Human Perception in Social Tasks: A Comparative Evaluation of Autonomous and Teleoperated Robots

Lavinia Hriscu, Alberto Sanfeliu, and Anaís Garrell

Abstract—**Robots and artificial intelligence technologies are becoming increasingly integrated into our daily lives. The introduction of humanoid robots into everyday settings is a gradual but ongoing process—one that society is already beginning to navigate. Yet this shift raises important questions: Who or what is truly behind these physical agents? And can we, as users, perceive differences in our interactions depending on whether a robot acts autonomously or it's teleoperated by a human? In this study, we present the results of an experiment in which participants interacted with a robot under two control conditions—autonomous and teleoperated—while it performed two distinct tasks in both static and dynamic movement scenarios. In our results, human operators outperformed autonomous systems in tasks requiring spatial awareness and contextual reasoning. Conversely, the autonomous robot—powered by a Large Language Model and operating without visual input—was perceived more favorably in tasks that demanded rapid access to broad and diverse information.**

Index Terms—**Social HRI, natural dialog for HRI, telerobotics and teleoperation.**

I. Introduction

The emergence of generative models, particularly Large Language Models (LLMs) [1], has significantly advanced the ability of robots to adapt to new tasks and environments [2]. This development is especially relevant for social robots, which are designed to operate in real-world scenarios and interact naturally with humans. These robots may function autonomously or can be teleoperated by a human. This duality raises a key question: how does the type of controller influence the way users perceive and experience the interaction?

Received 3 July 2025; accepted 16 October 2025. Date of publication 12 November 2025; date of current version 20 November 2025. This article was recommended for publication by Associate Editor K. Rogers and Editor S. Chernova upon evaluation of the reviewers' comments. This work was supported in part by JST Moonshot R, D under Grant JPMJMS2011-85, in part by the Spanish Ministry of Science, Innovation and Universities under the project LENA under Grant PID2022-142039NA-I00, funded by MCIU/AEI/10.13 039/50110001103, in part by the European project TORNADO under Grant HORIZON-CL4-2024-DIGITAL-EMERGING-01-101189557, and in part by Predoctoral Program AGAUR-FI ajuts under Grant 2025FI-101034 Joan Oró, through the Secretariat of Universities and Research of Department of Research and Universities of the Generalitat of Catalonia and European Social Plus Fund. *(Corresponding author: Lavinia Hriscu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Universitat Politècnica de Catalunya (UPC) under Application No. 2024.020.

The authors are with the Institut de Robòtica i Informàtica Industrial (CSIC-UPC), 08028 Barcelona, Spain, and also with the Universitat Politècnica de Catalunya (UPC), 08034 Barcelona, Spain (e-mail: lavinia.beatrice.hriscu@upc.edu; alberto.sanfeliu@upc.edu; anais.garrell@upc.edu).

Digital Object Identifier 10.1109/LRA.2025.3632091

While previous research has primarily investigated user perceptions of robots through observational studies [3], fewer studies have evaluated real-time interactions. Observational evidence indicates that autonomous robots can decrease user acceptance and increase perceived social distance [4]. Understanding these perceptual differences in real-world contexts is therefore critical for designing social robots that feel more natural and engaging. Studying user responses during free, real-time interactions can help bridge the gap between artificial and human-like behavior, fostering greater trust and acceptance.

In this work, we investigate how user perceptions differ when interacting with a robot that is either autonomous or teleoperated. We assess perceived intelligence, safety, and user satisfaction in two cognitively demanding interactive scenarios: an information assistance task and a package delivery task. Since physical movement and gestural behavior are known to influence expectations of robot capabilities [5], each task was carried out under two conditions: one in which the robot could move and follow the participant's gaze, and another in which it remained static.

A key feature of our study is that both autonomous and teleoperated robots were capable of generating free-form speech and actions. Although operators followed task-specific instructions, their verbal responses were not restricted to pre-scripted phrases. This setup enabled participants to evaluate the robot's capacity for natural conversation and context-sensitive behavior.

Each participant interacted with a single type of operator four times: once for each combination of task and movement. This repeated interaction allowed us to include familiarity as an additional dimension of analysis, examining how users' perceptions evolved across multiple interactions.

To support this investigation, we developed a unified framework for both operator types. This framework has been released as an open-source resource to encourage reproducibility and to support future research on real-time human-robot communication.

The structure of this paper is as follows: Section II discusses relevant prior research, Section III outlines the methodology, Section IV presents the results, Section V offers a discussion of the findings, and Section VI concludes the study.

II. Related Work

This section reviews recent research on the perception of robots being teleoperated or acting autonomously. Particular emphasis is placed on AI-driven control, especially the integration of LLMs in robotic design. We additionally highlight work exploring how movement affects the perception of robotic capabilities.

A. Perception of Teleoperated Robots

Teleoperated robots have been deployed across a range of domains in recent years [6], [7], [8], [9]. However, relatively little research has directly compared user perceptions of robots controlled by human operators versus those governed autonomously through embedded AI systems. Some existing studies have aimed to improve teleoperation interfaces by incorporating social interaction techniques to enhance usability and user engagement [10]. In child-robot interaction contexts, findings suggest that children's awareness of a remote human operator can significantly influence their levels of trust and learning outcomes [11]. More generally, user perceptions of robots are also shaped by demographic variables; for example, user gender has been shown to affect responses to different robot voice types [12].

