What Drives You to Interact?: The Role of User Motivation for a Robot in the Wild

Amy Koike*‡, Yuki Okafuji *†, Kenya Hoshimure*†, and Jun Baba*†

* *CyberAgent*, Tokyo, Japan

† *Osaka University*, Osaka, Japan

‡ *Univsesity of Wisconsin-Madison,* Wisconsin, United States

Email: ekoike@wisc.edu, {okafuji_yuki_xd, hoshimure_kenya, baba_jun}@cyberagent.co.jp

Abstract—In this paper, we aim to understand how user motivation shapes human-robot interaction (HRI) in the wild. To explore this, we conducted a field study by deploying a fully autonomous conversational robot in a shopping mall over two days. Through sequential video analysis, we identified five patterns of interaction fluency (Smooth, Awkward, Active, Messy, and Quiet), four types of user motivation for interacting with the robot (Function, Experiment, Curiosity, and Education), and user positioning towards the robot. We further analyzed how these motivations and positioning influence interaction fluency. Our findings suggest that incorporating users' motivation types into the design of robot behavior can enhance interaction fluency, engagement, and user satisfaction in real-world HRI scenarios.

Index Terms—human-robot interaction; social robots; service robots; field experiment; qualitative analysis

I. INTRODUCTION

Observing user behaviors in human-robot interaction (HRI) provides valuable insights for designing interactions to enhance a robot's acceptance, reliability, and user satisfaction.

In particular, such observations are often conducted in the context of HRI failures, where researchers aim to understand types of robot failures and corresponding user behaviors [*e.g.,* 1, 2, 3, 4]. While this focus on robot failures allows for a systematic analysis, much of the research has narrowly focused on the moments immediately before and after errors, often overlooking the broader dynamics of the entire interaction. Greater design insights could be gained by expanding the scope of observation to include the entire interaction, as past HRI studies have suggested that understanding how people encounter robots provides valuable design insights, particularly regarding the impact of first impression of robots [5, 6, 7, 8].

Moreover, prior HRI observation has been often conducted in controlled laboratory settings, where scripted scenarios and consistent user motivations drive the interaction with robots. These environments may hinder natural behaviors that emerge in the wild. For example, field studies revealed unique user behaviors, such as ignoring [9, 10] or disrupting robots [11], that are unlikely to be observed in a laboratory. This difference highlights the importance of studying user motivation as a key factor to shape real-world HRI.

To address the limitations of narrowly focusing on failure moments and controlling user motivation, it is crucial to study user behaviors throughout the entire interaction (*i.e.,* from the initial approach to the end of the interaction) in a field study.

Fig. 1. In this study, we deployed an autonomous conversational robot in a shopping mall. Our analysis revealed four types of user motivation for the robot: *Function*, *Experiment*, *Curiosity*, *Education*.

In this work, we designed an autonomous conversational robot driven by Large Language Models (LLMs), to provide route guidance in a shopping mall. During a two-day deployment in a mall in Japan, we collected video recordings of interactions. Through video analysis, we examined how individuals approached, engaged with, and departed from the robot. Our analysis identified five types of interaction fluency–Smooth, Awkward, Active, Messy, and Quiet; and four motivations for interacting with the robot –Function, Experiment, Curiosity, and Education (Fig. 1). We further analyzed how these user motivations influence interaction fluency.

In the remainder of the paper, we discuss related work, describe our system architecture and study design, and present the findings from our field study. We conclude with the implications of our findings and a discussion of potential limitations. Our work makes the following contributions:

1) *Empirical*: an understanding of *user motivations* by observing user behaviors throughout interactions with the robot in the wild;
2) *Design Implication*: a guidance for HRI researchers and practitioners on how to apply our findings regarding the influence of *user motivations* on interaction fluency.

II. RELATED WORK

A. HRI in the Wild

Since early 2000s, "HRI in the Wild" has been a key subject to study [*e.g.,* 12, 13, 14, 15, 16, 17]. Even in the 19th Annual ACM/IEEE International Conference on Human-Robot Interaction, the main theme was "HRI in the Wild" which aimed *"to bring human-robot interaction out of the lab and into everyday life [18]."*

That being said, laboratory experiments allow researchers to solely focus on specific variables within their research scope as laboratory experiments are typically conducted in a controlled environment with a predefined scenario. Moreover, laboratory experiments offer the advantage of controlling "participants' motivations" to interact with the robot, often by recruiting participants within the university or compensating them for the involvement. While such controlled setups offer consistency and rigor in study design, we often face challenges when applying the findings from laboratory studies into real-world HRI scenarios [19]. Moreover, it has been shown that users' behaviors to interact with a robot can be influenced by the surroundings and environment [20, 21, 22], which implies that users behave differently between lab studies and field studies.

Several previous studies that conducted field experiments have reported unique and intriguing human behaviors towards robots that are unlikely to be observed in laboratory experiments. For example, Rosenthal-von der Pütten et al. [23] reported people's "testing actions" such as waving in front of the robot's face, raising their eyebrows in an exaggerated manner, and showing a grimace face in order to test how the robot would react to them. Agrigoroaie et al. [24] reported one of their participants drew eyebrows for their robot. Service robots in real-world environments are often ignored by users due to their low social presence [9, 10], or even bullied [25]. Even when users start interactions, the users often interrupt interactions and leave during the robot is talking [11]. Brown et al. [26] studied a robotic trash can in order to observe spontaneous interaction in busy urban public spaces, and reported how people use the trash can differently. Unique interactions are also reported in the context of a child-robot interaction [*e.g.,* 27, 28]. For instance, Michaelis et al. [28] conducted an in-home deployment of a reading companion robot for a child, and found that there are family members involvements, which is unlikely to be observed in a lab study because the robot was designed for dyadic interactions.

Prior work has shown that field studies allow us to observe unexpected interactions in open-ended environments; however, there remain unexplored opportunities, particularly in understanding user motivations for interacting with a robot.

Furthermore, with large language models (LLMs), we now have greater flexibility to test conversational robots in the wild. As LLMs can facilitate spontaneous and improvised conversations, we are able to explore off-scripted human-robot interactions across diverse contexts and populations. This study uses a fully autonomous conversational robot to guide directions in a shopping mall, as well as engage in small talk.

