Project Proposal

Pookie: An AI-Driven Robot for Promoting Mental Well-being and Emotional Support

Authors: Tibet Buramarn, Kridbhume Chammanard, and Thitaya Divari

Advisor: Dr. Paulo Fernando Rocha Garcia, Ph.D., Assistant Professor of AI and Robotics at Chulalongkorn University and
Ms. Kunpariya Siripanit, Counseling Psychologist at Chula Student Wellness, Chulalongkorn University

Abstract

Pookie, an AI-driven robot for promoting mental well-being, developed under advisory with Chula Student Wellness, aims to create an AI-driven companion to enhance mental well-being by promoting positivity. The robot aims to address ***"Terror Outbursts,"*** a future concern in Thailand involving an anxiety–driven society, where the robot aims to alleviate feelings of stress and anxiety by providing a feeling of slowness and emotional attachment. The robot integrates computer vision, feature extraction, sensors, and actuators to address key customer needs in appearance, interactivity, and empathy. The general appearance of the robot is an anthropomorphic design with resemblance to a bear, with sensors, motors, and screens to deliver personalized interactions to the user.

Acknowledgements

Contents

1   Research Background

This section will provide justifications for the project and necessary knowledge for the reader in order to understand technical terms used throughout the report.

1.1   Justification of the Project

The project focuses on developing emotional well-being robots that promote mental wellness and positivity for users under the influence of stress and anxiety. In general, emotion-related robots are designed to respond to human emotions and can potentially achieve clinical outcomes similar to traditional therapy [1]. Research has shown that digital interventions, such as AI-powered mental well-being robots, can effectively reduce anxiety symptoms and address unmet mental health needs, offering a promising solution to supplement traditional therapeutic approaches [2].

The mental health industry faces significant challenges that cannot be fully addressed through human intervention alone [3]. Key issues include loneliness and social isolation, which are major contributors to depression, anxiety, and overall deterioration in mental health [4], as well as therapeutic challenges, where patients with dementia or other cognitive impairments often struggle with traditional therapeutic activities [5]. In a study by the National Innovation Agency (NIA) based in Thailand, it was identified that the concept of **Terror Outbursts** would become a pressing issue in Thailand by the year 2033 [6]. To elaborate, terror outburst refers to a society driven by constant fear and anxiety. Consequently, traditional methods of addressing anxiety, such as therapy and medication, may not be accessible or appealing to everyone. This creates a significant pain point for individuals seeking immediate, non-invasive support. Our target customer segment includes young adults and professionals aged 18-35 who experience mild to moderate anxiety but may be hesitant to seek conventional treatment, where we provide an innovative alternative to support their mental well-being.

To ensure emotional well-being robots meet user needs and deliver effective support, three key pillars are essential: appearance, interactivity, and empathy. First, the robot's appearance should strike a balance between human-like and machine-like traits, fostering both comfort and trust in users [7]. High interactivity is also crucial; the robot should provide adaptive feedback through various stimuli to engage users effectively and enhance their emotional states [8]. Moreover, a robot's perceived empathic abilities play a significant role in how users interact with it, as these perceptions directly influence their willingness to attribute mental states to the robot, thereby impacting the overall quality of the interaction [9]. On the concept of emotion detection, traditional emotional detection methods utilize verbal and non-verbal cues to accurately detect and respond to human emotions. Verbal cues like pitch variations, volume, and speech rate [10] are critical indicators of emotional states. For example, higher pitch and increased volume often signal heightened emotional arousal, as seen in both American English and Mandarin Chinese, where pitch and speed are essential for expressing emotions. Additionally, contextual understanding—interpreting emotions based on situational cues—further refines the robot's emotional recognition capabilities [11]. Non-verbal cues, such as facial expressions and body language, also play a vital role. For instance, a smile usually denotes happiness, while crossed arms might suggest defensiveness [12]. By integrating these verbal and non-verbal indicators, mental well-being support robots can offer tailored responses, thereby improving the overall effectiveness of their interactions with users.

1.2   Necessary Knowledge

The development of a mental well-being support robot with emotional detection capabilities requires a strong foundation in various advanced concepts within artificial intelligence, machine learning, robotics, and human-computer interaction. Below is an overview of the essential knowledge areas for this project:

Machine learning models are the backbone of emotion detection systems. Convolutional Neural Networks (CNNs) [13] are widely used for tasks such as facial emotion recognition, where they excel at analyzing image data to identify patterns corresponding to different emotional states. One specific architecture, VGGNet, has proven effective for emotion detection due to its deep, layered structure and ability to capture fine-grained facial features. VGGNet's simplicity in design [13], using smaller 3x3 filters stacked in depth, makes it particularly useful for recognizing subtle facial expressions that correspond to emotions. This capability enhances the accuracy of emotion detection, which is crucial for the mental well-being support robot to respond appropriately to a user's emotional state.

In the visual domain, key facial features like eyes, mouth, and eyebrows are extracted and analyzed by CNNs to detect emotions from facial expressions. However, effective emotion recognition often requires the consideration of temporal patterns in sequences of images, such as micro-expressions that unfold over time. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) [14] networks are essential for processing such sequential data. LSTMs, in particular, are highly effective at retaining

information over extended time periods, enabling the robot to identify subtle changes in facial expressions or gestures that might otherwise go unnoticed.

Speech recognition is crucial for enabling the robot to understand and interpret human speech, which is key to detecting emotions from spoken input. Speech recognition techniques allow the robot to process audio data, extracting meaningful insights such as tone, pitch, and speech patterns. These insights help the robot assess the emotional tone and context of the user's communication, making it possible to respond appropriately to their emotional needs. Similar to visual emotion recognition, LSTMs are also invaluable in analyzing sequential audio features, ensuring that variations in tone or pitch over time are captured effectively.

Effective emotion detection relies on extracting meaningful features from raw data. For instance, Mel-Frequency Cepstral Coefficients (MFCCs) [15] are a crucial feature extraction technique in speech emotion recognition, capturing the essential characteristics of the audio signal that correlate with emotional states. Similarly, in the visual domain, CNNs extract and analyze key facial features like eyes, mouth, and eyebrows to detect emotions from facial expressions. The combination of CNNs for spatial analysis and LSTMs for temporal analysis creates a robust framework for identifying emotions from both static and dynamic data.

The design of emotionally intelligent robots requires an understanding of Human-Robot Interaction (HRI) principles. These principles guide the development of robots that can interact naturally and empathetically with humans. Concepts such as user-friendly interface design, adaptive behavior, and empathetic response mechanisms ensure that the robot's interactions are socially acceptable and supportive.

