Empowering Users in Human-Robot Interaction through End-User Development and Large Language Models

Luigi Gargioni[1]

[1]University of Brescia - Department of Information Engineering, Via Branze 38, Brescia, 25123, Italy

Abstract

This paper presents a PhD research project aimed at enhancing Human-Robot Interaction (HRI) by empowering non-technical users in both the definition and execution of robot tasks. The project addresses two main objectives: (i) enabling end users, such as domain experts without programming skills, to define, validate, and modify robot tasks through End-User Development (EUD) enhanced by Large Language Models (LLMs); and (ii) improving robot adaptability during collaborative task execution by integrating human cognitive factors (e.g., beliefs, desires, and intentions) into robot decision-making. The findings aim to contribute to the ongoing shift toward more adaptable and human-aware robotic systems.

Keywords

Human-Robot Interaction, Human-Computer Interaction, End-User Development, Large Language Models

1. Introduction

The field of robotics is currently in a rapid expansion, driven by significant technological advancements across multiple disciplines. In industrial automation, innovations in control systems, materials, and technical standards are transforming manufacturing processes. Robots in this domain are capable of executing highly precise and repeatable operations at high speed over long periods, making them ideal for tasks such as precision welding, heavy load handling, and complex assembly line operations. In parallel, the rise of collaborative robots (cobots) is reshaping the way humans and robots interact. Designed to operate safely alongside human workers without the need for physical barriers, cobots are enabling more flexible, adaptive, more sustainable and human-centric manufacturing environments [1].

Furthermore, recent advances in Artificial Intelligence (AI) have introduced a new wave of computational capabilities to mobile robotics [2], including machine learning, automated planning, and computer vision. These developments have enabled mobile robots to autonomously navigate both known and unknown environments. Applications include hazardous area exploration, such as radioactive zones, and object recognition in sensitive contexts like airports.

Another area undergoing significant transformation is social robotics [3]. This is largely pushed by the advent of Large Language Models (LLMs), which have acquired conversational abilities far beyond traditional systems. Social robots are now capable of engaging in human-like interactions aimed at acting as a conversational partner, supporting mental well-being, and facilitating cognitive exercises.

While the robotics field has shown impressive progress and earned substantial attention, challenges persist, particularly in the domains of collaborative and social robotics, where interaction with humans remains a central aspect. Human-Robot Interaction (HRI) occurs at multiple levels: from task definition to real-time task execution and adaptation. These interactions often require a robot to understand and respond to human actions, preferences, and emotional states, for which traditional model-based robotic systems are not well suited.

In collaborative robotics, one key challenge lies in enabling domain experts to define the tasks that robots must perform. Given the increasing versatility of collaborative robots, such systems must be adaptable to frequent changes in production workflows. This demand aligns with the principles of

---

CHItaly 2025: Technologies and Methodologies of Human-Computer Interaction in the Third Millenium, Doctoral Consortium, 6-10 October 2025, Salerno, Italy

luigi.gargioni@unibs.it (L. Gargioni)
0000-0003-4354-916X (L. Gargioni)

End-User Development (EUD) [4, 5], which promotes empowering non-expert users to modify or extend technological artefacts. In this context, end-user robot programming has emerged as a growing field of research, aiming to make robotic systems programmable by users without formal expertise in robotics. However, as highlighted in [6], programming robots involves additional complexity compared to traditional software, due to the need to reference physical objects and locations, ensure safe operation, and coordinate physical movement in shared environments with humans. Two main approaches have been identified in end-user robot programming: (i) programming by demonstration [7], which requires multiple iterations for skill generalization, and (ii) visual programming interfaces [8], designed to lower the cognitive and technical barriers for users. In parallel, the ubiquity of natural language in human communication has inspired efforts to enable natural language programming for robots. This paradigm has been applied to tasks such as navigation, assembly, and manipulation in industrial contexts [9], and is now being explored for social and service robotics [10].

