examined two prominent AI-driven mental health chatbots, WYSA and SERMO. Wysa Inc., headquartered in the USA, is an everyday mental health application, and Sermo, also headquartered in the USA, is a social platform for physicians to collaborate and stay informed. WYSA detects negative moods and integrates features like depression assessments and meditation exercises, while SERMO addresses psychological impairments using Cognitive Behavioral Therapy (CBT) techniques. Quantitative analysis revealed that frequent users of WYSA experienced greater mood improvements than occasional users. Experts acknowledged that SERMO is well-suited for patients struggling with face-to-face communication, although challenges remain, including issues with data retention, dataset generation, and the inability of AI systems to handle emergencies effectively.

Potts et al.[2] addressed the lack of accessible mental health services in rural areas by introducing a multilingual chatbot, ChatPal. It was developed by the academic consortium led by Ulster University (UK), which collaborated with partners in Ireland, Scotland, Sweden and Finland. Funded by the Northern Periphery and Arctic (NPA) Programme, ChatPal aimed to provide support in English, Scottish Gaelic, Swedish, and Finnish. The study employed a single-arm pre-post intervention design, enrolling participants from rural areas to use ChatPal over 12 weeks, with well-being measured via scales like SWEMWBS. ChatPal, developed with Rasa (backend) and PhoneGap (frontend), features mood logging and mindfulness exercises. While ChatPal is seen as a complementary tool for mental health services, the study calls for further research to confirm its effectiveness. The chatbot's multilingual capabilities improve accessibility, but technical issues remain, particularly with integration and functionality.

Social desirability and social support are factors that contribute to positive mental health outcomes. Similarly, a need to belong, which is heightened among young adults, is associated with negative outcomes. Thus, looking at the current landscape of mental health issues, it is clear that providing emotional support through interventions can help people, but is no replacement for a true emotional connection with others. It can serve to help people deal with issues in a healthier way, leading to better outcomes.

Another study aimed to evaluate the effectiveness of ChatGPT in providing mental health support, with a particular focus on anxiety and depression. The primary objective was to assess the quality of responses generated by ChatGPT to user queries related to these mental health conditions. The study specifically analyzed the model's responses to three queries: two relating to anxiety management and medication, and a third regarding alternative treatment options. Moreover, the study examined the consistency and reliability of ChatGPT's responses across successive interactions.

It was found that ChatGPT did not provide information about medication.[3] Some advantages include its ability to offer personalized advice based on a person's history, improved accessibility in remote locations, and lower costs compared to traditional therapy. However, the information must be cross-checked with professionals due to some inconsistencies, and it cannot substitute for mental health care. The model also cannot provide prescriptions. ChatGPT's responses to prompts can be inappropriate, possibly due to the type of questions being asked. Thus, while ChatGPT is useful, the model developed for the user must address its shortcomings in some way.

Yang et al.[4] investigated the capabilities of large language models (LLMs) within the healthcare domain, focusing on both their potential applications and inherent limitations. The primary goal was to assess the effectiveness of general LLMs in healthcare settings and to identify areas where domain-specific models could offer improved performance. General-purpose LLMs often lack the specialized knowledge required for healthcare applications due to the disparity between the general text used in their training and the professional, domain-specific content needed for clinical use. The study highlighted the promise of domain-specific LLMs. For instance, BioBERT was developed by Korea University and trained on PubMed data. Another example, SCIBERT, was created by the Allen Institute for AI and was trained on broad scientific texts from Semantic Scholar. Similarly, PubMedBERT, by Microsoft Research, was specifically trained on PubMed abstracts. These models are all based on the BERT architecture and require significant computational resources for operation. Despite their promise, the study acknowledged the challenges that remain in their clinical implementation.

The performance of domain-specific LLMs was found to be superior compared to general models, particularly in patient interactions. A tailored model called ChatDoctor, which is a fine-tuned large language model based on Llama and trained on 100,000 real-world patient-doctor dialogues from an online consultation platform, and supported by an NIH grant. It demonstrated enhanced efficacy in clinical settings. However, the study also identified significant challenges in deploying LLMs in healthcare, particularly concerning data integrity, interpretability, and the high costs associated with developing these models. The integration of LLMs into clinical practice as supplementary tools was explored, emphasizing the need for improvements in task optimization and conversational assistance.

The challenges related to interpretability, data limitations, and ethical considerations must be addressed to fully realize the potential of LLMs in clinical practice. The research suggests that future efforts should focus on optimizing these models for specific tasks, improving data diversity, and ensuring the accuracy and reliability of the content generated by LLMs, particularly when cross-referenced with professional expertise.

User retention is also reported to be a critical factor, emphasizing the importance of highly engaging interactions with chatbots. It was demonstrated that inadequate retention rates often stem from a lack of personalization, which impedes the effectiveness of mental health apps. To address these challenges, the paper proposed several solutions, including the personalization of chatbots by utilizing the user's previous conversations. Additionally, incorporating peer communication methods was found to enhance both engagement and effectiveness. The developed app should adapt based on user feedback, with the overall goal of creating a more user-centric, adaptable, and effective platform.