Additionally, emotionally intelligent robots rely on a combination of advanced hardware and software to accurately detect and respond to human emotions. Key hardware components, including cameras, are essential for capturing detailed facial expressions in real-time, allowing systems to effectively analyze emotional states [16]. Microphones and audio sensors play a crucial role in gathering vocal cues, which are vital for emotion detection [17]. Processors and GPUs manage the heavy computational tasks, while actuators and motors control the robot's physical movements, such as gestures and facial expressions, enabling the robot to convey empathy and respond to users effectively. Haptic sensors further enhance this interaction by reacting to touch, contributing to a more interactive and supportive user experience.

Lastly, Bayesian Networks provide a robust framework for decision-making [18], enabling the robot to infer emotional states and choose appropriate responses. These graphical models represent variables and their dependencies through directed acyclic graphs (DAGs). For the robot, observable inputs like facial expressions, vocal cues, and contextual data are nodes, while hidden nodes represent inferred emotional states such as sadness or anxiety.

Bayesian Networks allow the robot to integrate prior knowledge and update beliefs with new information. A belief represents the robot's degree of confidence in a particular state or outcome, based on available evidence and prior knowledge. For example, if vocal cues indicate frustration but facial expressions appear neutral, the network can combine these inputs to infer the true emotional state. This approach assists the robot in making informed decisions and avoiding ambiguity or conflicting signals in data.

2 Objective

2.1 Main Objective Statement

The primary objective of the project's second phase is to continue the design, development, and deployment of an emotional wellness robot capable of recognizing and responding to stress and anxiety symptoms in users aged 18 and older through the integration of AI technologies such as computer vision and speech emotion recognition. In particular, this phase focuses on finishing a minimum viable product that can be appropriately tested on multiple end users in a controlled environment.

2.2 Specific Goals

- Fully 3D print the robotic outer shell following the design concept and specifications accomplished in the first semester.

- Integrate input processing, model serving, and actuator outputs into the board Jetson Nano.

- Develop 15 action bubbles (set of interactions given a certain input) for the robot. In other words, the robot will have 15 different sets of interactions.

- Conduct a 2-week testing period in a live environment to test the robot's effectiveness in identifying and improving positivity and anxiety.

2.3 Measurable Outcomes

- Given the testing sample size, the team hopes to achieve more than a 50% improvement in before–and–after positive well-being and anxiety ratings, indicated by the questionnaire.

- Achieve a sample size of more than 50 recipients.

3 Literature Survey and Review

This section covers the literature survey related to the content defined in the objectives section. It is important to note that this survey encompasses only additional literature that was not initially covered in the first project proposal.

3.1 Related Works: Existing Products and Technologies

1. **ElliQ**

   ElliQ, developed by Intuition Robotics, is an AI-driven social robot designed to address loneliness and promote well-being in older adults. ElliQ is a proactive and conversational companion that facilitates engagement through voice interaction, touch-screen activities, music, video calls, and cognitive games. The robot's primary goal is to reduce social isolation by fostering meaningful interactions and promoting an active lifestyle. ElliQ features a sleek, immobile tabletop design with an expressive lamp-like head that swivels to indicate engagement. Using proprietary AI algorithms, the robot autonomously initiates and personalizes suggestions based on the user's learned behaviors, sentiment analysis, and past interactions. Over time, the AI adapts its interactions to align with the user's preferences and routines, fostering a sense of companionship and trust.

   ElliQ has demonstrated significant potential in reducing loneliness and improving emotional well-being. Studies conducted in collaboration with healthcare organizations, including the New York State Office for the Aging (NYSOFA), revealed that 80% of users reported feeling less lonely with ElliQ, while 74% noted an improvement in their overall quality of life. These findings highlight the effectiveness of social robots as emotional support tools, providing daily engagement, mental stimulation, and social connection. Unlike traditional loneliness interventions that require human facilitators, ElliQ's autonomous nature allows for scalable deployment. By proactively initiating interactions, the robot encourages users to engage in activities that promote mental health, such as guided mindfulness exercises, cognitive challenges, and storytelling. Additionally, ElliQ supports social connection by facilitating video calls with family members, further reinforcing its role as a social catalyst.

   Despite its promising impact, ElliQ faces certain challenges, including user hesitation in accepting robotic companionship, technical support requirements, and limitations in conversational fluidity compared to human interactions. Additionally, further research is needed to evaluate its long-term effects on mental health and well-being. Future iterations of ElliQ may integrate more advanced AI-driven conversational capabilities, greater mobility, and enhanced customization features to further optimize user experience. The success of ElliQ underscores the broader potential of social robots in addressing emotional well-being beyond elderly care. Its application can extend to individuals experiencing social isolation due to disability, remote work environments, or other circumstances where human interaction is limited. As AI technology continues to advance, social robots like ElliQ will play an increasingly vital role in promoting mental wellness and emotional resilience [19].

Figure 1: ElliQ, Source: Adapted from [20]

2. **LOVOT**

LOVOT, developed by Groove X in Japan, is a social robot designed to provide companionship, particularly for older adults experiencing loneliness. Unlike stationary robots, LOVOT features a mobile, pet-like design equipped with AI-driven learning capabilities, allowing it to recognize users, respond to touch, and engage in affectionate interactions. The robot's design incorporates emotional expressiveness, including eye contact, physical warmth, and responsive movement, making it an appealing alternative to traditional social companionship.

A study by Tan et al. [21] examined the impact of LOVOT on single older adults' social well-being. Participants in the study interacted with LOVOT independently in their homes over a week and later shared their experiences in interviews. The study identified four key themes from these interactions: caring for the social robot, finding companionship, forming meaningful connections, and comparing the robot with traditional pets. Users reported that LOVOT provided emotional comfort and reduced feelings of loneliness, reinforcing the idea that social robots can serve as viable companions for older adults who live alone. Additionally, the participants expressed a preference for LOVOT over pets due to its lower maintenance requirements and its ability to provide companionship without the need for feeding or grooming.

LOVOT's adaptive AI enables it to tailor its behavior to individual users, reinforcing a sense of personal connection. This ability to form unique interactions based on user behavior sets LOVOT apart from other social robots. The study further emphasized the importance of designing robots that foster meaningful social engagement rather than serving as mere technological novelties. These findings suggest that social robots like LOVOT could play a crucial role in mitigating loneliness among aging populations and improving overall well-being.

