PyLoT Robotics 2026 Team Description Paper

Towa Yamashita Riku Kinoshita Katsuya Honda
Koki Muramoto

January 9, 2026

Abstract. This paper introduces Runa, a mobile manipulator newly developed for RoboCup@Home 2026 by the PyLoT Robotics team at Kaijo Junior and Senior High School, and outlines our software development and contributions. We designed an affordable, modular hardware platform and, on top of Robot Operating System 2 (ROS 2), built a software stack for object recognition and grasping, speech recognition and synthesis, and task planning using Large Language Models (LLM). Together, these components aim to deliver a system capable of general-purpose service tasks.

1 Introduction

PyLoT Robotics is a RoboCup team affiliated with Kaijo Junior and Senior High School. Founded in 2023 by junior and senior high school students, the team is fully student-led, covering administration, development, and outreach. We participate in regional RoboCup@Home leagues and provide STEM programs for middle school students, offering hands-on learning in robotics and autonomous systems. We have steadily improved our stack for perception, grasping, and planning. This paper reports our progress toward RoboCup@Home 2026, including spatial object detection using image and point-cloud processing, improved arm control, and more accurate navigation. We also work to make robotics more accessible and to grow the community among middle and high school students.

2 Hardware

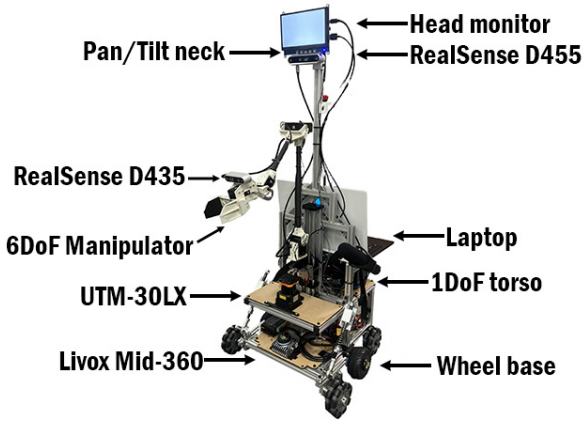

Fig. 1. Runa robot platform

PyLoT Robotics develops compact robots designed for high maintainability, as shown in Figure 1. The chassis utilizes a differential two-wheel drive system, while the main body features a disassemblable design using an aluminum frame and Medium Density Fiberboard (MDF), achieving light weight and low cost. Additionally, 3D-printed parts are used for various connection points, ensuring excellent maintainability and easy component replacement. The robot's underside houses a Hokuyo UTM-30LX laser scanner for self-localization and navigation. The central section mounts a 6 Degrees of Freedom (DOF) robotic arm based on the Openmanipulator-x with modifications, along with various circuit boards, a small monitor, and an emergency stop button. The upper section features an Intel RealSense RGB-Depth (RGB-D) camera and an orthogonal two-axis pan-tilt unit, enabling a wide range of functions including object recognition and obstacle detection. A laptop PC is mounted at the rear center of the robot body, serving as the computational resource for controlling the robot.

3 Software

3.1 Software Overview

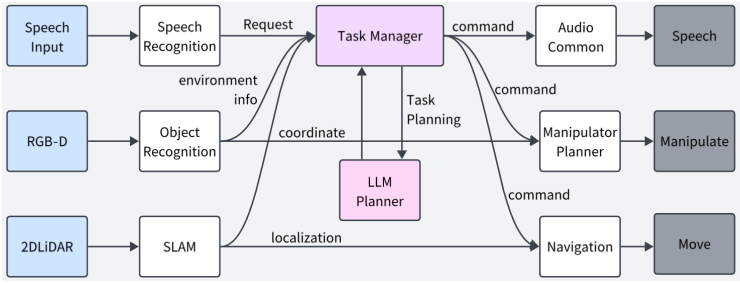

Fig. 2. System Overview

As illustrated in Figure 2, our system integrates foundation models for end-to-end task execution. Speech input is transcribed by Whisper and processed by an LLM to extract user intent. For perception, we combine Detic-based full-scene segmentation with CLIP zero-shot classification, enabling recognition of unknown objects in RoboCup environments. Semantic information from these Vision-Language Models (VLM) is fused into a semantic map that supports navigation and manipulation planning.

