RESEARCH-ARTICLE

Impact of Design Transparency on Trust and Data Sharing During Human-Robot Interactions in Public Places

Accepted: 01 December 2025
Revised: 13 October 2025
Received: 06 May 2025

Impact of Design Transparency on Trust and Data Sharing During Human-Robot Interactions in Public Places

AZRA ARYANIA, Institut de Robòtica i Informàtica Industrial-CSIC-UPC, Spain and Innovation Value Institute-Maynooth University, Ireland
SABARATHINAM CHOCKALINGAM, Institute for Energy Technology, Norway
HANNE KRISTINE RØDSETHOL, Institute for Energy Technology, Norway
GUILLEM ALENYA, Institut de Robòtica i Informàtica Industrial-CSIC-UPC, Spain

The prevalence of social robots is increasing, with examples such as customer service robots in malls and airports. This trend highlights the importance of transparency, particularly in data-sharing interactions with social robots operating in public spaces, where users may be asked to provide personal information to receive personalized experiences. This paper investigates how design transparency influences user trust and data-sharing behavior in human-robot interactions. We conducted an experiment with 143 participants who interacted with the social robot *ARI* under two transparency conditions: low and high transparency. In the low transparency condition, participants were informed about the data being collected and could choose to save or delete it. In the high transparency condition, the robot additionally indicated the sensitivity level of each data item: low (e.g., scenario preference), medium (e.g., name and email), and high (e.g., religious beliefs), allowing participants to make more informed decisions. Participants were presented with two scenarios: exploring city events and discovering local attractions. They received personalized recommendations based on their preferences, with the option to provide personal data (name, phone number, email) for possible future communication. After the interaction, participants decided whether to save or delete the data they had shared. The results indicated that while transparency did not significantly affect trust in the robot, it influenced data-sharing behavior. In particular, participants in the high transparency condition demonstrated more cautious behavior, opting to save less data and delete more. Furthermore, the results showed that both sensitivity levels and transparency influenced the participants' data-sharing choices. Low-level sensitivity data led to the highest rates of saving and the lowest rates of deleting, while medium-level sensitivity data showed the opposite pattern. These findings highlight the need to align data categorization with user perceptions to address data sharing concerns more effectively.

CCS Concepts: • **Social and professional topics** → *Privacy policies*; • **Human-centered computing** → *Interaction devices*.

Additional Key Words and Phrases: Data sharing, Human-Robot Interaction, Transparency, Trust, Social Robots

1 Introduction

The increasing use of social robots, especially their ability to interact with people, is becoming more prominent in private and public spaces. These systems frequently collect personal data to provide personalized experiences, such as dietary restrictions [1], medical information [46], or restaurant recommendations [2]. This introduces significant ethical and legal challenges [12]. Compliance with European data protection laws such as the General

Authors' Contact Information: Azra Aryania, Institut de Robòtica i Informàtica Industrial-CSIC-UPC, Barcelona, Spain, aaryania@iri.upc.edu and Innovation Value Institute-Maynooth University, Maynooth, Ireland, azra.aryania@mu.ie; Sabarathinam Chockalingam, Institute for Energy Technology, Halden, Norway, Sabarathinam.Chockalingam@ife.no; Hanne Kristine Rødsethol, Institute for Energy Technology, Halden, Norway, Hanne.Rodsethol@ife.no; Guillem Alenya, Institut de Robòtica i Informàtica Industrial-CSIC-UPC, Barcelona, Spain, galenya@iri.upc.edu.

ACM 2573-9522/2025/12-ART
https://doi.org/10.1145/3785152

Data Protection Regulation (GDPR) [6] and alignment with ethical guidelines such as the Engineering and Physical Sciences Research Council (EPSRC) Principles of Robotics [4] and trustworthy AI [34] are critical.

Transparency is a foundational concept in these frameworks particularly in the context of personal data handling [12, 13]. Despite its emphasis on regulatory and ethical standards, the role of transparency in Human-Computer Interaction (HCI) and Human-Robot Interaction (HRI) remains complex. Prior work shows that transparency can improve trust, awareness, and user engagement [3, 5, 35], but it can also reduce trust or create discomfort when it is poorly designed or overly detailed [10, 22, 30, 44]. The effects vary according to context, task, modality, and user expectations [11, 27, 37, 42]. For instance, visual versus verbal transparency channels have different impacts [37], and the balance between clarity and overload remains a key design concern.

Motivated by these mixed findings, we investigate how different transparency designs influence trusting beliefs and the sharing of personal information during HRI. We implemented two levels of transparency, low and high, based on the following three approaches: 1) providing explicit information about data collection and use, 2) indicating the sensitivity of personal data, and 3) providing users control over saving or deleting shared data. In the low transparency condition, participants received limited information, but could save or delete their data. In the high transparency condition, participants received detailed explanations, sensitivity indicators, and the same control options. Before the main experiment, we conducted a pre-testing phase (A/B testing) during the Project Expo in Fredrikstad, Norway, with 18 participants. This pre-test aimed to validate the transparency design and interface clarity by having participants interact with two display conditions and provide feedback on their understanding of privacy-related information. In addition to pre-testing, we conducted a field study at a public festival in Barcelona with 143 participants who interacted with an ARI social robot. During the interaction, the robot asked questions about city events and attractions, collected personal preferences and contact information, and allowed users to choose which data to save or delete. The participants then completed a questionnaire measuring their trusting beliefs in the robot.

The results indicated that while transparency did not affect trusting beliefs, it did affect the sharing of personal information. Participants in the high transparency group chose to delete more data items on average, whereas those in the low transparency group saved more. Interestingly, data categorized as medium-level sensitivity had the lowest save rate, contrary to the expectation that highly sensitive data would be shared less.

The rest of the paper is organized as follows: Section 2 provides background on definitions of transparency, reviews state-of-the-art research on how transparency influences trust and information sharing in HCI and HRI contexts, and discusses transparency design within these domains. Section 3 outlines the research methodology, including the research questions (RQs), hypotheses, and experimental design. Section 4 presents the experimental results, while Section 5 discusses the findings and highlights the study's limitations. Finally, Section 6 concludes the paper and suggests directions for future work

2 Related Works

2.1 Definitions of Transparency

While the concept of transparency remains somewhat ambiguous, it has been interpreted and defined in various ways across different guidelines. The transparency requirement in GDPR was first recognized at a 2009 conference in Madrid and later included in the 2012 GDPR proposal [13]. It is now a core principle in Article 5(1)(a) of the GDPR, which states that personal data must be "processed lawfully, fairly, and in a transparent manner concerning the data subject" [13]. According to GDPR [6], transparency is crucial for disclosing personal data collection and processing, requiring individuals to be clearly informed about when, how, and by whom their data is handled. This is especially important in complex technological contexts, such as social robots that interact with the public and provide recommendations based on users' preferences and, in some cases, their personal data. Beyond legal requirements, transparency also plays an ethical role, as it supports human autonomy by

enabling users to understand data usage, facilitating informed consent, and allowing them to make meaningful choices about technology use. Transparency should provide users with valuable information to support informed decision-making, as its absence can undermine the perceived trustworthiness of information providers.

Transparency plays a crucial role in building trust in AI systems by ensuring that users and stakeholders can understand and assess the system's functionality, decisions, and limitations. The definition of transparency according to the High-Level Expert Group on Artificial Intelligence (HLEG) [34] includes three key elements: traceability, explainability, and communication. Traceability involves maintaining a log of the system's development and operations to enable retrospective analysis, while explainability focuses on making the technical processes and rationale behind the AI's decisions understandable. Communication emphasizes informing users about the AI system's characteristics, limitations, and the fact that they interact with an AI, not a human. This definition aims to ensure user awareness and help users accurately calibrate their trust in the system by addressing the AI's explainability and limitations.

Transparency in AI has been discussed in different types and typologies [13, 25, 36, 45]. Felzmann et al. [13] classified transparency as prospective and retrospective. Prospective transparency requires organizations (data controllers) to inform individuals (data subjects) about data processing before it begins, using clear, accessible language. This includes details such as who is processing the data, what data is being processed, why, and how long it is being used, especially when targeting vulnerable groups such as children. Retrospective transparency allows individuals to trace decisions made by automated systems, potentially involving human intervention and explanations. In another study, the notion of transparency in AI is divided into two categories: outward and functional transparency [36]. Outward transparency focuses on the relationship between the AI system and external entities, such as developers, users, and media, addressing how the system is developed, presented, and deployed. Functional transparency concerns the system's inner workings, such as how decisions are made or factors are weighted. Both aspects are often intertwined but differ in their ease of implementation. Zhang et al. [45] associated transparency with explainability and presented two definitions of explanations: local and global explanations. Local explanations refer to the rationale behind a single, specific outcome of a model, providing insight into the factors that influenced that particular decision. In contrast, global explanations focus on how the algorithmic model works, offering a broader understanding of the model's overall decision-making process. In a similar study, Preece et al. [25] linked transparency to explainability by distinguishing between transparency-based explanations, which focus on understanding the inner workings of an AI model and aims to provide insight into how the model functions, and post-hoc explanations, which clarify specific decisions after they are made without revealing the model's internal processes or unpacking its underlying mechanisms.