In social interview tasks, autonomous robots are often regarded as more intelligent than teleoperated ones, although teleoperated robots tend to be perceived as having greater social presence [13]. However, in task-learning contexts, [14] found that robots—regardless of being autonomous or teleoperated—are generally viewed as less capable and less intelligent compared to human learners.

Building on this previous work, our study consists in performing two tasks to explore how perceived intelligence, safety and satisfaction change when interacting with a robot that is either teleoperated or autonomous. To our knowledge, this is the first study with a setup that allows the robot—regardless of operator type—to listen to the participant and respond freely, without relying on pre-scripted responses. Although both operators followed task-related instructions, they had full flexibility in communication.

B. Robotic Frameworks With Large Language Models

A growing number of studies have focused on integrating Large Language Models (LLMs) into popular robotic frameworks like the Robot Operating System (ROS) to improve natural language understanding and generation capabilities.

The work of [15] presents a theoretical sketch of a framework to enable robot communication with LLMs and speech recognition and generation models. Similarly, [16] describes a ROS 2-integrated assistant that leverages LLMs to support users in programming micro-assembly tasks.

To improve accessibility for non-expert users, [17] introduces a ROS 2-based system that bridges the gap between natural language commands and robot navigation instructions using LLMs. In addition to language understanding, some robotic applications benefit from integrating visual context with Vision-Language Models [18].

To the best of our knowledge, we present the first publicly available end-to-end tested ROS 2 framework that enables autonomous full conversational interaction—including speech recognition, natural language generation, and speech synthesis—designed to support experimentation and reproducibility in LLM-driven human-robot interaction.

(a) (b)

Fig. 1. (a) User interacting with the robot in static mode. (b) Operator controlling the robot through the Rviz interface.

C. Impact of Robot Movement

Previous studies have demonstrated that non-verbal cues, such as gestures and gaze, play a significant role in shaping users' perceptions of embodied systems, particularly in terms of perceived intelligence and behavior [5], [19], as well as increasing user engagement and enjoyment [20].

Additionally, gaze behaviors have proven effective in managing social dynamics during interactions. For instance, human-like gaze mechanisms have enabled robots to signal addressee roles and influence users' rapport and attentiveness in multiparty conversations [21].

Building on these findings, our study incorporates robot movement and gaze tracking as key dimensions of analysis, alongside the influence of operator type, to better understand their combined effect on interaction and user perception.

III. Methodology

This study aims to investigate how several factors influence the perception of a robot's behavior, including the type of task performed, the level of user familiarity (measured by the number of interactions), the robot's autonomy, and its mobility capabilities. To examine these variables, we conducted a user study within a structured experimental framework using IVO [22], a bimanual robot equipped with object manipulation, autonomous navigation, and multimodal interaction abilities through vision, hearing, and speech.

A. Experiment Design

The experiment featured two modalities of movement for the robot. In dynamic mode, the robot autonomously identified and moved toward the participant, actively tracking and following their head movements throughout the interaction. In contrast, during static mode, the robot remained stationary as shown in Fig. 1(a), only extending its arm to accept objects when needed.

Participants completed two tasks, each performed twice consecutively—once in each movement mode. The order of the four task-mode combinations was randomized to ensure a balanced experience across all conditions. Fig. 2 illustrates the

