CHULA INTERNATIONAL SCHOOL OF ΣNGINEERING
The ingenuity of CHULA ΣNGINEERING

Progress Report 2

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 and
Ms. Kunpariya Siripanit

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

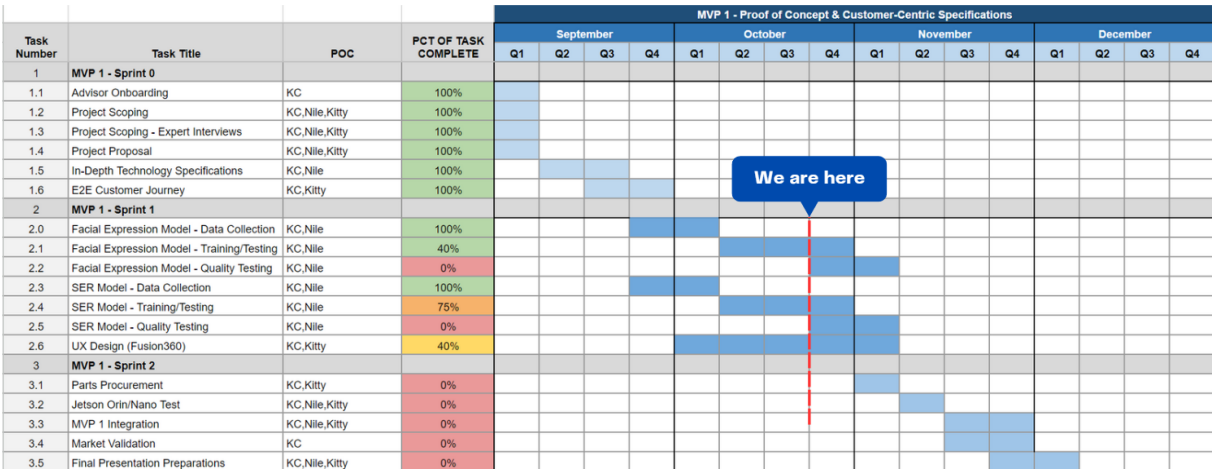
October 25, 2024

Contents

1 Introduction

This progress report provides an update on the development of Pookie, an AI-driven robot designed to promote mental well-being, developed in assistance from Chula Student Wellness. 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.

2 Overview



Task Number	Task Title	POC	PCT OF TASK COMPLETE
1	**MVP 1 - Sprint 0**		
1.1	Advisor Onboarding	KC	100%
1.2	Project Scoping	KC,Nile,Kitty	100%
1.3	Project Scoping - Expert Interviews	KC,Nile,Kitty	100%
1.4	Project Proposal	KC,Nile,Kitty	100%
1.5	In-Depth Technology Specifications	KC,Nile	100%
1.6	E2E Customer Journey	KC,Kitty	100%
2	**MVP 1 - Sprint 1**		
2.0	Facial Expression Model - Data Collection	KC,Nile	100%
2.1	Facial Expression Model - Training/Testing	KC,Nile	40%
2.2	Facial Expression Model - Quality Testing	KC,Nile	0%
2.3	SER Model - Data Collection	KC,Nile	100%
2.4	SER Model - Training/Testing	KC,Nile	75%
2.5	SER Model - Quality Testing	KC,Nile	0%
2.6	UX Design (Fusion360)	KC,Kitty	40%
3	**MVP 1 - Sprint 2**		
3.1	Parts Procurement	KC,Kitty	0%
3.2	Jetson Orin/Nano Test	KC,Nile,Kitty	0%
3.3	MVP 1 Integration	KC,Nile,Kitty	0%
3.4	Market Validation	KC	0%
3.5	Final Presentation Preparations	KC,Nile,Kitty	0%

Figure 1: Project GANTT Chart

Overall, the project progress has taken the correct trajectory ever since the project proposal was submitted. The deliverables from September, including project scoping and specifications for the project have all been completed. The deliverables for October, on the other hand, have been almost completed, but have not been refined to perfection as expected. However, a working prototype for both facial expression recognition and speech emotion recognition are now available, but room for improvement still exists. Lastly, the UX design in Fusion360 is almost half done, expecting a finish within early November.

3 Facial Expression Recognition

The development of the Pookie AI-driven robot incorporates a crucial element: detecting and interpreting user emotions, particularly stress and anxiety, using Facial Emotion Recognition (FER). This section outlines the current progress, challenges, and future approaches for designing the emotion detection system, including relevant datasets, models, and methodologies.