The concept of transparency in robotics has been interpreted differently in guidelines, such as the EPSRC Principles of Robotics [4], which emphasizes the need for transparency with respect to the machine-like nature of a robot to avoid misleading users. These guidelines assert that robots should not be designed to exploit vulnerable individuals by concealing their true nature. However, the impact of transparency in real-world applications is more complex and context-dependent, varying according to the robot's purpose and application. Transparency in robotics, particularly in therapeutic robots within home environments, requires effective privacy mechanisms that ensure data integrity, prevent misuse, and allow users to adjust their preferences in these sensitive settings [32]. The study [32] examined privacy frameworks for therapeutic robots in home settings, focusing on Nissenbaum's Contextual Integrity (CI) [24] and Fair Information Practices (FIPs) [16]. CI emphasizes the norms that govern the flow of information within specific contexts, while FIPs advocate universal rights such as transparency, access, and control over personal data. Applied to robots, these frameworks emphasize the need for transparency in data collection, storage, and use, as well as for providing users with control over sensors, data storage locations, and access.

2.2 Effects of Transparency on Trust and Information Sharing in HCI and HRI

The complexity of transparency in robotics, especially in non-industrial settings, can be understood in terms of trust and utility. Although the EPSRC Principles of Robotics advocate transparency to avoid deception, the relationship between transparency and trust is not straightforward. While increased transparency can enhance trust by clarifying decision-making, it may also reduce a robot's utility in emotionally engaging roles, such as companion robots, indicating the need for further research to explore the balance between transparency, utility, and trust [42].

The impact of transparency on trust, awareness, user perceptions, and engagement have been investigated in HCI [3, 5, 27]. Rader et al. [27] focused on the effects of four types of explanations (what-explanations, how-explanations, why-explanations, and objective explanations) of the Facebook News Feed on user beliefs and judgments about the News Feed. They found that all explanations increased participants' awareness of the algorithm's effects and understanding of their News Feed content. However, what- and how-explanations worked better than why- and objective explanations. In addition, while brief explanations effectively support awareness and accountability, they struggled to address interpretability and correctness. Chen et al. [5] conducted an initial exploration of how personalization and information transparency enhance user engagement and product involvement. They investigated the effects of personalization (overt vs. covert) and transparency (high vs. low) on data collection on perceived control and trust in mobile apps. They demonstrated that overt personalization increases perceived control, while high transparency enhances trust, reduces privacy concerns, and increases user engagement. Bitzer et al. [3] examined how Algorithmic Transparency (AT) affects user adoption, comprehension, and trust in contact-tracing apps. An online experiment with 116 participants showed that greater disclosure of the app's inner workings (transformation AT) positively influences user adoption, understanding, and trust.

In HRI, several studies have highlighted the positive impact of transparency on trust, user experience, and human-robot interaction performance [35, 38, 39]. Wang et al. in their studies [38, 39] demonstrated that clear explanations of a robot's decision-making process enhance trust, transparency, and human-robot team performance. Moreover, Vitale et al. [35] examined the impact of transparency in privacy-sensitive data collection systems using a humanoid robot and a screen-based kiosk. The analysis demonstrated strong evidence that a transparent system leads to a more positive user experience compared to a non-transparent system for both cases.

The extent to which transparency can enhance awareness and build trust was explored in several studies [10, 11, 22, 43]. In a recent study, Ezenyilimba et al. [11] explored how robot transparency and explanations influence human trust and situation awareness in a simulated environment. The study found that context-driven robot explanations significantly improved situation awareness and trust, allowing human teammates to make quicker decisions and take more active roles. Although transparency alone did not yield the highest situation awareness and trust scores, combining transparency with detailed explanations led to more trust and situation awareness. However, Yang et al. [43] reported different findings. They explored transparency in a task where participants controlled drones and monitored their visualizations for threats, with transparency manipulated through the drones' alert systems. Participants received either a low-transparency binary warning ("Danger" or "Clear") or a higher-transparency, nuanced alert system. Their findings revealed that higher transparency had a stronger effect on participants' trust, leading to more calibrated responses compared to the low-transparency condition.

Although transparency is generally seen as a sign of openness and accountability and can enhance perceived trustworthiness, it does not always lead to increased accountability or trust in practice. The ethical literature recognizes the complex effects of transparency, including both positive and negative outcomes [18]. Wanner et al. [40] proposed a Technology Acceptance Model (TAM) that highlights how transparency strongly influences trust in the system's ability, though its direct effect on trust propensity is not significant. Cramer et al. [7] in their study, indicated that while transparency improves perceived understanding and acceptance, it does not directly enhance trust. Research on transparency's impact on trust has revealed mixed results. In the field of

public policy, Grimmelikhuijsen [17] and De Fine Licht [8] reported that the impact of transparency on trust was not significant. In areas involving sensitive trade-offs, such as human well-being against financial considerations, transparency can have negative effects on public legitimacy [8]. This suggests that transparency does not always enhance trust, and decision-makers must carefully consider the context of decision-making.

Moreover, contrary to the prevailing narrative in explainable AI (XAI) research, while transparency is important, it must be carefully calibrated to avoid negative effects on trust and ensure the responsible use of AI systems. Schmidt et al. [30] through a behavioral experiment involving a machine learning-based decision support tool for text classification, revealed that increased transparency can reduce trust in some cases, particularly when AI's predictions are correct. The study also found instances of misplaced trust when the AI's predictions were incorrect. Furthermore, the study by Straten et al. [33] explored the effects of transparency regarding a robot's lack of human psychological capacities on children's perceptions and relationships with the robot. The results showed that transparency decreased children's human-like perceptions of the robot and reduced their trust. However, transparency did not affect children's feelings of closeness toward the robot. In another study, Yu et al. [44] reported contradictory results. They explored how AI decision-making transparency affects human trust, using an experimental vignette to compare transparent and non-transparent conditions. They found that while transparency enhances perceived effectiveness and trust, it can also increase discomfort, which can reduce trust. This dual effect highlights that transparency does not always have a universally positive impact, as discomfort can negate the benefits of perceived effectiveness, depending on the context.

While some studies have explored the impact of transparency on information sharing in HCI [20, 26], to the best of our knowledge, research on how transparency influences information sharing in HRI remains scarce [28]. Pu et al. [26] examined how Social Network Service transparency impacts self-disclosure intentions, finding that network transparency reduced perceived information control, while social presence and monitoring enhanced it. Hesselmann et al. [20] explored privacy concerns in ride-sharing services by developing a browser extension to enhance transparency, finding that while it improved user awareness and decision reflection, it had little impact on data disclosure rates.

When exploring attitudes toward sharing information with mobile home assistant robots, Reinhardt et al. [28] conducted a survey to understand participants' comfort and preferences regarding data collection, processing, and sharing by mobile home assistant robots, with a focus on informational privacy. The results revealed that the participants generally preferred local data processing, were uncomfortable with sharing location data even for emergency purposes, and showed interest in transparency regarding data usage. However, they were less comfortable with the robot predicting their next location, highlighting the need for further exploration of transparency and explainability features.

2.3 Design Transparency and Privacy Mechanisms for Trust and Data Sharing in HCI and HRI

Implementing transparency in robotic technology involves clearly communicating the robot's actions, capabilities, and data handling in an easily understandable way. Users often express varied privacy concerns about social robots in home settings, particularly about video recording. Although some studies, such as De Graaf et al. [9], allowed users to consent to filming by clicking a button. However, transparency regarding whether the robot was actively recording remained insufficient, leaving users uncertain and uneasy. Heuer et al. [21] emphasized the importance of addressing privacy throughout robot development to balance functionality with ethical considerations, ensure transparency in the collection of personal data, and provide users with the ability to intervene, for example, by disabling specific features. Furthermore, studies [15, 31] emphasized the need to balance sensor activation, functionality, and sensory transparency, highlighting the importance of trust, especially in older adults' homes. This transparency requires clear visual or auditory signals, such as status lights or indicators, to inform users when cameras or microphones are actively recording [15, 31].

Building on this, the study by Sedenberg et al. [32] tackled the privacy challenges posed by therapeutic robots, focusing on enhancing transparency and granting users greater control over data access, storage, and review processes, while ensuring data integrity using robust mechanisms. For users with diminished autonomy, clear guidelines for caregiver access are crucial to balance oversight with respect to personal privacy. Transparency regarding government access to data and adherence to legal frameworks is also essential. Moreover, enabling responsible data sharing through privacy-preserving systems, such as personal data stores, empowers users to manage consent and control access while supporting research.

Similarly, studies [19, 29] underscored a strong user preference for enhanced privacy indicators, local data processing, and control mechanisms that enable users to monitor and actively manage sensitive data collection, such as health and location information, across smart devices in both work and home settings. The study by Richter et al. [29] investigated employees' preferences for privacy indicators in smart workplaces, particularly smartwatches, and found that employees prefer splash-screen indicators to raise awareness about data collection and value the ability to control or interrupt it. Hernandez Acosta et al. [19] found a demand for greater transparency, local data storage, and deletion controls in voice assistant users, with participants expressing a need for customizable settings, "private modes," and notifications for data usage. This underscores the universal need for enhanced privacy controls across diverse cultural contexts.

