INTELLIGENT COMPANION: AN AUTONOMOUS ROBOT FOR REAL-TIME TARGET TRACKING AND INTERACTIVE VOICE ENGAGEMENT

Group Members

1. UMARU _ 12431495
2. 江炜韬 _ 12431497
3. KHOWAJA _ 12431493
4. 朱俊希 _ 12432460

MARCH 14, 2025

SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY [SUSTECH]

School of Design

Table of Content

List of Tables

List of Figures

1. Project Introduction

1.1 Product Research Background

RAISA, which stands for Robots, Artificial Intelligence, and Automation of Services, has become increasingly prevalent in our society. Over the past decade, the convergence of artificial intelligence (AI), robotics, and the Internet of Things (IoT) has transformed how interactive systems are designed and deployed [1]–[3]. Thanks to their advanced technological infrastructures, regions with dynamic manufacturing sectors, such as Guangdong Province in China, are at the forefront of robotics innovation. Our project is poised to meet the growing demand for affordable, autonomous, and interactive robotic systems. We are developing an intelligent companion robot that excels in real-time target tracking and engages users through interactive voice communication. This initiative is driven by academic interest and aligns with market needs, setting the stage for enhanced human-machine interaction across diverse environments. [4]–[6].

1.2 Product Introduction

The Intelligent Companion Robot is a cutting-edge, autonomous AI-driven system that excels at tracking targets in real-time and engaging users through intuitive voice commands. With advanced visual recognition, natural language processing (NLP), and autonomous mobility, this robot is optimally designed to thrive in diverse environments such as shopping malls, museums, and homes. Its innovative design delivers exceptional navigation support, enables personalised user interactions, and offers a scalable platform that adapts seamlessly to various applications. This product sets a bold new standard in robotic assistance, expertly combining deep learning techniques with robust hardware integration to redefine the future of user experience [7], [8].

2. Technological Feasibility and Innovation

2.1 Key Technologies: Visual Recognition, Voice Interaction, and Autonomous Mobility

The success of this project relies on integrating three key technologies: visual recognition, voice interaction, and autonomous mobility.

- Visual Recognition: The robot will have a high-resolution camera and advanced image processing algorithms to detect and track targets in real time. This capability is essential for following a person, recognizing objects, and navigating crowded environments.

- Voice Interaction: The robot will feature a microphone array, speaker system, and NLP algorithms to enable natural language conversations. This will allow users to interact with robots using voice commands, making it more intuitive and user-friendly.

- Autonomous Mobility: The robot will be mounted on a mobility chassis with wheels or tracks, allowing it to move autonomously in various environments. The mobility system will be equipped with sensors, such as LiDAR and ultrasonic sensors, to detect obstacles and navigate safely.

2.2 Deep Learning Algorithms for Target Tracking

Target tracking is a critical component of the robot's functionality. We will employ deep learning algorithms to achieve real-time tracking, specifically Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

- Convolutional Neural Networks (CNNs): CNNs are highly effective for image recognition tasks. We will use a pre-trained CNN model, such as YOLO (You Only Look Once) or SSD (Single Shot MultiBox Detector), to detect and classify objects in the robot's field of view. These models are known for their high accuracy and real-time performance, making them suitable for our application.

- Recurrent Neural Networks (RNNs): RNNs, particularly Long Short-Term Memory (LSTM) networks, will be used to track the movement of objects over time. By analysing the temporal sequence of frames captured by the camera, the robot can predict the future position of a target and adjust its movement accordingly.

2.3 Voice Recognition and Natural Language Processing

Voice interaction is a key feature of the robot, enabling users to communicate with it naturally. To achieve this, we will implement voice recognition and NLP algorithms.

- Voice Recognition: The robot will use a microphone array to capture voice commands. We will employ a voice recognition system based on Hidden Markov Models (HMMs) or Deep Neural Networks (DNNs) to convert speech into text. This system will be trained on a large dataset of voice samples to ensure high accuracy and robustness.