Although HRI shares many objectives with Human-Computer Interaction (HCI), some of the main challenges in building a human-aware collaboration with robots come from the field of autonomous robotics. These challenges often involve how robots make decisions and interpret their surroundings, areas traditionally explored through research on autonomous behavior and artificial cognition [11]. Several studies have addressed the architectural and cognitive challenges involved in designing adaptive and interactive robotic systems [12]. These systems typically integrate decisional components responsible for high-level reasoning, such as planning and situation assessment, with functional components that manage real-time perception and action [13]. Despite this progress, current architectures often depend on predefined models of human mental states, like beliefs, desires, and intentions, limiting their ability to adapt to unpredictable behavior. This model-based limitation introduces uncertainty and hinders the robot's capacity to anticipate or respond to nuanced human actions. Building accurate, generalizable models of human cognition remains an open challenge.

2. Research Objectives and Contributions

Building on the themes discussed above, this PhD project is structured around two main research objectives. *Objective 1* was explored in collaboration with Prof. Daniela Fogli and Prof. Pietro Baroni at the University of Brescia. *Objective 2* was developed during a research visit period at CNRS-LAAS in Toulouse (France), under the supervision of Dr. Rachid Alami, and with continued support from Prof. Daniela Fogli.

2.1. Objective 1: End-User Development for Robot Task Definition

Starting from the study reported in [14], this objective aimed to explore and support the ability of end users to autonomously define, edit, and verify robotic tasks. This encompassed not only the specification of task logic but also the identification and modeling of relevant domain items of the real-world environment. This objective is based on the idea that giving end users more control can make robotic systems easier to use and more adaptable to different and dynamic situations. Specifically, the goal is to investigate design approaches and interaction paradigms that facilitate such empowerment, including natural language interfaces and multimodal interaction techniques. Key aspects of interest include:

1. Supporting the definition of both tasks and domain elements directly by end users, reducing dependency on technical developers.
2. Exploring the viability and limitations of natural language as an approach for task specification.
3. Evaluating the effectiveness of multimodal interaction in enhancing the end-user experience across the different stages of the task definition workflow.

2.1.1. Methodology

To guide this investigation, the following research questions were formulated:

- **RQ1**: To what extent can domain experts define domain-specific tasks and items through dedicated user interfaces designed for non-programmers?
- **RQ2**: What are the capabilities and limitations of natural language interfaces in supporting users to accurately specify, validate, and refine structured robotic tasks, and how can multimodal interaction enhance this process during task definition?

To address the research questions, several interaction design activities were carried out. Whenever feasible, the investigations focused on use cases in the healthcare domain, particularly on galenic formulations, which are personalized medications manually prepared by pharmacists to meet specific patient needs, such as allergies, tailored dosages, or specific administration forms.

The healthcare domain was selected as a reference context since this PhD project is part of the *Technology for Health* doctoral program at the University of Brescia. The interaction design activities revolved around two main projects:

