Effects of privacy warning on the intention to disclose personal information during interaction with a robot in public spaces

Azra Aryania[1*], Ruben Huertas-Garcia[2], Santiago Forgas-Coll[2], Cecilio Angulo[1,3], Guillem Alenyà[1]

[1*]Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08020, Barcelona, Spain.
[2]Faculty of Economics and Business, Universitat de Barcelona, Av Diagonal 690, 08034 , Barcelona, Spain.
[3]Intelligent Data Science and Artificial Intelligence Research Centre, Universitat Politècnica de Catalunya, Jordi Girona 31, 08034, Barcelona, Spain.

*Corresponding author(s). E-mail(s): aaryania@iri.upc.edu;
Contributing authors: rhuertas@ub.edu; santiago.forgas@ub.edu; cecilio.angulo@upc.edu; galenya@iri.upc.edu;

Abstract

Social robots in public spaces can potentially gather personal data. This leads to privacy concerns and prompts users to consider whether to disclose their information to the robot during interaction with the robot. This paper aims to explore the effects of privacy warnings on people's intentions to disclose their personal information and their actual disclosure when interacting with a robot in public spaces. We propose a model to estimate individuals' disclosure intentions toward a robot and evaluate the impact of privacy warnings during interaction. We conducted an experiment with more than 100 participants interacting with a robot to assess the proposed model. Our results indicate that factors such as risk beliefs, trusting beliefs, perceived enjoyment, and social influence significantly influence individuals' intentions to disclose information. In general, a robot equipped with privacy warnings receives greater acceptance than a robot without privacy warnings, and participants who receive warnings disclose more low-privacy information. However, there is no significant difference in disclosure between participants who interacted with a robot with privacy warnings and those who interacted with a robot without privacy warnings. Overall, disclosure behavior does not significantly differ between the two groups, suggesting that privacy warnings do not effectively reduce disclosure information.

Keywords: Social robot, privacy warnings, intention to disclose personal information, interaction

1 Introduction

Nowadays, the utilization of social robots has significantly expanded across diverse application domains, including healthcare and therapy [1–3], education [4, 5], domestic environments [6, 7], and public spaces [8–11]. They have been employed in various application contexts to perform tasks in public spaces, such as museums [12], restaurants/bars [13, 14], and train stations [15], which

has rapidly increased their acceptance among users [16, 17].

Social robots that interact with people in public spaces can access and collect personal information, which raises privacy concerns when disclosing sensitive information [8, 18] and highlighting the significance of privacy-sensitive robotics [19]. Several studies have investigated factors influencing the intention to disclose personal information in various contexts, reporting the significance of perceived benefits as a positive effect [20–22], trusting beliefs as a positive impact [8, 23–25], and risk perceptions as a negative influence [8, 22]. Wang et al. [22] reported the positive and negative impacts of perceived benefits and perceived risks, respectively, on disclosure intention, while Harborth and Pape [24] investigated information trusting and risk beliefs in a marketing context. Additionally, Song et al. [25] discovered that factors such as service quality and enjoyment significantly influence consumers' willingness to share information with robots, while Fan et al. [26] suggested that perceived benefits and subjective norms shape disclosure intentions.

While the models address the intention to disclose information and focus on individuals' willingness to share personal data, technology acceptance models provide a more comprehensive understanding of user attitudes and behaviors regarding adopting new technologies. These models offer valuable insights into the overall acceptance and usage patterns of technology. The initial technology acceptance models [27] focused on enhancing job performance, and these models evolved to incorporate factors such as perceived enjoyment [28, 29] and social influence [30]. Subsequent versions integrated additional constructs such as perceived behavioral control [31]. Proposed model [32] emphasized key determinants such as performance and effort expectations, social influence, and facilitation conditions. Subsequent adaptations have been applied to various contexts, including elder care [16], revealing insights into user attitudes and factors affecting technology adoption, such as anxiety and attitudes toward technology use [1, 33].

Warnings serve multiple purposes, enhancing safety, influencing behavior, and providing information for informed decision-making [34]. Effective warning labels emphasize noticeable features and include a clear description of the threat and its consequences [35]. Studies have shown that privacy warnings can reduce users' intentions to disclose personal information [36–42]. In addition, mentions of privacy or security threats in web forms significantly reduce personal information disclosure [43, 44], with actionable warnings giving users greater control over their disclosure. However, some studies have suggested that privacy warnings may not effectively reduce disclosure [45–48] or may even lead to adverse effects, such as increased disclosure despite warnings [49, 50]. These studies indicate that users often ignore security warnings, even when explicitly instructed to pay attention, suggesting that mismatched perceptions of risk can lead to avoidable risks [47].

The privacy paradox [51, 52] refers to the perceived inconsistency between people's stated privacy concerns and their willingness to disclose personal information. Consumers are often assumed to either not care about privacy or misunderstand the implications of disclosure, despite understanding privacy practices through notices. However,Martin [53] argues that consumers misinterpret privacy notices as more protective than they are, leading to the belief that their privacy expectations are included in these notices. Therefore, disclosing information does not mean they are relinquishing those expectations. Similarly, the study by Lutz and Tamó-Larrieux [17] found a privacy paradox in the context of social robots, where the perceived benefits of these robots overcame users' privacy concerns, showing that the same misinterpretation of privacy assurances can occur in human-robot interaction (HRI).

Although there is extensive research on HRI and the social connections between humans and social robots, the field of HRI is still in its early stages of addressing privacy concerns and privacy warnings [19]. Lee et al. [54] investigated how privacy notices influence users' willingness to share data with robots. They found that the way private information is communicated significantly affects users' comfort with disclosure. Recently, Sakai et al. [55] reported mixed results, revealing that while robot notifications (without reasons) did improve the permission rate, the addition of rationale did not have the expected positive effect on increasing personal information usage permissions.

