

TRD for Project Alfred

1. Person Detection and Tracking:

- **Objective:** Detect when someone enters the Innovation Lab's reception area and track their movements until they exit.
- **Technical Requirements:**
 - Computer vision algorithms for person detection and tracking.
 - Integration with cameras or sensors to monitor the reception area.
 - Real-time processing for continuous tracking.
- **Technology Stack:**
 - Computer Vision: OpenCV, TensorFlow, or PyTorch.
 - Hardware: High-definition cameras or depth sensors.
- **Functional Specifications:**
 - Detect and track individuals entering and moving within the reception area.
 - Maintain continuous monitoring until the individual exits the area.

2. Multimodal Interaction:

- **Objective:** Enable Alfred to see, hear, and speak with expressions.
- **Technical Requirements:**
 - Audio processing for speech input and output.
 - Emotion recognition for expressions through audio cues, head movements, or antennas.
- **Technology Stack:**
 - Audio Processing: Speech recognition libraries, audio playback libraries.

- Emotion Recognition: OpenCV for facial expression analysis, or custom machine learning models.
- **Functional Specifications:**
 - Capture and process audio inputs for conversational interactions.
 - Display expressions through predefined audio cues, head movements, or antenna movements.

3. Conversational Interaction:

- **Objective:** Engage guests in conversations, gather information, and provide relevant notifications.
- **Technical Requirements:**
 - Natural Language Understanding (NLU) for processing guest inquiries and commands.
 - Integration with WhatsApp API for notifications.
- **Technology Stack:**
 - NLU: Transformers-based Language Models (LLMs) like BERT, GPT.
 - Messaging API: WhatsApp Business API.
- **Functional Specifications:**
 - Greet guests and inquire about their purpose of visit.
 - Provide information about Scaler Innovation Labs and SST.
 - Detect guest presence and initiate engaging conversations.
 - Notify relevant individuals via WhatsApp about guest visits.

4. Gesture-based Actions:

- **Objective:** Perform actions such as waving hand or nodding head while interacting with guests.
- **Technical Requirements:**
 - Robotic control system for hand and head gestures.

- Integration with conversational interactions.
- **Technology Stack:**
 - Robotic Control: ROS, PyRobot.
- **Functional Specifications:**
 - Perform gestures such as waving hand while greeting or nodding head while listening to guests.
 - Coordinate gestures with speech and conversational interactions.

5. Technology Stack:

- **Vision Intelligence:** OpenCV, TensorFlow, PyTorch, Scikit Image.
- **Language Models:** Transformers-based models like BERT, GPT.
- **Speech-to-Text and Text-to-Speech:** Speech recognition and synthesis libraries.

Assumptions and Dependencies:

- Reliable internet connectivity for accessing external APIs (e.g., WhatsApp).
- Adequate hardware resources for real-time processing and interaction.
- Availability of training data for machine learning models (e.g., person detection, emotion recognition).
- Compliance with privacy regulations when handling guest information and communications.