RAFEQI Design and Implementation of a Cost-Effective Personal Assistant Robot

Shahd Elagan [1], Gaser El khodary [2], Salma Amer [3], Zeyad Magdy [4], Islam Mohamed Arafa [5], Osama Abdelrazek [6], and Hany Hanafy [7]

[1,2,3,4,5,6,7] College of Artificial Intelligence, Arab Academy for Science and Technology and Maritime Transport, Al-Alamein, Egypt.

shahd.elagan@student.aast.edu, gaserehab@student.aast.edu, salma.amer@student.aast.edu, zeyad.mansour@student.aast.edu, islam.abdallah@student.aast.edu, osamahesham357@aast.edu, hanysaid2000@aast.edu

Received on, 25 December 2024 Accepted on, 24 February 2025 Published on, 06 March 2025

ABSTRACT

This paper introduces RAFEQI, an Artificial Intelligence (AI)-driven personal assistant robot designed for seamless human-robot interaction. RAFEQI features an advanced vision system for real-time object detection, age and gender classification, emotion recognition, and sign language interpretation. Its auditory system operates both online and offline, utilizing a rule-based chatbot for offline interactions and an AI-powered language model for real-time responses. The robot is equipped with a mapping and localization system that ensures efficient navigation and obstacle avoidance. A user-friendly touchscreen Graphical User Interface (GUI) enhances accessibility and interaction. Additionally, a mechanical upgrade allows the use of a high-quality feedback servo motor at a lower cost, improving efficiency without compromising performance. Built with 3D-printed materials, RAFEQI offers an affordable yet advanced robotic solution for applications in education, research, and commercial environments.

Key-words: Personal Assistant Robot, Humanoid Robots, 3D Printing, Poppy Robot, Reachy Robot, ChatGPT, Sign Language, Localization

I. INTRODUCTION

The integration of AI and robotics is increasingly shaping modern industries, with humanoid robots playing vital roles in healthcare, education, customer service, and home assistance. The demand for intelligent, userfriendly robotic solutions has driven the development of robots such as Nao [1], Pepper [2], Poppy [3], and Reachy [4].

Nao, developed by SoftBank Robotics, is widely used in research and education for its speech recognition and programmable behavior [1]. Pepper, also from SoftBank Robotics, specializes in social interaction and is deployed in retail, hospitality, and customer service settings [2]. Poppy, an open-source humanoid, supports education and research with its modular, customizable design [3]. Reachy, created by Pollen Robotics, serves as a personal assistant, performing object manipulation and visitor interactions [4].

This paper introduces RAFEQI, an AI-powered personal assistant robot that integrates key functionalities of existing humanoid robots while introducing enhanced capabilities. Unlike its predecessors, RAFEQI features AI-driven interaction, real-time mapping, and intuitive user engagement via an accessible GUI.

RAFEQI's upper body adopts Poppy's modular framework for adaptability, while its lower body incorporates Reachy's mobility for efficient navigation. This hybrid design enhances agility and intelligence, making it more practical for real-world applications. Equipped with OpenAI's ChatGPT-3.5 Turbo, RAFEQI provides accurate, context-aware responses in real time. Its vision system, powered by deep learning and a Stereo Vision System (SVS), enables gender, age, and emotion recognition, as well as sign language interpretation for improved accessibility.

The robot employs Simultaneous Localization and Mapping (SLAM) for efficient navigation in dynamic environments. A mechanical upgrade introduces high-quality feedback servo motors at a lower cost, optimizing efficiency without compromising performance. These enhancements make RAFEQI a versatile and cost-effective solution for education, research, and personal assistance.

II. LITERATURE REVIEW

The development of RAFEQI was inspired by existing humanoid robots designed for research, education, and personal assistance. Key influences include Nao, Pepper, Poppy, and Reachy, depicted in Fig. 1 and Fig. 2.

Nao and Pepper, developed by SoftBank Robotics, are widely used in research and commercial applications. Nao is designed for education and programming, featuring 25 degrees of freedom and advanced speech recognition [1]. Pepper is a socially interactive robot with voice recognition and an interactive tablet commonly used in customer service [2]. However, both are closed source, limiting hardware modifications.

In contrast, Poppy and Reachy are open-source platforms, offering modularity and customization. Poppy, developed by the Poppy Project, features 27 degrees of freedom and a flexible humanoid design, making it ideal for robotics research [3]. Reachy, created by Pollen Robotics, has 18 degrees of freedom, a wheeled base, and an advanced stereo vision system for interaction and object manipulation [4].

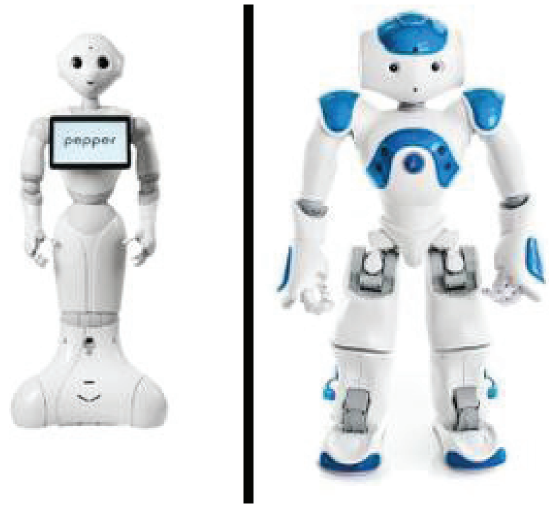

Figure 1. Pepper robot (left), Nao robot (right) [2] [1]

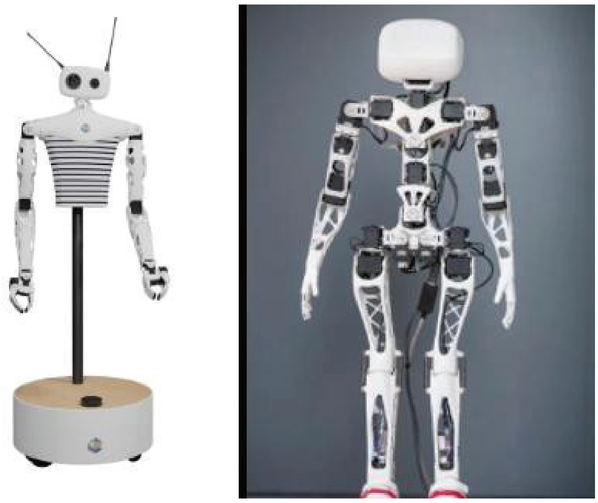

Figure 2. Reachy robot (left) Poppy robot (right) [3] [4]

RAFEQI integrates and improves upon the strengths of Poppy and Reachy while introducing key modifications for enhanced personal assistance. Its upper body is based on Poppy's modular design, adapted with structural and mechanical enhancements such as optimized servo motors and adjusted dimensions for better efficiency. The lower body, inspired by Reachy's mobility system, is redesigned for improved navigation and localization.

By leveraging open-source flexibility, RAFEQI provides a cost-effective and customizable alternative to closed-source robots like Nao and Pepper. Its AI-driven interaction, advanced vision system, and adaptable framework make it a versatile solution for research and real-world applications.

III. DESIGN AND HARDWARE MODIFICATIONS

The development of RAFEQI involved an extensive evaluation of various robotic platforms to identify the optimal combination of features. Once the key design elements were selected, modifications were implemented to enhance efficiency, reduce costs, and improve overall functionality. The design is divided into two main components: the Upper Body and Head Design and the Lower Body and Locomotion System.

The first prototype of RAFEQI was 3D printed using polylactic acid (PLA) material, as shown in Fig. 3 on the left side. Later iterations adopted resin-based 3D printing to enhance structural integrity and refine the finer details of the design, as shown on the right side of the image.

Figure 3. PLA material (left), Resin material (right)

A. *Upper Body and Head Design*

The upper body of RAFEQI houses twelve actuators, compared to thirteen actuators in Poppy, along with a vision system, an interactive screen, and a head assembly, contributing to a total weight of approximately 2 kg as shown in Fig. 4, on the right side, RAFEQI with the arrow pointing to the motor that was removed, motor abs_z(33), which appears in the full Poppy motor diagram shown on the left side. Selecting suitable motors was a critical design decision to ensure efficient movement, real-time feedback capabilities, and cost-effectiveness without compromising performance.