Some studies reported the importance of designing robots with privacy-sensitive behaviors to increase transparency and user comfort in interactions. Yang et al. [56] found that certain behaviors of a social robot (MyBom2), such as turning away or informing users about video recording, significantly reduce privacy concerns. Fernandes et al. [57] addressed privacy concerns associated with social robots used for elderly and disabled care, demonstrating that most individuals prefer when the robot avoids looking at them during privacy-sensitive situations, although some discomfort persists due to the robot's presence. The study by Martelaro et al. [58] revealed that when robots disclosed their vulnerabilities, it fostered a sense of trust and companionship in users. Similarly, in the context of conversational assistants, Saffarizadeh et al. [59] investigated how voice-based devices such as Alexa and Siri balance privacy concerns with self-disclosure.

As previously mentioned, the exploration of privacy issues within HRI, particularly particularly in relation to privacy warnings, remains relatively underdeveloped. To the best of our knowledge, there is a scarcity of empirical studies that specifically examine the impact of privacy warnings on information disclosure in HRI. While some studies, such as [17], explore privacy in terms of the intention to use robots, and [25], discuss the willingness to share information with sales robots, these works do not directly address privacy warnings or notifications related to disclosing information to robots. To fill this gap, this study proposes a research model to investigate the effects of inserting privacy warnings on people's intentions to disclose their personal information and their actual disclosure when interacting with a robot in public spaces. During interactions with the robot, users frequently have to decide whether to disclose their personal information. Those who prioritize privacy may hesitate to share their information with the robot. We propose that a robot incorporating privacy warnings during interaction will be more accepted than a robot without privacy warning messages. In addition, participants who interact with a robot delivering privacy warnings disclose less information than those who interact with a robot without privacy warnings. Therefore, we formulate the following research questions (RQs):

RQ1: To what extent do 'risk beliefs', 'trusting beliefs', 'perceived ease of use', 'perceived usefulness', 'perceived enjoyment', and 'social influence' affect the intention to disclose personal information to a robot in public spaces?

RQ2: To what extent, if any, does a robot in public spaces that delivers privacy warning messages during interactions influence its acceptance compared to a robot without privacy warnings?

RQ3: To what extent, if any, do participants interacting with a robot in public spaces delivering privacy warnings disclose less information than those interacting with a robot without privacy warnings?

2 Related Work

This section reviews relevant theories and frameworks that contextualize the research focus on models of intention to disclose personal information. It also discusses technology acceptance models adapted for social robots and explores the impact of privacy warnings on disclosure behavior particularly within HRI.

2.1 Intention to disclose personal information

The best-known theory explaining the motivations behind the behavior is the Theory of Reasoned Action (TRA) [60], which links attitudes to behaviors. The Theory of Planned Behavior (TPB) [61, 62] extends this by incorporating perceived control, while Privacy Calculus Theory [63, 64] suggests that individuals assess risks and benefits before sharing personal information, favoring disclosure when the perceived benefits outweigh the risks.

Numerous studies in the area of disclosing personal information have utilized the privacy calculus, TRA, and TPB [8, 20–26]. Malhotra et al. [23] applied TRA to introduce a causal model built upon the trust-risk framework. They revealed that the presence of trust and risk beliefs can serve as mediators in the relationship between privacy concerns and the behavioral intention to disclose personal information. In a study by McKnight et al. [20], a privacy calculus framework was introduced for Facebook users, incorporating privacy concerns, information sensitivity, and benefits such as perceived usefulness, enjoyment,

and trust. Their findings suggested that certain factors, such as trusting beliefs, privacy concerns, and information sensitivity, impact information disclosure, while others, such as usefulness and enjoyment, affect continuance intention. Xu et al. [21] introduced an integrated model incorporating the privacy calculus and TPB to examine factors impacting information disclosure on social networking sites. This model emphasized the role of perceived benefits and privacy concerns in determining the self-disclosure of personal information.

Wang et al. [22] found that factors such as self-presentation and personalized services positively impact consumers' perceived benefits, thus increasing the intention to disclose personal information. Conversely, perceived severity and perceived control had a negative effect on this intention. Harborth and Pape [24] proposed a model for behavioral intention to use, based on the model used in [23] to investigate the role of information trusting beliefs, and risk beliefs in the context of a marketing service provider. Song et al. [25] revealed that factors such as service quality, enjoyment, usefulness, representing self-interest, and trust, which are indicative of social interaction, were predictors of consumers' willingness to share personal information with robots. They found that service quality and enjoyment were the most influential factors in determining consumers' willingness to share personal information. Fan et al. [26] suggested that intentions to disclose personal information are shaped by perceived benefits and subjective norms, while attitudes toward privacy correlate with perceived risks. The results of our recent paper [8] demonstrated that risk beliefs, trust, enjoyment, and social influence play crucial roles in shaping individuals' intentions to disclose personal information. Moreover, although only 6.2% initially intended to share information, 98% of participants ultimately disclosed their personal data.

2.2 Technology acceptance models

Cognitive models commonly used to understand behavior have been applied to investigate the adoption of new technologies. Similarly, we believe that adapting such models can assist in examining the factors impacting users' intentions to disclose their personal information to a system, to ensure user acceptance.

The Technology Acceptance Model (TAM) [27], is widely used to study how users accept and adopt new technologies. Later revisions of TAM [28] introduced the concept of perceived enjoyment to explain long-term system use, particularly in both productivity and pleasure-oriented contexts. Studies [29] have shown that users are more likely to continue using a system if they find it enjoyable and useful, enhancing both job performance and personal satisfaction. TAM2 [30] extends the original TAM by incorporating social influence and cognitive instrumental processes to better explain perceived usefulness, while TAM3 [31] further refines the model by adding perceived behavioral control to explain ease of use. These enhancements offer a more comprehensive understanding of the acceptance of technology by users.

Venkatesh et al. [32] developed the Unified Theory of Acceptance and Use of Technology (UTAUT), focusing on four key determinants of technology use: performance expectation, effort expectation, social influence, and facilitation conditions. Later, UTAUT2 [65] expanded the model by adding hedonic motivation, price value, and habit, while accounting for the moderating effects of factors such as age, gender, and experience on technology adoption.

