Capturing and Shaping Human-Robot Intent Expression

Ting Li
George Mason University
Fairfax, Virginia, USA
tli21@gmu.edu

David Porfirio
George Mason University
Fairfax, Virginia, USA
dporfiri@gmu.edu

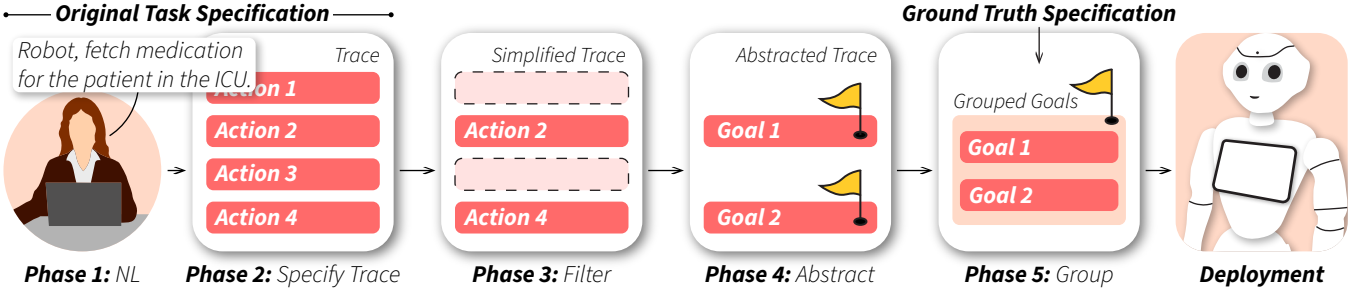

Figure 1: Our pipeline for capturing human intent (left) and actively shaping it to a form that the robot is better able to understand (right).

Abstract

Robot platforms must accurately understand user intent, including both hard constraints and soft preferences. This remains challenging because human communication is often ambiguous and imprecise, a property that carries over to interactions with robots. Consequently, much recent robotics research has focused on improving a robot's ability to passively infer intent from imperfect input. We argue, however, that understanding user intent should be a shared responsibility. Rather than relying solely on raw intent expression, robot platforms can and should be designed to actively shape user input when appropriate. In this paper, we present recent findings on how humans naturally communicate tasks to robots and demonstrate how users can be guided to refine their intent into forms that are more interpretable and useful for robotic systems.

CCS Concepts

• **Human-centered computing** → **Human-robot interaction**; *Interactive systems and tools*; User studies; • **Robotics** → *Task planning*.

Keywords

human-robot interaction, end-user programming, task planning

ACM Reference Format:
Ting Li and David Porfirio. 2026. Capturing and Shaping Human-Robot Intent Expression. In *Proceedings of ACM/IEEE International Conference on Human-Robot Interaction (HRI '26).* ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

HRI '26, Edinburgh, Scotland, UK

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

As robots become increasingly ubiquitous in everyday life, end users rely on intuitive communication paradigms to direct robot behavior. Most commonly, these paradigms fall into two categories: natural language and end-user programming (EUP). Natural language interaction typically involves users issuing real-time commands of varying complexity to their household robot companions, such as "While I'm gone remember to deliver medication upstairs, and make sure that dinner is served." In contrast, EUP enables users to specify the same task by providing a hand-crafted sequence of symbolic instructions, typically via a graphical user interface, such as `moveTo`(*medicine cabinet*), `grab`(*medication*), `moveTo`(*upstairs*), `handoff`(*medication, parent*), and so on.

Although both paradigms are widely regarded as intuitive, they often obscure the user's true intent. Natural language, in particular, is inherently imprecise and ambiguous [6]. Robots and human-computer interaction systems powered by large language models (LLMs) and vision language models (VLMs) have been proven to be exceptionally capable at inferring user intent [1, 3, 4, 7, 10, 11], but the burden of inferring domain-specific details (*e.g.,* the medication delivery is much more urgent than the dinner delivery) is on the robot rather than on the appropriate source—the *actual human domain expert.* This is especially important in safety-critical settings, where fast, intuitive communication with the robot is equally as important as the user's intent being interpreted correctly. EUP, by contrast, suffers from the opposite problem. By requiring users to hand-craft symbolic sequences of potentially branching and looping actions—that is, *programs*—the resulting specifications are highly context-specific, overly precise, and brittle to changes in context or execution conditions [5].