B. HRI Failures

Robot failures are a major concern when bringing human-robot interaction in real-world settings. This is because such failures can significantly affect interaction fluency and user satisfaction. To address these impacts, numerous HRI studies have explored robot failures from various perspectives such as their effects on user's perceptions towards the robot [*e.g.,* 29, 30, 31, 32, 33, 34], error mitigation and recovery strategies [*e.g.,* 35, 36, 37, 38, 39, 40, 41], explainable errors [*e.g.,* 42, 43], or error-aware systems [*e.g.,* 44, 45, 46, 47].

Another area of HRI failure research focuses on identifying the types of errors or failures that robots can encounter [*e.g.,* 1, 2, 3, 4]. Such taxonomy-building research enables researchers, developers, and interaction designers to systematically understand and analyze failures and even ideate improved HRI designs, such as recovery strategies and robot behavior design. For instance, Honig and Oron-Gilad [1] proposed an inclusive HRI failure taxonomy that combines system- and human-oriented classifications, specifically technical failures and interaction failures. Similarly, Tolmeijer et al. [3] introduced a comprehensive taxonomy categorizing failures into four types: Design, System, Expectation, and User failures. Recent works have extended the development of taxonomies by deepening their scope, focusing on specific contexts, or introducing novel dimensions. For example, Tian and Oviatt [2] conducted an in-depth exploration of social errors, resulting in five categories and 30 subcategories. In the context of service robots, Choi et al. [48] classified failures as Outcome Failures and Process Failures, drawing inspiration from marketing literature, while Xing et al. [49] identified two types of failures for service chatbots: Core Service Failures and Interactive Service Failures. Lastly, Kamino et al. [50] proposed five distinct interaction failure types based on severity: Terminal, Non-critical, Recoverable, and Favorable.

Our method is inspired by those taxonomy-building research, but we extend the context by focusing on users' behavior throughout the interaction. As field studies allow us to observe a diverse users with uncontrolled motivations, we believe our sequential analysis of user behaviors offers invaluable findings and insights for understanding HRI failures.

III. METHOD

In this section, we describe our system and study design. All research activities were reviewed and approved by the Ethical Review Board of Osaka University.

A. Robot and Equipment

We used a humanoid robot named Sota for our study (Fig. 2). Sota robots have been widely employed in various HRI studies as service robots [*e.g.,* 51, 52]. Sota features a torso, head, and two arms, allowing it to perform body gestures. To complement Sota's verbal interactions, we integrated a 13-inch display that shows recognized text and content synchronized with Sota's current state. Both the robot and the display were mounted on a 3D-printed container, which also housed a 180-degree fisheye camera, microphone, speaker, and cables.

Fig. 2. (a) Sota Robot: The Sota robot, a 27 cm-tall tabletop social humanoid, used in our study, is capable of performing bodily gestures. (b) System Setup: The system consisted of the robot and a display, which complemented the robot's verbal interactions by showing usage instructions, internal states, and route guidance. (c) Study Setting: The system was deployed on the ground floor of a shopping mall, positioned next to a mall map. A video camera was installed behind the robot, and an experimental notice was placed beside the system.

B. Autonomous Conversational System

We implemented a fully autonomous conversational system for shopping mall guidance. The system comprises four components: Recognition, Dialogue Management, Action Management, and Modality Control (Fig. 3).

1) Recognition: A 180-degree fisheye camera captures videos to detect visitors' presence, and Posenet [53] extracts pose keypoints of visitors. Distances between these keypoints are calculated to estimate the visitor's proximity to the robot. The visitor with the closest distance is selected, and changes in this distance determine whether the visitor is nearby, detected, or absent. Speech input is captured by a microphone, processed through the Google Speech-to-Text API, and passed to the system for further interaction.

2) Dialogue Management: To manage dialogue, we employed a state transition model that defines three states based on the visitors' proximity: when no visitors are present (**S1**), when visitors are detected (**S2**), and when visitors are close enough to interact (**S3**). To prevent misrecognition in noisy environments, the speech recognition system is switched on or off depending on the current state. When the current state is **S1** or **S2**, speech recognition is deactivated and only enabled when the current state is **S3**. Upon triggering **S2**, Sota turns its head toward the users and greets them with template utterances. When the current state is **S3** and speech recognition results are received, the system responds to the user's voice input by executing the LLM Action described later.

3) Action Management: The system includes two types of actions: a rule-based action that produces template responses, and a GPT-4o-powered generative model [54] based action that generates speech. The rule-based action is proactively triggered when the state transition happens. Sota uses template phrases based on its state; for example, Sota says "Hello! Welcome to our shopping mall! Feel free to ask any questions!" when it is transitioned to **S2**.

The generative model was used to generate Sota's responses to user's speech input. The system incorporated a database containing comprehensive mall information for providing route guidance. When Sota recognized the speech, the system combined the dialogue history with the mall information to generate an appropriate response. The prompt provided to GPT-4o follows the structure outlined in Table I in the appendix and consists of the role description, task policy instructions, the mall information, and the dialogue history. The role specifies that the model should take on the persona of Sota and act in accordance with the task policies. The task policy includes 1) calling a navigation function when users are looking for stores or facilities and 2) responding appropriately to other inquiries based on mall information and dialogue history. The pre-programmed navigation function displays an animated route from the current location to the destination on a screen, accompanied by a template-based spoken utterance. System responses were generated each time a new user utterance was made as long as the system state was in either **S2** or **S3**.

