|  |  |  |
| --- | --- | --- |
| amblem | **ÇANKAYA UNIVERSITY**  **Software Engineering Department** | **A circular logo with colorful arrows  Description automatically generated** |

**SENG 491 – 492 Graduation Project**

**Literature / Market**

**Research**

**SenseAI**

**Ece Heval Ünal**

**Mustafa Tozman**

**Özge Alkan**

**Sadrettin Anıl Karaçay**

**Zeliha Aybüke Baştürk**

**Version 1.1**

Table of Contents

[1 Introduction 1](#_Toc181140826)

[1.1 Purpose 1](#_Toc181140827)

[1.2 Scope of this document 1](#_Toc181140828)

[1.3 Definitions 1](#_Toc181140829)

[1.4 Version History 1](#_Toc181140830)

[2 Proposed Project 1](#_Toc181140831)

[2.1 Description 1](#_Toc181140832)

[2.2 The Need 2](#_Toc181140833)

[2.3 Technology 2](#_Toc181140834)

[2.4 Risks 2](#_Toc181140835)

[3 Existing/Similar Solutions 2](#_Toc181140836)

[3.1 Only Text-Based Chatbots 2](#_Toc181140837)

[3.1.1 Solution Description 2](#_Toc181140838)

[3.1.2 Shortcomings 3](#_Toc181140839)

[3.2 Ellie (SimSensei) 3](#_Toc181140840)

[3.2.1 Solution Description 3](#_Toc181140841)

[3.2.2 Shortcomings 3](#_Toc181140842)

[3.3 Replika AI 4](#_Toc181140843)

[3.3.1 Solution Description 4](#_Toc181140844)

[3.3.2 Shortcomings 4](#_Toc181140845)

[3.4 Moodfit 4](#_Toc181140843)

[3.4.1 Solution Description 4](#_Toc181140844)

[3.4.2 Shortcomings 4](#_Toc181140845)

[3.5 Article- A survey of Deep Learning-Based Multimodal Emotion Recognition: Speech](#_Toc181140840), Text, and Face…………………………………………………………………………………………………………..4

[3.5.1 Solution Description…………………………………….](#_Toc181140841)……………………………………………..4

[3.5.2 Shortcomings…………………………………………………………………………………………….5](#_Toc181140842)

[3.3 Replika AI 4](#_Toc181140843)

[3.3.1 Solution Description 4](#_Toc181140844)

[3.3.2 Shortcomings 4](#_Toc181140845)

[3.4 Moodfit 4](#_Toc181140843)

[3.4.1 Solution Description 4](#_Toc181140844)

[3.4.2 Shortcomings 4](#_Toc181140845)

[4 Conclusion 4](#_Toc181140846)

[5 References 5](#_Toc181140847)

# Introduction

## Purpose

The purpose of this document is to provide a summary of the market research we conducted for our project SenseAI. [1]

## Scope of this document

The scope of the document includes the definition of our project, the need it meets, the technologies to be used, risks and solutions similar to our project.

## Definitions

|  |  |
| --- | --- |
| Term | Description |
| CBT | Cognitive Behavioral Therapy |
| AI | Artificial Intelligence |
| NLP | Natural Language Processing |
| CNN | Convolutional Neural Network |
| VR | Virtual Reality |
|  |  |
|  |  |

## Version History

|  |  |  |
| --- | --- | --- |
| **Version No** | **Description of change** | **Date** |
| 1.0 | Initial Release | 16.10.2024 |
| 1.1 | More details are added about the similar projects | 23.10.2024 |
| 1.2 | Added articles about similar projects. | 30.10.2024 |

# Proposed Project

## Description

This project aims to develop a multimodal psychology chatbot that analyzes emotional states and user contexts through inputs such as speech, facial expressions, and biometric data from wearable devices (e.g. smartwatches). The system will perform real-time sentiment analysis to provide personalized feedback. Integrating multiple data formats for a comprehensive sentiment analysis and combining different types of inputs will provide an enriched sentiment analysis based on multimodal data processing, personalized, real-time response, and enable tracking of long-term mood patterns for sustainable emotional development. The chatbot will help users track their emotional well-being over time and provide tailored recommendations for stress management and mental health improvement.

## The Need

This project was developed to meet the needs of individuals seeking instant psychological support. The stress of modern life and daily challenges cause individuals to need urgent emotional support from time to time. This chatbot analyzes users' current moods and provides personalized suggestions and comforting feedback.

The system evaluates the user's emotional state in real time using multiple data modes such as tone of voice, facial expressions and biometric data. In this way, it allows individuals to quickly receive support and reach comforting suggestions in times of stress.

