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.

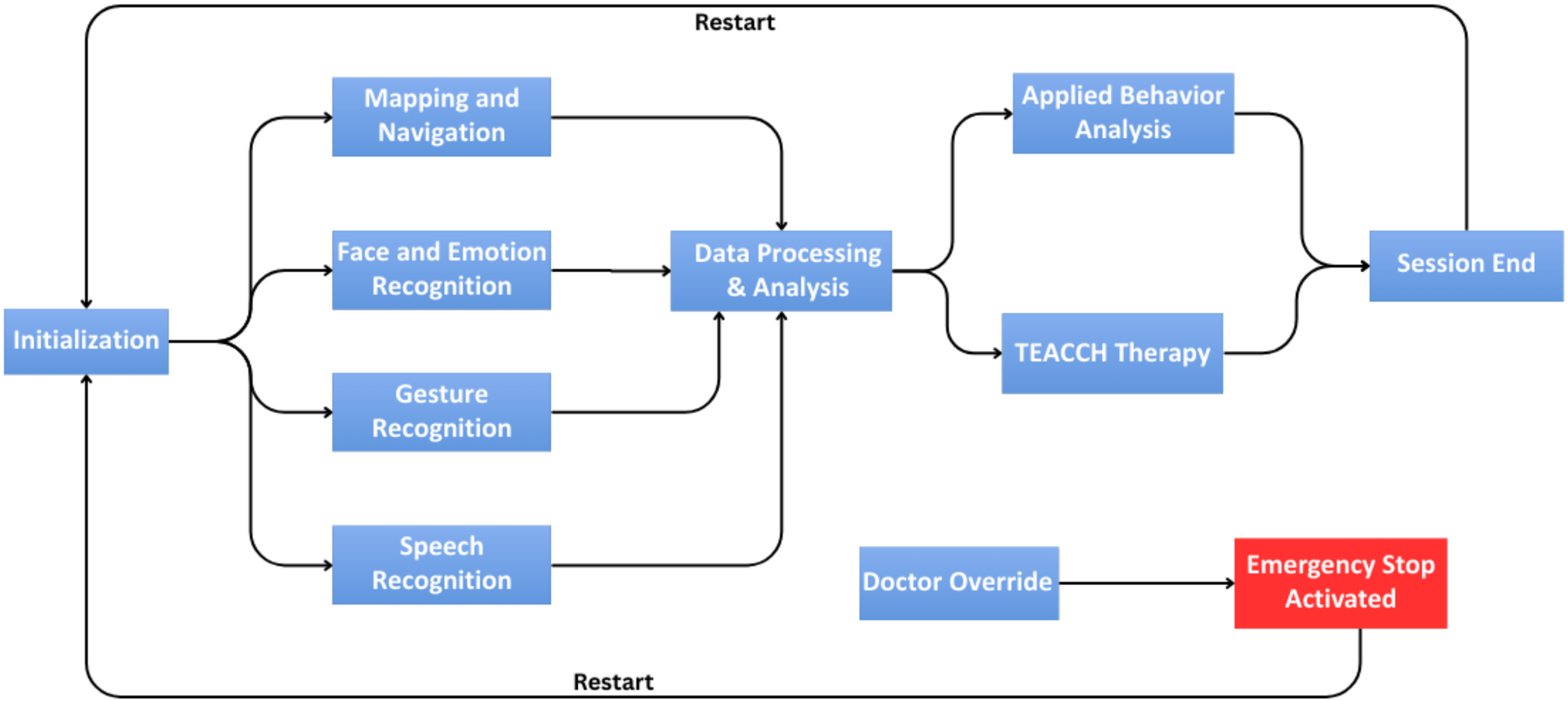
Technical Decomposition

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

Audio Pipeline

STT → Speaker ID → Response generation → TTS.

Visual Pipeline

Face/Object Detection → Recognition

Fusion

Combines outputs from both pipelines for final action.

Dashboard

Web-based interface for monitoring and settings.

Functional Requirements

Visual Perception Feature (Face Detection, Gaze Tracking)

Audio Sensing & Voice Activity Detection (VAD)

Joint Attention Module (sharing focus)