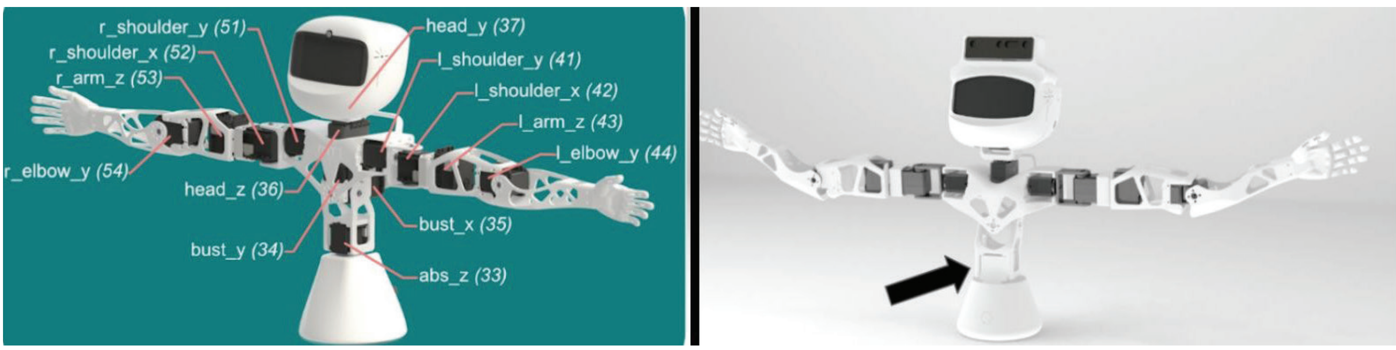

Figure 4. Poppy actuators (left), RAFEQI actuators (right) [4]

The Poppy robot originally used 8× Robotis Dynamixel MX28-AT [5] motors and 3× Robotis Dynamixel MX64-AT motors [6]. While these actuators offer high-precision movement and real-time feedback, they significantly increase the overall cost. To optimize affordability without sacrificing essential features, RAFEQI employs Waveshare ST3215-HS servo motors [7], which offer comparable functionality at a fraction of the cost. Fig. 5 illustrates a direct comparison between the Dynamixel MX-28AT motor (left) used in Poppy and the Waveshare ST3215-HS motor (right) used in RAFEQI.

Figure 5. MX-28AT motor (left), ST3215-HS motor (right) [5] [7]

A key factor in motor selection is the baud rate, which determines the speed of data transmission in a communication system. The MX-28AT motors operate via RS-485 or TTL (UART), supporting high-speed multi-drop networking and allowing multiple servos to communicate over the same bus. The ST3215-HS operates using UART communication, which, while slightly slower, is sufficient for RAFEQI's movement requirements. The differences between the two motors are shown in Table I.

TABLE I. PRESENTS A DETAILED COMPARISON OF THE ORIGINAL AND MODIFIED MOTOR CONFIGURATIONS

Feature	Robotis Dynamixel MX-28AT	Waveshare ST3215-HS
Communication Protocol	RS-485 or TTL (UART)	UART
Baud Rate	4.5M bps (bits per second)	1M bps
Feedback Mechanism	High-resolution encoder	360° magnetic encoder
Positional Accuracy	±0.5°	±0.5° to ±1°
Control Interface	Advanced control via RS-485/TTL	UART communication
Real-Time Feedback	Position, Speed, Temperature, Voltage, Torque, Current	Position, Speed, Load, Voltage
Cost per Motor	$290	$22

Since the dimensions and mounting interfaces of the ST3215-HS motors differ from the original Dynamixel MX-28AT, modifications were required to the Poppy-inspired upper body to ensure proper integration. The servo mounts were redesigned to accommodate the different shaft positioning and housing dimensions of the new motors. Additionally, structural adjustments were made to maintain joint alignment and articulation, ensuring that the robot's arms retain their intended range of motion. Furthermore, because the ST3215-HS motors have different torque characteristics, slight reinforcements were added to key load-bearing areas of the upper body to maintain mechanical stability during movement. These modifications allowed seamless integration of the cost-effective motors while preserving the humanoid aesthetics and functionality of RAFEQI.

By implementing these modifications, RAFEQI achieved a 90% cost reduction in actuator selection, making it more accessible for research and development while maintaining essential motor functions such as position tracking, movement error calculation, and smooth servo control.

The head assembly underwent significant modifications to accommodate the stereo vision camera and interactive touchscreen display. These upgrades enhance both the computer vision capabilities and user interaction. The original Poppy and Reachy robots used simpler vision setups with a single Raspberry Pi (RPI) camera positioned at the center of the head. However, during testing, this placement was found to be suboptimal for depth perception and object detection accuracy. To improve performance, the camera was relocated to the top of the head, optimizing the field of view and recognition accuracy.

RAFEQI replaces the standard raspberry pi (RPI) camera on the left [8] with an OpenCV AI Kit-D (OAK-D) Pro [9] stereo vision system on the right, as shown in Fig. 6, significantly enhancing depth perception, object detection, and facial recognition capabilities. This upgrade allows the robot to track faces, interpret sign language, and estimate user age and emotions more accurately.

Figure 6. RPI camera (left), OAK-D PRO (right) [8] [9]

The screen size was increased from 3 inches to 5 inches, enhancing the (graphical user interface) GUI's visibility and improving user experience. Since the screen serves as RAFEQI's interface, this upgrade improves interaction by making visual feedback more expressive and accessible. Fig. 7 provides a comparison between the original Poppy head on the left design and the modified RAFEQI head on the right, highlighting the structural and functional enhancements made to accommodate the upgraded vision system and larger display.

Figure 7. shows the Poppy head design (left) while the RAFEQI head design (right) [7]

B. *Lower Body and Locomotion System*

RAFEQI's lower body is designed for stable and efficient mobility while optimizing space for electronic components. Inspired by Reachy and Pepper, it adopts a 4WD omnidirectional wheel system instead of a bipedal design, which enhances stability, navigation, and localization. Unlike bipedal locomotion, Omni wheels enable seamless movement in any direction without rotation, improving maneuverability, response time, and weight distribution while simplifying path planning.

The base integrates essential components in a compact and ergonomic layout. It includes an emergency stop button, four sonar sensors for obstacle detection, a reset button, and a charging port for battery recharging. A real-time power monitoring screen displays system voltage and current draw, ensuring continuous diagnostics. Fig. 8 illustrates the wheel configuration (left) and component placement within the base (right), showcasing its accessible and compact design.

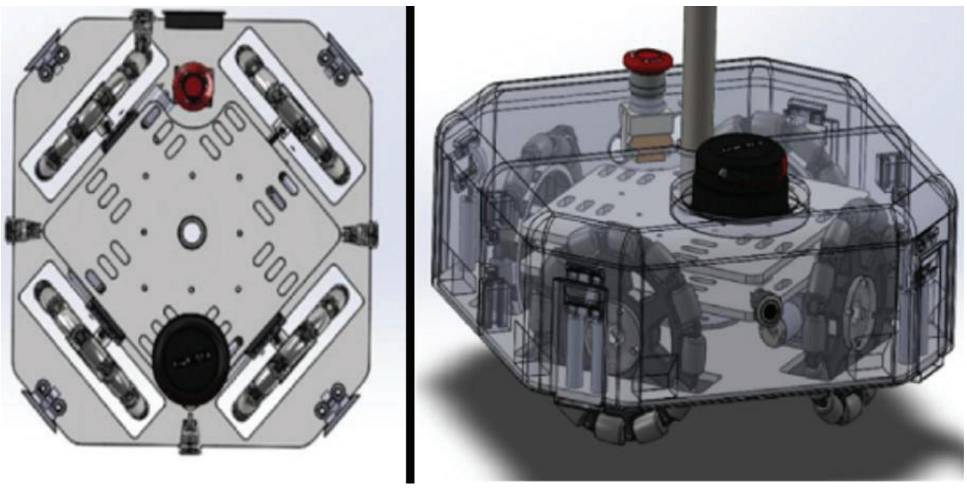

Figure 8. shows the wheel configuration (left) and base-level design (right)

Designing RAFEQI's lower body required integrating all electronic components while ensuring structural integrity and balance. To optimize space and prevent overheating, the internal structure is divided into three isolated levels, each serving a distinct function.