Furthermore, transparency-enhancing tools, such as the Data Track, play a key role in supporting user control over personal data by enabling ex-post transparency [14]. The Data Track allowed users to monitor their data disclosures, manage consent, and access their data history, aligning with GDPR requirements for data portability and consent revocation. Additionally, Wang et al. [37] developed a three-level transparency model for a robotic driving assistant, using varying amounts of information through auditory and visual communication channels. However, their findings indicated that the same level of transparency produced different results depending on the task, even within the same driving context.

Finally, Zhong et al. [46] explored the impact of varying transparency levels (low and high), based on collected data about user perception and system efficacy in the context of socially assistive robots for perinatal depression screening. They found that increased transparency enhanced the robot's perceived clarity and persuasiveness, improving its ethical acceptance. However, despite efforts to improve transparency by conveying explicit information about the robot's decision-making process, such as tone of voice, participant understanding remained low across all conditions.

To further investigate the impact of transparency on trust and data sharing during user interaction with robots, we aim to explore the separate manipulation of designed transparency at two levels (low and high) in the context of a robot as a service provider in public spaces. Building on the findings from the studies mentioned above, which emphasize the need for transparency in data collection and user control, we propose a combination of three distinct approaches to transparency: 1) informing participants about the data being collected, 2) describing the sensitivity level of each data item, and 3) enabling users to control their data by saving or deleting the collected information. In the low transparency condition, the robot informs participants about the collected data and provides options to save or delete their data. In the high transparency condition, the robot goes further by additionally highlighting the sensitivity level of each data item, categorized as low, medium, or high. This approach builds on existing work on transparency while providing a structured framework to evaluate its impact on user trust and data sharing.

3 Research Methodology

3.1 Research Questions and Hypotheses

This paper aims to evaluate participants' trusting beliefs and data-sharing behavior in interactions with a social robot in public spaces. In line with these objectives, we seek to present the design transparency in a realistic

manner that can be applied to real-world interactions. This leads us to formulate the following Research Questions (RQs).

RQ1: How do different levels of designed transparency (low vs. high) affect the trusting beliefs of participants who interact with a social robot as a service provider in public spaces?

RQ2: How do different levels of designed transparency (low vs. high) affect participants' data sharing with a social robot as a service provider in public spaces?

RQ3: How does communicating different sensitivity levels of personal information affect participants' data sharing with a social robot a service provider in public settings?

In addressing the RQs and objectives of this study, we designed the following hypotheses (Hs):

H1: The average value of trusting beliefs among participants interacting with a robot designed for high transparency is higher than that of those interacting with the robot designed for low transparency.

H2: Participants interacting with a robot in the high transparency condition make fewer "save" choices and more "delete" choices on average compared to those interacting with a robot in the low transparency condition.

H3: Participants interacting with a robot in the high transparency condition make fewer "save" choices and more "delete" choices across all levels of sensitivity (low, medium, and high) compared to those interacting with a robot in the low transparency condition.

H4: Participants make the most "save" choices and the fewest "delete" choices for low-level sensitivity information compared to medium- and high-level sensitivity information.

H5: Participants make the fewest "save" choices and the most "delete" choices for high-level sensitivity information compared to low- and medium-level sensitivity information.

3.2 Experimental Design

3.2.1 Overview and Setup. The study consisted of two parts: a pre-testing (A/B testing) conducted during the Project Expo in Fredrikstad, Norway, and a subsequent main experiment carried out at the public festival (Festa de la Ciència 2024) in Barcelona, Spain. Participation was open to visitors over the age of 18. Eighteen participants engaged in the A/B test during the Project Expo. The pre-testing was conducted in a public space, where participants interacted with the robot and evaluated the two display conditions in a naturalistic environment. The main experiment had 143 participants, comprising 61.5% females, 37.1% males, and 1.4% who preferred not to disclose their gender. The demographic distribution of the participants is outlined in Table 3. Both parts used the ARI social robot (PAL Robotics), equipped with a touchscreen interface for participant interaction. The pre-testing was intended to validate the transparency design and interface clarity before conducting the large-scale study. The main experiment then evaluated the finalized setup with a larger and more diverse sample. The two participant groups were completely independent. This structure also provided an opportunity to test the setup across diverse participants and different environmental contexts.

3.2.2 Pre-testing. The pre-testing phase of this study aimed to evaluate the feasibility and effectiveness of the planned activities, particularly focusing on user engagement and human-robot interaction. A significant component of this phase was an A/B testing activity designed to assess participants' understanding of privacy-related information and their perceptions of transparency. The methodology related to A/B testing is outlined below:

A/B Testing Setup

Two screens were prepared for the A/B testing:

1. Standard Information Display: This screen presented the personal information collected by the system without any indication of sensitivity levels or color coding, as shown in Fig. 1.

2. Enhanced Transparency Display: A screen presented the same information with color-coded sensitivity levels (high, moderate, and low sensitivity) as shown in Fig. 2.

Participants were asked to interact with both displays and provide feedback on their clarity and perceived

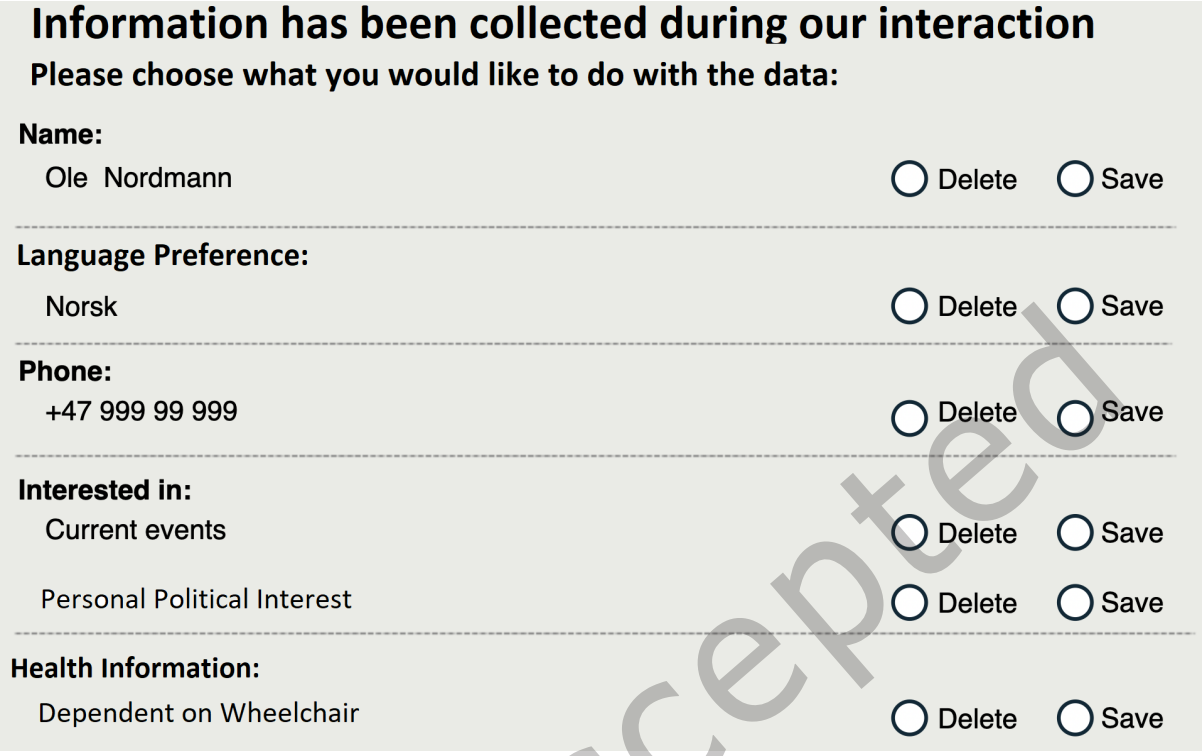

Fig. 1. A/B Testing - Example (Standard)

transparency. Participants' responses to the A/B testing were captured using a structured format. Two key questions were asked:
Q1. Did the color-coding make the sensitivity of information shared more understandable for you?
Q2. Did the color-coding on the sensitivity of information shared make you feel that the system is more transparent?
Both questions were followed by open-ended prompts, such as "Why/Why not?" to gather detailed qualitative feedback. The interaction process included: i. Introducing participants to the robot's potential applications, such as providing recommendations for local events or attractions, ii. Explaining the data-sharing scenario, where the municipality collects personal information to tailor recommendations, and iii. Presenting the two displays and asking participants to evaluate the information.

3.2.3 Main Experiment. In this study, participants interacted with the robot through a touchscreen interface, while the robot simultaneously read out the questions and guided them through each step. The robot's speech recognition and dialogue management capabilities were not employed because the outdoor setting—with high background noise and many attendees—could have compromised the reliability of such features. This configuration ensured consistent delivery of the scenarios across participants and supported the study's focus on transparency-by-design. Nevertheless, the interaction was less natural than fully multimodal or embodied HRI settings, resembling screen-based interaction in some respects, although the robot's embodied presence and verbal guidance provided an experience distinct from that of a standalone screen.