Subsequent studies have adapted the UTAUT model to address specific contexts. For instance, Heerink et al. [16] introduced the Almere model, which was designed to evaluate the acceptance of assistive social agents among elderly users. Building upon the UTAUT, the Almere model extends beyond functional evaluation factors such as perceived usefulness and ease of use, integrating variables related to social interaction. This includes new constructs specific to social robots, such as anxiety and attitudes toward technology use. The Almere model has been applied and adapted in elderly care settings employing social robots [1, 33]. Cobo Hurtado et al. [1] found that elderly users showed a positive attitude toward the robot but found it less user-friendly. Although they perceived it to be highly useful, they also reported increased anxiety and fewer facilitating conditions. Another model, RAM-care [33], extended the Almere model, incorporating perceived compatibility between the use of robots,

personal moral values, and perceived technological unemployment. Attitude and enjoyment were identified as key factors influencing the intention to use robots.

2.3 Effects of privacy warnings on information disclosure

Warnings serve as safe communications and are designed to alert individuals to potential hazards, aiming to prevent or reduce undesirable consequences [34]. Rousseau and Wogalter [35] proposed design principles for warning labels, emphasizing features such as noticeable colors, left-sided text, and framing to enhance recognition. Effective labels include "warning" in the heading, along with a description of the threat and its potential consequences. According to the model proposed by Wogalter [34], a warning is effective when individuals pay attention to it, understand its content, and influence their attitudes, encouraging them to refrain from revealing information and engaging in appropriate safety behaviors.

LaRose and Rifon [36] conducted a study to analyze the effects of privacy warnings and privacy seals on self-disclosure intentions and perceived negative consequences on websites. Their findings indicated that privacy warnings led to a decrease in users' intentions to disclose personal information, whereas privacy seals positively influenced disclosure intentions without altering perceptions of negative consequences. Braunstein et al. [37] conducted three distinct indirect surveys with users, addressing privacy indirectly through increasing privacy warnings in instructions and question-wording. They explored the impact of privacy warnings on the willingness to share personal information, revealing the strong effects of privacy warnings on question wording. These effects extended not only to users' reported willingness to share or retrieve sensitive content but also to their self-reported sharing behaviors.

Carpenter et al. [38] investigated computer-mediated warnings' effectiveness in reducing individuals' online disclosure behavior during online attack scenarios. They found that various warning terms, such as "warning," "caution," or "hazard," successfully decreased disclosure, with "hazard" being the most effective. Moreover, warnings were more effective at reducing the disclosure of driver's license numbers than email addresses. In another study by Carpenter et al. [39], the impact of warning sources (Google, FBI Cyber Division, and Department of Justice) on online information disclosure decisions was assessed, with the FBI Cyber Division being the most effective source. Carpenter et al. [40] also confirmed that warning conditions effectively reduced birth data disclosure compared to a no-warning condition. However, the authors emphasized the importance of careful design and testing of warnings for optimal effectiveness in reducing personal information disclosure.

Krol et al. [43] reported that warnings can be used to effectively manage personal information disclosure on web forms. In addition, mentioning privacy or security threats significantly decreased personal information disclosure. Furthermore, the most effective warnings give the user greater control, enabling them to choose which items they want to remove from a web form. Similarly, Mamonov et al. [44] suggested that individuals naturally take protective measures when they are aware of information security threats and limit the disclosure of sensitive information. Moreover, Feri et al. [41] found that data breach notifications prompt individuals to disclose less information to firms, especially if they are sensitive. Epstein et al. [42] conducted a survey to evaluate the impact of Terms of Service (TOS) warnings' strength and the inclusion of a click requirement on user disclosure. The results showed that strong warnings were more effective than weak warnings in reducing personal disclosures, and the click requirement enhanced the effectiveness of both weak and strong warnings. However, the commonly used TOS warning did not affect disclosure.

In contrast, various studies have indicated that privacy warnings do not effectively reduce the disclosure of personal information [45–48], with some even demonstrating adverse effects [49, 50]. For instance, Wu et al. [45] observed participants ignoring toolbar warnings despite explicit instructions to pay attention. Similarly, browser warnings in studies by Egelman et al. [46, 47] did not yield positive outcomes, with participants disregarding warnings even after multiple exposures. They concluded that when warnings do not match the users' perceptions of risk, they may engage

in avoidable risks [47]. In another experiment by Krol et al. [48], participants expressed trust in antivirus software and disregarded security warnings during PDF file downloads.

In a study on mobile site warnings, Zhang et al. [49] observed an adverse effect: despite the presence of security cues indicating the absence of a trusted security certificate, participants disclosed more social media information, including Facebook friends, Twitter IDs, and followers, compared to scenarios without the security cue. Similarly, Junger et al. [50] surveyed 100 users, requesting various details such as email addresses, recent purchase lists, online shopping sources, and partial bank account numbers. Despite warnings, 43.3% revealed banking details, and most users disclosed information, with privacy warnings showing no significant impact on disclosure.

A study by Lutz and Tamò-Larrieux [17] examined how privacy concerns affect user intentions when interacting with social robots. A survey was conducted to assess the proposed model and explore various antecedents, revealing that respondents primarily worried about data protection by manufacturers, followed by social and physical privacy concerns.

Although relatively few studies have specifically examined the role of privacy warnings in information disclosure within HRI, some research has begun to address this issue [54, 55]. Lee et al. [54] explored users' perceptions of privacy in interactions with a workplace social robot (Snackbot). Participants generally accepted data recording if informed. However, opinions were divided on accidental recordings, with some being comfortable if notified, while others expressed concerns about misuse, such as tracking their location. A recent study by Sakai et al. [55] revealed contradictory results. The authors explored the use of robot (CommU robot) notifications to acquire consent for using personal information. The research focused on developing an interactive mechanism through dialogue strategies aimed at making users feel safe when providing data. Two experiments examined the effects of reminders and rationale on users' understanding and willingness to give consent. Results showed that reminders improved users' understanding of data usage, though they did not significantly enhance feelings of safety or agreement. Additionally, robot notifications increased the permission rate, while providing rationale did not significantly impact it, possibly due to insufficient persuasiveness.

