

DHYAN – AI Powered Robot for Visual and Audio Sensing

Faculty of Computer Sciences and Engineering, Ghulam Ishaq Khan
Institute of Engineering Sciences and Technology, Pakistan

Supervised By:

Dr. Muhammad Hanif

Co Supervised By:

Dr. Khurram Jadoon

Dr. Zaiwar Ali

Team Members:

- Qasim Ali – 2022414
- Ali Raza – 2022664
- Abdullah Awan – 2022274
- Sara bint Bilal – 2022529





Problem Statement

- Autistic children in Pakistan lack access to affordable and engaging therapy sessions.^[1]
- Due to different behavior, not all children adapt to the therapy sessions quickly.

[1] <https://scientiamag.org/the-challenges-of-pakistani-parents-of-autistic-children-a-double-edged-sword/>





Proposed Solution

- 
- **AI Assisted Therapy**

- **Emotional Recognition & Gesture Commands**

- **Speech Recognition**

- **Real time User Profiling**

Requirements Gathering Methodology



Literature Review

Studied autism-focused robotics and therapy research.

Explored Socially Assistive Robots (SAR) for ASD, emotion recognition, and human–robot interaction models



Existing Systems Analysis

Analyzed NAO, Kaspar, Milo robots and autism support apps.

Examined interaction methods, sensing capabilities, and limitations.



Technical Decomposition

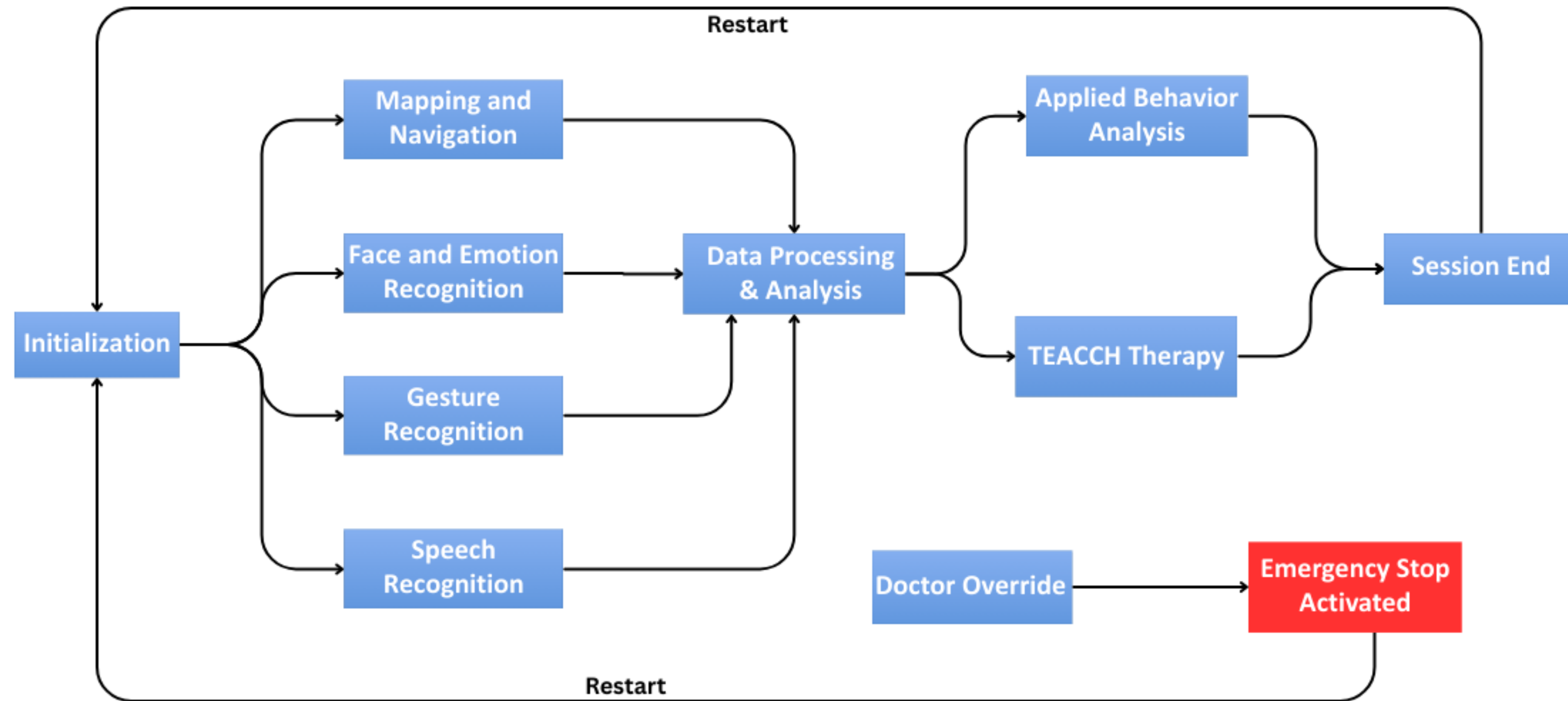
System requirements were defined by breaking the solution into core modules (Audio, Vision, AI Agent, Dashboard) following published design frameworks