In another study, data from the chatbot interactions, including session details and mood logs, were analyzed to extract features such as tenure, mood logging frequency, and conversation interactions. K-means clustering was employed to categorize users into three groups: abandoning, frequent transient, and sporadic users. This analysis compared user behaviors, engagement, and retention metrics with those of other mental health apps. The study emphasizes the importance of high engagement and retention metrics and the need for personalized user experiences.[5]

The effectiveness of the chatbot heavily depends on the underlying language model that powers it. Llama 2, an advanced open-source language model from Meta AI, can generate text similar to that of a human and is useful for chatbots, translation, content production, and other applications. Llama 2 offers notable advantages in versatility and adaptability when fine-tuned for specific domains, addressing limitations observed in previous models.[6]

Roumeliotis et al.[6] investigated the challenges and opportunities faced by developers when deploying and fine-tuning Llama 2, with the hypothesis that the open-source nature of Llama 2 facilitates faster development compared to closed-source models. Early adopters' experiences in deploying and fine-tuning Llama 2 were observed over a 10-day period, with particular attention given to the medical domain, a primary area of interest for fine-tuning efforts.

Data on model deployment, fine-tuning, and other relevant factors were gathered during this period. Textual data was then processed through keyword identification, K-means clustering, and word cloud visualization. The resulting analysis reveals that Llama 2 can be seamlessly deployed and fine-tuned to the domain-specific requirements of various industries, thereby addressing challenges encountered with earlier models.

Yang et al.[7] compared four Large Language Models (LLMs) for mental health analysis, focusing on the effectiveness of prompting strategies such as Chain of Thought prompting, emotion-enhanced prompts, and few-shot learning. The findings emphasized the importance of domain-specific fine-tuning for improved results in building mental health solutions.

Regular large language models often fall short in specialized areas like medicine, where domain-specific knowledge is crucial.[8] The authors proposed PMC-Llama, an open-source language model specifically tailored for medical applications.8 They systematically analyzed the process of adapting a general-purpose LLM to the medical domain by integrating 4.8 million biomedical academic papers and 30,000 medical instructional materials, and extensively fine-tuned it for compliance with the domain-specific knowledge base.

Mental health professionals use screening questionnaires (SQs) to identify symptom areas that require further exploration. Regular screening can enable the early identification of individuals in high-stress professions who may require mental health support. Data indicate that a significant percentage of public safety personnel screen positive for at least one mental health disorder, highlighting the advantages of frequent screening.[9]

Integrating LLMs into chatbots could enhance their ability to provide tailored support, especially when fine-tuned for specific screening tasks within high-stress populations.

Incorporating Virtual Reality (VR) and Augmented Reality (AR) into mental health interventions offers a transformative approach to enhancing user engagement and interaction. They can improve access to and availability of therapy due to their personalized nature. Proper training for mental health professionals, rigorous scientific research, and strict adherence to data privacy and ethical guidelines are essential for the responsible use of mental health apps, making them more engaging, targeted, and therapeutically effective.

Integrating Augmented Reality (AR) into chatbot platforms represents a promising advancement in mental health care. Current mental health chatbots, while offering useful features, have limitations. One major bottleneck is the need to ensure data privacy, especially since chatbots that provide personalized suggestions must store previous user interactions.[2] Ensuring that user data is protected from unauthorized use is crucial for trust and widespread adoption.

Another challenge is the need for relevant content and fine-tuning large language models to provide helpful and contextually appropriate responses. It must converse with the user in a manner that is genuinely helpful to them. Fine-tuning the model requires resources, and the model needs continuous updates to stay relevant.[3] A drawback is that the way the model arrives at decisions is not explained clearly.

User retention is another significant challenge for chatbot applications. Generalizations about user behavior have led to the development of different user archetypes.4 This information about how users interact with the chatbot can be used to improve the retention metrics of the application we build.

For chatbots to be effective for emotional support, they must understand how users feel and react empathetically. This is possible by using libraries with words and the emotions associated with those words. However, chatbots are only as good as their prompts, and the sentiment and emotion lexicons used for emotion-enhanced prompts suffer from annotation bias and limited vocabulary, which may not reflect the evolving language used in recent datasets.[4]

Recent advancements in training large language models across multiple languages have increased accessibility and improved the generalizability of studies.[5] Although this is important, the focus of this paper is on developing a robust mental health chatbot primarily for English-speaking users, as they represent the majority of current users.

Given the limitations of existing mental health chatbots and the recent advancements in technology, this paper proposes the development of a chatbot using the Llama 2 model integrated with Augmented Reality technology. The proposed chatbot will prioritize high user retention, accuracy, privacy, and address current chatbot limitations while incorporating the additional functionalities discussed.

This study proposes an Augmented Reality-enabled Mental Health Chatbot that can provide supplementary support between visits to mental health professionals. Although chatbots cannot replace traditional therapy, they can offer continuous mental health support, helping individuals declutter their thoughts and providing accessible care at any time.

The aim was to analyze user engagement with a mental health chatbot, focusing on its potential to improve user retention through interactive and personalized experiences. User retention was examined to identify challenges and optimize engagement by understanding different user archetypes. The impact of the chatbot on users' mental health needs to be monitored over time when used alongside professional medical guidance.

The chatbot was developed as an app, making it accessible to a broader audience with user-friendliness. This study used LLM models that enable personalization, which is critical for effective content delivery. Various personalization techniques, such as retaining the memory of previous conversations, were implemented to create a more tailored and continuous user experience.

Retaining the memory of previous conversations helped continue interactions from the last episode, rather than starting from the beginning each time.

METHODOLOGY

Workflow

The large language model (LLM) is customized with extensive medical literature and fine-tuned for optimal effectiveness. The application processes auditory input to provide information to the LLM. Additionally, the application features a character that users interact with, designed to appear friendly and empathetic. This companion, integrated with the LLM, facilitates user interaction and contributes to mental health improvement through the application's functionalities.

