# ÇANKAYA UNIVERSITY


## Computer Engineering Department


## SenseAI - Literature Review


**Ece Heval Ünal**

**Mustafa Tozman**

**Özge Alkan**

**Sadrettin Anıl Karaçay**

**Zeliha Aybüke Baştürk**

TABLE OF CONTENTS

## Abstract
This paper reviews the literature on multimodal, AI-driven psychology chatbots that aim to assess and support emotional well-being by integrating inputs from speech, facial expressions, text, and biometric data. Leveraging advances in Artificial Intelligence (AI), Natural Language Processing (NLP), computer vision, and signal processing, such systems enable real-time emotional analysis and personalized feedback. This literature review covers key research and technologies underpinning the project, including sentiment analysis, computer vision, wearable devices, and multimodal data fusion, and examines their applications in psychological support tools. The review highlights advances in sentiment analysis with transformers, facial recognition with convolutional neural networks (CNNs), and biometric analysis through machine learning models, providing a roadmap for developing comprehensive, real-time emotional tracking systems.

## Introduction
Recent years have seen significant advances in digital mental health tools, particularly AI-driven systems that analyze and interpret emotional states for improved psychological support. The growth in wearable technologies and machine learning models has created new opportunities to integrate multimodal data sources, such as text, speech, facial expressions, and biometric signals, allowing for a holistic understanding of emotional well-being. This project aims to build a psychology chatbot that leverages multimodal data to analyze emotional states, offer personalized feedback, and track users' emotional patterns over time. This literature review outlines relevant studies on multimodal emotional analysis, describes technologies such as NLP and computer vision, and discusses the ethical considerations involved in mental health chatbot development.

## 1. Background and Motivation
### 1.1 The Role of Chatbots in Mental Health Support
In recent years, digital tools have increasingly played a significant role in supporting mental health, with AI-powered chatbot systems emerging as key tools for addressing emotional needs. Mental health support chatbots provide users with a safe, non-judgmental environment to express their emotions and thoughts, offering a valuable space for self-reflection, stress management, and basic psychological support.

Research has shown that even text-based chatbots have a positive impact on users' mental well-being. For instance, applications like Wysa and Woebot employ Cognitive Behavioral Therapy (CBT) techniques to help users manage stress and anxiety effectively. However, these text-based systems have inherent limitations, as they rely solely on verbal input to interpret users' emotions and thoughts. This project, SenseAI, aims to overcome these limitations by incorporating multimodal data—such as voice tone, facial expressions, and biometric signals—to enable a more comprehensive emotional assessment, thus fostering more empathetic and accurate interactions.

The potential for chatbots to offer mental health support is immense, particularly as they can provide immediate assistance in times of need, help track emotional changes over time, and offer coping strategies that may reduce the burden on traditional mental health services. By incorporating multimodal emotional analysis, SenseAI seeks to deliver a more human-like, empathetic interaction that surpasses the capabilities of standard text-based chatbot solutions.

### 1.2 Advancements in Multimodal Emotional Analysis

The field of emotion recognition has progressed significantly, moving beyond traditional text-based sentiment analysis to sophisticated multimodal approaches that integrate visual and physiological data. Multimodal emotional analysis enables more accurate predictions by combining inputs from various sources, such as voice tone, facial expressions, and heart rate variability (HRV).

Multimodal emotion recognition systems leverage physical and behavioral indicators—such as speech tone, facial cues, and biometric signals—to construct a nuanced understanding of users' emotional states. For instance, HRV is a valuable biometric indicator often associated with stress levels, while facial expression analysis can reveal emotions like happiness, sadness, or anger. By combining these modalities, multimodal systems achieve higher accuracy and allow for a deeper understanding of the user’s emotional context, ultimately enabling more empathetic responses in chatbot interactions. This project leverages these advancements to build a mental health chatbot that provides robust emotional support by integrating multimodal data sources to improve both the relevance and empathy of its responses.

## 2. Key Technologies and Methodologies

### 2.1 Natural Language Processing (NLP) for Sentiment Analysis

Natural Language Processing (NLP) plays a fundamental role in understanding the emotional tone conveyed in a user’s text or speech input. In recent years, advancements in transformer-based models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have significantly improved the ability to capture complex linguistic patterns, making them ideal for sentiment analysis tasks. Transformers use self-attention mechanisms that allow models to consider the context of each word within a sentence, capturing subtle nuances in language that can reflect emotional states.