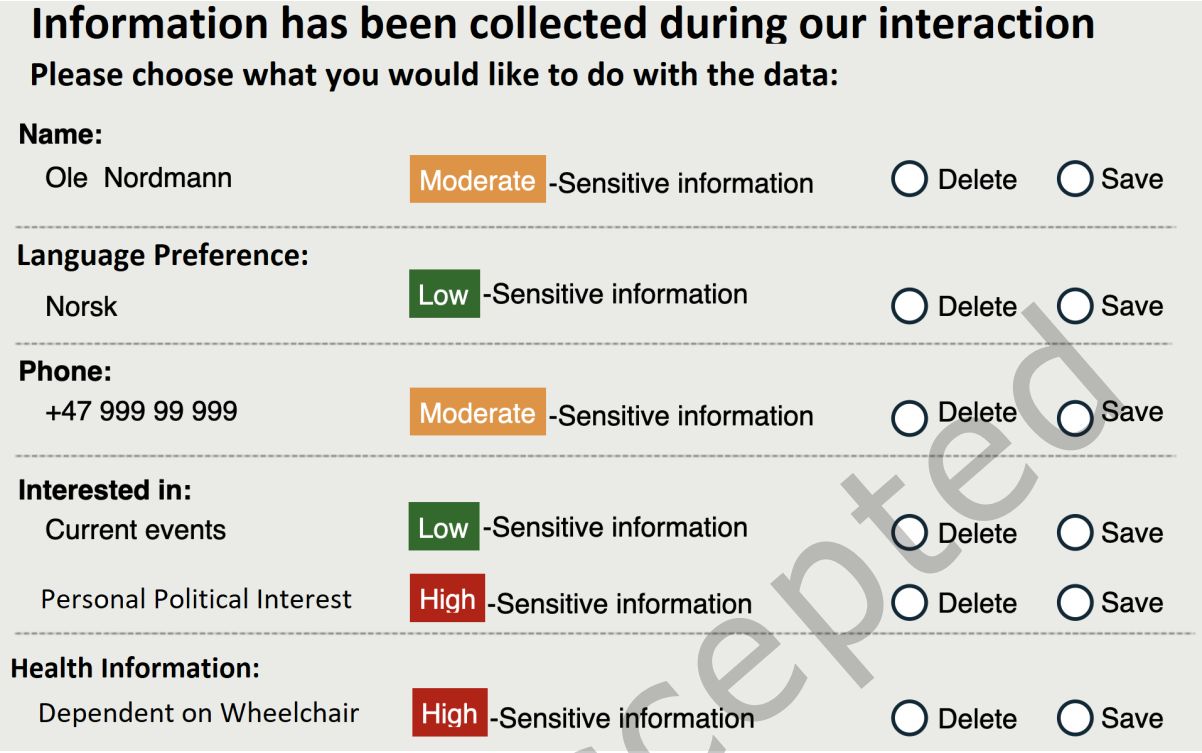

Fig. 2. A/B Testing - Example (Enhanced Transparency Display)

To address the RQs, we deployed two distinct service scenarios that the social robot (ARI) could provide: current events (Scenario 1) and interesting attractions (Scenario 2). Participants chose one of the two scenarios based on their interests. The robot then asked relevant questions tailored to the chosen scenario, prompting participants to share their preferences and personal data. Using this information, ARI provided personalized recommendations aligned with their interests. Each set of recommendations was organized into specific categories:

Scenario 1: Current events, with recommendations categorized as Cinema, International/Expat Group-based, or Religious.

Scenario 2: Local attractions, with recommendations categorized as Museums/Galleries or Religious Places.

In addition to preferences, participants were asked for personal information (name, phone number, and email) to enable future communication and updates. However, sharing this data was optional. All preferences and personal data were classified by sensitivity level: low, medium, and high.

i. Low-level sensitivity: Information such as preferences for scenario categories or specific types of events.

ii. Medium-level sensitivity: Personal details such as name, email, or phone number.

iii. High-level sensitivity: Information about religious beliefs, ethnicity, children, or preferred location.

These data points were categorized based on the European Commission's guidelines for determining high-level sensitivity information [41]. For instance, questions about preferences for specific scenario categories were considered low-level sensitivity, while questions requesting name, email, or phone number were deemed medium-level sensitivity. Finally, inquiries about religious beliefs, ethnicity, children, or preferred location were classified

Table 1. Scenario questions and categories with their sensitivity levels

Scenario	Category	Questions
Scenario 1: Current events in the city (Low-Level)	Cinema (Low-Level)	• Are you in a group with kids? (High-Level) • Which age group should I take into account when recommending a film to you? (High-Level) • Which genre of movies are you interested in? (e.g., Action, Adventure, Comedy, Documentary; Drama, SCI-FI, Thriller) (Low-Level)
	International/Expat Group-based (Low-Level)	• Are you originally from abroad? (High-Level) • Are you interested in events with expat groups/internationals? (Low-Level)
	Religious (Low-Level)	• Are you interested in events with a Christianity connection? (High-Level)
Scenario 2: Interesting attractions in the city (Low-Level)	Museums/Galleries (Low-Level)	• In which district of the city do you want to visit a museum or gallery? (High-Level)
	Religious Places (Low-Level)	• Do you have beliefs/interests in Christianity? (High-Level)

as high sensitivity. Table 1 shows the specific questions for each category in both scenarios, along with their sensitivity levels.

Based on the two transparency conditions (low vs. high), we divided participants into two groups. In the low transparency condition, the robot notified participants of the collected data and allowed them to choose whether to save or delete it. In the high transparency condition, the robot provided the same options but also indicated the sensitivity level of each data item using the low/medium/high categories, as mentioned above.

3.2.4 Procedure. The main experiment procedure consists of three steps: pre-interaction, interaction, and post-interaction, as shown in Fig. 3. During the pre-interaction phase, we gave participants information about the research project and the scenarios and asked them to sign the consent form. In the first part of the interaction phase, participants interacted with the robot and answered the questions presented on its touchscreen. The questions were related to the scenarios (see subsection 3.2.3 and Table 1 for more details). Then, depending on their responses, the robot displayed either a list of current events or a list of interesting attractions. In the second part of the interaction phase, which focused on design transparency, the robot asked participants whether they would like to share personal or contact details, including name, phone number, and email, to receive a monthly newsletter about current events, updates on attractions, or a list of interesting events or attractions via email or SMS. Depending on their preferences, participants were then asked to provide their personal information. Subsequently, the robot displayed a summary of the information collected during the interaction (e.g., category

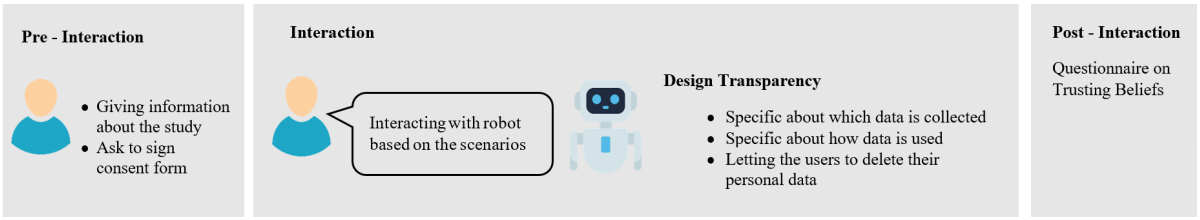

Fig. 3. Overview of the main experiment procedure

preferences and personal information) and asked participants whether they wanted to save or delete this data on the robot. They were required to decide for each item whether to save or delete it, ensuring that their preferences were explicitly recorded. Although no data were actually stored on the robot, even when participants chose to save a data point, they were not informed of this during the study to ensure genuine behavior. This approach allowed us to capture their true preferences. After the interaction, participants were debriefed and informed that no data had been saved. In the third phase, we asked participants to complete a short questionnaire about their level of trusting beliefs in the robot regarding data sharing (see Table 2 for the questionnaire items). The estimated duration of the whole experiment for each participant was about 5-10 minutes. Fig. 4 demonstrates a participant interacting with the robot.

3.2.5 Measurements. To address the study's RQs and assess the proposed hypotheses, we evaluated the trusting beliefs of the participants based on a questionnaire and the data sharing based on their responses to saving or deleting the information gathered during interaction with the robot. Trusting beliefs are defined as the degree to which users believe the robot is protecting their personal information [1, 2]. We used the same items to assess the trusting beliefs as in our previous studies [1, 2]. It contains five items (see Table 2) rated on a 7-point Likert scale ranging from (1) strongly disagree to (7) strongly agree, adapted from [23].

To determine participants' data-sharing behavior, we computed the total number of save/delete choices made during the second part of the interaction phase. In addition, we calculated the total number of save/delete choices for each participant according to the sensitivity level associated with each data item.

4 Experimental Results

4.1 Pretesting Results

Perceived Understandability of Color-Coding (related to Q1): 15 out of 18 participants (83%) reported that the color-coding made the sensitivity of shared information more understandable, while 3 participants (17%) indicated that it did not enhance their understanding.
Perceived Transparency of Color-Coding (related to Q2): 13 out of 18 participants (72%) stated that the color-coding of sensitivity levels made the system feel more transparent. Four participants (22%) did not find that the color-coding to improve transparency, and 1 participant (6%) responded with "Maybe".
Participants provided additional feedback on the information display and color-coding:
i. All presented information was labeled as "sensitive," suggesting a need to reformulate and refine the sensitivity categorization.
ii. Participants proposed the use of different shapes in addition to color-coding to improve accessibility for individuals with color vision deficiencies.
iii. Participants suggested defining the color meanings in a header or legend to reduce textual descriptions and improve clarity.

