



CHULA INTERNATIONAL SCHOOL OF ENGINEERING
The ingenuity of CHULA ENGINEERING

Pookie: An AI-Driven Robot for Promoting Mental Wellbeing 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

2147416 Final Project I
International School of Engineering (ISE)
Chulalongkorn University

December 11, 2024

Abstract

Pookie, an AI-driven robot for promoting mental wellbeing, developed under advisory with Chula Student Wellness, aims to create an AI-driven companion to enhance mental wellbeing 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 features an anthropomorphic form with LED displays and sensors for interaction, drawing inspiration from existing robots like Kiki, with future iterations expected to refine user experience and functionality based on feedback.

Acknowledgements

The team expresses their sincere gratitude to 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. Dr. Paulo's expertise in artificial intelligence and robotics was instrumental in guiding the technical development of the emotional wellness robot, while Ms. Kunpariya's insights into counseling psychology ensured the project's alignment with mental health principles. The team also acknowledges the support provided by the faculty and staff of the International School of Engineering, whose assistance was invaluable throughout this project.

Contents

1 Research Background	4
1.1 Justification of the Project	4
1.2 Necessary Knowledge	4
2 Objective	6
2.1 Main Objective Statement	6
2.2 Specific Goals	6
2.3 Measurable Outcomes	6
2.4 Relevance or Significance	6
3 Literature Review	7
3.1 Related Works: Existing Products and Technologies	7
3.2 State of the Art	8
4 Project Development	10
4.1 Overview	10
4.2 Design Concept	11
4.2.1 Design Reference	11
4.3 Limitations and Scope	11
4.4 Software Implementation	12
4.4.1 Process Flowchart	12
4.4.2 Software Architecture and Integration	12
4.4.3 Key Components	13
4.4.4 Core Integration Framework	14
4.4.5 Component Interaction	14
4.5 Facial Emotion Recognition	14
4.5.1 Faces Dataset	14
4.5.2 Facial Emotion Recognition Model Development	15
4.5.3 Facial Emotion Recognition Model	16
4.6 Speech Emotion Recognition	17
4.6.1 Speech Emotion Recognition Dataset and Model	17
4.7 Software Challenges	17
4.8 Hardware Implementation	18
4.8.1 Initial Design	18
4.8.2 Revised Design	19
5 Testing	23
6 Summary and Real Benefit to Industry	24
6.1 Summary	24
6.2 Benefits Direct Impact on the Industry	24
7 Team Roles and Responsibilities	25
References	26

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 wellbeing 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 wellbeing 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 wellbeing.

To ensure emotional wellbeing 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 [10], and speech rate [11] are critical indicators of emotional states. For example, higher pitch and increased volume often signal heightened emotional arousal [12], 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 [13]. Non-verbal cues, such as facial expressions and body language, also play a vital role. For instance, a smile usually denotes happiness [14], while crossed arms might suggest defensiveness [15]. By integrating these verbal and non-verbal indicators, mental wellbeing 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) [16] 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 [16], 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. Natural Language Processing (NLP) is crucial for enabling the robot to understand and interpret human language, which is key to detecting emotions from

text or spoken input. NLP techniques allow the robot to process language data, extracting meaningful insights such as sentiment and intent. 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.

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 [17]. Microphones and audio sensors play a crucial role in gathering vocal cues, which are vital for emotion detection [18]. 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.

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. Microphones and audio sensors play a crucial role in gathering vocal cues, which are vital for emotion detection. 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.

2 Objective

2.1 Main Objective Statement

The primary objective of this project is to design, develop, and deploy an emotional wellness robot capable of recognizing and responding to stress and anxiety symptoms in users through the integration of AI technologies such as computer vision and natural language processing. The robot must fulfill all three key pillars of customer expectations in emotional wellness robots: design, interactivity, and empathy.

2.2 Specific Goals

- Design an anthropomorphic robotic outer shell that resonates with the target customer segment.
- Design interactive verbal and non-verbal features for the robot, such as reaction when the user smiles, with expert supervision from Chula Student Wellness (CUSW).
- Develop an accurate emotion detection algorithm that captures the user's state of stress through speech emotion recognition and facial expression recognition.
- Develop empathetic human-robot interactions that promote emotional wellness within the customer.
- Conduct extensive testing and refinement based on user feedback.

2.3 Measurable Outcomes

- Perform end user testing in at least 3 users for 7 days each
- Obtain a reduction in self-reported anxiety score after using the robot (e.g., the self-reported anxiety score before using Pookie is 7, but after using Pookie it becomes 6)
- Obtain an increase in self-reported positivity score after using the robot (e.g., the self-reported positivity score before using Pookie is 7, but after using Pookie it becomes 8)
- Achieve a benchmark in accuracy metrics for emotion detection.

2.4 Relevance or Significance

With **Terror Outbursts** being one of the major societal challenges in Thailand, there is a pressing need for accessible positivity promotion. Our robot aims to bridge the gap between traditional therapy sessions by providing immediate support to individuals struggling with anxiety.

3 Literature Review

3.1 Related Works: Existing Products and Technologies

1. PARO

PARO, a robotic baby seal developed by the National Institute of Advanced Industrial Science and Technology (AIST) in Japan, is one of the most well-known emotional support robots. It is designed to provide companionship to elderly individuals, particularly those with dementia or cognitive impairments. PARO incorporates several advanced features to enhance its therapeutic effectiveness: it is equipped with tactile sensors that allow it to respond to touch, microphones to recognize voices and sounds, and artificial intelligence to learn and adapt to the user's preferences over time. It also has movement and behavior patterns that mimic a real seal, such as blinking its eyes, moving its flippers, and making sounds, which helps in creating a comforting and engaging interaction. Studies have shown that PARO can reduce stress and anxiety, foster socialization, and enhance overall mood among users [19].



Figure 1: PARO, Baby Seal Robot. Source: Adapted from [19]

2. Moxie

Moxie is an emotional support robot designed for children, especially those needing help with anxiety or emotional development. Equipped with AI and machine learning, Moxie interacts empathetically to help children build social, emotional, and cognitive skills. It engages in personalized conversations, encourages emotional expression, and uses storytelling, breathing exercises, and interactive games to foster learning. Moxie also tracks progress and adapts to each child's needs, providing daily activities and challenges that promote emotional growth [20].



Figure 2: Moxie, a learning robot with heart. Source: Adapted from [20]

3.2 State of the Art

The development of mental wellbeing support robots with advanced emotional detection capabilities is an interdisciplinary endeavor that requires significant advancements in artificial intelligence (AI), robotics, and human-computer interaction (HCI). Recent research and technological progress in these areas have laid the foundation for creating robots that can not only interact with humans but also understand and respond to their emotional states. This section reviews the state of the art in emotional detection, AI-driven emotional responses, and the integration of these technologies into robotic platforms.