In addition, the chatbot does not only provide instant support, but also analyzes individuals' mood changes over time by monitoring their long-term emotional state. It contributes to the spiritual development of users by sustainably monitoring their emotional well-being and offers personalized recommendations on topics such as stress management and mental health improvement.

## Technology

Leveraging AI, NLP, computer vision, and signal processing techniques, the system integrates existing solutions such as transformers for sentiment analysis, CNNs for facial recognition, and machine learning algorithms for biometric data analysis. Labeled data from existing datasets will be available to train and evaluate models, ensuring high emotional accuracy, fast response time, and user satisfaction.

## Risks

* Not being able to find sufficient labeled data
* A product similar to our project came out before us
* The accuracy of the results since there are no psychology students in our team

# Existing/Similar Solutions

## Only Text-Based Chatbots

### Solution Description

Although some of these applications have different purposes and support, the common point of all of them is that they are only text-based. Although there are many psychology chatbots that are text-based, we have listed the most used ones and their features.

* + WOEBOT[2]: The focus area is to provide emotional support to users using CBT techniques.
  + WYSA[3]: It helps users manage stress and anxiety through CBT, mindfulness exercises, meditation, and other techniques. In the paid version, communication with therapists is possible.
  + YOUPER[4]: It helps users with mood tracking and emotional awareness using CBT techniques. The app can direct users to therapists.
  + HEADSPACE[5]: The main focus of the app is stress and anxiety management, and users can instantly connect with mental health professionals and receive guidance.
  + X2AI[6]: It provides support to users on issues such as stress, anxiety and depression with CBT techniques.

### Shortcomings

The difference between our project and the projects above is that, in addition to being text-based, it is a multimodal chatbot that receives image, audio and biometric data from users.

## Ellie (SimSensei)

### Solution Description

Ellie is supported by DARPA (The Defence Advanced Research Projects Agency) which is under US Department of Defense and developed at the University of Southern California to assess the mental health of military veterans. It [7][9] analyzes users' moods by only capturing voice tone and facial expressions. It leverages tools like Stanford NLP, OpenFace, FACS (Facial Action Coding System), OpenCV, and Praat, along with AI frameworks like TensorFlow and PyTorch, to offer personalized interactions. Unity3D or Unreal Engine powers its 3D avatar and techniques like Pitch Tracking, Vocal Timbre Analysis, Feature Fusion and Pose Estimation.

### Shortcomings

It is the project most similar to ours, but our project will provide analysis reports and image-voice analysis to the user by obtaining biometric data. Ellie does not offer text-based support. Also, since Ellie was a project only for soldiers, its target audience was quite limited and difficult to access. However, our project will be accessible to everyone.

## Replika AI

### Solution Description

Replika AI [8] offers a virtual companion that keeps the user company and meets their needs. Here, users cannot communicate with the application through images while communicating their requests via text and voice. The application offers VR support to make the virtual friendship environment more realistic.

### Shortcomings

Our project will support the user by also receiving image data and biometric data from the users. Our project, which aims to provide psychological support instead of a friend, will not have VR support.

**3.4 Moodfit**

**3.4.1 Solution Description**

Moodfit [10]  allows users to track their mood by asking questions at regular intervals. The app offers tools such as mood diary, thought diary, and stress management. Users can evaluate themselves with surveys and activities. Moodfit also provides graphs and reports that analyze users' moods.

**3.4.2 Shortcoming**

This application does not provide artificial intelligence support to users. Only the answers to the questions asked by the user are taken as data. The text, voice, image and biometric data entry possibilities in our project are not available in this application.

**3.5  Article- A Survey of Deep Learning-Based Multimodal Emotion Recognition: Speech, Text, and Face[11]:**

**3.5.1 Solution Description**

This article examines multiple emotion recognition and developments in this field through deep learning techniques. It aims to recognize emotions by combining data such as text, audio, and facial expressions. The article focuses on how deep learning techniques and algorithms are applied, and aims to examine data sets and help researchers choose the appropriate data set with their advantages and disadvantages.

**3.5.2  Shortcomings**

In this article, unlike our project, biometric data is not used. In our project, thanks to the biometric data we receive from the user, other types of data received from the user will be interpreted more accurately.

**3.6 Article- Depression Detection Using Multimodal Analysis with Chatbot Support [12]:**

**3.6.1**  **Solution Description**

While this article aims to detect early depression, it is aimed to make more comprehensive depression detection by combining image, audio and text data from users.

**3.6.2 Shortcomings**

In this article, unlike our project, biometric data is not used. In our project, thanks to the biometric data we receive from the user, other types of data received from the user will be interpreted more accurately.