The first level houses the microcontroller and motor driver, enabling real-time motor control for smooth movement. The second level contains the Battery Management System (BMS) and a 2P3S-configured lithium-ion (Li-ion) battery pack. This setup consists of six 3.7V 8000mAh 65C cells, providing a final output of 11.1V 16000mAh. Power distribution boards and step-down voltage converters ensure stable power delivery to the robot's subsystems.

The third level includes the Jetson Nano [10], an edge tensor processing unit (TPU) [11], and an emergency stop system for safety. An A2 LiDAR [12] mounted on the base enables real-time mapping and localization, while four MB1020 LV-MaxSonar-EZ2 ultrasonic sensors [13] enhance obstacle detection, improving navigation in dynamic environments.

To enhance safety and minimize electromagnetic interference, the battery pack and controller components are housed in isolated compartments. This design reduces power fluctuations, thermal buildup, and noise interference, ensuring stable system operation.

The multi-layered structure optimizes space efficiency, power distribution, and real-time processing, enhancing RAFEQI's performance. This integration enables seamless movement, navigation, and interaction, making RAFEQI a versatile and efficient personal assistant robot.

IV. KINEMATICS AND MOTION CONTROL OF RAFEQI

The kinematics and motion control system of RAFEQI is designed to ensure precise, smooth, and adaptive movements for both its upper body and mobile base. The upper body kinematics governs the robot's ability to interact with users and objects, while the mobile base kinematics enables efficient omnidirectional movement in dynamic environments. The control strategies integrate forward and inverse kinematics models, Proportional-Integral-Derivative (PID) controllers, and trajectory optimization algorithms to maintain stability, accuracy, and efficiency in real-time operations.

RAFEQI's upper body consists of 12 actuated joints, which provide controlled movements for the arms, head, and interactive screen. The kinematics follows a serial-link manipulator configuration, allowing the robot to perform gestures such as pointing, waving, and head tilting. The Forward Kinematics (FK) of the upper body determines the end-effector position (e.g., the hand or screen) given a set of joint angles [14]. The transformation matrix T from the base frame to the end-effector is computed using Denavit-Hartenberg (DH) parameters as shown in (1):

$$T = \prod_{i=1}^{n} T_i = \prod_{i=1}^{n} [R_z(\theta_i) T_z(d_i) T_x(a_i) R_x(\alpha_i)] \quad (1)$$

The joint angles θi govern rotational motion, while di, ai, and αi define link offsets, lengths, and twist angles, shaping the robotic arm's spatial configuration. Rz and Rx represent rotations about the z-axis and x-axis, influencing joint orientation, while Tz and Tx define translations along these axes for precise link positioning. Together, these parameters establish RAFEQI's upper body kinematics, ensuring accurate end-effector control in both position and orientation.

For Inverse Kinematics (IK), an iterative numerical approach is employed due to the non-linear nature of multi-DOF arm motion. The Levenberg–Marquardt optimization method is used to compute joint angles required to reach a given Cartesian target while minimizing joint stress and motion constraints. The upper body motion is controlled via a Proportional-Integral-Derivative (PID) controller, ensuring smooth and precise trajectory execution. The control law for each joint is shown in (2):

$$u_i(t) = K_p e_i(t) + K_i \int e_i(t) dt + K_d \frac{de_i(t)}{dt} \quad (2)$$

The error ei(t) represents the difference between the desired and actual joint position, enabling precise motion correction. Kp, Ki, and

Kd are the proportional, integral, and derivative gains that dynamically adjust control effort to regulate position errors. The control signal ui(t) drives the actuator, ensuring smooth joint adjustments while maintaining stability and minimizing oscillations. To enhance motion fluidity, a velocity profiling strategy prevents abrupt accelerations, particularly during interaction-based gestures.

RAFEQI's mobile base employs a 4WD omni-wheel drive system, enabling holonomic motion, which allows movement in any direction without conducting rotation first. This design is particularly beneficial for maneuvering in confined spaces. The velocity kinematics of RAFEQI's base follows the standard Mecanum wheel inverse kinematics model. The relationship between global velocity (V_x, V_y) and angular velocity (ω) is shown in (3):

$$\begin{bmatrix} v_x \\ v_y \\ \omega_z \end{bmatrix} = \frac{r}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ -\frac{1}{(l_x+l_y)} & \frac{1}{(l_x+l_y)} & -\frac{1}{(l_x+l_y)} & \frac{1}{(l_x+l_y)} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \end{bmatrix} \quad (3)$$

Where r represents the wheel radius, which influences the robot's movement efficiency and speed, the parameters L and W denote the base length and width, respectively, which define the robot's physical dimensions and affect its stability and maneuverability. Lastly, ω_1, ω_2, ω_3, and ω_4 correspond to the four individual wheel speeds, controlling the motion dynamics and enabling omnidirectional movement through precise velocity adjustments.

This configuration allows RAFEQI to achieve Linear motion (forward, backward, left, right), Diagonal motion, Rotational motion (on-the-spot turns), and Complex curvilinear trajectories. The base movement is controlled using a closed-loop PID velocity controller, ensuring smooth acceleration and precise stopping. The control system dynamically adjusts individual wheel velocities to maintain trajectory stability. The velocity control equation is shown in (4):

$$\omega_i = K_p e_v + K_i \int e_v dt + K_d \frac{de_v}{dt} \quad (4)$$

The desired wheel speed ωi defines the rotational velocity for the robot's motion, while ev represents the velocity error—the difference between desired and actual speed—used for motor control adjustments. The Kp, Ki, and Kd PID parameters ensure precise velocity regulation, minimizing overshoot and stabilizing movement.

For path execution, RAFEQI employs an A* search algorithm for path planning, as shown in Fig. 9 [15], whereas the Dynamic Window Approach (DWA) for real-time obstacle avoidance is shown in Fig. 10 [16].

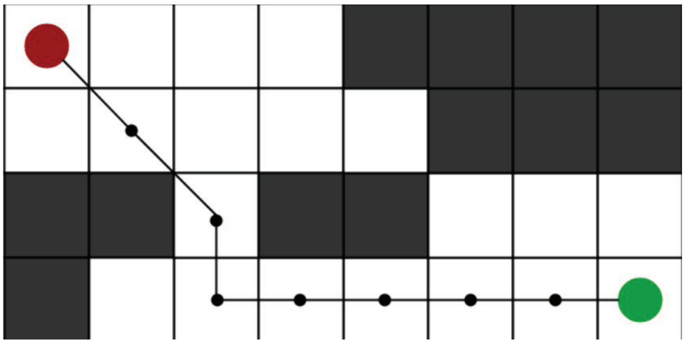

Figure 9. A* search algorithm [15]

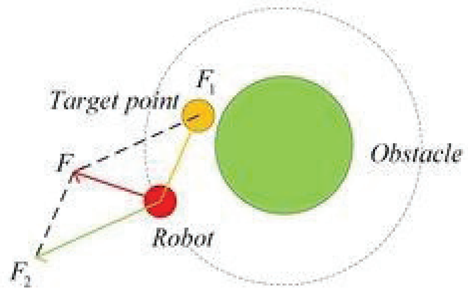

Figure 10. Dynamic Window Approach [16]

V. ARTIFICIAL INTELLIGENCE ARCHITECTURE AND SYSTEMS

A. *Vision System*

The vision system of RAFEQI is designed to enable depth perception, object recognition, emotion detection, and sign language interpretation using a stereo vision camera setup. The system incorporates the OAK-D Pro stereo vision camera, which leverages DepthAI models and libraries trained on platforms such as Roboflow.

This vision architecture is responsible for two primary tasks: depth estimation and high-level image processing, both of which are essential for robotic navigation, human interaction, and object manipulation.