Imitation Module

Emotion Recognition

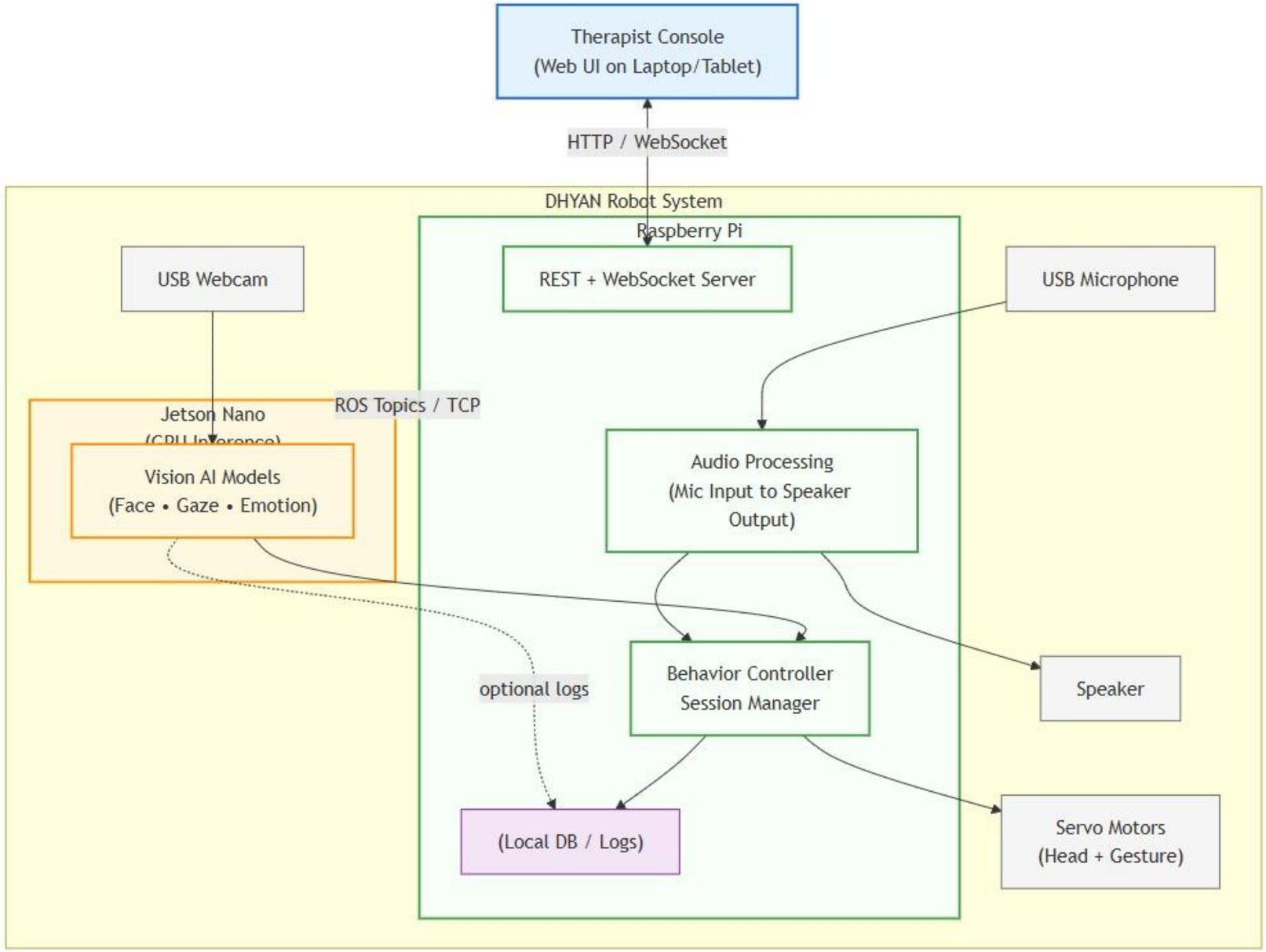
Therapist Console & Session Control

Logging, Storage & Data Management

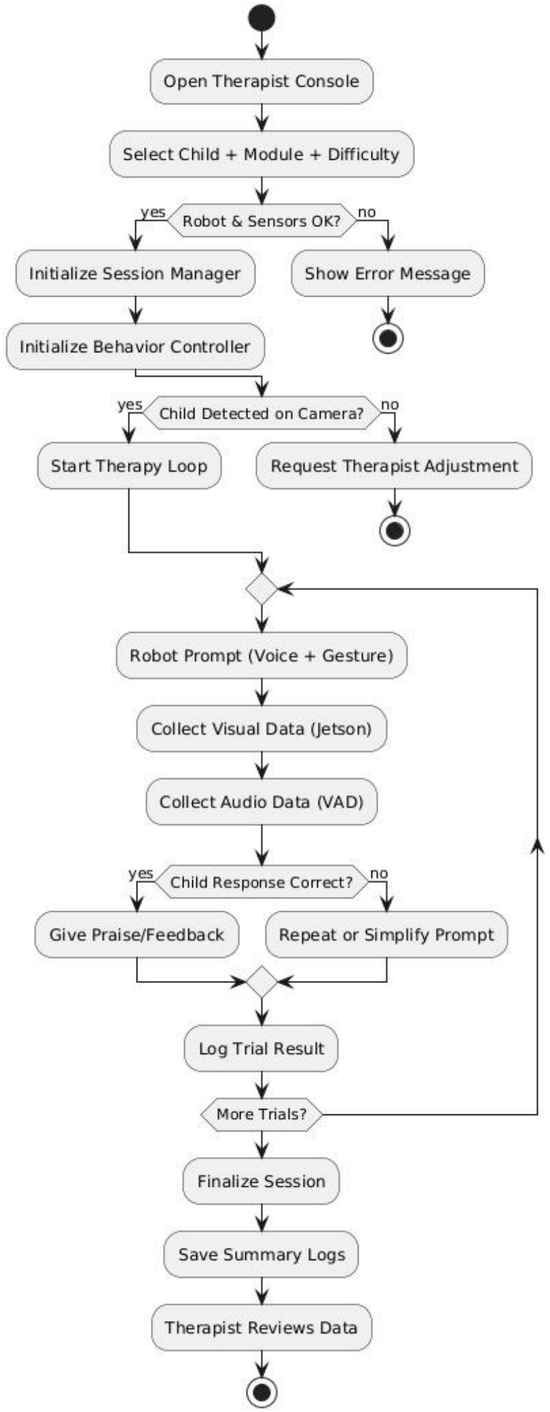
Non-Functional Requirements

- Performance Requirements
- Safety Requirements
- Security Requirements
- Software Quality Attributes
- Business Rules

