Context is Cue-cial: Assessing the Interpretation of Social Signals from Non-Anthropomorphic Robots in Different Contexts*

Aparajita Chowdhury[1][0000-0003-1663-1774], Ana Carrasco[2][0000-0002-2506-8605], Florian Müller[3][0000-0002-9621-6214], Aino Ahtinen[1][0000-0002-6886-3756], Kaisa Väänänen[1][0000-0002-3565-6021], Albrecht Schmidt[4][0000-0003-3890-1990], and Jan Leusmann[4][0000-0001-9700-5868]

[1] Computing Sciences, Tampere University, Tampere, Finland
[2] Instituto Superior Técnico, INESC-ID, Lisbon, Portugal
[3] TU Darmstadt, Darmstadt, Germany
[4] LMU Munich, Munich, Germany

Abstract. Humans excel at understanding social cues in communication, but robots struggle. Social cues are crucial for humans to interpret the intentions of their communication partners. Research indicates that we typically interpret the actions of anthropomorphic robots analogously to their human counterparts, paving a clear path to the design of appropriate social cues. For non-anthropomorphic robots, however, it is an open question how humans interpret social cues with different output modalities and in different contexts. Our study investigates whether social cues signaled by typical non-anthropomorphic modalities such as lights, sounds, and gestures are consistently interpreted across people and contexts. We, therefore, conducted a contextual investigation in a hospital, derived scenarios from co-design workshop, and tested 103 cues collected from the literature in a large online survey (N=1545). Our results demonstrate that most human interpretations vary by context, highlighting the need to design dynamic and adaptive social cues for interactive robotic systems.

Keywords: human computer interaction, human-robot interaction, user-adaptive interaction, social signals

1 Introduction

From a joyful smile when meeting a friend to a sarcastic one in awkward moments, non-verbal social cues are complex and highly context-dependent. These

* This is the author's accepted manuscript. Published as: Chowdhury, A. et al. (2026). Context Is Cue-Cial: Assessing the Interpretation of Social Signals from Non-anthropomorphic Robots in Different Contexts. In: *Social Robotics + AI. ICSR+AI 2025.* Lecture Notes in Computer Science, vol. 16131. Springer, Singapore. https://doi.org/10.1007/978-981-95-2379-5_21

cues, such as facial expressions, body language, and eye contact, help humans communicate intent without words [39]. As robots become more integrated into everyday life, interactions with them will increasingly involve non-verbal cues in various contexts [11,38]. However, current robots, especially non-anthropomorphic ones, struggle to adapt such cues to different contexts, making their intentions unclear.

While anthropomorphic robots benefit from human-like forms that support intuitive cue design, non-anthropomorphic robots, like delivery bots or robotic arms, are built for efficiency and typically lack features that facilitate human-like communication [6]. As these robots take on more collaborative or social roles, it is essential they convey intent through alternative means such as lights, movements, or sounds [26]. Yet, since these modalities are not familiar channels of social interaction for humans, it is common to misinterpret them. Prior research has explored non-verbal cues for non-anthropomorphic robots in contexts such as manufacturing, rescue, and logistics [12,13,36,43]. Yet, it is unclear how well these cues transfer across settings. This highlights a key gap in designing context-aware, intuitive social signals for non-anthropomorphic robots to support effective human-robot interaction. To address these limitations, we investigated how humans perceive social signals from non-anthropomorphic robots in different contexts. We chose a hospital as an example of a (semi-)public setting in which users might encounter non-anthropomorphic robots in the future, and extracted a set of distinct contexts. While we explored these in the hospital setting, we consider them generic enough to generalize to other settings. Our goal is to understand how humans perceive the social signals conveyed by a robot through lights, sounds, and gestures in various contexts, especially when they have not been previously trained or informed about the meaning of those signals. For our exploration, we chose the hospital setting, as it offers a wide range of scenarios involving people from diverse backgrounds and varying levels of knowledge. Additionally, it provides opportunities to study interactions not only with logistics personnel but also with visitors, nurses, and patients. To guide our investigation, we formulated the following research questions:

RQ1: What interaction scenarios can be developed for different contexts inside the hospital setting?

RQ2: What interaction modalities and social signals are appropriate for a non-anthropomorphic robot to effectively express its intent?

RQ3: How do different contexts influence the perception of social signals expressed by the robot within the hospital setting?

To address RQ1, we conducted a contextual inquiry in a Finnish hospital to explore human-robot interactions with a logistics robot in various contexts within the hospital. Following this, we organized a workshop with 15 students from a Finnish university to refine the insights gathered and to define five scenarios that could be presented to participants to assess the perception of social cues. The interaction scenarios were developed based on the data gathered from contextual inquiry at the hospital and the co-design workshop with the students. To answer RQ2, we asked the same workshop participants to redesign the logistics robot using modalities they believed would be suitable for this context. This exploration resulted in the selection of lights, sounds, and robotic arm gestures as means to express the robot's intent and social cues. Based on these designed interaction modalities, we reviewed 189 peer-reviewed articles to identify the social signals conveyed through light, sound, and gesture. This review yielded 103 signals, which we then recreated according to guidelines found in the literature. Finally, for RQ3, we conducted an online study with 1,545 participants to assess the perception of these social signals in various hospital contexts. Our findings indicate that there is no universal understanding of social signals for non-anthropomorphic robots. These signals are interpreted differently depending on the context, and their meanings may shift when adapted for different settings. This highlights the need for a dynamic and context-sensitive design of social cues tailored to specific interaction environments.

2 Related Work

In this section, we reflect on key topics relevant to this research. Section 2.1 explores the importance of social cues and signals for humans across different contexts. Section 2.2 addresses the challenges of designing social signals for non-anthropomorphic robots. Finally, Section 2.3 reviews existing knowledge on social signals for non-anthropomorphic robots in various contexts.