To achieve efficient real-time inference, RAFEQI's vision system is powered by a Jetson Nano, which is equipped with an Edge Tensor Processing Unit (TPU) accelerator. This hardware configuration optimizes computational efficiency, enabling low-latency processing of deep learning models for age and gender classification, emotion recognition, object detection, and sign language interpretation. The integration of the Edge TPU accelerators enhances inference speed and energy efficiency, ensuring that RAFEQI operates with high accuracy and minimal latency in real-world applications.

1. Age and gender classification model

The age and gender classification model is deployed on the OAK-D Pro camera, leveraging the DepthAI framework built on Intel's OpenVINO toolkit. The model is a lightweight deep convolutional neural network (CNN) optimized for real-time inference on edge devices. It is based on a ResNet-50 backbone trained on large-scale datasets, including Adience and IMDB-WIKI, containing diverse facial images labeled with age and gender [17].

The gender classification task is treated as a binary classification problem (male/female), with the model achieving an accuracy of 95.3% on the Adience dataset. The age estimation task is handled as a multi-class classification problem. The model achieves a Mean Absolute Error (MAE) of 4.2, meaning that the predicted age is, on average, within 4.2 years of the actual age.

The model was optimized for deployment on edge TPU accelerators by quantizing the weights to INT8, reducing computational complexity while maintaining accuracy. The inference latency on the OAK-D Pro is 7.2 milliseconds per frame, ensuring real-time processing capabilities. Fig. 11 shows the real-time output for gender and age estimation, displaying the detected faces along with predicted labels and age values.

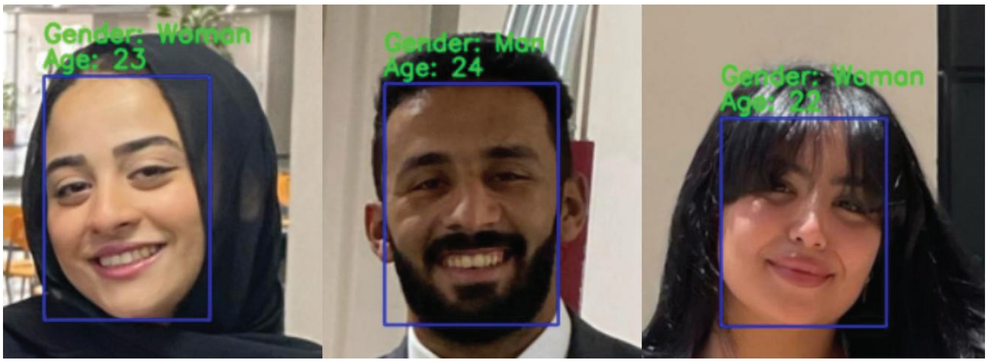

Figure 11. Real-time output for gender and age DepthAI model

2. Object detection and emotion recognition model

For object detection, RAFEQI utilizes a YOLOv8n (You Only Look Once version 8 - nano variant) model, optimized for low-latency real-time inference on edge devices. The model was trained on a custom dataset containing 60 object classes relevant to RAFEQI's interaction environment, including human faces, everyday objects, and obstacles. The final trained model achieved a mean average precision (mAP) of 92.1% at Intersection over Union (IoU)=0.5. The inference time on the OAK-D Pro camera is 6.8 milliseconds per frame, enabling high-speed object detection [18]. Some of the images used during model training are shown on the left side of Fig. 12, while the right side of the Figure shows YOLOv8n object detection test images where objects in the scene are labeled accordingly.

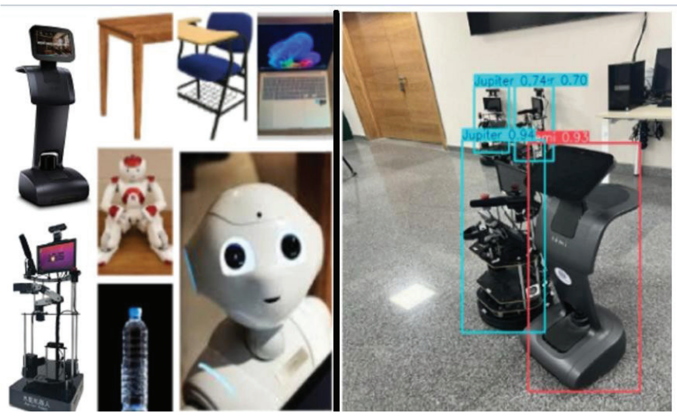

Figure 12. Images from the training dataset (left) and the YOLOv8n output (right)

For emotion detection, RAFEQI employs a fine-tuned EfficientNet-B3 model, trained on the Facial Expression Recognition 2013 (FER-2013) dataset, which consists of 35,887 grayscale facial images labeled with seven emotion classes (anger, disgust, fear, happiness, sadness, surprise, and neutrality). The model was trained using transfer learning, with the final trained version achieving an accuracy of 87.5% on the test set [19]. To enhance robustness to lighting variations and occlusions, the Multi-Task Cascaded Convolutional Network (MTCNN) is used as a preprocessing step to align and crop faces before emotion classification. This ensures that the model processes only relevant facial regions, reducing false predictions. The average inference time for emotion detection is 11.6 milliseconds per frame, enabling near real-time facial expression analysis. Fig. 13 illustrates the robot's output for emotion recognition, where detected faces are annotated with predicted emotions.

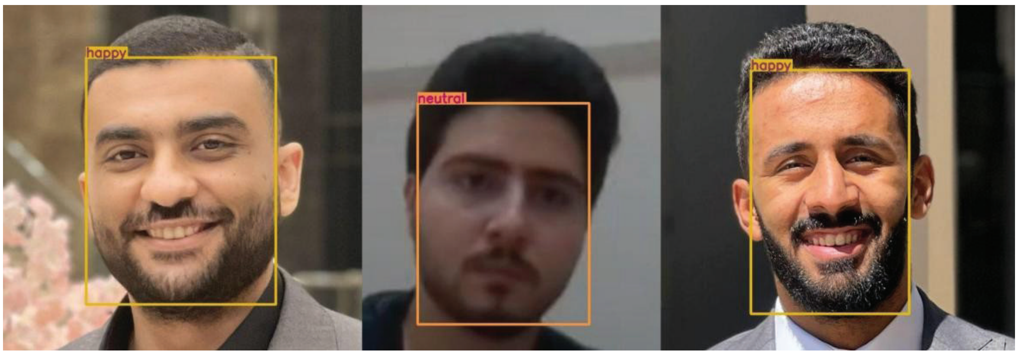

Figure 13. Real-time emotion detection on RAFEQI's screen

3. Depth estimation model

RAFEQI's depth estimation utilizes the OAK-D Pro stereo vision system with DepthAI's stereo depth estimation framework. Unlike monocular depth estimation, which relies on learned priors, stereo vision computes depth from disparity measurements, ensuring higher accuracy and robustness. The system applies semi-global matching (SGM) and depth post-processing techniques such as median filtering, bilateral filtering, and temporal disparity filtering to enhance precision and reduce noise.

For object grasping, real-time depth perception enables accurate positioning and stable manipulation. The OAK-D Pro achieves an absolute depth error below 2% for distances up to 4 meters, enhancing spatial awareness for grasping tasks. Subpixel disparity refinement improves fine-detail reconstruction, which is particularly useful for small or partially occluded objects. Fig. 14 illustrates depth estimation accuracy across distances and presents depth maps overlaid on the camera feed, demonstrating scene understanding and precision [20].