In this project, transformer models will be fine-tuned specifically for detecting emotional cues in conversational text. Research suggests that when sentiment analysis is combined with other emotional indicators, such as voice and facial expressions, the overall accuracy and reliability of emotion detection improve. This approach provides a solid foundation for a multimodal chatbot system like SenseAI, enabling it to interpret and respond to user inputs with greater emotional sensitivity and accuracy.

## 2.2 Computer Vision for Facial Expression Analysis

Computer vision, particularly using Convolutional Neural Networks (CNNs), is essential for analyzing facial expressions that convey emotions. CNNs have proven highly effective in identifying facial features and understanding emotions such as happiness, sadness, or anger through the recognition of subtle changes in expressions. Studies have shown that combining facial expression data with other modalities, like text and biometric signals, enhances the accuracy of emotion recognition.

For the SenseAI project, pre-trained CNN models, such as those trained on the FER+ or AffectNet datasets, can be adapted to recognize real-time facial expressions. This functionality enables the chatbot to respond in a manner that reflects empathy and aligns with the user's emotional state. Facial recognition techniques are crucial for providing non-verbal cues, which add depth to the chatbot's responses and allow for a more human-like interaction.

## 2.3 Biometric Data Analysis from Wearable Devices

Biometric data, such as heart rate variability (HRV) and skin conductance, provide valuable insights into a user's emotional arousal and stress levels. Wearable devices can capture this data in real time, offering an additional layer of information that complements text and visual inputs. Machine learning algorithms can process these biometric signals to detect emotional states, such as anxiety or calmness, by analyzing patterns over time.

The literature indicates that combining wearable-derived data with other modalities improves the accuracy of emotional assessments. For instance, HRV is a commonly used metric in stress detection, as fluctuations in heart rate are associated with changes in emotional arousal. In the SenseAI project, biometric data will be processed using time-series machine learning models to identify stress patterns, allowing the chatbot to offer timely interventions and personalized feedback based on the user's physiological state.

# 3. Multimodal Data Fusion

### 3.1 The Importance of Multimodal Fusion in Emotion Recognition

Multimodal fusion combines data from various sources—text, visual, and biometric—to create a comprehensive emotional profile. Research has demonstrated that multimodal fusion techniques, such as feature-level and decision-level fusion, yield significantly higher accuracy in emotion detection by integrating complementary data streams. Each modality brings unique insights: while text may reveal explicit thoughts or concerns, facial expressions and biometric data provide non-verbal cues that deepen the understanding of a user's emotional state.

For psychological applications like SenseAI, multimodal fusion is crucial to improving both response relevance and empathic accuracy. By integrating various emotional indicators, the chatbot can respond in a manner that aligns more closely with the user's overall emotional profile. This feature is essential in mental health contexts, where nuanced and context-aware responses foster a supportive and safe environment for the user.

## 3.2 Data Fusion in Multimodal Emotion Recognition

Data fusion in multimodal emotion recognition involves combining insights from various data sources to enhance the accuracy, reliability, and robustness of emotional assessments. In SenseAI, fusion techniques are applied across different levels to create a nuanced understanding of emotional states. Below are the relevant fusion levels utilized within the project:

- **Low-Level Fusion (Raw Data Combination):** At this level, raw data from each modality—such as speech tone, facial expressions, and heart rate—is collected and processed in parallel. This approach provides a holistic view of the user's physiological and behavioral states, forming a foundation for more detailed feature extraction and emotional analysis.
- **Feature-Level Fusion:** In feature-level fusion, specific features indicative of emotions are extracted from each modality and then combined to create a unified representation of the user's emotional state. For example, speech features may include pitch and tone, facial expressions may focus on key landmarks like eyebrow movements or mouth curvature, and biometric data may capture HRV or skin conductance. By merging these features, the system leverages the complementary strengths of each modality to provide a more accurate and dynamic emotional profile.
- **Decision-Level Fusion:** In decision-level fusion, separate predictions are generated from each modality and then combined to form a final emotional assessment. This level relies on algorithms, such as voting schemes or weighted decision models, that merge outputs from each modality, allowing for increased accuracy by mitigating the limitations of individual data sources. This method enhances robustness in emotional recognition, as it synthesizes

information from text, visual, and biometric sources to deliver consistent emotional assessments.