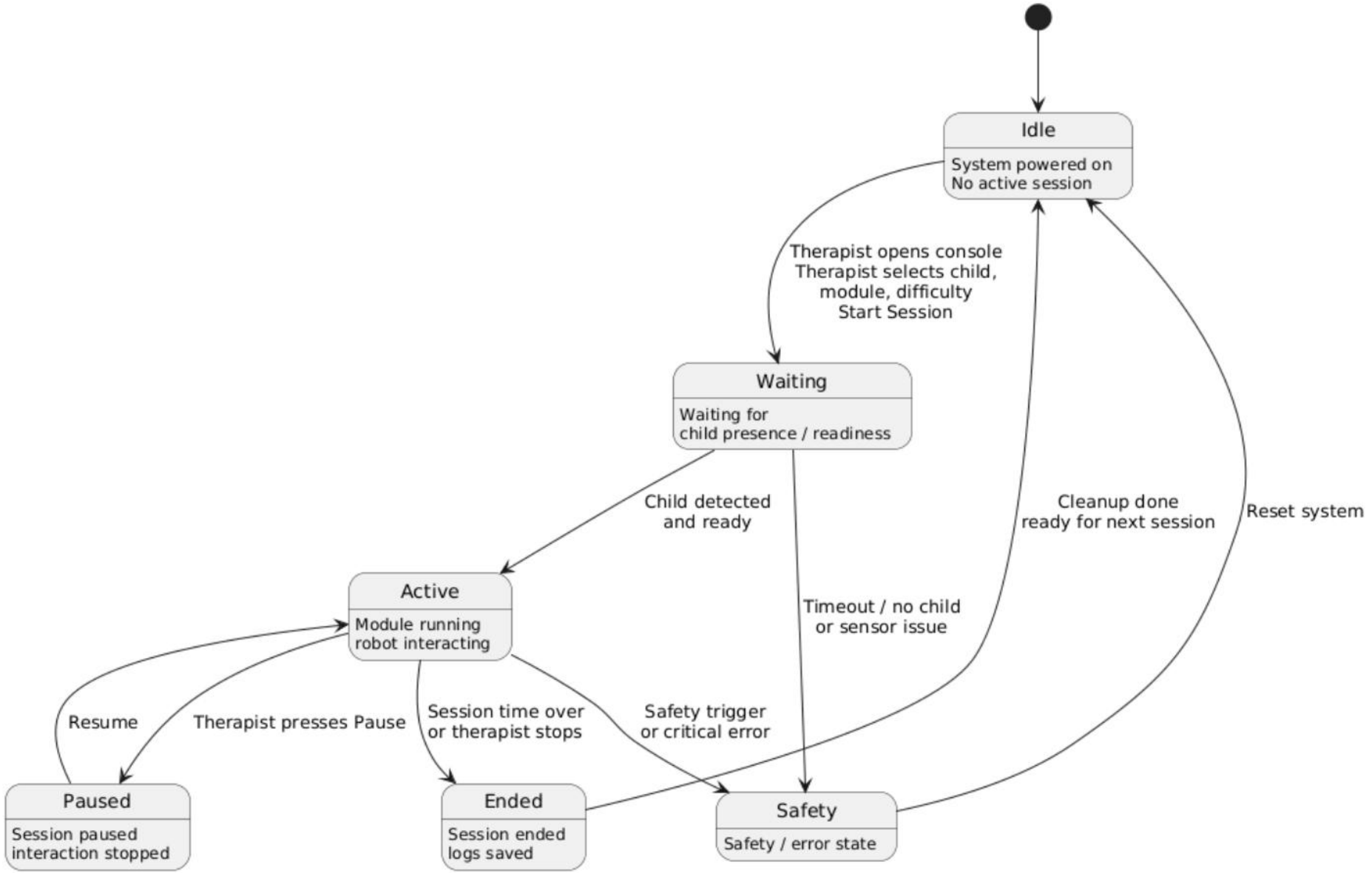
System Architecture Design



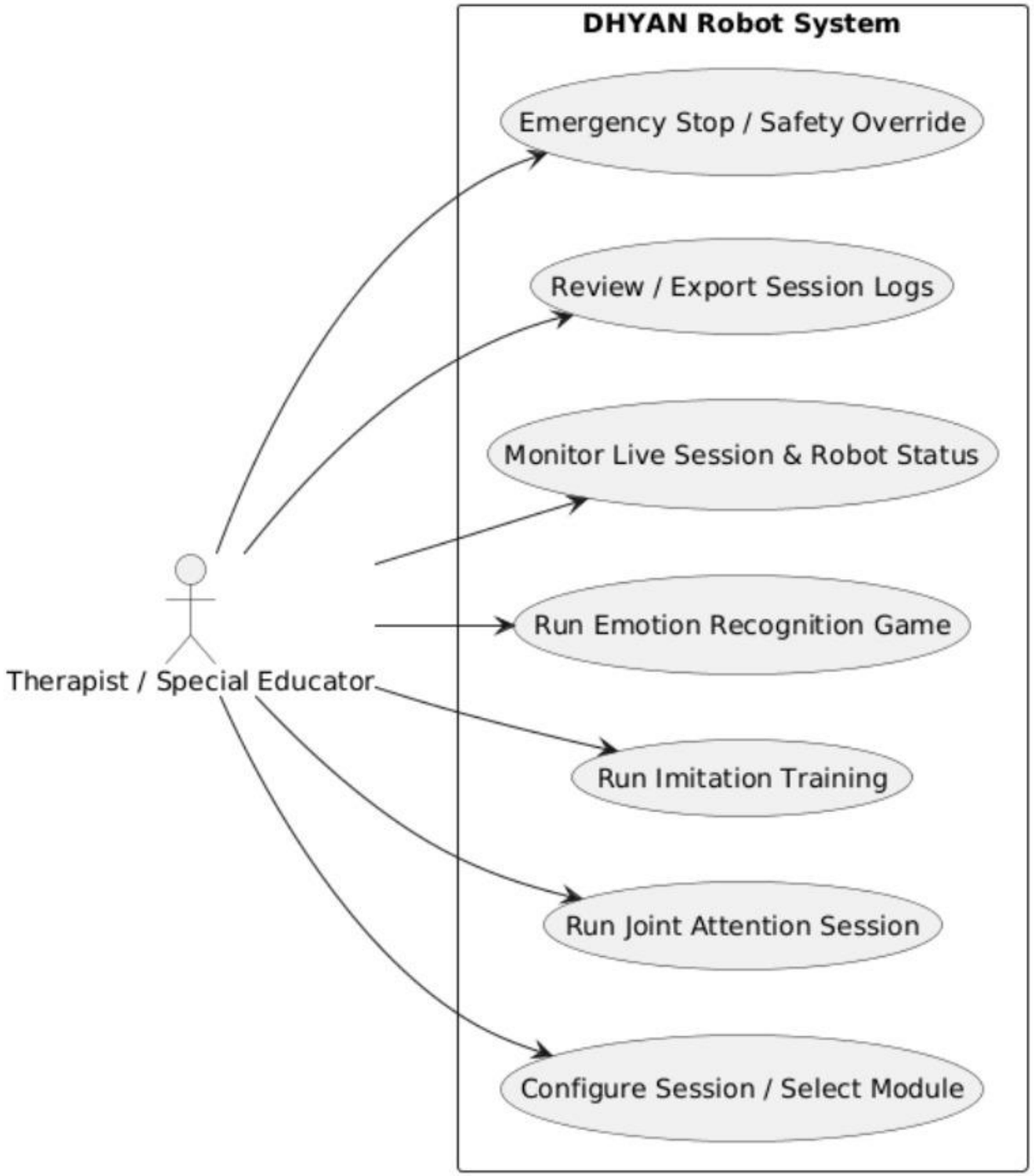
Activity Diagram



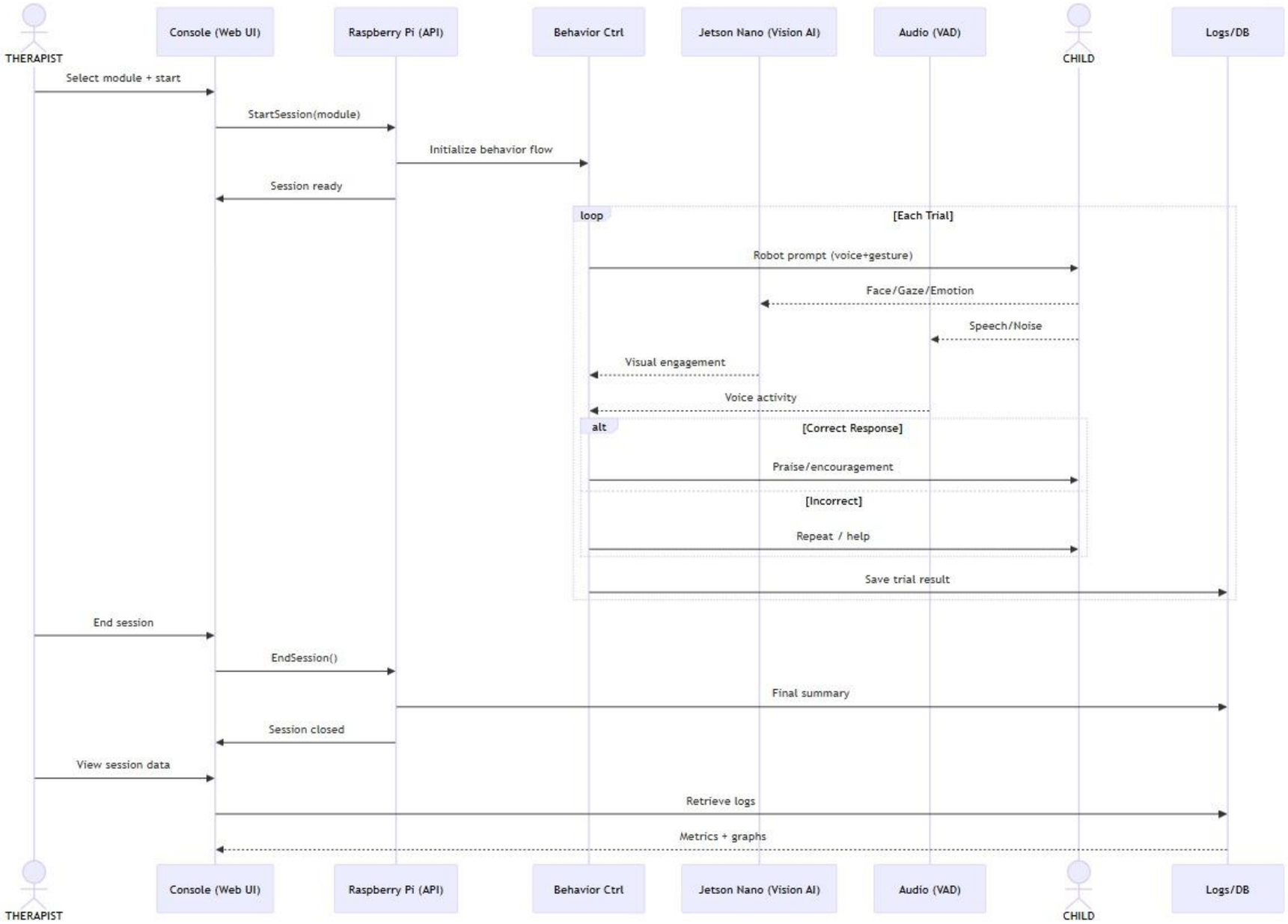