Emotional detection in robots is primarily driven by advancements in machine learning and deep learning techniques, particularly in the fields of computer vision, natural language processing (NLP), and speech recognition.

Facial expression recognition, a critical component, often employs convolutional neural networks (CNNs) to classify emotions based on facial features. Significant progress has been made in this domain with models like OpenFace [21] and VGGFace [22], which offer high accuracy in identifying a wide range of emotions from facial data.

Speech-based emotion detection is another key area, where prosodic features such as tone, pitch, and rhythm are analyzed to infer emotional states [23]. Techniques like Long Short-Term Memory (LSTM) networks and more recently, transformer-based models such as BERT [24], have shown promising results in understanding the emotional content of spoken language. These models can be fine-tuned for specific languages and contexts, making them adaptable to diverse environments.

In natural language processing (NLP), sentiment analysis and emotion classification techniques are used to detect emotions from text. Advanced models like BERT [25] and GPT [26] have revolutionized NLP tasks, enabling the robot to understand not just the semantic content but also the emotional undertone of the conversation. This is particularly useful in chatbots and conversational agents, where understanding the user's emotions can significantly enhance the interaction quality.

Creating emotionally intelligent responses in robots involves the integration of advanced AI techniques that can interpret and respond to human emotions in a meaningful way. The development of these responses relies on recent advancements in natural language processing (NLP) and deep learning models, particularly those designed to handle the complexity of human emotions.

Transformer-based models, such as BERT and DialoGPT, have become central to generating emotionally appropriate responses. BERT, with its bidirectional encoder representations from transformers, has been extensively used in sentiment analysis tasks, enabling the model to understand and generate contextually relevant emotional responses from textual data. Research highlights the effectiveness of BERT models, especially when fine-tuned, in enhancing the accuracy and depth of emotional understanding, which is critical in generating appropriate responses [25].

DialoGPT, a model developed by Microsoft and fine-tuned on empathetic dialogue datasets, further extends these capabilities by producing responses that are not only contextually appropriate but also emotionally resonant. This model's ability to handle nuanced emotional states and produce empathetic responses makes it particularly suited for applications in mental wellbeing support robots [27].

The effectiveness of these AI-driven emotional responses is often evaluated using metrics like perplexity, which measures the coherence of generated responses. Models like the transformer-based approach have demonstrated lower perplexity scores compared to others, indicating their superior ability to generate contextually and emotionally appropriate responses [27]. Continuous optimization and fine-tuning of these models are necessary to improve their performance further, making them more reliable and effective in real-world applications.

Human-Robot Interaction (HRI) is a multidisciplinary field that encompasses various aspects of interaction between humans and robots. This field brings together knowledge from social sciences, cognitive sciences, robotics, engineering, and human-computer interaction to design, evaluate, and understand robotic systems intended for interaction with humans. The core of HRI lies in understanding how humans and robots can interact effectively, whether in a one-on-one scenario or in situations involving multiple humans and robots.

Human-robot collaboration has evolved significantly, with a focus on developing robots that can work alongside humans seamlessly. Advanced techniques such as compliance control, human performance-based approaches, model learning-based methods, and synergy models have been explored to improve collaborative interactions. These techniques allow robots to adapt to human actions, making them more responsive and effective in shared tasks [28].

Brain-Computer Interfaces (BCIs) represent a transformative approach in HRI by enabling direct communication between humans and robots through brain signals. This technology has been applied in various domains, including rehabilitation and communication, where it helps individuals with motor impairments regain control over robotic devices. The integration of BCI in HRI paves the way for more intuitive and efficient interactions, particularly in scenarios requiring real-time decision-making and adaptive responses [28].

Emotional intelligence in robotics is crucial for creating more natural and effective human-robot interactions. By recognizing and responding to human emotions through facial expressions, body gestures, and eye-tracking, robots can engage with humans on a deeper, more empathetic level. This capability is particularly important in applications such as assistive robots, where understanding and reacting to a user's emotional state can significantly enhance the interaction's quality and effectiveness [28].

4 Project Development

This section details the development of Pookie, an AI-driven robot designed to promote mental well-being, developed under design supervision from Chula Student Wellness, a platform for free psychological counseling for Chula students. Pookie aims to act as a companion to help alleviate feelings of stress and anxiety, especially in response to future societal concerns like ***Terror Outbursts***, an anxiety-driven phenomenon anticipated to affect Thailand. The project focuses on enhancing positivity and emotional attachment by creating a robot that interacts with users in an empathetic and calming manner.

4.1 Overview

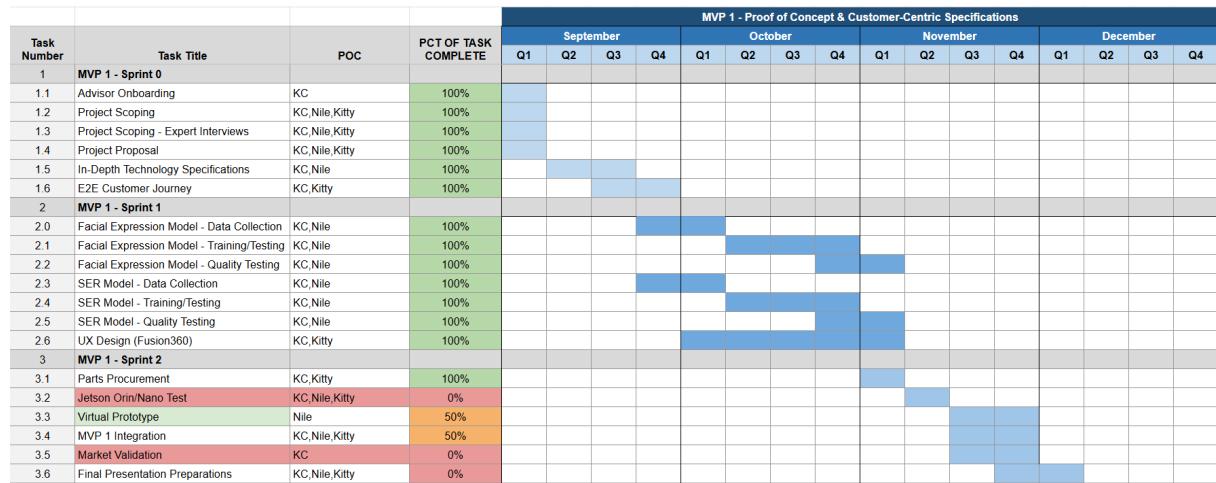


Figure 3: Project GANTT Chart

The project comprises a total of two scopes, one for each semester, where each scope will have its respective expectations. The two scopes are MVP 1 (Proof of Concept) and MVP 2 (Market Prototype). In general, MVP 1 aims to initially collect specifications and design inspiration through primary (expert interviews) and secondary research. Next, a minimum viable product including a working set of AI algorithms, software integration, and CAD mockup was created. On the other hand, MVP 2 aims to elevate the proof of concept to a market prototype, where the software aspects such as AI and integration would be quality tested and dynamic, and the hardware would be readily available for testing.