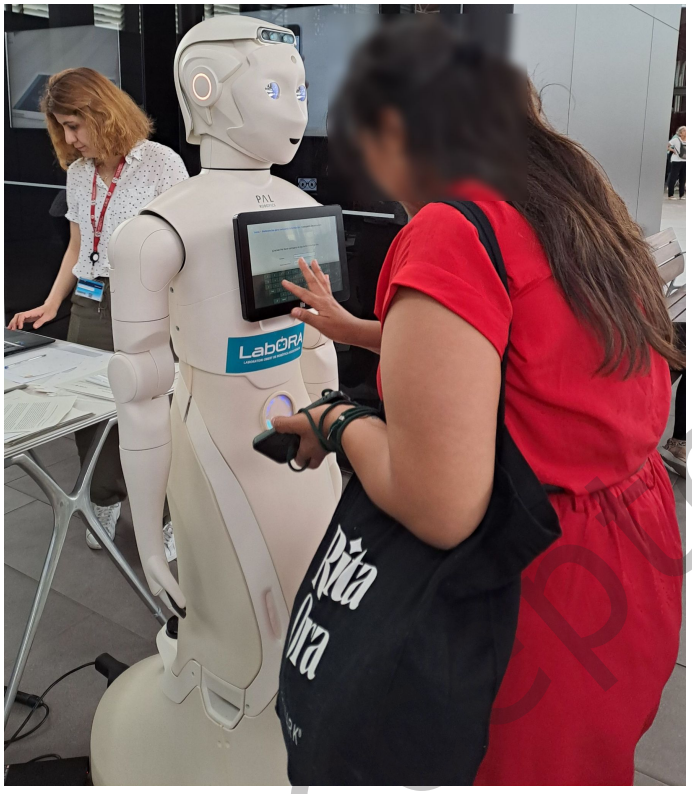

Fig. 4. A participant interacting with the robot

These findings highlight the effectiveness of color-coding in improving understandability and transparency for most participants while also revealing areas for further refinement to enhance accessibility and clarity.

4.2 Psychometric characteristics of trusting beliefs

We assessed the psychometric characteristics of dimensionality, reliability, and validity of the trusting beliefs items as a factor using Confirmatory Factor Analysis (CFA), as demonstrated in Table 2. This analysis is crucial for understanding the underlying structure and measurement properties of the items related to trusting beliefs in our study. The Average Variance Extracted (AVE) measures the extent to which a group of items within a construct aligns with the intended concept. Internal consistency reliability was evaluated by using Cronbach's Alpha, providing insights into the consistency of responses across these items. The accepted values for AVE and Cronbach's Alpha are 0.5 and 0.7, respectively. Additionally, validity was assessed through factor loadings, indicating how well each item measures the underlying construct of trusting beliefs, with each item ideally having a value exceeding 0.60. Moreover, a T-value represents the ratio of the difference between a sample mean and a population mean to the standard error of that difference, assessing the statistical significance of the mean difference. As demonstrated in Table 2, all the constructs met their criteria.

Table 2. Analysis of the dimensionality, reliability, and validity of the trusting beliefs items

	Factor loading	T	Mean	SD
Trusting Beliefs (AVE: 0.8; Cronbach's Alpha: 0.87)				
I think the robot is trustworthy in handling information.	0.91	9.61	5.50	1.19
I think the robot tells the truth and fulfills promises related to the information provided by me.	0.90	7.58	5.63	1.21
I trust that the robot would keep my best interests in mind when dealing with information.	0.99	7.91	5.01	1.45
I think the robot is in general, predictable and consistent regarding the usage of information.	0.70	7.93	5.58	1.13
I think the robot is honest with users when it comes to using the provided information.	0.94	10.50	5.43	1.20

Table 3. Demographic profile of the participants

Variable	Description	Frequency	Percentage
Gender	Female	88	61.5
	Male	53	37.1
	Prefer not to say	2	1.4
Age	18-24	7	4.9
	25-34	22	15.3
	35-44	40	28.0
	45-54	61	42.7
	More than 54	13	9.1

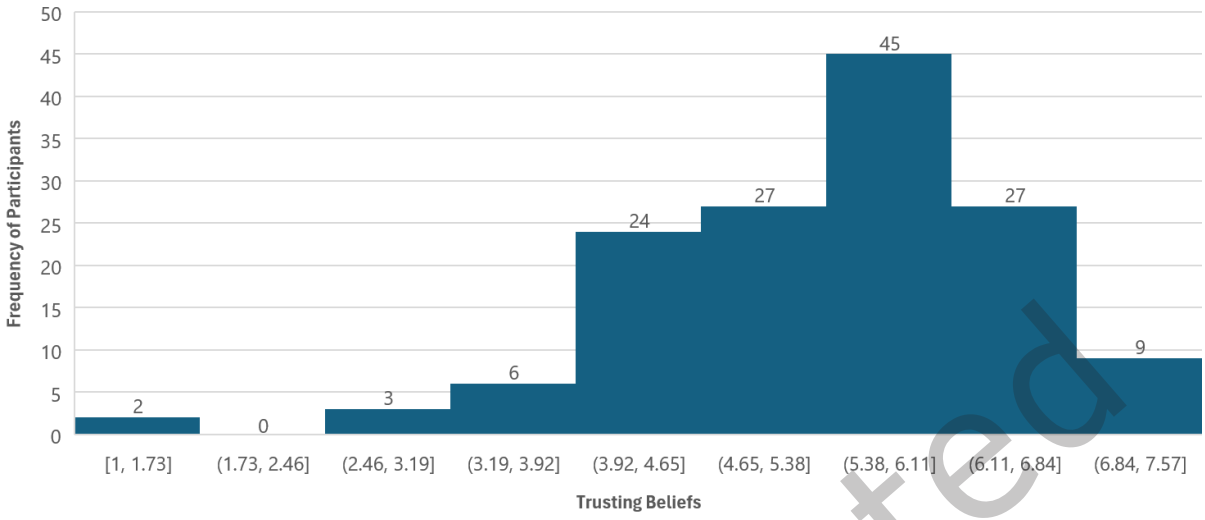

Fig. 5. Histogram showing the distribution of trusting beliefs among participants

4.3 Assessing the results of the questionnaire

As mentioned earlier, we evaluated the trusting beliefs of the participants based on a questionnaire to assess the study hypotheses. Fig. 5 shows the distribution of trusting beliefs across participants. The trusting beliefs results for all participants, with *M=5.36* and *SD = 1.09*, demonstrate that 81 individuals (56.64%) showed their agreement (or strong agreement) with the trusting beliefs questions and thus a high level of trust in the robot. While 51 persons (35.66%) were in the middle range [3.19-5.38], only 11 individuals (7.70%) showed a low level of trust in the robot.

According to the results of the one-way ANOVA test for trusting beliefs, the average value of the trusting beliefs for participants interacting with a robot in both conditions (high and low transparency design) is almost equal $M(Trusting\ Beliefs)_{High\ Transparency} = 5.33$, $M(Trusting\ Beliefs)_{Low\ Transparency} = 5.38$ and their difference does not reach statistical significance ($F = 0.06$, $p = 0.79$), which is confirmed by Fig. 6.

Hence, the findings suggest that participants who interact with the robot in a high transparency design do not exhibit increased trusting beliefs in the robot compared to those who interact with the robot in a low transparency design, which contradicts H1.

4.4 Assessing data sharing behavior in participant-robot interactions

We evaluated participants' data-sharing behavior by analyzing the total number of save and delete choices they made at the end of the interaction phase. Additionally, we computed the number of save and delete choices for each participant across different data sensitivity levels—Low-Level (LL), Medium-Level (ML), and High-Level (HL)—and conducted one-way ANOVA tests for all measurements, as presented in Table 4.

The one-way ANOVA test on the total Number of Save (NS) and Delete (ND) choices made by participants indicated a statistically significant difference between those who interacted with the robot in the high transparency design ($NS_{High\ Transparency}$, $ND_{High\ Transparency}$) and those who interacted with the robot in the Low Transparency design ($NS_{Low\ Transparency}$, $ND_{Low\ Transparency}$). Therefore, participants who interacted with the