2.1 Social Cues and Signals for Humans across Different Contexts

Charles Darwin argued that social cues reflect our internal states, facilitating interaction and cooperation [10]. Social signals, verbal or non-verbal cues guiding interactions, help shape our impressions of others [19]. From early childhood, humans learn to interpret and respond to these cues through experience and context [4,27]. For example, a genuine smile may express joy, while a polite smile can mask disappointment to maintain social harmony. Social signals—primarily non-verbal expressions like facial cues, body language, and gaze—are central to communication, trust-building, and emotional understanding [48]. Their interpretation depends on cultural norms, relationships, and situational context [5,19]. Through observing trusted individuals and shared social knowledge, humans develop an intuitive grasp of how to convey and interpret these signals appropriately across different settings [16]. Our lifelong experiences and cognitive abilities enable us to adapt social cues to specific contexts. This skill has been crucial to human evolution, supporting survival, cooperation, and effective communication [10,16].

2.2 Designing Social Cues for Robots

The use of appropriate social signals is essential for both humans and robots, not only to convey intent but also to build trust and influence user response [8,48,45]. However, robots lack the innate human ability to adapt social cues to different contexts, which can lead to user discomfort. For instance, in [35], Pepper's prolonged eye contact was meant to boost engagement but felt intrusive in quiet settings like libraries. Similarly, Baxter's negative head shakes were perceived as rude on assembly lines [2]. Both robots were eventually discontinued due to poor user acceptance [46], highlighting how misaligned social cues can increase mistrust [47]. Non-anthropomorphic robots face added difficulty in expressing social signals, as their design prioritizes function and cost over human-like features [6]. While efforts to incorporate social cues into such robots exist, outcomes vary. Baxter, for example, was praised in assistive roles but criticized in industrial contexts for being slow and overly human-like [2,46]. Its failure, along with that of its successor, Sawyer, underscores the need for context-appropriate cues in robot design [30]. Although anthropomorphic features aid communication [7,32], integrating them into task-focused robots remains impractical due to cost and performance trade-offs [6].

2.3 Social Cues and Modalities for Non-Anthropomorphic Robots in Various Contexts

As non-anthropomorphic robots enter public spaces, it is crucial they use appropriate social cues to communicate their intentions, especially in settings like healthcare, where robots assist with tasks such as delivering equipment and supporting well-being [24,29,34]. Many hospital staff and visitors may have limited experience with robots, so these robots must express their intentions in ways that are understandable and acceptable [22]. Robots can use lights [3,44], audio signals [33], and gestures [41] to communicate effectively. Studies have explored light as a communication tool for robots, such as using different colors and animations to express actions like "wait," "progress," and "help" [3]. Similarly, audio signals have been used to convey emotions and intent, with examples like R2-D2's and Wall-E's sound design [17]. However, there is limited guidance on replicating these audio signals in real-world robots [20]. Robot arm gestures have also been studied, especially in collaborative tasks, with research exploring how movements can convey emotions and intent [41,18]. These gestures, inspired by human or animal movements, are used to communicate social signals like directions and commands [14,15]. However, most studies have focused on specific contexts, and it remains unclear how gestures like "stop" would be interpreted in different settings, such as a hospital during medicine delivery. Although research has explored non-anthropomorphic social signals using light, sounds, and gestures, there is a gap in understanding how these signals are perceived across various contexts.

3 Methodology

To address our research questions, we employed a two-phase methodological approach consisting of a pre-study and an online user study. The following sections provide a detailed account of each phase.

3.1 Pre-Study

We conducted a multiphase study to explore human-robot interaction in a hospital setting with a medicine delivery robot. In the first phase, we observed the robot during its daily routes through corridors, elevators, and reception areas, documenting interactions with logistics workers, nurses, receptionists, and visitors. Two logistics workers participated in short unstructured interviews, and two members of the robot's software team supported us during the study. These observations helped us understand the robot's tasks, stakeholders, navigation, and goals.

Next, we organized a co-design workshop with 15 Master's students (7 female, 8 male) from a human-robot interaction course, selected due to prior experience designing for social robots and familiarity with the hospital as patients or visitors. Recruiting hospital staff was not feasible due to their busy schedules. This phase was conducted to analyze the findings from the contextual inquiry and to develop scenarios and define effective communication modalities for the robot. Since the insights gathered from the hospital contextual inquiry were limited (including four participants altogether), the co-design workshop played a crucial role in helping us explore potential scenarios in greater detail. It enabled us to brainstorm what situations might arise and what the robot could communicate to different people in the hospital, such as staff, patients, and visitors. The workshop began with a 30-minute presentation of the observation findings and robot operation. Participants worked in small groups to develop five hospital interaction scenarios, discussing the robot's tasks, human interaction strategies, and expression techniques. Data from notes, audio recordings, and design canvases were thematically analyzed, revealing lights, sounds, and gestures as suitable expressive modalities. The five distinct hospital interaction scenarios were defined as:

Scenario 1 Loading Medicine: The robot starts its day when one of its human colleagues load medicines into the robot and asks it to go to deliver it to one of the hospital wings.

Scenario 2 Obstacle Encounter: The robot starts its journey and encounters closed door.

Scenario 3 Passing Through: As the robot goes to deliver medicine, it needs to pass through a lobby where the robot meets human colleagues in a logistics vehicle.

Scenario 4 Elevator Interaction: The robot needs to take the elevator to reach its destination. Human colleagues can share the elevator with the robot.

Scenario 5 Delivering Medicine: The robot reached its destination and needs to inform someone that the robot is here to deliver.

We also created visualizations for the scenario to support the clarity and understandability of the interactions (see Figure 1.

Based on the workshop findings, the participants identified three main communication modalities: lights, sounds, and robot arm gestures as the most suitable to convey the robot's intent, while screens were considered less effective in large hospital areas. Based on these insights, we developed a prototype of the

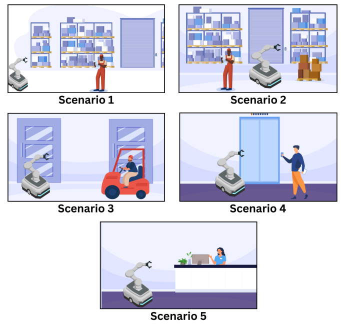

Fig. 1: Hospital scenarios resulted from the pre-study

robot that incorporates the selected modalities (see Fig. 2). These findings highlight the importance of simple understandable non-verbal cues to ensure robot safe and effective integration into hospital workflows. Ethical approval was obtained for both the hospital observation and the workshop, with all data handling conducted in compliance with EU GDPR regulations.