3.1 As-is Behavior

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. The foundation model used was a VGGNet architecture (Figure 2), a multi layer convolutional neural network often used for feature extraction and classification tasks, with a notable variety of emotion detection models stemming from it.

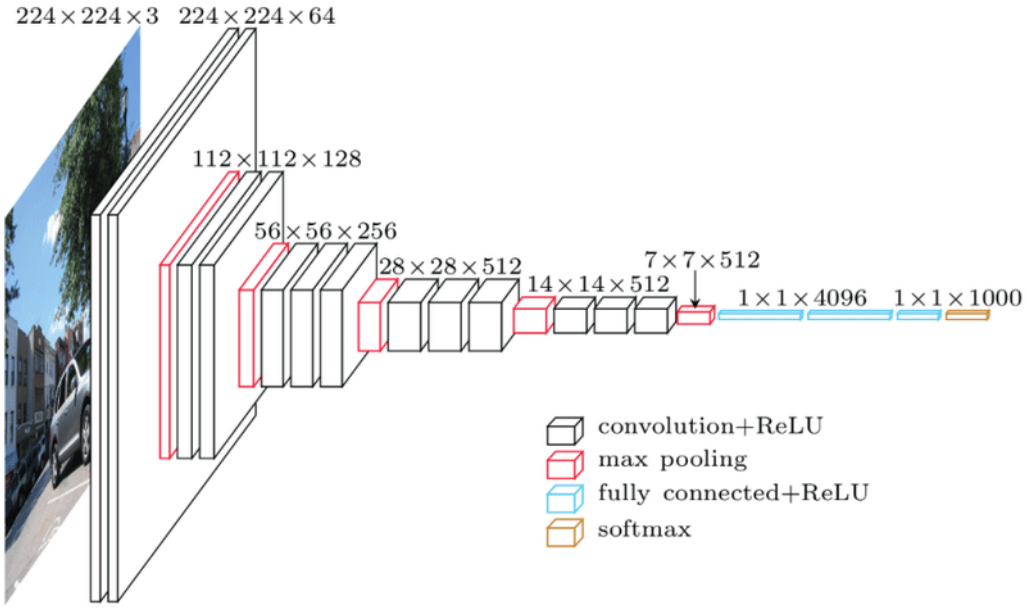


Figure 2: VGGNet Architecture

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 3, 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 3: Fine Tuning Results

3.2 Refined Approach

After consultation with our advisor, we were recommended to revise our research on facial emotion recognition, where some of the assumptions and approaches we had initially proven to be wrong. Initially, the fine tuning approach was meant to familiarize the pre-trained VGGNet with a new dataset that could be ignored or underrepresented. However, the use case of fine tuning and transfer learning requires careful consideration. Thus, instead of fine tuning the model over the Chinese dataset, we instead included the Chinese dataset in the training and validation sets, where instead of using pre-trained weights, we trained an emotion detection model following the VGGNet architecture from scratch, which yielded significantly better results, as shown in Figure 4.