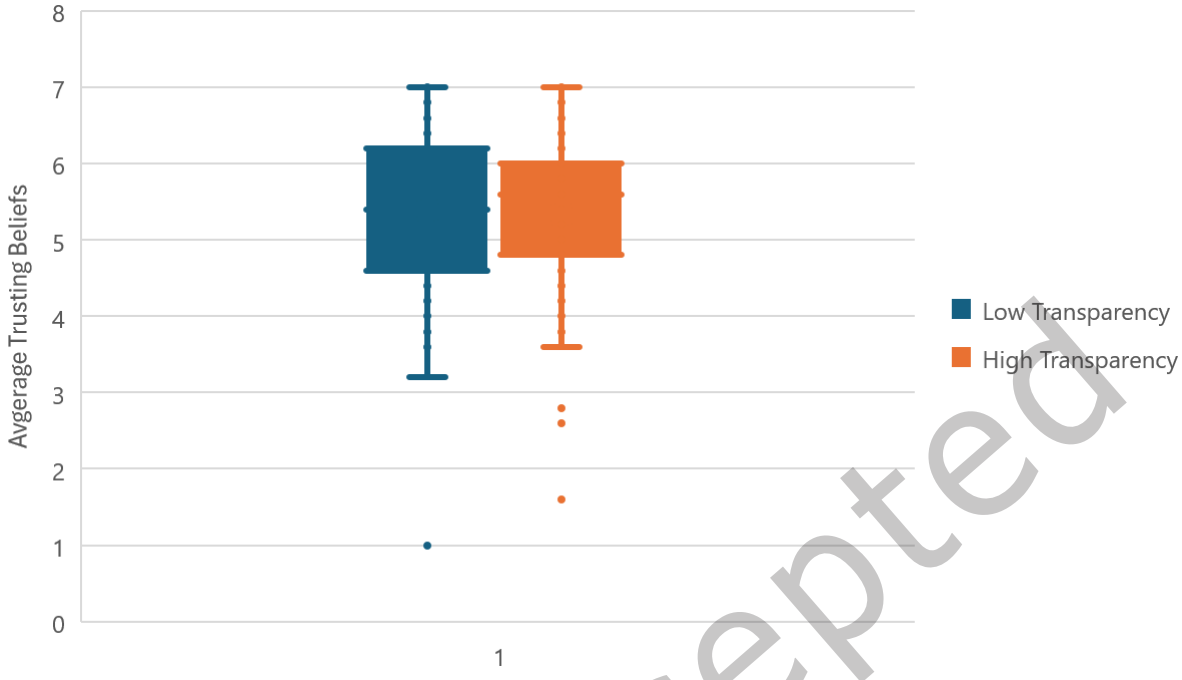

Fig. 6. The average value of trusting beliefs for each group

robot in the high transparency design appeared to make fewer save choices and more delete choices than those in the low transparency design, suggesting that H2 is supported (see Table 4).

In addition, the one-way ANOVA test on the number of save and delete choices across different sensitivity levels of the questions revealed statistically significant differences between participants who interacted with the robot in the high transparency design and those in the low transparency design. As shown in Table 4 and Fig. 7, the average number of save choices across all sensitivity levels was lower for participants in the high transparency condition than for those in the low transparency condition. In contrast, the average number of delete choices was higher for participants in the high transparency condition. These findings collectively support H3.

Moreover, to examine the average number of save and delete choices across different sensitivity levels of the data collected during the interaction with the robot, we conducted a one-way ANOVA test for the three sensitivity levels (see Table 5). The findings indicated statistically significant differences for both the average number of save and delete choices: F=215.83, p=1.85e−65 for NS and F=16.92, p=8.51e−8 for ND.

The analysis of the average levels of save choices (NS) and delete choices (ND) across different sensitivity levels revealed that low-level sensitivity data had the highest average number of save choices ($NS - LL$), while medium-level sensitivity data ($NS - ML$) had the lowest. Similarly, the lowest average number of delete choices was observed for low-level sensitivity data ($ND-LL$), whereas medium-level sensitivity data ($ND-ML$) exhibited the highest. These findings support hypothesis H4, as participants made the most ”save” choices and the fewest ”delete” choices for low-level sensitivity information. However, they contradict hypothesis H5, as participants made the fewest ”save” choices and the most ”delete” choices for medium-level sensitivity information rather than for high-level sensitivity information.

Table 4. One-Way ANOVA test results of the data sharing for each group of the participants and in terms of the sensitivity levels of the data

Measurement	Mean	SD	F-Statistic	p-value
$NS_{High\ Transparency}$	2.61	5.24	19.06***	2.43e−5
$NS_{Low\ Transparency}$	4.2	4.21		
$ND_{High\ Transparency}$	2.46	6.10	19.24***	2.24e−5
$ND_{Low\ Transparency}$	0.92	2.90		
$(NS-LL)_{High\ Transparency}$	1.86	2.20	13.81***	0.0003
$(NS-LL)_{Low\ Transparency}$	2.71	1.51		
$(ND-LL)_{High\ Transparency}$	0.41	2.47	5.70*	0.02
$(ND-LL)_{Low\ Transparency}$	0.14	1.23		
$(NS-ML)_{High\ Transparency}$	0.19	0.34	6.20*	0.01
$(NS-ML)_{Low\ Transparency}$	0.52	0.89		
$(ND-ML)_{High\ Transparency}$	1.34	0.70	12.26***	0.0006
$(ND-ML)_{Low\ Transparency}$	0.55	0.26		
$(NS-HL)_{High\ Transparency}$	0.55	0.52	14.08***	0.00026
$(NS-HL)_{Low\ Transparency}$	0.96	0.33		
$(ND-HL)_{High\ Transparency}$	0.70	0.49	21.53***	7.86e−6
$(ND-HL)_{Lw\ Transparency}$	0.22	0.28		

Note: NS and ND represent the number of save and delete choices, respectively. Suffixes -LL, -ML, and -HL indicate data with low, medium, and high sensitivity levels, respectively. Additionally, * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$.

5 Discussion

5.1 Summary of key findings and interpretation of results

To address RQ1, we analyzed the questionnaire results on trusting beliefs in the robot, revealing that most participants exhibited high trust. However, the ANOVA showed no significant difference in trust between the high and low transparency designs. This suggests that the transparency level does not significantly influence participants' trusting beliefs, contradicting H1. These results reflect the complex, context-dependent relationship between transparency and trust. While transparency is often associated with openness and accountability, it does not always translate into higher trust, as evidenced by the lack of a significant difference between the two designs. This aligns with previous studies, which found mixed effects of transparency on trust [7, 8, 17, 40]. The findings suggest that the effects of transparency on trust are more complex than initially assumed. While higher transparency was hypothesized to enhance trust, the results indicated no such effect in human-robot interactions. This lack of effect may result from the questionnaire's inability to accurately capture participants' trusting perceptions or the transparency designs not aligning with their expectations. These findings underscore the importance of calibrating transparency to the specific context and audience to balance user expectations, comfort, and the perceived necessity of transparency while avoiding unintended negative outcomes.

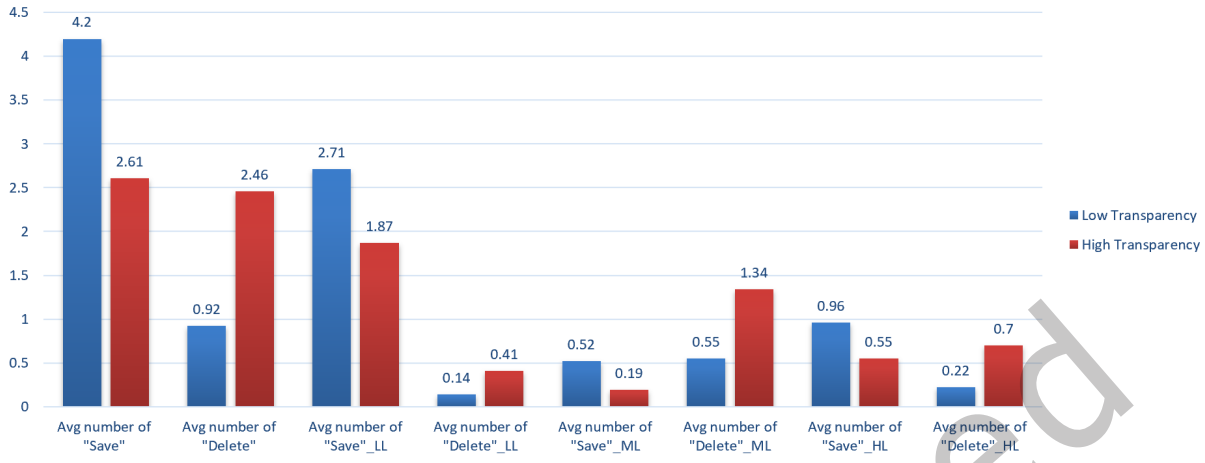

Fig. 7. The average number of save/delete for each group in terms of the sensitivity level of the data

Table 5. One-way ANOVA test results of the data sharing in terms of the sensitivity levels of the data

Measurement	Mean	SD	F-Statistic	p-value
$NS-LL$	2.47	1.35	***	
$NS-ML$	0.37	0.66	215.83	1.85e−65
$NS-HL$	0.77	0.46		
$ND-LL$	0.27	1.96	***	
$ND-ML$	0.92	0.48	16.92	8.51e−8
$ND-HL$	0.45	0.43		

Note: NS and ND represent the number of save and delete choices, respectively. Suffixes -LL, -ML, and -HL indicate data with low, medium, and high sensitivity levels, respectively. Additionally, * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$.

To address RQ2, our study found that participants interacting with a high transparency robot exhibited more cautious data-sharing behaviors, as evidenced by fewer save choices and more delete choices than those in the low transparency condition. These results support H2, suggesting that higher transparency increases users' awareness of data sensitivity, prompting more critical evaluations of the robot's data-sharing requests. This finding is consistent with previous research that highlights the role of transparency in influencing self-disclosure behaviors and decision-making [20, 26, 28]. In this context, the high transparency design encouraged participants to assess data-sharing requests, leading to reduced willingness to disclose personal data. This suggests that transparency can effectively promote cautious data sharing, aligning with users' preferences for control and accountability over their data handling practices.