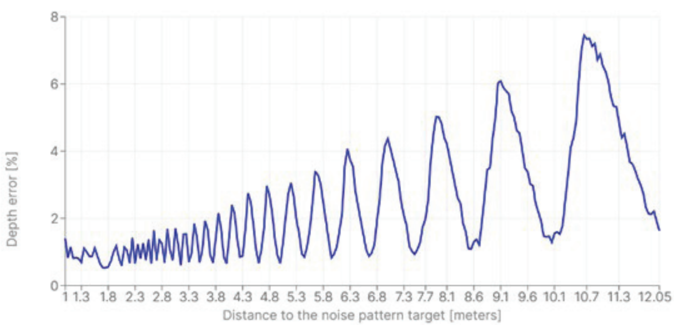

Figure 14. Depth estimation accuracy [20]

To ensure real-time performance, the system processes stereo images at 30 FPS, employing confidence thresholding and edge-aware filtering to minimize disparity inconsistencies. Point cloud processing further refines depth data, enhancing object recognition and grasp planning. This stereo-vision system significantly improves 3D spatial awareness, allowing adaptive grasping strategies based on environmental conditions. By integrating stereo vision, depth refinement algorithms, and real-time grasping techniques, RAFEQI achieves precise, stable, and adaptive manipulation in dynamic environments. Fig. 15 illustrates depth estimation output where depth maps overlaid on the camera feed.

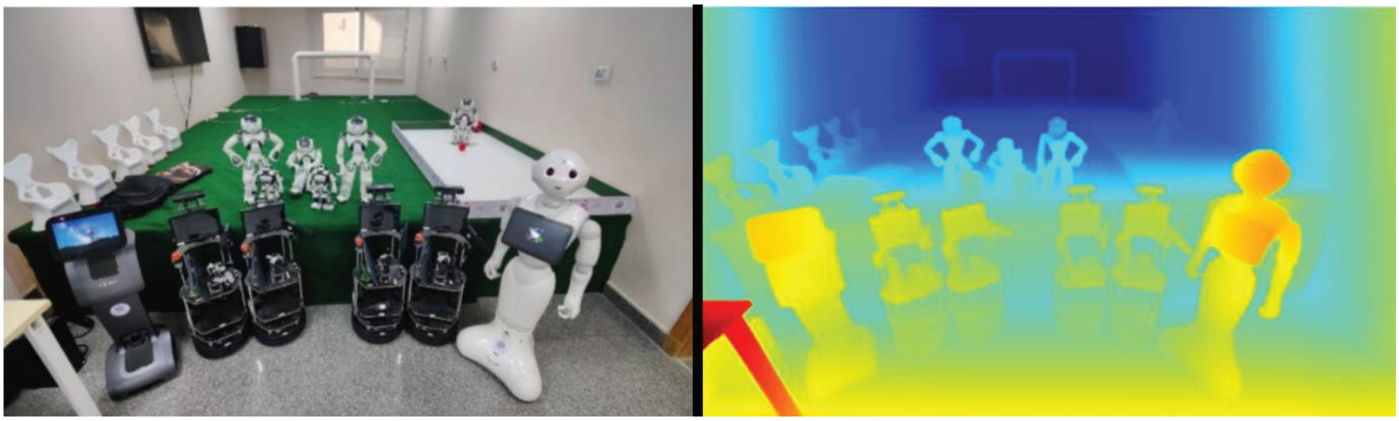

Figure 15. Depth estimation results

4. Sign language recognition and pose estimation

To enable sign language recognition, RAFEQI employs a hybrid model combining YOLOv8-Pose and MediaPipe Hand Tracking. The YOLOv8-Pose model, optimized for real-time keypoint detection, is used to track full-body poses and detect 17 skeletal key points, including hand and finger positions. The model was trained on the OpenSign dataset, which contains labeled pose sequences for American Sign Language (ASL).

For hand gesture recognition, RAFEQI integrates MediaPipe Hand Tracking, which detects 21 hand landmarks per hand using a combination of a palm detection model and a regression-based landmark estimation model. The real-time inference speed of the MediaPipe model is 19.3 milliseconds per frame, ensuring efficient tracking of complex hand movements.

The YOLOv8-Pose model achieves a mean Average Precision (mAP) of 73.7% for full-body pose estimation, while the MediaPipe-based sign recognition system achieves an accuracy of 93.4% when tested on ASL gestures. The combined system allows RAFEQI to interpret sign language, recognize human postures, and facilitate non-verbal communication, making it more accessible to individuals with hearing impairments.

[18][19]. Fig. 16 demonstrates the robot's ability to recognize hand signs, highlighting detected gestures with corresponding classification labels in real time.

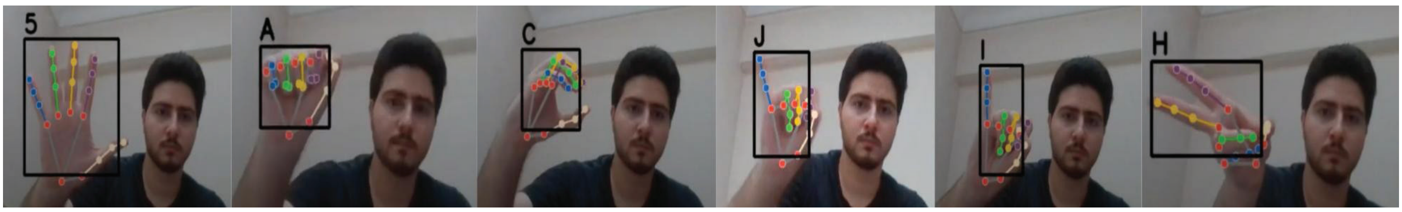

Figure 16. Hand signs real-time detection

B. *Auditory System*

Effective communication between humans and robots remains one of the most critical challenges in artificial intelligence and robotics. To achieve better user engagement, the robot's ability to comprehend natural language and generate meaningful responses must be of high quality. This is particularly crucial for RAFEQI, which functions as a personal assistant robot where natural language understanding (NLU) and response generation are essential for human interaction. The effectiveness of a conversational AI system directly influences user engagement, trust, and the overall usability of the robot. When users perceive the robot's comprehension as reliable and accurate, they are more likely to interact with it, exploring its features and utilizing its interface.

To achieve robust and adaptable speech interaction, RAFEQI integrates two language models: an offline model for local processing and an online model for cloud-based advanced conversational AI. This dual-model setup ensures uninterrupted natural language processing (NLP) capabilities, even in cases where network connectivity is unavailable.

1. Offline conversational AI model

RAFEQI includes an offline language model, ensuring continuous functionality even when an internet connection is unavailable. This model is based on the ChatterBot library, an NLP framework designed for local conversational AI. The model was trained on a dataset of Twitter dialogues and other public conversational corpora, making it robust in handling basic human interactions.

The ChatterBot model was further fine-tuned with domain-specific JSON files, incorporating customized responses tailored to RAFEQI's operational context. These modifications improved the model's coherence and response relevance, allowing it to effectively answer frequently asked questions, engage in small talk, and provide assistance in predefined scenarios.

The offline model operates with an intent-based architecture, where each query is matched to a predefined intent using TF-IDF (Term Frequency-Inverse Document Frequency) vectorization. When a user inputs a query, the system selects the most appropriate response based on its trained similarity metrics. This approach allows for low-latency responses, making the robot's conversational abilities function smoothly without internet dependency. The inference time for the ChatterBot-based model on RAFEQI's hardware is 7.2 milliseconds per query, ensuring real-time interaction [21].

2. Online conversational AI model and Text-to-Speech (TTS) integration

To extend RAFEQI's conversational capabilities beyond the offline model, the robot is also connected to the ChatGPT-3.5 Turbo API, a state-of-the-art cloud-based language model developed by OpenAI. The ChatGPT-3.5 Turbo model enables context-aware conversation, dynamic response generation, and the ability to handle open-domain interactions. Unlike intent-based systems, ChatGPT's architecture is based on the Transformer model, which uses self-attention mechanisms to generate human-like responses.

The online model is enhanced with GPT-TTS1-HD, a fine-tuned Text-To-Speech (TTS) system that provides high-definition speech synthesis with customizable voice parameters. The TTS model offer an adaptive intonation and prosody, making the speech more natural, variable speech rate control, allowing adjustments based on user preference. Context-aware emphasis on handling comprehension of critical information [22].

RAFEQI's speech pipeline follows a sequential process where user input is transcribed, processed, and responded to using either the offline or online model. The inference time for ChatGPT-3.5 Turbo is approximately 16.4 milliseconds per query, allowing near-instantaneous response generation. The TTS inference latency is 12.7 milliseconds, ensuring that RAFEQI's spoken responses are delivered smoothly without noticeable delay[23].