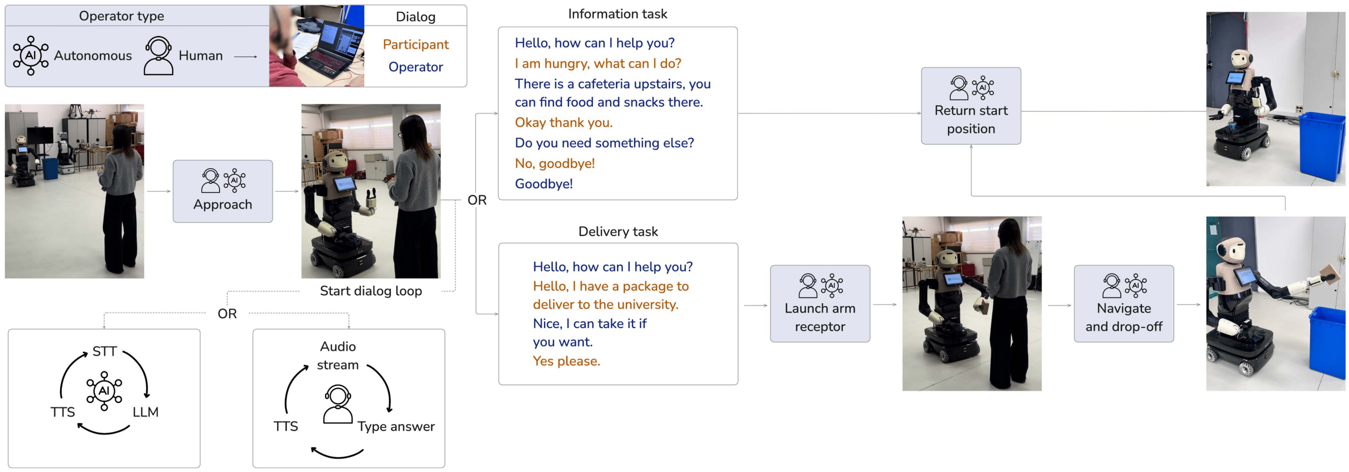

Fig. 2. Experimental workflow in dynamic mode. The diagram depicts the execution of two tasks—package delivery and information assistance—under two control conditions: AI-driven and human-operated. Although the overall task structure remains consistent, the method of controlling the robot differs significantly. In the AI-driven condition, the robot autonomously detects and approaches the participant. Participant speech is transcribed using Speech-to-Text (STT), processed by a Large Language Model (LLM), and the resulting output is either vocalized via Text-to-Speech (TTS) or used to activate specific robot actions (e.g., extending the robot's arm). In contrast, under human control, the operator manually navigates the robot, listens to the participant through an audio stream, and responds by either typing a message to be delivered via TTS or directly triggering the appropriate robot module. Upon task completion, both control modes involve instructing the robot to return to its starting position.

experimental process, including a simple example of interaction for each task in dynamic mode.

1) Task 1: Information Assistance: In this experimental task, the robot operated in the role of an information assistant. Both the human operator and the Large Language Model (LLM) were provided with a document in English containing fictionalized data pertaining to a university building. The LLM employed a Retrieval-Augmented Generation (RAG) framework [23], integrated with the Language-agnostic BERT Sentence Embedding (LaBSE) model [24], to interpret and generate responses to user queries across three languages: English, Catalan, and Spanish. Human operators were instructed to generate responses strictly grounded in the contents of the provided document, excluding extraneous or speculative information. Prior to interaction, participants received contextual briefings about the university environment and were encouraged to formulate questions typical of those directed to a concierge or front-desk assistant.

2) Task 2: Package Delivery: In this scenario, participants were given a small box and interacted freely with the robot to decide whether to hand it over, without scripted instructions. If they consented, the robot reached out to take the item. In the static condition, it held the item; in the dynamic condition, it delivered it to a drop-off point and returned.

B. Experimental Setup

For both robot control modes, we developed an architecture capable of capturing participant input, processing the responses, and synthesizing audio output from the corresponding text.

1) Autonomous Framework: To harness autonomous capabilities in the robot, we designed the architecture of PARLAM (Personalized Assistant Retriever Language Model), implemented using ROS 2 Humble Hawksbill and made publicly available at Github. The process begins with a speech input node, which captures incoming audio from the user and returns a text transcription using Vosk models [25]. This transcription is sent to a LLM-OpenAI's GPT-4o-mini [26]- node that generates a textual response, which can be an answer to the user or a flag to trigger robot actions. If it's an answer, it is sent to a speech output node, which employs gTTS [27] to convert it into audio for the robot to play back to the user. These three nodes function as ROS 2 service servers, with their interaction coordinated by a dedicated service client node. This client node also acts as an action server for a separate ROS 1 action client—part of an internal control system not included in the public repository. The ROS 1 action client initiates the interaction process and continuously monitors the dialogue state, including phases such as listening, reasoning, and responding. A ROS 1/ROS 2 bridge enables this hybrid architecture, which is necessary because key control modules—such as arm actuation and internal behavior routines—remain implemented in ROS 1.

2) Human Framework: We developed an RViz-based [28] interface that allowed the operator to observe the participant and control robot's movement and speech as shown in Fig. 1(b). The operator could hear the participant in real time but responded by typing messages. A ROS 1 node acting as a service client transmitted the typed text through the ROS bridge to the speech output node of PARLAM running in ROS 2.

C. Robot Behavior

The goal of this study was to examine potential differences in behavior between autonomous and human-operated robots, specifically in terms of language use and the response process to move the robot. Similarly to the voice, the robot's physical actions, embedded in a behavior tree, were unaffected by the operator type and were triggered as needed during the interaction, whether by the human or the LLM.

1) Human-Operated Robot: An initial exploratory study involving nine individuals (eight men, one woman) was conducted to gather insights into natural human operator behavior patterns. Participants responded to a mix of written and spoken queries. We found that converting spoken operator responses into the robot's voice required a transcription step, which was error-prone and inconsistent. To ensure a uniform robot voice and minimize bias in the user experience, we chose to implement only the option for operators to type their responses.

Two human operators, proficient in the three languages, were thoroughly briefed on their tasks and the identity of the robot (IVO). For the information assistance task, they had access to the same document as the language model, supplemented with Catalan and Spanish translations. In package reception, the goal was to establish trust with the user and complete the delivery effectively.