- **Project 1 - Preparation of Personalized Medicines through Collaborative Robots**: In [15], a survey was conducted to investigate how medications are organized and dispensed in pharmacies and patients' homes, with a specific focus on the use of pill dispensers. This initial analysis offered a contextual analysis of pharmacy workflows and provided the opportunity to explore the feasibility of a future end-to-end service for galenic preparation and delivery. Subsequently, the next step of this activity (related paper under review) focuses on the application of a human-centred methodology adopted to design an EUD environment by involving representative end users (experts in the pharmaceutical sector) from system ideation to its evaluation. The study identified repetitive and low-value tasks that could be effectively delegated to collaborative robots, allowing pharmacists to focus on higher-priority activities. The primary conclusions of the work pertain to the design implications associated with AI-enabled EUD for collaborative robots. Based on these findings, in [16], low-fidelity prototypes of an interactive system were developed. Following the emerged implications and the designed mockups, a web-based prototype has been developed [17, 18]. This application is called PRAISE (Pharmaceutical Robotic and AI System for End users). The purpose of the application is to provide support to end users (i.e., pharmacists) in defining robot programs that are suitable for the specific case of galenic preparations. The application is conceived as an EUD environment, which implements a hybrid interaction approach based on a natural language interface leveraging LLMs and a domain-oriented graphical interface to check and revise the robot programs created. A user study conducted with nine pharmacists has demonstrated the validity of the approach with positive feedback.
- **Project 2 - Natural Language and Multimodal Interfaces for End-User Robot Programming**: The work reported in [19] explores the potential of OpenAI ChatGPT in enhancing natural language processing understanding. This work is based on the starting point of the PhD project, namely the CAPIRCI application reported in [14]. CAPIRCI is designed to support non-technical users in defining robot tasks. The work introduced the possibility for users to verify and modify natural language-generated programs using a block-based interface, with the aim of enhancing transparency and control. The goal of this work was to begin an exploration of the capabilities of the new technologies based on LLMs in identifying user intents and structuring the desired output. Extending the previous work, [20] presented a prototype environment that allows users with no technical background to define domain-specific elements (objects, actions, locations) and create pick-and-place tasks using a combination of natural language and graphical programming via Blockly. The goal was to assess the feasibility of multimodal interaction in defining and revising robot tasks. As a continuation of the previous work, a new prototype (related paper under review) integrated a digital twin into the application. Once a task is defined, it can be simulated virtually, allowing users to preview the robot's behavior and identify potential execution issues before deployment.

2.1.2. Main Findings

These projects advance the field of end-user robot programming by introducing a human-centered AI approach that integrates LLMs, multimodal interaction, and digital twins. The resulting workflow empowers domain experts to conceptualize, validate, and refine robot tasks through intuitive, accessible tools, without requiring programming or robotics expertise.

Central to the methodology is the active involvement of end users throughout the design process, ensuring that the environment aligns with their practices, expectations, and cognitive models.

The activities carried out aim to demonstrate that, when properly supported, non-technical users can effectively engage in robot task definition, ultimately enabling more adaptive, usable, and trustworthy robotic systems in real-world environments.

2.2. Objective 2: Human-Robot Collaboration for Task Accomplishment

The second objective of the PhD project was to investigate how a robot can effectively collaborate with a human partner by taking into account human beliefs, intentions, and emotional states during task execution. This goal addresses the limitations of rigid model-based approaches, which often fall short when facing the complexity and unpredictability of real-world human behavior.

Key challenges addressed in this objective include:

1. Overcoming the limitations of traditional model-based approaches to robotics when interacting with humans, whose behavior is often non-deterministic and emotionally driven.
2. Maintaining a robust and rigorous task planning framework, even in dynamic and uncertain environments.

2.2.1. Methodology

To guide this investigation, the following research questions were formulated:

- **RQ3**: How can robots adapt to the unpredictable behavior of human partners in collaborative scenarios, while maintaining task control and ensuring safety during execution?

To address RQ3, a hybrid system architecture was designed that integrates deterministic techniques (e.g., task planning and rule-based approach) with the flexibility of LLMs. The preliminary conceptualization was introduced in [21], and a complete description of the architecture, along with initial evaluations, is provided in a paper currently under review.

The approach is based on the concept of *shared goals* between the human and the robot. A task is defined as a cooperative activity aimed at achieving a mutually agreed-upon goal.

The assumed application context is a coaching scenario where the robot assists a person in completing a task composed of several activities. To illustrate this scenario, it can be considered a patient in a care facility who is required to take medications and engage in both cognitive and physical activities. Effective collaboration requires adapting the robot's actions not only to the requirements of the task but also to the human's evolving behavior, preferences, and affective state.

The proposed architecture is modular and designed to ensure seamless collaboration, adaptive behavior, and safe execution across the following components:

- *Natural Language Understanding and Knowledge Acquisition*: Interaction begins with natural language input from the user, which is transformed into semantic vector embeddings and stored in a dedicated Vector Database.
- *Human-Robot Task Synthesizer*: This module, based on an LLM, processes user input and contextual data to generate a structured task representation, capturing goals, constraints, and required actions.