3.2 Online User Study

The observation and co-design workshop provided an initial understanding of interaction modalities, techniques, and social signals. To achieve better results from data analysis of the user study, we further explored social signals from previous literature to identify those that conveyed specific intentions in the context of non-anthropomorphic robots. We then recreated these signals for our delivery robot prototype and conducted an online study to examine how different contexts influence the perception of social signals expressed by the robot (RQ3). The subsequent sections are organized as follows: *Robot Prototype*, *Participants*, *Procedure*, *Data Collection and Analysis*

Robot Prototype . We developed a prototype delivery robot to recreate expressions identified from the literature. For the mobile base, we used the Elephant Robotics MyAGV 2023 PI [5]. On top of it, we mounted a MyCobot 280 M5Stack robotic arm with a gripper from Elephant Robotics[6], which was used to perform arm gestures for expressing social cues. For light-based gestures, we attached an 8×8 Neopixel LED matrix housed in a 3D-printed case at the front of the robot (see Figure 2). Audio expressions were generated using external USB speakers. All expressions were recorded in a well-lit room with a white background to ensure consistency.

[5] https://shop.elephantrobotics.com/products/myagv-2023-pi
[6] https://www.elephantrobotics.com/en/mycobot-280-m5-2023-en/

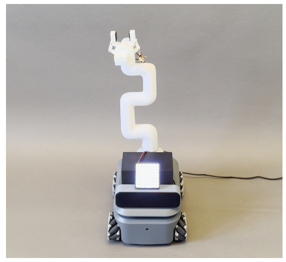

Fig. 2: Robot Setup used for recreating the expressions from literature.

Participants. We recruited 2,193 participants for the online study via Prolific, aiming for three participants per unique combination of SCENARIO (5) × EXPRESSION (103). After excluding 648 low-effort responses, identified through the text summaries participants were required to provide, we were left with 1,545 valid participants (784 female, 725 male, 36 non-binary). The participants' average age was 32.32 years ($SD = 10.41$), representing 86 nationalities. The largest groups came from South Africa (229), the UK (169), the United States (155), Poland (117), Canada (87), Portugal (84), Mexico (68), Zimbabwe (55), and Spain (43). 353 participants had never interacted with a robot, 909 reported between one and seven interactions, and 283 had more than seven interactions, mostly with social robots or robotic arms.

Procedure. Based on the workshop findings, we collected 189 expressions for lights, sounds, and gestures from the literature. After excluding expressions that lacked sufficient detail or could not be replicated on our robot, we finalized a set of 103 expressions: 31 gestures, 32 light cues, and 40 audio cues. Gesture selection followed [42], light cues were drawn from 119 Web of Science articles (filtered for clear color/pattern descriptions), and audio expressions were based on [51], using only those with publicly available files. To ensure consistency in the recordings, we standardized all expressions. For gestures, we controlled arm speed, face/trunk position, and direction. For lights, we fixed speed, pattern type (e.g., blinking, sweeping), and RGB color. For audio expressions, we used original files, maintained a consistent volume, and limited their duration to 35 seconds. All videos are available in the supplementary material. The study was conducted online. After providing informed consent, participants answered two demographic questions about their prior experiences with robots and the types of robots they had encountered. Each was assigned a scenario and condition, shown an illustration and a simplified scenario description (based on pre-study findings). Instructions clarified that the robot communicates non-verbally. Participants then watched a video of one expression and provided two open-ended responses: (1) what the robot did, and (2) in one word, what the expression meant in context. These interpretations were analyzed as detailed in section 4.

Data Analysis. Four researchers initially developed a codebook to ensure triangulation, after which two researchers collaboratively coded the data to ensure reliability and consistency. In the first iteration, coding 1,545 descriptions resulted in 47 unique codes, covering behaviors (e.g., "robot is performing a task," "robot opens elevator/door," "processing"), requests (e.g., "come here," "asking human to do something," "needs help"), intentions (e.g., "robot wants to pick something up," "wants to get in the elevator," "showing information"), and emotions (e.g., "curiosity," "happy," "confused"). A second iteration refined this to 35 unique codes. Finally, we discussed these results among four researchers and categorized these labels into 10 distinct code groups for a final analysis:

Task-Oriented Communication: *Intent* (to communicate intended actions, e.g., "robot delivering package", "robot pick up", "let human pass", "giving permission".), *Request* (asking human for help or to do something, e.g., "asking to pass/move aside", "robot asks to open elevator/door", etc.), and *Status* (displaying robot's internal states - "error", "ready/waiting", "robot is performing a task", etc.).

Social Interaction: *Emotional State* (e.g, "happy", "amazed", "confused", "distress"), *Politeness* (e.g., "greeting", "deference", "showing gratitude", "appreciation"), and *Interest* (showing enthusiasm or curiosity).

Directive Communication: *Attention* (to direct human's focus, e.g., "call for attention", "warning", "alert"), *Agreement* (acceptance, consent, approval, acknowledgment) and *Disagreement* (refusing or negating something).

All samples where participants' interpretations could not be understood were classified as *None*. Details about this coding process are provided in the supplementary material.

Metrics. We measure the level of agreement of an expression using a pairwise similarity metric that checks how many interpretations of the 3 participants match for each scenario-expression combination. We call this metric the Agreeability Score (AS):

$$AS = \frac{1}{N} \sum_{i \neq j}^{N} (E_i = E_j), \qquad 1 \leq i < j \leq N \tag{1}$$

where E_i and E_j are the coded interpretations of the participants i and j, and N is the number of participants per scenario-expression combination (in our case, $N = 3$). $E_i = E_j$ takes the value of 1 if the codes are equal, and 0 otherwise. AS can be either 0 for no agreement, 1/3 for partial agreement and 1 for a perfect agreement.
Additionally, to analyze the impact of the scenario and the modality of the expression in the participants' interpretations, we fitted multinomial logistic regression models to predict code groups, with scenario and modality as independent variables. We looked at Likelihood ratios and calculated the Nagelkerke pseudo-R^2 =[37] as a metric of these models' explanatory power.

Table 1: Updated agreement scores for each scenario and modality, including both average (M) and the percentage of perfect agreement (P.A.) values ($AS = 1$).