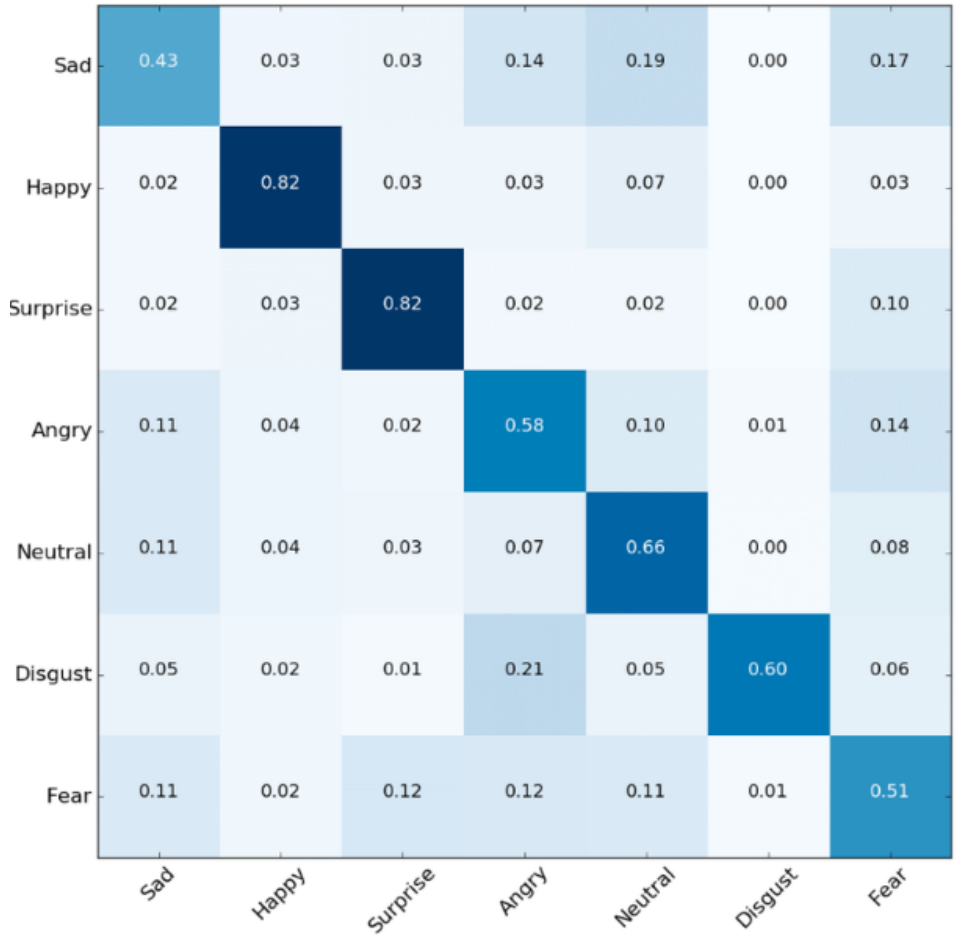


Figure 4: Refined Approach Results

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.

3.3 Next Steps

The next steps for facial emotion recognition is to integrate this model with the SER server to have a common ground for communication and output. Although the model is not perfect, it is viable enough to be used for a prototype for the rest of the semester, where other parts will be prioritized from now on.

4 Speech Emotion Recognition

Another 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 current progress, challenges, and future approaches for designing the speech emotion recognition model, including relevant datasets and methodologies.

4.1 Uvicorn Server

Significant progress has been achieved in the backend infrastructure development with the successful implementation of the Uvicorn server. The server now effectively facilitates HTTP-based communication with the Speech Emotion Recognition (SER) model. This implementation has notably enhanced our testing capabilities by enabling real-time inference processing.

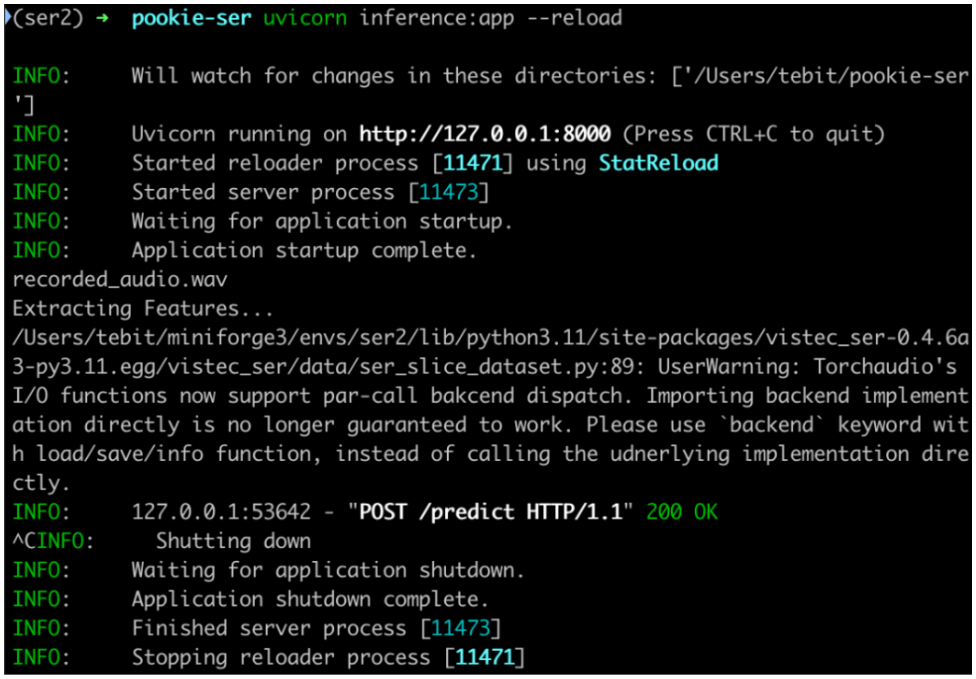


Figure 5: Local Uvicorn Server

4.2 Simple Client

A simple client application has been successfully developed with the implementation of essential functionalities. The terminal-based interface allows users to control audio recording durations through keyboard inputs. The client manages the workflow from audio capture to server communication, handling the transmission of recorded files and displaying emotion recognition results as they are processed.