State Machine Diagram



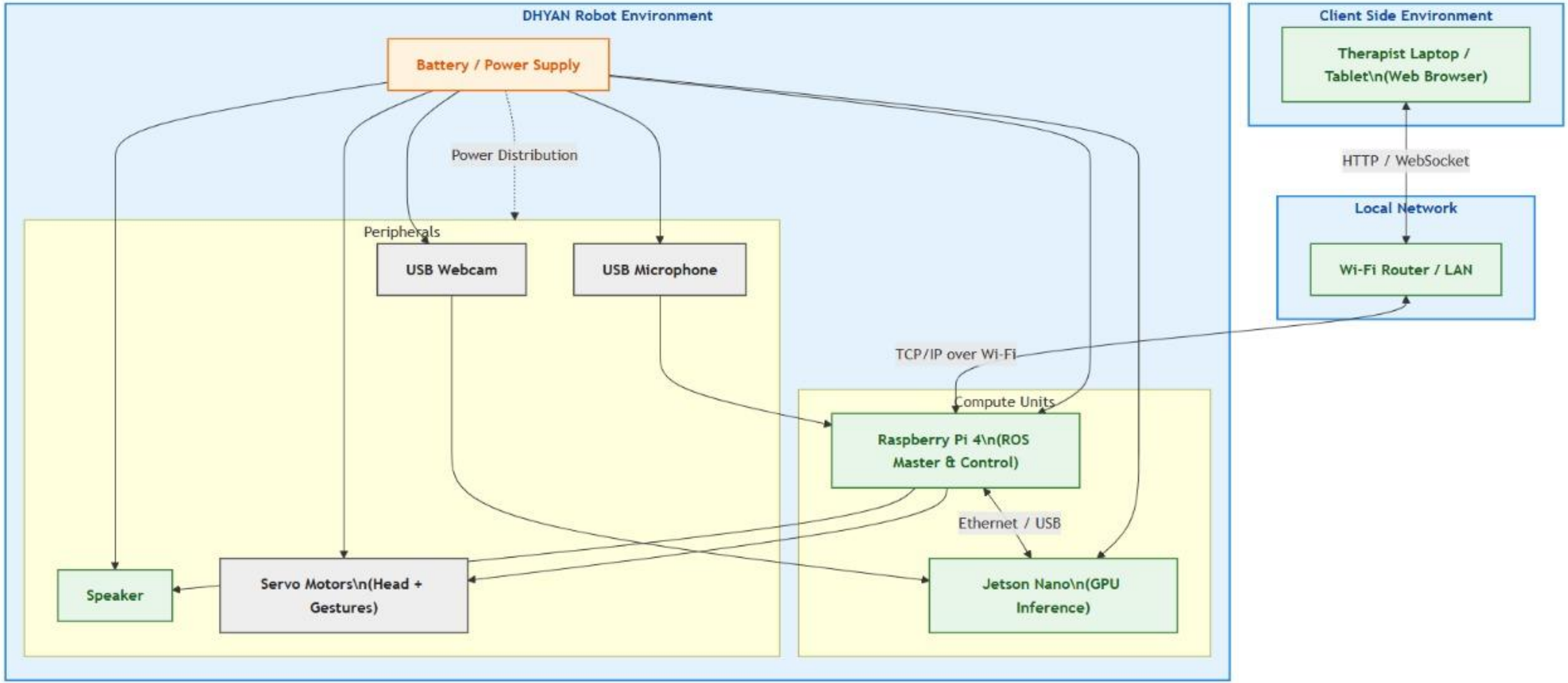
Use Case Diagram



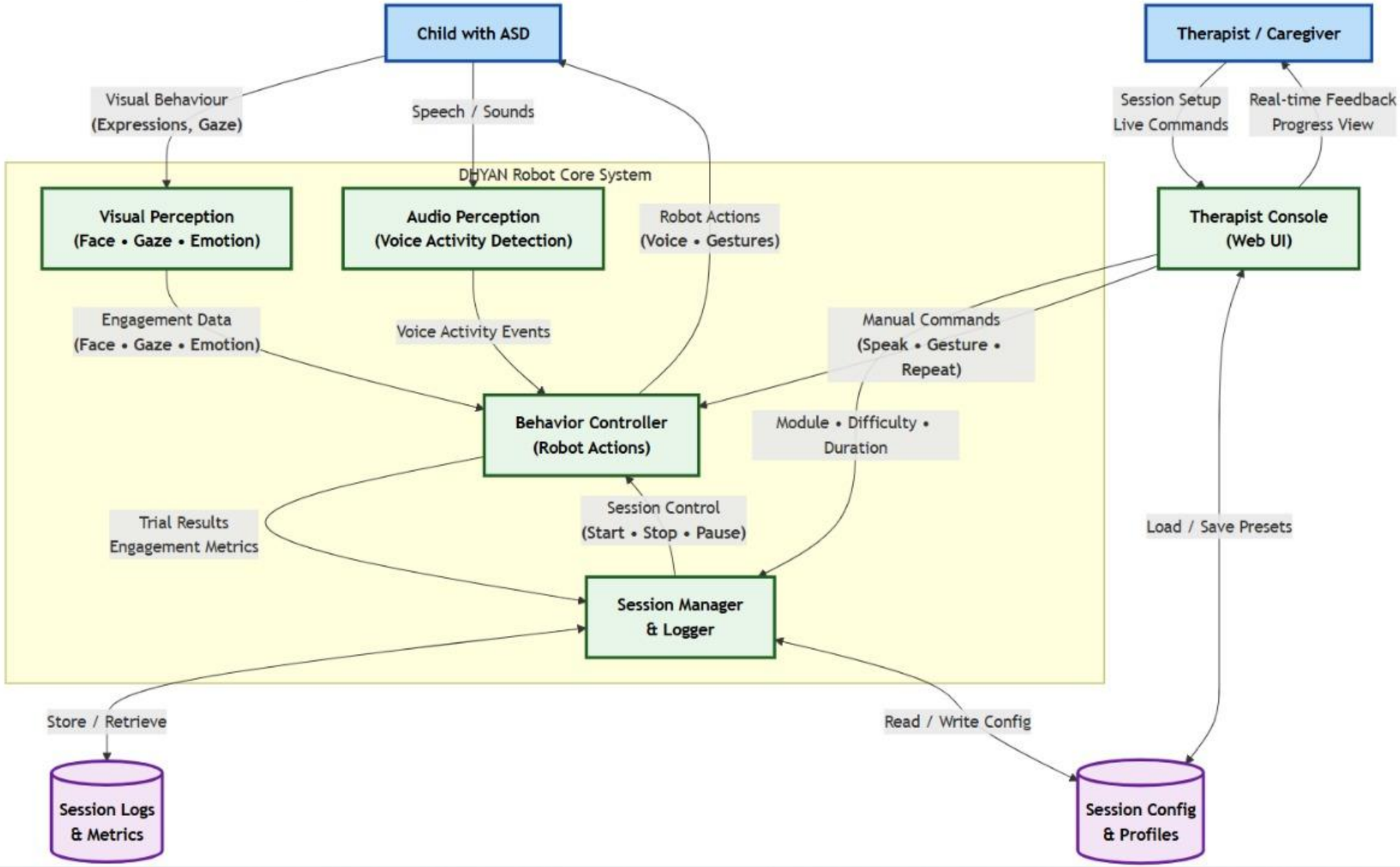
Sequence Diagram