To optimize cloud interaction and reduce latency, RAFEQI's auditory system caches frequently used responses, enabling localized retrieval of common dialogues. This hybrid approach balances processing efficiency, network dependency, and conversational fluency.

3. Hardware and audio processing pipeline

RAFEQI's auditory hardware setup includes a high-fidelity microphone array for accurate speech recognition and noise cancellation, dual stereo speakers. These settings provide a crisp, natural-sounding speech output, Digital Signal Processing (DSP) enhancement, improving voice clarity and background noise suppression.

To enhance speech recognition accuracy, the Whisper Automatic Speech Recognition (ASR) model by OpenAI was integrated for robust voice-to-text conversion. The ASR model was fine-tuned using multispeaker datasets, achieving a Word Error Rate (WER) of 3.8% under real-world conditions. The end-to-end speech recognition pipeline allows for continuous listening, enabling hands-free interaction with the robot.

Fig. 17 illustrates the complete pipeline for user interaction with RAFEQI's large language models (LLMs), including speech recognition, NLP processing, and TTS synthesis.

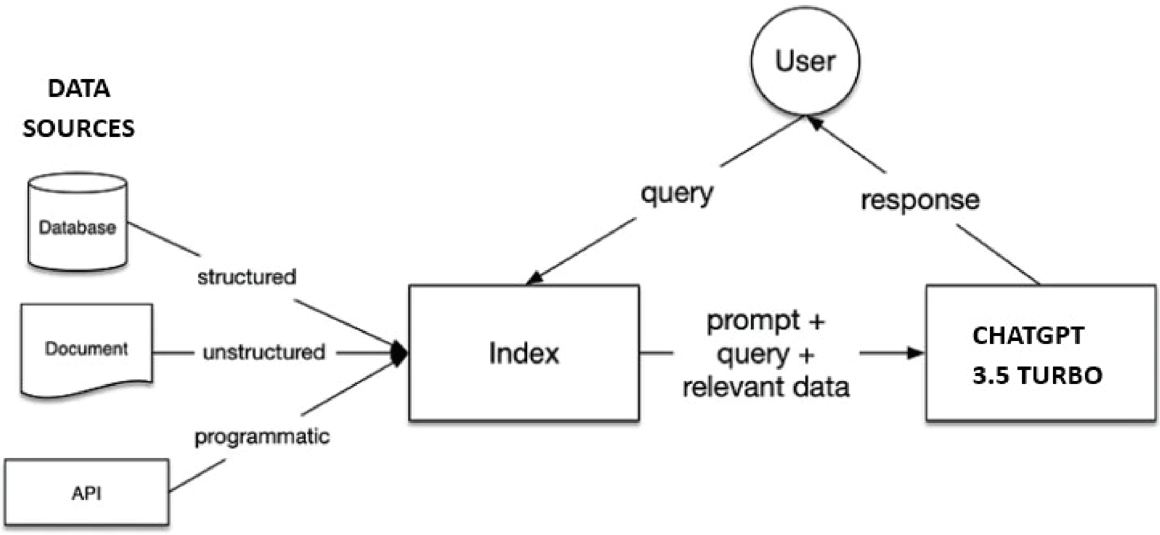

Figure 17. Natural language processing pipeline

VI. LOCALIZATION & MAPPING

Localization and mapping are crucial for RAFEQI's autonomous navigation, enabling precise environmental interaction and movement. To achieve high accuracy, RAFEQI integrates multiple sensors and advanced algorithms.

RAFEQI is equipped with an A2 LiDAR, sonar sensors, and an Inertial Measurement Unit (IMU). The A2 LiDAR provides detailed spatial data for mapping and obstacle detection, while sonar sensors enhance low-lying obstacle detection with a range of 20–750 cm. The IMU contributes orientation and movement data, complemented by motor encoder feedback, which provides odometry data for speed and distance estimation. Sensor fusion improves localization and mapping accuracy for complex environments.

For mapping, RAFEQI employs GMapping [24], a widely used Simultaneous Localization and Mapping (SLAM) algorithm. For localization, an improved Adaptive Monte Carlo Localization (AMCL) algorithm integrates 3D LiDAR, IMU, and odometry data to enhance indoor positioning accuracy [25].

The localization process begins with multi-sensor fusion, where wheel odometry and IMU data are processed using an Extended Kalman Filter (EKF) to refine the motion model. AMCL then distributes a particle set based on estimated motion dynamics.

To further refine localization, the Point-to-Line Iterative Closest Point (PL-ICP) algorithm aligns the 3D laser point cloud using pose differences from AMCL. The Levenberg–Marquardt method then computes a high precision laser odometry estimate. Finally, the AMCL-based position estimate is corrected using refined laser odometry data, followed by particle re-weighting and resampling for improved stability.

Simulation and real-world tests confirm that this enhanced AMCL framework significantly improves positioning accuracy in dynamic indoor environments. Fig. 18 illustrates the mapping, localization, and final map update steps, while Fig. 19 presents the map and localization outputs.