This report details the implementation of Pookie MVP 1 (Proof of Concept) over the span of four months during the first semester. Overall, the project has taken the correct trajectory since conceptualization. As shown in Figure 3, MVP 1 comprises three sprints (Sprint 0, Sprint 1, Sprint 2). In Sprint 0, the team focused on overall product design, reaching out to Dr. Kunparinya Siripanit from Chula Student Wellness for top-down design specifications to accommodate users of general anxiety. Additionally, the team spent time addressing feasibility aspects of the project and literature review, which was discussed in the project proposal. Next, in Sprint 1, the team split into working on their respective tasks, where the software engineering members researched and trained AI models on emotion recognition, including speech emotion recognition and facial emotion recognition, whereas the hardware engineering team focused on developing a mockup of the robot's physical appearance and acquiring necessary pieces of equipment. Finally, in Sprint 2, the team attempted to integrate all fundamental pieces of software and hardware together. While the project expected integration of software with the Jetson Orin microcomputer, it was recommended by Dr. Paulo instead to focus on developing a virtual prototype, to which a Fusion360 render of the robot's basic interactions and movements were created. Additionally, the task of *market validation* was also canceled, where initially it was planned that a customer survey and customer interview would be collected to obtain further feedback and refine Pookie's concept.

4.2 Design Concept

4.2.1 Design Reference

This section illustrates the overall design concept and references for Pookie. Pookie, the emotional wellbeing robot, is conceptualized as a responsive, AI-driven companion designed to improve mental well-being, specifically catered to customers under the influence of stress and anxiety. To be specific, anxiety in this project is indicated by signs of stress or worry, and the definition of *emotional wellbeing* is *positivity promotion*. It is important to note that the scope of the project, in terms of emotional support, is identified as *promotion*, meaning promotion of positive wellbeing through the use of robotics, rather than *prevention*, which refers to a specific goal of preventing long term issues such as depression or suicide.

With Pookie, it was advised by Dr. Fah Kunpariya (Counseling Psychologist from Chula Student Wellness) to implement an animal-like design with anthropomorphic features. The reason is to reference designs that stimulate a sense of calmness, such as a pet-like design. As such, Pookie resembles a teddy bear, which hopes to stimulate the user with a sense of calmness and childhood nostalgia. After brainstorming and conceptualization, a Fusion360 CAD mockup for Pookie was created. Pookie's outer shell was 3D printed to hide and house a series of motors, microcontrollers and other necessary electronic components as illustrated in 4. Within the scope of this semester, the hardware aspects only comprised a Fusion360 CAD Mockup and virtual prototype. Physical hardware assembly and complete integration was expected to be out of scope, and will begin early next semester in MVP2.

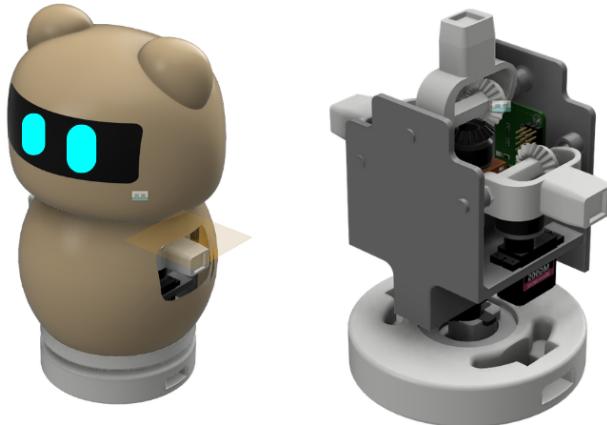


Figure 4: Pookie CAD Mockup. *Left:* Physical Appearance Design, *Right:* Internal Hardware Design.

4.3 Limitations and Scope

This project aims to develop a proof of concept for an emotional well-being robot, drawing inspiration from existing models such as Kiki. However, there are several limitations and scope considerations for this project:

- **Security:** The primary focus of this project is to create a prototype that demonstrates the feasibility of an emotional well-being robot. As such, the security measures implemented will be at a basic level. Comprehensive security features, including data encryption and advanced user authentication, are beyond the scope of this project.
- **Safety:** While the robot will undergo rigorous testing to ensure fundamental safety in terms of electronics, heat output, and physical design, the scope of safety considerations will be limited to these basic aspects. Detailed safety protocols, including long-term durability and fail-safes for unforeseen hazards, will not be extensively addressed in this prototype phase.
- **Functionality:** The robot will focus on core emotional well-being functionalities, such as basic interaction and mood assessment. Advanced features, such as personalizing therapeutic interventions or integration with external health systems, will not be included in this prototype.

- **User Experience:** The prototype will provide a foundational user experience but may lack the polish and customization of fully developed models. User interface and experience enhancements will be considered in future, scaled development phases.
- **Scalability:** The project will not address scalability concerns for mass production or widespread deployment. The prototype is intended to demonstrate initial concepts and feasibility rather than full scale implementation. Integration: This project will not explore extensive integration with other technologies or platforms. The focus will remain on the standalone capabilities of the robot, with minimal emphasis on interoperability with existing systems.

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

4.4 Software Implementation

Software implementation for Pookie comprises two key pillars: 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. For each AI model, the robot utilizes feature labels of the seven universal emotions (happiness, neutral, sadness, fear, anger, disgust, surprise), which are used as a baseline for many emotion classification tasks, including stress and anxiety detection. In particular, stress and anxiety symptoms are shown to be associated with negative emotions being fear, disgust, and anger, as discussed with Dr.Fah Kunpariya, our psychology advisor. 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 nonverbal interactions comprise physical actions from the robot such as arm movement changes in the LED display resembling its eyes. In terms of progress, the goal for the first semester was a minimum viable software product, comprising two working AI models and an integration backend that combines the with physical actions and other processing. In terms of AI model implementation, the facial emotion recognition and speech emotion recognition models produced decent results, being more than 50% accurate in all emotions' classifications. The integration of these two models uses a server-client architecture to deliver physical interaction, which will be discussed in the software architecture section.

4.4.1 Process Flowchart

The process flowchart for Pookie MVP 1 is illustrated in Fig. 5. Overall, Pookie will detect the user's face and voice and process the input to induce an action. Based on the emotions detected, such as happy, sad, or negative emotions (inferred as stress), then Pookie will carry out a set of physical actions, including sound, eyes, and actuator.