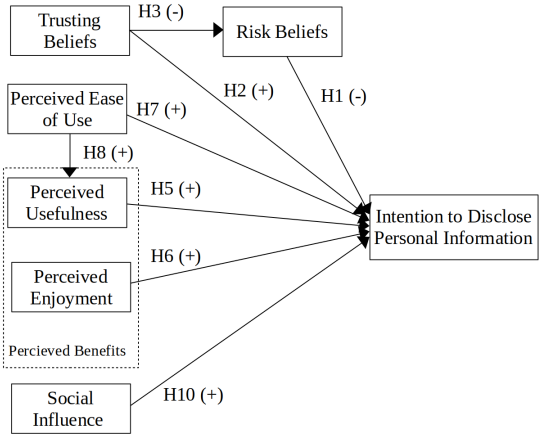

Fig. 1: Research model

Some studies have examined the impact of privacy notices on privacy concerns in interactions with social robots. Yang et al. [56] examined how robot social behaviors—specifically gaze, distance, and clarity of intent—can reduce privacy concerns in sensitive situations. Their findings highlighted that actions such as turning away or notifying users about video recording effectively alleviate these concerns. In a related study by Fernandes et al. [57], the findings demonstrated that most participants prefer when the robot avoids looking at them in sensitive situations. Martelaro et al. [58] and Saffarizadeh et al. [59] found that when robots or conversational assistants reveal personal details, it builds trust in users and increases their willingness to share personal information. The proposed research model by Saffarizadeh et al. [59] explained how conversational assistants' self-disclosure influences users' willingness to share personal information through cognitive and emotional trust.

3 Research Model and Hypotheses

In addressing the research questions and objectives of the study, we developed a research model (see Fig. 1) to estimate users' intention to disclose personal information to the robot based on existing models [20, 23, 25], and a modified version of the UTAUT model [16, 66].

As illustrated, the Intention To Disclose Personal Information (ITDPI) is a dependent construct. Six drivers are considered antecedents of intention to disclose personal information: trusting beliefs, risk beliefs, perceived benefits (integration of perceived usefulness and perceived enjoyment), perceived ease of use, and social influence. Risk beliefs and perceived usefulness are assigned a mediating role. We describe the constructs and the associated hypotheses as follows:

Risk Beliefs (RB)

It refers to users' expectations of losses associated with the disclosure of personal information [21–23, 26]. Based on the theory of privacy calculus, users perform calculations to balance perceived benefits and perceived risk during personal information disclosure [63]. Individuals who are worried about information disclosure are concerned about who might access their personal information [23]. Therefore, beliefs about potential risks associated with experience negatively influence the intention to disclose personal information [22, 23, 63]. In other words, when users perceive greater risks, they will have lower intentions to disclose personal information. Therefore, we propose the following:

- **H1:** Risk beliefs negatively influence the intention to disclose personal information.

Trusting Beliefs (TB)

It refers to the degree to which users believe the robot is protecting their personal information [23–25]. According to [23, 24], increased trust in an online company reduces the expectation of risks associated with sharing personal information. Therefore, trusting beliefs play a significant role in predicting personal information disclosure between humans and robots, leading to a positive effect on the intention to disclose personal information [23, 25, 67]. Furthermore, several studies have found that risk mediates the impact of trust on consumer purchase intentions [68, 69] and the intention to disclose personal information [23, 24] through a negative relationship. Therefore, we can expect the following:

- **H2:** Trusting beliefs positively influence the intention to disclose personal information.
- **H3:** Trusting beliefs negatively influence risk beliefs.
- **H4:** Trusting beliefs indirectly affect the intention to disclose personal information, which is mediated by risk beliefs.

Perceived Benefits (PB)

Privacy calculus proposes that when consumers are asked to disclose personal information, they will conduct a risk-benefit analysis to evaluate the benefits they will receive in return for disclosing information [63]. Perceived benefits refer to the expected returns from information disclosure, such as financial incentives [70], entertainment [20], customized and personalized services [22, 25, 71], and social connections [21], motivating consumers to trade off some level of privacy. Numerous studies have demonstrated the significant influence of perceived benefits on personal information disclosure, showing a positive relationship between users' perceived benefits and their intentions to disclose personal information [21, 22, 25, 26]. Additionally, recent findings [17] have identified a robot privacy paradox, suggesting that users' privacy concerns regarding social robots can be overridden by perceived benefits, even when privacy risks with institutions such as robot manufacturers are high. This implies that emphasizing the benefits of social robots may reduce privacy concerns in sensitive situations. In our study, we focus on two primary benefits associated with social robots: perceived enjoyment and perceived usefulness [25, 66].

- **Perceived Usefulness (PU)**
  It refers to the degree to which a user perceives a product as effective and beneficial in improving performance [16]. Perceived usefulness has been extensively studied across different contexts, including social networks, social robots, and computer-mediated environments [16, 25, 66]. In social networks, for instance, sharing information enhances the network's effectiveness and productivity, suggesting that limiting access to personal information may reduce its usefulness. Therefore, perceived usefulness can increase the intention to disclose personal information.
- **Perceived Enjoyment (PENJ)**
  It refers to the pleasure someone feels when using a product, independent of its anticipated performance consequences [29]. Enjoyment is a significant reason people use

social robots [16, 66]. Studies indicate that enjoyment can lead to increased disclosure of personal information on social networks, as users derive pleasure from sharing their activities and interests [72]. Similarly, when users find interactions with robots enjoyable, they are more likely to engage in information disclosure activities and interact with them [25].