4) Modality Control: Along with the verbal response, Sota performs hand gestures. To design the gestures, we followed previous HRI work that pairs Sota's gestures with its generated responses [52]. In total, we defined approximately 40 pairs of words and gestures; *e.g.,* raising a hand when the generated response includes "Hello," or raising both hands for "Thank you." If a generated response does not include any of the predefined words, the system performs small, random arm and neck movements, which are meant to be "secondary motions" to signal to users that the robot is active. Sota is equipped with LEDs as visual indicators, installed in both its eye lines and mouth. The LED colors change based on the generated response. The display content is synchronized with the robot's current state. In its initial state, it shows instructions on how to interact with Sota. During response generation, it displays "I am thinking...," and when providing the route guidance, it presents a map with the suggested animated route. The robot's speech synthesis is generated using VoiceText Web API [55], with the Japanese character voice "Hikari" selected to match Sota's appearance.
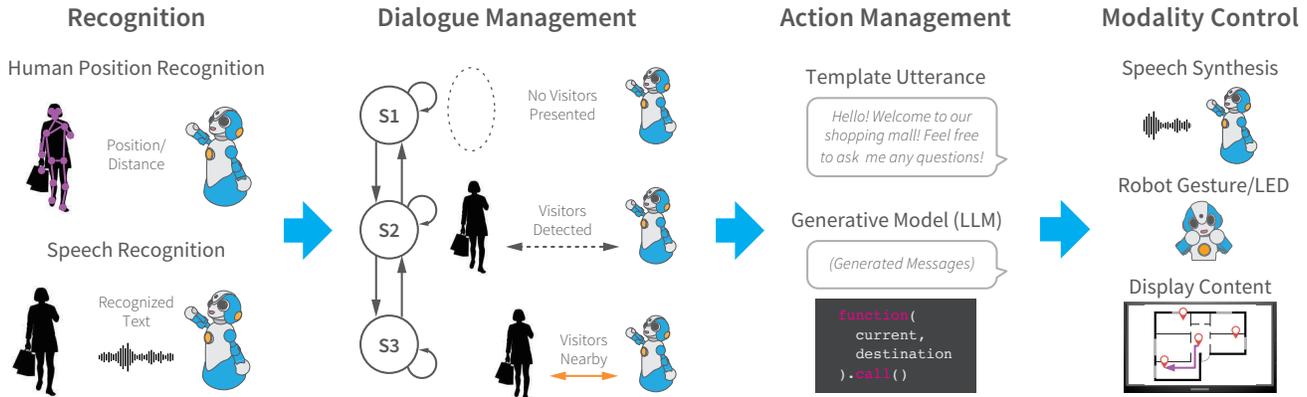
Fig. 3. System Architecture: Our autonomous conversational system consists of four components, from left to right: Recognition, Dialogue Management, Action Management, and Modality Control. The system can recognize human position and speech. Dialogue is managed using a state transition model, which defines three states based on visitors' proximity. Two types of actions are used for action management: template utterances and a generative model. Template utterances are triggered during a state transition, while the generative model is activated when user speech is recognized in S3, where users are nearby. Lastly, modalities such as speech synthesis, robot gestures, and display content are aligned with the action and the robot's current state. The only error the robot can detect is a Text-to-Speech failure, either from network or API issues. In response, Sota prompted users to retry with phrases such as "Hmm, sorry. Could you try again?" and displayed instructions on its screen.

C. Data Collection through Field Experiment

Our field study was conducted at a shopping mall located in Osaka, Japan. The mall is one of the largest in the area with three floors. The system was installed next to a mall map on the ground floor (Fig. 2(c)). The experiment lasted for two days with the system operating for eight hours each day, and recorded as video footage. The users were informed that video recording would be collected and consented to their videos being used as part of the research via instructions placed next to the robot. At least one experimenter was present in the space to address any issues. Although the experimenter stood slightly apart from the system, they continuously monitored the system and interactions.

As a result, we obtained 17 hours of video recording. From the recording, we extracted individual clips where interactions occurred. In this paper, we define an interaction as *any sequence in which visitors stopped to look at Sota regardless of whether a conversation initiated*. In total, we identified 232 interactions, which the duration ranged from six seconds to five minutes. Screencaptures from the clips are shown in Fig. 1. Due to the nature of field study, we could not collect users' biographical information. Instead, we analyzed all the clips to estimate the perceived ages of the individuals. The process involved three researchers (R1 – R3) and one descriptor (D1). Two researchers primarily led the process, while the others participated in discussions as needed. We identified 565 people who interacted with the robot, including 126 preschoolers; 39 elementary school students; 29 middle and high schoolers; 51 individuals in their twenties; 180 in their thirties; 117 middle-aged adults; and 23 older adults.

D. Analysis Procedure

We adopted the video analysis process outlined by Kamino et al. [50] that utilizes thick description [56]. Following their approach, R1 – R3 and D1 generated detailed description of

the interaction captured in each clip. After completing the descriptions, R1 and R2 examined them to ensure the researchers and descriptor agreed on the interpretations. Subsequently, the researchers engaged in coding followed a thematic analysis approach [57]. To build codes, we employed a deductive and inductive approach: a part of the codes was derived from our research focus while a part of the codes emerged by open-coding. R1 first reviewed all descriptions to familiarize themselves with the data, then conducted open coding for 20% of the dataset. The codes were then categorized and refined. This refinement process was conducted iteratively and R2 was involved in this process from the second iteration. R1 coded 75% of the data while R2 coded the remainder. Cohen's Kappa was calculated to measure the agreement between the two raters by using 10% of the data. The reliability was sufficiently high ($\kappa = 0.71$). Disagreements were resolved through discussion. Once the coding was completed, the researchers examined the frequency and patterns of the codes and iteratively discussed to develop themes.

E. Codebook and Coding Process

In this section, we describe the codes and coding process. The codebook is structured into four high-level categories: Behavioral Codes, Conflict Types, Recovery Actions, and Timing to Leave. While Conflict Types and Recovery Actions were primarily informed by prior literature, Behavioral Codes and Timing to Leave emerged through open coding. Additionally, proxy actions, which are unique to our study, provide a novel perspective on the observed phenomena. For our final codebook, please refer to Table II in the appendix. Our coding process followed three sequential phases: the *Encounter Phase*, *Interaction Phase*, and *Leave Phase*. The *Encounter Phase* marks the initial moment when the user first encounters the robot. The *Interaction Phase* captures the sequence of events following the user's decision to stop or

initiate a conversation. Finally, the *Leave Phase* occurs when the user departs, concluding their interaction with the robot.

First, Behavioral Codes, which were used for annotating visitors' behavior, were assigned in all the three phases. Researchers selected codes from a shared list across the three phases. We defined 35 behavioral codes including four non-verbal social signals; eight verbal social signals; five proxy actions; four state actions; and 13 event actions. Multiple codes were assigned when several behaviors were observed.

In addition to the Behavioral Codes, *Interaction Phase* was coded as follows: 1) Conflict Types, 2) Recovery Actions by the user, and 3) whether the robot initiated error recovery. For coding Conflict Types, we identified four types of *conflicts*: *overlapping speech and poor timing*, *misrecognizing and responding to unintended input*, *guiding incorrect routes*, and *system errors*. For Recovery Actions taken by users, we identified five patterns: *wait for the robot to finish speaking*, *speak to the robot again*, *enhance audibility*, *improve clarity*, and *check the display*. Similar to the Behavioral Codes, multiple codes were assigned when several Conflict Types and Recovery Actions were observed.

Finally, we identified three types of timing when users left the robot upon completion of the interaction (Timing to Leave): specifically completion of *route guidance*, *conversation* or *observation*. We also recognized cases where users left the robot mid-conversation, which are *leaving while the robot is processing*, *leaving immediately after the robot starts providing route guidance*, *leaving due to a conflict*, and *leaving due to interruption by a side participant*.