Figure 2: LOVOT, Source: Adapted from [22]

3.2 State of the Art

Bayesian Networks (BNs) are a proven tool for decision-making under uncertainty, as demonstrated by Rothmund et al. [23], where Dynamic Bayesian Networks (DBNs) were applied to enhance the autonomy of industrial drones. The drones utilized DBNs to infer internal faults, assess environmental conditions, and make proactive decisions to avoid failures while executing independent tasks. By integrating information over time and dynamically updating beliefs, the drones optimized task execution while minimizing risks and the consequences of failures. Our project draws upon similar principles to develop a mental wellness robot designed to promote positivity and reduce stress. Although our robot operates with predefined action bubbles, which are structured interactions tailored to various user emotions, Bayesian Networks can play a vital role in determining which action bubble to deploy based on the user's current emotional state.

Similar to the drone's ability to assess environmental conditions, our robot can use a Bayesian Network to infer a user's emotional state from multiple observable inputs such as facial expressions, vocal tone, and speech patterns. For instance, if vocal cues suggest frustration while facial expressions indicate

neutrality, the Bayesian model can combine these observations to probabilistically identify the user's dominant emotional state. Following the decision-making approach in the drone study, the Bayesian Network can evaluate which predefined action bubble, such as a greeting, playful movement, or verbal feedback, would most effectively promote wellness in the user. By assessing probabilistic relationships between input signals and predefined user emotions, the system ensures that interactions feel relevant and positive.

Emotional indicators are often incomplete or conflicting, such as vocal tone indicating stress while facial expressions suggest calmness. The Bayesian framework excels in such scenarios by integrating prior knowledge and real-time evidence to make confident decisions, ensuring that the robot's interactions remain meaningful and appropriate. Rothmund et al. [23] emphasized the importance of minimizing risks in decision-making. Similarly, our robot uses Bayesian methods to weigh the likelihood of success for various action bubbles. For example, if the evidence suggests high uncertainty in emotional detection, the robot can select neutral or universally positive interactions to avoid a mismatch between the user's needs and the robot's response. The inclusion of Bayesian Networks ensures that the mental wellness robot adapts dynamically to user states, optimizing its responses to enhance emotional support and well-being.

4 Project Concept Development

4.1 Acknowledgement

The project concept was developed under consultation with Ms. Kunpariya Siripanit, Counseling Psychologist at Chula Student Wellness (CUSW), Chulalongkorn University.

4.2 Project Overview

Pookie, the emotional well-being robot, is conceptualized as a responsive, AI-driven companion designed to improve mental well-being, specifically catered to customers under stress and anxiety. In this project, **anxiety** is indicated by signs of stress or worry, and **emotional well-being** is defined as *"positivity promotion."*

The project focuses on **promotion**, meaning the promotion of positive well-being through the use of robotics, rather than **prevention**, which refers to the goal of preventing long-term issues such as depression or suicide.

The robot is envisioned to satisfy three key pillars of customer needs:

- **Appearance**

- **Interactivity**

- **Empathy**

These will be achieved through the intuitive integration of computer vision, feature extraction, sensors, and actuators.

4.3 Limitations and Scope

This project aims to develop a proof of concept for an emotional well-being robot. However, several limitations and scope considerations apply:

1. **Security:**

   - The focus is on creating a prototype that demonstrates the feasibility of an emotional well-being robot.
   - Security measures will be basic, and advanced features like data encryption and user authentication are beyond the project's scope.

2. **Safety:**

   - Fundamental safety, including electronics, heat output, and physical design, will be ensured through rigorous testing.
   - Detailed safety protocols, such as long-term durability and fail-safes for unforeseen hazards, are outside the scope of this prototype phase.

3. **Functionality:**

   - The robot will provide core emotional well-being functionalities such as basic interaction and mood assessment.
   - Advanced features, like personalized therapeutic interventions or integration with external health systems, will not be included in this prototype.

4. **User Experience:**

   - The prototype will offer a foundational user experience, but may lack the polish and customization of fully developed models.
   - User interface enhancements will be addressed in future, scaled development phases.

5. **Scalability:**

   - The project will not focus on scalability for mass production or widespread deployment.
   - The prototype demonstrates initial concepts and feasibility, not full-scale implementation.

6. **Integration:**

- Extensive integration with other technologies or platforms is not within the project's focus.
- Emphasis will be on the standalone capabilities of the robot, with minimal focus on interoperability with existing systems.

By acknowledging these limitations and scope considerations, this project sets clear expectations for its initial development phase. Future iterations may address these areas in greater detail based on feedback and further research.

5 Project Planning and Timeline

The overall project will span a total of two academic semesters of the senior year (a total of approximately 8 months) and will comprise a set of goals for each semester. The project will be managed using an agile methodology, where by the end of the project, two deliverables will be obtained: MVP1 and MVP2. This section will break down the project planning and timeline for only MVP2, as well as expected deliverables for each phase.

5.1 Channels

Throughout the project, two essential tools will be used to facilitate communication and task delegation within the project. The first tool is Discord, a multi-functional communication tool that is practical for meetings, scheduling events, and so on. Discord will be used as the primary communication tool for the members in the project, as well as for some advisors. The second tool is Jira, an agile project management tool that facilitates task delegation and software project management. Jira will be used to track the tasks of each member in the project, as well as to track software features and bugs within the project in the form of tickets for ease of audit. Additionally, it will also comprise the customer journey of each feature of the robot in the form of "user stories."