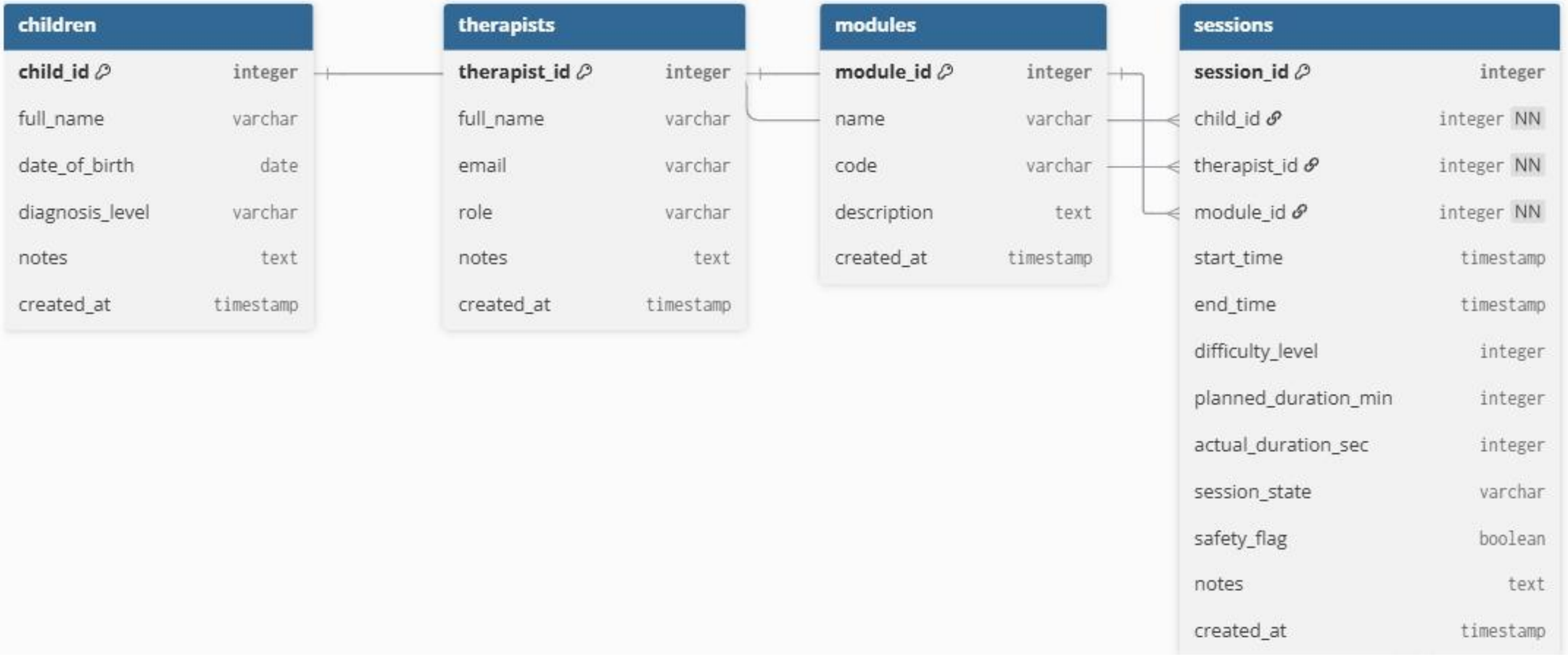
Deployment Diagram



Dataflow Diagram



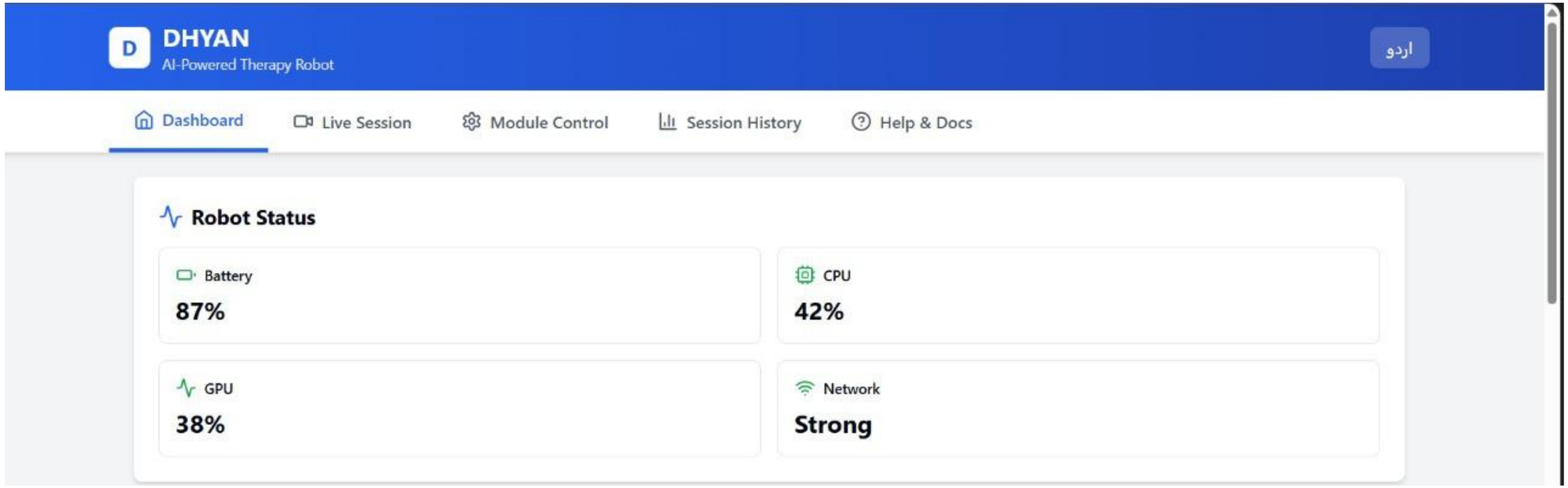
Entity Relationship Diagram