## 4. Data Sources and Datasets

A variety of datasets are essential for training multimodal emotion recognition models, as they provide a foundation for developing accurate and reliable AI-driven psychology chatbots. SenseAI will utilize datasets that span textual, visual, and biometric data to ensure comprehensive emotional understanding across different modalities.

- **Text and Speech Datasets:** For sentiment and emotional tone detection, datasets such as the **IEMOCAP** (Interactive Emotional Dyadic Motion Capture) and **EmoReact** are instrumental. These datasets contain annotated conversational text and speech samples labeled with emotions like happiness, sadness, and anger, providing a robust foundation for training Natural Language Processing (NLP) models to detect emotional cues. Multilingual resources, such as **EmoReact**, help the model adapt to users from diverse linguistic backgrounds, enhancing the inclusivity and applicability of the chatbot.
- **Facial Expression Datasets:** Datasets like **FER+ (Facial Expression Recognition)** and **AffectNet** provide large collections of labeled facial images that are crucial for facial emotion recognition models. These datasets are used to train Convolutional Neural Networks (CNNs) to analyze facial expressions in real time, supporting emotion recognition based on visual input. By incorporating these datasets, SenseAI can assess non-verbal cues like smiles, frowns, and other subtle facial movements, making its responses more empathetic and accurate.
- **Biometric Data Datasets:** For emotional detection through biometric signals, the **WESAD (Wearable Stress and Affect Detection)** dataset is a valuable resource. WESAD includes sensor data collected from wearable devices, capturing physiological indicators such as heart rate, skin conductance, and respiratory rate. This dataset allows machine learning models to understand and interpret stress and arousal levels in real-time, enabling SenseAI to respond to the physiological states of users dynamically and with heightened emotional sensitivity.

By combining these diverse datasets, SenseAI will achieve a balanced understanding of user emotions across modalities, allowing for holistic emotional assessments. The diversity and quality of these datasets play a critical role in building a chatbot that adapts accurately to various emotional cues and supports a broad range of users.

## 5. Model Training

Training the emotion recognition model in SenseAI involves processing multimodal data—including text, facial expressions, and biometric signals—to accurately assess and track emotional states. Each data modality is processed individually to harness its unique features, followed by integration to create a cohesive emotional profile. Below is an outline of the training approach for each modality:

- **Text Data Processing (NLP):** Text data, derived from user interactions, is tokenized and fed into a transformer-based model, such as BERT or GPT. These models are fine-tuned to detect subtle emotional cues in conversational text, leveraging attention mechanisms to capture the context and sentiment behind user statements. Transformer models are particularly effective in understanding nuances, making them an essential component for analyzing language that conveys emotional depth.
- **Facial Expression Processing (Computer Vision):** Facial expression data is processed using Convolutional Neural Networks (CNNs), which are pre-trained on extensive facial emotion datasets like FER+ and AffectNet. These CNN models are then fine-tuned to detect relevant facial expressions in real-time video feeds, enabling SenseAI to capture emotional changes in users' facial expressions during conversations. By recognizing expressions such as smiles, frowns, and surprise, the chatbot can adjust its responses to better align with the user's current emotional state.
- **Biometric Signal Processing (Time-Series Analysis):** Biometric signals, including heart rate and skin conductance, are analyzed using time-series machine learning models that can detect patterns associated with emotional arousal and stress. For instance, fluctuations in heart rate variability (HRV) are closely associated with stress levels, making HRV a key metric for emotional analysis. By training on the WESAD dataset, these models learn to interpret biometric signals in real-time, adding an additional dimension of emotional awareness to the chatbot's response generation.

**Performance Metrics and Model Evaluation:** The performance of each model is evaluated using accuracy, precision, and recall metrics, aiming for a balanced trade-off between true positive and false positive rates. Cross-modal consistency is also monitored to ensure that emotional predictions remain coherent across different data types, enhancing the model's reliability. This cross-checking mechanism helps refine the chatbot's responses, making them more aligned with the user's overall emotional state.