3.2 Speech Recognition

Our system uses Whisper[1] to convert speech into text. Upon dialog start, the spoken input is converted into text based on predefined prompts, which are then fed into an LLM such as GPT-4[2]. This enables the robot to simultaneously understand the user's intent and determine what tasks to perform.

3.3 Object Detection

To address the unpredictability of RoboCup objects, we adopt a two-stage VLM pipeline. First, Detic provides open-vocabulary segmentation of all visible objects, and then CLIP performs prompt-based zero-shot classification. Carefully crafted prompts further improve recognition accuracy. The resulting semantically labeled point cloud is used for downstream planning, as shown in Figure 3.

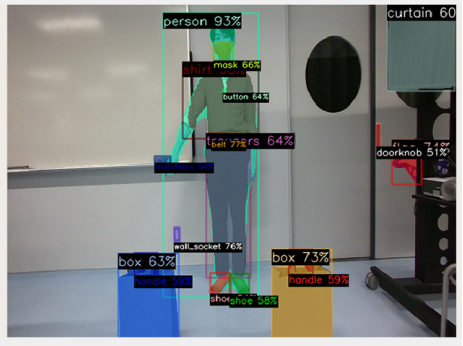

Fig. 3. Object Detection using Detic

3.4 Grasping

We implemented a custom inverse-kinematics module for our robot arm. Target object coordinates from semantic perception are transformed into the arm frame, and joint and Cartesian trajectories are planned in parallel while avoiding obstacles.

3.5 Task Planning

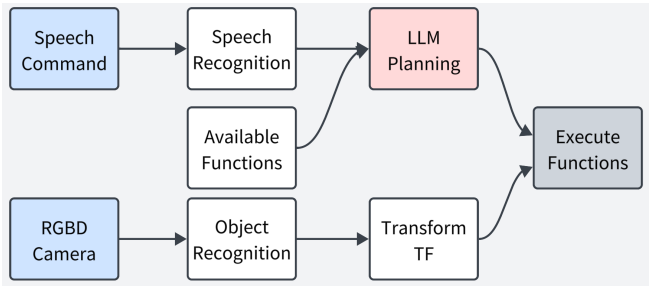

Fig. 4. Overview of an LLM-based speech-to-plan system

As shown in Figure 4, an LLM-based planner converts natural-language commands into executable action sequences. Leveraging predefined action functions, object names, and environment information, the system produces structured plans capable of solving General Purpose Service Robot (GPSR) and Enhanced GPSR (EGPSR) tasks.

3.6 Following a person

In our "Following a Person" module, we developed a navigation framework that couples A*-based global path generation with nonlinear Model Predictive Control (MPC) for robust trajectory tracking. The A* planner yields collision-free routes, while the nonlinear MPC controller ensures smooth, stable following in dynamic environments. This approach improves the robot's ability to maintain distance, react to human motion, and navigate safely in cluttered spaces. Our system performed strongly in the 2025 RoboCup@Home Open Platform League (OPL) Salvador, achieving one of the highest scores in the Help Me Carry task and demonstrating the effectiveness of our navigation approach.

3.7 Active SLAM

We developed an Active Simultaneous Localization and Mapping (SLAM) system that runs SLAM and navigation concurrently in previously unknown environments. By updating the map and planning motion in real time with SlamToolbox[3], the robot localizes accurately and generates reliable paths without prior environmental information. This enables precise, adaptive movement and stable navigation in dynamic or unstructured spaces. The approach earned high scores at both the 2025 RoboCup@Home Japan Open and the 2025 RoboCup@Home OPL Salvador, demonstrating its effectiveness in real competition settings.