IV. FINDINGS

Our field study revealed 163 out of 232 interaction groups initiated a conversation with the robot. Within those 163 groups, 110 interactions included a conversational failure due to the *conflicts*. The most frequent conflict was *overlapping speech and poor timing*, occurring 81 times. This was followed by *system errors* from the robot, which were recorded 52 times. Additionally, we observed 18 instances of the robot *provided incorrect route guidance*, and 11 cases of the robot *misrecognizing and responding to unintended input*. Although we report statistics on conflicts, this paper does not delve into the specifics of individual conflicts. Instead, our focus is on how these conflicts impacted interaction fluency, and how users' reactive behaviors and recovery actions reflect their motivations for the robot. This approach aligns with our research scope aim to holistically understand in-the-wild interactions, as outlined in Section I.

In the rest of this section, we report our findings under the following three categories: *interaction fluency*, *user's motivation to use the robot*, and *user's positioning towards the robot*. While *interaction fluency* categorizes interaction patterns, *user's motivation* and *user's positioning* are categories that describe user traits. Our findings are also diagrammatically described in Fig. 4.

A. Interaction Fluency

We identified five patterns of interaction flow in terms of fluency: **Smooth** (22.8%), **Awkward** (23.7%), **Active** (19.0%), **Messy** (4.7%), and **Quiet** (29.7%). The percentages in parentheses indicate the proportion of groups categorized under each pattern. For each pattern, we provide illustrative quotes from our descriptions. The quotes are marked with interaction group IDs (*e.g.*, I5 denotes interaction group 5). It is important to note that the genders and ages mentioned in descriptions are based on the researchers' perceptions.

Smooth pattern represents the ideal interaction where everything proceeds as expected and does not include any of the *conflicts* listed in Section III-E. For example, when a user asked for directions, the robot correctly recognized the question and provided appropriate route guidance, as the following example from I79 illustrates:

> A man (M1) and a woman (W1) approached from the left, stopped, and looked at Sota. M1 asked, "Where is the pet supply store?" Sota showed the direction, and after listening, W1 pointed that way, and they walked off to the left.

We also included interactions where users asked non-task-related questions, and the robot responded in a reasonable way. For instance, the following from I11 illustrates *Smooth*:

> A man (M1) and a woman (W1) approached from the left, with M1 waving lightly. W1 asked, "Aren't you lonely?" Sota replied, "No worries, I'm fine! But thank you for asking." W1 cheerfully responded, "That's good!" When Sota invited questions, W1 said, "Good luck!" and Sota replied, "Thank you."

Awkward pattern, in contrast, refers to situations where meaningful conversation was *not* established. This occurred when users encounter *conflicts* and left without recovery actions from the *conflicts* as the example from I52 illustrates:

> A man (M1) stopped and asked Sota for McDonald's location three times, but his questions overlapped with Sota's speech, preventing recognition. M1 looked frustrated, and gave up and walked away.

This type of interaction flow included *conflicts* in the following proportions: 43.6% were *system errors*, 27.3% were *overlapping speech and poor timing*, 18.2% were *incorrect route guidance*, and 9.1% were *misrecognizing and responding to unintended input*. In 70.9% of Awkward interactions, no recovery actions were taken.

Active pattern describes interactions where *conflicts* arose between users and the robot, but the users took recovery actions, ultimately receiving a reasonable response from the robot, for example, I26 had the following interaction:

> A woman (W1) stopped in front of Sota and asked for the bus stop location, but Sota was speaking and didn't respond. [...] After pausing briefly, W1 leaned forward and asked again, prompting Sota to recognize her and provide directions.

88.0% of the conflicts observed in Active pattern involved *overlapping speech and poor timing*. Most of these conflicts
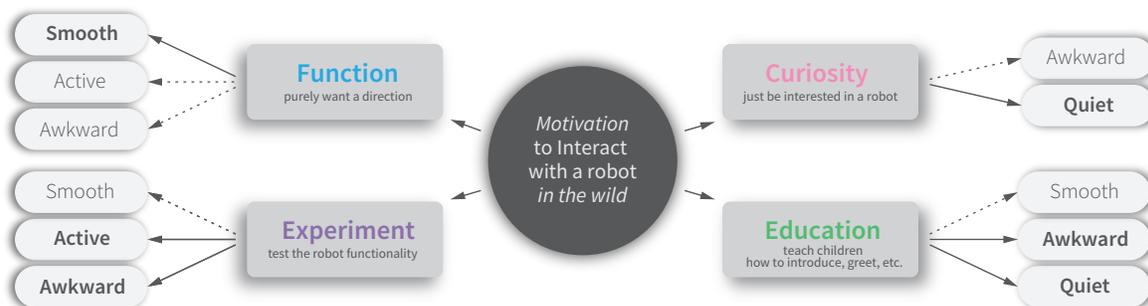
Fig. 4. A diagrammatic model of our findings: We found that user motivation–specifically, why they interacted with the robot–is a key factor in shaping human-robot interaction (HRI) in real-world settings. We identified four types of motivations: Function, Experiment, Curiosity, and Education, and examined how each motivation influences the interaction flow. By analyzing 232 interactions, we uncovered five distinct patterns of interaction flow: Smooth, Active, Awkward, Messy, and Quiet. This diagrammatic model illustrates how each motivation is connected to these interaction flow patterns except for Messy. We observed interactions where multiple motivations were likely present within the same group, often resulting in Messy interactions.

were addressed by either *waiting for the robot to finish speaking* or *speaking to the robot again*.

Quiet pattern refers to cases where users observed the robot without initiating any conversation as the following quote from I98 illustrates:

> *A woman (W1) walking with a child (C1) noticed Sota and pointed it out. C1 stared at Sota with her mouth opened wide, as W1 picked her up for a better view and appeared to whisper something. C1 continued staring with her mouth open.*

Lastly, **Messy** pattern involves inconsistent interactions where conversations were intermittently established and disrupted by multiple users. The following quote (I142) represents the typical interaction flow observed in Messy pattern:

> *Sota announced to two women (W1, W2) and two children (C1, C2). C2 moved to Sota's side, while C1 followed when Sota turned left, continuing to watch. W2 asked, "Where is the restroom?" as C1 moved restlessly and C2 repeatedly said, "Hello, hello," to Sota. As the children wandered, W2 chased after them, and W1 picked up C2 and left.*

In 45.0% of **Messy** interactions, *overlapping speech* was observed as a *conflict*, while four cases involved *system errors*. In six of these interactions, recovery actions included *waiting for the robot to finish speaking* or *speaking to the robot again*.