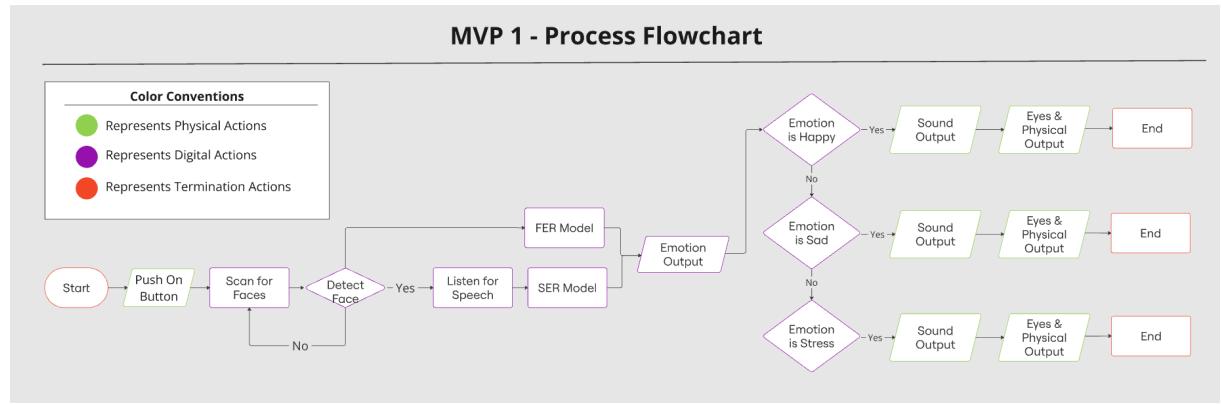


Figure 5: MVP 1 Process Flowchart

4.4.2 Software Architecture and Integration

To go into further detail on the process flowchart, this section illustrates the solution architecture, involving core technologies used, including servers, clients, libraries and so on. **System Overview**

The architecture employs a three-tier structure consisting of:

- **Data Acquisition Layer:** Interfaces with sensors and hardware for input collection (e.g., facial images and voice).
- **Processing Layer:** Handles emotion analysis using AI models and decision-making algorithms.
- **Interaction Layer:** Manages the robot's responses, including audio, visual, and physical movements.

This layered design ensures that each module operates independently, facilitating parallel development and easier debugging.

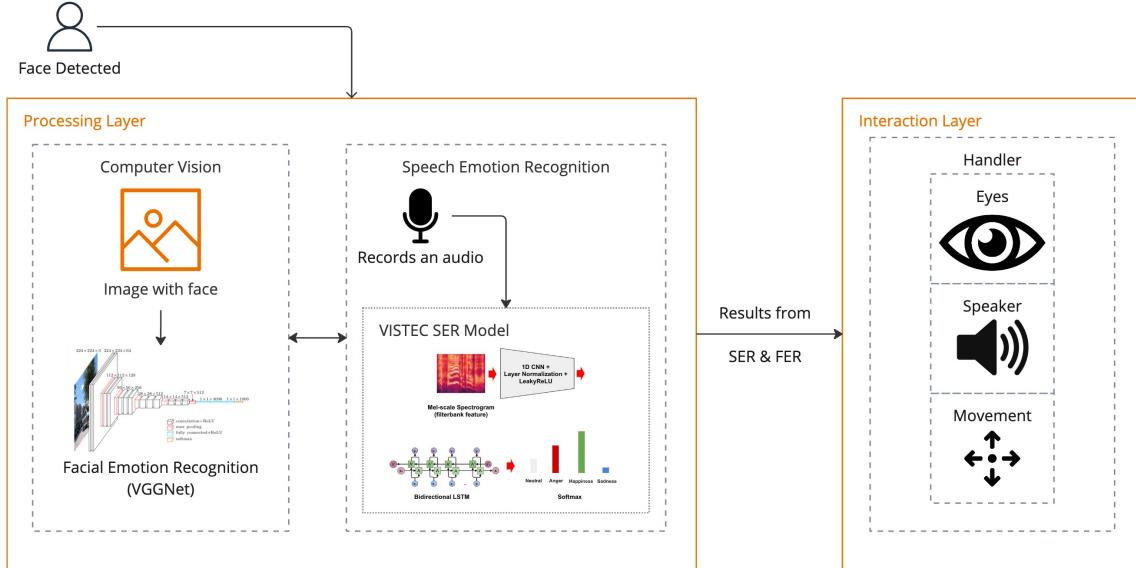


Figure 6: Pooke Software Architecture

4.4.3 Key Components

1. Emotion Detection Module

This module is responsible for identifying the user's emotional state based on facial expressions and speech patterns.

- **Facial Emotion Recognition (FER):** Utilizes a fine-tuned VGGNet model trained on the Chinese Faces Dataset. It processes real-time video feeds to classify emotions such as happiness, sadness, and stress.
- **Speech Emotion Recognition (SER):** Leverages a pre-trained model developed in collaboration with VISTEC, analyzing voice inputs for tonal features like pitch and intensity to detect emotions.

Both components output structured emotion data to the Core Integration Framework for further processing.

2. Interaction Module

The Interaction Module generates empathetic responses based on detected emotions:

- **Verbal Interactions:** Uses pre-recorded audio and tone synthesis to produce comforting sounds or words.
- **Non-Verbal Interactions:** Controls physical actions such as arm movements, eye animations on the LED display, and body rotation using servo motors.

4.4.4 Core Integration Framework

This middleware orchestrates communication between the emotion detection and interaction modules. Key responsibilities include:

- Synchronizing data streams from the FER and SER modules.
- Deciding responses using a rule-based system or decision tree logic.
- Triggering corresponding interaction commands.

4.4.5 Component Interaction

The interaction between components is depicted in the process flowchart. The steps include:

1. **Data Acquisition:** Real-time inputs (video and audio) are captured using a USB camera and a lavalier microphone.
2. **Emotion Detection:** Data is processed by FER and SER modules to output an emotional state.
3. **Decision Making:** The Core Integration Framework maps the detected emotion to predefined response rules.
4. **Response Execution:** Interaction Module activates verbal and non-verbal responses.

4.5 Facial Emotion Recognition

The development of the AI-driven robot incorporates a crucial element: detecting and interpreting user emotions, particularly stress and anxiety, using Facial Emotion Recognition (FER). In particular, stress will be mapped to negative emotions: fear, anger, and disgust.

4.5.1 Faces Dataset

The facial expression recognition model was trained on a Chinese Faces Dataset , chosen for its similar characteristics to the hard-to-obtain Thai dataset. Utilizing an intuitive approach to synthetic data generation, the dataset was based on a **Chinese Face Dataset with Dynamic Expressions and Diverse Ages Synthesized by Deep Learning** [29], where a team of researchers created a facial expression image generation model for various Chinese faces belonging to different age groups, genders, and face structure, as shown in Figure 7, labeled as “Ours”.