Scenario	**Arm Gestures**		**Light**		**Audio**		**All Modalities**	
	M	**P.A.**	**M**	**P.A.**	**M**	**P.A.**	**M**	**P.A.**
Loading Medicine	33.33	19.35	34.38	21.88	23.33	10.00	30.35	17.08
Obstacle Encounter	26.88	12.90	35.42	18.75	29.17	12.50	30.49	14.72
Passing Through	30.11	19.35	42.71	28.12	25.83	7.50	32.88	18.32
Elevator Interaction	41.94	25.81	25.00	12.50	21.67	7.50	29.54	15.27
Delivering Medicine	32.26	16.13	39.58	28.12	26.67	10.00	32.84	18.08
All Scenarios	32.90	18.71	35.42	21.88	25.33	9.50		

4 Results

4.1 Agreeability Analysis

Applying the Agreement Score (Equation 1) on the coded participant interpretations, we found that, out of all the 103 expressions, across the 5 scenarios ($103 \times 5 = 515$ total combinations), 226(43.88%) of them obtained a partial agreement (2 out of 3 equal, $AS = 0.33$) and 83(16.12%) of them obtained a perfect agreement ($AS = 1$). As the results from Table 1 suggest, audio gestures are generally less often agreed on than arm or light gestures. Furthermore, the agreeability is highest for the passing through and delivering medicine scenarios.

When looking at perfect agreement (Table 1), we see a slight difference between scenarios (highest for the passing through scenario - 18.32%), and a bigger difference between modalities (highest for lights - 21.88%). Additionally, for expressions that had a perfect agreement across multiple scenarios (17 out of 103 - 16.5%), we found that only 4 (3.9%) had a perfect agreement between the scenarios (meaning, all the participants, across the 5 scenarios, agreed on the interpretation of that expression). These results suggest that both scenario and modality have an impact on how participants interpret the robot's expressions.

Table 1 shows the average AS and percentage of perfect agreement for each scenario $\times$ modality combination. Furthermore, Figure 3 shows the distribution of agreed-on code groups for each scenario. It shows that for all scenarios, around 40% of the expressions had an $AS = 0$. For the loading medicine scenario, the majority of expression has been agreed on as *Status* (30%), for the obstacle encounter scenario *Request* (40%), for passing through *Attention* (20%), elevator interaction *Request* (23%), and delivering medicine *Attention* (21%). Furthermore, we see that the agreed-on code distribution per modality is even more widespread, indicating, that the scenario had a larger impact on the interpretation than the modality.

Next, we investigated the distribution of how often each code group was agreed on across scenarios, see Figure 4. Here, we found that the same gesture is agreed on with the same meaning across our five scenarios.

Agreeability with original meaning. We looked into how participants' interpretations matched the original interpretation of the expression from their reference for all arm gestures and light expressions, as the literature source for the audio [51] did not provide this information. Thus, we conducted this analysis on the 31 arm gestures and 32 light gestures out of 103 total expressions. We categorized the original meanings using our 10 code groups and compared them with the participants' interpretations. We found that for 10 arm gestures and 11 light signals, the original meaning matches the agreed code group ($AS \geq 33\%$) implying that only 20% of the expressions matched their original meaning. As most expressions were derived from different settings and were originally performed by significantly more anthropomorphic robots, we expected that the original meaning would be lost. Once again, the strong correlation between the scenario and the meaning of these expressions is evident.

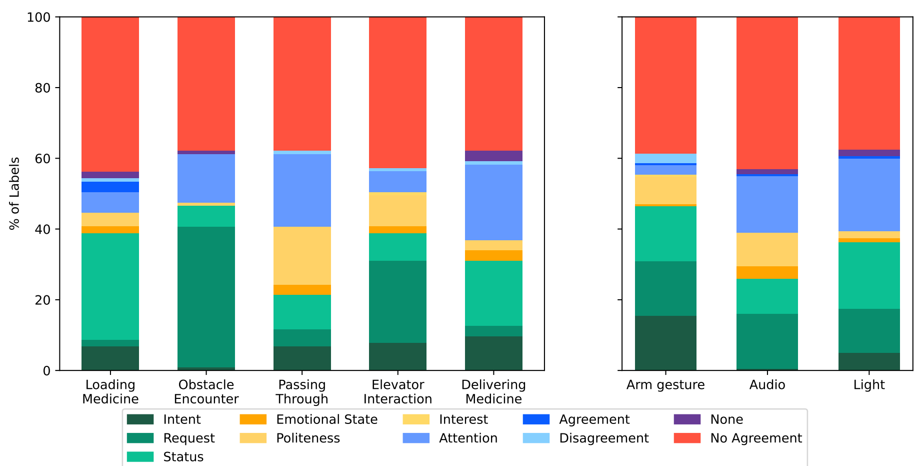

Fig. 3: Distribution of agreed on Code Groups across different scenarios (left) and modalities (right). No Agreement means that all 3 participants' interpretations were coded into different code groups. Thus, the expression was not agreed on.

4.2 Statistical Analysis

We fitted multinomial models to explain code groups by our independent variables. Due to the limited size of the dataset, all models containing expressions did

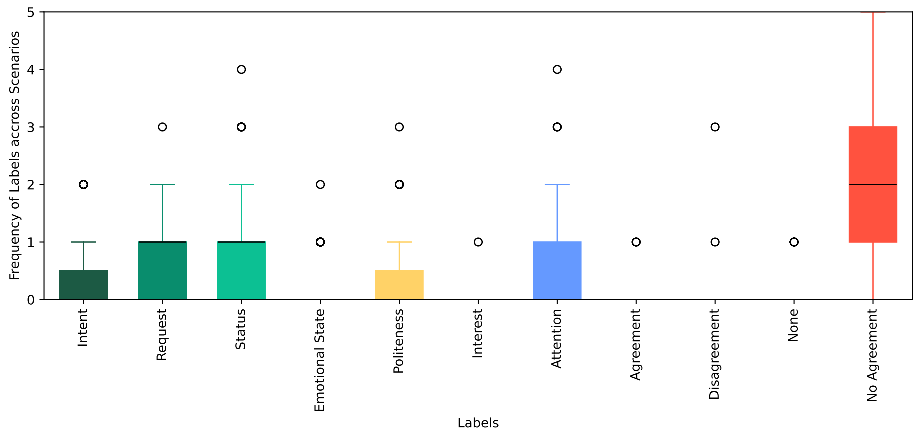

Fig. 4: Distribution of Code Groups across scenarios.

not converge. Therefore, we removed this factor for all further analyses and instead focused on reduced models to analyze the influence of scenario and modality on the dependent variables.