**Cross-Modality Training:** A unique feature of SenseAI's model training is the integration of cross-modality learning, where the predictions from one modality (e.g., text sentiment) are cross-validated with data from another (e.g., facial expressions or biometric signals). This approach helps improve the accuracy and consistency of the emotional assessment, providing a holistic view that draws from multiple data points. This training approach enables SenseAI to deliver nuanced, data-driven emotional insights, making it a reliable tool for mental health support.

# 6. Challenges

The development of a multimodal, AI-driven chatbot for mental health support presents various challenges that must be addressed to ensure effectiveness, reliability, and user trust. Below are some of the primary technical, ethical, and practical challenges encountered in this project, along with proposed solutions:

### 6.1 Data Privacy and Security

Handling multimodal data—particularly sensitive information such as biometric signals and facial expressions—raises significant privacy concerns. Users may be reluctant to share personal data if they are unsure of its security or potential misuse. Ensuring the confidentiality of this information is paramount in mental health applications.

- **Solution:** Data anonymization and secure data storage techniques, such as end-to-end encryption, can mitigate privacy risks. Additionally, implementing data minimization practices ensures that only essential information is collected. Informed consent and transparency about data use further build trust. Techniques such as federated learning, where data is processed locally on the user's device, can be explored to enhance privacy while minimizing the transfer of sensitive data.

### 6.2 Computational Efficiency and Real-Time Processing

Multimodal data fusion and real-time emotion recognition require substantial computational resources, especially when integrating multiple complex models, such as CNNs for facial recognition and transformers for NLP. For the system to deliver timely and responsive feedback, computational efficiency is critical.

- **Solution:** Optimization strategies such as model pruning and quantization can reduce the computational load, making real-time processing more feasible. Additionally, exploring edge computing—where data processing occurs locally on the device rather than in the cloud—can reduce latency and enhance user experience. By using lightweight, pre-trained models and adapting them for on-device processing, SenseAI can improve responsiveness without sacrificing accuracy.

### 6.3 Emotional Sensitivity and User Safety

In mental health applications, incorrect or insensitive responses can lead to emotional distress, undermining the effectiveness of the chatbot. If the chatbot misinterprets or overreacts to certain emotions, it could inadvertently exacerbate a user's emotional state rather than providing support.

- **Solution:** SenseAI incorporates a feedback mechanism where users can rate responses, helping the system improve over time based on user feedback. Additionally, the chatbot's responses are calibrated to avoid overly assertive or diagnostic language, maintaining a supportive tone that prioritizes empathy and encouragement. Safeguards such as a crisis detection module can alert users to seek human assistance in cases where severe distress is detected.

#### 6.4 Ethical Concerns in Emotional Analysis

Automated emotional analysis introduces ethical considerations, especially regarding potential biases in the models. For instance, emotional recognition models trained on specific demographic datasets may not generalize well to diverse populations, potentially leading to inaccurate assessments.

- **Solution:** To address this, SenseAI utilizes diverse, inclusive datasets for model training to minimize biases and enhance the generalizability of emotional assessments across different demographic groups. Regular audits and model evaluations are conducted to detect any unintended biases. Furthermore, the project aligns with established ethical frameworks, such as the IEEE's Ethically Aligned Design guidelines, to ensure responsible development and deployment.

#### 6.5 Model Interpretability

The complex nature of multimodal systems—especially those involving neural networks—can make it challenging to interpret model decisions. This lack of transparency can be problematic in mental health applications, where understanding the reasoning behind certain assessments is valuable for user trust.

- **Solution:** Explainable AI (XAI) techniques can be applied to enhance the interpretability of the model's outputs. For example, attention visualization for transformer models in NLP or heatmaps for CNN-based facial recognition can provide insights into which inputs influenced the emotional prediction. By offering transparent explanations for its assessments, SenseAI builds user confidence and allows for greater accountability.

## 7. Ethical and Practical Considerations

The development of a multimodal psychology chatbot like SenseAI raises critical ethical and practical concerns, especially given the sensitive nature of emotional and biometric data. This section outlines the primary ethical guidelines followed and practical strategies adopted to ensure that SenseAI remains a responsible, trustworthy, and supportive mental health tool.