2) Autonomous Robot: The autonomous robot was crafted to emulate human interaction patterns, drawing on the pilot study for realistic language use and timing behavior. Prompts were written in English and varied depending on the nature of the task but included shared foundational components: the robot's identity (IVO), dialogue history, user queries, the interaction language, and its role as a university assistant. The conversational style was tailored to each task:

- Information assistance: The answers had to be based only on the context retrieved via RAG. The model had to avoid speculation, maintain linguistic precision, and not disclose its artificial origin.
- Package delivery: The model had to answer in casual tone, prioritizing efficiency over small talk. It had to refrain from initiating personal or unrelated dialogue and focus strictly on the task, unless the participant steered the conversation otherwise.

To prevent robot response time from affecting participants' experience, the autonomous robot included a delay before responding, simulating natural human timing. This delay was based on both thinking and typing times measured in the pilot study. Since our pilot results closely matched the typing speed distribution of [29], we used a similar equation to calculate the total response delay in autonomous mode:

$$delay = 1 + \mathcal{N}(0.3, 0.03) \times n\_char \tag{1}$$

where n_char denoted the number of characters in the input query.

D. Participants and Procedure

A total of 34 individuals took part in the study (24 male, 10 female), ranging in age from 18 to 64 years (M = 33.23, SD = 10.88). Participants were informed that the robot could be controlled either by an AI system or a human. This dual framing was central to the study's design, which collected data both on participants' perceptions of the robot and their ability to infer the operator type; however, in the present study we focus only on the perception aspect. A complementary analysis can be found in [30]. After each interaction, participants answered questions about their experience and indicated which operator they believed was in control, serving as a manipulation check. They were instructed to base their evaluations solely on the speech, without making assumptions about the operator's identity.

Volunteers engaged in single-turn interactions for as long as they wished, and were free to choose among three languages—Catalan, English, or Spanish—according to their language preferences. Overall, 14 participants selected Catalan, 10 chose Spanish, and 10 opted for English. The assignment of operator types was balanced across languages, yielding 7 sessions in Catalan, 5 in English, and 5 in Spanish for each operator group.

Participants were randomly assigned to a single operator condition for all four interactions, with 17 individuals per group, ensuring a balanced design for this between-subjects factor. Task type and movement modality were treated as within-subjects factors following a mixed-factorial design to account for these variables.

Before starting the experiment, all participants provided written informed consent using a form approved by the institution. The study protocol was reviewed and endorsed by the Ethics Committee of the Universitat Politècnica de Catalunya (UPC), and adhered to all applicable ethical guidelines and regulations (ID: 2024.020).

E. Measures and Analysis

Before starting the experiment, participants answered demographic questions regarding age and gender. After each interaction, participants were asked to complete a questionnaire designed to capture their subjective experience. To assess perceived intelligence and perceived safety of the robot, we used 5-point semantic differential scales based on the Godspeed Questionnaire [31]. Following the approach in [32], we used a modified version of the perceived safety subscale. Specifically, we reversed the miscoded "still-surprised" item to ensure consistency with the other items in the subscale, such as "anxious–relaxed" and "agitated–calm". We also measured user satisfaction using the 7-point subscales from the Usefulness, Satisfaction, and Ease of Use (USE) Questionnaire [33]. At the end of the four interactions, participants could provide additional feedback or suggestions to understand better their answers.

To examine how perceived intelligence, perceived safety, and satisfaction were affected by experimental factors—namely, number of interaction, task type, movement modality, and operator type—we used a linear mixed-effects model. Average scores for each dimension served as the dependent variables, with participant identifier included as a random effect to account for individual variability. We first analyzed a model including only main effects, then tested two additional models incorporating interaction terms: one between task type and operator type, and another between movement modality and operator type, to explore potential interactions involving the operator.

IV. Results

This section presents the findings of the questionnaire data analysis, structured around the three main dimensions explored

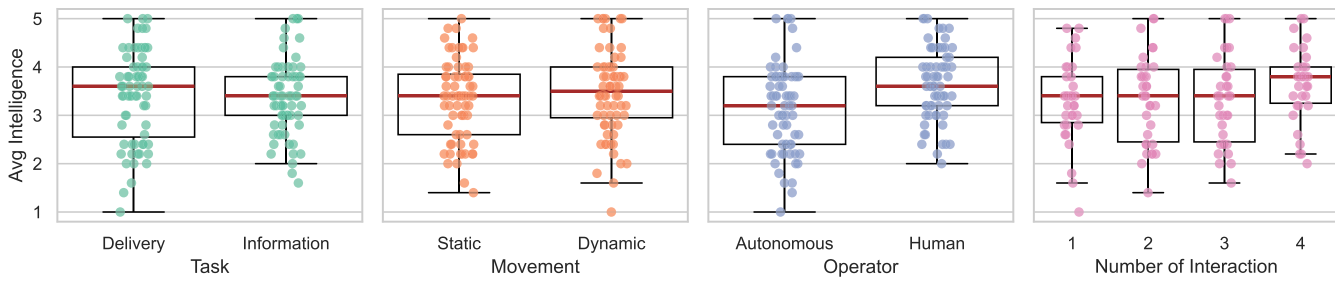

Fig. 3. Average perceived intelligence. Average ratings of perceived intelligence of all the experiments across all four dimensions and categories. Brown lines indicate the median values.

in the study: perceived intelligence, perceived safety, and user satisfaction.