To address RQ3, the study revealed significant differences in the save and delete choices of the participants based on the sensitivity of the data. Participants who interacted with the high transparency robot made fewer save choices and more delete choices across all sensitivity levels, supporting hypothesis H3. Data sensitivity also influenced their behavior, with the highest number of save and delete choices recorded for low-level sensitivity

data and the lowest for medium-level sensitivity data, which confirms hypothesis H4 but contradicts H5. These results suggest that users are more cautious about sensitive data and less likely to share it when transparency is higher. The contradiction with H5 may be due to differences between how we categorized the sensitivity of the data and how the participants perceived it. This highlights the need to align the sensitivity levels with user perceptions to obtain more accurate results. Furthermore, the findings show the importance of designing transparency features that account for data sensitivity to better address privacy concerns and help users make informed decisions about sharing their data.

Reflecting on these results, it is evident that while transparency plays a significant role in promoting cautious behavior and encouraging critical evaluation of data requests, its impact on trust and sharing behaviors is not straightforward. The effects observed across different levels of transparency and data sensitivity suggest the need for a thoughtful design approach that considers not only the transparency features themselves but also the context, user expectations, and the nature of the data being shared. Ensuring that transparency features are appropriately aligned with the data type and users' preferences are crucial for fostering trust while maintaining privacy.

5.2 Limitations

While this study provides valuable insights, it is important to acknowledge certain limitations that could influence the findings and guide future research.

Perception of sensitivity: A key limitation of this study lies in the potential mismatch between the predefined sensitivity levels of the data and participants' subjective perceptions of that sensitivity. While we categorized data into low, medium, and high sensitivity based on predefined criteria, participants may have interpreted the sensitivity of certain data types differently, depending on personal, cultural, or contextual factors. This discrepancy could have influenced their data-sharing behavior in ways not fully captured by the experimental design. For example, some participants may have perceived medium-level sensitivity data as either more sensitive or trivial, leading to unexpected patterns in saving or deleting data.

Sample Limitations: The experiment was conducted at a public science festival open to everyone and offered free admission, attracting a diverse audience. However, voluntary participation may have introduced self-selection bias, as those who chose to participate might have been more interested or motivated, potentially affecting the generalizability of the findings. To mitigate this limitation, we ensured a relatively large sample size (143 participants), which helps improve the representativeness of the sample. The diversity of festival attendees and the open nature of recruitment also contributed to capturing a wide range of perspectives. While some bias may remain, these factors enhance the generalizability of our findings.

Interaction Constraints: The robot used in the study had advanced features such as speech recognition and dialogue management, but these were not activated due to the outdoor setting. With over 1,000 attendees and high background noise levels, relying on speech recognition would have risked inaccuracies. Instead, participants interacted with the robot through a touchscreen, which ensured consistency but reduced the naturalness and interactivity of the experience. This could have influenced how participants engaged with the robot and, consequently, the study as a whole.

Finally, while pretesting revealed useful suggestions for refining the user interface (e.g., color coding, sensitivity categorization, and accessibility), these refinements were not integrated into the current experiment to maintain consistency across participants. Instead, they will inform future design iterations and follow-up studies.

6 Conclusions and Future Work Directions

This study investigated the influence of design transparency on trust and data-sharing behavior during human-robot interactions. Participants interacted with a social robot under high or low transparency conditions, answering questions of varying sensitivity levels and deciding whether to save or delete the collected data. Although transparency did not significantly impact trust (contradicting H1), it had a significant effect on data-sharing behaviors, with high transparency leading to fewer save choices and more delete choices (supporting H2). Regarding data sensitivity, participants exhibited less cautious behavior with low-level sensitivity data, saving it more frequently and deleting it less often, which confirmed H3 and H4. Interestingly, the medium-level sensitivity data were saved and deleted the least, partially contradicting H5.

These findings suggest that while transparency can raise awareness and influence cautious data sharing, the predefined sensitivity categories used in the study did not always align with participants' perceptions. This highlights the importance of tailoring transparency features to data sensitivity levels to effectively address privacy concerns.

Building on these insights, future research could collect participants' subjective ratings of data sensitivity prior to the main experiment and adjust the sensitivity labels accordingly, ensuring that the classification better reflects participants' perceptions. Alternatively, providing brief explanations or examples for why certain data are considered low, medium, or high sensitivity could help standardize participants' understanding. Incorporating these measures would reduce the mismatch between predefined sensitivity levels and participants' subjective perceptions, leading to more accurate insights into data-sharing behavior.

Further studies might explore ways to improve sample representativeness by ensuring a broader and more diverse sample that reflects the general population. This could involve conducting experiments in various public spaces to reduce participant selection bias and better capture the range of perspectives, experiences, and demographics present in the wider population. Furthermore, future research could conduct experiments in quieter and more controlled environments to better understand the impact of interactive features, such as speech recognition. Exploring different modes of interaction (e.g., touchscreens, voice, or gestures) could also help determine the most effective ways to engage users and assess robot trust and transparency in various contexts.

Finally, conducting longitudinal studies could provide valuable insights into how users' perceptions and behaviors change over time as they interact with robots in real-world, public settings. These studies would help us understand how sustained interaction with robots influences various aspects of human-robot dynamics, such as trust, sharing information, and the broader impacts of robots in public spaces. Investigating long-term effects would reveal how users adapt to and form relationships with robots over time and how factors such as transparency, data control, and robot functionality affect these ongoing interactions.

Ethical Approval and Consent to Participate

This study was approved by the Ethics Committee of the Spanish National Research Council (CSIC) under reference number 152/2024. The ethical evaluation confirmed that the research involving human participants, including the processing of personal data, complied with all applicable ethical standards. Informed consent was obtained from all participants prior to their participation. Participants were informed of the voluntary nature of the study, their right to withdraw at any time without consequence, and the measures taken to ensure data confidentiality and anonymity. No personal data were stored, and all responses were anonymized in compliance with EU Regulation 2016/679, Directive 95/46/EC (GDPR), and Spanish Organic Law 3/2018, of 5/12, on the Protection of Personal Data and guarantee of digital rights.

Acknowledgments

This work is partially funded by the Research Council of Norway (RCN) through the project "User-centred Security Framework for Social Robots in Public Space (SecuRoPS)," with the project number 321324 (INT-NO/0875), by MCIN/AEI/10.13039/501100011033, and by the European Union NextGenerationEU/PRTR under the project ROB-IN (PLEC2021-007859).

References

[1] Azra Aryania, Ruben Huertas-Garcia, Santiago Forgas-Coll, Cecilio Angulo, and Guillem Alenya. 2024. How do people intend to disclose personal information to a social robot in public spaces?. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*. IEEE, Pasadena, CA, USA, 1809–1814. https://doi.org/10.1109/RO-MAN60168.2024.10731370

[2] Azra Aryania, Ruben Huertas-Garcia, Santiago Forgas-Coll, Cecilio Angulo, and Guillem Alenya. 2025. Effects of Privacy Warning on the Intention to Disclose Personal Information During Interaction with a Robot in Public Spaces. *International Journal of Social Robotics* (2025). https://doi.org/10.1007/s12369-025-01244-3

[3] Tobias Bitzer and Martin Wiener. 2021. Algorithmic Transparency and Contact- tracing Apps – An Empirical Investigation. Montreal.

[4] Margaret Boden, Joanna Bryson, Darwin Caldwell, Kerstin Dautenhahn, Lilian Edwards, Sarah Kember, Paul Newman, Vivienne Parry, Geoff Pegman, Tom Rodden, Tom Sorrell, Mick Wallis, Blay Whitby, and Alan Winfield. 2017. Principles of robotics: regulating robots in the real world. *Connection Science* 29, 2 (April 2017), 124–129. https://doi.org/10.1080/09540091.2016.1271400

[5] Tsai-Wei Chen and S. Shyam Sundar. 2018. This App Would Like to Use Your Current Location to Better Serve You: Importance of User Assent and System Transparency in Personalized Mobile Services. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. https://doi.org/10.1145/3173574.3174111

[6] European Commission. 2016. General Data Protection Regulation (GDPR). https://doi.org/10.5040/9781782258674

[7] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction* 18, 5 (Nov. 2008), 455–496. https://doi.org/10.1007/s11257-008-9051-3

[8] Jenny De Fine Licht. 2014. Policy Area as a Potential Moderator of Transparency Effects: An Experiment. *Public Administration Review* 74, 3 (May 2014), 361–371. https://doi.org/10.1111/puar.12194

[9] Maartje M.A. De Graaf, Somaya Ben Allouch, and Tineke Klamer. 2015. Sharing a life with Harvey: Exploring the acceptance of and relationship-building with a social robot. *Computers in Human Behavior* 43 (Feb. 2015), 1–14. https://doi.org/10.1016/j.chb.2014.10.030

[10] Motahhare Eslami, Sneha R. Krishna Kumaran, Christian Sandvig, and Karrie Karahalios. 2018. Communicating Algorithmic Process in Online Behavioral Advertising. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. https://doi.org/10.1145/3173574.3174006

[11] Akuadasuo Ezenyilimba, Margaret Wong, Alexander Hehr, Mustafa Demir, Alexandra Wolff, Erin Chiou, and Nancy Cooke. 2023. Impact of Transparency and Explanations on Trust and Situation Awareness in Human–Robot Teams. *Journal of Cognitive Engineering and Decision Making* 17, 1 (March 2023), 75–93. https://doi.org/10.1177/15553434221136358

[12] Heike Felzmann, Eduard Fosch-Villaronga, Christoph Lutz, and Aurelia Tamo-Larrieux. 2019. Robots and Transparency: The Multiple Dimensions of Transparency in the Context of Robot Technologies. *IEEE Robotics & Automation Magazine* 26, 2 (June 2019), 71–78. https://doi.org/10.1109/MRA.2019.2904644

[13] Heike Felzmann, Eduard Fosch Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. 2019. Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society* 6, 1 (Jan. 2019), 205395171986054. https://doi.org/10.1177/2053951719860542

[14] Simone Fischer-Hübner, Julio Angulo, Farzaneh Karegar, and Tobias Pulls. 2016. Transparency, Privacy and Trust – Technology for Tracking and Controlling My Data Disclosures: Does This Work?. In *Proceedings 10*. Springer International Publishing, Darmstadt, Germany, July 18-22, 3–14. https://doi.org/10.1007/978-3-319-41354-9_1 Series Title: IFIP Advances in Information and Communication Technology.