### 7.1 Privacy and Data Security

Ensuring the privacy and security of user data is fundamental, as SenseAI collects sensitive multimodal data, including text inputs, facial expressions, and biometric signals. Unauthorized access or misuse of this information could lead to privacy violations and erode user trust.

- **Privacy-by-Design:** SenseAI incorporates privacy from the outset by embedding security features directly into the system architecture. Data encryption, secure data storage, and anonymization techniques help protect user information. For instance, all user data is encrypted both in transit and at rest, reducing the risk of unauthorized access.
- **User Control and Consent:** Transparency and user control are prioritized, allowing users to choose which data modalities they wish to share (e.g., opting out of facial expression analysis). Clear consent forms inform users of how their data will be used, processed, and stored, allowing them to make informed choices.
- **Federated Learning for Enhanced Privacy:** To minimize the need for transferring sensitive data, federated learning approaches are explored, where models are trained on-device, and only anonymized updates are shared with central servers. This method significantly reduces data transfer, enhancing privacy without compromising model effectiveness.

### 7.2 Emotional Sensitivity

Chatbots designed for mental health must be particularly cautious in handling user emotions. Misinterpreting or inadequately responding to emotions can result in user distress or harm. Emotional sensitivity is critical, especially when addressing users experiencing negative emotional states.

- **Empathetic Response Design:** SenseAI’s response system is calibrated to prioritize empathy, validation, and encouragement, rather than providing definitive diagnostic statements. The chatbot’s responses are designed to be supportive, reassuring, and reflective of the user's emotional state, ensuring that users feel heard and validated.
- **Crisis Detection and Escalation:** Recognizing that some emotional states may require professional intervention, SenseAI includes a crisis detection module. This module identifies severe emotional distress (e.g., high levels of anxiety or depressive indicators) and provides resources, such as helpline

information or the option to speak with a mental health professional, if available.

### 7.3 Bias and Fairness

Bias in AI systems can lead to inaccuracies, especially if the training data lacks diversity. For a mental health chatbot, biased emotion recognition could result in unintentional misunderstandings, particularly for users from different cultural backgrounds.

- **Inclusive and Diverse Training Data:** SenseAI's emotion recognition models are trained on datasets that represent diverse populations across age, gender, ethnicity, and cultural backgrounds. Regular audits and evaluations are performed to ensure that the chatbot remains fair and accurate for all users, regardless of demographic differences.
- **Continuous Bias Monitoring:** The models are continually monitored and adjusted based on user feedback to identify and address any emerging biases. By integrating explainable AI methods, SenseAI can make model predictions more transparent, allowing developers to understand and mitigate biases effectively.

### 7.4 Ethical Framework Compliance

SenseAI aligns with established ethical frameworks to ensure that its development and deployment adhere to accepted standards in AI ethics, especially those concerning user welfare and accountability.

- **IEEE Ethically Aligned Design and WHO Guidelines:** SenseAI follows ethical guidelines set forth by reputable organizations like the IEEE and the World Health Organization (WHO). These frameworks provide principles for ensuring user welfare, fairness, and accountability in AI systems, particularly within healthcare and mental health applications.
- **Informed Consent and Transparency:** Ethical AI deployment requires that users are fully informed about the data collected, how it will be used, and their rights to access or delete their data. SenseAI ensures that all users are aware of these aspects and provides them with access to their personal data history, ensuring transparency and fostering trust.

## 8. Related Articles

Research in multimodal emotion recognition and AI-driven mental health support has expanded in recent years. This section reviews notable studies and advancements that provide a foundation for SenseAI's approach to multimodal emotional analysis.