Figure 6: Simple Client

4.3 Results

Performance testing of the SER model has yielded mixed results. While the system demonstrates exceptional computational efficiency, achieving inference times between 16ms to 36ms when tested on an M1 MacBook Air, the accuracy of emotion recognition requires further refinement. Initial validation testing, conducted using a sample audio clip from timestamp of a reference emotional speech video, revealed significant discrepancies between expected and actual emotion classifications. The audio segment, which clearly exhibits characteristics of sadness, produced inconsistent recognition results. Several factors may contribute to this performance gap, including the model's training on dramatized voice datasets, variations in team members' vocal characteristics, or potential configuration issues in the local deployment environment. These findings indicate a need for model optimization and further investigation into the impact of different voice characteristics on recognition accuracy.

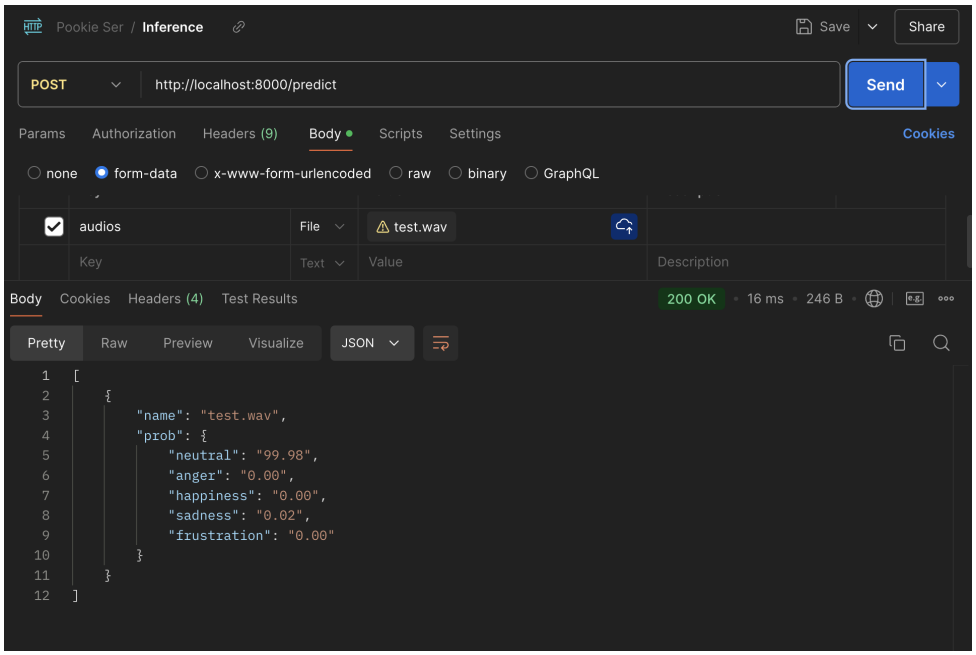


Figure 7: 16ms Inference Time

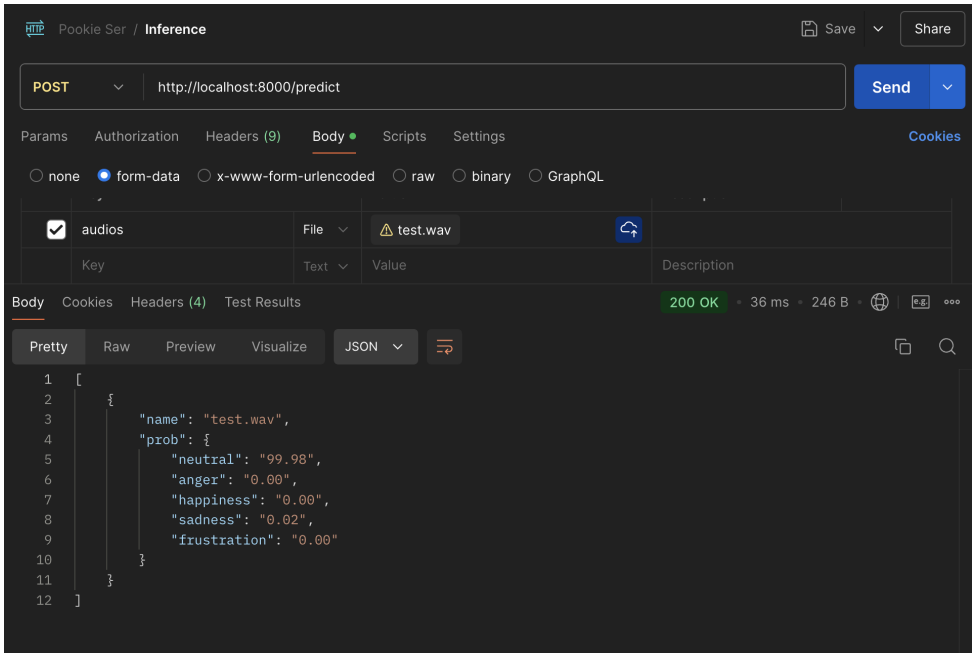


Figure 8: 36ms Inference Time

4.4 Future Actions and Implementation

The completion of the initial SER system implementation has highlighted several critical areas for future development and optimization. Primary among these objectives is the integration of the Speech Emotion Recognition system with the existing Facial Emotion Recognition (FER) framework through our server infrastructure. Additionally, a comprehensive investigation into the current model's performance discrepancies has been prioritized. This investigation will focus on analyzing the underlying causes of prediction inaccuracies and implementing necessary improvements to enhance the model's reliability.

5 Robot Design

5.1 Fusion360 CAD Implementation

Significant progress has been achieved in the hardware development phase of the Pookie robot. The outer shell design, now at 80% completion, has been crafted using Fusion360 CAD software, ensuring precise dimensional accuracy and manufacturability. This development phase has focused on creating a structurally sound and aesthetically cohesive external framework.