- Natural Language Processing (NLP): Once the voice command is converted into text, NLP algorithms will understand the user's intent and respond appropriately. We will use pre-trained NLP models, such as BERT (Bidirectional Encoder Representations from Transformers) or GPT (Generative Pre-trained Transformer), to handle complex language tasks, including sentiment analysis, question answering, and dialogue management.

3. Hardware Architecture and Integration

3.1 Overview of Core Hardware Components

The hardware architecture of the robot is designed to support its core functionalities, including visual recognition, voice interaction, and autonomous mobility. The key components include:

- **Camera Module**: A high-resolution camera with a wide field of view will be used for visual recognition and target tracking. The XIAO ESP32S3 Sense integrates a camera and supports model training services, making it an ideal choice for visual recognition tasks.

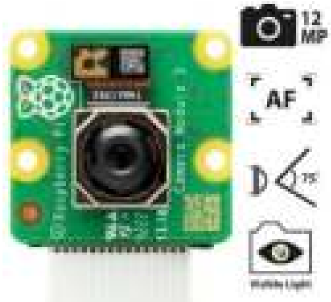


Figure 1. Raspberry Pi Camera Module V3

- **Sensing Module**: The robot will be equipped with a combination of LiDAR, ultrasonic sensors, and infrared sensors to detect obstacles and navigate the environment. The RPLIDAR A1, a low-cost 2D LiDAR sensor, is incorporated in this application.

Function: The ***RPLIDAR A1 is a 2D*** laser radar that can perform 360° omnidirectional laser distance measurement and scanning, suitable for robot navigation, environmental modelling, and map building. As an advantage, it features high precision, high resolution, long service life, safety, and plug-and-play convenience, making it suitable for various environments.

Theory: The ***HC-SR04 ultrasonic sensor*** calculates distance by measuring the propagation time of ultrasonic pulses in the air and using the known speed of sound. It offers high precision, stability, and ease of integration at a low cost, making it suitable for applications like robot obstacle avoidance, smart home automation, and industrial automation.

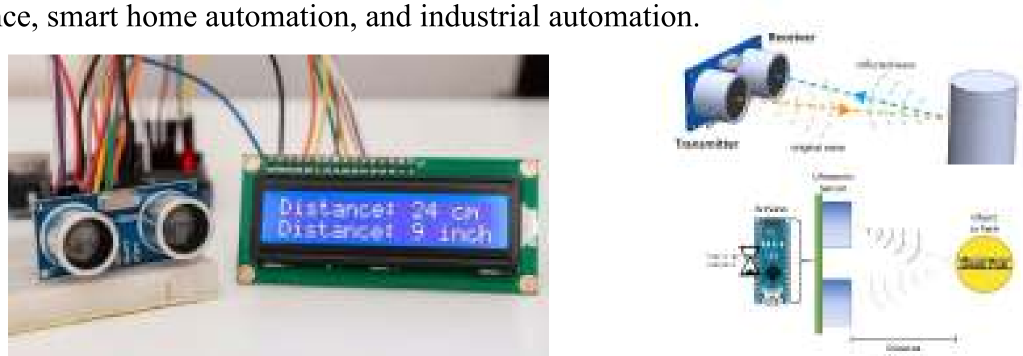


Figure 2. HC-SR04 ultrasonic sensor with Arduino

- **Mobility Chassis**: Depending on the terrain, the robot will be mounted on a four-wheeled or tracked chassis. The Dagu Wild Thumper 6WD chassis is a robust option that provides excellent mobility and stability.

Figure 3. Mobility Chassis types

- **Microcontroller**: The robot's control system will be based on the Seeed Studio XIAO ESP32S3 Sense, which offers sufficient processing power for running deep learning algorithms and controlling the robot's movements. The XIAO's integrated camera and model training services will streamline the development of visual recognition and voice interaction capabilities.

Figure 4. Seeed Studio XIAO ESP32S3

- **Microphone and Speaker**: A microphone array, such as the ReSpeaker 4-Mic Array, will be used for voice recognition. This project will use a high-quality speaker, such as the Adafruit I2S 3W Class D Amplifier, for audio output.