4 Our Research

4.1 Self Recovery Planning using VLM

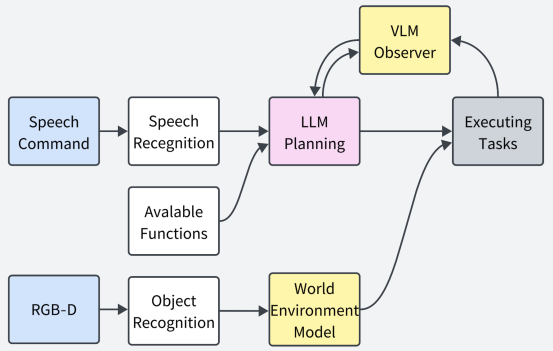

Fig. 5. Diagram of a robot planning system using VLM

As shown in Figure 5, we aim to boost planning robustness by pairing an LLM planner with a VLM verifier. The LLM generates an action plan, the Task Executor runs the corresponding functions, and upon state completion, the LLM

queries the VLM for visual confirmation. Using camera input, the VLM reports task status, and the LLM performs replanning when needed, enabling self-recovery. Because the VLM has no built-in robot semantics, it focuses on the target object and target location and generates prompts to assess the running function. A paper on this topic is currently under conference review.

4.2 Semantic Map using VLM

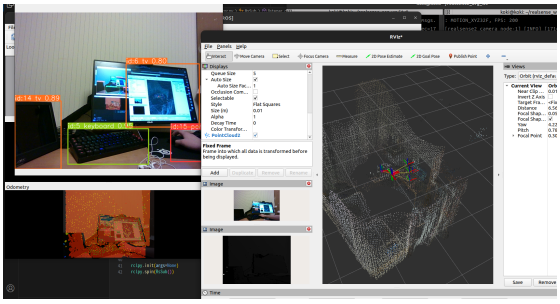

Fig. 6. Semantic map constructed using VLM

As illustrated in Figure 6, we aim to construct a world environment model with semantic information so that robots can execute general tasks efficiently. The world model, which combines location and semantic information, is inspired by CLIP-Fields[4], and organizes knowledge in a cohesive space.

4.3 Adaptive Pan-Tilt Camera for Target Recognition

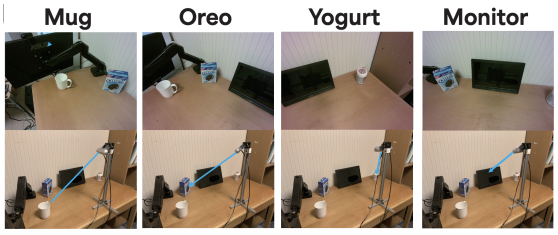

Fig. 7. Viewpoint control for target recognition

As shown in Figure 7, prior work assumed the camera always captured all information needed for self-recovery. We remove that assumption and build a system

that actively gathers the missing observations required to complete a task, targeting mobile manipulators that need 360-degree monitoring.

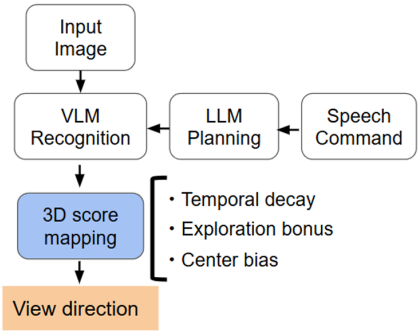

Fig. 8. Viewpoint control system overview

We have constructed a system as shown in Figure 8. Given verbal instructions, the VLM assesses object relevance and selects camera angles based on the object distribution. The distribution decays over time, and a gain toward long-unobserved directions drives simultaneous search.

5 Contribution to STEM Education

PyLoT Robotics expands access to robotics through outreach programs and step-by-step curricula, contributing to STEM education in our community. We provide structured courses, documentation, and tools that enable middle school students to gain early, practical experience in robot development. This hands-on approach builds skills in programming, mechanical design, and system integration.