Therefore, we propose the following:

- **H5:** Perceived usefulness positively influences the intention to disclose personal information.
- **H6:** Perceived enjoyment positively influences the intention to disclose personal information.

Perceived Ease of Use (PEOU)

It is defined as the degree to which a user believes that using a product will require minimal effort [16]. Perceived ease of use, can affect personal information disclosure. Given the novelty of robotic technology, users may perceive interacting with a robot as challenging [16]. Therefore, if interaction with the robot is intuitive and does not require advanced skills, users are more likely to engage with it [1, 16], thus increasing its perceived usefulness [16, 66] and the intention to disclose personal information [25]. Furthermore, similar to many prior studies, perceived ease of use might indirectly influence the intention to disclose personal information, mediated by perceived usefulness through a positive relationship [16, 66]. Thus, we propose the following:

- **H7:** Perceived ease of use positively influences the intention to disclose personal information.
- **H8:** Perceived ease of use positively influences perceived usefulness.
- **H9:** Perceived ease of use will indirectly affect the intention to disclose personal information, which is mediated by perceived usefulness.

Social Influence (SI)

It derives from a concept known as the subjective norm, which concerns how social factors are perceived to affect an individual's decision to engage in a particular action [62]. This component was incorporated into the UTAUT model [32], indicating the extent to which an individual considers their social environment when adopting a new system. Additionally, in the Almere model [16], it reflects individuals' perceptions that significant others believe they should or should not use the system. Since social robots are not yet widely adopted, their usage is likely to be encouraged within supportive communities. Therefore, individuals in social networks that favor robot usage are more likely to demonstrate a willingness to adopt this technology [17].

We considered specific factors within the social influence construct, including 1) the influence of friends and peers regarding the adoption of the robot, 2) the desire to be viewed positively by others when using the robot, and 3) the significance of respected individuals' opinions in shaping one's willingness to adopt the robot.

Several research studies have revealed a positive correlation between social influence and the intention to disclose personal information [26], especially when motivated by perceived benefits such as financial incentives [70]. Therefore, we can anticipate:

- **H10:** Social influence positively influences the intention to disclose personal information.

3.1 Hypotheses of RQ2 and RQ3

The primary objective of the study is to explore the effects of delivering privacy warnings on people's intentions to disclose their personal information to the robot and their actual disclosure during interaction with the robot.

Several studies have indicated that privacy warnings do not significantly impact trust [39, 48]. In an experiment involving warnings before downloading a PDF file, the majority of participants disregarded the warnings and trusted the antivirus program [48]. Moreover, the study by Carpenter et al. [39] revealed that the most trusted source of warning may not be the most effective for disclosure, suggesting that effective warnings should be evaluated based on disclosure outcomes rather than individuals' perceptions of trust alone. Studies by Wu et al. [67] and Liu et al. [73] have shown significant relationships between privacy notices and trust. Wu et al. [67] proposed that integrating privacy notices into

company website designs can enhance trust levels and customers' willingness to provide personal data. Pan and Zinkhan [74] discovered that in high-risk scenarios, consumers perceive greater trust when online privacy notices are present compared to when they are absent. According to the findings in Wogalter et al. [34], social influence represents an external factor linked to warnings, potentially influencing system usage. Observing others paying attention to a warning increases the likelihood of individuals attending to it themselves [34], which can affect their intention to disclose information.

In addition, the review in Section 2.3 confirms that privacy warnings reduce users' intentions to disclose personal information. LaRose and Rifon [36] found that privacy warnings heighten perceptions of privacy risks associated with information usage and decrease user disclosures on websites. Braunstein et al. [37] found that privacy warnings strongly influenced users' willingness to share personal information, affecting both the phrasing of questions and their self-reported sharing behavior.

Since our sample size is relatively small (approximately 55 individuals per group), an exploratory (non-conclusive) study is proposed at this stage. Therefore, to address RQ2, the following hypotheses are Working Hypotheses (WH) as proposed:

- **WH1:** The RB is a more influential antecedent of the ITDPI for participants who interact with the robot that delivers privacy warnings than for those who interact with the robot without privacy warnings.
- **WH2:** The TB is a more influential antecedent of the ITDPI for participants who interact with the robot that delivers privacy warnings than for those who interact with the robot without privacy warnings.
- **WH3:** TB is a more influential antecedent of the RB for participants who interact with the robot that delivers privacy warnings than for those who interact with the robot without privacy warnings.
- **WH4:** The PU is a more influential antecedent of the ITDPI for participants who interact with the robot that delivers privacy warnings than for those who interact with the robot without privacy warnings.
- **WH5:** The PENJ is a more influential antecedent of the ITDPI for participants who interact with the robot that delivers privacy warnings than for those who interact with the robot without privacy warnings.
- **WH6:** The PEOU is a more influential antecedent of the ITDPI for participants who interact with a robot that delivers privacy warnings than for those who interact with the robot without privacy warnings.
- **WH7:** The PEOU is a more influential antecedent of the PU for participants who interact with a robot that delivers privacy warnings than for those who interact with the robot without privacy warnings.
- **WH8:** The SI is a more influential antecedent of the ITDPI for participants who interact with the robot that delivers privacy warnings than for those who interact with the robot without privacy warnings.
- **WH9:** Participants who interact with the robot that delivers privacy warnings will intend to disclose their information less than those who interact with the robot without privacy warnings.

Findings from Carpenter et al. [38, 39] suggested that detailed consequence descriptions can evoke emotional responses, increase risk perception, and discourage users from disclosing excessive personal information online. In a subsequent work [40], the authors emphasized the importance of crafting customized warning messages to enhance privacy protection. Krol et al. [43] revealed that when users are given more control over what data to share, privacy warnings or security threats significantly reduce personal information disclosure. Mamonov et al. [44] found that users naturally take protective actions when aware of security risks, while Feri et al. [41] highlighted the impact of data breach notifications in reducing sensitive disclosures. Epstein et al. [42] further demonstrated that strong Terms of Service warnings and click requirements effectively decrease personal disclosures.