Figure 5. Adafruit I2S 3W Class D Amplifier

Function: The Adafruit I2S 3W Class D Amplifier is a digital audio amplifier module that receives digital audio input in the I2S standard, decodes it into an analogue signal, and directly amplifies it to drive a speaker. This module uses Class D amplification technology, which is highly efficient and suitable for portable and battery-powered projects. It also features over-temperature and over-current protection.

3.2 Camera, Sensing, and Mobility Modules

The camera, sensing, and mobility modules are critical for the robot's ability to navigate and interact with its environment.

- **Camera Module**: The Seeed Studio XIAO ESP32S3 Sense will be used for visual recognition and target tracking. The integrated camera will capture real-time video, which will be processed by the CNN and RNN algorithms for target detection and tracking. XIAO’s model training services will allow us to train custom models for specific tasks, enhancing the robot's visual recognition capabilities.

- **Sensing Module**: The RPLIDAR A1 will create a 2D environment map, enabling the robot to navigate autonomously. Ultrasonic and infrared sensors will be placed around the robot to detect obstacles and avoid collisions. These sensors will provide real-time data to the microcontroller, adjusting the robot's movement accordingly.

- **Mobility Chassis**: The Dagu Wild Thumper 6WD chassis will be controlled by the Seeed Studio XIAO ESP32S3 Sense, which commands the motor drivers to move the robot forward, backwards, or turn. The chassis will have encoders to monitor the robot's speed and position, ensuring precise movement.

3.3 Integration of Control Systems (Microphone, Speaker, Mobility Chassis)

Integrating control systems is essential for the robot's functionality. The microphone, speaker, and mobility chassis will connect to the central processing unit, the Seeed Studio XIAO ESP32S3 Sense.

- **Microphone and Speaker**: The ReSpeaker 4-Mic Array will connect via I2C to capture voice commands, which the voice recognition system will process. An Adafruit I2S 3W Class D Amplifier will output audio responses from the natural language processing (NLP) system.

- **Mobility Chassis**: The Dagu Wild Thumper 6WD chassis will be controlled using a Dual H-Bridge Motor Driver. The microcontroller will send PWM signals to manage wheel speed and direction, while encoders will provide real-time feedback for movement adjustments.

4. Market Demand Analysis

The increasing need for personalised assistance in public and private spaces underpins the market demand for intelligent companion robots. Key trends include likewise domain of application as shown in Table 1 below:

- **Retail and Navigation:** In large shopping malls, consumers benefit from guided navigation and real-time information, reducing confusion and enhancing shopping experiences.
- **Cultural and Educational Venues:** Museums and exhibitions are exploring interactive guides to improve visitor engagement and provide detailed exhibit information.
- **Home and Personal Assistance:** As the population ages, there is a rising demand for affordable, interactive home companions that can assist with daily tasks and provide social interaction.

Market research indicates that integrating AI and robotics enhances customer experience and generates significant cost savings through automation and efficiency improvements.

Table 1 Application Domains [9]

Application domains	Description
Guard robot	Deployed by private agencies in settings such as campuses, airports, and markets to provide guard and protection.
Police robot	Deployed by official police departments to fight crime and ensure safety in public areas.
Military/peace-keeping robot	Deployed in the military and army to protect the safety of civilians or soldiers.
Service robot	1. Public service robot: Deployed by agencies in hotels or markets to provide comprehensive services, including protecting the safety of people by reminding them to wear masks and patrolling to monitor their surroundings.
	2. Private/home service robot: Deployed by individuals or families for house security tasks to protect the safety of family members or individuals.
General security robot	Deployed to perform various security tasks without specific application scenarios or potential employers.

5. Competitive Analysis

5.1 Competitor Analysis

The competitive landscape for intelligent companion robots features several key players:

- **SoftBank Robotics Pepper:** A humanoid robot known for its social interaction capabilities, Pepper offers natural language processing and visual recognition. However, its high cost and limited mobility in complex environments are drawbacks.
- **Anki's Cozmo:** Designed primarily for entertainment and education, Cozmo excels in facial recognition and playful interactions but is limited in real-world navigation and broader functional applications.
- **Amazon's Astro:** Astro integrates sensor technology with voice interaction as a home assistant and security robot. While it offers robust connectivity and monitoring, its market penetration is still early.