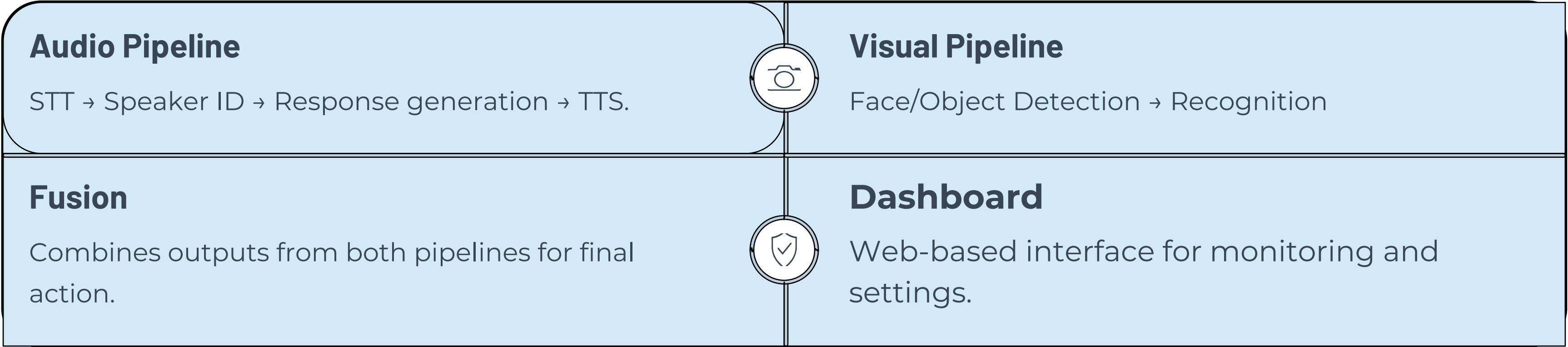
Stakeholder Analysis

Identified diverse user groups from different industries to understand their unique needs and constraints

