2nd German Robotics Conference (GRC), Cologne, Germany, 2026.

Adaptive Domestic Service Robotics through Foundation Models for Perception, Interaction, and Action

Raphael Memmesheimer, Dmytro Pavlichenko, Evgenii Kruzhkov, Jonas Bode,
Fynn Schilke, Vitalii Tutevych, Christian Lenz, Michael Schreiber, and Sven Behnke

I. INTRODUCTION

This extended abstract presents an overview of the approaches and results of the NimbRo@Home team, winner of the RoboCup@Home 2024 Open Platform League (OPL) in Eindhoven, Netherlands [1], [2] and runner-up of the RoboCup@Home 2025 OPL in Salvador, Brazil. The competition evaluates domestic service robots in realistic household scenarios, requiring robust perception, interaction, manipulation, and autonomous task execution. Our work focuses on advancing generalization and robustness through the integration of foundation models for perception, human-robot interaction, and planning.

A central contribution is the deployment of foundational models like Large Language Models (LLMs) for human-robot interaction and task planning, Vision Language Models (VLMs) for scene state analysis, and open-vocabulary object segmentation for perception and grasping. Traditional RoboCup systems rely on closed-set, supervised vision pipelines that require extensive labeling and retraining for each competition environment. We demonstrate that promptable vision models enable segmentation and manipulation of previously unseen objects described only by natural language. This significantly reduces labeling overhead and allows the robot to adapt on the fly. We combine these models with a closed-set pipeline based on YOLO and MaskDINO, yielding a hybrid perception system that balances robustness, speed, and flexibility across tasks.

II. HARDWARE

Our hardware platform is a modified PAL Robotics TIAGo++ Omni robot [3] (see Fig. 1) equipped with dual 7-DoF arms, an Orbbec Gemini 2 RGB-D camera, an Ouster OSDome 3D LiDAR, a wide-FoV camera registered against the 3D LiDAR, and an onboard Zotac ZBOX QTG7A4500 GPU workstation. An additionally installed Zoom AM7 microphone, a Bose SoundLink Max speaker, and a front-facing screen support the human-robot interaction. Custom silicone-coated grippers improve handling of deformable and slippery objects. We focus on local model inference; however, employing a hybrid Wi-Fi/5G connection ensures reliable access to online services, including LLMs and VLMs. To optimize the manipulability in domestic environments, we extended the height of the robot by 200 mm.

All authors are with Autonomous Intelligent Systems, Computer Science Institute VI, University of Bonn, 53115 Bonn, Germany, nimbro-athome@ais.uni-bonn.de

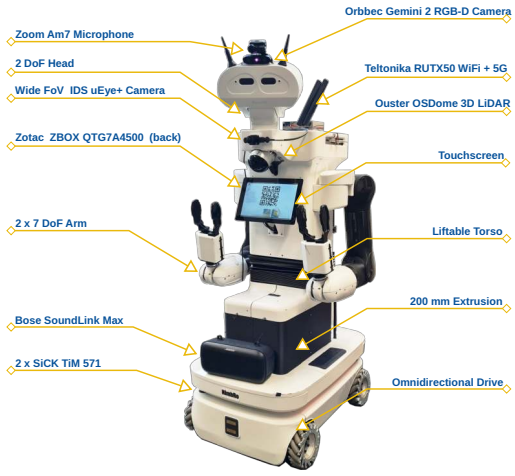

Fig. 1. Modified PAL Robotics TIAGo++ robot.

III. APPROACHES

For manipulation, segmented depth data is transformed into partial point clouds, optionally completed via 3D model registration. Grasp poses are sampled, collision-checked against fused RGB-D and LiDAR data, and ranked using a heuristic cost function considering reachability, clearance, and workspace margins. This pipeline enables reliable grasping of both known and unknown objects.

NimbRo@Home's 2026 approach centers on robust, adaptable perception driven by semi-automated learning pipelines. Object understanding relies on Mask DINO for pixel-wise instance segmentation, pre-trained on large-scale datasets and rapidly adapted on-site through a two-stage labeling process. Pretrained models propose object segments, which are refined and labeled using an ensemble of vision foundation models (CLIP, ViT, Theia) and minimal per-class examples. Interactive correction with Segment Anything and CVAT completes the loop, enabling fast creation of task-specific datasets. This workflow allows detectors to be reconfigured during competition, supporting new objects and environments with minimal manual effort.

Human perception integrates multiple modalities. YOLO V8 pose estimates full-body poses, while RetinaFace and metric embeddings enable face recognition and re-identification. Person tracking fuses RGB-D detections with 2D LiDAR leg tracks, allowing recovery after occlusions and reliable long-term interaction. For manipulation, grasp

(a) Aligning to the door handle (b) Pressing and pulling the door (c) Pushing the door with two arms

Fig. 2. Door opening to the inside of the apartment during the RoboCup@Home OPL 2025 finals in Salvador, Brazil.