A. Perceived Intelligence

The internal consistency of the perceived intelligence scale was high, with a Cronbach's alpha of 0.87.

The boxplots in Fig. 3 illustrate the distribution of the scores across the four experimental variables. Data indicates slightly higher intelligence ratings for human-operated robots compared to those autonomous. In contrast, no notable differences were observed between static and dynamic movement modalities, nor for task type distributions. Perceived intelligence remained relatively stable across repeated interactions, although a modest increase was noted during the fourth interaction.

1) Simple Mixed Effects Model: The analysis revealed that the fourth interaction had a significant positive effect on perceived intelligence ($\beta = 0.345, 95\%$ CI$= [0.015, 0.675], p = .04$). This effect was also present in the mixed effects with interaction terms presented below. Similarly, human-operated robots were associated with significantly higher intelligence ratings compared to autonomous ones ($\beta = 0.512, 95\%$ CI$= [0.127, 0.897], p = .009$). Other factors did not show statistically significant effects.

2) Model With Interaction Movement-Operator: For autonomous robots, the static modality led to lower perceptions of intelligence compared to the dynamic one ($\beta = -0.436, 95\%$ CI$= [-0.757, -0.115], p = .008$). In dynamic mode, operator condition did not significantly influence the perceived intelligence, with a confidence interval indicating uncertainty in the magnitude and direction of the effect ($\beta = 0.203, 95\%$ CI$= [-0.244, 0.651], p = .373$). However, the interaction term tells us that, in static mode, the presence of a human operator seemed to partially compensate the lack of movement, resulting in higher perceived intelligence than when the same static robot was autonomous ($\beta = 0.617, 95\%$ CI$= [0.161, 1.073], p = .008$).

3) Model With Interaction Task-Operator: When the robot was autonomous, it was perceived as more intelligent during the information task than during the delivery task ($\beta = 0.464, 95\%$ CI$= [0.163, 0.765], p = .003$). In contrast, for human-operated robots, the opposite pattern emerged: they were rated as more intelligent in the handover task ($\beta = 1.005, 95\%$ CI$= [0.565, 1.445], p < .001$). The interaction analysis confirmed this crossover effect—when a human operator was involved, the information task actually led to lower intelligence ratings than the delivery task ($\beta = -0.987, 95\%$ CI$= [-1.413, -0.561], p < .001$).

B. Perceived Safety

The perceived safety scale demonstrated high internal consistency, with a Cronbach's alpha of 0.83. As shown in Fig. 4, median scores were high and remained similar across task, movement and operator type, although information, dynamic mode, and autonomous condition exhibited slightly greater variability and more low outliers, suggesting occasional reductions in perceived safety. No clear trend emerged between number of interaction and perceived safety, as distributions were consistent across interaction levels.

Across the three fitted models, we found no statistically significant evidence of a relationship between the four variables and perceived safety, and the confidence intervals indicated substantial uncertainty in both the direction and magnitude of the effects.

C. Satisfaction

The internal consistency of the satisfaction scale was very high, with a Cronbach's alpha of 0.9. Fig. 5 shows generally similar satisfaction levels across most categories, but operator type stands out as the most influential factor, with human operators rated slightly higher on average.

1) Simple Mixed Effects Model: The analysis revealed a significant negative effect of performing the information task compared to the delivery ($\beta = -0.388, 95\%$ CI$= [-0.705, -0.071], p = .016$). Human-operated robots were confirmed to be higher in satisfaction compared to autonomous ones ($\beta = 0.682, 95\%$ CI$= [0.071, 1.293], p = .029$). No significant effects were found for the number of interactions or movement, suggesting these factors did not meaningfully influence satisfaction.

2) Model With Interaction Movement-Operator: The results confirmed a significant negative effect of the information task on satisfaction ratings ($\beta = -0.388, 95\%$ CI$= [-0.704, -0.072], p = .016$). The fourth interaction showed a positive but non-significant trend ($\beta = 0.336, 95\%$ CI$= [-0.112, 0.784], p = .141$). The main effects of movement and

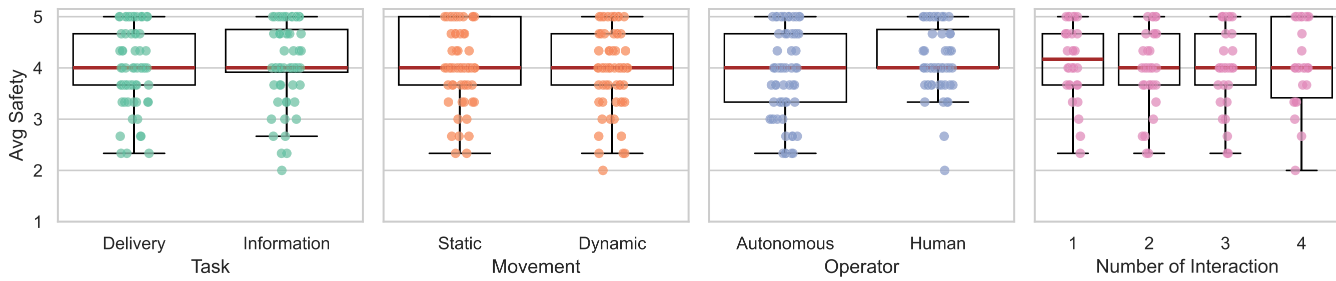

Fig. 4. Average perceived safety. Mean ratings of safety across all experiments, covering all four dimensions and categories. The median is indicated in brown.

