CHULA INTERNATIONAL SCHOOL OF ΣNGINEERING
The ingenuity of CHULA ΣNGINEERING

# Project Proposal: *Emotional Wellness Robot*

**Authors:** Tibet Buramarn, Kridbhume Chammanard, and Thitaya Divari

**Advisor:** Dr. Paulo Fernando Rocha Garcia and
Ms. Kunpariya Siripanit

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

September 6, 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 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 features an anthropomorphic form with LED displays and sensors for interaction, drawing inspiration from existing robots like Kiki and Eilik, 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

# Pookie: An AI-Driven Robot for Promoting Mental Wellbeing and Emotional Support

## 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 Outburst" 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. Given this, 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 an mental wellbeing 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. In contrast, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks [17] are essential for processing sequential data, such as audio signals, enabling the detection of emotions through vocal features.

Effective emotion detection relies on extracting meaningful features from raw data. For instance, Mel-Frequency Cepstral Coefficients (MFCCs) [18] are a crucial feature extraction technique in speech emotion recognition, capturing the essential characteristics of the audio signal that correlate with emotional states. 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 [19].

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 [20].

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 [21]. Microphones and audio sensors play a crucial role in gathering vocal cues, which are vital for emotion detection [22]. 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 making sounds when the robot is touched on the head, with expert supervision from Chula Student Wellness (CUSW).
- Develop an accurate emotion detection algorithm that captures the user's state of anxiety through speech emotion recognition and facial expression recognition.
- Develop empathetic human-robot interactions that promote emotional wellness within the customer.
- Finalize a customer-centric design obtained through market validation efforts.
- Conduct extensive testing and refinement based on user feedback.

## 2.3 Measurable Outcomes

- Achieve a relative reduction in self-reported anxiety.
- Achieve a benchmark in accuracy metrics for emotion detection.
- Obtain an improved before-and-after positivity score.

## 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, increase socialization, and improve overall mood among users [23].

Figure 1: PARO, baby seal robot adapted from [24]

2. **Therabot**

   Therabot is a robot designed to engage users in therapeutic interactions by responding to their emotions through computer vision and providing appropriate feedback, helping individuals process their feelings and improve their mood. It features a soft-textured exterior, which participants have expressed a strong preference for, as it creates a sense of safety and emotional support, making the robot more approachable and pleasant to interact with. Additionally, the integration of a mobile app allows for more interactive and personalized sessions, enhancing user engagement and providing a more beneficial therapeutic experience [25].


Figure 2: Therabot, the robotic beagle puppy adapted from [26]

3. **Pepper**

   Pepper is a humanoid robot designed to recognize emotions through facial expressions, vocal tones, and body language, enabling empathetic interactions in both personal and business settings. Using natural language processing (NLP), it engages in conversations and tailors its responses based on context and past interactions. Pepper features 360-degree mobility, an interactive touchscreen, and cloud-based AI for real-time updates. It integrates with business systems, making it versatile for retail, healthcare, and education, where it assists with tasks like greeting customers, monitoring patients, and engaging children. Customizable and adaptive, Pepper excels as an emotional support and service robot [27].