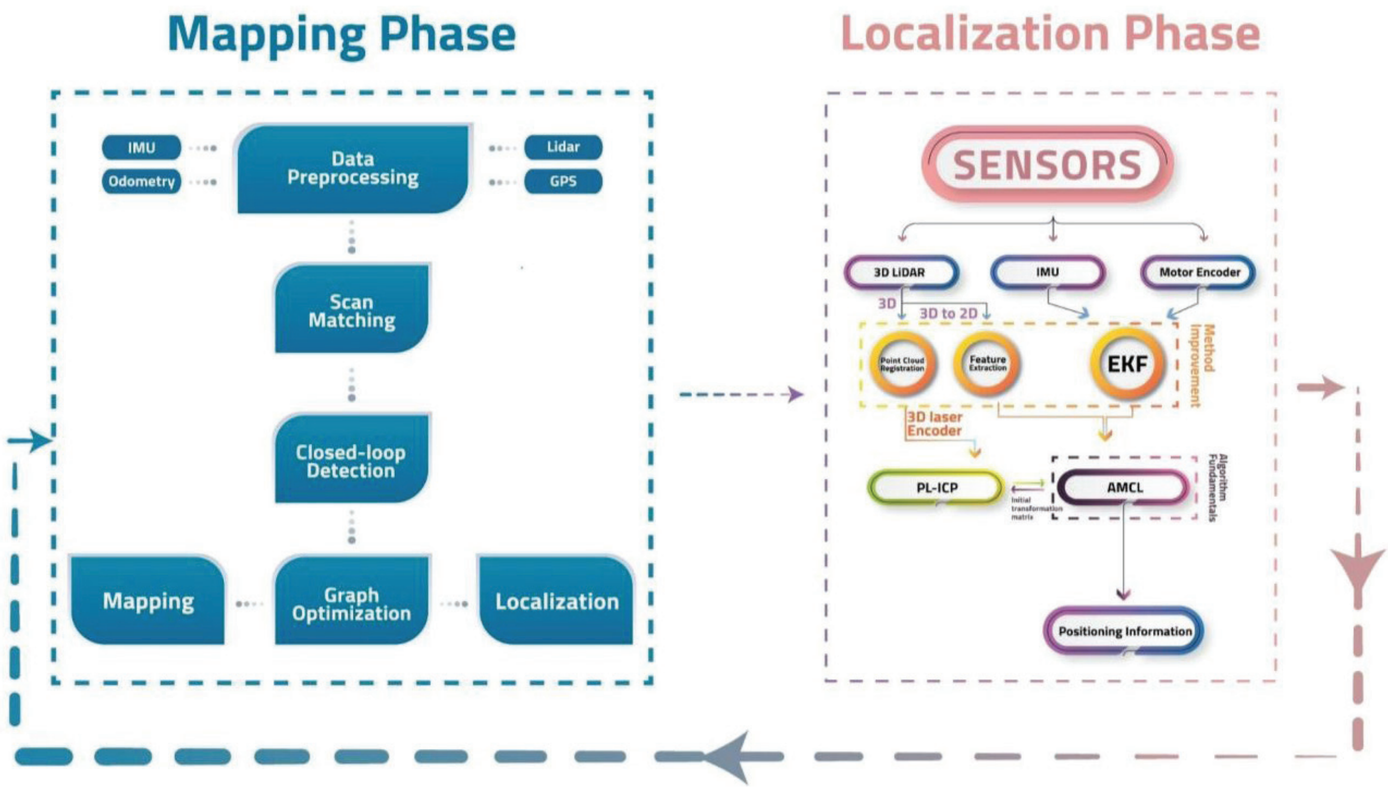

Figure 18. Mapping and Localization phases

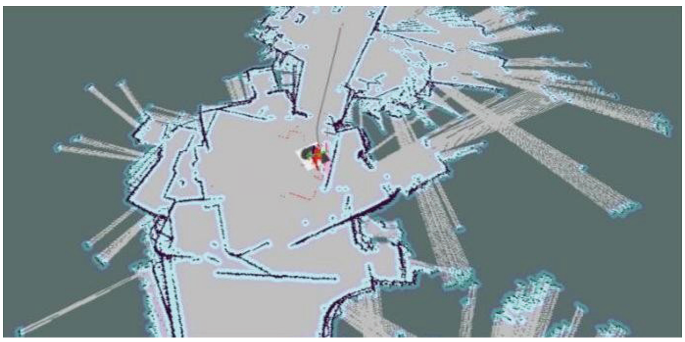

Figure 19. Mapping and Localization Output

VII. HUMAN-ROBOT INTERACTION INTERFACE: GUI & WEB APPLICATION

Enhancing Accessibility and Human-Robot Interaction (HRI) was a key design objective for RAFEQI. Unlike traditional interaction models relying solely on facial expressions, RAFEQI features a Graphical User Interface (GUI) on its interactive touchscreen, complemented by a mobile-friendly web application. This hybrid approach enables multimodal communication, allowing both local and remote access to system functionalities. By integrating visual feedback, real-time diagnostics, and direct control mechanisms, RAFEQI ensures a seamless user experience, particularly in assistive and service robotics applications.

The GUI and web application support real-time data monitoring, system configuration, and remote control. Both interfaces employ a multi-layered architecture, dynamically presenting sensor data, logs, and user settings for enhanced transparency and usability. Categorized menus allow users to adjust voice interaction settings, view object detection results, and manage navigation controls in real time. A feedback display provides continuous monitoring of performance metrics, battery status, and environmental updates for optimized interaction.

Beyond the GUI, RAFEQI's web application extends core functionalities, enabling multi-device accessibility across smartphones, tablets, and PCs. Users can remotely control movements, configure the system, and manage AI-driven features. The analytics dashboard provides insights into battery performance, system diagnostics, and activity logs, ensuring efficient supervision. A key feature is live streaming of the robot's vision system, allowing real-time situational awareness and decision-making.

To enhance user engagement, the GUI and web interface integrate emotion-driven animations, delivering responsive interactions. Adaptive scaling techniques optimize readability under varying lighting conditions, while multi-touch support enables smooth scrolling, selection, and gesture-based navigation. The Jetson Nano-powered GUI employs hardware-accelerated rendering for low-latency performance, while the web app runs on an embedded server with secure encryption and authentication.

RAFEQI's modular GUI and web interface allow future expansions without hardware modifications, supporting AI-driven analytics, cloud-based control, and advanced collaboration tools. Fig. 20 illustrates the GUI layout, showcasing interactive elements and remote access features, including live-streaming vision and control options.

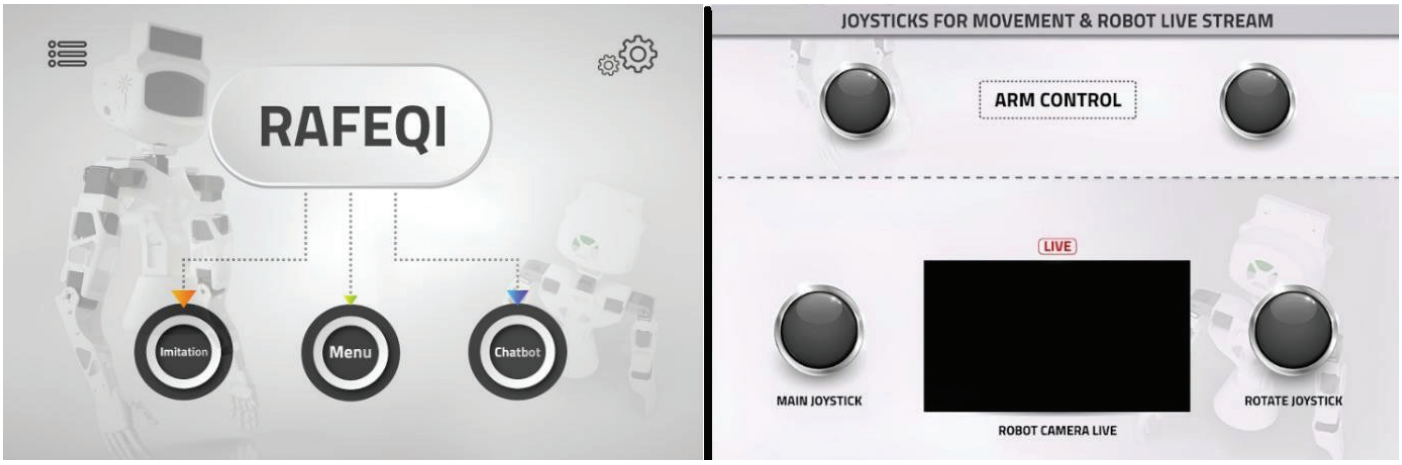

Figure 20. illustrates the GUI (left), WEB application control, and live streaming (right)

RAFEQI also supports predefined arm movements, as shown in Fig. 21, which users can activate via dedicated buttons for an interactive experience. The right-side interface includes buttons for advanced AI models across computer vision, natural language processing (NLP), and localization, enhancing task accuracy and efficiency. The user-friendly design ensures intuitive navigation, allowing users to fully explore RAFEQI's capabilities.

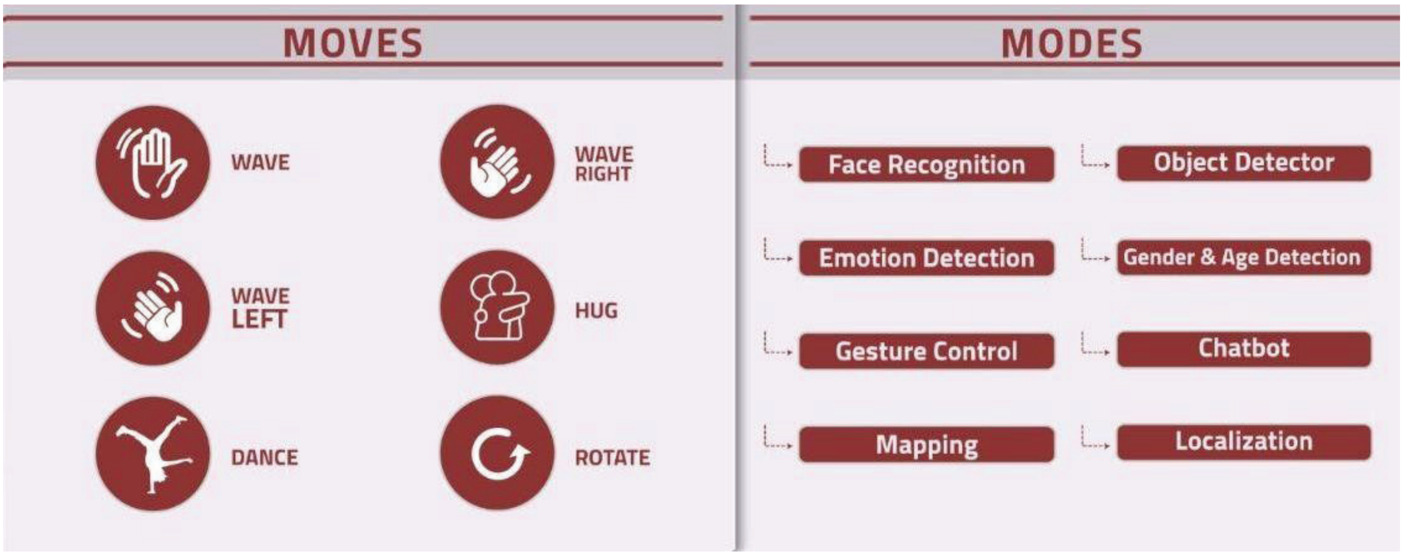

Figure 21. Predefined body moves and AI models

VIII. CONCLUSION

RAFEQI represents a significant advancement in personal assistant robotics, integrating AI-driven interaction, vision-based perception, adaptive navigation, and remote accessibility. The robot's multimodal communication system, combining offline and online conversational AI models, ensures seamless human-robot interaction. The GUI and web application further enhance accessibility, allowing users to monitor system data, control movements, and configure settings remotely, expanding RAFEQI's usability beyond physical interaction.