5.2 MVP 2 - Non-Commercial Prototype

MVP 2 will span the entirety of academic semester 2 (from January until April) and will focus on implementing a non-commercial prototype. To elaborate, the team expects a prototype that is viable for testing, but does not consider factors such as security, quality testing, etc to a deep level. MVP 2 will consist of four sprints, each focusing on distinct milestones. Sprint 0, spanning Weeks 1–3, will focus on initial planning, including project scope definition, task delegation, parts acquisition, and 3D-printed CAD designs for components such as the head, arms, base, and outer shell. Experimentation with the Jetson Nano will include configuring the LCD screen, integrating code, and setting up motor interfacing. Preliminary research on Bayesian networks will also be conducted, culminating in a raw prototype. Sprint 1, lasting from Week 4 to Week 7, will focus on assembling a functional prototype and implementing at least four of the planned 15 action bubbles, including greetings, subtle robot movements, positive reinforcement for happiness, and stress-response actions. Debugging and fine-tuning will also take place during this phase. Sprint 2, spanning Weeks 9–12, will prioritize completing all 15 action bubbles, engaging with potential end-user testing candidates, and preparing for iterative testing. Sprint 3, covering Weeks 13–16, will emphasize team testing, quality assurance, and two rounds of end-user testing in Weeks 14 and 15, followed by preparation for the final presentation. By the end of MVP 2, the project will deliver an integrated prototype that successfully combines hardware and software, validated through comprehensive team and end-user testing, establishing a strong foundation for future development.
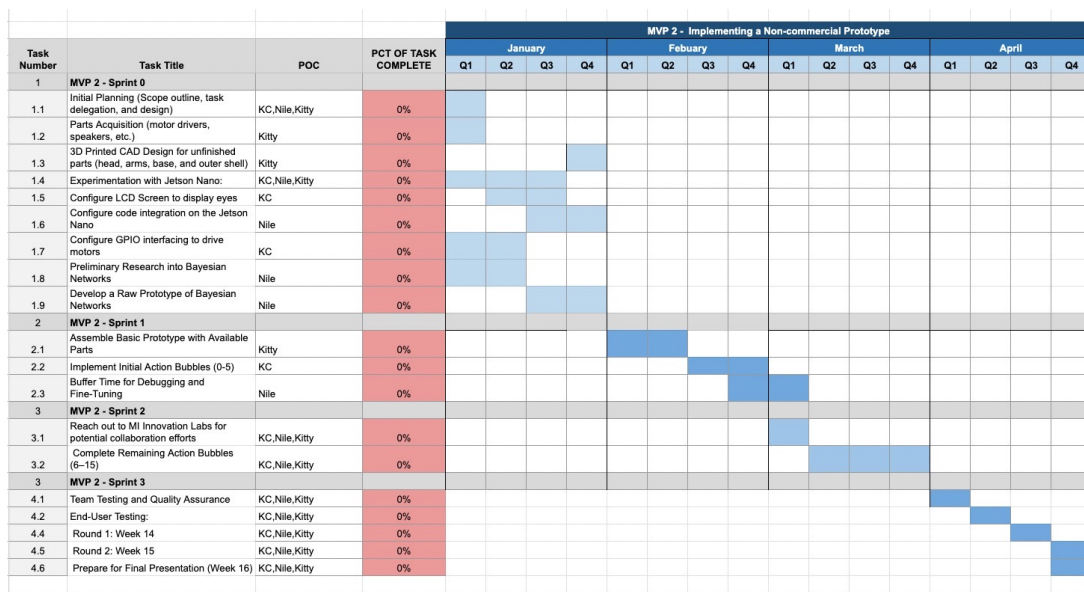

Figure 3: Project GANTT Chart for MVP 2

1. **MVP 1 - Sprint 0 (Weeks 1–3)**

    1.1 Initial Planning (Scope outline, task delegation, and design)

    1.2 Parts Acquisition (motor drivers, speakers, etc.)

    1.3 3D Printed CAD Design for unfinished parts (head, arms, base, and outer shell)

    1.4 Experimentation with Jetson Nano:

    1.4.1 Configure LCD Screen to display eyes

    1.4.2 Configure code integration on the Jetson Nano

    1.4.3 Configure GPIO interfacing to drive motors

    1.5 Preliminary Research into Bayesian Networks

    1.6 Develop a Raw Prototype of Bayesian Networks

2. **MVP 1 - Sprint 1 (Weeks 3–7)**

    2.1 Assemble Basic Prototype with Available Parts

    2.2 Implement Initial Action Bubbles (0-5)

    2.3 Buffer Time for Debugging and Fine-Tuning

3. **MVP 1 - Sprint 2 (Weeks 9–12)**

    3.1 Reach out to MI Innovation Labs for potential collaboration efforts

    3.2 Complete Remaining Action Bubbles (6–15)

4. **MVP 1 - Sprint 3 (Weeks 13–16)**

    4.1 Team Testing and Quality Assurance

    4.2 End-User Testing:

    4.2.1 Round 1: Week 14

    4.2.2 Round 2: Week 15

    4.3 Prepare for Final Presentation (Week 16)

5.3 Current Progress and Challenges

In terms of current progress, the robot is still in its early development. Throughout the last semester, the team focused on project concept development and software architecture, where the hardware components were only considered nearing the end of the semester. As such, the main challenge is configuring the hardware components to work correctly, where the main concern currently is the LCD screen that represents the eyes. Currently, the eyes of the robot used a Python library called OpenCV, which is an image processing library allowing the drawing of various shapes. The eyes were hard coded as various shapes and patterns drawn on a black canvas, as shown in Figure 4. The main challenge with this approach was the scalability and implementation time. Since each frame is hard coded, it does not allow for much flexibility and complex shapes. However, the upside is that it can easily be built in as a class into the main code.
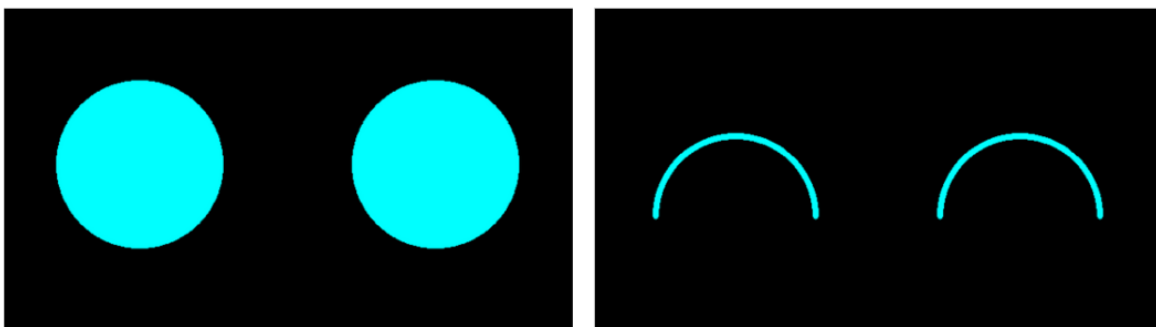

Figure 4: OpenCV Eyes Example for "Happy"

The team has not yet finalized the approach to displaying the eyes on the screen, deciding between OpenCV shapes like in semester 1, or to use a more conventional method like using game frames. Additionally, the system has not been migrated or tested on the Jetson Nano, which will be the main processor for the project. As such, unprecedented challenges such as incompatibility may be faced later. In response to this, the team has allocated buffer time to each sprint to ensure a feasible outcome.

Another key challenge is defining the "action bubbles". These action bubbles indicate the set of all predefined interactions the robot may have with the user. For instance, an action bubble may include the robot's eyes changing color and shape to resemble a wink, one of its arms raised up, and its head tilting a little bit, to form a "greetings!" action bubble. The team must not only implement these, but also define the objectives of these actions so that they have an actual positive impact on the user. These actions will be finalized after a feedback loop in the testing phase.