Data

A mental health chatbot requires vast amounts of data to provide accurate and reliable solutions to individuals' mental health challenges. Developing a comprehensive database that integrates both local knowledge and online resources is essential.[7] This database also contains research articles tailored to the individual's specific needs, such as those addressing emotional support, depression, anxiety disorders, and eating disorders. General information regarding interventions can also be obtained from reputable online sources, complementing the personalized data.

Profiling

Mental health professionals initially assess and profile individuals based on their diagnosis, which is then fed as input to the model (Figure 1). This profiling enhances the personalization of the chatbot. Users are prompted to select a broad category that aligns with their experiences, and the application provides relevant information based on both the user's selection and the therapist's assessment

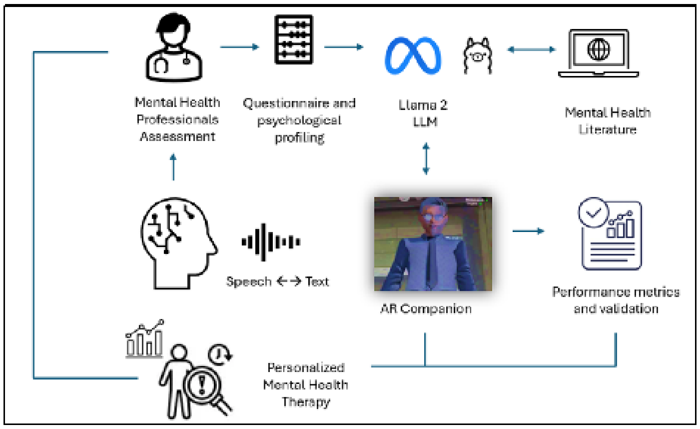
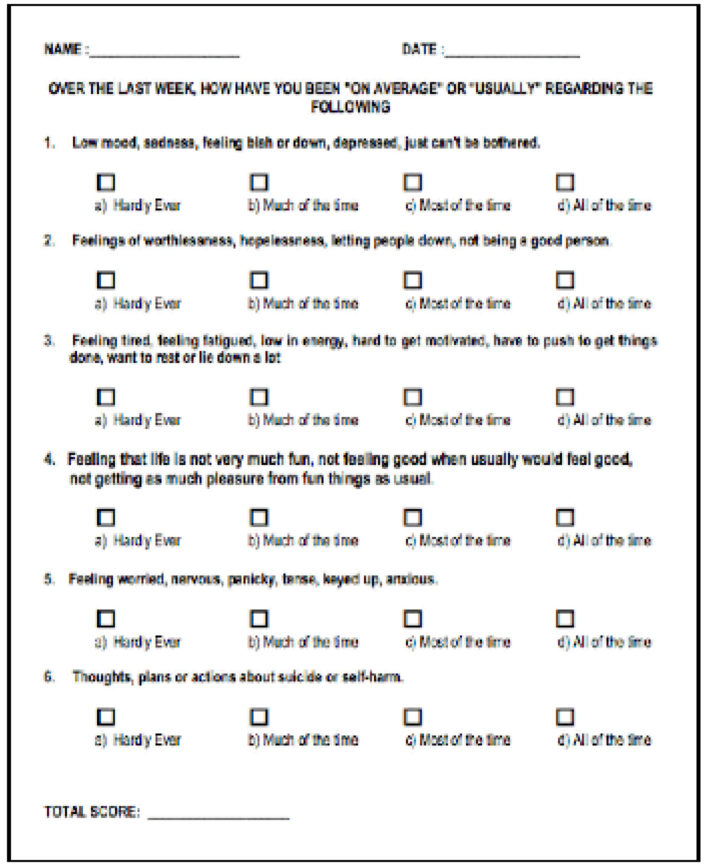


FIGURE 1. Workflow of AR Companion (ARden) assisted personalized therapy.

This work aims to understand individual perspectives and provide relief from mental health conditions, such as stress, anxiety, and depression, through personalized solutions. To effectively address the needs of users experiencing mental health issues, collaboration with professionals is essential. Through multiple sessions with a counseling psychologist, they assess a person's characteristics, drawing upon their experience and intuition developed through years of study and practice. The chatbot can then create a tailored solution or personalized interaction method for each individual, taking professional input into account. For instance, if a person tends to respond only to non-confrontational communication, the chatbot will recognize this and deliver information in a compassionate, non-threatening manner to ensure the individual is receptive.

A questionnaire is developed, as shown in Figure 2, based on input from the psychologist, with its design undergoing thorough consultation and multiple levels of review by the therapists. This personalized profiling ensures that users receive appropriate, targeted support, rather than generic interventions.



NAME :________________ DATE :________________

OVER THE LAST WEEK, HOW HAVE YOU BEEN "ON AVERAGE" OR "USUALLY" REGARDING THE FOLLOWING

1. Low mood, sadness, feeling blah or down, depressed, just can't be bothered.

☐ a) Hardly Ever ☐ b) Much of the time ☐ c) Most of the time ☐ d) All of the time

2. Feelings of worthlessness, hopelessness, letting people down, not being a good person.

☐ a) Hardly Ever ☐ b) Much of the time ☐ c) Most of the time ☐ d) All of the time

3. Feeling tired, feeling fatigued, low in energy, hard to get motivated, have to push to get things done, want to rest or lie down a lot

☐ a) Hardly Ever ☐ b) Much of the time ☐ c) Most of the time ☐ d) All of the time

4. Feeling that life is not very much fun, not feeling good when usually would feel good, not getting as much pleasure from fun things as usual.