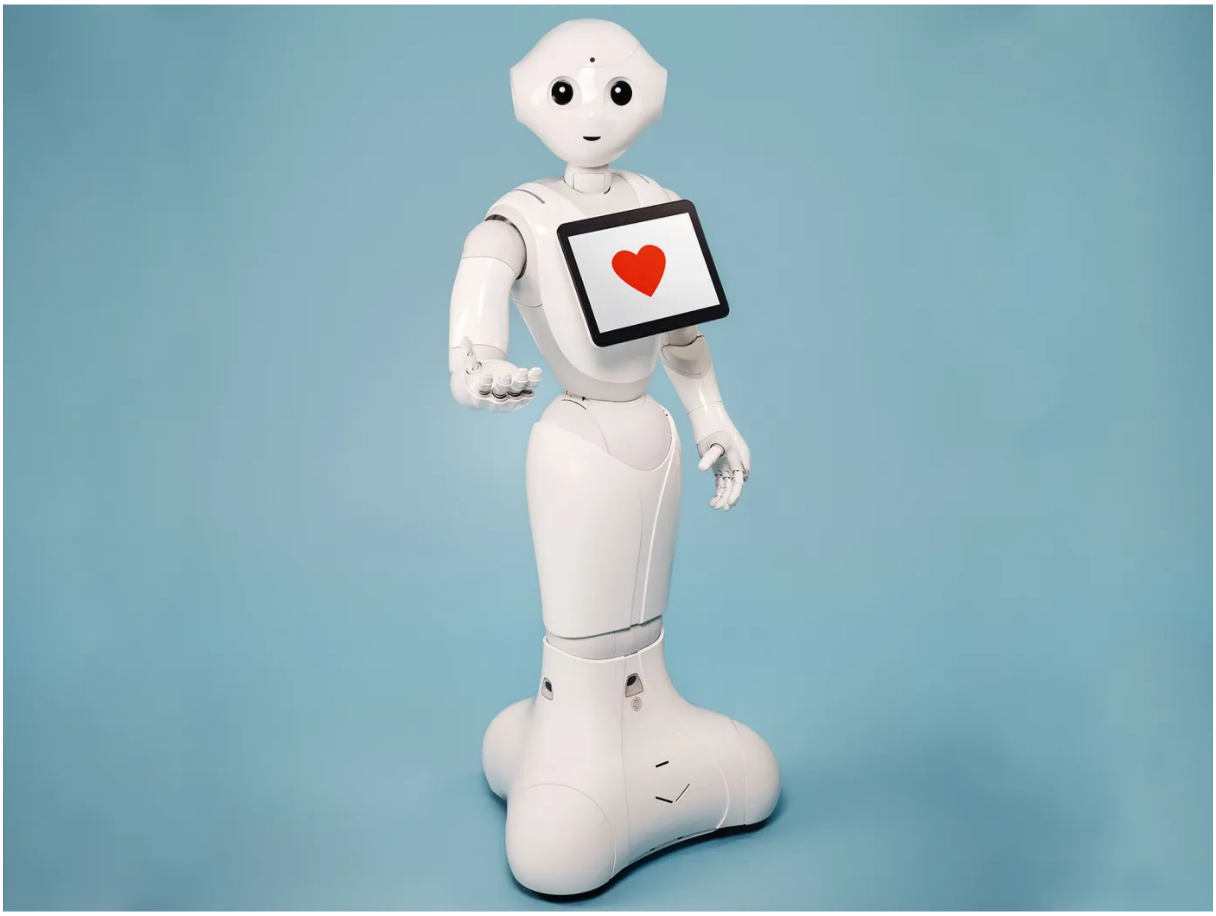

Figure 3: Pepper, the humanoid robot adapted from [28]

4. **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 [29].

Figure 4: Moxie, a learning robot with heart adapted from [30]

### 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 is a critical component, where convolutional neural networks (CNNs) are often employed to classify emotions based on facial features. Significant progress has been made in this domain with models like OpenFace [31] and VGGFace [32], 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 [33]. Techniques like Long Short-Term Memory (LSTM) networks and more recently, transformer-based models such as BERT [34], 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 [35] and GPT [36] 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 [35].

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 [37].

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 [37]. 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 [38].

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 [38].

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 [38].

# 4 Project Concept Development

## 4.1 Acknowledgement

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 wellbeing robot, is conceptualized as a responsive, AI-driven companion designed to provide 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.

As mentioned, the project envisions a robot that satisfies three key pillars of customer needs: appearance, interactivity, and empathy. This will be accomplished by 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, drawing inspiration from existing models such as Kiki and Eilik. 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 personalized 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.

# 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 each MVP, 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 1 - Proof of Concept & Customer-Centric Specifications

MVP 1 will span the entirety of academic semester 1 (from September until December) and will focus on delivering a proof of concept of the project, as well as feature specifications that focus on the customers' needs. MVP 1 will comprise three sprints, each lasting for around a month.

The project starts at MVP 1 Sprint 0, which focuses on preliminary research and feature definition. This sprint will span the entirety of September, and comprises the project proposal, in-depth customer journey, and technology specifications (such as specifically which AI models to be used). MVP 1 Sprint 0 will have two deliverables: the written project proposal and first progress report.

In the next phase, MVP 1 Sprint 1, which lasts from October to early November, we will focus on preliminary software development aimed towards providing a proof of concept for the emotion detection algorithm. For each AI model that will be used, we will allocate time for data collection, training, and iterative quality checks using real data. In general, emotion detection will comprise detection of facial expressions, Speech Emotion Recognition (SER), and Gesture Recognition technologies, utilizing computer vision and feature extraction technologies. In addition, there are plans to integrate context understanding into the robot using NLP technologies, but we will not include it in the scope. Additionally, the UX design will also be drafted, consisting of a complete customer journey for each feature, as well as a basic outer shell for the robot that fits the design specifications. Sprint 1 will have three deliverables: second progress report, third progress report, and completed prototypes for emotion detection.

