# Demonstration of an Open-source ROS 2 Framework and Simulator for Situated Interactive Social Robots

Sara Cooper, Raquel Ros, Séverin Lemaignan, Ferran Gebellí, Lorenzo Ferrini, Luka Juričić
*PAL Robotics*
Barcelona, Spain
`name.surname@pal-robotics.com`

***Abstract*—We introduce an open-source ROS 2 architecture for situated social robots, along with a simulator that allows mixed-reality development and interactions. The architecture is a hybrid symbolic/subsymbolic system that integrates explicit ontology semantics for perception, reasoning, and execution, with LLMs. It features multimodal social perception by leveraging the open source ROS4HRI framework; LLMs (both edge- and cloud-based) to facilitate natural language interaction between the user and system; KnowledgeCore, an open-source knowledge base, to reason about facts in the world; and an intent-based controller to supervise the execution of parallel/sequential tasks and skills. We demonstrate our system architecture with a social robot running the mixed-reality system.**

***Index Terms*—Situated social robots; ROS 2 framework; mixed-reality simulator**

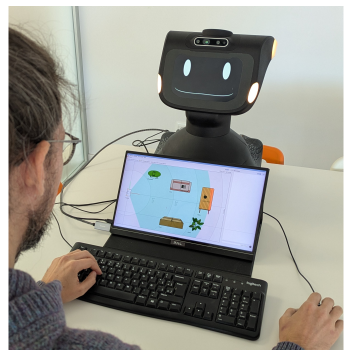

Fig. 1. User interacting with the robot and the mixed-reality simulator.

## I. Introduction

This demonstration showcases a modern take on hybrid symbolic/sub-symbolic architectures for social robots. Our architecture is open-source and is based on the Robot Operating System (ROS) 2. It integrates a variety of open-source tools and libraries to support the development of situated social robots. We demonstrate our architecture with a social robot running in a mixed-reality system (Fig. 1).

Our work is motivated by the recent surge of robotic systems that use Large Language Models (LLMs) to act as "statistical controllers" for robots [1]–[3]. While these systems have shown impressive capabilities in natural language understanding and generation, they lack situatedness, i.e. the ability to reason about the physical and – perhaps more importantly – the social world, and to interact with it in a meaningful way. In addition, LLMs are often used as black boxes, making it difficult to understand and debug the robot's behavior, and falling short of the transparency and accountability that are essential for social robots [4].

Our hybrid architecture combines explicit symbolic modeling of the robot's physical and social environment, similar to e.g. [5], with the strengths of LLMs. We re-use and extend an established ontology (the OpenRobots Ontology [6]) whose semantics are aligned on a human-level taxonomy of the world, and thus semantically close to the LLMs representation space. By describing at run-time the robot's environment with this ontology, we effectively leverage symbolic world modeling as a bridge between, on the one hand, the robot's situated perception of a continuous world and, on the other hand, the abstract LLMs.

Our open source architecture is built on top of the ROS 2 middleware, thus allowing integration of ROS 2 modules; it leverages the ROS4HRI [7] framework to represent humans interacting with the robots; and integrates several model-based design principles, both based on the work by Stampfer [8] and our own research on modeling user's *intents*. The architecture is general and can be applied to a wide range of social robots. It is currently deployed on several off-the-shelf robots like the PAL Robotics TIAGo Pro and TIAGo Head (pictured in Fig. 1) robots, as well as part of a novel interaction simulator that allows mixed-reality development and testing of social robots in semantic rich environment. The main goal is provide tools for researchers to build cognitive systems exploiting symbolic contextual world modeling in decision-making and execution.

## II. The architecture

The system architecture is depicted in Figure 2 (right). We distinguish three main sub-systems described next.

### A. Social Perception

The social perception pipeline is based on the ROS4HRI framework [7]. While the framework allows multiple sources of information to represent people in the environment, in this work, we focus on two main features: 'faces' (`/human/faces`) and 'voices' `/human/voices`. The former allows us to represent a person based on the detected faces