In this paper, we examine recent findings on how humans naturally communicate tasks to robots and how designers can leverage

these tendencies when building robot platforms. Specifically, we ask how robot platforms and interfaces can guide users to refine their task specifications in ways that make their underlying intent clearer to the robot. We present a preliminary pipeline (shown in Figure 1) that does exactly this, guiding users to clarify their true intent after an initial, natural task expression. Finally, we discuss several principles of human task communication that can inform the design of future human–robot communication interfaces.

2 Principles of Human Task Specification

We conducted an IRB-approved crowdsourcing study that elicited (a) natural-language task specifications from users and (b) corresponding *traces*, defined as step-by-step sequences of instructions derived from those specifications. The goal of this study was to examine how humans naturally communicate tasks to robots using two of the most common modalities in human–robot interaction. Based on the collected data, we describe several principles of human–robot task communication. Note that these results should be taken as preliminary findings for now. Future work is necessary for further validation.

Human natural language communication is often (though not always) sequential. Participants responded to two task prompts: an open-ended prompt and a structured prompt in which they were given an explicit but unordered set of goals to communicate to an imagined robot. In the structured condition, participants overwhelmingly employed lexical markers of sequentiality—such as *then*, *next*, and *after*—to organize their instructions. In contrast, these same markers appeared far less frequently in the open-ended condition. **Implication:** In realistic, open-ended human–robot interaction, robots cannot reliably depend on explicit sequential language to infer users' intended ordering constraints.

Human natural language communication is predominantly linear. In the same natural language instructions—both structured and open-ended—participants *rarely* employed conditionals marked by keywords such as *if*, *when*, etc. **Implication:** users naturally express tasks linearly. If tasks are branching, robots may need to infer such control structure.

Humans include implied or redundant steps when conveying instructions to a robot. When asked to assemble step-by-step task instructions, participants consistently included more procedural detail than was strictly necessary for task execution. In particular, users frequently specified actions that a robot could reasonably infer on its own (e.g., the action **`moveTo`**(*cereal*) is implied by **`grab`**(*cereal*)). This tendency aligns with and extends findings from prior work [8, 9]. **Implication:** In end-user programming for robots, not all user-provided steps should be treated as essential constraints on execution.

2.1 Proposed Solution for Eliciting True Human Intent

Aligned with these principles, we propose a solution that guides users through a sequence of interaction phases designed to uncover the essential meaning behind robot task requests. The approach begins by eliciting an initial task specification in natural language and translating it into a step-by-step procedural trace of actions (Phases 1 and 2 in Figure 1). Users are then prompted to identify steps that are unnecessary for the robot to know, or that are redundant given the robot's autonomous planning capabilities, thereby reducing the risk of over-constraining execution (Phase 3). Next, users abstract remaining actions to their higher-level goals, disentangling what must be achieved from how it is accomplished (Phase 4). Finally, the approach relaxes ordering constraints by allowing users to express preferences over execution order and parallelization rather than enforcing strict sequences (Phase 5).

We implemented and tested an early version of this pipeline within the same exploratory crowdsourcing study that we discussed above. Our results show that our pipeline effectively aligns user input with intended robot behavior by progressively reducing unnecessary procedural detail while preserving task correctness. User-created traces initially contained substantial redundancy, and our pipeline successfully removed approximately half of the specified actions, with minimal reintroduction during user validation. Distilled task representations led to shorter and more robust execution plans, particularly in novel environments where user-created traces were brittle to contextual changes.