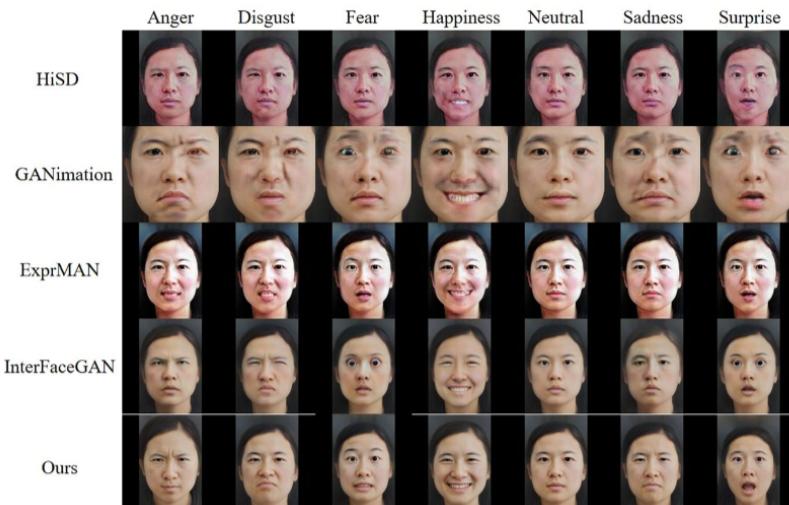


Figure 7: A Chinese Face Dataset with Dynamic Expressions and Diverse Ages Synthesized. Source: Adapted from [29]

4.5.2 Facial Emotion Recognition Model Development

Initially, the facial expression recognition model for this project was expected to use a model fine tuning approach over a Chinese Faces Dataset, as it provides the most symmetric resemblance to a Thai dataset, which is difficult to obtain. For our approach, we took a pre-trained model for emotion recognition using VGGNet architecture, then fine tuned the inference layers on a Chinese dataset in order to get more accurate representation for Thai faces. The other layers were frozen, and served as foundation parameters for transfer learning. After multiple versions of the model, however, it could be seen that this approach did not yield great results. As shown in Figure 8, the model yielded results far worse than simple guessing, most likely due to inappropriate usage of transfer learning on an already imperfect model.

	precision	recall	f1-score	support
anger	0.23	0.25	0.24	24
disgust	0.11	0.11	0.11	27
fear	0.17	0.07	0.10	28
happiness	0.10	0.09	0.10	22
neutral	0.18	0.27	0.22	26
sadness	0.06	0.04	0.05	26
surprise	0.12	0.19	0.15	21
accuracy			0.14	174
macro avg	0.14	0.15	0.14	174
weighted avg	0.14	0.14	0.14	174

Figure 8: Initial Model Training Results

4.5.3 Facial Emotion Recognition Model

Utilizing a VGGNet [30] architecture (Figure 9) —a multi-layer convolutional neural network renowned for feature extraction and classification, particularly in emotion detection models—the research aimed to develop an accurate facial recognition approach. The Chinese faces dataset was trained over the VGGNet architecture with little hyperparameter tuning.

As seen in the results, the model is great at distinguishing between positive emotions such as happiness or surprise, but it is quite lacking in negative emotions. However, given the scope of the project, where positive and neutral are associated with specific outputs, and negative emotions are all classified as stress, then the model is generally enough to use as a minimum viable product for the rest of the semester.

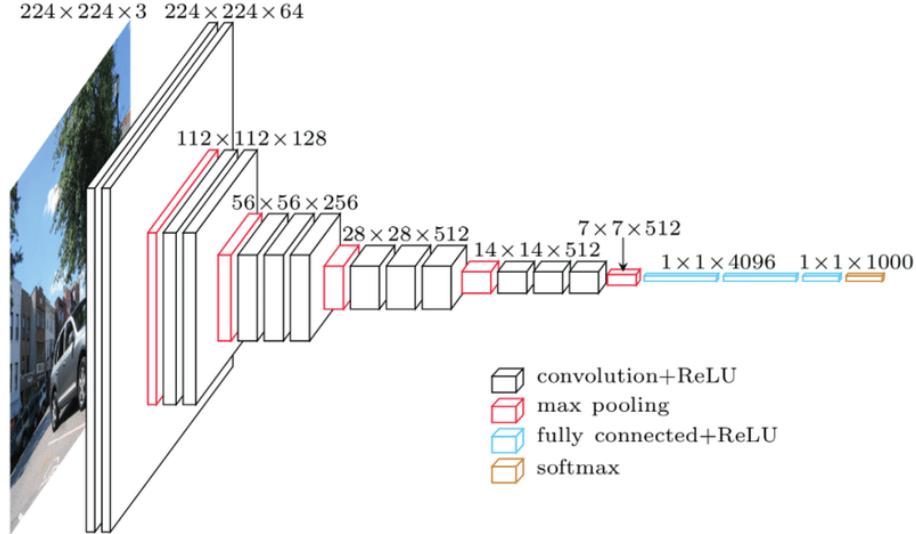


Figure 9: VGGNet Architecture



Figure 10: Validation Set Confusion Matrix

4.6 Speech Emotion Recognition

A key element of Pookie’s AI is the Speech Emotion Recognition (SER) system, which is used for recognizing the user’s emotions based on vocal patterns. By analyzing factors such as pitch, tone, and intensity, the system detects emotions: neutral, anger, happiness, sadness, and frustration. This section outlines the results and challenges faced during the implementation of the speech emotion recognition model, starting with relevant datasets and methodologies.

4.6.1 Speech Emotion Recognition Dataset and Model

We are utilizing a dataset created by Chulalongkorn University in collaboration with VISTEC, DEPA, and AIS, containing 41 hours and 36 minutes of audio recordings labeled with five emotions: neutral, anger, happiness, sadness, and frustration. The SER model used within Pookie is also developed by VISTEC which is a speech emotion recognition model based on the aforementioned dataset.

The confusion matrix visualizes the performance of an emotion recognition model across five classes: neutral, anger, happiness, sadness, and frustration. The diagonal elements indicate correct predictions, with the highest accuracies being for neutral (0.72), anger (0.73), and frustration (0.61). Sadness and happiness show slightly lower accuracies at 0.6 and 0.62, respectively. However, some significant misclassifications occur, such as frustration being confused with sadness (0.31) and happiness being confused with frustration (0.17). The model’s weighted accuracy is 66.12%, while its unweighted accuracy is 65.67%, reflecting a modest overall performance with some variability in recognizing different emotions accurately.