A major contribution of RAFEQI lies in its vision system, which utilizes a stereo vision camera with deep learning models for age and gender classification, emotion detection, object recognition, and sign language interpretation. By leveraging edge AI processing on a Jetson Nano with an Edge TPU accelerator, the system achieves low latency and real-time performance, making it suitable for dynamic environments.

For navigation, RAFEQI employs a hybrid localization and mapping system, integrating A2 LiDAR, sonar sensors, and an IMU. The robot utilizes GMapping SLAM for mapping, followed by an optimized AMCL algorithm with EKF and PL-ICP for precise localization. This approach enables autonomous navigation and obstacle avoidance in structured and unstructured environments.

RAFEQI's cost-efficient yet high-performance hardware design modifies the original Poppy and Reachy structures, enabling high-feedback servo motors while significantly reducing costs. The 4WD omni-wheel base enhances seamless omnidirectional movement, further improving mobility.

The integration of a web application significantly extends RAFEQI's usability, allowing remote supervision, teleoperation, and real-time diagnostics. The expanded analytics dashboard provides insights into battery levels, performance metrics, and operational history, ensuring ease of management in various applications.

Future work will focus on enhancing context awareness, multimodal learning, and reinforcement-based decision-making to expand RAFEQI's capabilities in autonomous assistance and adaptive learning. The web application will continue evolving with AI-driven analytics and cloud-based collaboration tools, solidifying RAFEQI's role as a next-generation AI-powered personal assistant.

ACKNOWLEDGEMENT

The development of RAFEQI would not have been possible without the support and resources provided by the Robotics Lab at the College of Artificial Intelligence (CAI), Arab Academy for Science, Technology and Maritime Transport (AASTMT) Alamein Campus. The lab's state-of-the-art research environment, access to advanced robotic platforms, and collaborative expertise played a crucial role in designing, testing, and optimizing RAFEQI's mechanical structure, AI-driven interaction systems, and autonomous navigation framework.

We extend our sincere gratitude to the faculty members and researchers at the CAI Robotics Lab for their valuable insights, feedback, and technical contributions throughout the project. Their expertise in robotics, artificial intelligence, and embedded systems significantly enriched the development process, enabling RAFEQI to evolve into a versatile, efficient, and human-centered personal assistant robot.

Additionally, we acknowledge the support from the AASTMT administration, which facilitated access to cutting-edge hardware, computational resources, and testing environments. The collaborative efforts within the Robotics Lab have been instrumental in advancing robotic research and AI-driven automation, and we look forward to further innovations and contributions in this field.

REFERENCES

[1] S. Michieletto, D. Zanin, and E. Menegatti, "NAO Robot Simulation for Service Robotics Purposes," in *2013 European Modelling Symposium*, IEEE, Nov. 2013, pp. 477–482. doi: 10.1109/EMS.2013.80.

[2] A. K. Pandey and R. Gelin, "A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind," *IEEE Robot Autom Mag*, vol. 25, no. 3, pp. 40–48, Sep. 2018, doi: 10.1109/MRA.2018.2833157.

[3] M. Lapeyre, P. Rouanet, and P.-Y. Oudeyer, "Poppy: a New Bio-Inspired Humanoid Robot Platform for Biped Locomotion and Physical Human-Robot Interaction," *in Proceedings of the 6th International Symposium on Adaptive Motion in Animals and Machines (AMAM)*, 2013.

[4] S. Mick *et al.*, "Reachy, a 3D-Printed Human-Like Robotic Arm as a Testbed for Human-Robot Control Strategies," *Front Neurorobot*, vol. 13, Aug. 2019, doi: 10.3389/fnbot.2019.00065.

[5] "ROBOTIS," 2025. [Online]. Available: https://emanual.robotis.com/docs/en/dxl/mx/mx-28/

[6] "ROBOTIS e-Manual." [Online]. Available: https://emanual.robotis.com/docs/en/dxl/mx/mx-64/

[7] "20kg.cm Bus Servo Motor, 106PRM High Speed, Large Torque, With 360 Degrees High Precision Magnetic Encoder," 2025. [Online]. Available: https://www.waveshare.com/st3215-hs-servo-motor.htm

[8] R. P. Ltd, "Buy a Raspberry Pi High Quality Camera." [Online]. Available: https://www.raspberrypi.com/products/raspberry-pi-high-quality-camera/

[9] "OAK-D Pro." [Online]. Available: https://shop.luxonis.com/products/oak-d-pro?variant=42455252369631

[10] "Buy the Latest Jetson Products." [Online]. Available: https://developer.nvidia.com/buy-jetson

[11] "Products | Coral," 2020, *Coral*. [Online]. Available: https://coral.ai/products/

[12] "RPLIDAR-A2 Laser Range Scanner_ Solid Laser Range Scanner|SLAMTEC." [Online]. Available: https://www.slamtec.com/en/Lidar/A2

[13] "Maxbotix," 2025. [Online]. Available: https://maxbotix.com/products/mb1020

[14] T. N. Canh, B. P. Nguyen, H. Q. Tran, and X. HoangVan, "Development of a Human-Robot Interaction Platform for Dual-Arm Robots Based on ROS and Multimodal Artificial Intelligence," 2024.

[15] R. Belwariar, "A* Search Algorithm - GeeksforGeeks," Mar. 2018. [Online]. Available: https://www.geeksforgeeks.org/a-search-algorithm/

[16] J. Sun, G. Liu, G. Tian, and J. Zhang, "Smart Obstacle Avoidance Using a Danger Index for a Dynamic Environment," *Applied Sciences*, vol. 9, no. 8, p. 1589, Apr. 2019, doi: 10.3390/app9081589.

[17] V. Sheoran, S. Joshi, and T. R. Bhayani, "Age and Gender Prediction Using Deep CNNs and Transfer Learning," 2021, pp. 293–304. doi: 10.1007/978-981-16-1092-9_25.

[18] "YOLOv8: A New State-of-the-Art Computer Vision Model." [Online]. Available: https://yolov8.com/

[19] L. Li, K. Ota, and M. Dong, "Deep Learning for Smart Industry: Efficient Manufacture Inspection System With Fog Computing," *IEEE Trans Industr Inform*, vol. 14, no. 10, pp. 4665–4673, Oct. 2018, doi: 10.1109/TII.2018.2842821.

[20] "Depth accuracy," 2025. [Online]. Available: https://docs.luxonis.com/hardware/platform/depth/depth

[21] S. Kim *et al.*, "Squeezeformer: An Efficient Transformer for Automatic Speech Recognition," Jun. 2022.

[22] K. Tripathi, R. Gothi, and P. Wasnik, "Enhancing Whisper's Accuracy and Speed for Indian Languages through Prompt-Tuning and Tokenization," Dec. 2024.

[23] K. S. Kalyan, "A Survey of GPT-3 Family Large Language Models Including ChatGPT and GPT-4," Oct. 2023.

[24] "gmapping - ROS Wiki." [Online]. Available: https://wiki.ros.org/gmapping

[25] Y. Liu, C. Wang, H. Wu, Y. Wei, M. Ren, and C. Zhao, "Improved LiDAR Localization Method for Mobile Robots Based on Multi-Sensing," *Remote Sens (Basel)*, vol. 14, no. 23, p. 6133, Dec. 2022, doi: 10.3390/rs14236133.