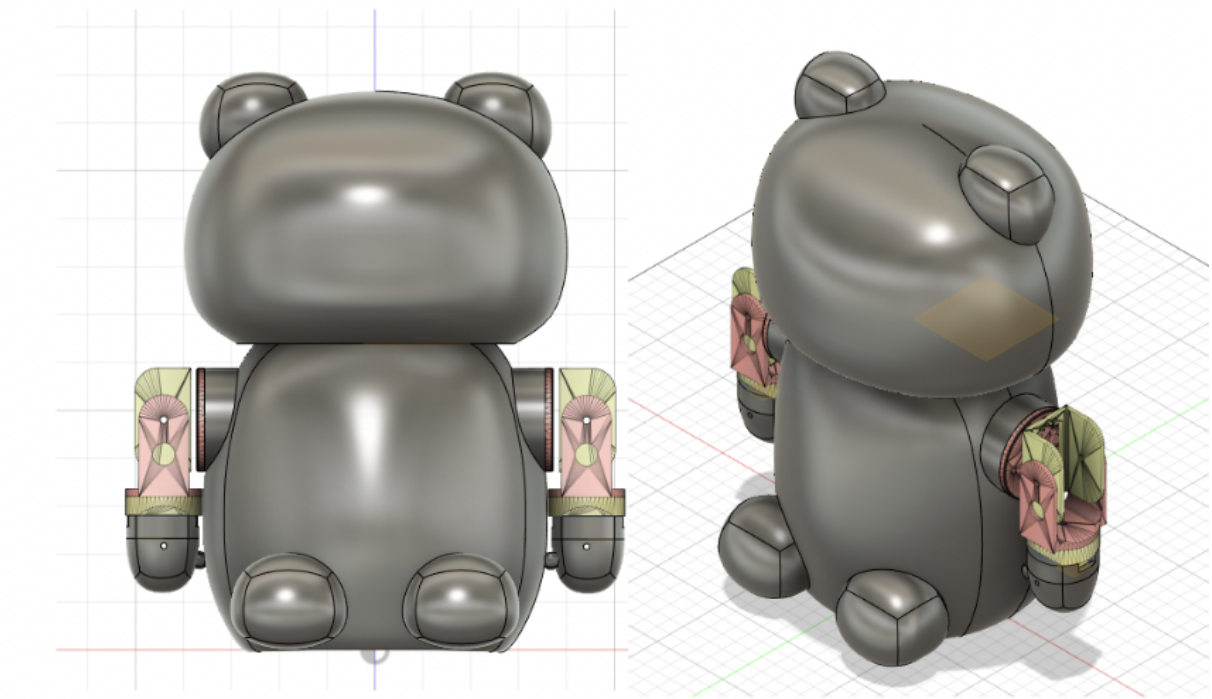


Figure 9: Pookie CAD Implementation

The robotic arm mechanism represents a completed milestone in the project's development cycle. Through comprehensive mechanical analysis and iterative design refinement, the arm assembly has been successfully engineered to meet all operational requirements. The integration of DS3255 servo motors has been a crucial element in this design phase, with their specifications carefully incorporated to ensure optimal torque delivery and precise movement control.

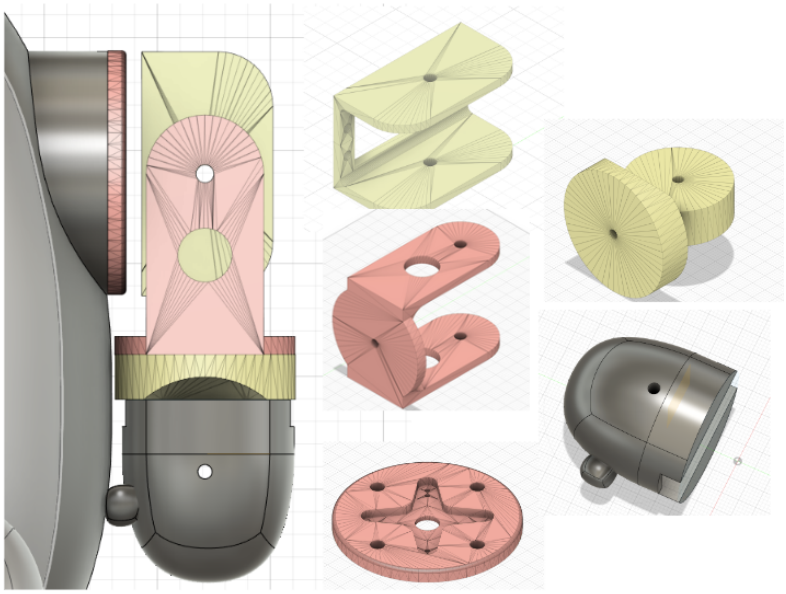


Figure 10: Pookie Robotics Arm Mechanism