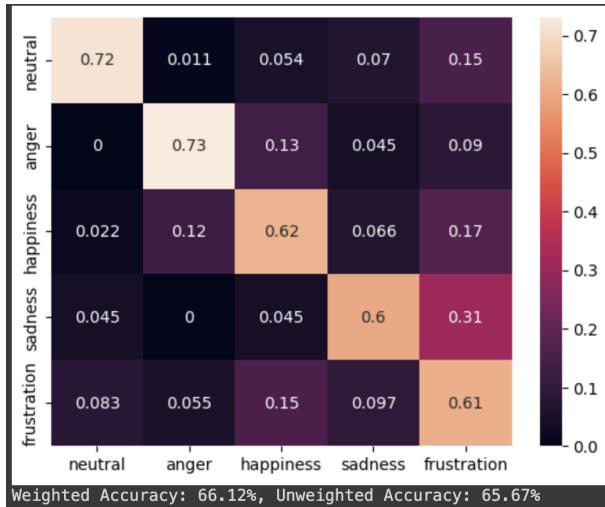


Figure 11: SER Confusion Matrix

4.7 Software Challenges

In this project, multiprocessing was chosen over threading for displaying images with OpenCV’s `imshow()` function due to key technical constraints. OpenCV’s GUI functions, including `imshow()`, are not thread-safe, which leads to unresponsive windows and crash risks if used in a multithreaded environment. Additionally, Python’s Global Interpreter Lock (GIL) limits threading’s effectiveness in CPU-bound tasks, preventing full utilization of multicore processors. By using multiprocessing, each process runs independently with its own memory and execution context, allowing true parallelism and better fault isolation. Though multiprocessing introduces overhead, it ensures stable GUI rendering and optimal use of computational resources, making it a better fit for this real-time image processing task.

4.8 Hardware Implementation

This section details the hardware implementation for Pookie. Overall, the design consists of a robotic shell housing an array of electronic components including the microprocessor, motors, sensors, and so on. A fully functional hardware implementation was not within the scope for this semester, which focused on a basic software proof of concept, and physical design.

4.8.1 Initial Design

1. Outer Shell Design

In the initial phase, a preliminary design is developed for the outer shell. The design is inspired by Rilakkuma, a well-known Japanese character created by San-X in 2003. Rilakkuma is recognized for its comforting and cute aesthetic, which represents relaxation and embodies the essence of kawaii culture. Its rounded features and laid-back attitude have made it a popular symbol of comfort and appeal [31].

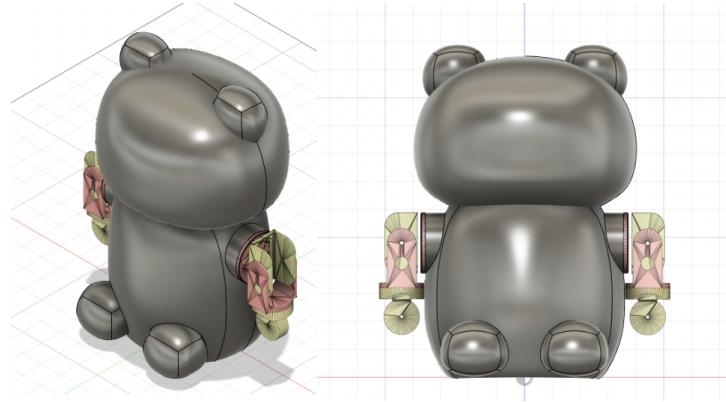


Figure 12: Initial Outer Shell Design of Pookie

2. Inner Shell Mechanism Design

The arms are designed to move along the x-axis revolute joint and are powered by CS002 20 Kg servo motors.

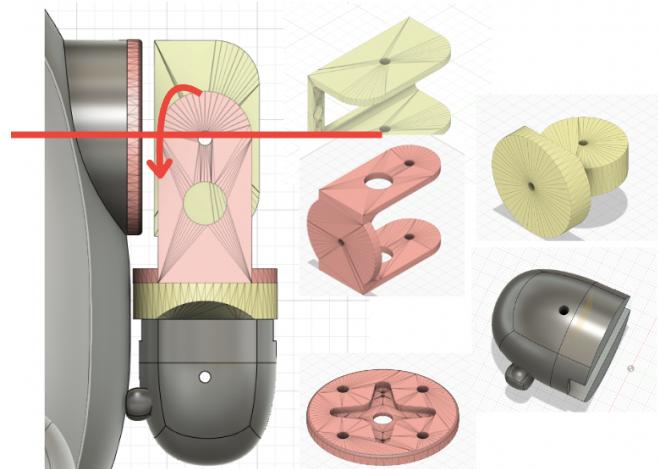


Figure 13: Initial Arm Mechanism of Pookie

4.8.2 Revised Design

The revised hardware design of the Pookie robot focuses on anthropomorphic aesthetics, user-centric functionality, seamless mechanical integration, changes in the arm movement axis, and a more compact size.

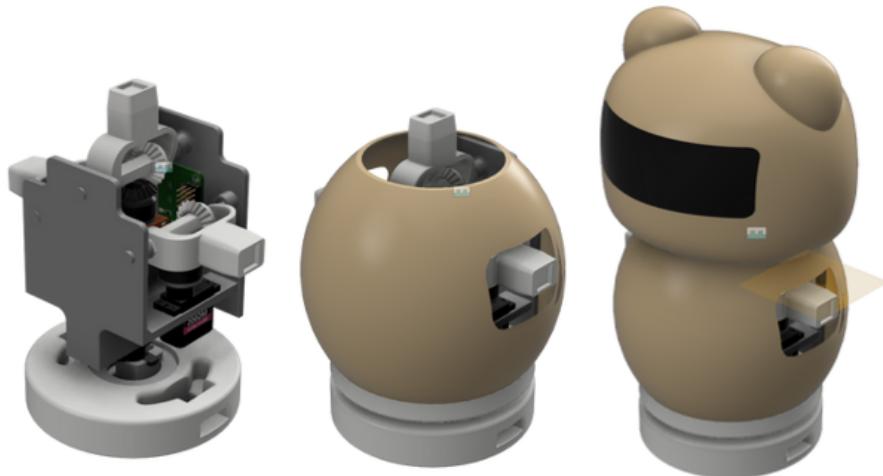


Figure 14: Revised Design of Pookie

1. Outer Shell Design

The outer shell of Pookie is designed to enclose all internal components in a compact form. It consists of two main sections: the body and the head. The body houses the internal systems, while the head contains an LED screen for displaying the robot's eyes and serves as the mounting point for the head movement mechanism.



Figure 15: Revised Outer Shell Body of Pookie



Figure 16: Revised Outer Shell Head of Pookie

2. Inner Shell Mechanism Design

The movement mechanisms in Pookie include systems for the head, arms, and base. Each actuator is represented as a cylinder, indicating rotational motion driven by MG90s servo motors. The motion occurs along defined axes, with the z-axis primarily supporting rotational movements.