Proposed System Design



Project Scope & Core Modules



Functional Requirements

**Visual Perception
Feature (Face
Detection, Gaze
Tracking)**

**Audio Sensing &
Voice Activity
Detection (VAD)**

**Joint Attention
Module (sharing
focus)**

Imitation Module

**Emotion
Recognition**

**Therapist Console
& Session Control**

**Logging, Storage
& Data
Management**

Non-Functional Requirements

**Performance
Requirements**

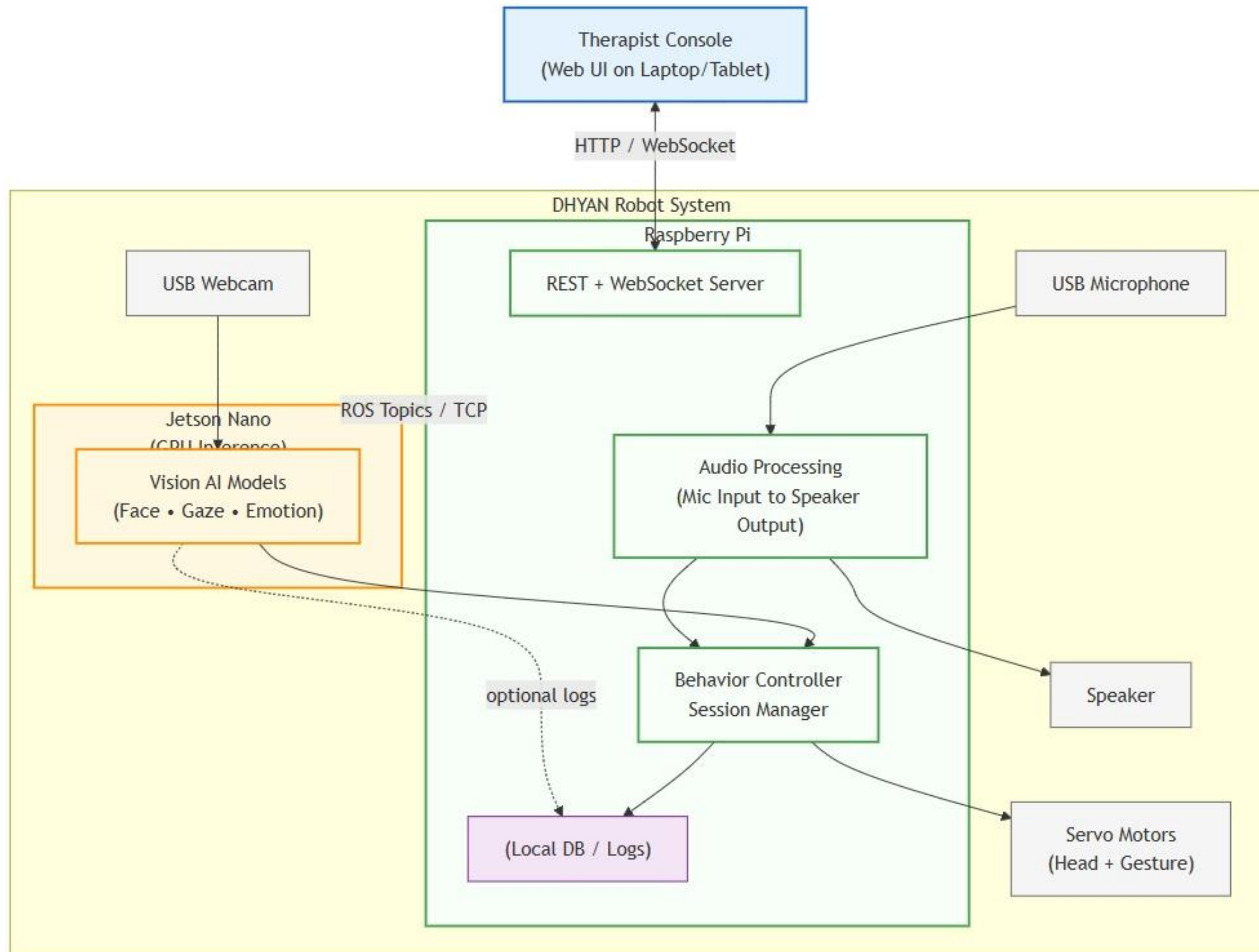
**Safety
Requirements**

**Security
Requirements**

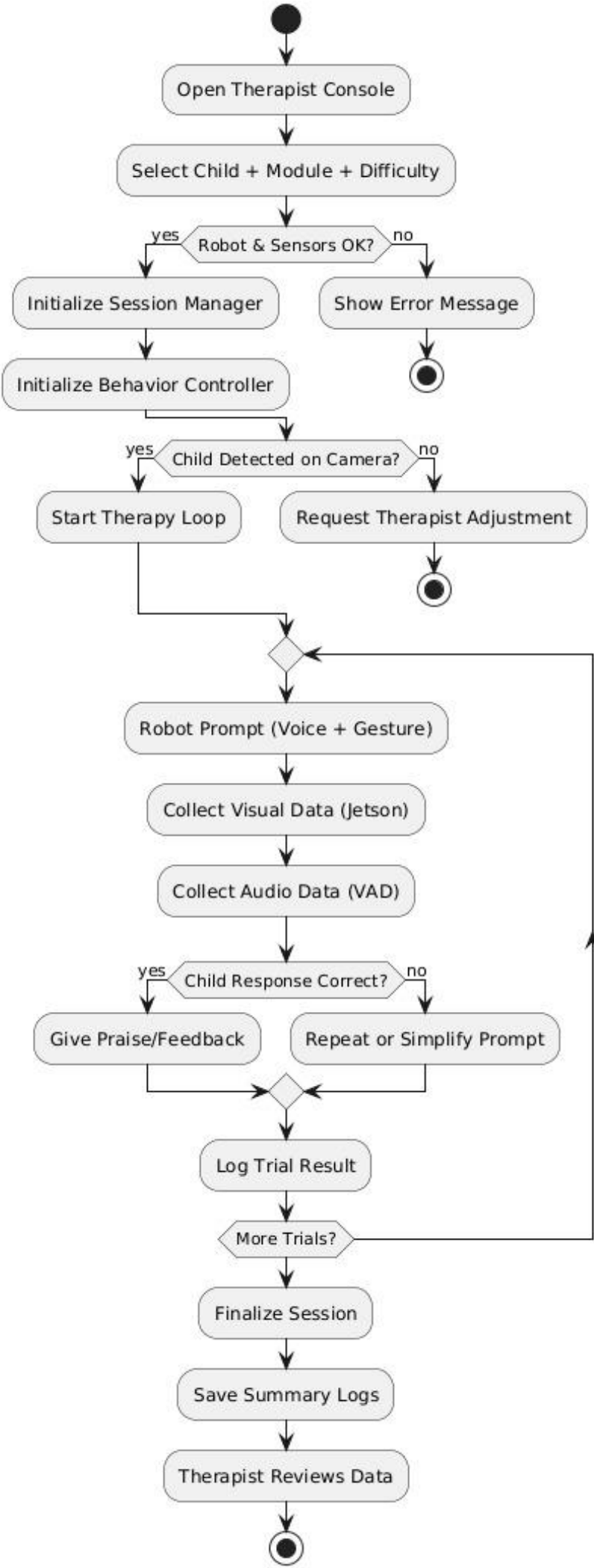
**Software
Quality
Attributes**

**Business
Rules**

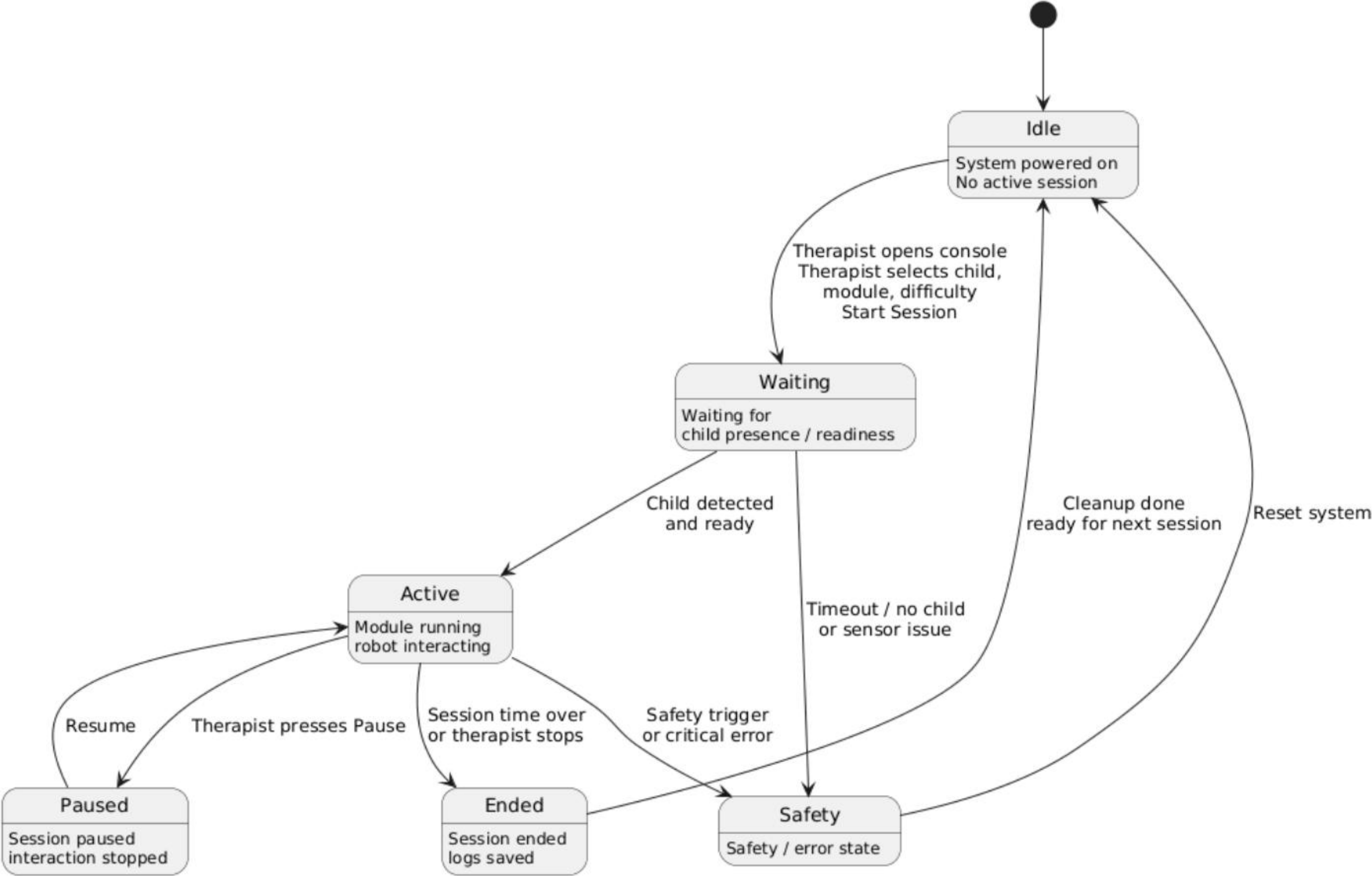
System Architecture Design



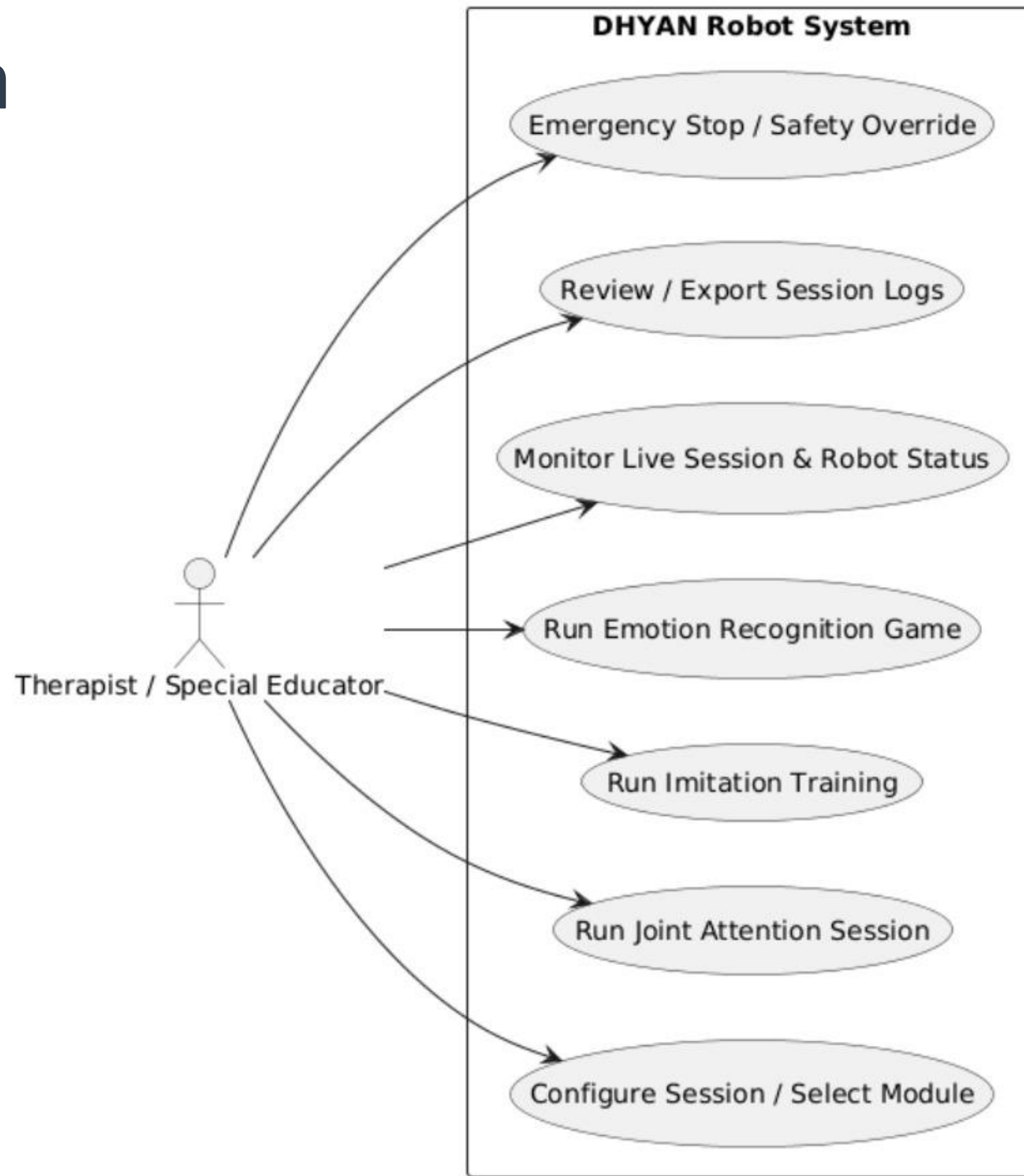
Activity Diagram