B. Motivation to Interact with the Robot

We identified four types of user motivations to interact with the robot, namely (1) **Function** (28.4%), (2) **Experimenters** (29.3%), (3) **Curiosity** (36.6%), and (4) **Education** (12.1%). The percentages in parentheses indicate the proportion of groups categorized under each type. Note that there were interaction groups where we recognized multiple motivations.

Function-focused users approached the robot with a clear goal, *i.e.,* asking for directions. In their *Encounter Phase*, these users often approached from the direction of the mall map. While studying the map, they became aware of the robot through its announcement and then began using it. As they focused on achieving their goal, their behavior typically

involved interacting more with the display than with the robot itself, and conversations were minimal in *Interaction Phase*. The most frequent conflict they encountered was *overlapping speech and poor timing*. Their most common recovery action was *speaking to the robot again*, followed by no recovery action, and *waiting for the robot to finish speaking* as the third. In their *Leave Phase*, they waited for *Completion of route guidance* by the robot when they obtained the correct guidance. They also looked around to understand the route suggested by the robot. In our study, 25 out of 65 Function-focused groups experienced Smooth interactions, 17 groups had Active, and 16 groups encountered Awkward.

Experimenters were interested in testing the robot's functionality. In the *Encounter Phase*, commonly observed user behaviors included pointing at the robot and observing it from a distance to understand its actions. In *Interaction Phase*, they often asked for directions following the system's instructions, however, they did not necessarily need the directions. They were interested in how the robot would respond; therefore, they tended to ask several directions or ask challenging questions such as "*Where is my boss? (I83)*" In *Leave Phase*, they typically walked down in the opposite direction to the way the robot announced or interrupted the robot as soon as they confirmed the robot started guiding. Similar to the Function-focused group, the most common *conflict* they faced was *overlapping speech and poor timing*. Their most frequent recovery action was *speaking to the robot again*, followed by taking no recovery action, with *waiting for the robot to finish speaking* as the third most common response. In our study, 23 out of 73 groups of Experimenters resulted in Active interactions, 22 groups faced Awkward, and 19 groups experienced Smooth interactions.

Curiosity-driven users, as the name suggests, were motivated by curiosity. These users focused on the robot itself, engaged in non-goal-oriented, playful conversation instead of asking for directions, while Experimenters were aware of contextual cues (*i.e.,* seeing Sota as a receptionist) and often asked "where is...?" questions to test the robot's function. In their *Encounter Phase*, they typically seemed interested by

running to the robot, pointing at it, gazing at it, or calling their accompanying person to show the robot. In their *Interaction Phase*, more than half of people in this group did not initiate a conversation with the robot at all. They observed the robot for a while and then left without engaging in conversation. In another instance, we observed users showing interest in the robot and attempting to get their accompanying person to interact with it. In the *Interaction Phase* for those who initiated conversations in this group, similar to Experimenters, they did not always need directions but were more interested in the robot itself. For example, they often engaged in non-goal-oriented conversations, saying things like, "*You're so cute* (I19)," "*Good boy* (I32)," or "*What is your name?* (I94)" *Conflicts* they often encountered was the robot *misrecognizing and responding to unintended input*, which was likely due to they often chat about the robot with their accompanying person in their *Interaction Phase*. More than half of this group did not take recovery actions and simply left the robot when a conflict happened. We observed 51 out of 85 groups of Curiosity-driven users resulted in Quiet interactions, 15 groups in Awkward, and nine groups engaged in Smooth interactions.

Finally, **Education**-oriented users approached the robot for educational purposes, often for their children. This group typically consisted of families, such as a pair of an adult and a child, where the adult encouraged or guided their child in engaging with the robot, while Curiosity-driven users approached the robot for their own experience. In the *Encounter Phase*, we frequently observed adults pointing it out to their children. They often saw the child if they were interested in it. Once the child recognized the robot, their behavior commonly included pointing or gazing at it. We also identified several children hiding behind the adult or refusing to approach the robot. In their *Interaction Phase*, adults often took *proxy actions*, teaching children how to greet the robot or encouraging them to introduce themselves, *e.g.,*, in I111, a woman said "*Hello*" in a way that encouraged her child to mimic her. Their *Leave Phase* was typically depended on the child. If the child lost interest in the robot, they began to leave. If the child remained interested, they tended to stay a little longer. We were not able to see a significant tendency in *conflict* types, although we observed the robot failed to distinguish between speech meant for the child and directed to the robot. When encountering *conflicts*, they did not initiate recovery actions in most cases. We identified 12 out of 28 Education-oriented groups resulted in Quiet interactions, eight groups experienced Awkward interactions, seven had Smooth interactions, and one group engaged in Active interaction.

C. Positioning

By observing user behavior across the phases, we identified two major groups based on how users positioned themselves in relation to the robot, which we defined as *positioning*: **Friendly** and **Dry**. The **Friendly** group interacted with the robot by following human-like social norms, and individuals from this group were observed across all four motivation groups, though they were particularly notable in the Exper-

iment group. These users engaged with the robot in a relaxed, informal manner, speaking as if to a close friend. Even during conflicts, they maintained a cheerful tone. For instance, when encountering a conflict, a participant from I10 said, "*Hey, listen to me!*" Similarly, when I88 faced a system error and received an apology from the robot, they responded with laughter, saying, "*You must be tired.*" The **Dry** group, in contrast, viewed the robot not as a sociable but as "a tool." This positioning was commonly observed among Function-motivated users and occasionally among Experimenters. Users in this group displayed minimal social signals toward the robot. Instead, they expressed practical impressions, often sharing remarks with their companions, such as, "*Oh, the world is getting more convenient.*" (I64).

In addition to these positioning, we observed inappropriate behaviors by a few users, which we named **Abuse** group. This group was found in Experiment and Curiosity types of motivation. They showed rude and offensive attitudes to the robot by saying "*You are ball head!* (I6)" or "*Go to hell* (I105)." A person from I233 showed middle fingers to the robot when the robot was not able to respond. Lastly, there was a group of people whose positioning could not be identified. We frequently categorized them under Curiosity or Education in terms of motivations, as they seemed to be interested in their *Encounter Phase*. However, as they did not initiate any conversations or display social signals, leaving us insufficient information to determine their positioning.