Moreover, while research on the impact of privacy warnings in HRI is limited, a study by Lee et al. [54] found that participants generally accepted data recording by the Snackbot if informed, though views

on accidental recordings were mixed, with some users comfortable with notifications and others concerned about potential misuse. Therefore, to address RQ3, we propose the following working hypothesis:

- **WH10:** Participants who interact with the robot that delivers privacy warnings will disclose their information less than those who interact with the robot without privacy warnings.

4 Research Methodology

To assess the research questions, we deployed two scenarios on a social robot (ARI)[1]: showing a recommended list of restaurants (Scenario 1) and taking a selfie with the robot (Scenario 2), in which the participants interacted with the robot and responded to the questions related to each scenario. In further detail, in Scenario 1, the robot asks questions related to the type and theme of the restaurant the participants would prefer, whether they follow a special diet, have food allergies, or are accompanied by family members or individuals with disabilities. After answering these questions, the robot provides restaurant recommendations based on the responses. In Scenario 2, the robot inquires about participants' social network usage, their favorite social networks, and whether they would like to take a selfie with the robot and share it on their profiles.

The questions of each scenario were categorized into three privacy levels: low, medium, and high. This categorization was based on the guidelines from the European Commission regarding sensitive information [75]. We classified questions that, for instance, inquire about specific dietary restrictions, allergies, or personal situations related to disabilities as high privacy due to their sensitive nature. Questions considered to have medium privacy include inquiries such as 'How much do you spend on each meal?' or 'How much time do you usually spend on social networks per day?' In contrast, questions like 'Are you interested in social networks?' or 'What type of theme would you prefer?' were classified as low privacy.

Participants were divided into two groups: the study group and the control group. The study group received privacy warnings before each high-privacy-level question, while the control group did not receive any warnings. Fig. 2 depicts a participant interacting with the robot while responding to the scenario questions. The scenario questions and a video of the interaction with the robot are available at: https://www.iri.upc.edu/groups/perception/#SecuRoPS. Further details about the procedure are outlined in the following subsection.

Fig. 2: A participant interacting with the robot

4.1 Procedure and setting

The experimental procedure of this study consists of three primary phases: Pre-Interaction, Interaction, and Post-Interaction, as illustrated in Fig. 3 and described below.

During the Pre-Interaction phase, participants received information about the research project, its objectives, and procedures, and were requested to sign a consent form. The estimated duration for this phase is approximately 5 minutes. In the Interaction phase, participants answered questions via the robot's touch screen, with the study group receiving a privacy warning before each

[1] https://pal-robotics.com/robots/ari/

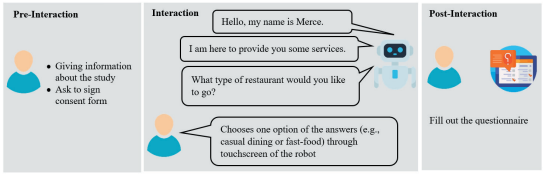

Fig. 3: Overview of the three-phase study procedure

Table 1: Demographic profile of the participants

Variable	Description	Frequency	Percentage
Gender	Female	53	46.9
	Male	58	51.3
	Prefer not to say	2	1.8
Age	18-24	5	4.4
	25-34	17	15.0
	35-44	36	31.9
	45-54	44	39.0
	More than 54	11	9.7

high-privacy-level question. This phase also lasted approximately 5 minutes per participant. Finally, in the Post-Interaction phase, participants completed a questionnaire involving factors affecting their intention to disclose personal information, with an estimated duration of approximately 10 minutes per participant.

The experiment took place at the public science festival (16a Festa de la Ciència) in Barcelona. The event was open and free to the general public, and participation in our experiment was available to all visitors over the age of 18. We collected data from those who were interested and voluntarily participated, totaling 113 individuals by the end of the festival. The demographic details of the participants are outlined in Table 1.

4.2 Measurement development

We employed a questionnaire comprising 32 statements to evaluate participants' intentions to disclose personal information. These statements formed eight constructs and were rated on a 7-point Likert scale ranging from (1) strongly disagree to (7) strongly agree, drawn from various sources [16, 22, 23, 66, 71]. Details of the constructs, their items, and their sources are outlined in Table 2.

To determine the participants' actual disclosure of personal information, we collected their responses and computed a metric named the Disclosure Index (DI) for the total number of questions answered by each participant. In addition, we should note that some questions followed a hierarchical format, meaning the subsequent questions for each participant may vary depending on their responses. For instance, if a participant answered "yes" to the question "Is there any food that you have an allergy to?", the following question would ask them to specify the food. If the answer was "no," the conversation would move to another topic, such as restaurant preferences. Therefore, the total number of questions presented to each participant can differ from others. Hence, we define DI as follows:

$$DI = \frac{NAQ}{TAQ} \tag{1}$$

where NAQ indicates the total Number of Answered Questions for each participant, whereas TAQ denotes the Total number of Asked Questions by the robot for each participant.

Similarly, we computed three metrics for each category of questions individually: DI_L (Disclosure Index for low privacy), D_M (Disclosure Index for medium privacy), and DI_H (Disclosure Index for high privacy). These metrics represent the ratio of the number of answered questions by the participant in each privacy category to the total number of questions asked by the robot for each participant within the corresponding category.