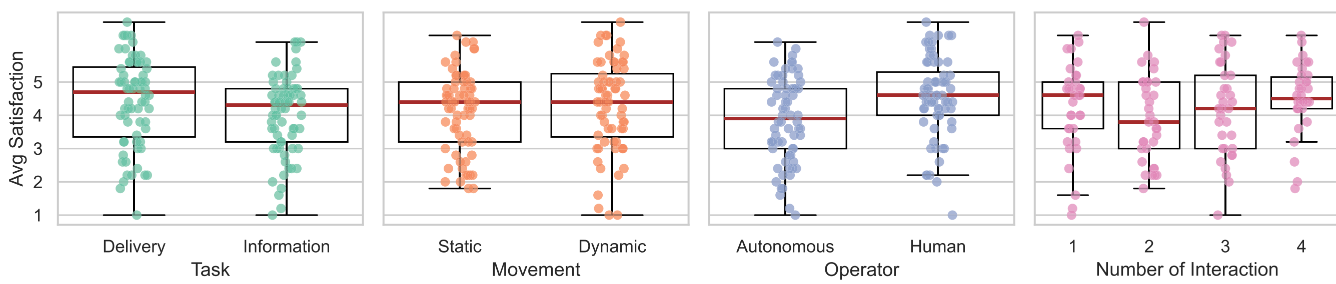

Fig. 5. Average perceived satisfaction. Evaluations of the satisfaction for each dimension averaged across all the experiments. Brown lines indicate the median values.

operator type were also not significant, nor was the interaction term between static modality and human operator ($\beta = 0.432, 95\%$ CI$= [-0.205, 1.068], p = .183$).

3) Model With Interaction Task-Operator: The results showed a significant main effect of operator type during the delivery task, with human operators associated with higher satisfaction ratings compared to autonomous robots ($\beta = 1.190, 95\%$ CI$= [0.508, 1.872], p = .001$). No significant main effect of task type was found for AI-driven robots ($\beta = 0.119, 95\%$ CI$= [-0.309, 0.547], p = .585$), indicating similar satisfaction ratings across tasks for this condition. However, a significant interaction term between task and operator type was obtained ($\beta = -1.015, 95\%$ CI$= [-1.62, -0.409], p = .001$), revealing that while human operators were rated more satisfying in the delivery, they were perceived as less satisfying than autonomous for information assistance.

V. Discussion

In this study, we investigated user perception in a human-robot interaction scenario involving two distinct tasks, two robot mobility modes, and two control types: AI-driven and human-operated. A total of 34 participants interacted with a single operator type while experiencing both robot modalities and completing both tasks. User perceptions were evaluated along three key dimensions—intelligence, safety, and satisfaction—which were analyzed across four factors: operator type, mobility mode, task type, and number of interactions. While safety perceptions remained relatively stable across conditions, ratings of intelligence and satisfaction showed significant variation depending on specific experimental factors. In the following discussion, we examine potential explanations for these results and consider their implications for the design and deployment of future human-robot interaction systems.

Although our study resulted in 136 observations across conditions, the effective sample size for estimating the effect of each factor remains relatively modest. As such, the present findings should be regarded as an initial step toward understanding these relationships rather than as definitive or broadly generalizable conclusions.

With regard to multilingual capabilities, an objective analysis of the conversations conducted by both types of operators revealed no qualitative differences in speech across languages. A more detailed examination of conversational aspects, such as user feedback on the interaction, linguistic style, social and emotional features, and knowledge and reasoning abilities, is presented in [30].

A. Mimicking Human Behavior Patterns

After analyzing both the questionnaire data and participant feedback—including comments and recorded conversations—we conclude that the system architecture performed effectively for both autonomous and human-operated robots.

In the autonomous version, speech recognition accurately transcribed user input in most cases, with errors mainly due to non-standard accents. The LLM responded appropriately to user prompts and triggered action modules in real time. Text-to-Speech output was fluent and clear.

For the human operator, interactions also occurred in real time. Most of the time of the interaction was spent thinking and typing responses rather than on system latency. Participants noted that replies felt slow for both operators, suggesting that the modeled human-like delay was realistic. However, some users were bewildered by this fact and even found this delay bothersome, especially during the information assistance task. This suggests a need to either signal to users that the system is processing a response or eliminate the delay entirely for faster autonomous interactions. A more extensive analysis of quantitative conversational data can be found in [30].

The fact that participants perceived the robot as significantly more intelligent during the fourth interaction suggests that familiarity with the system enhances perceived intelligence, possibly due to adjusted expectations.