V. DISCUSSION

A. Comparison with Controlled Lab Settings

1) Interaction Fluency: User behaviors seen in **Smooth**, **Awkward**, and **Active** interactions are often studied in controlled lab environments [*e.g.,* 3, 48, 50]. On the other hand, **Quiet**–where users may remain passive or refrain from engaging with the robot and **Messy**–where users "go wild" during the interaction are unlikely to be observed in lab settings. In lab studies, participants are typically motivated and guided to interact, making it challenging to observe more subtle or hesitant behaviors like those seen in our field study.

2) User Motivation: **Function**-motivated users are perhaps the closest to population that we want to focus on a controlled lab study. However, unlike lab environments where participants often display polite and cooperative behavior due to experimenter bias, **Function** group in our study interacted in a more straightforward and pragmatic manner, which we describe as Dry positioning. On the other hand, **Experimenters** and **Curiosity**-motivated users, whose behavioral manner is more exploratory, should be the populations showing unique behaviors, which we are able to encounter only in the wild studies. The similar behaviors were also observed in the prior real-world HRI work on testing behaviors [23] and robot bullying [24, 25]. Users with Friendly positioning were often observed within these motivation groups. They tended to display attentive behaviors when encountering *conflicts*, which aligns with prior research suggesting that social robot failures can have positive effects on user perception [58, 50, 29]. Finally,

we were able to observe interesting interaction dynamics, in interactions by **Curiosity**- and **Education**- motivated groups: *proxy actions*. We often saw proxy actions, where one group member–often a parent and a child or accompanying person–was forced to interact with the robot while others observed. Similar results, especially in where family dynamics play a role in child-robot interactions, are observed in several prior works [28, 59]. These patterns are less common in controlled lab settings, where participants are often instructed to engage directly with the robot themselves.

B. Design Implications

In this section, we discuss how user motivations inform design implications and how to minimize failures while creating engaging, motivation-specific interactions.

Function-motivated users approached the robot with a clear goal, relying on the display to quickly understand the system, leading to the highest rate of Smooth and Active interactions. For these users, interaction design can be minimal, focusing on essential information such as internal states, processing steps, and error types [*e.g.,* 42, 43]. As their behavior is minimal, distinguishing them from other groups (*e.g.,* Experimenters) during the *Encounter Phase* can be challenging. Contextual information, like the robot's location, is essential for identifying Function-motivated users. For instance, a robot near a mall map could anticipate that a user is seeking directions.

Experimenters had moderate success obtaining directions but faced more Awkward moments than the Function-focused group due to their exploratory behavior. Their goal was to test the robot's capabilities (*i.e.*, what it can and cannot do) rather than its internal processes. Since they often test unexpected functions, interactions should include flexible behaviors or improvised responses, which LLMs may excel at [60]. Based on our findings, they tend to take a moment to observe the robot to learn how to use the system before approaching. We may able to use this moment to anticipate Experimenters.

Curiosity-driven group engaged less in verbal interaction, though half displayed non-verbal social signals. These users were often passive observers, showing interest through gestures like pointing or running during the *Encounter Phase*. To encourage engagement, the robot could use reactive interactions, such as waving or dancing, as suggested by prior HRI research emphasizing non-instrumental functions to enhance user experience [61]. Additionally, demonstrating interactions with others nearby could satisfy their curiosity and help them remember how to use the robot in the future.

Education-oriented group showed mixed interaction fluency, including Active, Awkward, and Quiet interactions. While no clear conflict tendencies were identified, proxy actions–like an adult pointing out the robot to a child during their *Encounter Phase*–emerged as a key design principle. When such behaviors are detected, the robot should focus on engaging with the child, taking proactive steps to involve them in the interaction.

In the wild, user motivations vary, which requires tailored robot behaviors for each motivation group. Understanding user motivation is essential to prevent failures and enhance interaction fluency. Prioritizing specific motivation groups is important for effective and sustainable HRI design, as designing for all groups may not be feasible. For HRI researchers, especially in the wild, understanding the distribution of motivation groups is also crucial for meaningful outcomes.

C. Limitation and Future Work

Our work has a number of limitations. First, we based our descriptions on "perceived gender and age," as we did not collect biographical information. While we made careful judgments through discussions, the statistics based on these perceptions may not be entirely accurate and it does not reflect the users' actual identities. Second, our findings may have been influenced by the robot's appearance. We believe the Function group, often classified as the Dry positioning, was less impacted by the robot's appearance, as their focus was on "using" the robot's functions. In contrast, other positioning and motivation groups might respond to the robot's cute and toy-like appearance, encouraging playful or even teasing behaviors. We suspect such reactions may differ when a different robot embodiment is used, as a prior work suggests robot's appearance provides a "locus of attention" for human-robot interaction and has impact on social outcomes [62], and the robot form significant impacts user perceptions of the robot [63]. We need further exploration with different robots to extend our findings. Third, the setting of a shopping mall may have attracted users with more leisure time or a greater willingness to engage with the robot. These people might be more open to exploration, social interaction, or casual experimentation, which could have influenced their motivation for the interaction. To broaden the scope of our results, future research should be conducted in a variety of real-world settings. Finally, we acknowledge potential cross-cultural differences that may influence our findings. Since our study was conducted in Japan and analyzed in Japanese, cultural factors could have played a role [64]. Japanese social norms often encourage reserved and quiet behavior in public spaces [65], which may lead people to interact with robots in a more restrained manner. Individuals from different cultures may approach the robot in different ways. Such cultural variations can impact the interaction dynamics. Future work should incorporate cross-cultural examinations to further understand the role of motivation for HRI in the wild.

VI. CONCLUSION

This paper aims to understand how people approach, interact, and leave a robot in the wild. To conduct field study, we deployed an autonomous conversational robot in a shopping mall. We engaged in video analysis and identified interaction fluency, user motivation types to interact with the robot, and user positioning towards the robot. Our findings suggest user's motivation can be a key factor to designing HRI in the wild.

REFERENCES

[1] S. Honig and T. Oron-Gilad, "Understanding and resolving failures in human-robot interaction: Literature review and model development," *Frontiers in Psychology*, vol. 9, 2018.

[2] L. Tian and S. Oviatt, "A taxonomy of social errors in human-robot interaction," *J. Hum.-Robot Interact.*, vol. 10, no. 2, Feb. 2021.

[3] S. Tolmeijer, A. Weiss, M. Hanheide, F. Lindner, T. M. Powers, C. Dixon, and M. L. Tielman, "Taxonomy of trust-relevant failures and mitigation strategies," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 3–12.