☐ a) Hardly Ever ☐ b) Much of the time ☐ c) Most of the time ☐ d) All of the time

5. Feeling worried, nervous, panicky, tense, keyed up, anxious.

☐ a) Hardly Ever ☐ b) Much of the time ☐ c) Most of the time ☐ d) All of the time

6. Thoughts, plans or actions about suicide or self-harm.

☐ a) Hardly Ever ☐ b) Much of the time ☐ c) Most of the time ☐ d) All of the time

TOTAL SCORE: ________________

FIGURE 2. Questionnaire.

Mental Health Chatbot

Large Language Model

The process begins by uploading appropriate resources as PDF files, followed by the application of a text extraction algorithm to retrieve the content while removing any non-text elements (Figure 3).[8] The extracted text is then segmented into manageable pieces to facilitate meaningful embeddings. Llama 2, implemented via the Langchain framework, is employed to generate embeddings that capture the semantic essence of each text fragment. These embeddings are stored and managed in Pinecone, a cloud-based vector store. This fully managed service handles hardware infrastructure required for efficiently storing and searching vector data. Once the embeddings are uploaded to Pinecone's cloud storage, natural language queries can be processed. Pinecone performs a similarity search by converting queries into text representations and generating embeddings via Langchain Llama 2. This search identifies the most similar embeddings, retrieving the corresponding text fragments. These fragments are then combined to form a cohesive natural language response, enabling smooth and effective user interaction in question-and-answer scenarios.

FIGURE 3. Knowledge base construction.

Model Selection

Llama 2 was selected for this preliminary study due to several compelling factors. Its open-source framework supports accessibility and collaborative development, offering a parameter range from 7 billion to 70 billion. Additionally, Llama 2 distinguishes itself through its speed, which outperforms earlier models. This is particularly advantageous for time-sensitive tasks and applications requiring rapid processing, like mental health chatbots. Moreover, Llama 2 offers comprehensive documentation and a supportive open-source community, facilitating its integration. As large language models evolve rapidly, future iterations of this work will consider adopting more advanced models.[9]

Integration with Chatbot

Firstly, a Unity project was set up with the necessary dependencies installed to enable the companion to be deployed into a mobile application. Inworld AI is an engine that was used to create a character prefab to import into Unity. To begin with, the Inworld AI software had to be downloaded and interfaced with the Unity platform. Then, after the avatar was generated using Ready Player Me, a cross-game avatar platform that enables avatar creation and seamless integration into other platforms, its characteristics were customized using the Inworld AI portal. The API keys were then configured within the Unity project to import the character into the environment. The built-in speech of the character was replaced with the mental health chatbot. Visual customizations from Ready Player Me were refined, and Inworld AI's tools were used to adjust baseline emotional expressions and idle animations to suit a mental health companion.

The imported character in the project had to be positioned, rotated, and scaled to the correct proportions relative to the room. This process required trial and error, and it had to be made such that it aligned with the real-world surfaces and surroundings based on the camera position. The avatar was also fitted with a script to make it move wherever the user desired by just clicking on the spot. This was done to ensure the companion was placed in the position where the user was most comfortable. This level of personalization was aimed at giving the user the best experience possible.

Inworld AI, a character engine developed by TheGist, Inc. (dba Inworld AI, New York, NY, USA, with its platform publicly launched c.2022 and continuously updated), is used to create non-playable characters in games using AI, natural language processing, emotional simulation, and behavior modeling. It generates expressions and facial features based on the character's behavior using AI models that mimic human gestures.[10] The avatars were also modified in this study. This behavior is determined by the information we feed it. Characteristics such as anger, confidence, and aggressiveness, for example, can be changed using the user interaction tools.

These characteristics influence how the avatars communicate and interact with the user. This can be brought to the user by our custom mental health chatbot, which bypasses the default conversational settings.

The first step was to create the chatbot, which has already been described. Post-chatbot creation, it is to be interfaced with the character instead of the in-built GPT-3 model. Emotion detection is another important aspect of this application that needs to be improved upon in the next iterations of it.[11]

User Interface (UI)

The initial version of the UI had a canvas with buttons to allow the user to navigate between different functionalities of the application. However, this was removed to truly make the application even more user-friendly and non-frustrating. The idea of the avatar companion fulfilling all of the user's commands and the avatar acting as the interface was more appropriate. Thus, the UI was changed to make the companion-user relationship the most important aspect of the project. This also enables future iterations of the application to have more interesting use cases and functionalities.[12]

Natural Interactions

Talking with the character mimics human conversation. This is an intuitive approach that lowers the cognitive load for the user. This lower barrier to entry can be the difference between the app being used or not.

Convenience

As the interface is completely voice-based, people with disabilities, such as motor or vision impairments, can use the application. Going forward, it can use the information from the user to guide them to take medication on time.

Efficiency

The commands are much faster when voice-based compared to navigating through menus and typing the queries, as with conventional chatbots.

Personalization

As the chatbot learns from previous interactions with the user, this brings a level of personalization to the user interface that is just not possible through hard-coded menus. The chatbot has the ability to give personalized information in less time. It is hypothesized that, as the relationship between the user and ARden grows, the application will achieve higher retention and usage rates. This is much more than what a traditional interface can do.