5 Experimental Results

In this section, we examine users' intentions to disclose personal information by assessing our proposed research model using data collected from participants' responses to the questionnaire. We assess the psychometric properties of dimensionality, reliability, and validity of the questionnaire (Subsection 5.1). Next, we evaluate the overall structural model and compare models between the study and control groups, addressing the validity of our hypotheses in Subsections 5.2 and 5.3. Finally, we analyze participants' actual disclosure

Table 2: Constructs and items of the questionnaire

Construct	Items	Code	Source
Risk Beliefs (RB)	• Providing the robot with personal information would involve many unexpected problems.	RB1	[71]
	• It would be risky to give personal information to the robot.	RB2	
	• There would be a high potential for loss in disclosing my personal information to the robot.	RB3	
Trusting Beliefs (TB)	• I think the robot is trustworthy in handling information.	TB1	[23]
	• I think robot tells the truth and fulfill promises related to information provided by me.	TB2	
	• I trust that the robot would keep my best interests in mind when dealing with information.	TB3	
	• I think the robot is in general predictable and consistent regarding the usage of information.	TB4	
	• I think the robot is honest with users when it comes to using the provided information.	TB5	
Perceived Usefulness (PU)	• I think the robot is useful to me.	PU1	[16, 66]
	• It would be nice to use the robot for booking from a restaurant.	PU2	
	• I think the robot can be used to book from a restaurant and do other things.	PU3	
Perceived Enjoyment (PENJ)	• It is fun to talk to the robot.	PENJ1	[16, 66]
	• It is fun to do things with the robot.	PENJ2	
	• The robot looks enjoyable.	PENJ3	
	• The robot seems charming.	PENJ4	
	• The robot seems boring.	PENJ5	
Perceived Ease of Use (PEOU)	• I immediately learned how to use the robot.	PEOU1	[16, 66]
	• The robot seems easy to use.	PEOU2	
	• I think I can use the robot without any help.	PEOU3	
	• I think I can use the robot with someone's help.	PEOU4	
	• I think I can use the robot if I have some good instructions.	PEOU5	

Continue on next page

Table 2: Constructs and items of the questionnaire (continued)

Construct	Items	Code	Source
Social Influence (SI)	• I think my friends would like me to use the robot.	SI1	[16, 66]
	• I think it would give a good impression if I use the robot.	SI2	
	• I think that people whose opinion I value would look favorably upon me using the robot.	SI3	
Intention To Disclose Personal Information (ITDPI)	• I am likely to disclose my personal information to the robot.	ITDPI1	[22]
	• I am willing to disclose my personal information to the robot.	ITDPI2	
	• Disclosing my personal information to the robot for its services is unlikely for me.	ITDPI3	

of personal information based on their responses during the interaction (Subsection 5.4).

5.1 Psychometric characteristics

We assessed the psychometric characteristics of dimensionality, reliability, and validity of the constructs using the average variance extracted method [76] as demonstrated in Table 3.

The Average Variance Extracted (AVE) measures the extent to which a group of items within a construct aligns with the concept it intends to measure. It assesses whether the variance explained by these items surpasses the associated error variance. An AVE value exceeding 0.50 signifies satisfactory convergent validity. As demonstrated in Table 3, all the constructs met this criterion. Another metric, Composite Reliability (CR) and Cronbach's alpha (Alpha), assesses the consistency with which a group of items within a construct measures the same concept. They signify the effectiveness of these items in measuring the construct, with a value ideally exceeding 0.70, as depicted in Table 3. Factor loading, another metric, measures the correlation between an item and a construct, revealing how effectively the item reflects or is linked to the underlying construct. It indicates the strength and direction of this relationship, with each item having a value ideally exceeding 0.60 [77]. Moreover, T represents the ratio of the variance in means to the standard deviation of the variance in means, to determine the statistical significance of the variance between a sample mean and a population mean. As depicted in Table 3, all factor loading values surpassed 0.6, while T exhibited a significantly high value in accordance with the existing literature [77].

Table 4 presents the discriminant validity of the quality dimensions, assessed using AVE [76]. This indicates that a construct should have more variance in common with its indicators than with other constructs in the model. This is confirmed when the square root of the AVE for each construct (represented on the diagonal of the matrix) is greater than its correlation with the other constructs (represented by the remaining values of each row). The findings in Table 4 support this criterion.

5.2 Results of the general structural model

We utilized Structural Equation Modeling (SEM) to examine the intention to disclose personal information, assessing the psychometric properties of the items, model modifications, causal connections between constructs, and hypotheses. This involved estimating an SEM model using variance and covariance matrices through Maximum Likelihood Estimation (MLE) with EQS6 [78]. In other words, the SEM determines the impact of various

Table 3: Analysis of the dimensionality, reliability, and validity of the items.

	Factor loading	**T**	**Mean**	**SD**
Risk Beliefs (AVE: 0.66; CR: 0.82; Alpha: 0.78)				
RB1	0.7	2.97	4.15	1.73
RB2	0.88	5.47	4.46	1.73
RB3	0.74	3.72	3.80	1.67
Trusting Beliefs (AVE: 0.66; CR: 0.88; Alpha: 0.88)				
TB1	0.64	5.65	5.04	1.45
TB2	0.79	8.25	5.27	1.43
TB3	0.86	8.54	5.39	1.42
TB4	0.81	8.10	5.21	1.33
TB5	0.76	8.11	5.27	1.41
Perceived Usefulness (AVE: 0.66; CR: 0.82; Alpha: 0.87)				
PU1	0.77	9.38	4.72	1.59
PU2	0.83	10.18	5.41	1.71
PU3	0.74	7.12	5.69	1.59
Perceived Enjoyment (AVE: 0.71; CR: 0.91; Alpha: 0.90)				
PENJ1	0.86	12.45	4.89	1.82
PENJ2	0.80	10.00	5.19	1.58
PENJ3	0.75	6.87	5.66	1.44
PENJ4	0.86	11.52	4.94	1.82
PENJ5	0.79	12.46	4.48	1.83
Perceived Ease of Use (AVE: 0.66; CR: 0.88; Alpha: 0.874)				
PEOU1	0.69	5.00	6.21	1.27
PEOU2	0.73	5.27	6.11	1.45
PEOU3	0.79	6.75	6.08	1.43
PEOU4	0.87	5.54	5.55	1.01
PEOU5	0.78	5.25	6.04	1.16
Social Influence (AVE: 0.64; CR: 0.81; Alpha: 0.80)				
SI1	0.79	8.88	4.96	1.46
SI2	0.64	6.16	4.53	1.49
SI3	0.85	11.40	4.95	1.63
Intention to Disclose Personal Information (AVE: 0.82; CR: 0.93; Alpha: 0.92)				
ITDPI1	0.9	14.75	4.27	1.78
ITDPI2	0.91	14.66	4.16	1.77
ITDPI3	0.89	13.12	4.28	1.58