2.2 Future Work

Our immediate future work involves implementing the pipeline on a robot and investigating whether our results translate to the spoken domain. We envision such a system being a mixed-initiative user interface [2] with control being shared between both agents. For example, given a user's task specification in natural language, the robot can automatically convert the specification to a trace, "distill" the trace to its critical steps, abstract critical steps to higher-level goals, and relax the total ordering of these steps, all the while making sure to confirm each decision with the user.

Acknowledgments

This work was supported by George Mason University.

References

[1] Anthony Brohan, Yevgen Chebotar, Chelsea Finn, Karol Hausman, Alexander Herzog, Daniel Ho, Julian Ibarz, Alex Irpan, Eric Jang, Ryan Julian, et al. 2023. Do as i can, not as i say: Grounding language in robotic affordances. In *Conference on robot learning*. PMLR, 287–318.

[2] Eric Horvitz. 1999. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 159–166.

[3] Subbarao Kambhampati, Karthik Valmeekam, Lin Guan, Mudit Verma, Kaya Stechly, Siddhant Bhambri, Lucas Paul Saldyt, and Anil B Murthy. 2024. Position: LLMs Can't Plan, But Can Help Planning in LLM-Modulo Frameworks. In *Proceedings of the 41st International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 235)*, Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (Eds.). PMLR, 22895–22907. https://proceedings.mlr.press/v235/kambhampati24a.html

[4] Christine P Lee, David Porfirio, Xinyu Jessica Wang, Kevin Chenkai Zhao, and Bilge Mutlu. 2025. Veriplan: Integrating formal verification and llms into end-user planning. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–19.

[5] Yuan-Hong Liao, Xavier Puig, Marko Boben, Antonio Torralba, and Sanja Fidler. 2019. Synthesizing environment-aware activities via activity sketches. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6291–6299.

[6] Steven T Piantadosi, Harry Tily, and Edward Gibson. 2012. The communicative function of ambiguity in language. *Cognition* 122, 3 (2012), 280–291.

[7] David Porfirio, Vincent Hsiao, Morgan Fine-Morris, Leslie Smith, and Laura M. Hiatt. 2025. Bootstrapping Human-Like Planning via LLMs. In *2025 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. 665–670. doi:10.1109/RO-MAN63969.2025.11217637

[8] David Porfirio, Mark Roberts, and Laura M Hiatt. 2024. Goal-oriented end-user programming of robots. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 582–591.

[9] David Porfirio, Mark Roberts, and Laura M Hiatt. 2025. An Interaction Specification Language for Robot Application Development. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 1062–1066.

[10] Benedict Quartey, Eric Rosen, Stefanie Tellex, and George Konidaris. 2025. Verifiably following complex robot instructions with foundation models. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 1–8.

[11] Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, Ayzaan Wahid, Quan Vuong, Vincent Vanhoucke, Huong Tran, Radu Soricut, Anikait Singh, Jaspiar Singh, Pierre Sermanet, Pannag R. Sanketi, Grecia Salazar, Michael S. Ryoo, Krista Reymann, Kanishka Rao, Karl Pertsch, Igor Mordatch, Henryk Michalewski, Yao Lu, Sergey Levine, Lisa Lee, Tsang-Wei Edward Lee, Isabel Leal, Yuheng Kuang, Dmitry Kalashnikov, Ryan Julian, Nikhil J. Joshi, Alex Irpan, Brian Ichter, Jasmine Hsu, Alexander Herzog, Karol Hausman, Keerthana Gopalakrishnan, Chuyuan Fu, Pete Florence, Chelsea Finn, Kumar Avinava Dubey, Danny Driess, Tianli Ding, Krzysztof Marcin Choromanski, Xi Chen, Yevgen Chebotar, Justice Carbajal, Noah Brown, Anthony Brohan, Montserrat Gonzalez Arenas, and Kehang Han. 2023. RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control. In *Proceedings of The 7th Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 229)*, Jie Tan, Marc Toussaint, and Kourosh Darvish (Eds.). PMLR, 2165–2183. https://proceedings.mlr.press/v229/zitkovich23a.html