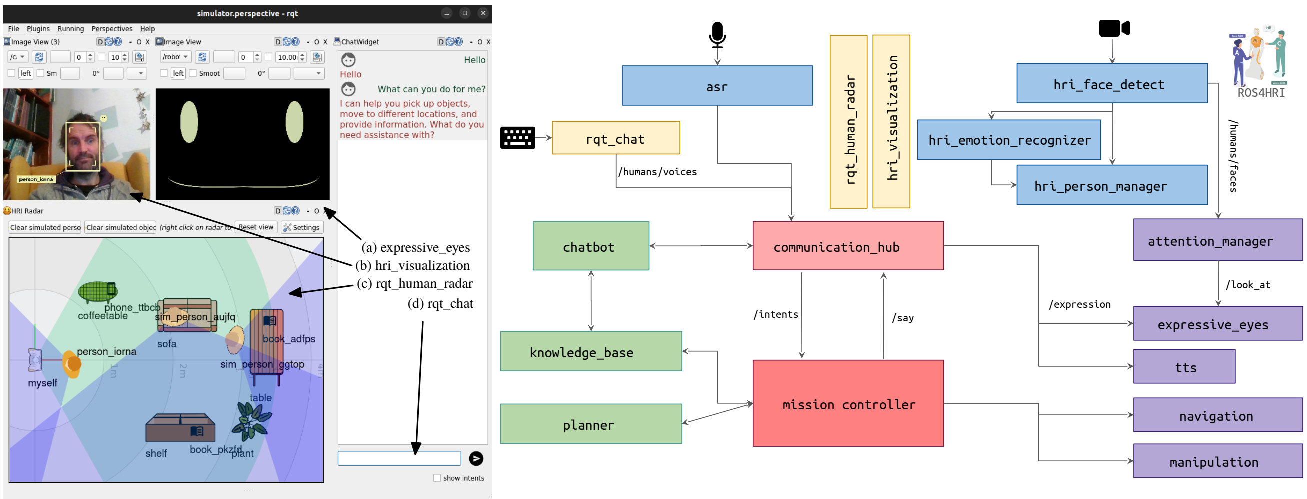


Fig. 2. (left) Screenshot of the Interaction simulator; (right) Overview of the ROS 2 architecture. Blue blocks correspond to the perception sub-system; green blocks, to the reasoning sub-system; red blocks, to the supervision sub-systems; purple blocks, to the skills (functional) system capabilities (navigation and manipulation represent simulated skills in the current work). Blocks in yellow are components of our open-source interaction simulator.

in the images from the camera (assigning ids and estimated pose) while expanding on their facial expressions. The latter provides us with verbal input to the system, which can be obtained either from the microphone or through the keyboard.

### B. Situated Reasoning

The Situated Reasoning subsystem integrates any OpenAI-compatible LLM (e.g. ollama, ChatGPT) with the robot's knowledge base to enable natural interaction and context-aware decision-making.

The robot uses the OpenRobots Ontology [6] for symbolic reasoning with OWL2 RL semantics, providing a structured vocabulary to describe the environment and infer relationships.

The system monitors the environment every two seconds, identifying furniture objects, and their attributes (e.g., color, contents, name). Environmental data is formatted into a structured ROS topic prompt (Fig. 3b), so that updates reflect changes in objects, furniture, or people, including their addition, removal, or movement. It includes the main user interacting with the robot (`user_id`), corresponding to the first person detected by the perception system.

This prompt is passed to the LLM every time the user talks to the system verbally, or through the simulator's chat interface (Fig. 2 left, (d)), which processes the information and generates actionable intents. Fig. 3 shows an example of the reasoning input-output based on the user's request.

### C. Intent-based Execution

We abstract users' 'desires' into *intents* [9] that the robot can process. Intents can vary in complexity, from simply speaking a utterance to the user, to sequences of tasks that shall be performed while adapting to the environment at execution time. The mission controller is handling incoming intents, and delegates to a planner the generation of a specific task plan based on the available skills of the robot. For demonstration purposes, we have limited the skills to simulated `grab`,

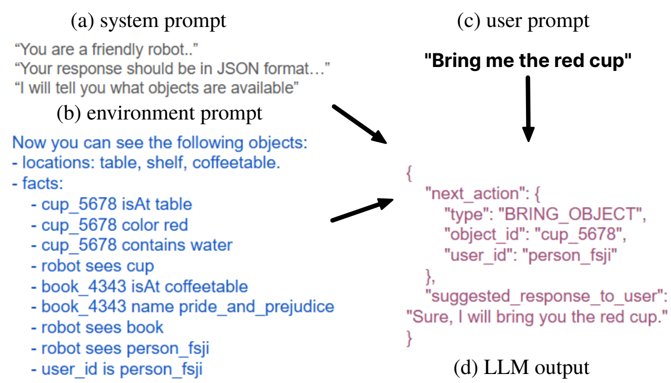


Fig. 3. Example of three prompts being combined and submitted to ChatGPT `4o-mini`. It returns an intent and suggested response as a JSON output.

`go_to`, `release` and `say`. From these, we can resolve goals such as bringing an object, giving it to someone, taking it somewhere, or picking or releasing it, while communicating with the human.

## III. Mixed-Reality Interaction Simulator

The interaction simulator that we will demonstrate is designed to emulate significant components of a typical ROS4HRI pipeline, as well as semantic scene understanding (by placing objects with known semantics in the environment).

The simulator is implemented as a collection of both standard and custom ROS RQt plugins (Fig. 2 left). It is an hybrid simulator in two ways: on the one hand, it integrates multiple actual ROS4HRI modules such as `hri_face_detect`, `hri_person_manager`, `knowledge_core`; on the other hand, it provides a mixed-reality environment, where the robot can interact both with virtual humans and objects, as well as real user, detected and tracked from their webcam.