In the last phase, MVP 1 Sprint 2, which lasts from November to early December, we will focus on acquiring the first batch of hardware components for the project and testing the emotion detection models on our selected microcontroller. The microcontroller must have enough compute power to perform inferences in real time. Additionally, Sprint 2 will involve market validation, which comprises testing on the customer segment to obtain feedback for improvement in the following semester. By the end of MVP 1, we expect a prototype that integrates software and hardware components on a feasible level.

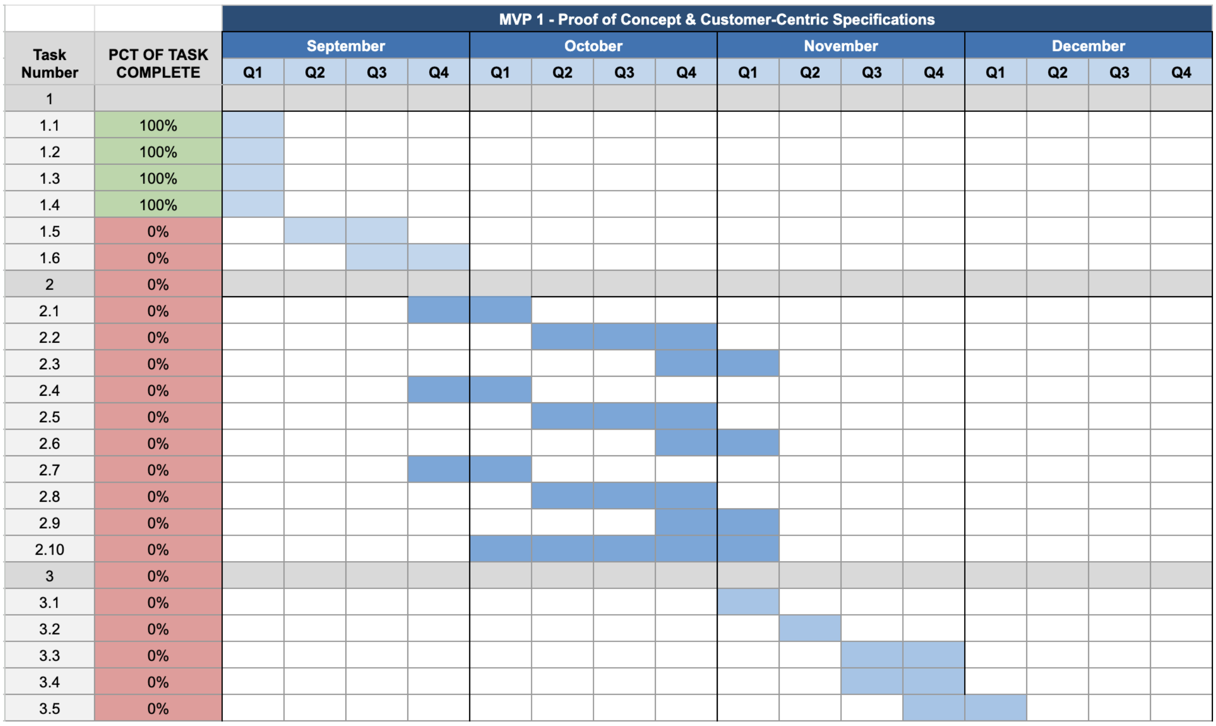


Figure 5: Gantt Chart for MVP 1 Timeline

## Task Number:

1. **MVP 1 - Sprint 0**
   1.1 Advisor Onboarding
   1.2 Project Scoping
   1.3 Project Scoping - Expert Interviews
   1.4 Project Proposal
   1.5 In-Depth Technology Specifications
   1.6 E2E Customer Journey

2. **MVP 1 - Sprint 1**
   2.1 Facial Expression Model - Data Collection
   2.2 Facial Expression Model - Training/Testing
   2.3 Facial Expression Model - Quality Testing
   2.4 SER Model - Data Collection
   2.5 SER Model - Training/Testing
   2.6 SER Model - Quality Testing
   2.7 Gesture Model - Data Collection
   2.8 Gesture Model - Training/Testing
   2.9 Gesture Model - Quality Testing
   2.10 UX Design (Fusion360)