As a result, two junior high school students joined our team and successfully completed the Receptionist task at the 2025 RoboCup Japan Open @Home Bridge Competition. Additionally, we achieved one of the highest scores in the Help Me Carry task at the 2025 RoboCup@Home OPL Salvador, demonstrating that junior and high school students can perform at an international level. These outcomes highlight our educational impact and the effectiveness of our mentorship in fostering the next generation of roboticists.

6 Conclusions

This paper presented PyLoT Robotics's robot platform, scientific contributions, and approach for the RoboCup@Home 2026 competition. Through participation in RoboCup@Home, we aim to advance autonomous home service robots and make robotics more accessible to the public. Aligned with open-source principles, we will continue to release our code, documentation, and results, and contribute to the RoboCup community beyond the competition.

Acknowledgements

This paper is supported by Shimon Ajisaka, Dai Komukai, Yoshihiro Shibata and Takumi Nakamura.

References

1. Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, pages 28492–28518. PMLR, 2023.
2. OpenAI. Gpt-4 technical report. Technical report, OpenAI, 2023.
3. Steve Macenski and Ivona Jambrecic. Slam toolbox: Slam for the dynamic world. *Journal of Open Source Software*, 6(61):2783, 2021.
4. Nur Muhammad Mahi Shafiullah, Chris Paxton, Lerrel Pinto, Soumith Chintala, and Arthur Szlam. Clip-fields: Weakly supervised semantic fields for robotic memory. In *Robotics: Science and Systems*, 2023.
5. Xingyi Zhou, Rohit Girdhar, Armand Joulin, Philipp Krähenbühl, and Ishan Misra. Detecting twenty-thousand classes using image-level supervision. In *European Conference on Computer Vision (ECCV)*, 2022.
6. Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021.
7. Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Yolo by ultralytics. `https://github.com/ultralytics/ultralytics`, Jan 2023.
8. Mathieu Labbé and François Michaud. Rtab-map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. *arXiv preprint arXiv:2403.06341*, 2024.
9. Steve Macenski, Tully Moore, David V Lu, Alexey Merzlyakov, and Michael Ferguson. From the desks of ros maintainers: A survey of modern and capable mobile robotics algorithms in the robot operating system 2. `https://github.com/ros-navigation/navigation2`, 2023.

Robot Runa Hardware Description

Robot Runa has the custom garbage collection mechanism. Specifications are as follows:

- Base: wheel base (differential pair), 2.5m/s max speed.
- Arm: 7DOF(1 DOF torso, 6DOF manipulator)
- Neck: 2DOF
- Head: Depth Camera, monitor display
- Robot dimensions: height: 1.4m (max), width: 0.6m depth 0.8m
- Robot weight: 30kg.
- RGB-D Sensors: Intel RealSense D435, Intel RealSense D455.
- LiDAR: Hokuyo UTM-30LX
- Microphones: CVM-V30PRO

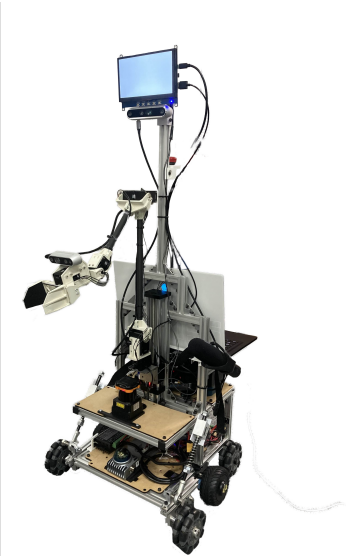

Fig. 9. Robot Runa

Robot's Software Description

For our robot we are using the following software:

- Platform: ROS2 Humble
- Navigation: Navigation2,EMCL
- Face recognition: Yolov8-Pose,Detic,GPT-4o
- Speech recognition: Whisper,Vosk.
- Speech generation: Audio-Common.
- Object recognition: Detic,CLIP,Yolov8
- Arms control: In-house arm motion planner

The following are the specifications of the laptop mounted on our Robot

- CPU: Ryzen7
- GPU: NVIDIA Geforce RTX 3050ti
- Memory: 16GB