Enhancing Collaboration in Human-Robot Teams: Grounding via Implicit Communication

Yan Zhang[1], Jarrod Knibbe[2], Jorge Goncalves[1] and Wafa Johal[1]

[1]*School of Computing and Information Systems, University of Melbourne, Australia*
[2]*School of Electrical Engineering and Computer Science, The University of Queensland, Australia*

Abstract

Implicit communication plays a crucial role in human collaboration, where contextual cues, such as intentions and situational awareness, foster shared understanding and seamless coordination. However, translating these capabilities to robots in collaborative tasks remains challenging. This work investigates the role of implicit communication in human-robot teams through three integrated studies. First, we analyse how linguistic implicatures influence performance and user experience in human-robot physical collaboration. Second, we explore how multi-modal implicit cues enhance team effectiveness and alignment when used proactively and reactively by robots. Building on these findings, the final phase focuses on designing and evaluating a multi-LLM robotics system that learns and adapts human teammates' communication strategies. By advancing robots' capacity to interpret and generate contextually appropriate implicatures, this research bridges gaps in human-robot collaboration, enabling the natural establishment, maintenance, and dynamic adaptation of common grounds.

Keywords

Human-Robot Collaboration, Implicit Communication, Multi-agentic, LLM

1. Background

Collaborative robots (cobots) have demonstrated the potential to assist in professional and personal contexts, though their integration as co-workers in daily human activities is still evolving and expanding. A crucial aspect of human-robot collaboration (HRC) is the seamless coordination between humans and robots, where effective and natural communication plays a pivotal role in ensuring successful collaboration.

In the broader nature of human communication, much of communication operates implicitly and relies heavily on subtle cues to convey intentions and achieve goals [1]. Drawing from linguistic pragmatics, such cues are termed implicatures, context-dependent meanings inferred by receivers [2]. Not limited to verbal expressions, for example, applause may overtly signal approval but also implicitly suggest the conclusion of an event [3]. In human-robot interaction (HRI), classifying implicit communication is often oversimplified (e.g., equating gaze with "implicit" and speech with "explicit" [4, 5]). This work adopts a nuanced perspective: implicit communication is a way to convey latent information that receivers can interpret under the same mutual understanding.

Implicit communication is essential in collaborative tasks in which teammates iteratively build shared understanding over time [6, 7, 8]. Drawing from Clark's common ground theory, human collaboration relies on the establishment, maintenance, and dynamic adaptation of common ground, which is the mutual knowledge, beliefs, and assumptions that underpin shared intentionality [9, 10]. Common ground is not static; it evolves through continuous, bidirectional efforts to align understanding during interactions. This research explores how implicit communication facilitates the three steps of the grounding process, thus enhancing human-robot collaboration.

A notable example of implicit communication is the use of indirect speech act (ISA), a pragmatic feature that requires the listener to infer the speaker's intended meaning from related context [11]. For

CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan
yan.zhang.1@unimelb.edu.au (Y. Zhang); j.knibbe@uq.edu.au (J. Knibbe); jorge.goncalves@unimelb.edu.au (J. Goncalves); wafa.johal@unimelb.edu.au (W. Johal)
0000-0003-2142-5094 (Y. Zhang); 0000-0002-8844-8576 (J. Knibbe); 0000-0002-0117-0322 (J. Goncalves); 0000-0001-9118-0454 (W. Johal)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2025-11-20
CEUR-WS.org/Vol-4101/paper16.pdf