3. **MVP 1 - Sprint 2**
   3.1 Parts Procurement
   3.2 Jetson Orin/Nano Test
   3.3 MVP 1 Integration
   3.4 Market Validation
   3.5 Final Presentation Preparations

## 5.3 MVP 2 - Non-Commercial Prototype

In this project proposal, the specific details of MVP 2 will not be disclosed. However, in general, MVP 2 will comprise three sprints, similar to MVP 1, but will focus on integration of the emotion detection algorithm with hardware components to create an output. MVP 2 will last the entirety of academic semester 2, and is expected to deliver a fully functional prototype, but will not be implemented to the extent of commercialization. Additionally, MVP 2 will focus on user testing, involving measurable metrics mentioned in the objectives section. The prototype will not encompass full consideration of security, safety, and real usage, but will focus on fulfilling the customer demands of appearance, interactivity, and empathy.

## 5.4 Resource Allocation

The project team comprises three senior engineering students. The roles for the project will be designated as follows: project manager, project engineer for product development, and project engineer for software development. The project manager is responsible for overseeing the entire project, delegating tasks, managing the timeline, and creating engagements with advisors from ISE and Chula Student Wellness. They are expected to streamline work processes for the project engineers and professionally check deliverables before submission. The project engineers are separated into two categories: product development and software development. The product development engineer will be responsible for UX design through CAD software, gathering hardware components, and managing the budget for doing so. The software development engineer will be responsible for overseeing all software initiatives, as well as managing DevOps practices within the project.

## 5.5 Budget Allocation

While the exact hardware components for the robot cannot be deduced yet, the table below will illustrate an approximate allocation of the budget provided for the project. Note that the budget specified is the maximum that was allocated for expenditure. The budget for each component may vary, but should never exceed the maximum allocated budget.

| **Component** | **Maximum Allocated Budget (THB)** |
|---|---|
| Microcontroller | 30,000 |
| Camera | 5,000 |
| Microphone | 5,000 |
| Electronics | 15,000 |
| Chassis and Framework | 5,000 |
| Decorative Components | 5,000 |
| LED Display | 5,000 |
| Miscellaneous | 5,000 |
| **Maximum Expenditure** | **75,000** |

Table 1: Approximate Budget Allocation for Hardware Components

# 6 Theory and Technical Backup

## 6.1 Hardware Features

The physical design of the robot is anthropomorphic-centric, with elements of animals as well. The robot will be around 12 inches tall, comprising various integrated hardware components. Starting from the top, the robot's head will consist of a 3D printed sphere, and eyes made from an integrated LED display, which will be used as a form of interaction with the user. In addition, the head will also house the camera and microphone, used to receive image and sound inputs for processing within the microcontroller. Next, the robot's body will consist of a large chassis to house the electrical components and the microcontroller. The body will also consist of motors located on the arms to allow for minor arm movement. Lastly, touch sensors will be placed in certain parts of the robot to imitate petting interaction. This design is akin to many desktop companion robots for promoting mental wellness, such as Kiki or Eilik, with the core difference being in the integration of various emotion detection methods.

## 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.

## 6.3 Emotion Detection Model - 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. OpenFace [31] and VGGFace [32] are among the most prominent models in this field. These models extract key facial features—such as eye movement, mouth shape, and eyebrow positions—from images and videos to classify emotions like happiness, anger, and sadness. CNNs have demonstrated strong performance by learning spatial hierarchies of features, allowing them to detect subtle changes in facial expressions, even in complex or dynamic environments. Such models are pivotal in creating emotionally responsive robots, as they allow real-time emotion tracking through visual input.

## 6.4 Emotion Detection Model - Speech Recognition

Speech recognition for emotional detection focuses on analyzing vocal characteristics to infer emotional states. Key prosodic features such as pitch, tone, and rhythm are crucial indicators of emotions in speech. Recent methods have employed Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks for processing audio data, while newer approaches utilize transformer models like BERT [34]. These models capture the temporal dependencies in spoken language, enabling the system to interpret emotions more accurately. LSTMs are particularly effective at maintaining contextual information over time, which is vital for understanding emotions that are expressed through vocal modulations. The integration of these techniques allows robots to engage in emotionally intelligent conversations by understanding the user's mood through their voice.