**3.7 Article- Multimodal Emotion Recognition Based on Facial Expressions, Speech, and EEG [13]:**

**3.7.1 Solution Description**

In this study, a deep learning based multimodal emotion recognition (MER) method called "Deep-Emotion" is proposed, which adaptively combines the most distinctive features such as facial expressions, voice and electroencephalogram (EEG). It is aimed to provide high accuracy and durability by collecting electrical signals from brain activity through electrodes placed on the user's scalp.

**3.7.2 Shortcoming**

Unlike this study, our project will not receive EEG data. Instead, text and biometric data will be entered to support analysis report results.

**3.8 Article- Multimodal Emotion Recognition Based on Facial Expressions, Speech, and Body Gestures [14]:**

**3.8.1 Solution Description**

In this study, it uses voice, facial expressions and body gestures data for emotion recognition. The results obtained by using these three modalities together have higher success rates than the results obtained by using only two modalities.

**3.8.2 Shortcoming**

Unlike this study, we will not use body gestures data. We aim to return more accurate results to the user by adding text and biometric data and using more types of data together.

# Conclusion

Among the projects we mentioned above, there is no application that collects text, audio, video, and biometric data together as we will implement in our project. We believe that with a multimodal chatbot that receives many data from various areas and gives analysis reports, the data received from the user will be interpreted more accurately and will help the user in a better way. When we combine the biometric data we collect with other types of data, we will obtain meaningful results and support them through many channels. Another advantage of our project is that it is an open and easily accessible application that everyone can use from their phone, and the user can access the analysis results and their past data.

# References

|  |  |
| --- | --- |
| [1] | Software Engineeering Department, "Graduation Projects," SENG, [Online]. Available: https://seng.cankaya.edu.tr/graduation-projects/. [Accessed 28 June 2024]. |

[2] Woebot Health, "Woebot Health - Mental Health for the Digital Age," Available: <https://woebothealth.com/>. [Accessed: 24-Oct-2024].

[3] Wysa, "Reports," Available: <https://www.wysa.com/reports#6>. [Accessed: 24-Oct-2024].

[4] Youper, "AI Therapy," Available: <https://www.youper.ai/>. [Accessed: 24-Oct-2024].

[5] Headspace for Work, "Headspace for Work," Available: <https://organizations.headspace.com/>. [Accessed: 24-Oct-2024].

[6] Cass.ai, "X2AI Home," Available: <https://www.cass.ai/x2ai-home>. [Accessed: 24-Oct-2024].

[7] News.com.au, "Meet Ellie, the robot therapist treating soldiers with PTSD," Available: <https://www.news.com.au/technology/innovation/meet-ellie-the-robot-therapist-treating-soldiers-with-ptsd/news-story/0201fa7cf336c609182cffd637deef00>. [Accessed: 24-Oct-2024].

[8] Replika, "Replika - My AI Friend," Available: https://replika.ai/. [Accessed: 24-Oct-2024].

[9] Whatscreativeluc, "Hi, I’m Ellie," Available: <https://whatscreativeluc.blogspot.com/2018/11/hi-im-ellie.html>. [Accessed: 24-Oct-2024]

[10] Moodfit. (n.d.). *Moodfit: Your mental health companion.* [Online]. Available:<https://www.getmoodfit.com/>. [Accessed: Oct. 31, 2024].

[11 H. Lian, C. Lu, S. Li, Y. Zhao, C. Tang, and Y. Zong, "A Survey of Deep Learning-Based Multimodal Emotion Recognition: Speech, Text, and Face," *Entropy*, vol. 25, no. 10, pp. 1440, 2023. doi: 10.3390/e25101440

[12] Archana Sharma, Anuradha Saxena, Ashok Kumar, ve Divyansh Singh, "Depression Detection Using Multimodal Analysis with Chatbot Support," *2023 IEEE International Conference on Computational Intelligence and Data Science (ICCIDS)*, Chennai, Hindistan, 2023, ss. 1-6, doi: 10.1109/ICCIDS56901.2023.10489080

[13]  S. Wang, Z. Zhang, B. Chen, Z. Zhang, W. Fang, and J. Pan, "Multimodal Emotion Recognition Based on Facial Expressions, Speech, and EEG," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 142-154, Jan.-March 2022, doi: 10.1109/TAFFC.2021.10026861.

[14] J. Yan, P. Li, C. Du, K. Zhu, X. Zhou, Y. Liu, and J. Wei, "Multimodal Emotion Recognition Based on Facial Expressions, Speech, and Body Gestures," *Electronics*, vol. 13, no. 18, p. 3756, Sep. 2024, doi: 10.3390/electronics13183756.