[15] Nora Fronemann, Kathrin Pollmann, and Wulf Loh. 2022. Should my robot know what's best for me? Human–robot interaction between user experience and ethical design. *AI & SOCIETY* 37, 2 (June 2022), 517–533. https://doi.org/10.1007/s00146-021-01210-3

[16] Robert Gellman. 2014. Fair Information Practices: A Basic History. *SSRN Electronic Journal* (2014). https://doi.org/10.2139/ssrn.2415020

[17] Stephan Grimmelikhuijsen. 2012. Linking transparency, knowledge and citizen trust in government: an experiment. *International Review of Administrative Sciences* 78, 1 (March 2012), 50–73. https://doi.org/10.1177/0020852311429667

[18] David Heald. 2006. *Transparency: The Key to Better Governance?* British Academy.

[19] Luca Hernández Acosta and Delphine Reinhardt. 2024. "Alexa, How Do You Protect My Privacy?" A Quantitative Study of User Preferences and Requirements About Smart Speaker Privacy Settings. In *In IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer Nature Switzerland, Cham, 148–161. https://doi.org/10.1007/978-3-031-65175-5_11 Series Title: IFIP

Advances in Information and Communication Technology.

[20] Carsten Hesselmann, Delphine Reinhardt, Jan Gertheiss, and Jörg P. Müller. 2022. Data Privacy in Ride-Sharing Services: From an Analysis of Common Practices to Improvement of User Awareness *(Lecture Notes in Computer Science, Vol. 13700)*. Cham: Springer International Publishing, 20–39. https://doi.org/10.1007/978-3-031-22295-5

[21] Tanja Heuer, Ina Schiering, and Reinhard Gerndt. 2017. Privacy and Socially Assistive Robots - A Meta Study, Vol. Revised Selected Papers 12 (2018). Ispra, Italy, 265–281. https://doi.org/10.1007/978-3-319-92925-5_18 Series Title: IFIP Advances in Information and Communication Technology.

[22] René F. Kizilcec. 2016. How Much Information?: Effects of Transparency on Trust in an Algorithmic Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 2390–2395. https://doi.org/10.1145/2858036.2858402

[23] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research* 15, 4 (Dec. 2004), 336–355. https://doi.org/10.1287/isre.1040.0032

[24] Helen Nissenbaum. 2004. Privacy as Contextual Integrity. *Washington Law Review* 79 (2004), 119.

[25] Alun Preece, Dan Harborne, Dave Braines, Richard Tomsett, and Supriyo Chakraborty. 2018. Stakeholders in Explainable AI. http://arxiv.org/abs/1810.00184 arXiv:1810.00184 [cs].

[26] Wenxi Pu, Siyuan Li, Gregory J. Bott, Marie Esposito, and Jason Bennett Thatcher. 2022. To Disclose or Not to Disclose: An Evaluation of the Effects of Information Control and Social Network Transparency. *Computers & Security* 112 (Jan. 2022), 102509. https://doi.org/10.1016/j.cose.2021.102509

[27] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–13. https://doi.org/10.1145/3173574.3173677

[28] Delphine Reinhardt, Monisha Khurana, and Luca Hernández Acosta. 2021. "I still need my privacy": Exploring the level of comfort and privacy preferences of German-speaking older adults in the case of mobile assistant robots. *Pervasive and Mobile Computing* 74 (July 2021), 101397. https://doi.org/10.1016/j.pmcj.2021.101397

[29] Alexander Richter, Patrick Kühtreiber, and Delphine Reinhardt. 2022. Enhanced Privacy in Smart Workplaces: Employees' Preferences for Transparency Indicators and Control Interactions in the Case of Data Collection with Smart Watches. In *In IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer International Publishing, Cham, 34–50.

[30] Philipp Schmidt, Felix Biessmann, and Timm Teubner. 2020. Transparency and trust in artificial intelligence systems. *Journal of Decision Systems* 29, 4 (Oct. 2020), 260–278. https://doi.org/10.1080/12460125.2020.1819094

[31] Trenton Schulz, Jo Herstad, and Harald Holone. 2018. Privacy at Home: An Inquiry into Sensors and Robots for the Stay at Home Elderly. In *Human Aspects of IT for the Aged Population*. Springer International Publishing, Las Vegas, NV, USA, 377–394. https://doi.org/10.1007/978-3-319-92037-5_28 Series Title: Lecture Notes in Computer Science.

[32] Elaine Sedenberg, John Chuang, and Deirdre Mulligan. 2016. Designing Commercial Therapeutic Robots for Privacy Preserving Systems and Ethical Research Practices Within the Home. *International Journal of Social Robotics* 8, 4 (Aug. 2016), 575–587. https://doi.org/10.1007/s12369-016-0362-y

[33] Caroline L. Van Straten, Jochen Peter, Rinaldo Kühne, and Alex Barco. 2020. Transparency about a Robot's Lack of Human Psychological Capacities: Effects on Child-Robot Perception and Relationship Formation. *ACM Transactions on Human-Robot Interaction* 9, 2 (June 2020), 1–22. https://doi.org/10.1145/3365668

[34] The European commission and Expert Group on Artificial Intelligence. 2020. *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment.*

[35] Jonathan Vitale, Meg Tonkin, Sarita Herse, Suman Ojha, Jesse Clark, Mary-Anne Williams, Xun Wang, and William Judge. 2018. Be More Transparent and Users Will Like You: A Robot Privacy and User Experience Design Experiment. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Chicago IL USA, 379–387. https://doi.org/10.1145/3171221.3171269

[36] Joel Walmsley. 2021. Artificial intelligence and the value of transparency. *AI & society* 36, 2 (June 2021), 585–595. https://doi.org/10.1007/s00146-020-01066-z

[37] Jianmin Wang, Yujia Liu, Tianyang Yue, Chengji Wang, Jinjing Mao, Yuxi Wang, and Fang You. 2021. Robot Transparency and Anthropomorphic Attribute Effects on Human–Robot Interactions. *Sensors* 21, 17 (Aug. 2021), 5722. https://doi.org/10.3390/s21175722

[38] Ning Wang, David V Pynadath, and Susan G Hill. 2015. Building Trust in a Human-Robot Team with Automatically Generated Explanations. In *Proceedings of the interservice/industry training, simulation and education conference (I/ITSEC)*, Vol. 15315. Los Angeles, 1–12.

[39] Ning Wang, David V. Pynadath, and Susan G. Hill. 2016. Trust calibration within a human-robot team: Comparing automatically generated explanations. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, Christchurch, New Zealand, 109–116. https://doi.org/10.1109/HRI.2016.7451741

[40] Jonas Wanner, Lukas-Valentin Herm, Kai Heinrich, and Christian Janiesch. 2022. The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study. *Electronic Markets* 32, 4 (Dec. 2022), 2079–2102. https://doi.org/10.1007/s12525-022-00593-5

[41] European Commission Website. 2025. What personal data is considered sensitive? https://commission.europa.eu/law/law-topic/data-protection/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en
[42] Robert H. Wortham and Andreas Theodorou. 2017. Robot transparency, trust and utility. *Connection Science* 29, 3 (July 2017), 242–248. https://doi.org/10.1080/09540091.2017.1313816
[43] X. Jessie Yang, Vaibhav V. Unhelkar, Kevin Li, and Julie A. Shah. 2017. Evaluating Effects of User Experience and System Transparency on Trust in Automation. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Vienna Austria, 408–416. https://doi.org/10.1145/2909824.3020230
[44] Liangru Yu and Yi Li. 2022. Artificial Intelligence Decision-Making Transparency and Employees' Trust: The Parallel Multiple Mediating Effect of Effectiveness and Discomfort. *Behavioral Sciences* 12, 5 (April 2022), 127. https://doi.org/10.3390/bs12050127
[45] Yunfeng Zhang, Q. Vera Liao, and Rachel K. E. Bellamy. 2020. Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 295–305. https://doi.org/10.1145/3351095.3372852 arXiv:2001.02114 [cs].
[46] Mengyu Zhong, Marc Fraile, Ginevra Castellano, and Katie Winkle. 2023. A case study in designing trustworthy interactions: implications for socially assistive robotics. *Frontiers in Computer Science* 5 (June 2023), 1152532. https://doi.org/10.3389/fcomp.2023.1152532