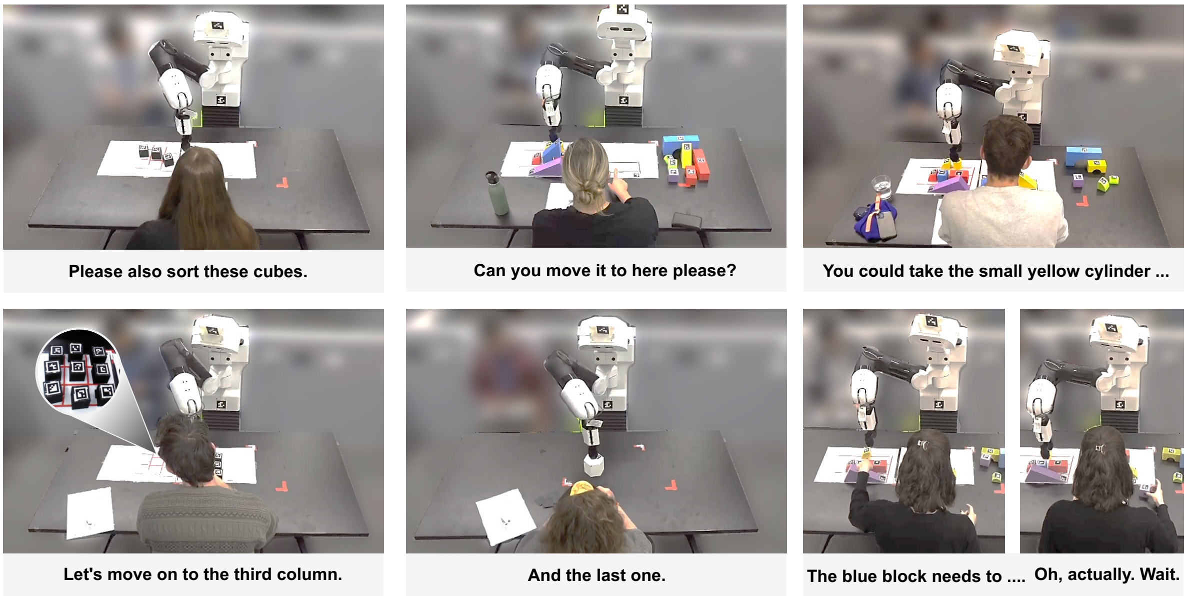

Figure 1: This figure presents images from our experiment, featuring representative participant utterances to illustrate the types of requests used. The top left image depicts a direct speech act, while the rest of the images showcase various indirect speech acts.

instance, the statement “I’m cold” can imply a request to close the window, depending on the situation. ISAs exemplify how humans efficiently establish common ground, as they rely on mutual knowledge and situational context while enabling listeners to actively infer unstated goals. Studies suggest humans tend to use implicit ways to convey intentions when interacting with robots at frequencies similar to those used with other humans [12, 13]. However, in real-world collaboration scenarios, the appropriateness of implicit intentions is less obvious due to the lack of clarity. While ISAs hold promise for fostering natural interaction, there remains a gap in empirical evidence regarding their effectiveness in establishing common ground and improving team performance and user experience, particularly in dynamic, physically human-robot collaborative settings.

Prior work in HRI has focused on implicit human-to-robot communication [14, 15, 16]. However, collaboration thrives on bidirectional flow, where both partners contribute to sustaining shared understanding [17, 18]. There is still a need for further exploration of how robots can convey implicit information to humans and how this affects HRC [19, 20]. To address this gap, our next work investigates how robots can leverage multi-modal implicit cues to serve dual purposes, including proactively conveying task-relevant context and reactively backchannelling to acknowledge human intent. We examine how these cues influence both objective team outcomes and subjective human perceptions, and understand the establishment and maintenance of common ground.

Building on the understanding of implicit communication, we aim to enhance robotic systems by integrating multimodal implicit cue interpretation and generation, enabling robots to dynamically adapt to evolving common ground. By leveraging multimodal large language models (LLM) that infer different modalities such as speech, gaze, and motion, robots can potentially improve their ability to understand and respond to implicit intentions in a more nuanced manner. Additionally, by adapting to human behaviours and learning to generate appropriate implicit cues themselves, robots may become more adept at facilitating natural, bidirectional communication. However, relying on a single LLM presents several challenges. Due to knowledge gaps or misinformation, LLM outputs may suffer from hallucination or bias [21, 22]. A cooperative approach that combines decisions from multiple LLMs could enhance the robustness and reliability of the system [23]. Furthermore, when a single LLM is tasked with complex scenarios, extensive fine-tuning is required for different task components (e.g., perception, planning, execution), along with integration of external tools (e.g., API calls) [24, 25]. By adopting a multi-LLM approach, where each LLM focuses on a specific component, robots could extend