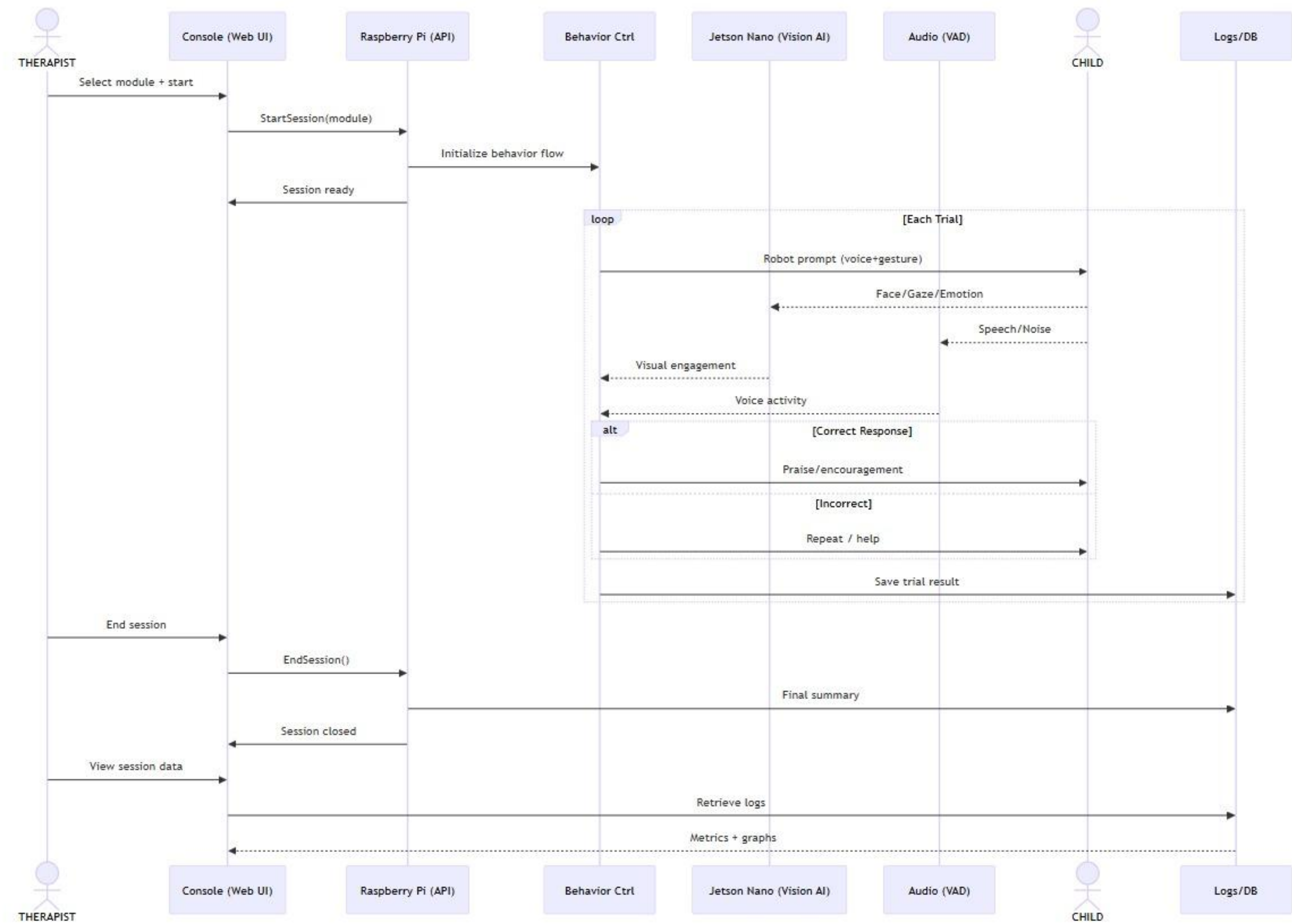
State Machine Diagram



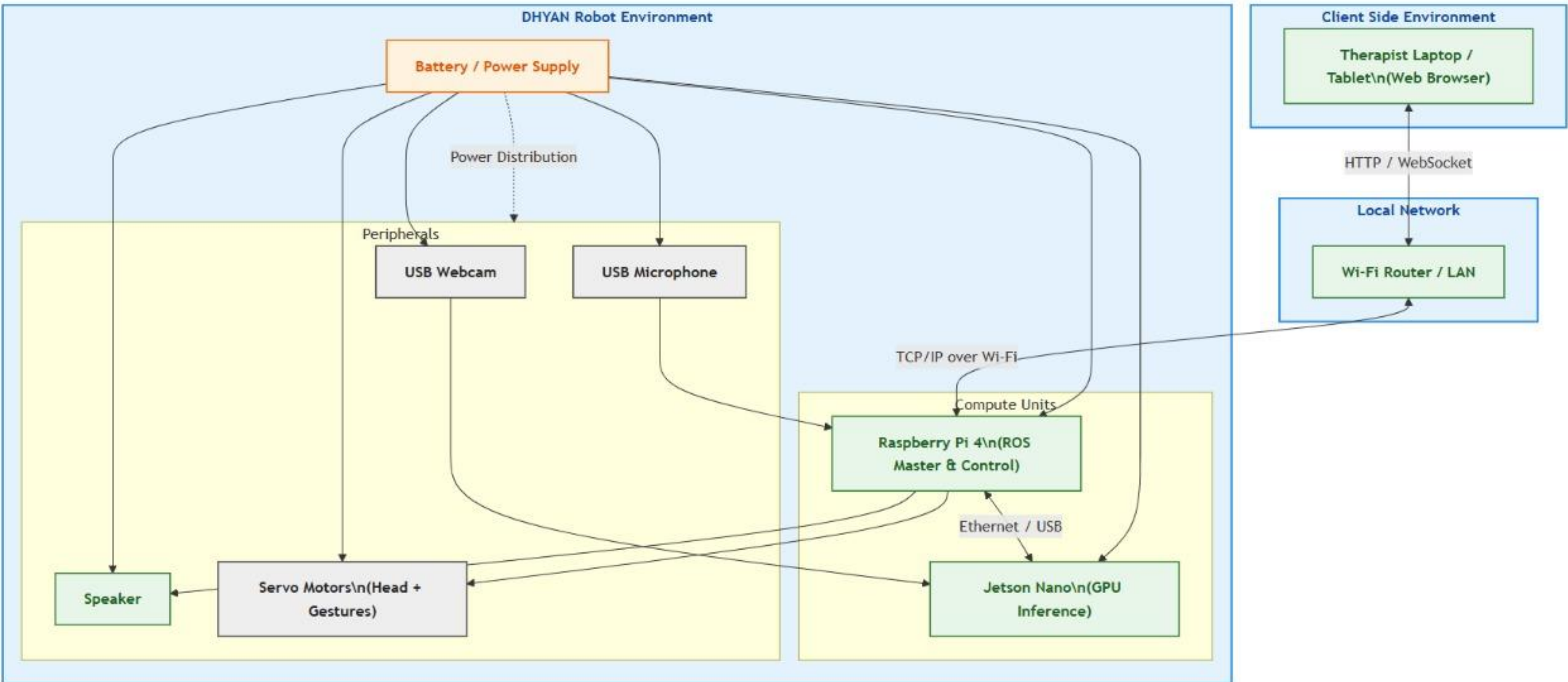
Use Case Diagram



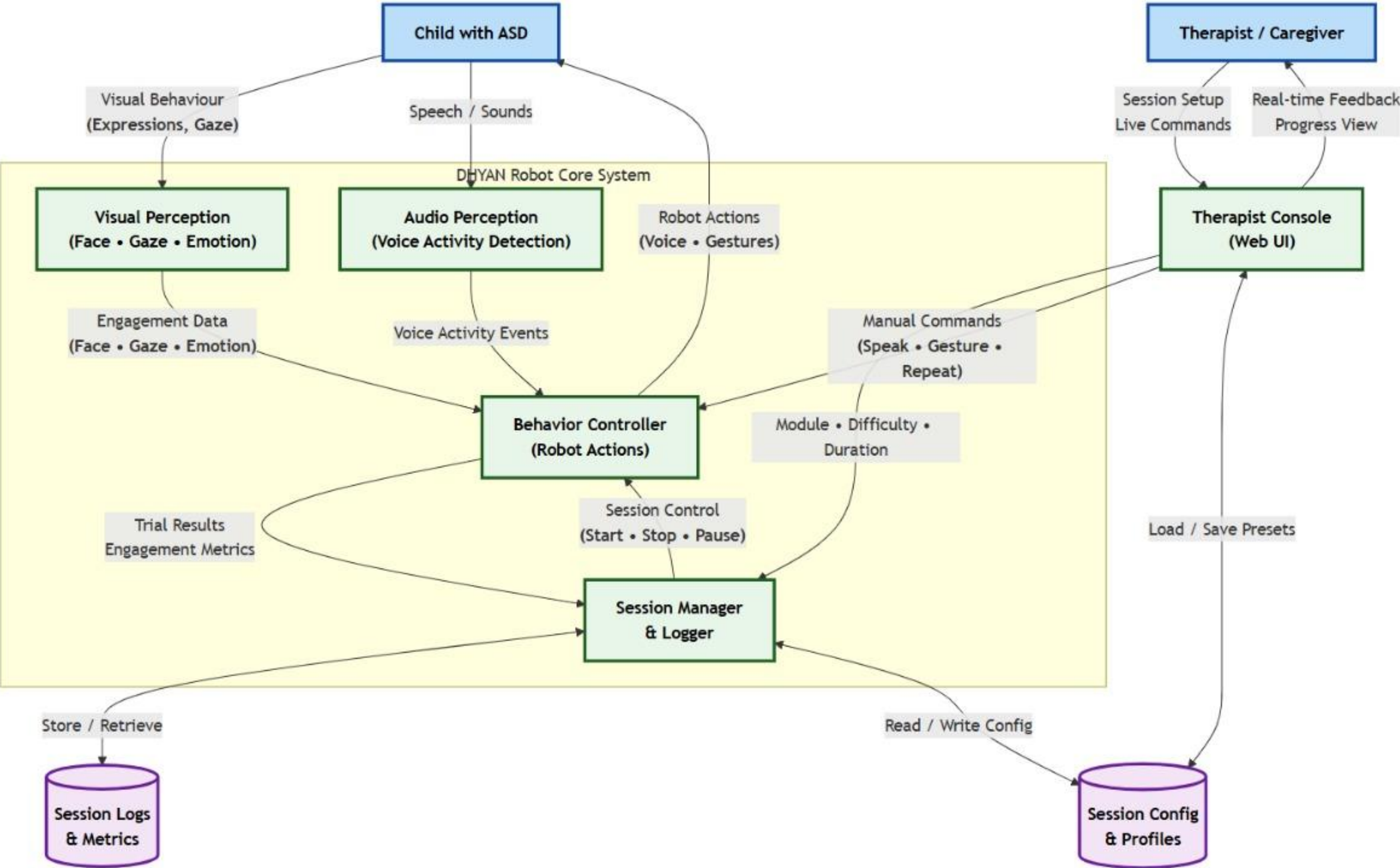
Sequence Diagram



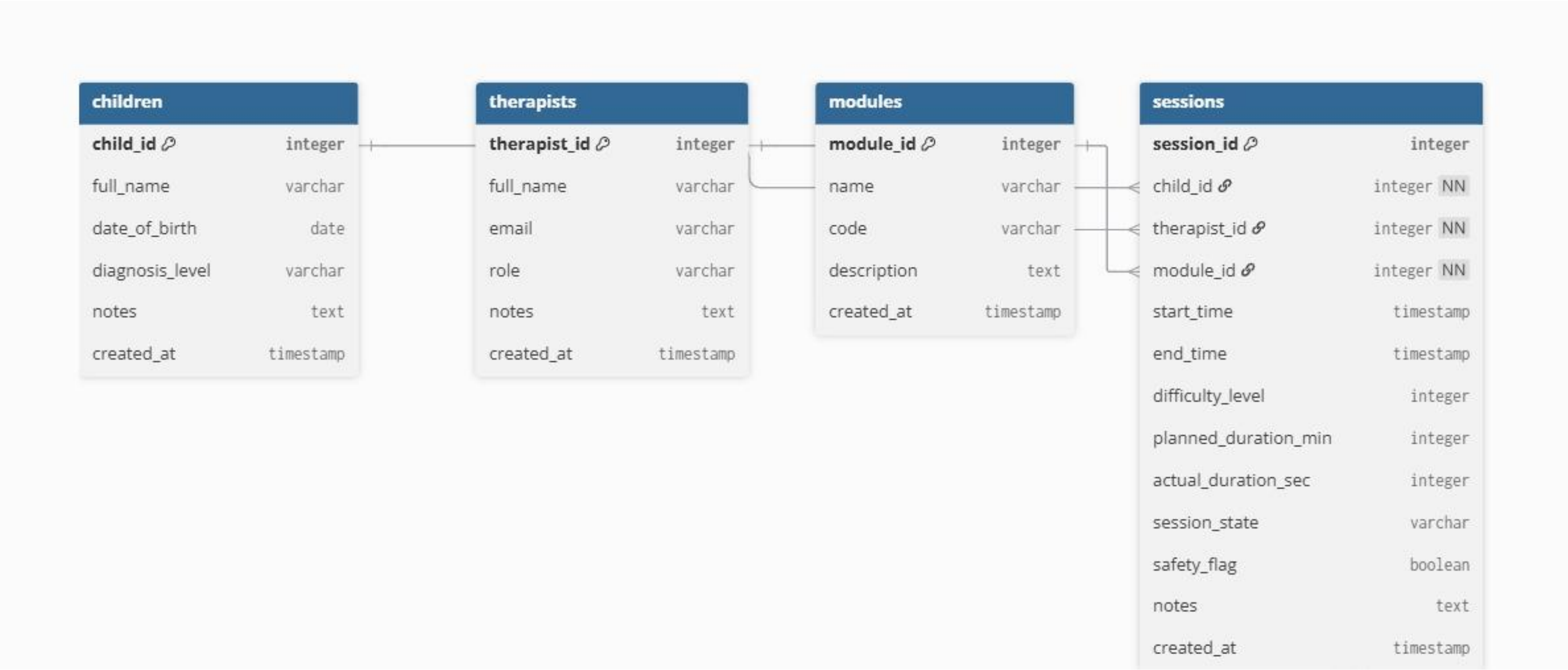
Deployment Diagram



Dataflow Diagram



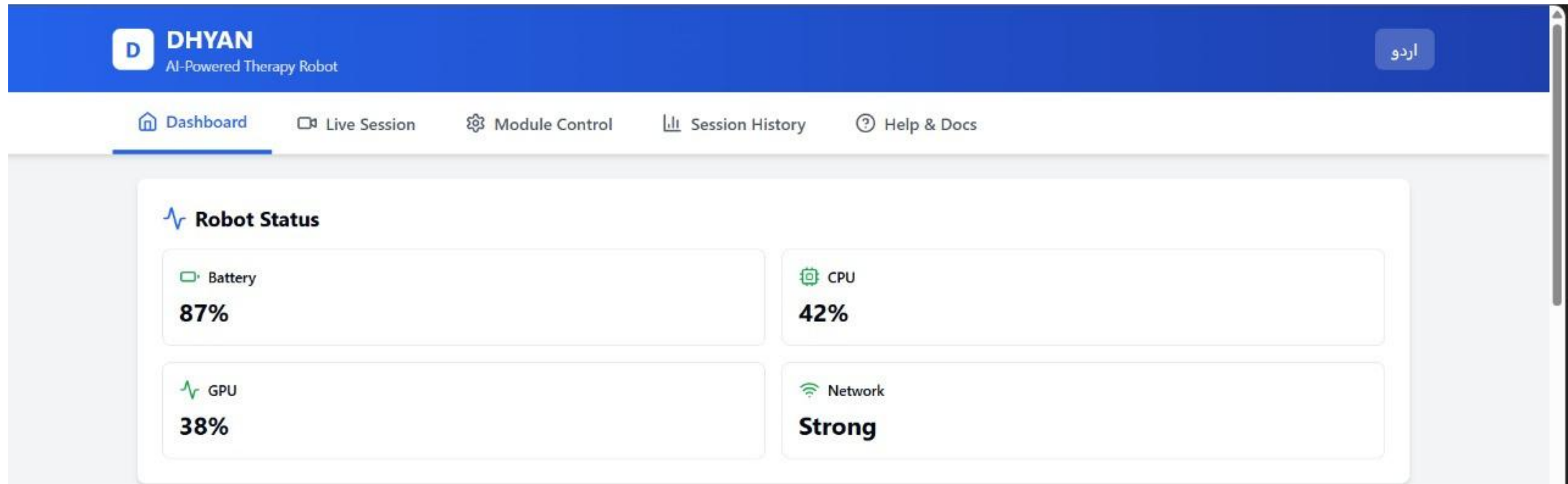
Entity Relationship Diagram



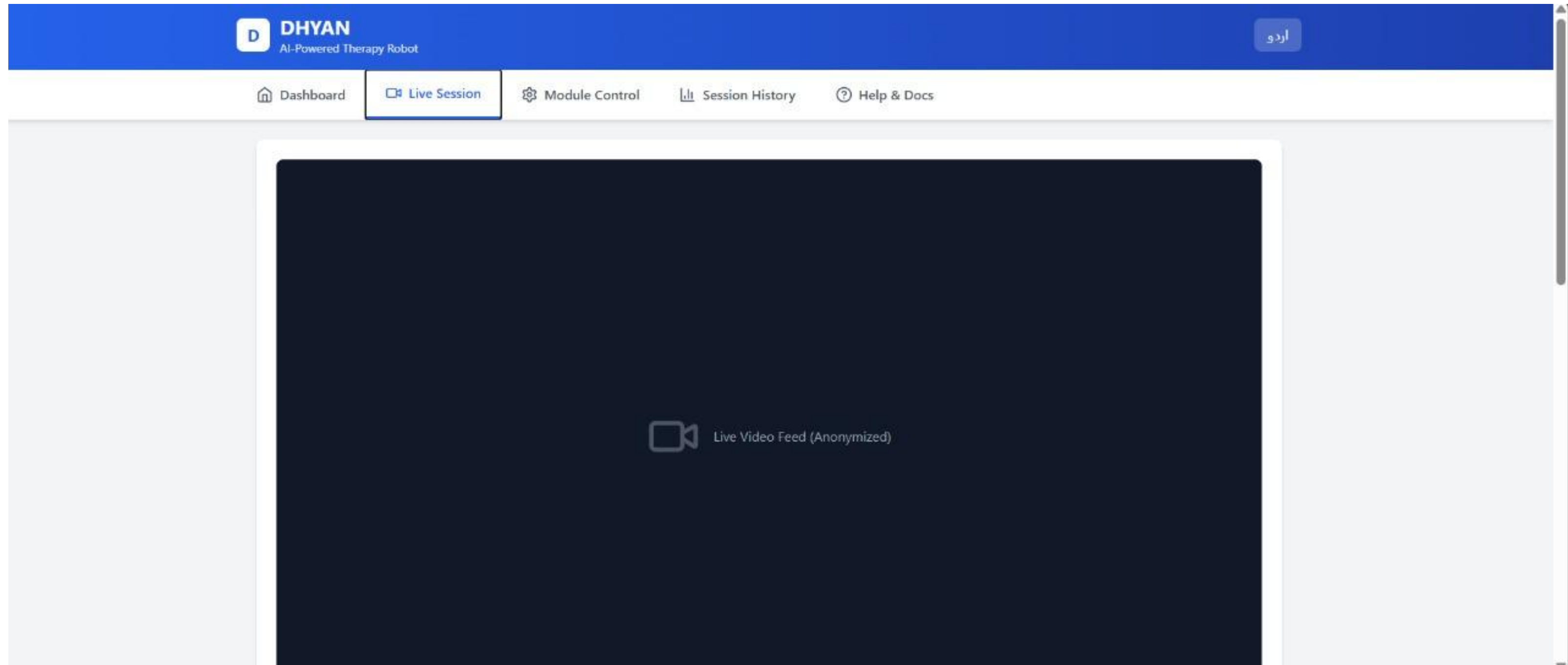
Proposed AI/ML Models

Model	Purpose	Input	Output	Advantage
Whisper	Speech-to-text	Raw audio	Text transcript	Fast, accurate STT
MFCC + CNN	Speaker Identification	Audio features	User ID	Lightweight, Pi-compatible
GTTS	Text-to-speech	Text	Audio output	Fast response
MTCNN	Face Detection	Video frame	Face bounding box	Lightweight, real-time