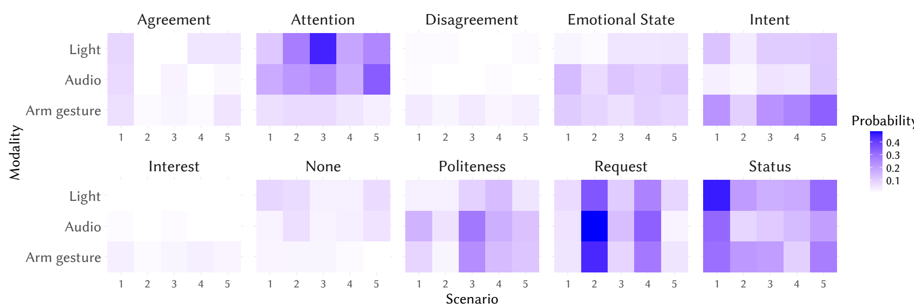

Fig. 5: Predicted probabilities of choosing a particular code group for different combinations of Scenario and Modality. Darker shades indicate higher probabilities.

Predicting code group from scenario and modality We fitted a multinomial logistic regression model with scenario, modality and their interaction as fixed effects to predict code group. We further fitted two reduced models, one excluding scenario and one excluding modality to evaluate the influence of the factors. Likelihood ratio tests confirmed that removing scenario ($\chi^2(108) = 436.89$, $p < .001$) or modality ($\chi^2(90) = 318.45$, $p < .001$) significantly degraded the model fit.

Accordingly, we continued with the full model including both factors and their interaction.

As a measure of the explanatory power of the model, we calculated the Nagelkerke pseudo-R^2. The result $R^2 = 0.36$ represents a moderate fit and indicates that a substantial part of the variance in the data can be explained by scenario and modality alone. This indicates that the interpretation of an expression by our participants depended not only on the specific expression but also to a large extent on what modality was used and in which scenario it was performed.

Figure 5 presents the predicted probabilities from the full model for selecting a code group based on different combinations of scenario and modality. The most frequently assigned interpretations by participants were "Status" ($M = 0.22$, $SD = 0.10$), "Request" ($M = 0.19$, $SD = 0.16$), and "Attention" ($M = 0.18$, $SD = 0.12$).

A more detailed analysis of these code groups shows that arm gestures rarely lead to the interpretation "Attention" ($M = 0.06, SD = 0.02$), but that audio ($M = 0.23, SD = 0.07$) (especially in scenario 5, 0.34) and light ($M = 0.25, SD = 0.12$) (especially in scenario 3, 0.45) do so more frequently. For "Request" we found no strong dependence on the selected modality, but a strong dependence on the scenario. In scenarios 2 ($M = 0.43, SD = 0.07$) and 4 ($M = 0.29, SD = 0.04$), all expressions across all modalities are primarily interpreted as "Request". For "status", we found a more complex picture with strong connections between scenario 1 ($M = 0.36, SD = 0.09$) and 5 ($M = 0.26, SD = 0.06$), as well as the modalities light ($M = 0.26, SD = 0.12$) and arm gestures ($M = 0.22, SD = .08$). Across all conditions, we found the strongest connection for audio in scenario 2, where roughly half of our participants (0.48) interpreted any gesture as a request.

These probabilities support the interpretation that both modality and scenario exhibit a strong impact on the interpretation of participants.

Predicting agreement from scenario and modality We employed a similar approach to predict agreement from scenario and modality. For this analysis, we focused only on expressions with perfect agreement, as explained in section 3.2. For the analysis, we fitted a binomial regression model with scenario, modality and their interaction as fixed effects. Again, we fitted two reduced models excluding scenario and, respectively, modality to evaluate the influence of the factors. The model fit did not significantly decline by removing scenario ($\chi^2(12) = 6.10$, $p = 0.9$) or modality ($\chi^2(10) = 17.18$, $p = 0.07$). This result indicates that both independent variables have no significant influence on the explanatory power of the model. Accordingly, the Nagelkerke pseudo-R^2 of the complete model of $R^2 = 0.06$ also demonstrates a very low explanatory power of the model. From this we conclude that the influence of scenario and modality on whether there is a unique interpretation of gestures across several people is very small or non-existent. Thus, predicting the agreement without knowledge of the actual expression seems infeasible.

4.3 Distribution of Code and Code Groups

In the final iteration of coding, participants' interpretations of the expressions were coded into a total of 10 code groups. Figure 6 shows the distribution of these group codes between scenarios. The results show uneven distributions of the code groups. For most scenarios, approximately 60% of the interpretations fell into 3 (or less) out of 10 code groups. Calculating the normalized entropy[7] of these frequencies, we see that scenario 2 has the lowest value, 0.718 (where 1 would correspond to a perfect distribution of the 10 code groups), with almost a majority of participants interpreting "Request". On the other hand, scenario 1 has the highest entropy, 0.851, with the majority of interpretations distributed between "Attention", "Status" and "Intent". These results could be easily explained by the scenario in which they are interpreted, once again highlighting the codependency of scenarios in the interpretation of these expressions.

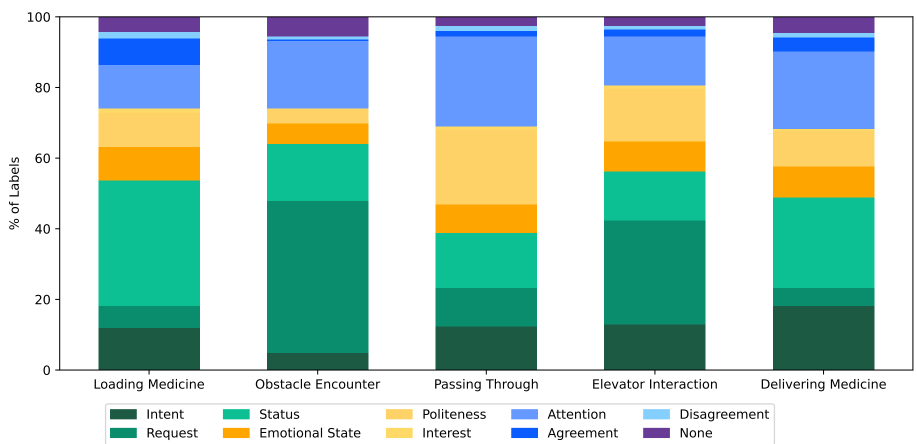

Fig. 6: Distribution of Code Groups across different scenarios

5 Discussion

Our study offers key insights into how social signals from non-anthropomorphic robots are interpreted across contexts and modalities. We discuss these findings in relation to our research questions and propose design implications, including how AI can support the development of adaptive, context-aware robot signals for more intuitive human-robot interaction.