[4] D. Y. Geiskkovitch and J. E. Young, "Social robots don't do that: Exploring robot-typical errors in child-robot interaction," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 200–202.

[5] M. Iwasaki, J. Zhou, M. Ikeda, T. Kawamura, and H. Nakanishi, "A customer's attitude to a robotic salesperson depends on their initial interaction," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2018, pp. 300–305.

[6] J. Xu and A. Howard, "The impact of first impressions on human-robot trust during problem-solving scenarios," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE Press, 2018, p. 435–441.

[7] B. Petrak, K. Weitz, I. Aslan, and E. Andre, "Let me show you your new home: Studying the effect of proxemic-awareness of robots on users' first impressions," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE Press, 2019, p. 1–7.

[8] C. P. Lee, B. Cagiltay, and B. Mutlu, "The unboxing experience: Exploration and design of initial interactions between children and social robots," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, ser. CHI '22. New York, NY, USA: Association for Computing Machinery, 2022.

[9] M. K. Lee, S. Kiesler, J. Forlizzi, and P. Rybski, "Ripple effects of an embedded social agent: a field study of a social robot in the workplace," in *SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 695–704.

[10] K. Tanaka, N. Yamashita, H. Nakanishi, and H. Ishiguro, "Teleoperated or autonomous?: How to produce a robot operator's pseudo presence in hri," in *ACM/IEEE International Conference on Human Robot Interaction*, 2016, pp. 133–140.

[11] Y. Okafuji, Y. Ozaki, J. Baba, J. Nakanishi, K. Ogawa, Y. Yoshikawa, and H. Ishiguro, "Behavioral assessment of a humanoid robot when attracting pedestrians in a mall," *International Journal of Social Robotics*, vol. 14, pp. 1731–1747, 2022.

[12] S. Sabanovic, M. Michalowski, and R. Simmons, "Robots in the wild: observing human-robot social interaction outside the lab," in *9th IEEE International Workshop on Advanced Motion Control, 2006.*, 2006, pp. 596–601.

[13] T. Salter, I. Werry, and F. Michaud, "Going into the wild in child–robot interaction studies: issues in social robotic development," *Intelligent Service Robotics*, vol. 1, no. 2, pp. 93–108, 2008.

[14] J. L. Burke, H. I. Christensen, R. Menassa, R. Koeppe, J. Dyer, and M. Munich, "Hri: the real world," in *Proceedings of the 6th International Conference on Human-Robot Interaction*, ser. HRI '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 103–104.

[15] S. Sabanović, S. M. Reeder, and B. Kechavarzi, "Designing robots in the wild: in situ prototype evaluation for a break management robot," *J. Hum.-Robot Interact.*, vol. 3, no. 1, p. 70–88, Feb. 2014.

[16] A. Rosenthal-von der Pütten, A. Weiss, and S. Sabanovic, "Hri workshop: The challenge (not) to go wild! challenges and best practices to study hri in natural interaction settings," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2016, pp. 583–584.

[17] M. Jung and P. Hinds, "Robots in the wild: A time for more robust theories of human-robot interaction," pp. 1–5, 2018.

[18] *HRI '24: Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: Association for Computing Machinery, 2024.

[19] J. M. Innes and B. W. Morrison, "Experimental studies of human-robot interaction: Threats to valid interpretation from methodological constraints associated with experimental manipulations," *International Journal of Social Robotics*, vol. 13, pp. 765–773, 2021.

[20] Y. Malhotra and D. Galletta, "Extending the technology acceptance model to account for social influence: theoretical bases and empirical validation," in *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, vol. Track1, 1999, pp. 14 pp.–.

[21] A. G. Cass, K. Striegnitz, N. Webb, and V. Yu, "Exposing real-world challenges using hri in the wild," in *Proc. of the 4th Workshop on Public Space Human-Robot Interaction at the Intl. Conf. on Human-Computer Interaction with Mobile Devices and Services*, 2018.

[22] Y. Okafuji, S. Song, J. Baba, Y. Yoshikawa, and H. Ishiguro, "Influence of collaborative customer service by service robots and clerks in bakery stores," *Frontiers in Robotics and AI*, vol. 10, 2023.

[23] A. M. Rosenthal-von der Pütten, N. C. Krämer, C. Becker-Asano, K. Ogawa, S. Nishio, and H. Ishiguro, "The uncanny in the wild. analysis of unscripted human–android interaction in the field," *International Journal of Social Robotics*, vol. 6, pp. 67–83, 2014.

[24] R. Agrigoroaie, S.-D. Ciocirlan, and A. Tapus, "In the wild hri scenario: influence of regulatory focus theory," *Frontiers in Robotics and AI*, vol. 7, p. 58, 2020.

[25] D. Brscić, H. Kidokoro, Y. Suehiro, and T. Kanda, "Escaping from children's abuse of social robots," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 59–66.

[26] B. Brown, F. Bu, I. Mandel, and W. Ju, "Trash in motion: Emergent interactions with a robotic trashcan," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24. New York, NY, USA: Association for Computing Machinery, 2024.

[27] P. Salvini, G. Ciaravella, W. Yu, G. Ferri, A. Manzi, B. Mazzolai, C. Laschi, S. Oh, and P. Dario, "How safe are service robots in urban environments? bullying a robot," in *19th International Symposium in Robot and Human Interactive Communication*, 2010, pp. 1–7.

[28] J. E. Michaelis, B. Cagiltay, R. Ibtasar, and B. Mutlu, ""off script:" design opportunities emerging from long-term social robot interactions in-the-wild," in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 378–387.

[29] M. Ragni, A. Rudenko, B. Kuhnert, and K. O. Arras, "Errare humanum est: Erroneous robots in human-robot interaction," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2016, pp. 501–506.

[30] B. Nesset, G. Rajendran, J. D. Aguas Lopes, and H. Hastie, "Sensitivity of trust scales in the face of errors," in *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '22. IEEE Press, 2022, p. 950–954.

[31] D. Kontogiorgos, A. Pereira, B. Sahindal, S. van Waveren, and J. Gustafson, "Behavioural responses to robot conversational failures," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 53–62.

[32] M. Desai, P. Kaniarasu, M. Medvedev, A. Steinfeld, and H. Yanco, "Impact of robot failures and feedback on real-time trust," in *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '13. IEEE Press, 2013, p. 251–258.

[33] X. Chang, Y. Li, S. Liu, L. Ma, and R. Lc, ""sorry to keep you waiting": Recovering from negative consequences resulting from service robot unintended rejection," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 96–105.