RESULTS

Recommendation From Professionals

The AI companion was developed in consultation with mental health professionals to ensure the ethical integrity of the application while validating the responsible AI training, clinical relevance, and accuracy of their responses. The application developed primarily complements the therapeutic practices, addressing the gap in mental health delivery systems.[13] In alignment with this approach, the proposed idea of this AR-enabled chatbot application was taken to a mental health professional at the National Institute for Empowerment of Persons with Multiple Disabilities (NIEPMD) in India for initial validation. The key suggestions that were implemented include: ease of use, efficient interaction without too many menus, and personalization.

The other recommendation was to target general mental health needs, thereby broadening the scope of the application. This further enhanced the efficacy of the large language model by allowing it to specialize in high-demand and specific areas of mental health. Subsequent steps focused on niche areas such as stress management, anxiety, and depression, to name a few.

Once a basic prototype was developed, it underwent further evaluation for feedback on clinical efficacy. The user base was defined in the review process of the prototype. The prototype was intended to be prescribed to selected users based on their psychological profiles, following consultation with a certified mental health professional. Consequently, the chatbot should be deployed as part of a hybrid healthcare model, ensuring that the mental health professional remains actively involved in the patient care loop.

Patient confidentiality was a major concern. Based on these inputs and further research, it was evident that any solution in this domain must strictly adhere to data privacy regulations.[14] While our study has not yet encountered situations that raise privacy concerns, future research and similar product developments will need to ensure robust measures that protect patient data.

Parameter Tuning

As outlined in Figure 4, the mental health chatbot offers a range of features that are designed to be beneficial

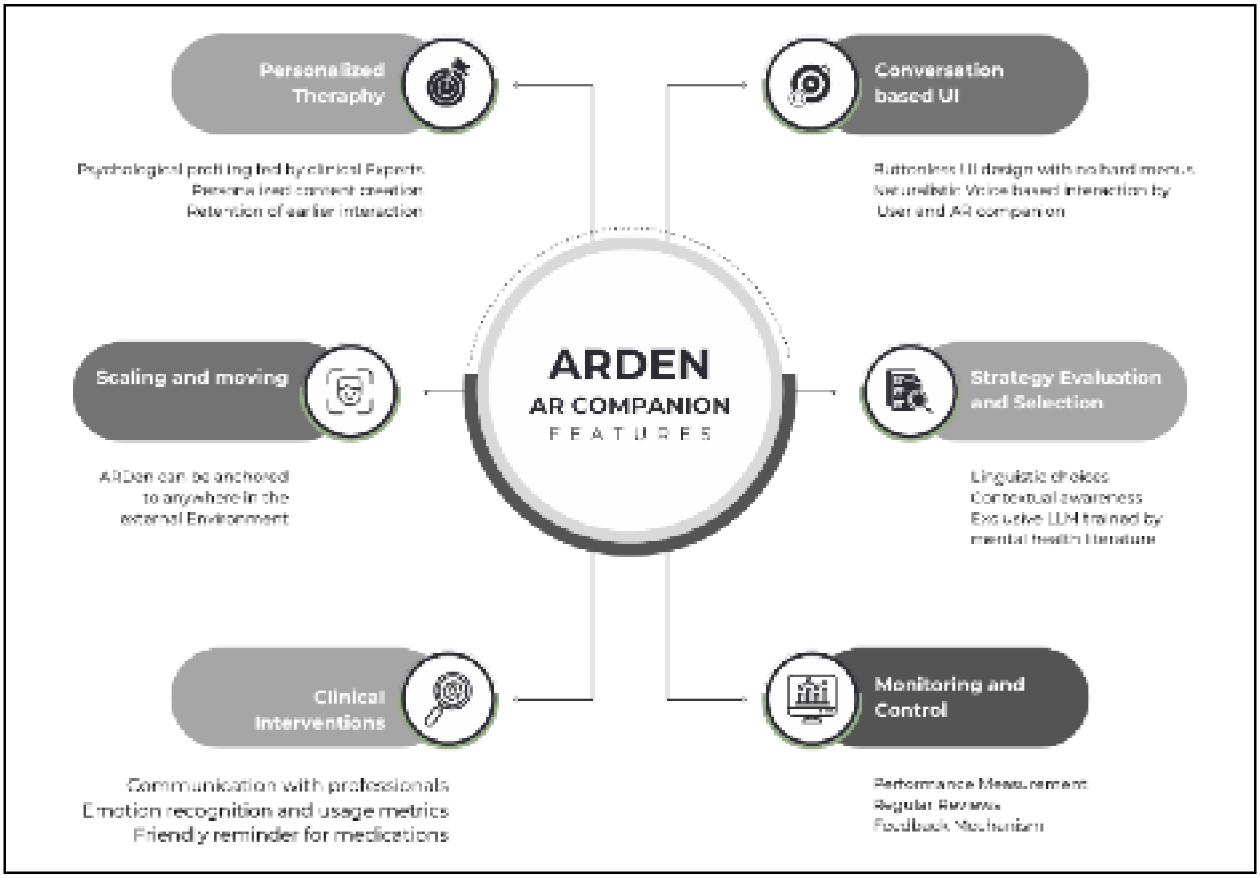


FIGURE 4. Features of AR companion.

to patients. To achieve this, various parameters are optimized to define the state-of-the-art large language model-based chatbot. Fine-tuning is accomplished by adjusting parameters such as temperature, maximum generation length, and sampling, among others.[15] Rigorous evaluations of the developed model are performed to mitigate risks and streamline responses to align with the user's needs and expectations.