- *Situation Assessment & Human-Robot Task Progress*: By integrating prior knowledge, task history, and current context, this module dynamically updates the task plan in response to environmental or human-driven changes. Also, this module is based on an LLM. It ensures continual alignment between system behavior and human expectations.
- *Event Planner & Next Steps*: Responsible for monitoring task progression and deciding subsequent actions, this component checks for discrepancies between expected and actual outcomes, triggering feedback loops when necessary.
- *Robot Perception & Robot Effector*: These modules ensure the system's responsiveness and safety. The perception component continuously interprets both the task state and the human's behavior, leveraging a Visual Language Model (VLM), while the effector module governs physical actions and verbal communication, maintaining alignment with user preferences and situational needs.

2.2.2. Main Findings

The developed architecture contributes to the advancement of HRI by blending deterministic techniques with the contextual flexibility offered by LLMs. The system is capable of adapting robot behavior in real time, maintaining alignment with human intentions and adjusting to unexpected changes in the environment or in user input.

Empirical case studies confirm the system's capacity to interpret human-provided goals in natural language, generate coherent and context-aware task plans, dynamically adapt execution strategies in response to ongoing interaction, and ensure safety and reliability despite behavioral uncertainty on the human side.

3. Discussion and Future Works

While the work presented in this paper introduces promising approaches to empowering end users in robot programming and enhancing human-robot collaboration, some limitations remain that offer implications for future development.

A first limitation relates to the realism of the deployed environments. While the digital twin component introduces a useful simulation layer, it currently simplifies many aspects of the real-world manipulation tasks. Moving forward, a more realistic digital twin, capable of simulating diverse objects and conditions (e.g., different objects, workspace constraints, and safety considerations), is needed to better reflect the complexity of actual work settings. Similarly, future user testing should incorporate real objects, such as test tubes and pharmaceutical tools, to more closely replicate the environment in which pharmacists operate and to validate the robustness of the task definition interface under practical constraints.

In the case of the hybrid architecture for collaborative task execution, current experiments rely on partially abstracted or scripted interaction scenarios. A natural next step is to bring this system into more defined and dynamic use cases (e.g., many activities with different priorities and schedule constraints, also considering human preferences), where human behavior is less predictable and the system's adaptability can be more rigorously challenged.

Finally, while recent advances in LLMs offer new opportunities for natural and flexible interaction, they must be approached with critical awareness. In the projects presented in this paper, LLMs are never used *as-is*; instead, they are systematically complemented by additional technologies and methodologies designed to interpret, verify, and ground user input. This hybrid strategy is essential to mitigate the known limitations of LLMs, such as inconsistencies or safety concerns, and to ensure that human-robot collaboration remains robust, reliable, and context-aware.

Overall, this work puts effort into enabling more human-aware and intelligent robotic systems. By demonstrating the feasibility of key concepts, this work provides a solid foundation for future advancements in the field. Building on these results, the next steps will involve refining the approach through real-world testing, enhancing simulation environments, and fostering continuous user engagement to ensure robust and practical deployment.

Declaration on Generative AI

During the preparation of this work, the author used ChatGPT and Grammarly in order to: grammar and spelling check, paraphrase, and reword. After using these services, the author reviewed and edited the content as needed, thus, he takes full responsibility for the publication's content.

Funding

The PhD scholarship is co-funded by the Italian Ministry, Piano Nazionale di Ripresa e Resilienza (PNRR) and Antares Vision.

References

[1] L. Gargioni, D. Fogli, P. Baroni, A meta-design approach to collaborative robotics to achieve sustainability goals, CEUR-WS 3978 (2025) 8.

[2] S. G. Tzafestas, Mobile robot control and navigation: A global overview, Journal of Intelligent & Robotic Systems 91 (2018) 35–58.