Compared to these competitors, our Intelligent Companion robot emphasises a balance of low-cost hardware with high-end software capabilities, delivering an adaptable platform that can be scaled across various sectors. The product's integrated approach to real-time tracking and voice engagement provides a unique advantage in terms of both performance and versatility.

6. Design and Development

6.0 Task Allocation

The project is divided among interdisciplinary teams:

- **Hardware Engineers:** Responsible for designing and integrating sensors, cameras, and mobility chassis.
- **Software Developers:** Focus on developing robust algorithms for visual processing, voice recognition, and autonomous navigation.
- **AI Specialists:** Develop and optimise deep learning models (such as CNNs for object detection and LSTM networks for tracking) and natural language processing frameworks.
- **Project Managers:** Oversee timeline, resource allocation, and risk mitigation strategies.

6.1 Appearance Design and Overall Structure

The robot features a streamlined, ergonomic design, emphasising user-friendliness and visual appeal. Its overall structure includes:

- A robust chassis that allows for smooth movement across various terrains.
- An integrated camera module that acts as the primary visual recognition sensor.
- A compact yet powerful processing unit that supports real-time deep learning computations.

6.2 Function Development

The development of core functionalities focuses on:

- **Real-Time Target Tracking:** Utilizing pre-trained deep learning models (e.g., YOLO or SSD) to detect and classify objects accurately.
- **Voice Interaction:** Implementing state-of-the-art NLP models (such as BERT or GPT) to interpret and respond to user commands in natural language.
- **Autonomous Navigation:** Integrating sensor inputs (from LiDAR, ultrasonic sensors, and infrared detectors) with path planning algorithms (like A* or Dijkstra's) to ensure smooth and collision-free movement.

6.1 Tactile Interaction

Although the primary mode of interaction is voice and visual, the robot is also designed with tactile interfaces:

- **Touch Panels:** Incorporated into the chassis to allow users to activate certain functions or receive haptic feedback.
- **Gesture Sensors:** Enable intuitive physical interactions and complement voice commands in noisy environments or for accessibility.

6.2 "Eyes" Dynamics

The visual system is a critical component:

- **Camera Module:** A high-resolution, wide field-of-view camera captures real-time video, acting as the "eyes" of the robot.
- **Dynamic Tracking:** Adaptive focus and real-time processing allow the robot to adjust its gaze and tracking parameters based on the movement of targets, thereby mimicking human eye behaviour.

6.3 Behavior Control

The behaviour control system is built on a robust software framework that:

- **Decision-Making Algorithms:** Process sensor data and user inputs to determine the robot's actions.
- **Real-Time Feedback:** Uses motor encoders and environmental sensor data to continuously refine movement and interaction patterns.
- **Integration with AI Models:** Allows adaptive learning so the robot improves its responses and navigation strategies over time.

6.4 Status and Intimacy Features

To foster a more personal connection with users, the robot incorporates:

- **Emotional Indicators:** LED displays and subtle mechanical expressions that indicate status (e.g., "listening," "processing," or "idle").
- **User-Centric Design:** Features that adjust based on user preferences and past interactions, enhancing the overall engagement experience.

6.3 Device Integration Process

6.3.1 Project Application

The integrated system is designed for diverse real-world applications:

- **Shopping Malls:** Assisting customers with navigation and store information.
- **Museums:** As an interactive tour guide, they provide exhibit information and answer questions.
- **Home Environments:** Acting as a personal assistant that can help with daily tasks and provide companionship.

6.3.2 Integration Process

The integration of hardware and software is achieved through:

- **Central Processing Unit:** A microcontroller (e.g., Seeed Studio XIAO ESP32S3 Sense) coordinates sensor inputs, voice commands, and movement control.
- **Interfacing Components:** Standardized communication protocols (such as I2C for microphone arrays and PWM for motor control) ensure smooth operation.
- **Iterative Testing:** Prototyping and real-world testing are conducted to refine integration, resolve technical challenges, and ensure reliable performance across all modules.