[7] $H_{\text{normalized}} = -\frac{\sum_{i=1}^{N} p_i \log_2 p_i}{\log_2(N)}, \qquad N = 10$, where p_i is the frequency of code group i

5.1 Interpretation of Expressions Vary Across Different Contexts

Our analysis revealed low agreement rates across all evaluated expressions (see subsection 4.1), suggesting that expressions from prior studies [51,42] may not generalize well to the hospital context. This supports Leusmann et al. [31], who highlight the importance of context in interpreting expressions and note the potential for confusion in different settings. Interestingly, lights achieved the highest Agreeability Score (see Table 1), possibly due to standardized color meanings (e.g., red for error, green for OK) defined by ISO standards [25]. This aligns with prior research showing that visual signals—especially when paired with sound—can enhance robot behavior legibility [1], and in some cases, light signals outperform gestures and audio in perception [17].

The inconsistency in gesture interpretation extended to comparisons with the original authors' intended meanings. Our results showed very low agreement with the expressions' original meanings (see section 4.1), suggesting that context and robot embodiment may have altered the interpretation. This reinforces the idea that expressions are not universally understood, and context and embodiment significantly affect how humans, especially users with little knowledge about robot signals, perceive social signals [49,40]. For example, the robot saying "Wall-e" (used in four of our audio expressions) was easily understood as the robot's name in the movie context, but only 18 out of 60 participants recognized it in our study context.

5.2 Influence of Modality and Scenarios for Expressions

Our results show that modality and scenario significantly explain the variability in the data (see subsection 4.2), suggesting that both the expression's modality and the environment strongly guide its interpretation. For example, in scenario 2, where the robot stands outside a closed door, 43% (Figure 3) of participants interpreted the expression as "Request" (likely asking to open the door). Similarly, in scenario 4, standing outside an elevator, 29.4% interpreted it as "Request". In scenario 1, where the robot is loading medicine, 35.6% interpreted it as "Status" (performing a task or waiting for instructions), while in scenario 5, interpretations were more evenly spread across "Status", "Attention", and "Intent" (indicating it was delivering medicine, wanted to, or simply alerting the human). Combining modality with scenario further strengthened these effects. For instance, light in scenario 1, regardless of pattern, led to the interpretation of "Status" in nearly half of cases. Further analysis is needed to optimize modality combinations for desired interpretations.

5.3 Needs to Design Dynamic and Adaptive Social Cues for Interactive Robotic Systems

Our findings highlight the impact that context might have in interpreting social signals, showcasing the importance of adapting these signals to successfully express different intentions in different contexts. Although there have been attempts to design and explore different social signals for non-anthropomorphic robots, these signals might not be generalizable to many contexts. In logistics robots, for example, each light color has a specific meaning for error, warning, progress, etc. [28]. For instance, a blue spotlight is a common way for robots to express "I am coming"; however, in the context of the works of Hoggermueller et al. [23], blue lights for robots expressed sadness. The task of mapping the correct cue to the intended expression becomes even more challenging when, besides context, people's preconceptions of expressions (e.g. green is for go) also have an impact on how these expressions are interpreted. Therefore, there is need for concrete design implications and guidelines to design and implement social cues and signals dynamically as well as including cues which have already been standardized in specific context.

Expressions cannot be universally used across contexts and non-anthropomorphic robots. When transferring the expressions from the literature to our context and robot, we found that most of the expressions were interpreted differently from what they were originally designed to express. This finding suggests that expressions cannot be generalized, as the same expression in a specific context and done by a specific robot can convey a very different meaning than in another context and from another robot. This seems to be especially true when the robot's level of anthropomorphism varies strongly. For instance, when transferring lights from studies with anthropomorphic robots[50] to our setup, emotional expressions were much more difficult to convey (see the low values for the code group "Emotional State" in Figure 3). Following these findings, designers and developers should consider developing techniques to transfer meaning and intention from one context to another.

Good understanding of the contexts and modalities are crucial in human-robot interaction. Our study emphasizes the importance of context-aware communication of non-anthropomorphic robots to effectively convey social cues. For robots to express social cues appropriately, it is essential to identify suitable modalities that align with the specific context of use. Therefore, designers should consider the environment and its constraints when selecting modalities for social cue expression. For instance, in noisy settings like warehouses or factories, sound or speech may not be effective for communication due to high ambient noise levels. In contrast, quieter environments such as hospitals may allow sound-based modalities to convey certain cues effectively. By tailoring interaction modalities to the context, designers can enhance the effectiveness and appropriateness of robot social interactions.

Sounds need to be carefully designed in different contexts. The study found that people interpret sound expressions differently. Among the three gestures, the sound expressions had the least agreement among the participants.

This aligns with the findings of Leusmann et al. [31] who also found that sound cues were confusing for participants during human-robot interaction. This leads to the conclusion that sound expressions might not be enough to express intent or social cues on its own. Designers should consider integrating other modalities with sounds to appropriately express intent. Otherwise, there might be misunderstanding in the context, which could decrease users enthusiasm, trust and acceptance towards the robot.

Using Artificial Intelligence to generate context appropriate social cues. Artificial intelligence (AI), such as large language models (LLMs), has been proven to be efficient in processing large data to achieve user-adaptive results [21], [52]. Thus, **integrating LLMs to dynamically generate social cues for interactive robots, based on knowledge from previous literature and standards regarding contexts and expressions,** would support the generation of adaptive social cues. Designers and AI engineers should collaborate and understand the possibilities and challenges of such LLM systems and how it could be integrated for robot systems for dynamic social cues expression. We also suggest that such LLM systems would not only be beneficial for non-anthropomorphic robots but also for anthropomorphic and social robots, enabling them to generate appropriate and adaptive social cues in several contexts, culture, and environment. Designing LLM systems to incorporate context, the interaction modalities available to the robot, and knowledge from the previous literature on robots would contribute to novel paradigms in human-robot interaction design.