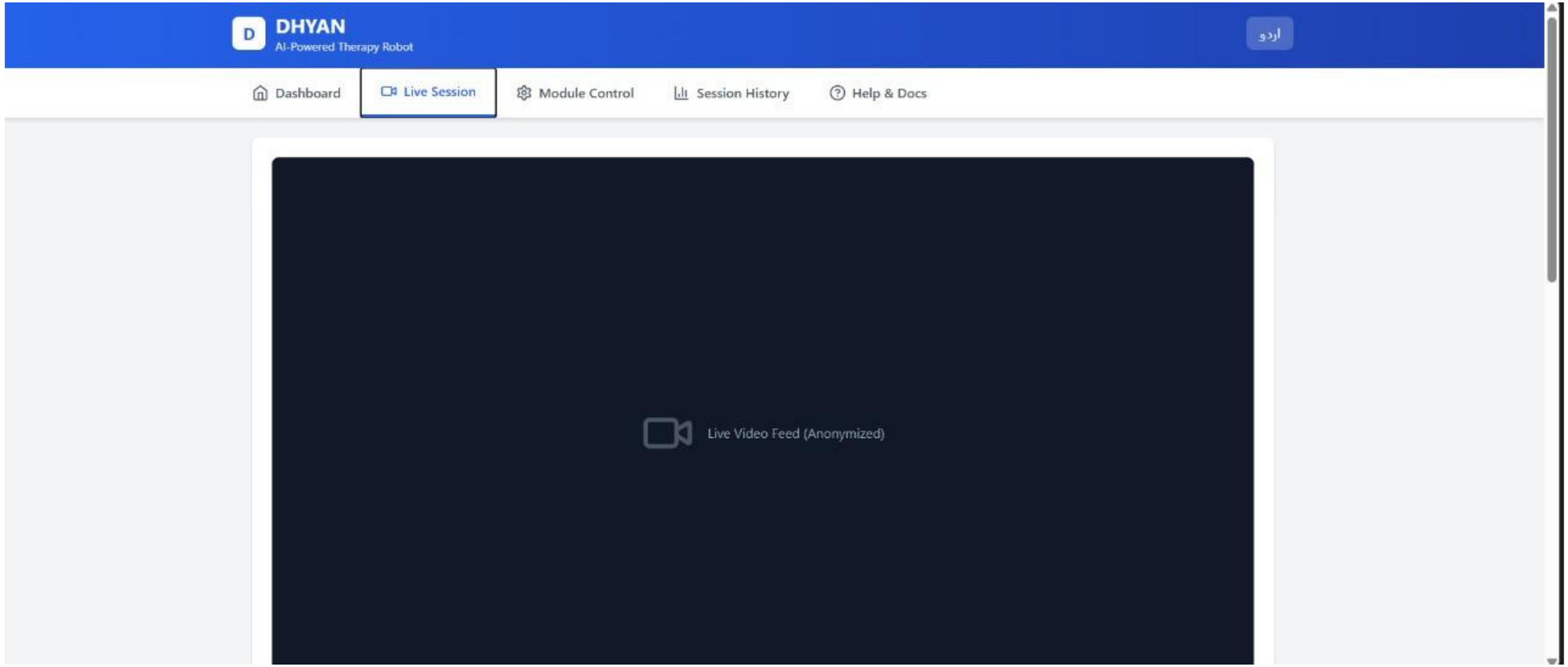
Proposed AI/ML Models

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

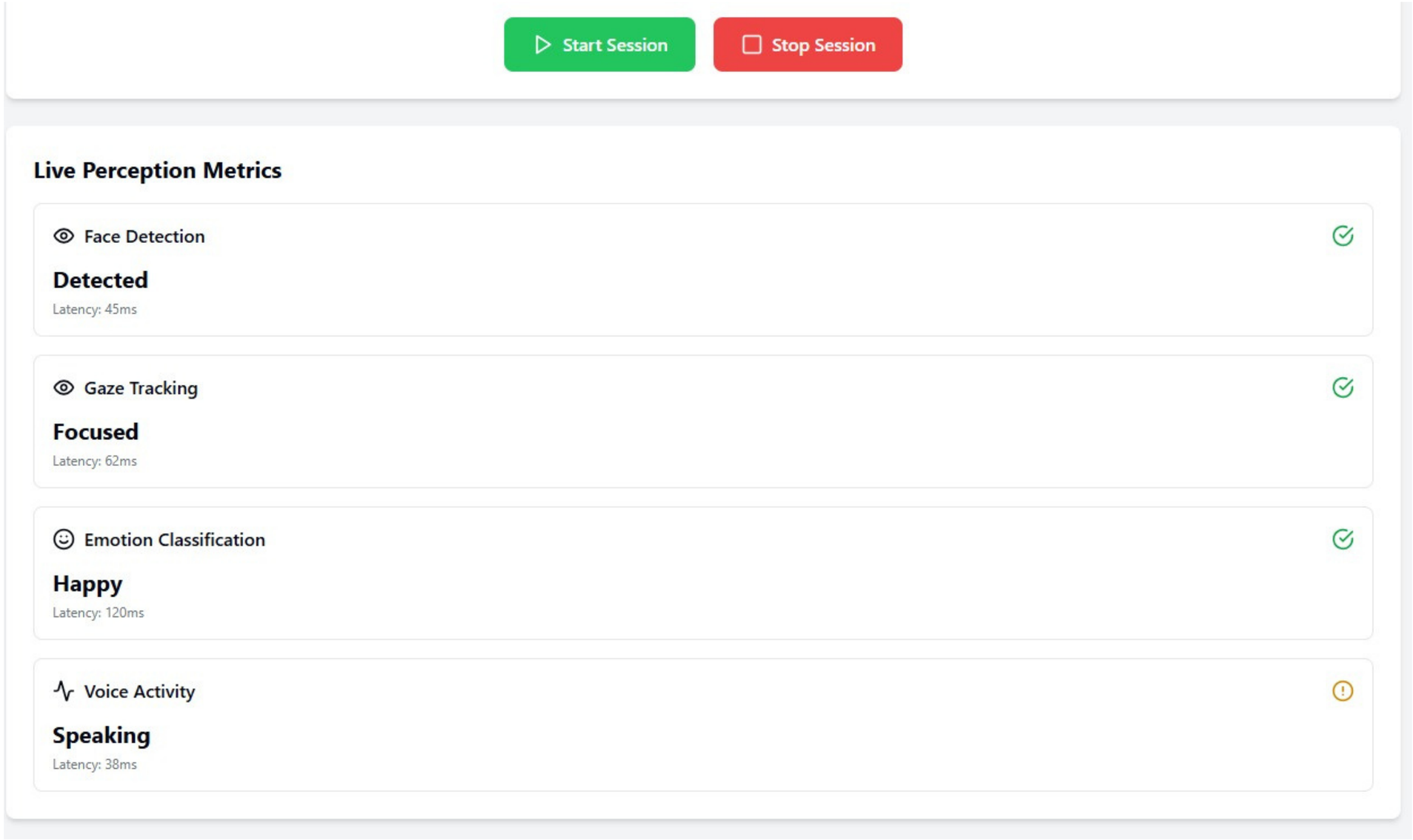