## 6.5 Testing

The testing of the robot will comprise both internal and external components. In terms of internal components: The emotion detection model should achieve a benchmark score in accuracy metrics, or a similar number on a different scale. Since both the computer vision and speech emotion recognition (SER) datasets will be discretely labeled, quantitative measurement of accuracy will be feasible. For instance, the speech recognition model may classify a person's voice as either normal or anxious.

On the other hand, in terms of external components, user testing will involve supervision and validation from Chula Student Wellness. Questionnaires will be developed to obtain the required information accordingly, representative questions include:

- **Self-Report on Anxiety**
  1. Before using Pookie, how would you rate your anxiety levels from 1-10? Please provide your rating now and when you're at your most anxious.
  2. After using Pookie, how would you rate your anxiety levels from 1-10? Please provide your rating now and when you're at your most anxious.
- **Positivity Promotion**
  1. Before using Pookie, how would you rate your general positive feelings in life from 1-10?
  2. After using Pookie, how would you rate your general positive feelings in life from 1-10?
- **Attachment**
  1. After using Pookie for 1 week, do you feel more attached to it?feelings in life from 1-10?

While these are examples of questions that will be asked, proper guidance from Chula Student Wellness will be given as well during the testing stage.

# 7 Project Outcome

Over the span of the project, the team will have many deliverables. During the first semester, there will be a total of seven deliverables: project proposal, first progress report, second progress report, final report draft, final presentation, final report, and a live demo of MVP1. As mentioned in project planning, MVP1 is expected to be a proof of concept of the basic emotion detection algorithms and their integration with other hardware components.

On the other hand, during the second semester, the number of deliverables in the form of reports has not been disclosed yet, but a functional, testable prototype will be completed. The team expects to perform intensive testing through the customer segment to obtain feedback for scaling the robot for commercial use. However, the outcome of the project does not envision a robot ready for commercial deployment.

To recap the overarching goals of the project, the robot must fulfill three key pillars of customer expectations for emotional wellness robots: appearance, interactivity, and empathy. By the end of the project, we will have completed:

- An anthropomorphic robotic outer shell that resonates with the target customer segment, which seamlessly houses and integrates electronic components.
- Interactive and feasible features for the robot through elements such as touch, under expert supervision from Chula Student Wellness (CUSW)
- An accurate emotion detection algorithm that effectively captures the user's anxiety state.
- Empathetic human-robot interactions that promote emotional wellness.
- A customer-centric design through market validation efforts.
- A thorough report on user feedback after testing the robot.

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

- A relative reduction in anxiety levels from user testing which will be measured as a self-report.
- Benchmark accuracy across various emotion detection metrics, indicating a relatively functional model.
- Improvement in before-and-after lifestyle positivity scores.

# 8 Summary and Real Benefit to Industry

## 8.1 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.

## 8.2 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 [39], 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.

# 9 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 preprint*, 2024. [Online]. Available: https://doi.org/10.1101/2024.07.17.24310551

[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, 2024. [Online]. Available: https://dx.doi.org/10.21275/SR24508214759

[3] J. T. Charles, S. Jannet, and C. S. J, "Mitigating occupational mental health-related factors to prevent manufacturing industry accidents," *Research Square preprint*, 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://doi.org/10.1016/j.jad.2023.08.035

[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, pp. 222–229, 2021. [Online]. Available: https://doi.org/10.5770/cgj.24.447

[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. D. 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 ACM Conversational User Interfaces 2024 (CUI '24)*, 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," in *Highlights in Science, Engineering and Technology, CSIC 2023*, vol. 85. CSIC, 2024, pp. 298–308. [Online]. Available: https://doi.org/10.54097/be97jg10

[9] A. Lillo, A. Saracco, E. Siletto, C. Mattutino, and C. Gena, "Investigating the relationship between empathy and attribution of mental states to robots," in *Workshop_AVI2024 Proceedings*. Department of Computer Science, University of Turin, 2024, pp. 298–308. [Online]. Available: https://doi.org/10.48550/arXiv.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://doi.org/10.1016/j.jvoice.2019.10.002

[11] T. Wang, Y. cheol Lee, and Q. Ma, "Within and across-language comparison of vocal emotions in mandarin and english," *Applied Sciences*, vol. 8, no. 12, p. 2629, 2018. [Online]. Available: https://doi.org/10.3390/app8122629