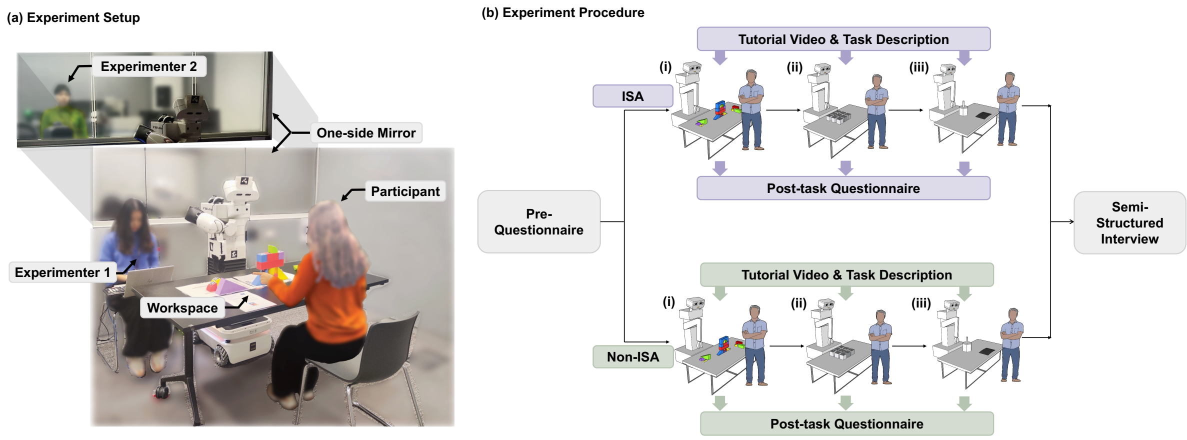

Figure 2: (a) This figure illustrates the experiment setup. The participant, robot, and experimenter 1 were all present in the same room. The participant and robot were seated on opposite sides of a table, with the shared workspace located in the centre. Experimenter 1 sat next to the robot, near the emergency button, and operated the robot's speech-WoZ interface. Experimenter 2 was positioned behind a one-side mirror, allowing for a clear view of the room, and was responsible for teleoperating the robot's arm movements. (b) This figure shows the experiment procedure. Each participant first completed a pre-questionnaire before being assigned to either the ISA or non-ISA group. The participant then performed three tasks with the robot in a counter-balanced order. Before each task, participants watched a tutorial video and read a task description. After completing each task, they filled out a post-task questionnaire. The experiment concluded with a semi-structured interview.

their implicit communication capabilities beyond basic tasks like object grasping to more complex, dynamic, and collaborative manipulations [26] and also facilitate easier model updates for long-term use [27].

In summary, the contribution of this research lies in enhancing human-robot collaboration through implicit communication, grounded in Clark's theory of common ground, by implementing a three-phase approach. We present preliminary findings from the first phase, demonstrating how robots' interpretation of ISAs improves task efficiency and user experience in collaborative manipulation tasks. A full analysis and discussion of these results are detailed in our full paper [28]. Future work will extend to bidirectional implicit communication (phase 2) and adaptive multi-LLM systems (phase 3), aiming to refine robots' ability to sustain and evolve common ground in dynamic settings. By bridging the gap between rigid, explicit robot communication and the fluidity of human teamwork, this research paves the way for cobots that integrate seamlessly into real-world collaborative environments.

2. Methodology

This research is based on lab empirical studies that involved a robot and a participant. We use TIAGo (Figure 2(a)), a mobile manipulator robot with anthropomorphic features, including a 2 Degrees of Freedom head, neck, torso, and 7 Degrees of Freedom arm. TIAGo is equipped with an RGB-D camera, stereo microphone, and speaker that is suitable for HRI research and tabletop manipulation tasks [29]. We built a motion teleoperation interface and speech control software using the Robotics Operating System (ROS) and the TIAGo API. The arm actions are implemented by inverse kinematics, which calculates the joint configuration based on the desired Cartesian coordinates of the end effector [30]. These customised tools can facilitate experiments to control the robot in pilots and user studies.

For the first phase, we conducted a lab study (Figure 1), engaging a participant and a robot in three collaborative physical tasks, to investigate the impacts of linguistic implicatures on HRC. Specifically, we aimed to answer these research questions: **(RQ1)** How does a robot's capability to understand indirect speech acts influence the perceived *team's performance*? **(RQ1.1)** How does a robot's capability

Table 1
The key results from CLMM analysis. Italics are covariates.