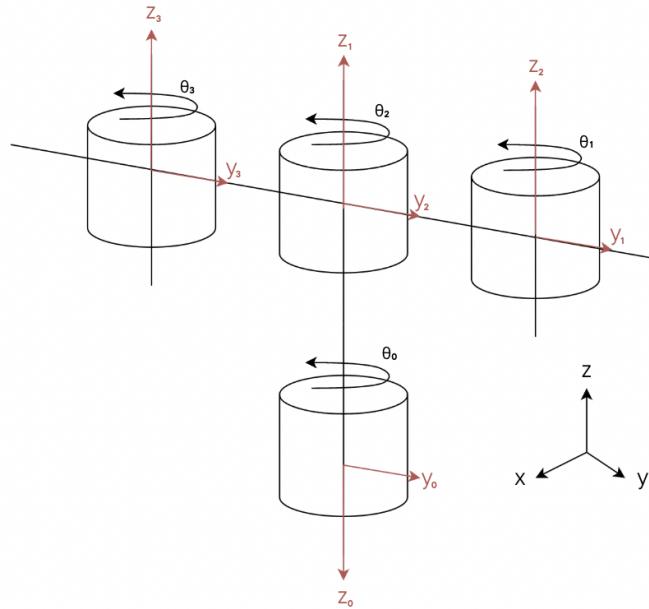


Figure 17: Motion Axes for Pookie's Movement Mechanisms

Both the head and arm movements use a bevel gear mechanism with a 1:1 transmission powered by MG90s servo motors. This system provides a single degree of freedom along the **y-axis** revolute joint. A magnet slot at the tip allows easy attachment to the outer shell head, simplifying assembly and maintenance.

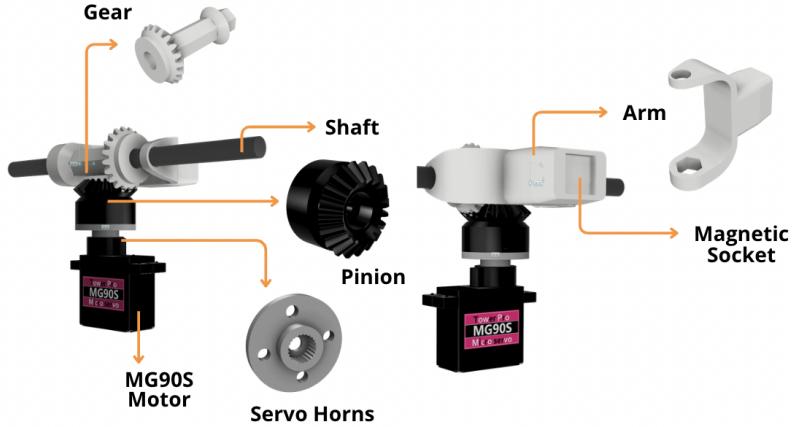


Figure 18: Components of the Head and Arm Mechanisms

The bevel gear mechanism transfers motion from the servo motor to the movement axis efficiently. With a 1:1 transmission ratio, the angular velocity (ω) of the driving gear is equal to that of the driven gear. This relationship is expressed as:

$$\omega_{\text{output}} = \omega_{\text{input}}$$

where ω_{input} is the angular velocity of the motor, and ω_{output} is the angular velocity of the output gear. Similarly, the angular acceleration (α) remains unchanged during transmission:

$$\alpha_{\text{output}} = \alpha_{\text{input}}$$

This design ensures that the servo motor's precise control over movement translates directly to the head and arm mechanisms. The torque transmitted by the system is given by:

$$\tau_{\text{output}} = \tau_{\text{input}}$$

where τ_{input} is the torque generated by the servo motor. The selection of a 1:1 transmission ratio supports direct coupling, allowing for seamless motion transmission without the need for additional adjustments or gear reduction.

For the base rotation, a bearing-supported system allows full 360-degree rotation along the z-axis. This system incorporates a low-friction bearing assembly, ensuring smooth and precise rotational motion.

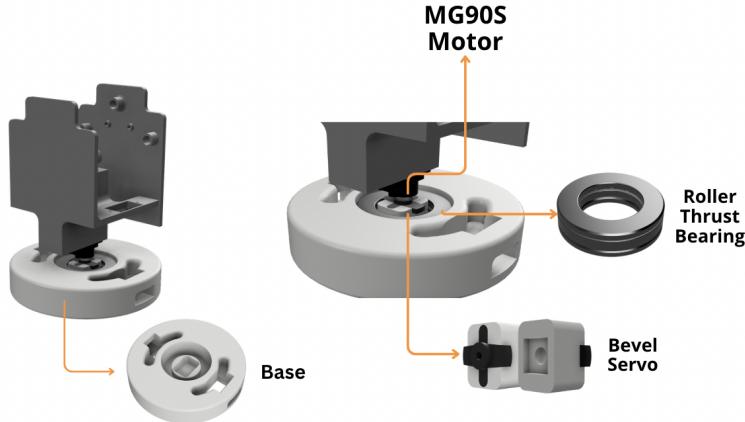


Figure 19: Components of the Base Mechanism

An armor has been designed to house and support the head, arm, and base rotation mechanisms. It includes compartments for the servo motor and the PWM servo motor driver, as well as openings to

facilitate wiring integration. This armor ensures structural integrity and maintains the functionality of all movement mechanisms within Pookie's compact design.

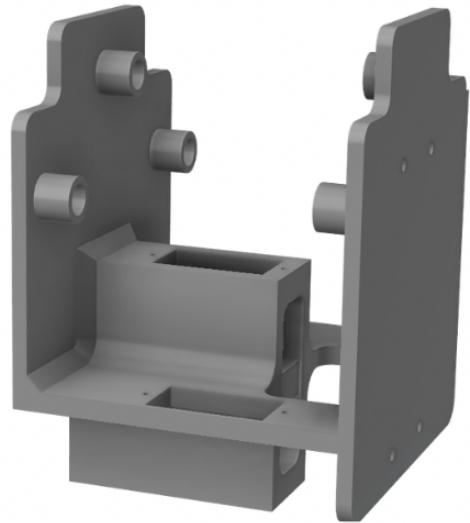


Figure 20: Internal Armor Design of Pookie

5 Testing

This section details the proposed approach to testing Pookie on end users. Overall, the testing phase will be conducted in Semester 2 under the supervision of our psychology advisor, Dr. Fah Kunpariya Siripanit. The testing will comprise 3-4 end users diagnosed with general anxiety. These end users will be volunteers provided by Chula Student Wellness, and in some cases, personal connections.

The selected users will be given Pookie for a total of seven days to incorporate the robot into their daily lives. Our advisor suggests that seven days is the minimum duration required for a user to develop an emotional attachment. Once they are given the robot, they will complete a questionnaire both before and after the seven-day period to evaluate whether Pookie has had a positive impact on their daily lives. The questions include self-ratings on current anxiety and positivity levels (on a scale from 1-10).