B. The Perceived Effect of Movement

Previous studies have shown that gestures can influence the perception of embodied systems' intelligence and behavior [5], [19]. In our study, the most notable differences between operator types emerged when movement was removed. In the static condition, human operators were perceived as more intelligent, consistent with the general tendency of participants to attribute higher intelligence to human agents. When movement was introduced, however, participants rated autonomous systems at a similar level of perceived intelligence to human-operated ones. This pattern suggests that dynamic embodiment, combined with coherent conversational and action sequencing, may have helped the autonomous system narrow the gap with human teleoperation in participants' perceptions.

C. Task Perception and Performance

Although human operators were generally rated as more intelligent, this trend reversed depending on the task type. In package delivery, humans were perceived as more intelligent and satisfying. However, in the information task, the autonomous system outperformed humans in both metrics.

For AI-driven systems, satisfaction remained consistent across tasks, but perceived intelligence was higher in the information task than in package reception. This may be linked to known limitations of LLMs in spatial and contextual reasoning [34], particularly since no environmental or visual input was provided. Adding real-time visual input—to equal human operator knowledge—or embodiment cues (e.g., specifying that the delivery is done with the left arm) could help improve autonomous performance in physically grounded tasks.

It is also worth noting that both type of operators were instructed not to invent information. Humans adhered strictly to this guideline, often resulting in hesitation or delays when unsure. While the LLM also respected this instruction, it tended to offer alternative suggestions rather than explicitly stating "I don't know." This difference in strategy and the inherent ability of vector search models to access rapidly large texts, may have contributed to the higher ratings for autonomous performance in the information task.

VI. Conclusion

This study investigated the impact of operator type—human-operated versus autonomous (AI-driven)—on user perceptions of robot intelligence, safety, and satisfaction. To comprehend these dynamics, we examined four key factors: number of interactions, task type, robot mobility mode, and operator type. Our findings revealed that task type significantly influenced perceived performance and user experience. Human-operated robots demonstrated superior performance in spatial navigation tasks, likely attributable to the human operator's capacity for real-time visual interpretation and situational awareness. In contrast, the AI-driven system outperformed humans in tasks that required rapid data retrieval and processing, highlighting the advantages of automation in cognitively intensive, information-based scenarios.

Robot mobility mode also played a substantial role in shaping user perceptions. The robot was consistently rated as more intelligent and engaging when it exhibited dynamic, responsive movement patterns than when it displayed static or rigid behaviors. These findings suggest that movement expressiveness can serve as a nonverbal cue that enhances perceived cognitive competence.

These findings highlight the respective strengths of human and AI control in interactive robotics and the importance of aligning operator type with task demands. They also show how robot embodiment and behavior shape user attitudes—critical considerations for designing adaptive, user-centered systems. While preliminary, the results point to promising directions that future studies with larger samples should further test and generalize. Additionally, future work could benefit from not explicitly indicating the operator duality, allowing for a more naturalistic assessment of user perceptions.

References

[1] W. X. Zhao et al., "A survey of large language models," 2023, *arXiv:2303.18223*.
[2] F. Zeng, W. Gan, Y. Wang, N. Liu, and P. S. Yu, "Large language models for robotics: A survey," 2023, *arXiv:2311.07226*.
[3] S. Woods, M. Walters, K. L. Koay, and K. Dautenhahn, "Comparing human robot interaction scenarios using live and video based methods: Towards a novel methodological approach," in *Proc. 9th IEEE Int. Workshop Adv. Motion Control*, 2006, pp. 750–755.
[4] Y. Liu and S. Liao, "The influence of robot autonomy on perception distance, acceptance and subjective norm," *J. Phys., Conf. Ser.*, vol. 2037, no. 1, 2021, Art. no. 012116.
[5] M. Choi, R. Kornfield, L. Takayama, and B. Mutlu, "Movement matters: Effects of motion and mimicry on perception of similarity and closeness in robot-mediated communication," in *Proc. Conf. Hum. Factors Comput. Syst.*, 2017, pp. 325–335.
[6] K. Darvish et al., "Teleoperation of humanoid robots: A survey," *IEEE Trans. Robot.*, vol. 39, no. 3, pp. 1706–1727, Jun. 2023.
[7] D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "Teleoperation of multiple social robots," *IEEE Trans. Syst., Man, Cybern.-Part A, Syst. Humans*, vol. 42, no. 3, pp. 530–544, May 2012.
[8] I. El Rassi and J.-M. El Rassi, "A review of haptic feedback in tele-operated robotic surgery," *J. Med. Eng. Technol.*, vol. 44, no. 5, pp. 247–254, 2020.
[9] G. Yang et al., "Keep healthcare workers safe: Application of teleoperated robot in isolation ward for COVID-19 prevention and control," *Chin. J. Mech. Eng.*, vol. 33, pp. 1–4, 2020.
[10] D. J. Rea, S. H. Seo, and J. E. Young, "Social robotics for nonsocial teleoperation: Leveraging social techniques to impact teleoperator performance and experience," *Curr. Robot. Rep.*, vol. 1, no. 4, pp. 287–295, 2020.