The simulator features the following main components:

| featuring capability | world state (KB facts) | dialogue | execution |
|---|---|---|---|
| chatting about the world state | a human, a romance book called Pride and Prejudice on the sofa, a fantasy book called Harry Potter on the table | H asks a suggestion for a relaxing activity<br>R suggests reading and describes the two books<br>H requests more info on the one on the sofa<br>R talks about Pride and Prejudice | – |
| fulfilling a human desire | a human, a book on the shelf | H asks for a book<br>R asks if any specific one<br>H responds any available one | R goes to the shelf, grabs the book,<br>R goes to the sofa where the user is<br>R releases the book |
| handling multiple objects | a human, the robot holding a book next to the shelf, a coffee cup on the coffee-table, a water cup on the table | H asks the robot to pick up the cup on the coffee table | R goes to the coffee table<br>R holds the coffee cup and the book<br>`[the book moves to the coffee-table area as well]` |
| | robot holding book and coffee cup next to coffee-table | H asks the robot to leave the book on the sofa | R goes to the sofa<br>R releases the book<br>`[the cup moves to the sofa area as well]` |
| | book at the sofa, robot holding coffee cup next to the sofa | H asks the robot to leave the cup on the table | R goes to the table and releases the cup |
| selecting objects based on context | a human, a coffee cup and a water cup on the table | H is thirsty and asks for a drink<br>R proposes two drinks<br>H says she's sleepy and asks to leave the drink on the coffee-table | R goes to table and picks up coffee cup<br>R goes to the coffee-table, releases it |
| handling multiple desires | a human, a coffee cup on the coffee-table, a water cup on the table, a book on the sofa | H asks the robot to bring the coffee cup | R goes to coffee table, picks up coffee cup<br>R approaches the user |
| | | H asks the robot to take the book to the shelf while still executing the previous plan | R gives the coffee cup to the user<br>R moves to the sofa, grabs the book<br>R moves to the shelf, and releases the book |

TABLE I

MIXED-REALITY USE CASES DEMONSTRATING THE BENEFIT OF INTEGRATING LLMS AND KNOWLEDGE REPRESENTATION IN SITUATED INTERACTIONS (R AS ROBOT, H AS HUMAN)

- `hri_visualization` (Fig. 2b): identifies and tracks faces and bodies in the environment, determining their poses. It assigns unique IDs to individuals and detects facial emotions, contributing to enriched interaction data;
- `rqt_human_radar` (Fig. 2c): monitors and localises nearby individuals, tracks relative positions and provides the capability to introduce new virtual objects and persons for simulated scenarios;
- `expressive_eyes` (Fig. 2a): facilitates the dynamic display of the robot's face, allowing the adjustment of expressions to simulate various emotional states[1];
- `rqt_chat` (Fig. 2d): a chat interface that allows the user to interact with the robot through text, providing a way to test the reasoning capabilities of the system.

## IV. USE CASES DEMONSTRATION

The interaction simulator allows us to quickly develop mixed reality environments to evaluate the reasoning and supervision capabilities of the system. Table I exemplifies the work presented through five situations that feature several key characteristics of the system. All of these scenarios will be demonstrated at the conference with a setup similar to Fig. 1.

## V. ACKNOWLEDGEMENTS

The work is funded by the Horizon Europe CoreSense (no. 101070254) and ARISE (no. 101135784) projects.

[1] `expressive_eyes` is not currently open-source.

## REFERENCES

[1] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, and et al, "Training language models to follow instructions with human feedback," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Curran Associates, Inc., 2022, pp. 27 730–27 744.

[2] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, J. Dabis, and et al., "RT-1: Robotics Transformer for Real-World Control at Scale," in *Robotics: Science and Systems XIX*. Robotics: Science and Systems Foundation, 2023.

[3] S. Vemprala, R. Bonatti, A. Bucker, and A. Kapoor, "ChatGPT for robotics: Design principles and model abilities," Microsoft, Tech. Rep., 2023.

[4] H. level Expert Group on Artificial Intelligence, "Ethics guidelines for trustworthy ai," European Commission, Tech. Rep., 2019. [Online]. Available: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[5] S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami, "Artificial cognition for social human-robot interaction: An implementation," *Artificial Intelligence*, 2017.

[6] S. Lemaignan, R. Ros, L. Mösenlechner, R. Alami, and M. Beetz, "Oro, a knowledge management platform for cognitive architectures in robotics," in *2010 IEEE/RSJ International conference on intelligent robots and systems*. IEEE, 2010, pp. 3548–3553.

[7] Y. Mohamed and S. Lemaignan, "Ros for human-robot interaction," in *2021 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2021, pp. 3020–3027.

[8] D. Stampfer, "Contributions to system composition using a system design process driven by service definitions for service robotics," Ph.D. dissertation, Technische Universität München, 2018. [Online]. Available: https://mediatum.ub.tum.de/1399658

[9] S. Lemaignan, "Intents," PAL Robotics, techreport, 2024. [Online]. Available: https://docs.pal-robotics.com/sdk-dev/development/intents