The paradigm of the response is facilitated by the availability of these parameters. Temperature is a variable that tells us how incidental the output or the response generated is.[16] The values can be interpreted as follows: if the temperature values are low, the model is more deterministic. This is ideal for maintaining consistent and reassuring conversations in mental health settings. Higher values may introduce more variability, which may not be required for this purpose. Due to this, the temperature was originally set to 0.5.[17]

Tokens with a combined likelihood exceeding a threshold (p) are examined in top-p sampling, which is sometimes referred to as nucleus sampling. This restricts the token selection process by regularly modifying the queries according to their probabilities. This approach preserves focus on the majority of tokens while ensuring diversity.

As shown in Table 1, the maximum length of the response generated is set to be 512 words. Seeing as shorter responses can keep the conversations focused but might lead to leaving out details, this is an area with tradeoffs. Other parameters, such as learning rate and beam search width, also play a crucial role in making the model dynamic and user-friendly. The existing learning rate of the large language model is 0.0002.[18]

TABLE 1. Hyperparameters of Llama 2.

Parameters	Value
Learning Rate	0.0002
Response Speed	120–250 s
Chunk Size	512
Temperature	0.5
Sampling	Top P or nucleus sampling
Software Development Kit	Boto3(AWS SDK for Python)

A model that is fine-tuned with relevant mental health data and uses emotion tracking tools to adjust the model's tone and word choice would be ideal for this purpose. Setting parameters that encourage the chatbot to ask questions and take pauses would ensure user engagement, thereby enhancing interaction.

The recent Llama 3 model launched by Meta represents a more advanced and efficient iteration of large language models, offering significant potential for future applications in the development of mental health chatbots. The following comparisons (Table 2) outline the key improvements in the model, as indicated by Meta's advancements.

TABLE 2. Comparison of Llama 2 and Llama 3.

Feature	Llama 2	Llama 3
Training Data	Trained on around 2.2 trillion tokens	Trained on approximately 15 trillion tokens
Model Sizes	Released in 7B, 13B, and 70B parameter sizes	Available in 8B and 70B parameter versions
Context Window	Supports up to 4,096 tokens	Supports up to 8,192 tokens
Performance	Better performance over Llama 2	Outperforms Llama 3 across all benchmarks

Response of the chatbot

The chatbot shows compassion towards the user by first validating what the user feels and gently suggesting what the user could do to improve their situation. Throughout the entire conversation, the companion constantly exhibits friendly body language.

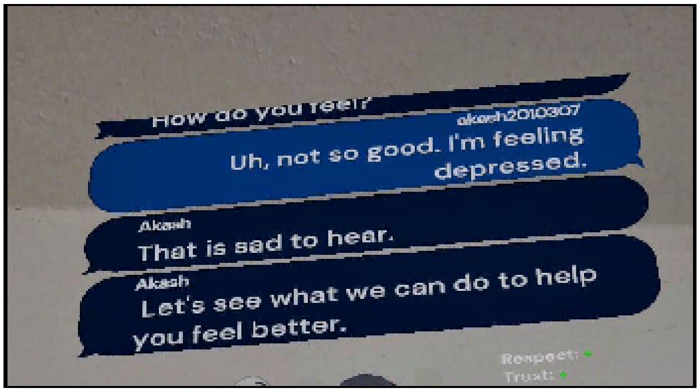


FIGURE 5. Sample response for depression.

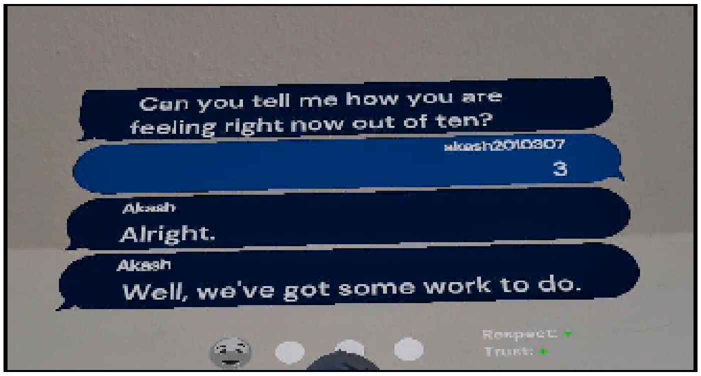


FIGURE 6. Sample response for anxiety.

As seen in Figure 5, the character interacts with the user with compassion and empathy, which is crucial when dealing with sensitive topics. In Figure 6, the chatbot asks the user to rate their feelings on a scale of 1 to 10, which lets the chatbot deal with the situation differently based on the user's response. When the situation is particularly difficult, the user is directed to consult with the therapist.

In Figure 7, the user is being suggested exercises to follow in order to feel better. In general, the exercises that therapists suggest to their patients can be reinforced in this app in order to ensure proper completion. Also, the app's interactive nature immerses the user in the exercise, leading to improved outcomes. In summary, the application interacts with the user in a compassionate way to provide insights and aims to improve mental health outcomes. An integrated mechanism for analyzing and