(a) Throwing away trash (b) Closing cupboard door (c) Opening door to the outside

Fig. 3. Various mobile manipulation capabilities as demonstrated during the RoboCup@Home OPL 2025 finals in Salvador, Brazil.

planning starts from segmented depth data, optionally registered to 3D object models. Candidate grasps are sampled, filtered for reachability and collisions, and ranked by a cost function considering approach feasibility, safety margins, and workspace constraints, symmetrically across both arms. Grasp success is verified through gripper encoder feedback and point-cloud evidence near the hand, ensuring reliability even with thin objects.

Navigation and interaction emphasize autonomy in dynamic, human-centered environments. Mapping and localization use SLAM Toolbox [4] and AMCL, with region and pose markers encoding semantic structure. Navigation builds on ROS 2 with global A* planning and a model predictive controller for smooth omnidirectional motion, fusing 2D and 3D LiDAR data for obstacle-aware costmaps. Audio processing combines denoising, voice activity detection, multilingual speech recognition, and neural text-to-speech synthesis. Large Language Models (LLMs) interpret free-form user commands into structured function calls [5] that reflect executable robot actions. Behavior control is organized through hierarchical state machines whose parameters can be derived from natural language, enabling flexible task execution today and paving the way for learning behaviors from demonstration in the future.

IV. RESULTS

Across Stage 1 and Stage 2 tasks—including Carry My Luggage, Receptionist, GPSR, Restaurant, and Clean the Table—the system demonstrated strong robustness and generalization. In the 2024 final demonstration, two robots showcased environment scanning via open-vocabulary perception and the manipulation of unlabeled objects, culminating in pouring an egg into a pan. NimbRo@Home achieved the highest overall score and received the "Waitress Captain" award for the best performance in the Restaurant task.

TABLE I

ROBOCUP@HOME 2024 AND 2025 RESULTS.

Year	Competition	Position
2025	RoboCup World Championship Sao Paulo, Brazil	2nd
2025	RoboCup German Open Nuremberg, Germany	1st
2024	RoboCup World Championship Eindhoven, Netherlands	1st
2024	RoboCup German Open Kassel, Germany	1st

In 2025 we showcased incorporated VLMs utilization for scene analysis and deployed it in our manipulation pipeline to constrain manipulation dependent on the analyzed scene state, e.g., by estimating the door state of furniture items like cupboards or shelves. We further showed advanced manipulation with two arms for opening apartment doors to the inside and to the outside (see Fig. 2). Further, we threw misplaced or trash items into the trash bin, closed cupboards, and opened doors to the outside (see Fig. 3).

Results of our team's performance from 2024 to 2025 are contained in Table I. Within this period we successfully performed in two RoboCup@Home World Championships and won two RoboCup@Home German Open competitions.

V. CONCLUSION

We conclude that open-vocabulary perception combined with LLM-based planning substantially advances domestic service robotics. Hybrid architectures that integrate foundation models with task-specific pipelines enable adaptability in unstructured environments, reduce manual engineering effort, and move RoboCup@Home closer to real-world deployment. We demonstrated the success of our system and approaches by winning the RoboCup@Home 2024 OPL and running up the RoboCup@Home 2025 OPL and further winning two RoboCup@Home German Open competitions.

References

[1] R. Memmesheimer, J. Nogga, B. Pätzold, E. Kruzhkov, S. Bultmann, M. Schreiber, J. Bode, B. Karacora, J. Park, A. Savinykh, et al., "Robocup@ home 2024 opl winner nimbro: Anthropomorphic service robots using foundation models for perception and planning," in *Robot World Cup*, Springer, 2024, pp. 515–527.

[2] R. Memmesheimer, J. Bode, M. Splietker, S. Bultmann, B. T. Imbusch, and S. Behnke, *NimbRo@Home 2023 Open Platform League team description*, RoboCup@Home Team Description Papers, 2023.

[3] J. Pages, L. Marchionni, and F. Ferro, "Tiago: The modular robot that adapts to different research needs," in *International workshop on robot modularity, IROS*, sn, vol. 290, 2016, p. 98.

[4] S. Macenski and I. Jambrecic, "Slam toolbox: Slam for the dynamic world," *Journal of Open Source Software*, vol. 6, no. 61, p. 2783, 2021.

[5] J. Bode, B. Pätzold, R. Memmesheimer, and S. Behnke, "A comparison of prompt engineering techniques for task planning and execution in service robotics," in *2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids)*, IEEE, 2024, pp. 309–314.