Since there is no universal interpretation of these scores, a control group (those without access to Pookie) will also answer the questions, allowing the team to establish baseline scores and draw meaningful insights. The key questions are summarized in Table 1. It is important to note that while the general rationale for these questions comes from our advisor, the specific wording and execution of the questionnaire will be finalized next year.

Table 1: List of Questions for Testing

Key Questions	Quantitative Metric
How would you rate your everyday level of anxiety?	Self-rated scale from 1-10
How would you rate your maximum level of anxiety?	Self-rated scale from 1-10
How would you rate your everyday level of positivity?	Self-rated scale from 1-10
How would you rate your maximum level of positivity?	Self-rated scale from 1-10

In regard to security and PDPA compliance, Pookie will require full consent from the end users for the temporary usage of face and voice data. The team is seeking assistance from Chula Regulations to ensure that Pookie adheres to all necessary legal and ethical requirements for testing.

6 Summary and Real Benefit to Industry

6.1 Summary

In summary, the project focuses on developing an emotional wellbeing robot that promotes mental wellness and positivity for users under the influence of stress and anxiety. The robot will utilize technologies such as computer vision and natural language processing to recognize and respond to emotional cues in real time, providing consistent and empathetic support. With a focus on design, interactivity, and empathy, the robot is tailored to meet the needs of Gen Z and younger millennials in promoting mental wellbeing. The project is structured around two academic semesters, employing an agile methodology to deliver two key prototypes, MVP1 (Proof of Concept) and MVP2 (A Fully Functional Design). The development process is guided 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 not only meets technical standards but also aligns with mental health principles. This project offers a scalable, innovative solution that addresses the increasing prevalence of anxiety disorders, positioning the robot as a sustainable alternative to traditional mental health support methods.

6.2 Benefits Direct Impact on the Industry

This project will significantly benefit the mental health field, specifically targeting patients with general anxiety and stress, which comprises a massive customer segment. In particular, mental wellbeing robots are an important initiative in countering **Terror Outbursts** in Thailand. By leveraging AI to detect and respond to emotions through facial expressions and voice analysis, these robots can reduce reliance on human intervention in mental positivity promotion. Additionally, the project will enhance service quality by offering consistent promotion of mental well-being, effectively addressing unmet mental health needs. Scalability and Long-Term Value: With recent studies indicating a 55% increase in anxiety disorders globally from 1990 to 2019, affecting approximately 301 million people worldwide, these robots will become increasingly essential. Their ability to offer real-time, personalized assistance will not only improve individual well-being but also contribute to the long-term sustainability of mental health care systems. While the project will remain in the proof of concept and prototyping stage, it aims to provide an innovative approach to mental wellbeing, potentially scaling to commercial use in the future.

7 Team Roles and Responsibilities

Although each student is assigned to 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. Hakkanpää, 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] T. Wang, Y.-c. Lee, and Q. Ma, “Within and across-language comparison of vocal emotions in mandarin and english,” *Applied Sciences*, vol. 8, no. 12, 2018. [Online]. Available: <https://www.mdpi.com/2076-3417/8/12/2629>
- [12] E. Rodero, “Intonation and emotion: Influence of pitch levels and contour type on creating emotions,” *Journal of Voice*, vol. 25, no. 1, pp. e25–e34, 2011. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0892199710000378>
- [13] 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>
- [14] H. Chapre, L. Choudhury, and M. Kamber, “Emotion / facial expression detection,” *International Journal of Science and Research (IJSR)*, vol. 12, no. 5, pp. 1395–1398, May 2023. [Online]. Available: <https://www.ijsr.net/>
- [15] H. Liu, “Emotion detection through body gesture and face,” 2024. [Online]. Available: <https://arxiv.org/abs/2407.09913>

- [16] 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>
- [17] 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>
- [18] 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.
- [19] K. Wada and T. Shibata, "Robot therapy in a care house - its sociopsychological and physiological effects on the residents," in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, 2006, pp. 3966–3971.
- [20] N. Hurst, C. Clabaugh, R. Baynes, J. Cohn, D. Mitroff, and S. Scherer, "Social and emotional skills training with embodied moxie," 2020. [Online]. Available: <https://arxiv.org/abs/2004.12962>
- [21] 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>
- [22] R. Zhang, "Facial emotion detection based on improved vgg-16," in *Proceedings of the 5th International Conference on Computing and Data Science*. Open Access Proceedings, 2023. [Online]. Available: <https://doi.org/10.54254/2755-2721/21/20231110>
- [23] S. A., G. T., J. Aravind, M. Venkatratna, and R. Gayathri, "Speech emotion recognition," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 9, no. VIII, pp. 2637–2640, 2021. [Online]. Available: <https://doi.org/10.22214/ijraset.2021.37375>
- [24] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2019. [Online]. Available: <https://doi.org/10.48550/arXiv.1810.04805>
- [25] Y. Wu, Z. Jin, C. Shi, P. Liang, and T. Zhan, "Research on the application of deep learning-based bert model in sentiment analysis," in *Proceedings of the 2nd International Conference on Software Engineering and Machine Learning*, 2024. [Online]. Available: <https://doi.org/10.54254/2755-2721/67/2024MA>
- [26] S. Shahriar, B. D. Lund, N. R. Mannuru, M. A. Arshad, K. Hayawi, R. V. K. Bevara, A. Mannuru, and L. Batool, "Putting gpt-4o to the sword: A comprehensive evaluation of language, vision, speech, and multimodal proficiency," *Preprints*, 2024. [Online]. Available: <https://doi.org/10.20944/preprints202406.1635.v1>
- [27] M. Abuhmida, M. J. Islam, and W. Booth, "Empathy in ai: Developing a sentiment-sensitive chatbot through advanced natural language processing," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 13, no. 3, pp. 140–147, 2024. [Online]. Available: <https://doi.org/10.54254/2755-2721/67/2024MA>
- [28] F. Safavi, P. Olikkal, D. Pei, S. Kamal, H. Meyerson, V. Penumalee, and R. Vinjamuri, "Emerging frontiers in human–robot interaction," *Journal of Intelligent & Robotic Systems*, vol. 110, no. 45, 2024. [Online]. Available: <https://doi.org/10.1007/s10846-024-02074-7>
- [29] S. Han, Y. Guo, X. Zhou, J. Huang, L. Shen, and Y. Luo, "A Chinese Face Dataset with Dynamic Expressions and Diverse Ages Synthesized by Deep Learning," *Scientific Data*, vol. 10, no. 1, 12 2023. [Online]. Available: <https://doi.org/10.1038/s41597-023-02701-2>
- [30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. [Online]. Available: <https://arxiv.org/abs/1409.1556>
- [31] Hinka, "The history of rilakkuma and friends: From stress buster to kawaii stationery," accessed: 2024-11-22. [Online]. Available: <https://www.hinka.co/blog/the-history-of-rilakkuma-and-friends-from-stress-buster-to-kawaii-stationery>