User Interface - 1



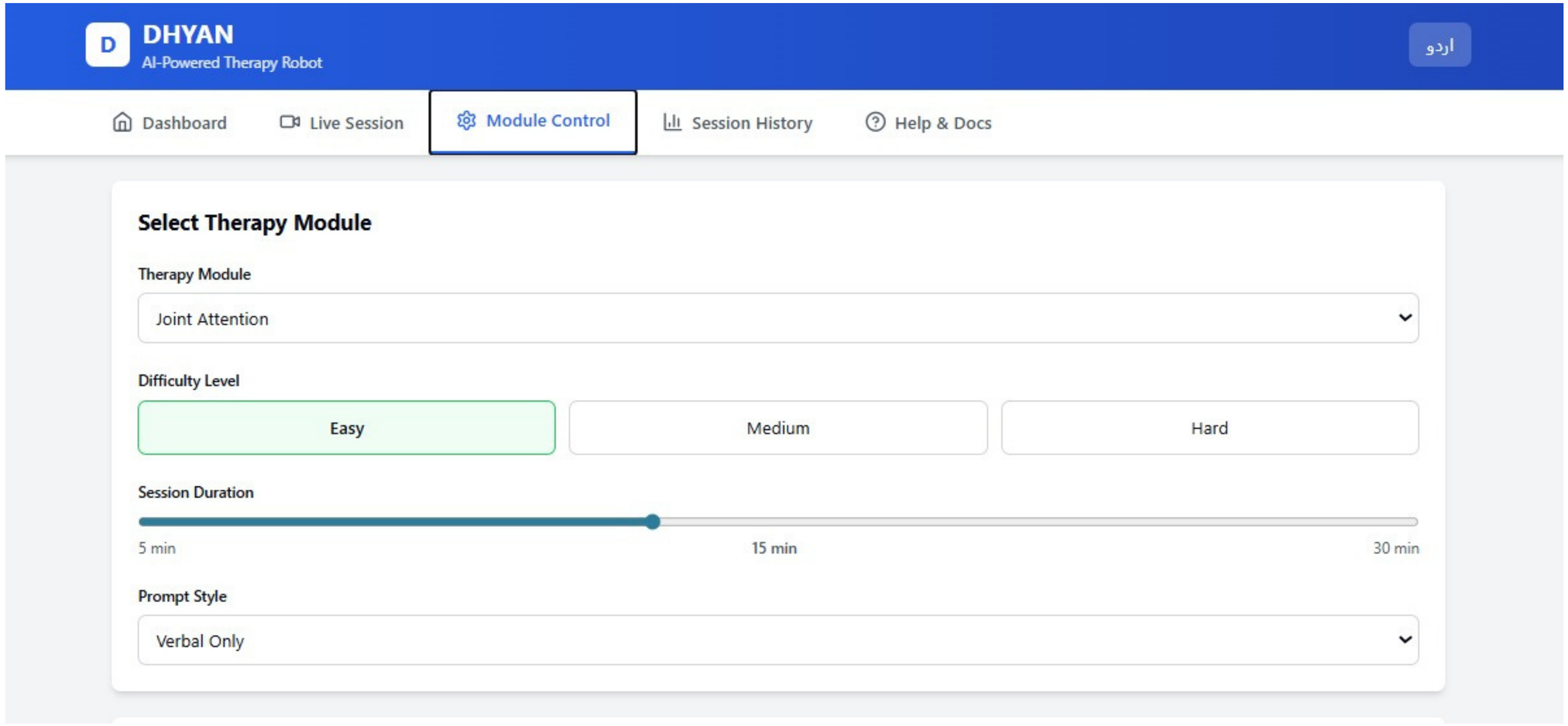
User Interface - 2



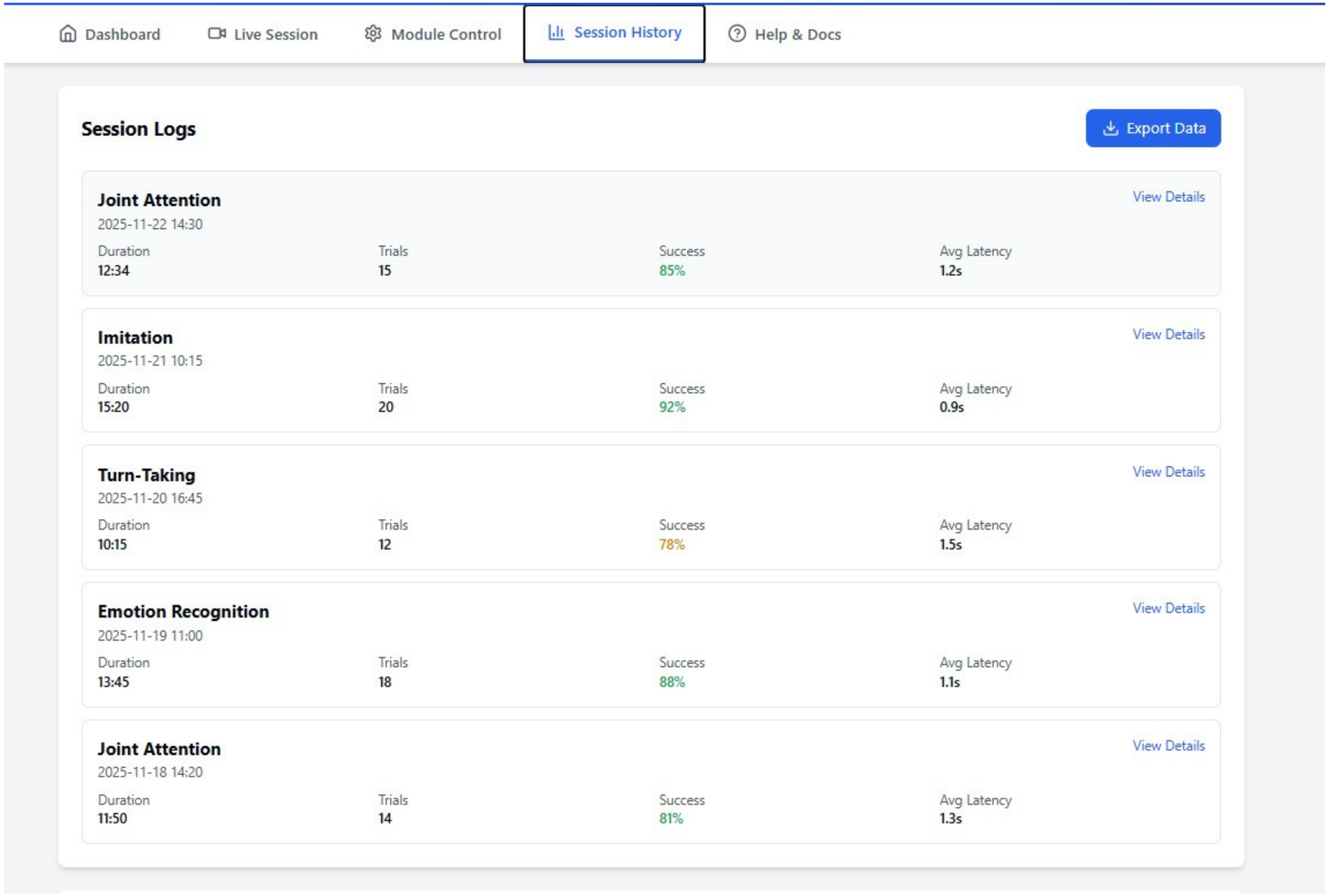
User Interface - 3