Additionally, significant progress has been made on the software side, particularly with the development of the FER (facial emotion recognition) and SER (speech emotion recognition) systems. These systems are integral to the robot's ability to understand and respond to the user's emotional state. More details on the methodologies, implementation, and technical challenges of these systems will be addressed in the next section, *Theory and Technical Backup.*

6 Theory and Technical Backup

6.1 Hardware Features

The robot's physical design is primarily anthropomorphic, incorporating elements inspired by animals as well. It stands approximately 12 inches tall and features various integrated hardware components. Starting from the top, the robot's head will be a 3D-printed sphere with LED display eyes, which will serve as the primary form of interaction with the user. Additionally, the head will house a camera and microphone to capture images and sounds for processing by the microcontroller. The robot's body will consist of a large chassis designed to hold the servo motors and the PWM servo motor, which powers the movement of the arms, head, and base. A key feature of the design is the use of bevel gears arranged perpendicularly, enabling a compact structure while ensuring efficient power transmission. The robot will have a total of 4 degrees of freedom, incorporating mainly revolute joints driven by servo motors, which equally transmit power to the bevel gears, driving the base, arms, and head. This design is similar to many desktop companion robots aimed at promoting mental wellness, such as Kiki or Eilik, with the primary distinction being the integration of various emotion detection methods. Throughout the first semester, the team iterated through many designs, and the final design was established as a bear-like robot shown in Figure 5.
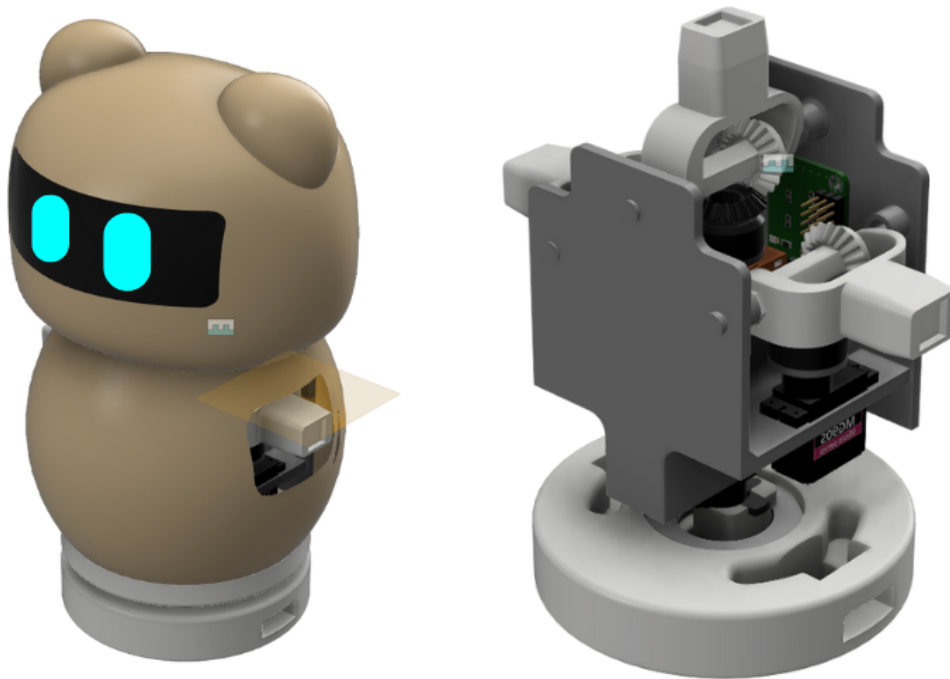

Figure 5: Pookie's Mockup

In terms of high level system design, the system receives input from two sources: audio input and video input. The Jetson Nano processes these inputs and provides output through the speaker. The LCD output displays Pookie's eyes. Additionally, the Jetson Nano controls the motor driver, enabling the movement of Pookie's arm, base, and head mechanisms. The system also communicates with remote devices. Figure 6 illustrates a high level diagram of all the components.
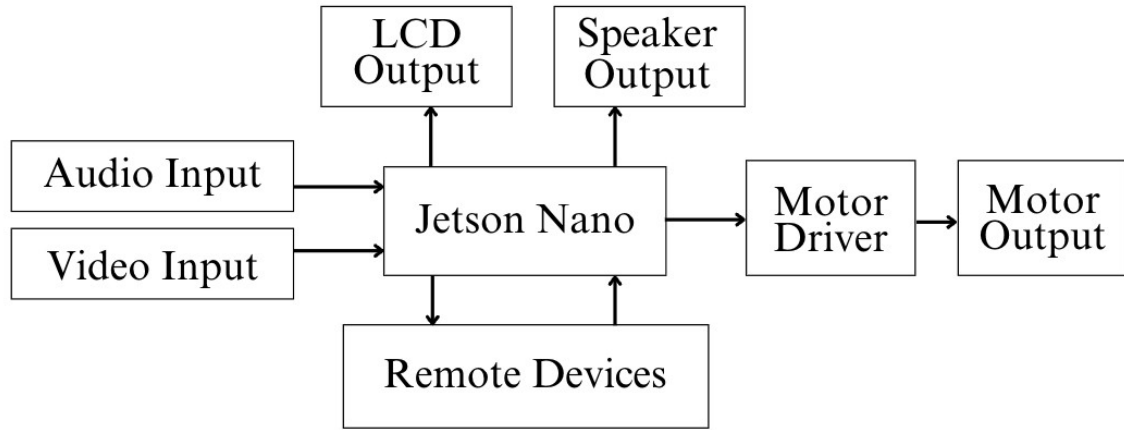
Figure 6: High Level Architecture of the System

6.2 Software Features - Overview

The robot comprises two main features: emotion detection and interaction. Emotion detection is an initiative to incorporate empathy for the customer experience with the robot, using computer vision to analyze facial expressions, as well as speech emotion recognition to analyze tone and pitch. Given predicted emotional status, the robot will be programmed to provide interaction in two forms: verbal and non-verbal. Verbal interactions consist of noises made by the robot, whereas non-verbal interactions comprise physical actions from the robot such as arm movement or changes in the LED display resembling its eyes. Figure 7 illustrates the software architecture of the robot.
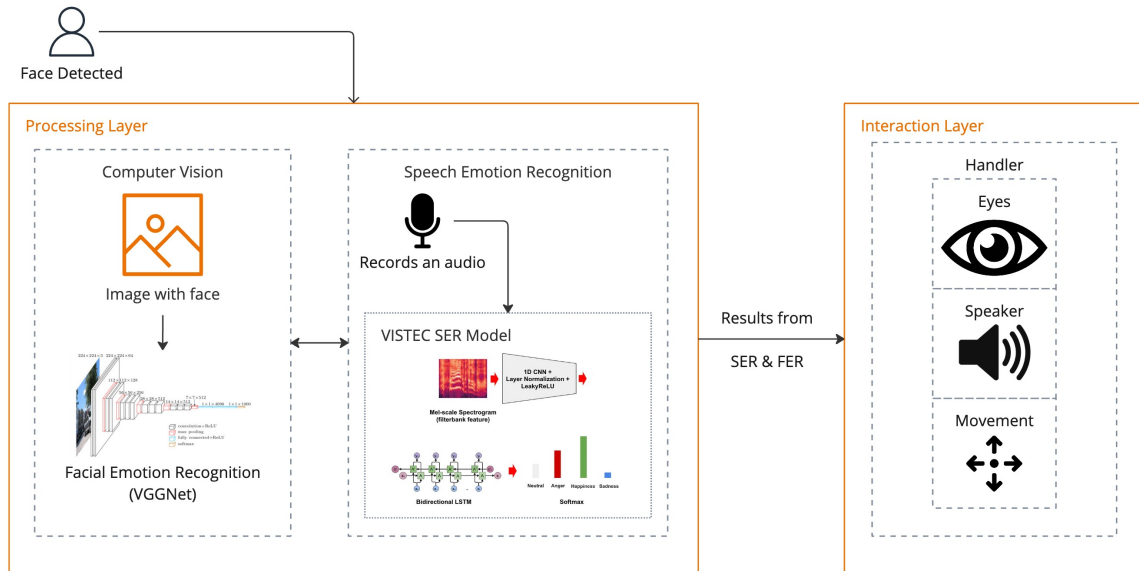