- **Multimodal Emotion Recognition Based on Facial Expressions, Speech, and EEG**
  Wang et al. (2022) propose a deep learning-based multimodal emotion recognition system called "Deep-Emotion," which combines facial expressions, voice, and electroencephalogram (EEG) data for high accuracy in emotion recognition. This study highlights the potential of integrating physiological signals for robust emotional analysis, which informs SenseAI's approach in utilizing biometric data alongside text and visual inputs to capture user emotions comprehensively.
- **A Survey of Deep Learning-Based Multimodal Emotion Recognition: Speech, Text, and Face**
  Lian et al. (2023) examine various emotion recognition techniques across modalities like text, audio, and facial expressions. This survey categorizes deep learning models based on their performance in emotion recognition and offers a comparison of popular datasets. The findings emphasize the importance of choosing appropriate datasets and techniques for each modality, a principle applied in SenseAI's model selection and data fusion methods.
- **Multimodal Emotion Recognition Based on Facial Expressions, Speech, and Body Gestures**
  Yan et al. (2024) explore a three-modality system for emotion recognition, integrating facial expressions, speech, and body gestures to enhance recognition accuracy. This study demonstrates the benefits of analyzing non-verbal cues like gestures in emotional detection, underscoring the value of multimodal fusion. Although SenseAI does not use body gestures, this study supports the efficacy of combining visual and auditory data to achieve a more nuanced emotional understanding.
- **Depression Detection Using Multimodal Analysis with Chatbot Support**
  Sharma et al. (2023) investigate an AI system for early depression detection through multimodal analysis, using text, audio, and image data. This approach aligns closely with SenseAI's objectives, as both systems aim to provide real-time emotional support using AI-driven insights. The study emphasizes the role of multimodal data in detecting nuanced emotional states, reinforcing SenseAI's focus on combining text, facial expressions, and biometric data for accurate emotional tracking.

# 9. Similar Projects

Several existing projects have explored the use of AI-driven chatbots in mental health support, each with unique features and limitations. Reviewing these projects provides context for SenseAI's unique approach to multimodal integration.

### 9.1 Only Text-Based Chatbots

Most psychology-focused chatbots currently rely solely on text input to gauge user emotions and provide support. These include:

- **Woebot**: Uses Cognitive Behavioral Therapy (CBT) techniques to help users manage emotions. It focuses on emotional support but lacks multimodal analysis capabilities, relying only on text-based interactions.
- **Wysa**: Offers tools for stress and anxiety management, including CBT and mindfulness exercises. The premium version provides access to therapists, though the chatbot's emotional understanding is limited to text.
- **Youper**: Combines mood tracking with CBT techniques to enhance emotional awareness. It directs users to therapists when needed, although its capabilities are restricted to analyzing textual inputs.
- **Headspace**: Primarily focused on stress and anxiety management, allowing users to connect with mental health professionals. While effective for guidance, it does not leverage multimodal data to improve emotional understanding.

### 9.2 Ellie (SimSensei)

Developed at the University of Southern California and supported by DARPA, **Ellie** is an AI-driven system designed to assess mental health, primarily for military veterans. It combines facial expression analysis and voice tone detection, offering a closer parallel to SenseAI's multimodal approach. Ellie leverages tools like Stanford NLP, OpenFace, and OpenCV to analyze user emotions, with techniques like Pitch Tracking and Vocal Timbre Analysis to improve emotional accuracy. While Ellie focuses on veterans' mental health assessment, its multimodal architecture validates the use of both facial and vocal data in emotion recognition.

### 9.3 Replika AI

**Replika AI** provides a virtual companion for users, allowing communication through text and voice, though it does not utilize image or biometric data for emotion recognition. While Replika aims to fulfill companionship needs, it does not offer advanced emotional support features based on real-time multimodal inputs. Replika's focus on companionship rather than psychological support sets it apart from SenseAI's goal of delivering empathetic and clinically supportive interactions.

## 10. Conclusion

The development of a multimodal, AI-driven psychology chatbot such as SenseAI represents a significant step forward in digital mental health support. By integrating advanced technologies in Natural Language Processing (NLP), computer vision, and biometric data analysis, SenseAI aims to provide users with personalized, context-aware emotional assistance that goes beyond traditional, text-only chatbots. Leveraging multimodal data fusion enables SenseAI to interpret users' emotional states with greater depth and accuracy, enhancing the empathy and relevance of its responses.

This literature review has explored the key technologies, datasets, fusion techniques, and ethical considerations essential for creating a chatbot capable of real-time emotional tracking. The use of transformer models for sentiment analysis, CNNs for facial expression recognition, and machine learning for biometric analysis provides a robust framework that addresses the complexities of emotional detection across multiple data sources. Additionally, the ethical guidelines and privacy measures adopted by SenseAI ensure that user data is handled with utmost care, fostering trust and transparency in a highly sensitive domain.