MobileFaceNet	Face Recognition	Detected face	User ID	High accuracy
YOLO-Nano	Object Detection	Video frame	Bounding boxes	Real-time detection

User Interface - 1



User Interface - 2



User Interface - 3

▶ Start Session

□ Stop Session

Live Perception Metrics

👁 Face Detection

Detected

Latency: 45ms

👁 Gaze Tracking

Focused

Latency: 62ms

😊 Emotion Classification

Happy


Latency: 120ms

🗣 Voice Activity


Speaking


Latency: 38ms


User Interface - 4


 **DHYAN**
AI-Powered Therapy Robot


اردو

 Dashboard

 Live Session

 **Module Control**

 Session History

 Help & Docs

Select Therapy Module

Therapy Module

Joint Attention

Difficulty Level

Easy

Medium

Hard

Session Duration

5 min

15 min

30 min

Prompt Style

Verbal Only

User Interface - 5

Dashboard

Live Session

Module Control

Session History

Help & Docs

Session Logs

Export Data

Joint Attention

2025-11-22 14:30

Duration

12:34

Trials

15

Success

85%

Avg Latency

1.2s

View Details

Imitation

2025-11-21 10:15

Duration

15:20

Trials

20

Success

92%

Avg Latency

0.9s

View Details

Turn-Taking

2025-11-20 16:45

Duration

10:15

Trials

12

Success

78%

Avg Latency

1.5s

View Details

Emotion Recognition

2025-11-19 11:00

Duration

13:45

Trials

18

Success

88%

Avg Latency

1.1s

View Details

Joint Attention

2025-11-18 14:20

Duration

11:50

Trials

14

Success

81%

Avg Latency

1.3s

View Details

Requirement to Design Mapping

Requirement	Design Component Addressing It