6.4 Results Display

The project's prototype demonstrations have yielded promising results:

- **Performance Metrics:** High accuracy in target tracking and swift response times in voice interactions.
- **User Feedback:** Positive reception from pilot tests in controlled environments such as simulated malls and homes.
- **Visual Aids:** Graphs, charts, and screenshots document the robot's performance data, highlighting improvements over iterative testing phases.

6.4 Economic Feasibility, Scalability and Sustainability

The economic feasibility of the intelligent companion robot is high, thanks to its low-cost hardware components and scalable software architecture. Using off-the-shelf components, such as the Seeed Studio XIAO ESP32S3 Sense and RPLIDAR A1, keeps the overall cost of the robot low. The robot's software can also be easily adapted for different applications, making it a scalable solution for various industries. The approximate economic analysis is as indicated in Table 2.

Table 2: Price list of hardware components

Name	Price（RMB）
HC-SR04	4.64
MAX98357A	10.00
XIAO ESP32 S3 Sense	79.00
RPLIDAR A1	927.84
Dagu Wild Thumper 6WD	2599.65
Total	3621.13

6.5 Risk Mitigation and Contingency Plans

Several risks could impact the project, including technical challenges, budget overruns, and delays in the timeline. To mitigate these risks, the team will implement contingency plans, such as:

- **Technical Challenges**: Regular testing and prototyping will be conducted to identify and address technical issues early in development.

- **Budget Overruns**: The project budget will be closely monitored, and cost-saving measures will be implemented if necessary.
- **Delays in Timeline**: The project timeline will be flexible, with buffer periods built into each milestone to accommodate delays.