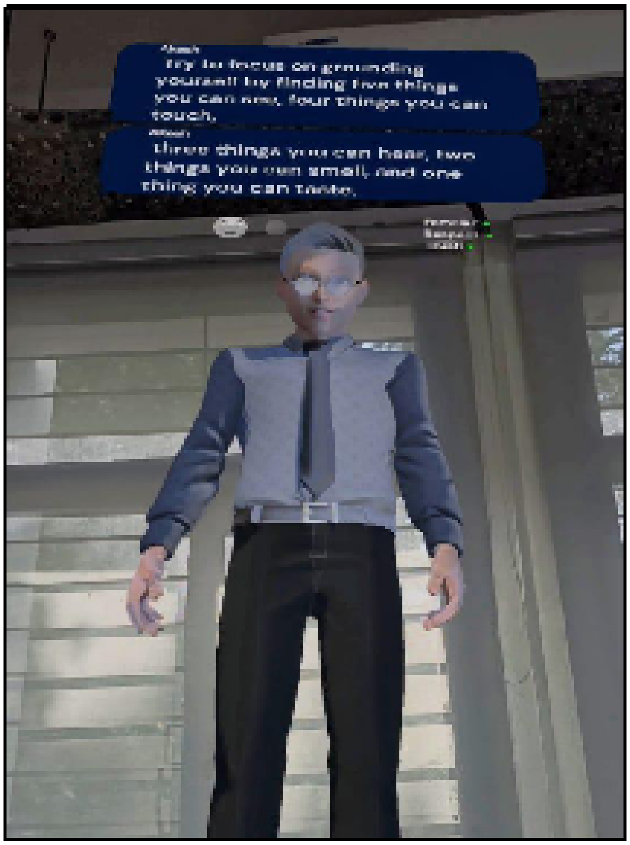


FIGURE 7. Sample exercise.

presenting emotions detected from the users' interactions with the chatbot proved to be useful. Throughout these interactions, prominent emotional states such as happiness, sadness, and anger could be identified to help understand the mental state of the user. It was evident that this analysis would help the user track their mood.

A formal study is planned to evaluate the therapeutic potential of the app. Participants' moods will be assessed over a 30-day period using daily self-report scales, with additional objective measures, such as physiological indicators or behavioral assessments, potentially incorporated. Data analysis will compare pre- and post-study mood scores, along with any improvements reported by participants. To further validate these findings, future research could focus on longitudinal studies and randomized controlled trials involving individuals diagnosed with mental health conditions, aiming to rigorously quantify the chatbot's impact on mental health outcomes.

Testing Feedback

The application was tested for user interaction (UI) experiences, and some bugs were observed within the app, such as the character moving on its own at times, the app crashing during longer interactions, and the character not always being anchored in the environment correctly. In terms of functionality, it was suggested to add more exercises and activities, like guided meditations.

Clinician Feedback

The designed application was evaluated by two different psychiatrists. They suggested further work on adding more evidence-based exercises and ensuring the clinical accuracy of the information provided. They also wanted more features in the app so that clinicians can, with patient consent, access summaries of patients' interactions or mood trends to better inform therapy sessions. Apart from this, they had questions regarding crisis management protocols within the app and the specifications of data privacy and security for sensitive user information. Along with these important considerations, the overall feedback was encouraging, with psychiatrists recognizing the potential of Arden to support individuals between interventions and overcome communication barriers.[19]

CONCLUSION AND FUTURE WORK

As discussed, the application was developed after analyzing the capabilities and limitations of existing solutions. It was determined that increased user engagement is essential for the application to be truly beneficial. With the advancement of AR technology and its capabilities, the study focused on designing a system that integrates both large language models (LLMs) and augmented reality (AR).

The companion application was successfully implemented using Unity and deployed on mobile platforms. The constructed avatar was equipped with an API interfacing with the LLM, which functions as a chatbot. This chatbot was fine-tuned with relevant medical literature to provide accurate and useful information to the user. Thus, the study produced an initial design of an AR-enabled avatar, equipped with an LLM, for emotional regulation purposes.

The design developed in this study serves as a reference for future advancements in the field. Designing an effective mental health chatbot involves several critical aspects, including, but not limited to, tracking the mood scores of users,[20] enhancing the LLM's capabilities,[21] maintaining data privacy, and ensuring the chatbot retains context from previous interactions with the user.

Mood scores are an effective method for monitoring the mood and mental health of a person over a period of time. This approach can be validated with the assistance of mental health professionals.[22] The system can offer more personalized support based on mood scores.

Collaboration with mental health professionals has offered important insights into the everyday challenges patients encounter. It is essential that products in this field are developed and tested alongside these professionals to ensure they effectively address real-world needs, particularly by improving the mental health of young adults.[23]

This study was conducted prior to the release of Llama 3, GPT-4, and other newer models. With further advancements in large language models and emotional recognition, future chatbots are expected to exhibit more human-like qualities and be capable of taking in multimodal inputs from users more effectively.

Throughout the development of this application, a few key findings emerged regarding emotion detection. It became apparent that an application of this sort must take into account the following for the emotion detection capabilities: the user's speech and voice tone patterns, facial expression changes, linguistic cues, contextual awareness of the model, and multimodal data integration capabilities. These are all areas where humans naturally excel and are currently superior compared to models.

Real-time emotion detection remains a challenging task, which requires further research. Taking emotion recognition capabilities alongside the other functional abilities of the application previously discussed, it is clear that further research is required for applications of this kind before they can be effectively deployed. We anticipate that, with further technological advancements, this design of an AR-enabled mental health chatbot will contribute to improved mental health support.

AUTHOR CONTRIBUTIONS

All authors contributed equally to this work.

FUNDING

Not applicable.

DATA AVAILABILITY STATEMENT

Not applicable.

CONFLICTS OF INTEREST

The authors declare they have no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

FURTHER DISCLOSURE

Not applicable.