[34] D. Kontogiorgos, M. Tran, J. Gustafson, and M. Soleymani, "A systematic cross-corpus analysis of human reactions to robot conversational failures," in *Proceedings of the 2021 International Conference on Multimodal Interaction*, ser. ICMI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 112–120. [Online]. Available: https://doi.org/10.1145/3462244.3479887

[35] S. Reig, E. J. Carter, T. Fong, J. Forlizzi, and A. Steinfeld, "Flailing, hailing, prevailing: Perceptions of multi-robot failure recovery strategies," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 158–167.

[36] H. N. Green, M. M. Islam, S. Ali, and T. Iqbal, "Who's laughing nao? examining perceptions of failure in a humorous robot partner,"

in *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '22. IEEE Press, 2022, p. 313–322.

[37] J. M. Kraus, J. Merger, F. Gröner, and J. Pätz, "'sorry' says the robot: The tendency to anthropomorphize and technology affinity affect trust in repair strategies after error," in *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 436–441.

[38] L. Hoffmann, M. Derksen, and S. Kopp, "What a pity, pepper! how warmth in robots' language impacts reactions to errors during a collaborative task," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 245–247.

[39] S. Engelhardt, E. Hansson, and I. Leite, "Better faulty than sorry: Investigating social recovery strategies to minimize the impact of failure in human-robot interaction." in *WCIHAI@ IVA*, 2017, pp. 19–27.

[40] M. K. Lee, S. Kiesler, J. Forlizzi, S. Srinivasa, and P. Rybski, "Gracefully mitigating breakdowns in robotic services," in *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2010, pp. 203–210.

[41] M. Axelsson, M. Spitale, and H. Gunes, ""oh, sorry, i think i interrupted you": Designing repair strategies for robotic longitudinal well-being coaching," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 13–22. [Online]. Available: https://doi.org/10.1145/3610977.3634948

[42] D. Das, S. Banerjee, and S. Chernova, "Explainable ai for robot failures: Generating explanations that improve user assistance in fault recovery," in *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 351–360.

[43] G. LeMasurier, A. Gautam, Z. Han, J. W. Crandall, and H. A. Yanco, "Reactive or proactive? how robots should explain failures," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 413–422.

[44] J. Ravishankar, M. Doering, and T. Kanda, "Zero-shot learning to enable error awareness in data-driven hri," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 592–601.

[45] M. Stiber, R. H. Taylor, and C.-M. Huang, "On using social signals to enable flexible error-aware hri," in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 222–230.

[46] S. van Waveren, "Leveraging non-experts and formal methods to automatically correct robot failures," in *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '22. IEEE Press, 2022, p. 1182–1184.

[47] M. Spitale, M. T. Parreira, M. Stiber, M. Axelsson, N. Kara, G. Kankariya, C.-M. Huang, M. Jung, W. Ju, and H. Gunes, "Err@hri 2024 challenge: Multimodal detection of errors and failures in human-robot interactions," in *Proceedings of the 26th International Conference on Multimodal Interaction*, ser. ICMI '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 652–656. [Online]. Available: https://doi.org/10.1145/3678957.3689030

[48] S. Choi, A. S. Mattila, and L. E. Bolton, "To err is human(-oid): How do consumers react to robot service failure and recovery?" *Journal of Service Research*, vol. 24, no. 3, pp. 354–371, 2021.

[49] X. Xing, M. Song, Y. Duan, and J. Mou, "Effects of different service failure types and recovery strategies on the consumer response mechanism of chatbots," *Technology in Society*, vol. 70, p. 102049, 2022.

[50] W. Kamino, N. Randall, T. Saga, L.-J. Hsu, K. M. Tsui, S. Sabanović, and S. Nagata, "We all make mistakes: Terminal, non-critical, recoverable, and favorable interaction failures between people and a social robot," in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2023, pp. 1028–1033.

[51] S. Song, J. Baba, J. Nakanishi, Y. Yoshikawa, and H. Ishiguro, "Costume vs. wizard of oz vs. telepresence: How social presence forms of tele-operated robots influence customer behavior," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2022, pp. 521–529.

[52] J. Baba, S. Song, J. Nakanishi, Y. Yoshikawa, and H. Ishiguro, "Local vs. avatar robot: Performance and perceived workload of service encounters in public space," *Frontiers in Robotics and AI*, vol. 8, 2021.

[53] A. Kendall, M. Grimes, and R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[54] OpenAI, "Hello gpt-4o," 2023, accessed: 2024-09-23. [Online]. Available: https://openai.com/index/hello-gpt-4o/

[55] HOYA, "Voicetext web api," 2020. [Online]. Available: https://cloud.voicetext.jp/webapi

[56] C. Geertz, "Thick description: Toward an interpretive theory of culture," in *The cultural geography reader*, 2008, pp. 41–51.

[57] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[58] K. Harrison, G. Perugia, F. Correia, K. Somasundaram, S. van Waveren, A. Paiva, and A. Loutfi, "The imperfectly relatable robot: An interdisciplinary workshop on the role of failure in hri," in *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 917–919.

[59] H.-R. Ho, E. M. Hubbard, and B. Mutlu, ""it's not a replacement:" enabling parent-robot collaboration to support in-home learning experiences of young children," in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24. New York, NY, USA: Association for Computing Machinery, 2024.

[60] C. Y. Kim, C. P. Lee, and B. Mutlu, "Understanding large-language model (llm)-powered human-robot interaction," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 371–380.

[61] Y. Chamoto, Y. Okafuji, K. Matsumura, J. Baba, and J. Nakanishi, "Investigating the influence of task-dependent and task-independent robot behavior on the impression of robots and the user experience," in *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2023, pp. 640–646.

[62] R. McDonnell and B. Mutlu, "Appearance," *The handbook on socially interactive agents: 20 years of research on embodied conversational agents, intelligent virtual agents, and social robotics volume 1: Methods, behavior, cognition*, pp. 105–146, 2021.

[63] M. Spitale, M. Axelsson, and H. Gunes, "Robotic mental well-being coaches for the workplace: An in-the-wild study on form," in *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 301–310. [Online]. Available: https://doi.org/10.1145/3568162.3577003

[64] M. Salem, M. Ziadee, and M. Sakr, "Marhaba, how may i help you? effects of politeness and culture on robot acceptance and anthropomorphization," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 74–81.

[65] O. Ikeno and R. J. Davies, Eds., *The Japanese Mind: Understanding Contemporary Culture*. Tuttle Publishing, 2002.