The inner shell architecture remains under active development, with significant progress made on critical mechanical components. The servo motor housing for the DS3255 units has been successfully designed in Fusion360, ensuring precise mounting and optimal operational performance. However, the upper head assembly is still in the design phase, requiring further refinement. Concurrent development is underway for the integration frameworks of essential components, including sensor mounting brackets, speaker housings, and LED eye assemblies. This systematic approach to the inner architecture ensures proper component placement while maintaining the structural integrity of the design.

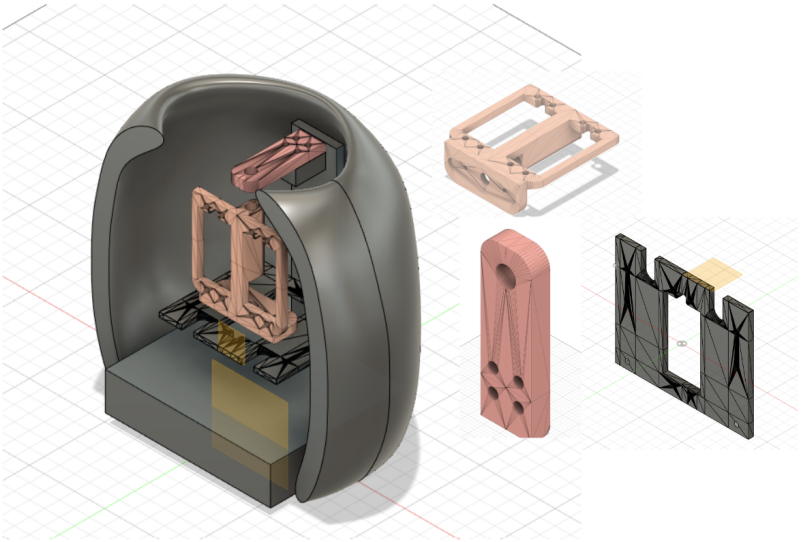


Figure 11: Pookie Inner Shell Design