SenseAI's approach stands out in its commitment to creating a truly empathetic and adaptive mental health tool. By focusing on both technological innovation and ethical integrity, SenseAI aims to set a new standard in AI-driven emotional support, providing users with a safe, responsive, and supportive virtual companion that can positively impact mental well-being.

## 11. References


1. Zeng, Z., et al. (2020). "Facial expression recognition: A survey." *Image and Vision Computing*.
2. Nassir, N., et al. (2020). "Real-time stress detection using multimodal data." *IEEE Transactions on Affective Computing*.
3. Kushal, T., et al. (2021). "Therapeutic chatbots in mental health care." *Psychiatric Services*.
4. Nguyen, T., et al. (2021). "Ensuring privacy and security in multimodal health data systems." *International Journal of Cybersecurity*.

5. Li, S., et al. (2022). "Advancements in multimodal data fusion techniques for emotion recognition." *Pattern Recognition*.
6. González, F., et al. (2021). "Ethical implications of AI in mental health applications." *AI & Ethics*.
7. Wang, S., Zhang, Z., Chen, B., Zhang, Z., Fang, W., and Pan, J. (2022). "Multimodal Emotion Recognition Based on Facial Expressions, Speech, and EEG." *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 142-154. doi: 10.1109/TAFFC.2021.10026861.
8. Lian, H., Lu, C., Li, S., Zhao, Y., Tang, C., and Zong, Y. (2023). "A Survey of Deep Learning-Based Multimodal Emotion Recognition: Speech, Text, and Face." *Entropy*, vol. 25, no. 10, pp. 1440. doi: 10.3390/e25101440.
9. Yan, J., Li, P., Du, C., Zhu, K., Zhou, X., Liu, Y., and Wei, J. (2024). "Multimodal Emotion Recognition Based on Facial Expressions, Speech, and Body Gestures." *Electronics*, vol. 13, no. 18, p. 3756. doi: 10.3390/electronics13183756.
10. Sharma, A., Saxena, A., Kumar, A., and Singh, D. (2023). "Depression Detection Using Multimodal Analysis with Chatbot Support." *2023 IEEE International Conference on Computational Intelligence and Data Science (ICCIDS)*, Chennai, India, pp. 1-6. doi: 10.1109/ICCIDS56901.2023.10489080.
11. Hussain, M., et al. (2021). "Fusion of multimodal data for emotion recognition: A comprehensive review." *Journal of Artificial Intelligence*.
12. Woebot Health. "Woebot Health - Mental Health for the Digital Age." Available: https://woebothealth.com/. [Accessed: 24-November-2024].
13. Wysa. "Reports." Available: https://www.wysa.com/reports#6. [Accessed: 24-November-2024].
14. Youper. "AI Therapy." Available: https://www.youper.ai/. [Accessed: 24-November-2024].
15. Headspace for Work. "Headspace for Work." Available: https://organizations.headspace.com/. [Accessed: 24-November-2024].
16. 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-November-2024].
17. Whatscreativeluc. "Hi, I'm Ellie." Available: https://whatscreativeluc.blogspot.com/2018/11/hi-im-ellie.html. [Accessed: 24-November-2024].
18. Replika. "Replika - My AI Friend." Available: https://replika.ai/. [Accessed: 24-Oct-2024].
19. Chen, J., et al. (2020). "A review of chatbot applications in mental health." *Journal of Medical Internet Research*, 22(11)
. doi: 10.2196/20293.

20. Aldarwish, M., & Ahmad, H. (2021). "Emotional recognition using multimodal data fusion: Applications in mental health monitoring." *Computers in Biology and Medicine*, 132:104296. doi: 10.1016/j.compbiomed.2021.104296.
21. Hassan, A., et al. (2022). "Advances in facial emotion recognition systems for affective computing." *IEEE Access*, 10:99836-99849. doi: 10.1109/ACCESS.2022.3200165.
22. Zhu, Y., et al. (2023). "Multimodal sentiment analysis: A survey and benchmark." *IEEE Transactions on Affective Computing*. doi: 10.1109/TAFFC.2023.3255142.
23. Park, G., & Baek, H. (2024). "Emotion-aware AI for mental health applications: A review of current technologies and ethical challenges." *AI & Society*. doi: 10.1007/s00146-024-01503-4.