User Interface - 4



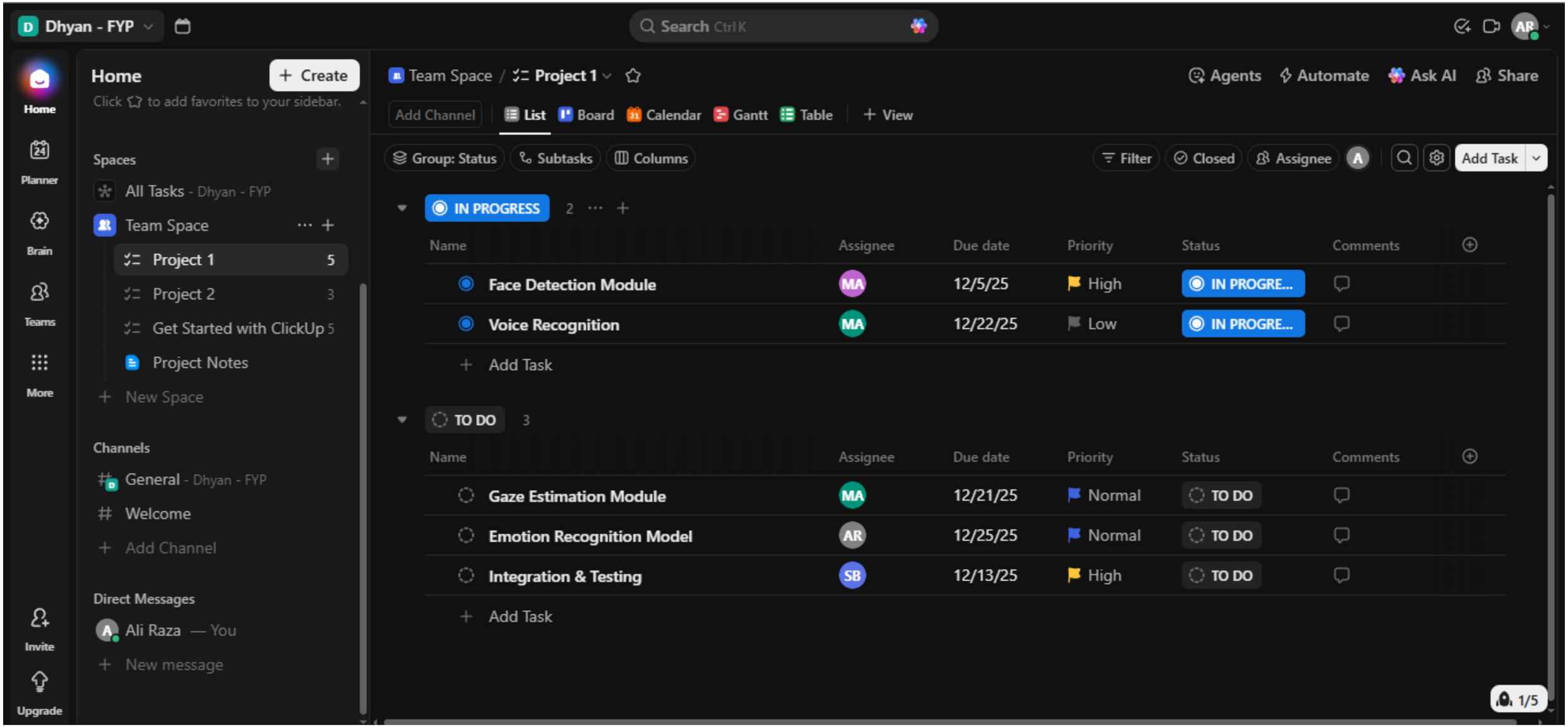
User Interface - 5



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



Thankyou