Figure 7: Pookie Software Architecture

6.3 Facial Expression Recognition

Facial expression recognition is a core technique in emotion detection systems, crucial for understanding non-verbal emotional cues in humans. Recent advancements have centered on using Convolutional Neural Networks (CNNs) to detect and classify facial emotions with high accuracy. CNNs are particularly effective because they learn spatial hierarchies of features, enabling them to detect subtle changes in facial expressions, even in complex or dynamic environments. This capability makes CNNs highly suitable for emotion detection tasks, especially in applications requiring real-time emotion tracking, such as emotionally responsive robots.

One example in emotion classification is the integration of CNN-LSTM architectures [24], which improves how these models handle both the details in individual images and changes over time in a sequence of images. This is especially relevant in our project, as the testing dataset comprises Thai ethnicity, requiring robust generalization across unique facial features.

In this project, the emotion classification model maps facial expressions into 7 universal emotions, which must be further classified as stressed or anxious to fit the scope of the project. Research has shown that fear, anger, and disgust [25] strongly correlate with stress and anxiety. Leveraging this finding, the robot detects these three emotions and, if present, performs specific action bubbles aimed at uplifting the user's mood.

6.4 Thai Speech Emotion Recognition

Speech Emotion Recognition (SER) is an integral component of Pookie's AI system, enabling it to interpret vocal emotional cues in Thai. By analyzing pitch, tone, and intensity, the SER system classifies emotions into five categories: neutral, anger, happiness, sadness, and frustration. SER is particularly effective because these vocal features are proven markers of emotional states, allowing accurate classification across a variety of speech patterns.

The model leverages a Thai-language dataset developed by Chulalongkorn University in collaboration with VISTEC, DEPA, and AIS, which contains 41 hours and 36 minutes of labeled audio recordings. Using this dataset, the SER model was developed by VISTEC and demonstrates a weighted accuracy of 66.12% and an unweighted accuracy of 65.67%.

The SER model performs best in recognizing neutral (0.72), anger (0.73), and frustration (0.61), while sadness (0.6) and happiness (0.62) exhibit slightly lower accuracies. However, notable misclassifications include frustration being confused with sadness (0.31) and happiness being confused with frustration (0.17). These misclassifications indicate areas for potential improvement, especially in distinguishing between similar emotions.
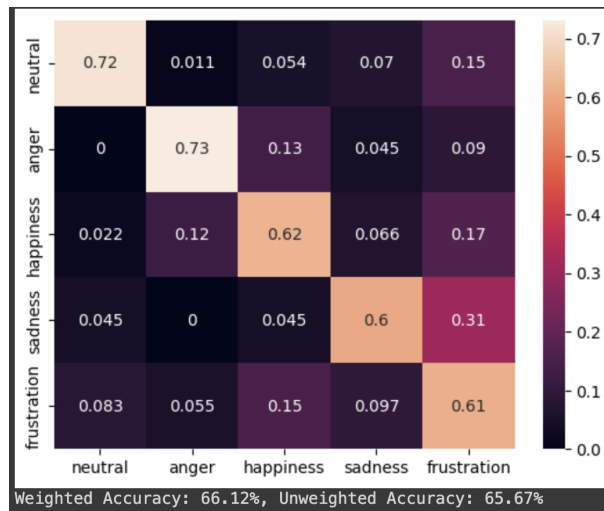

Figure 8: SER Confusion Matrix

6.5 Testing

This section outlines the revised methodology for end-user testing, focusing on assessing the robot's ability to understand human emotions and deliver responses that positively impact the user's mental state, particularly in reducing anxiety and promoting positivity. Since internal testing (e.g., model accuracy,

integration testing) was thoroughly conducted last semester, this semester will emphasize external testing with real users.

Initially, the plan for end-user testing involved two participants, each interacting with the robot, Pookie, over the course of seven days. This approach was recommended by our psychology advisor, who believed that seven days would allow users to form an emotional attachment to the robot, which is essential for improving positivity in daily life. However, after consulting with our engineering advisor, Dr. Paulo, we were advised that testing with only two users would not provide statistically significant results. Consequently, the team decided to prioritize gathering more data points and deprioritize the focus on user attachment.

The final testing methodology will take place in a controlled environment: Pookie will be placed in the MI Innovation Labs, located in the 100th Year Engineering Building, for a period of 1–2 weeks. During this time, students will have the opportunity to interact with the robot and provide feedback. The MI Labs were chosen because they attract a high concentration of our target demographic (users aged 18 and above) and offer a safe, enclosed environment for testing. A team member will supervise the robot at all times, administering pre- and post-interaction surveys to gather data on changes in the users' anxiety and positivity levels.

Although this approach provides a larger number of data points and allows for more robust data analysis, it does not fully replicate the intended user experience of incorporating Pookie into daily life, as originally envisioned. Given the time constraints, the team has opted for a method that prioritizes data collection over long-term user attachment.

One limitation of this method is the location itself: the MI Innovation Labs are frequented mainly by engineering students, which may skew the results. However, as our target demographic is users aged 18 and above with mild to moderate anxiety, engineering students can still serve as a reasonably representative sample for this segment.

The interactions will be brief, lasting approximately five minutes per user, excluding the time needed to explain the project. The specific parameters of the experiment are outlined in Table 1.