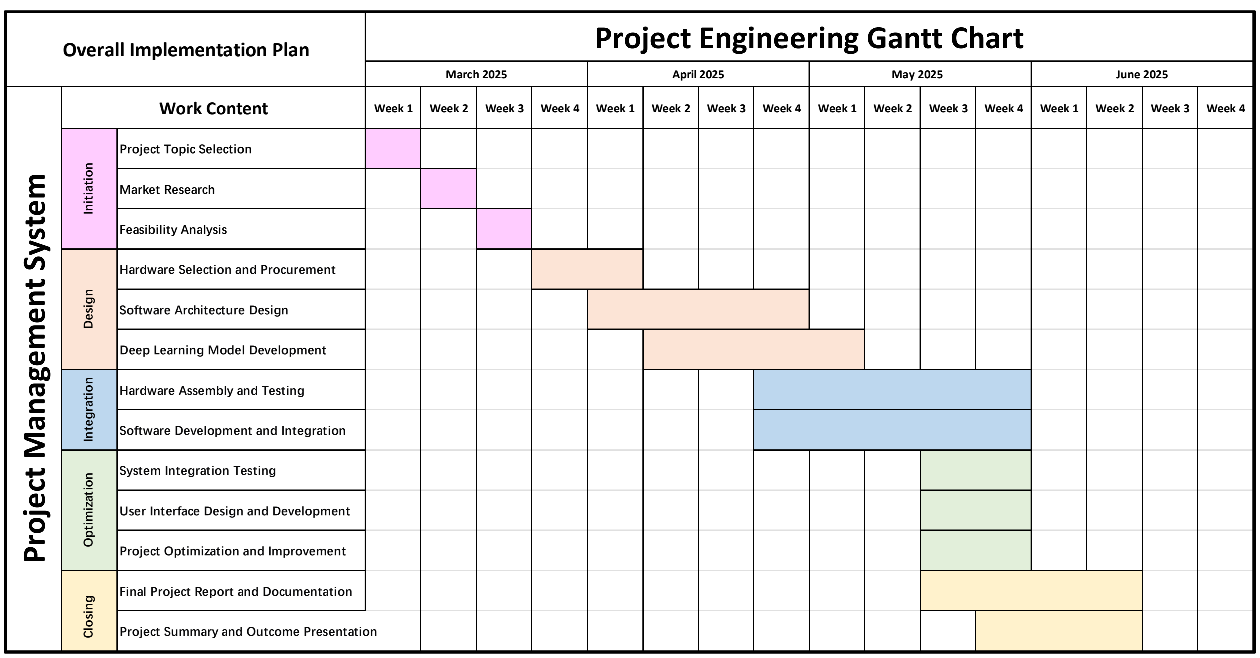
7. Project Summary

7.1 Project Progress and Summary

The Intelligent Companion project will undergo multiple development stages:

- **Research and Development:** Comprehensive research on AI algorithms and sensor integration laid a solid foundation.
- **Hardware and Software Integration:** Successful prototyping and iterative testing confirmed that the robot meets its functional requirements.
- **Future Roadmap:** Plans include further optimising deep learning models, expanding interaction modalities, and broader field testing in real-world environments.
- Remark: The project not only demonstrates technological feasibility but also shows strong market potential. Ongoing refinements will continue to enhance performance, scalability, and user engagement.

Table 3: Project time frame



In conclusion, modern Chinese achievements in intelligent manufacturing and robotics reflect a blend of **visionary policies**, **cutting-edge research**, and **industrial commitment**. These innovations will transform China and inspire global technological advancements as we move forward.

8. References

[1] A. Y. Asmoro, G. Butler, and G. Szili, "Exploring the Current Status and Future Potential of Robot, Artificial Intelligence, and Service Automation in the Indonesian Tourism Industry," *J. Kepariwisataan Destin. Hosp. dan Perjalanan*, vol. 7, no. 2, pp. 133–153, 2023, doi: 10.34013/jk.v7i2.1226.

[2] K. Miao, C. Chen, and X. Zheng, "A study on the application of artificial intelligence in the design of intelligent medical robots," *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, 2024, doi: 10.2478/amns.2023.2.01388.

[3] M. Soori, B. Arezoo, and R. Dastres, "Artificial intelligence, machine learning and deep learning in advanced robotics, a review," *Cogn. Robot.*, vol. 3, pp. 54–70, 2023, doi: https://doi.org/10.1016/j.cogr.2023.04.001.

[4] F. Arvin, J. Espinosa, B. Bird, A. West, S. Watson, and B. Lennox, "Mona: an Affordable Open-Source Mobile Robot for Education and Research," *J. Intell. Robot. Syst.*, vol. 94, no. 3, pp. 761–775, 2019, doi: 10.1007/s10846-018-0866-9.

[5] K. Berezina, O. Ciftci, and C. Cobanoglu, "Robots, Artificial Intelligence, and Service Automation in Restaurants," in *Robots, Artificial Intelligence, and Service Automation in Travel, Tourism and Hospitality*, S. Ivanov and C. Webster, Eds. Emerald Publishing Limited, 2019, pp. 185–219. doi: 10.1108/978-1-78756-687-320191010.

[6] J. Bowen and E. Whalen, "Trends that are changing travel and tourism," *Worldw. Hosp. Tour. Themes*, vol. 9, no. 6, pp. 592–602, Jan. 2017, doi: 10.1108/WHATT-09-2017-0045.

[7] E. Broadbent, M. Billinghurst, S. G. Boardman, and P. M. Doraiswamy, "Enhancing social connectedness with companion robots using AI," *Sci. Robot.*, vol. 8, no. 80, p. eadi6347, 2023, doi: 10.1126/scirobotics.adi6347.

[8] M. A. Goodrich and A. C. Schultz, "Human–Robot Interaction: A Survey," *Found. Trends® Human–Computer Interact.*, vol. 1, no. 3, pp. 203–275, 2008, doi: 10.1561/1100000005.

[9] X. Ye and L. P. Robert, "A Human–Security Robot Interaction Literature Review," *J. Hum.-Robot Interact.*, vol. 14, no. 2, Dec. 2024, doi: 10.1145/3700888.