[3] H. Mahdi, S. A. Akgun, S. Saleh, K. Dautenhahn, A survey on the design and evolution of social robots—past, present and future, Robotics and Autonomous Systems 156 (2022) 104193.

[4] H. Lieberman, F. Paternò, V. Wulf, End User Development (Human-Computer Interaction Series), Springer-Verlag, Berlin, Heidelberg, 2006.

[5] B. R. Barricelli, F. Cassano, D. Fogli, A. Piccinno, End-user development, end-user programming and end-user software engineering: A systematic mapping study, Systems and Software (2019).

[6] G. Ajaykumar, M. Steele, C.-M. Huang, A survey on end-user robot programming, ACM Comput. Surv. 54 (2021). doi:10.1145/3466819.

[7] Y. Liu, Z. Li, H. Liu, Z. Kan, Skill transfer learning for autonomous robots and human–robot cooperation: A survey, Robotics and Autonomous Systems (2020).

[8] A. Schoen, N. White, C. Henrichs, A. Siebert-Evenstone, D. Shaffer, B. Mutlu, Coframe: A system for training novice cobot programmers, in: HRI, 2022. doi:10.1109/HRI53351.2022.9889345.

[9] D. K. Misra, J. Sung, K. Lee, A. Saxena, Tell me dave: Context-sensitive grounding of natural language to manipulation instructions, Int. J. Rob. Res. (2016). doi:10.1177/0278364915602060.

[10] S. Gallo, G. Vaiani, F. Paternò, End-user personalisation of humanoid robot behaviour through vocal interaction, in: Workshop on Robots for Humans, volume 3794, CEUR-WS.org, 2024.

[11] S. Rossi, F. Ferland, A. Tapus, User profiling and behavioral adaptation for hri: A survey, Pattern Recognition Letters 99 (2017) 3–12.

[12] R. Alami, A. Clodic, V. Montreuil, E. A. Sisbot, R. Chatila, Toward human-aware robot task planning., in: AAAI spring symposium, 2006, pp. 39–46.

[13] P. Foggia, A. Greco, A. Roberto, A. Saggese, M. Vento, A social robot architecture for personalized real-time human-robot interaction, IEEE Internet of Things Journal (2023).

[14] D. Fogli, L. Gargioni, G. Guida, F. Tampalini, A hybrid approach to user-oriented programming of collaborative robots, Robotics and Computer-Integrated Manufacturing 73 (2022) 102234.

[15] L. Gargioni, D. Fogli, P. Baroni, A systematic review on pill and medication dispensers from a human-centered perspective, Journal of healthcare informatics research 8 (2024) 244–285.

[16] L. Gargioni, D. Fogli, P. Baroni, Designing human-robot collaboration for the preparation of personalized medicines, GoodIT '23, ACM, 2023, p. 135–140. doi:10.1145/3582515.3609527.

[17] L. Gargioni, D. Fogli, P. Baroni, Exploring the adoption of collaborative robots for the preparation of galenic formulations, Information 16 (2025) 559.

[18] L. Gargioni, D. Fogli, P. Baroni, Preparation of personalized medicines through collaborative robots: A hybrid approach to the end-user development of robot programs, ACM Journal on Responsible Computing (2025).

[19] G. Bimbatti, D. Fogli, L. Gargioni, et al., Can chatgpt support end-user development of robot programs?, in: CEUR Workshop Proceedings, volume 3408, 2023, p. 8.

[20] L. Gargioni, D. Fogli, Integrating chatgpt with blockly for end-user development of robot tasks, HRI '24, ACM, New York, NY, USA, 2024, p. 478–482. doi:10.1145/3610978.3640653.

[21] L. Gargioni, R. Alami, D. Fogli, Towards a hybrid llm/model-based architecture for robot coaching: An instance of human-machine collaboration, in: CEUR Workshop Proceedings, CEUR-WS, 2025.