Table 1: Experiment Parameters

Parameter	Details
Format	Quantitative Questionnaire; Total 10 Questions
Environment	MI Innovation Labs, 100th Year Building, M Floor
Experiment Time	Approximately 5 minutes per person
HTarget Segment	Students aged 18 or more

As previously mentioned, users will be asked a series of questions before and after interacting with the robot, including some preliminary ones. For example, questions like "What is your occupation?" will help determine whether the robot has a more significant impact on specific user segments. Additionally, the numerical scores assigned will have specific interpretations, which we will later discuss with the psychology advisor (e.g., an anxiety score of 8 might indicate that the individual begins to struggle with tasks when anxious).

Preliminary Questions

- **Consensual:** Could you spare 5 minutes of your time for our experiment? We will collect data from you in the form of questionnaires, and we will only use your face and voice data temporarily for processing, where it will be completely erased afterwards, do you consent?

- **Demographic:** What is your age? What is your occupation? If you are a student, from which faculty are you studying?

Questions asked before interaction

- **Psychographic:** On a scale from 1-10, how would you rate your everyday positivity?

- **Psychographic:** On a scale from 1-10, how frequently do you feel positive/happy?

- **Psychographic:** On a scale from 1-10, how would you rate your daily anxiety levels?

- **Psychographic:** On a scale from 1-10, how frequently do you feel anxious?

7 Project Outcome

Over the course of the project, the team has had several deliverables. During the first semester, we successfully completed seven key deliverables: the project proposal, first and second progress reports, final report draft, final presentation, final report, and a live demo of MVP1. As outlined in the project plan, MVP1 served as a proof of concept, showcasing the basic emotion detection algorithms and their basic integration with the robot's hardware components. Now, in the second semester and at the stage of MVP2, our focus has shifted to integration and expanding the robot's functionality. While the exact number of report deliverables is still being determined, we have already made significant progress toward developing a functional, testable prototype. The team is now engaged in extensive testing with our target customer segment to gather feedback that will guide further development and scale the robot for potential commercial use. However, it is important to note that the project's final outcome will not be a commercially deployable product at this stage. To summarize the project's overarching objectives, the robot must fulfill three core pillars of customer expectations for emotional wellness robots: appearance, interactivity, and empathy. By the end of the project, we will have accomplished the following:

- Created an anthropomorphic outer shell that resonates with our target market, seamlessly integrating the necessary electronic components.

- Implemented interactive features under the expert guidance of Chula Student Wellness (CUSW).

- Developed an accurate emotion detection algorithm capable of effectively assessing the user's anxiety state.

- Fostered empathetic human-robot interactions that promote emotional wellness.

- Validated our design through market research and user testing feedback.

- Compiled a comprehensive report detailing user feedback from testing.

Quantitatively, by the end of the project, we expect the following outcomes:

- A measurable reduction in anxiety levels as reported by users during testing.

- A measurable improvement in positivity levels as reported by users during testing.

- Achieve a satisfactory sample size for testing.

- Achieving benchmark accuracy across various emotion detection metrics, demonstrating a functional and effective model.

8 Summary and Benefits

8.1 Summary

This project centers on developing an emotional well-being robot designed to support mental health by helping users manage stress and anxiety. Using technologies such as computer vision and natural language processing, the robot can recognize and respond to emotional signals in real time, offering consistent and empathetic assistance. Its design emphasizes appearance, interactivity, and empathy, specifically catering to the needs of Gen Z and younger millennials in promoting mental wellness. The project is structured across two academic semesters, following an agile methodology to produce two major prototypes: MVP1 (Proof of Concept) and MVP2 (Fully Functional Design). The development is supported by insights from Dr. Paulo Fernando Rocha Garcia, Ph.D., Assistant Professor of AI and Robotics at Chulalongkorn University, and Ms. Kunpariya Siripanit, a counseling psychologist at Chulalongkorn University, ensuring that the robot meets both technical and mental health standards. Ultimately, this project provides an innovative, scalable solution that addresses the growing challenge of anxiety disorders, positioning the robot as a sustainable alternative to traditional mental health support.

8.2 Benefits

- **Direct Industry Impact:** This project will greatly contribute to the mental health field, particularly addressing the needs of patients dealing with stress and general anxiety—an enormous market segment. Mental well-being robots are a crucial tool in helping manage "Terror Outbursts" in Thailand. By utilizing AI to detect and respond to emotions through facial recognition and voice analysis, these robots can reduce dependence on human intervention in promoting mental positivity. Moreover, they will improve the consistency and quality of mental health support, effectively addressing gaps in current care.

- **Scalability and Long-Term Value:** With global anxiety disorders rising by 55% between 1990 and 2019, affecting an estimated 301 million people worldwide [39], these robots are poised to become increasingly vital. Their ability to provide real-time, personalized support will enhance individual well-being and contribute to the long-term sustainability of mental health care systems. Although the project is currently focused on proof of concept and prototyping, it offers a novel approach to mental well-being, with potential for future scalability and commercialization.

9 Team Roles and Responsibilities

Although each student is assigned a specific component of the project, it is important to recognize that we will collaborate on various aspects. As a result, each role may evolve and shift as the project progresses.

1. **Kridbhume Chammanard - Project Manager**

   As the Project Manager, Kridbhume Chammanard is responsible for overseeing the entire project, ensuring that all activities align with the established goals and deadlines. Kridbhume will delegate tasks to the project engineers, manage the project timeline, and make necessary adjustments to keep the project on track. Engaging with advisors from ISE and Chula Student Wellness is a key aspect of the role, including providing regular updates and consultations. Kridbhume will streamline the work processes for the project engineers to enhance efficiency and productivity, and will also professionally review and approve all deliverables before submission. Additionally, Kridbhume will oversee the overall project budget, ensuring that expenditures are within allocated limits and that resources are used effectively.

2. **Thitaya Divari - Project Engineer, Product Development**

   Thitaya Divari, as the Project Engineer for Product Development, will focus on designing the user experience (UX) of the robot, utilizing CAD software to develop and refine user interfaces. Thitaya will be responsible for sourcing and gathering all necessary hardware components required for the robot's physical construction, as well as working on the assembly and testing of these components to ensure they meet design specifications. Managing the budget allocated for hardware components and product development activities will be another critical responsibility, including tracking expenses and making adjustments as needed. Thitaya will also prepare detailed documentation of the product development process, including design specifications, component lists, and testing results.

3. **Tibet Buramarn - Project Engineer, Software Development**

   In the role of Project Engineer for Software Development, Tibet Buramarn will lead all software development initiatives, including coding, testing, and integrating the robot's software functionalities. Tibet will implement and manage DevOps practices within the project, ensuring smooth development, deployment, and maintenance of the software. A significant aspect of the role includes ensuring seamless integration of software with hardware components, addressing any compatibility issues, and optimizing performance. Tibet will conduct thorough testing and debugging of the software to identify and resolve issues, ensuring the reliability and functionality of the system.