	Fixed Effects	Estimates	Std Error	95% CI	z	p-value
Team Fluency	Speech Mode (Non-ISA)	0.961	0.403	0.17 – 1.751	2.382	0.017*
	Robot (No)	0.004	0.110	-0.211 – 0.218	0.032	0.974
	Voice Assistant (Never)	0.129	0.210	-0.283 – 0.541	0.615	0.538
	Physical Collaborative Tasks (Never)	0.193	0.174	-0.147 – 0.533	1.112	0.266
Goal Alignment	Speech Mode (Non-ISA)	2.309	0.656	1.023 – 3.596	3.518	<0.001***
	Robot (No)	0.120	0.170	-0.214 – 0.453	0.701	0.483
	Voice Assistant (Never)	-0.099	0.316	-0.719 – 0.521	-0.312	0.755
	Physical Collaborative Tasks (Never)	0.536	0.270	0.007 – 1.064	1.985	0.047*
Performance Trust	Speech Mode (Non-ISA)	1.105	0.493	0.138 – 2.072	2.240	0.025*
	Robot (No)	-0.041	0.136	-0.307 – 0.226	-0.298	0.766
	Voice Assistant (Never)	0.231	0.248	-0.255 – 0.717	0.932	0.351
	Physical Collaborative Tasks (Never)	0.400	0.211	-0.014 – 0.814	1.892	0.058
Anthropomorphism	Speech Mode (Non-ISA)	2.708	0.674	1.387 – 4.03	4.016	<0.001***
	Robot (No)	-0.168	0.184	-0.528 – 0.192	-0.915	0.360
	Voice Assistant (Never)	0.031	0.340	-0.635 – 0.697	0.092	0.927
	Physical Collaborative Tasks (Never)	-0.111	0.288	-0.676 – 0.454	-0.385	0.701

to understand indirect speech acts influence the *fluency* of human-robot teamwork? **(RQ1.2)** How does a robot's capability to understand indirect speech acts influence the establishment of *goal alignment* among the human-robot team? **(RQ2)** How does a robot's capability to understand indirect speech acts influence a human teammate's *trust* in the robot's performance? **(RQ3)** How does a robot's capability to understand indirect speech acts influence a human teammate's perception of the robot's *anthropomorphism*?

The experiment involved 36 participants, who were divided into two groups: one interacting with a robot capable of interpreting ISAs and another interacting with a robot that could only respond to direct commands. Each participant engaged in three tasks: assembly, sorting, and polishing. This was a Wizard-of-Oz (WoZ) study and employed a mixed-methods approach, using both quantitative data (team fluency [31], goal alignment [31], performance trust [32], and robot anthropomorphism [33] questionnaires) and qualitative data (semi-structured interviews). Cumulative Link Mixed Models (CLMMs) via the "ordinal" package in R [34] have been used for quantitative data analysis. This analysis is appropriate given the ordinal nature of dependent variables, while also allowing us to account for effects for each participant, the type of tasks, and each sub-item of the used scales. The interview results were transcribed and analysed using thematic analysis [35] by two researchers. After individually coding the transcripts and identifying initial themes, they discussed any discrepancies and refined the themes through consensus. The experiment setup and the procedure are shown in Figure 2.

3. Preliminary Results

Our findings demonstrate that the robot's capacity to interpret linguistic implicatures plays a pivotal role in establishing common ground during HRC, although its effectiveness can vary depending on the context. The quantitative results show the robot's ability to comprehend implicatures significantly enhances participants' perceived team fluency, goal alignment, trust, and anthropomorphism Table 1. Moreover, this alignment of understanding fosters a sense of joint agency, where humans perceive the robot not as a passive tool but as an active partner engaged in co-constructing shared context.

Other qualitative results show that participants felt more confident in having established common ground with the robot when they successfully used ISAs to communicate. These indirect requests often emerged subconsciously, reflecting a natural inclination to leverage shared context, such as prior commands, environmental affordance (e.g., object locations), and multimodal cues (e.g., gestures). However, the effectiveness of ISAs can be task- and context-dependent. While indirect requests enhanced trust in ideal scenarios, participants expressed concerns about error-prone real-world settings, where misinterpretations could lead to mistrust.

These insights highlight the inherent limitations of relying solely on explicit command-based interactions, which lack the subtlety required for establishing shared understanding and the sense of teaming. The results also emphasise the importance of using implicit communication in a contextually adaptive and appropriate manner. Therefore, the careful integration of explicit information and implicatures in verbal communication emerges as a critical factor in optimising the performance and overall experience of human-robot collaboration.