REFERENCES

1. Denecke, K., Abd-Alrazaq, A., Househ, M. Artificial Intelligence for Chatbots in Mental Health: Opportunities and Challenges. In *Multiple perspectives on artificial intelligence in healthcare: Opportunities and challenges*. Househ, M., Borycki, E., Kushniruk A. eds. Springer Nature, Switzerland; 2021; p.115–128; https://doi.org/10.1007/978-3-030-67303-1_10.
2. Potts, C., Lindström, F., Bond, R., et al. A multilingual digital mental health and well-being chatbot (ChatPal): pre-post multicenter intervention study. *J Med Internet Res*. 2023;25:e43051. https://doi.org/10.2196/43051.
3. Farhat, F. ChatGPT as a complementary mental health resource: A boon or a bane. *Ann Biomed Eng*. 2024;52(5):1111–1114. https://doi.org/10.1007/s10439-023-03326-7.
4. Yang, R., Tan, T.F., Lu, W., et al. Large language models in health care: development, applications, and challenges. *Health Care Sci*. 2023;2(4):255–263. https://doi.org/10.1002/hcs2.61.
5. Booth F., Potts, C., Bond, R., et al. A mental health and well-being chatbot: user event log analysis. *JMIR Mhealth Uhealth*. 2023;11:e43052. https://doi.org/10.2196/43052.

6. Roumeliotis, K., Tselikas, N., Nasiopoulos, D. Llama 2: Early adopters' utilization of Meta's new open-source pretrained model. *Preprints*. 2023. https://doi.org/10.20944/preprints202307.2142.v1.
7. Yang, K., Ji, S., Zhang, T., et al. Towards interpretable mental health analysis with large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing,* Singapore, 6-10 December, 2023. https://doi.org/10.18653/v1/2023.emnlp-main.370.
8. Wu, C., Lin, W., Zhang, X., et al. PMC-LLaMA: toward building open-source language models for medicine. *J Am Med Inform Assoc*. 2024;31(9):1833–1843. https://doi.org/10.1093/jamia/ocae045.
9. Shields, R.E., Korol, S., Carleton, R.N., et al. Brief mental health disorder screening questionnaires and use with public safety personnel: a review. *Int J Environ Res Public Health*. 2021;18(7):3743. https://doi.org/10.3390/ijerph18073743.
10. Singha, R., Singha, S. Mental health treatment: exploring the potential of augmented reality and virtual reality. In *Applications of Virtual and Augmented Reality for Health and Wellbeing*; Hiran, K.K., Doshi, R., Patel, M., eds. IGI Global: Pennsylvania, USA; 2024; pp.91–110. https://doi.org/10.4018/979-8-3693-1123-3.ch006.
11. Bakır, Ç.N., Abbas, S.O., Sever, E., et al. Use of augmented reality in mental health-related conditions: a systematic review. *Digit Health*. 2023;9:20552076231203649. https://doi.org/10.1177/20552076231203649.
12. Lecomte, T., Potvin, S., Corbière, M., et al. Mobile apps for mental health issues: meta-review of meta-analyses. *JMIR Mhealth Uhealth*. 2020;8(5):e17458. https://doi.org/10.2196/17458.
13. Kernkraut, A.M., Diniz Nagem Janot de Matos, L. The impact of negative effects on the high prevalence of anxiety, depression and stress in healthcare professionals. *Work*. 2024; 79(2), 857–866. Available Online: https://journals.sagepub.com/doi/abs/10.3233/WOR-230257?download=true#:~:text=It%20was%20observed%20that%20a,of%20depression%20and%20stress%20prevalence.
14. May, R., Denecke, K. Security, privacy, and healthcare-related conversational agents: a scoping review. *Inform Health Soc Care*. 2022;47(2):194–210. https://doi.org/10.1080/17538157.2021.1983578.
15. Sahoo, S.S., Plasek, J.M., Xu, H., et al. Large language models for biomedicine: foundations, opportunities, challenges, and best practices. *J Am Med Inform Assoc*. 2024;31(9):2114–2124. https://doi.org/10.1093/jamia/ocae074.
16. Boucher, E.M., Harake, N.R., Ward, H.E., et al. Artificially intelligent chatbots in digital mental health interventions: a review. *Expert Rev Med Devices*. 2021;18(sup1):37–49. https://doi.org/10.1080/17434440.2021.2013200.
17. Touvron, H., Martin, L., Stone, K., et al. Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288. 2023.
18. Zhang, R., Han, J., Liu, C, et al. LLaMA-adapter: Efficient fine-tuning of large language models with zero-initialized attention. In *the Twelfth International Conference on Learning Representations*, Vienna Austria, 7–11 May, 2024. Available Online: https://openreview.net/pdf?id=d4UiXAHN2W.
19. Carlson, C. Virtual and augmented simulations in mental health. *Curr Psychiatry Rep*. 2023;25(9):365–371. https://doi.org/10.1007/s11920-023-01438-4.
20. Hsu, M.C. The construction of critical factors for successfully introducing chatbots into mental health services in the army: Using a hybrid MCDM approach. *Sustainability*. 2023;15(10):7905. https://doi.org/10.3390/su15107905.
21. Dosovitskiy, G. and Bunge, E. Development of a chatbot for depression: adolescent perceptions and recommendations. *Child Adolesc Ment Health*. 2023;28(1):124–127. https://doi.org/10.1111/camh.12627.
22. Furukawa, T.A. Assessment of mood: guides for clinicians. *J Psychosom Res*. 2010;68(6):581–589. https://doi.org/10.1016/j.jpsychores.2009.05.003.
23. Berryman, C., Ferguson, C.J., Negy, C. Social media use and mental health among young adults. *Psychiatry Q*. 2018;89(2):307–314. https://doi.org/10.1007/s11126-017-9535-6.