Face detection must run at ≥ 10 FPS	Jetson Nano AI inference module + Vision Node (Architecture Diagram)
Robot must stop speaking when child speaks	VAD module + TTS controller (Sequence Diagram)
Therapist must start/pause/stop session	Session Controller + Therapist Console UI (UI Design)
Robot must perform gestures smoothly	Servo Control Module (Architecture Diagram)
Emotion recognition correctness feedback	Emotion Recognition Module + Feedback Logic (State Machine)
All session data must be logged	Logging Module + Database (DFD + ERD)

Task Breakdown & allocation using project management tool

D Dhyan - FYP

Home

Planner

Brain

Teams


More

Invite

Upgrade

Home

+ Create

Click  to add favorites to your sidebar.

Spaces

All Tasks - Dhyan - FYP

Team Space

Project 1

Project 2

Get Started with ClickUp 5

Project Notes

New Space

Channels

General - Dhyan - FYP

Welcome

Add Channel

Direct Messages

Ali Raza — You

New message

Search Ctrl K

AgentsAutomateAsk AISHare

Team Space / Project 1

Add ChannelListBoardCalendarGanttTable+ View

Group: StatusSubtasksColumnsFilterClosedAssigneeAAdd Task

IN PROGRESS2

Name	Assignee	Due date	Priority	Status	Comments
Face Detection Module	MA	12/5/25	High	IN PROGRE...	
Voice Recognition	MA	12/22/25	Low	IN PROGRE...	
+ Add Task					

TO DO3

Name	Assignee	Due date	Priority	Status	Comments
Gaze Estimation Module	MA	12/21/25	Normal	TO DO	
Emotion Recognition Model	AR	12/25/25	Normal	TO DO	
Integration & Testing	SB	12/13/25	High	TO DO	
+ Add Task					

1/5

Thankyou