5.4 Limitations

Some expressions from the literature were excluded due to limitations in our non-anthropomorphic robot setup. Adapting gestures from anthropomorphic robots often led to a loss of meaning, as complex gestures (e.g., involving heads or torsos) couldn't be replicated. Although our focus was on non-anthropomorphic robots, this adaptation challenge likely influenced participant interpretations.

The online format introduced ambiguity in responses. Despite careful question design and quality control, some answers were hard to code. To address this, a follow-up in-person study is planned.

Having only three interpretation samples per expression limited statistical analysis and increased susceptibility to outliers, potentially affecting result robustness.

Another limitation was the lack of original audio files from prior studies. While we sourced sounds (e.g., Wall-E, R2-D2) online, most papers didn't provide audio or defined meanings. This made it difficult to ensure accuracy in interpreting audio cues. In contrast, we confirmed meanings for lights and gestures, lending more reliability to those modalities.

6 Conclusion

This paper investigates how humans interpret social cues from non-anthropomorphic robots across different contexts. Through observations and co-design workshops in a hospital setting, we explored human-robot interactions in five scenarios. We gathered 103 gesture, light, and audio expressions from the literature and recreated them using a non-anthropomorphic robot. Our findings show that there is no universal understanding of social cues, interpretation is highly dependent on context and modality. The social situation often drives interpretation, regardless of the actual signal, making it difficult for non-anthropomorphic robots to convey varied meanings in the same context. As humans adapt social signals to context, robots must do the same, especially as they enter semi-public spaces. With advances in adaptive technology [9], tailoring social signals to specific contexts could enhance acceptance and trust. We offer design implications to help integrate adaptive cues, guiding designers in creating contextually appropriate, socially aware interactions for non-anthropomorphic robots.

Acknowledgments. This work is partially supported and funded by the HumanE AI Network under the European Union's Horizon 2020 ICT program (grant agreement no. 952026) and Mixed Fleet Research Project funded by Business Finland.

References

1. Angelopoulos, G., Vigni, F., Rossi, A., Russo, G., Turco, M., Rossi, S.: Familiar acoustic cues for legible service robots. 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) pp. 1187–1192 (2022), `https://api.semanticscholar.org/CorpusID:252625333`
2. Banh, A., Rea, D.J., Young, J.E., Sharlin, E.: Inspector baxter: The social aspects of integrating a robot as a quality inspector in an assembly line. In: Proceedings of the 3rd international conference on human-agent interaction. pp. 19–26 (2015)
3. Baraka, K., Rosenthal, S., Veloso, M.: Enhancing human understanding of a mobile robot's state and actions using expressive lights. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN). pp. 652–657. IEEE (2016)
4. Bassett, D.S., Gazzaniga, M.S.: Understanding complexity in the human brain. Trends in cognitive sciences **15**(5), 200–209 (2011)
5. Baumgarten, M., Süß, H.M., Weis, S.: The cue is the key. European Journal of Psychological Assessment (2015)
6. Bethel, C.L., Murphy, R.R.: Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **38**(1), 83–92 (2007)
7. Breazeal, C.: Designing sociable robots. MIT press (2004)
8. Burgoon, J.K., Guerrero, L.K., Manusov, V.: Nonverbal signals. Handbook of interpersonal communication pp. 239–280 (2011)
9. Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., et al.: A survey on evaluation of large language models. ACM Transactions on Intelligent Systems and Technology **15**(3), 1–45 (2024)
10. Darwin, C.: The descent of man, and selection in relation to sex, vol. 2. D. Appleton (1872)

11. De Graaf, M.M., Allouch, S.B., Klamer, T.: Sharing a life with harvey: Exploring the acceptance of and relationship-building with a social robot. Computers in human behavior **43**, 1–14 (2015)
12. Dwyer, J.: Using communication mechanisms derived from social cues in industrial autonomous mobile robots to improve interactions. Ph.D. thesis, Queensland University of Technology (2023)
13. Elprama, B., El Makrini, I., Jacobs, A.: Acceptance of collaborative robots by factory workers: a pilot study on the importance of social cues of anthropomorphic robots. In: International Symposium on Robot and Human Interactive Communication. vol. 7 (2016)
14. Embgen, S., Luber, M., Becker-Asano, C., Ragni, M., Evers, V., Arras, K.O.: Robot-specific social cues in emotional body language. In: 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication. pp. 1019–1025. IEEE (2012)
15. Ende, T., Haddadin, S., Parusel, S., Wüsthoff, T., Hassenzahl, M., Albu-Schäffer, A.: A human-centered approach to robot gesture based communication within collaborative working processes. In: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 3367–3374. IEEE (2011)
16. Feine, J., Gnewuch, U., Morana, S., Maedche, A.: A taxonomy of social cues for conversational agents. International Journal of human-computer studies **132**, 138–161 (2019)
17. Fernandez, R., John, N., Kirmani, S., Hart, J.W., Sinapov, J., Stone, P.: Passive demonstrations of light-based robot signals for improved human interpretability. 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) pp. 234–239 (2018), `https://api.semanticscholar.org/CorpusID:53051722`
18. Fiore, S.M., Wiltshire, T.J., Lobato, E.J., Jentsch, F.G., Huang, W.H., Axelrod, B.: Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior. Frontiers in psychology **4**, 859 (2013)
19. Fiske, S.T.T.: Social cognition: From brains to culture. SAGE Publications Ltd (2020)
20. Frederiksen, M.R., Stoey, K.: Augmenting the audio-based expression modality of a non-affective robot. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII). pp. 144–149. IEEE (2019)
21. Ghosh, A., Huang, B., Yan, Y., Lin, W.: Enhancing healthcare user interfaces through large language models within the adaptive user interface framework. In: International Congress on Information and Communication Technology. pp. 527–540. Springer (2024)
22. Hebesberger, D., Koertner, T., Gisinger, C., Pripfl, J.: A long-term autonomous robot at a care hospital: A mixed methods study on social acceptance and experiences of staff and older adults. International Journal of Social Robotics **9**(3), 417–429 (2017)
23. Hoggenmueller, M., Chen, J., Hespanhol, L.: Emotional expressions of non-humanoid urban robots: the role of contextual aspects on interpretations. In: Proceedings of the 9TH ACM International Symposium on Pervasive Displays. pp. 87–95 (2020)
24. Holland, J., Kingston, L., McCarthy, C., Armstrong, E., O'Dwyer, P., Merz, F., McConnell, M.: Service robots in the healthcare sector. Robotics **10**(1), 47 (2021)
25. Plastics – Determination of fracture toughness (G_{IC} and K_{IC}) – Linear elastic fracture mechanics (LEFM) approach. Standard, International Organization for Standardization, Geneva, CH (Mar 2000)