[11] J. M. K. Westlund and C. Breazeal, “Transparency, teleoperation, and children’s understanding of social robots,” in *Proc. 11th ACM/IEEE Int. Conf. Hum.-Robot Interaction*, 2016, pp. 625–626.
[12] S. Song, J. Baba, J. Nakanishi, Y. Yoshikawa, and H. Ishiguro, “Mind the voice!: Effect of robot voice pitch, robot voice gender, and user gender on user perception of teleoperated robots,” in *Proc. Extended Abstr. Conf. Hum. Factors Comput. Syst.*, 2020, pp. 1–8.
[13] J. J. Choi, Y. Kim, and S. S. Kwak, “The autonomy levels and the human intervention levels of robots: The impact of robot types in human-robot interaction,” in *Proc. 23rd IEEE Int. Symp. Robot Hum. Interactive Commun.*, 2014, pp. 1069–1074.
[14] M. Bennett, T. Williams, D. Thames, and M. Scheutz, “Differences in interaction patterns and perception for teleoperated and autonomous humanoid robots,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2017, pp. 6589–6594.
[15] H. Ye, “Auromix. ROS-LLM,” 2023. [Online]. Available: https://github.com/Auromix/ROS-LLM
[16] R. Wiemann, N. Terei, and A. Raatz, “Large language model for assisted robot programming in micro-assembly,” *Procedia CIRP*, vol. 130, pp. 244–249, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S221282712401240X
[17] J. P. Espada, S. Y. Qiu, R. G. Crespo, and J. L. Carús, “Leveraging large language models for autonomous robotic mapping and navigation,” *Int. J. Adv. Robotic Syst.*, vol. 22, no. 2, 2025, Art. no. 17298806251325965.
[18] “Ros2 nanollm,” 2024. [Online]. Available: https://github.com/NVIDIA-AI-IOT/ros2_nanollm
[19] C.-M. Huang and B. Mutlu, “Modeling and evaluating narrative gestures for humanlike robots,” in *Proc. Robot., Sci. Syst.*, 2013, vol. 2.
[20] A. Kim, J. Han, Y. Jung, and K. Lee, “The effects of familiarity and robot gesture on user acceptance of information,” in *Proc. 8th ACM/IEEE Int. Conf. Hum.-Robot Interaction*, 2013, pp. 159–160.
[21] B. Mutlu, T. Kanda, J. Forlizzi, J. Hodgins, and H. Ishiguro, “Conversational gaze mechanisms for humanlike robots,” *ACM Trans. Interactive Intell. Syst.*, vol. 1, no. 2, pp. 1–33, 2012.
[22] J. Laplaza et al., “Ivo robot: A new social robot for human-robot collaboration,” in *Proc. 17th ACM/IEEE Int. Conf. Hum.-Robot Interaction*, 2022, pp. 860–864.
[23] P. Lewis et al., “Retrieval-augmented generation for knowledge-intensive nlp tasks,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 9459–9474.
[24] F. Feng, Y. Yang, D. Cer, N. Arivazhagan, and W. Wang, “Language-agnostic BERT sentence embedding,” in *Proc. 60th Annu. Meet. Assoc. Comput. Linguist.*, vol. 1, 2022, pp. 878–891.
[25] N. V. Shmyrev, “Vosk speech recognition toolkit: Offline speech recognition API for android, iOS, raspberry pi, and servers with Python, Java, C#, and Node,” 2020. [Online]. Available: https://github.com/alphacep/vosk-api
[26] OpenAI, “GPT models: 4-o-mini,” 2024. [Online]. Available: https://platform.openai.com/docs/models/gpt-4o-mini
[27] P. N. Durette, “gTTS-Google text-to-speech,” 2024. [Online]. Available: https://pypi.org/project/gTTS/
[28] ROS Development Team, “RViz. robot operating system (ROS),” 2018. [Online]. Available: http://wiki.ros.org/rviz
[29] C. R. Jones, I. Rathi, S. Taylor, and B. K. Bergen, “People cannot distinguish GPT-4 from a human in a turing test, in *Proc. ACM Conf. Fairness, Accountability, Transparency*, 2025, pp. 1615–1639.
[30] L. Hriscu, A. Sanfeliu, and A. Garrell, “AI or human? Understanding perceptions of embodied robots with LLMs,” in *Proc. 34th IEEE Int. Conf. Robot Hum. Interactive Commun.*, 2025, pp. 917–923.
[31] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, “Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots,” *Int. J. Social Robot.*, vol. 1, pp. 71–81, 2009.
[32] C. Y. Kim, C. P. Lee, and B. Mutlu, “Understanding large-language model (LLM)-powered human-robot interaction,” in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interaction*, 2024, pp. 371–380.
[33] A. M. Lund, “Measuring usability with the use questionnaire12,” *Usability Interface*, vol. 8, no. 2, pp. 3–6, 2001.
[34] Y. Bang et al., “A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination,and interactivity,” in *Proc. 13th Int. Joint Conf. Natural Lang. Process. 3rd Conf. Asia-Pacific Chapter Assoc. Comput. Linguist.*, vol. 1, 2023, pp. 675–718.