4. Future Work

Based on the findings from phase one, where humans and robots established common ground by linguistic implicatures, phase two investigates how robots can maintain shared understanding through bidirectional implicit communication in physical collaborative tasks. We evaluate three interaction paradigms: 1) when the robot proactively uses implicatures to convey contextual information; 2) when the robot uses implicit cues as backchanneling to respond to human-implied intentions. Specifically, we compare four methods of communication: explicit speech, implicit speech, explicit motion, and implicit motion, across two scenarios: human-lead (robot backchannels) and robot-lead (robot initiates). Metrics include team fluency, efficiency, goal alignment, and trust. Overall, this phase aims to identify how implicit cues maintain common ground in physical collaborative tasks and how mismatched communication disrupt collaboration.

Phase three integrates insights from phases one and two into designing and evaluating a robotic system that employs cooperative multimodal multi-LLMs for human-robot collaborative manipulation tasks. In this system, each LLM will be initialised from a pre-trained base model and progressively fine-tuned for distinct roles, such as interaction perceiver, interaction generator, environment perceiver, task planner, and executor. The LLMs will cooperate to generate responses for each role, mitigating issues such as hallucinations. A user experiment will compare the performance of the multi-LLM system with that of a single LLM system on a simple pick-and-place task (low ambiguity) and a complex manipulation task (dynamic context shifts). The system evaluation will include both quantitative metrics, such as response time and task success rate, and qualitative analysis of failure cases. Additionally, we will assess the interaction through quantitative measures, including team fluency, efficiency, goal alignment, and trust, alongside qualitative insights into human perception of how effectively the system adapts to common ground. This phase bridges the iterative establishment and maintenance of shared understanding with real-time adaptation, advancing robots from rigid collaborators to partners capable of fluid, human-like coordination.

5. Conclusion

This research enhances human-robot collaboration by integrating implicit communication through three phases, guided by Clark's common ground theory. By advancing robots' ability to establish, maintain, and dynamically adapt shared understanding through multimodal implicit cues, we aim to enable more natural and context-aware teamwork. Our current findings demonstrate that implicit communication fosters the establishment of common ground, improving both team performance and user experience. Future work will explore how robots can maintain alignment using multimodal cues and dynamically adapt their strategies through multi-LLM architectures, ensuring robust collaboration in complex real-world environments. This research aims to transition HRC from rigid, direct communication to natural, nuanced, yet effective communication, ensuring adaptability and inclusivity for diverse populations in dynamic real-world environments.

6. Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT and Grammarly to check grammar and spelling. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] M. C. Frank, N. D. Goodman, Predicting pragmatic reasoning in language games, Science 336 (2012) 998–998.

[2] P. Grice, Studies in the Way of Words, Harvard University Press, 1991.

[3] R. Gilbert, Joyful noise: Reflections on applause, Southwest Review 86 (2001) 13–33.

[4] R. A. Bolt, "put-that-there" voice and gesture at the graphics interface, in: Proceedings of the 7th annual conference on Computer graphics and interactive techniques, 1980, pp. 262–270.

[5] S. Li, X. Zhang, Implicit intention communication in human–robot interaction through visual behavior studies, IEEE Transactions on Human-Machine Systems 47 (2017) 437–448.

[6] H. H. Clark, Discourse in production., Handbook of psycholinguistics (1994) 985–1021.

[7] D. R. Traum, Computational models of grounding in collaborative systems, in: Psychological Models of Communication in Collaborative Systems-Papers from the AAAI Fall Symposium, 1999, pp. 124–131.

[8] H. H. Clark, D. Wilkes-Gibbs, Referring as a collaborative process, Cognition 22 (1986) 1–39.

[9] H. H. Clark, S. E. Brennan, Grounding in communication., Perspectives on socially shared cognition (1991) 127–149.

[10] H. H. Clark, Using language, Cambridge university press, 1996.

[11] J. R. Searle, Indirect speech acts, in: Speech acts, Brill, 1975, pp. 59–82.

[12] M. K. Lee, S. Kiesler, J. Forlizzi, Receptionist or information kiosk: how do people talk with a robot?, in: Proceedings of the 2010 ACM conference on Computer supported cooperative work, 2010, pp. 31–40.

[13] M. Bennett, T. Williams, D. Thames, M. Scheutz, Differences in interaction patterns and perception for teleoperated and autonomous humanoid robots, in: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2017, pp. 6589–6594.

[14] K. Mahadevan, M. Sousa, A. Tang, T. Grossman, "grip-that-there": An investigation of explicit and implicit task allocation techniques for human-robot collaboration, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021, pp. 1–14.