26. Jung, M.F.: Affective grounding in human-robot interaction. In: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. pp. 263–273 (2017)
27. Langton, S.R., Watt, R.J., Bruce, V.: Do the eyes have it? cues to the direction of social attention. Trends in cognitive sciences **4**(2), 50–59 (2000)
28. Larsson, F., Wernestrand, A.: Transformation of internal logistics: with autonomous transportation (2023)
29. Law, M., Ahn, H.S., Broadbent, E., Peri, K., Kerse, N., Topou, E., Gasteiger, N., MacDonald, B.: Case studies on the usability, acceptability and functionality of autonomous mobile delivery robots in real-world healthcare settings. Intelligent Service Robotics **14**(3), 387–398 (2021)
30. Lawrence, C.: Rise and Fall of Rethink Robotics — asme.org. `https://www.asme.org/topics-resources/content/rise-fall-of-rethink-robotics` (2019), [Accessed 10-10-2024]
31. Leusmann, J., Villa, S., Liang, T., Wang, C., Schmidt, A., Mayer, S.: An approach to elicit human-understandable robot expressions to support human-robot interaction. arXiv preprint arXiv:2410.01071 (2024)
32. Liew, T.W., Tan, S.M.: Social cues and implications for designing expert and competent artificial agents: A systematic review. Telematics and Informatics **65**, 101721 (2021)
33. Martín, F.A., Castro-González, Á., Salichs, M.Á., et al.: Sound synthesis for communicating nonverbal expressive cues. IEEE Access **5**, 1941–1957 (2017)
34. Morgan, A.A., Abdi, J., Syed, M.A., Kohen, G.E., Barlow, P., Vizcaychipi, M.P.: Robots in healthcare: a scoping review. Current robotics reports **3**(4), 271–280 (2022)
35. Mubin, O., Kharub, I., Khan, A.: Pepper in the library" students' first impressions. In: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–9 (2020)
36. Murphy, R.R.: Human-robot interaction in rescue robotics. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) **34**(2), 138–153 (2004)
37. NAGELKERKE, N.J.D.: A note on a general definition of the coefficient of determination. Biometrika **78**(3), 691–692 (09 1991). `https://doi.org/10.1093/biomet/78.3.691`, `https://doi.org/10.1093/biomet/78.3.691`
38. Nourbakhsh, I.R.: Robot futures. Mit Press (2015)
39. Polzin, T.S.: Verbal and non-verbal cues in the communication of emotions. In: 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100). vol. 4, pp. 2429–2432. IEEE (2000)
40. Salem, M., Ziadee, M., Sakr, M.F.: Effects of politeness and interaction context on perception and experience of hri. In: International Conference on Software Reuse (2013), `https://api.semanticscholar.org/CorpusID:41081900`
41. Sauer, V., Sauer, A., Mertens, A.: Zoomorphic gestures for communicating cobot states. IEEE Robotics and Automation Letters **6**(2), 2179–2185 (2021)
42. Saunderson, S., Nejat, G.: How robots influence humans: A survey of nonverbal communication in social human–robot interaction. International Journal of Social Robotics **11**(4), 575–608 (Jan 2019). `https://doi.org/10.1007/s12369-019-00523-0`, `http://dx.doi.org/10.1007/s12369-019-00523-0`
43. Sauppé, A., Mutlu, B.: The social impact of a robot co-worker in industrial settings. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems. pp. 3613–3622 (2015)

44. Song, S., Yamada, S.: Designing expressive lights and in-situ motions for robots to express emotions. In: Proceedings of the 6th international conference on human-agent interaction. pp. 222–228 (2018)
45. Song, Y., Tao, D., Luximon, Y.: In robot we trust? the effect of emotional expressions and contextual cues on anthropomorphic trustworthiness. Applied Ergonomics **109**, 103967 (2023)
46. Tulli, S., Ambrossio, D.A., Najjar, A., Rodríguez-Lera, F.J.: Great expectations & aborted business initiatives: The paradox of social robot between research and industry. BNAIC/BENELEARN **1** (2019)
47. Villani, V., Pini, F., Leali, F., Secchi, C.: Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. Mechatronics **55**, 248–266 (2018)
48. Vinciarelli, A., Salamin, H., Pantic, M.: Social signal processing: Understanding social interactions through nonverbal behavior analysis. In: 2009 IEEE computer society conference on computer vision and pattern recognition workshops. pp. 42–49. IEEE (2009)
49. Wainer, J., Feil-Seifer, D., Shell, D.A., Matarić, M.J.: The role of physical embodiment in human-robot interaction. ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication pp. 117–122 (2006), `https://api.semanticscholar.org/CorpusID:12556718`
50. Xie, Y., Matsumoto, M.: Emotional expression for humanoid robot using led light and breathing simulator. IEEJ Transactions on Electrical and Electronic Engineering **17**(9), 1372–1374 (2022)
51. Zhang, B.J., Fitter, N.T.: Nonverbal sound in human-robot interaction: A systematic review. J. Hum.-Robot Interact. **12**(4) (Dec 2023). `https://doi.org/10.1145/3583743`, `https://doi.org/10.1145/3583743`
52. Zhang, C., Chen, J., Li, J., Peng, Y., Mao, Z.: Large language models for human-robot interaction: A review. Biomimetic Intelligence and Robotics p. 100131 (2023)