[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://doi.org/10.1016/j.jvoice.2010.02.002

[13] R. Abbas, B. Ni, R. Ma, T. Li, Y. Lu, and X. Li, "Context-based emotion recognition: A survey," *SSRN Preprints*, 2023. [Online]. Available: https://doi.org/10.2139/ssrn.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, 2023. [Online]. Available: https://doi.org/10.21275/SR23516180518

[15] H. Liu, "Emotion detection through body gesture and face," *arXiv preprint arXiv:2407.09913*, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2407.09913

[16] M. M. Taye, "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions," *Computation*, vol. 11, no. 3, p. 52, 2023. [Online]. Available: https://doi.org/10.3390/computation11030052

[17] 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

[18] 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

[19] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends, and challenges," *arXiv preprint arXiv:1708.05148*, 2017. [Online]. Available: https://doi.org/10.48550/arXiv.1708.05148

[20] C. Bartneck, T. Belpaeme, F. Eyssel, T. Kanda, M. Keijsers, and S. Sabanovic, *Human-Robot Interaction – An Introduction*, 2nd ed. Cambridge University Press, 2024. [Online]. Available: https://www.human-robot-interaction.org/

[21] S. Gupta, J. Das, K. Gupta, O. Bholankar, P. Date, and S. Kaushal, "Enhancing mood-based music selection through physiological sensing technology," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 12, no. IV, pp. 71–74, 2024. [Online]. Available: https://doi.org/10.22214/ijraset.2024.59658

[22] R. Rastogi, T. Anand, S. K. Sharma, and S. Panwar, "Emotion detection via voice and speech recognition," *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL)*, vol. 13, no. 1, 2023. [Online]. Available: https://doi.org/10.4018/IJCBPL.333473

[23] K. Wada and T. Shibata, "Robot therapy in a care house - its sociopsychological and physiological effects on the residents," in *Proceedings of the 2006 IEEE International Conference on Robotics and Automation.* Orlando, Florida: IEEE, 2006. [Online]. Available: https://doi.org/10.1109/ROBOT.2006.1642310

[24] T. Shibata, L. Hung, S. Petersen, K. Darling, K. Inoue, K. Martyn, Y. Hori, G. Lane, D. Park, R. Mizoguchi *et al.*, "Paro as a biofeedback medical device for mental health in the covid-19 era," *Sustainability*, vol. 13, no. 20, p. 11502, 2021.

[25] C. L. Bethel, Z. Henkel, S. Darrow, and K. Baugus, "Therabot-an adaptive therapeutic support robot," in *Proceedings of the 2018 IEEE International Conference on Robotics and Automation*, IEEE. IEEE, 2018. [Online]. Available: https://doi.org/10.1109/DISA.2018.8490642

[26] AI for Good, "Therabot - adaptive therapeutic support robot," 2024, accessed: 2024-09-06. [Online]. Available: https://aiforgood.itu.int/speaker/therabot/

[27] S. Robotics, "Pepper press kit 2022," SoftBank Robotics, Paris, France, Tech. Rep., 2022, accessed: 2024-09-06. [Online]. Available: https://robotik-info.ch/wp-content/uploads/2022/07/Pepper-press-kit-EN-2022_1.pdf

[28] SoftBank Robotics , "Pepper - robots: Your guide to the world of robotics," 2014, accessed: 2024-09-06. [Online]. Available: https://robotsguide.com/robots/pepper/

[29] N. Hurst, C. Clabaugh, R. Baynes, J. Cohn, D. Mitroff, and S. Scherer, "Social and emotional skills training with embodied moxie," *arXiv preprint arXiv:2004.12962v1*, 2020. [Online]. Available: https://doi.org/10.48550/arXiv.2004.12962

[30] I. Embodied, "Moxie robot official website," 2020, accessed: 2024-09-06. [Online]. Available: https://moxierobot.com/

[31] 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

[32] 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

[33] 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

[34] 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

[35] 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

[36] 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

[37] 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

[38] 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

[39] S. F. Javaid, I. J. Hashim, M. J. Hashim, E. Stip, M. A. Samad, and A. A. Ahbabi, "Epidemiology of anxiety disorders: Global burden and sociodemographic associations," *Middle East Current Psychiatry*, vol. 30, no. 44, 2023. [Online]. Available: https://doi.org/10.1186/s43045-023-00315-3