References

[1] C. E. Palmer, E. Marshall, E. Millgate, G. Warren, M. P. Ewbank, E. Cooper, S. Lawes, M. Bouazzaoui, A. Smith, C. Hutchins-Joss, J. Young, M. Margoum, S. Healey, L. Marshall, S. Mehew, R. Cummins, V. Tablan, A. Catarino, A. E. Welchman, and A. D. Blackwell, "Combining ai and human support in mental health: a digital intervention with comparable effectiveness to human-delivered care," *medRxiv*, 2024. [Online]. Available: https://www.medrxiv.org/content/early/2024/07/17/2024.07.17.24310551

[2] M. M, A. Bakkar, A. Hiremath, Amruthaiah, and J. Pranith, "Jarvis the companion," *International Journal of Science and Research (IJSR)*, vol. 13, no. 5, pp. 1672–1677, May 2024. [Online]. Available: https://www.ijsr.net/

[3] J. T. Charles, S. Jannet, and C. S. J, "Mitigating Occupational Mental Health-Related factors to prevent manufacturing industry accidents," *Research Square (Research Square)*, 1 2024. [Online]. Available: https://doi.org/10.21203/rs.3.rs-3771035/v1

[4] K. W. Goh, S. Ishikawa, R. Okubo, I. Kusumi, and T. Tabuchi, "Are patients with mental disorders more vulnerable to loneliness and social isolation during the covid-19 pandemic? – findings from the japan covid-19 and society internet survey," *Journal of Affective Disorders*, vol. 340, pp. 72–79, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165032723010145

[5] P. Sukhawathanakul, A. Crizzle, H. Tuokko, G. Naglie, and M. J. Rapoport, "Psychotherapeutic interventions for dementia: a systematic review," *Canadian Geriatrics Journal*, vol. 24, no. 3, p. 222–236, Sep. 2021. [Online]. Available: https://cgjonline.ca/index.php/cgj/article/view/447

[6] National Innovation Agency [NIA], *Futures of Mental Health in Thailand in 2033.* NIA Thailand, 2023. [Online]. Available: https://www.nia.or.th/bookshelf/view/237

[7] M. De Cet, M. Cvajner, I. Torre, and M. Obaid, "Do your expectations match? a mixed-methods study on the association between a robot's voice and appearance," in *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, ser. CUI '24. New York, NY, USA: Association for Computing Machinery, 2024. [Online]. Available: https://doi.org/10.1145/3640794.3665551

[8] J. Wang, "Enhancing user experience using a framework integrating emotion recognition and eye-tracking," *Highlights in Science, Engineering and Technology*, vol. 85, p. 298–308, Mar. 2024. [Online]. Available: https://drpress.org/ojs/index.php/HSET/article/view/18387

[9] A. Lillo, A. Saracco, E. Siletto, C. Mattutino, and C. Gena, "Investigating the relationship between empathy and attribution of mental states to robots," 2024. [Online]. Available: https://arxiv.org/abs/2405.01019

[10] T. Hakanpää, T. Waaramaa, and A.-M. Laukkanen, "Comparing contemporary commercial and classical styles: Emotion expression in singing," *Journal of Voice*, vol. 35, no. 4, pp. 570–580, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0892199719302097

[11] R. Abbas, B. Ni, R. Ma, T. Li, Y. Lu, and X. Li, "Context-based emotion recognition: A survey," *Neurocomputing*, 2024, preprint submitted to Elsevier, not peer-reviewed. [Online]. Available: https://ssrn.com/abstract=4657124

[12] H. Liu, "Emotion detection through body gesture and face," 2024. [Online]. Available: https://arxiv.org/abs/2407.09913

[13] M. M. Taye, "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions," *Computation*, vol. 11, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2079-3197/11/3/52

[14] R. M. Schmidt, "Recurrent neural networks: A gentle introduction and overview," *arXiv preprint arXiv:1912.05911*, 2019. [Online]. Available: https://doi.org/10.48550/arXiv.1912.05911

[15] P. P. Singh and P. Rani, "An approach to extract feature using mfcc," *IOSR Journal of Engineering*, vol. 4, no. 8, pp. 21–25, 2014. [Online]. Available: https://doi.org/10.9790/3021-04812125

[16] S. Gupta, "Enhancing Mood Based Music Selection through Physiological Sensing Technology," *International Journal for Research in Applied Science and Engineering Technology*, vol. 12, no. 4, pp. 71–74, 4 2024. [Online]. Available: https://doi.org/10.22214/ijraset.2024.59658

[17] C. C. V, G. K. S, and D. K. Y. C, "Emotion detection via voice and speech recognition," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 635–643, 2024.

[18] D. Heckerman, "A tutorial on learning with bayesian networks," *CoRR*, vol. abs/2002.00269, 2020. [Online]. Available: https://arxiv.org/abs/2002.00269

[19] E. Broadbent, K. Loveys, G. Ilan, G. Chen, M. M. Chilukuri, S. G. Boardman, P. M. Doraiswamy, and D. Skuler, "Elliq, an ai-driven social robot to alleviate loneliness: Progress and lessons learned," *Journal of Aging Research & Lifestyle*, vol. 13, pp. 22–28, March 2024. [Online]. Available: https://doi.org/10.14283/jarlife.2024.2

[20] I. Spectrum, "Elliq, a social home robot for older adults, now available for preorder," *IEEE Spectrum*, 2023, accessed: 2025-01-24. [Online]. Available: https://spectrum.ieee.org/elliq-a-social-home-robot-for-older-adults-now-available-for-preorder

[21] C. K. Tan, V. W. Q. Lou, C. Y. M. Cheng, P. C. He, and V. E. J. Khoo, "Improving the social well-being of single older adults using the lovot social robot: Qualitative phenomenological study," *JMIR Human Factors*, vol. 11, p. e56669, 2024. [Online]. Available: https://humanfactors.jmir.org/2024/1/e56669

[22] Robots Guide, "Lovot - robots guide," 2024, accessed: January 24, 2025. [Online]. Available: https://robotsguide.com/robots/lovot

[23] S. V. Rothmund, C. A. Thieme, I. B. Utne, and T. A. Johansen, "A bayesian approach to risk-based autonomy, with applications to contact-based drone inspections," *IEEE Robotics and Automation Letters*, 2021.

[24] E. Ryumina, D. Dresvyanskiy, and A. Karpov, "In search of a robust facial expressions recognition model: A large-scale visual cross-corpus study," *Neurocomputing*, vol. 514, pp. 435–450, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222012656

[25] T. Baltrušaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 59–66. [Online]. Available: https://doi.org/10.1109/FG.2018.00019