[15] R. A. Knepper, C. I. Mavrogiannis, J. Proft, C. Liang, Implicit communication in a joint action, in: Proceedings of the 2017 acm/ieee international conference on human-robot interaction, 2017, pp. 283–292.

[16] Y. Che, A. M. Okamura, D. Sadigh, Efficient and trustworthy social navigation via explicit and implicit robot–human communication, IEEE Transactions on Robotics 36 (2020) 692–707.

[17] V. V. Unhelkar, S. Li, J. A. Shah, Decision-making for bidirectional communication in sequential human-robot collaborative tasks, in: Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, 2020, pp. 329–341.

[18] L. Tian, K. He, S. Xu, A. Cosgun, D. Kulic, Crafting with a robot assistant: use social cues to inform adaptive handovers in human-robot collaboration, in: Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction, 2023, pp. 252–260.

[19] S. M. Fiore, T. J. Wiltshire, E. J. Lobato, F. G. Jentsch, W. H. Huang, B. Axelrod, Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior, Frontiers in psychology 4 (2013) 859.

[20] G. Lemasurier, G. Bejerano, V. Albanese, J. Parrillo, H. A. Yanco, N. Amerson, R. Hetrick, E. Phillips, Methods for expressing robot intent for human–robot collaboration in shared workspaces, ACM Transactions on Human-Robot Interaction (THRI) 10 (2021) 1–27.

[21] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. J. Bang, A. Madotto, P. Fung, Survey of hallucination in natural language generation, ACM Computing Surveys 55 (2023) 1–38.

[22] S. Feng, C. Y. Park, Y. Liu, Y. Tsvetkov, From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair nlp models, arXiv preprint

arXiv:2305.08283 (2023).
[23] S. Feng, W. Shi, Y. Wang, W. Ding, V. Balachandran, Y. Tsvetkov, Don't hallucinate, abstain: Identifying llm knowledge gaps via multi-llm collaboration, arXiv preprint arXiv:2402.00367 (2024).
[24] Y. Qin, S. Liang, Y. Ye, K. Zhu, L. Yan, Y. Lu, Y. Lin, X. Cong, X. Tang, B. Qian, et al., Toolllm: Facilitating large language models to master 16000+ real-world apis, arXiv preprint arXiv:2307.16789 (2023).
[25] A. Zeng, M. Liu, R. Lu, B. Wang, X. Liu, Y. Dong, J. Tang, Agenttuning: Enabling generalized agent abilities for llms, arXiv preprint arXiv:2310.12823 (2023).
[26] G.-C. Kang, J. Kim, J. Kim, B.-T. Zhang, Prograsp: Pragmatic human-robot communication for object grasping, in: 2024 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2024, pp. 3304–3310.
[27] W. Shen, C. Li, H. Chen, M. Yan, X. Quan, H. Chen, J. Zhang, F. Huang, Small llms are weak tool learners: A multi-llm agent, arXiv preprint arXiv:2401.07324 (2024).
[28] Y. Zhang, T. S. Ratnayake, C. Sew, J. Knibbe, J. Goncalves, W. Johal, Can you pass that tool?: Implications of indirect speech in physical, arXiv preprint arXiv:2502.11720 (2025).
[29] J. Pages, L. Marchionni, F. Ferro, Tiago: the modular robot that adapts to different research needs, in: International workshop on robot modularity, IROS, volume 290, 2016.
[30] S. Chitta, E. Marder-Eppstein, W. Meeussen, V. Pradeep, A. R. Tsouroukdissian, J. Bohren, D. Coleman, B. Magyar, G. Raiola, M. Lüdtke, et al., ros_control: A generic and simple control framework for ros, Journal of Open Source Software 2 (2017) 456–456.
[31] G. Hoffman, C. Breazeal, Effects of anticipatory perceptual simulation on practiced human-robot tasks, Autonomous Robots 28 (2010) 403–423.
[32] B. F. Malle, D. Ullman, A multidimensional conception and measure of human-robot trust, in: Trust in human-robot interaction, Elsevier, 2021, pp. 3–25.
[33] C. Bartneck, Godspeed questionnaire series: Translations and usage, in: International Handbook of Behavioral Health Assessment, Springer, 2023, pp. 1–35.
[34] R. H. B. Christensen, ordinal—regression models for ordinal data, R package version 10 (2019) 54.
[35] V. Braun, V. Clarke, Thematic analysis., American Psychological Association, 2012.