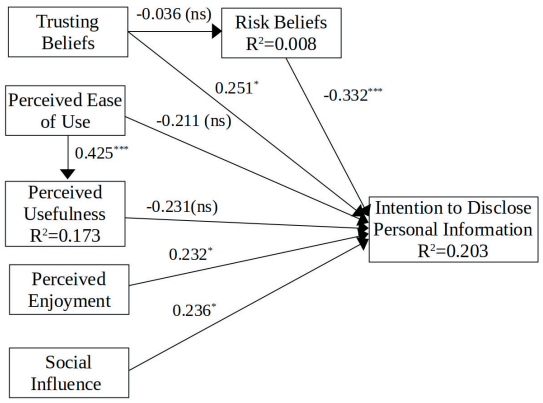

Fig. 4: General structure model

Note: ns denotes no significance, * denotes $p < .05$, ** denotes $p < .01$, *** denotes $p < .001$

constructs on their corresponding dependent variables. Fig. 4 illustrates the causal relationships among the constructs in the general model.

The respective goodness-of-fit (R^2) values were 0.008 for RB, 0.173 for PU, and 0.203 for ITDPI. Among the relationships in the general model, five out of ten reached statistical significance ($p < .05$), confirming hypotheses H1, H2, H6, H8, and H10 (refer to Table 5). However, three relations ($TB \rightarrow RB$; $\beta = -0.036$), ($PU \rightarrow ITDPI$; $\beta = -0.231$), and ($PEOU \rightarrow ITDPI$; $\beta = -0.211$) did not achieve statistical significance. Hence, hypotheses H3, H4, H5, H7, and H9 are rejected. In addition, while PEOU ($\beta = 0.425$, $p < .001$) has the highest weight value, it cannot be considered an influential construct. Thus, RB ($\beta = -0.332$, $p < .001$) was the most influential construct directly influencing ITDPI, followed by TB ($\beta = 0.251$, $p < .05$) and SI ($\beta = 0.236$, $p < .05$).

5.3 Results of groups models

Fig. 5 and Fig. 6 illustrate the causal relationships among the constructs in the control and study groups, respectively. The R^2 values obtained align with the sample sizes: $R^2 = 0.378$ for the ITDPI in the study group and $R^2 = 0.083$ for the ITDPI in the control group. Notably, when the robot was equipped with privacy warnings, the R^2 values significantly increased to explain the ITDPI ($R^2_{study-group} - R^2_{control-group} = 0.295$), of which 29.5% of the variance was explained. Additionally, after applying Fisher's transformation and assessing the difference in correlations, we determined that this difference was statistically significant ($z = 1.655$, $p < .05$). Thus, it can be inferred that delivering privacy warnings by the robot enhances its predictive power in explaining acceptance compared to when no privacy warnings are delivered.

Furthermore, the results reveal that, in the control and study groups' models, three and five relationships, respectively, achieved statistical significance ($p < .05$). Both models demonstrated significant relationships for $RB \rightarrow ITDPI$ and $PEOU \rightarrow PU$, while $TB \rightarrow RB$, $PEOU \rightarrow ITDPI$, and $PU \rightarrow ITDPI$ did not reach statistically significant. However, in terms of relationship weights, within the control model, the highest value was observed for PEOU mediated by PU ($\beta = 0.475$, $p < .001$), it did not emerge as the most influential construct. Thus, for participants interacting with a robot without privacy warnings, the ITDPI was primarily influenced by SI ($\beta = 0.275$, $p < .05$), followed by RB ($\beta = -0.269$, $p < .05$). Conversely, when participants interacted with a robot equipped with privacy warnings, their intention to disclose personal information was influenced mainly by RB ($\beta = -0.418$, $p < .001$), followed by TB ($\beta = 0.364$, $p < .01$), and PENJ ($\beta = 0.231$, $p < .05$).

In line with our proposed working hypotheses outlined in subsection 3.1, when the robot is equipped with privacy warnings, TB and PENJ demonstrate greater discriminatory effects on ITDPI compared to a robot without privacy warnings, indicating potential support for WH2 and WH5. However, contrary to our expectations for working hypothesis WH7, when participants interacted with a robot without privacy warnings, PEOU had a greater influence on PU ($PEOU \rightarrow PU$) than when privacy warnings were present. For WH1, the value of $RB \rightarrow ITDPI$ for the study group exceeded that of the control group; however, due to insufficient sample size, WH1 does not seem to be accepted. Furthermore, since the relationships $TB \rightarrow RB$, $PU \rightarrow ITDPI$, and $PEOU \rightarrow ITDPI$ did not achieve significance in either group, hypotheses WH3, WH4, and WH6 were rejected. Finally, WH8 was also rejected, as the SI did not reach statistical significance for the study group.

Table 4: Discriminant validity.

	RB	TB	PU	PENJ	PEOU	SI	ITDPI
RB	0.81						
TB	0.05	0.81					
PU	0.02	0.63***	0.81				
PENJ	-0.10	0.54***	0.53***	0.84			
PEOU	−0.23*	0.61***	0.50***	0.36**	0.81		
SI	-0.10	0.49***	0.54***	0.55***	0.33*	0.80	
ITDPI	−0.50***	0.10	0.02	0.18	-0.16	0.27*	0.91

Table 5: Hypotheses validation for general model

Hypothesis	Independent variable	Dependent variable	Beta[1]	T	Result
H1	RB	ITDPI	−0.332***	-3.856	✓
H2	TB	ITDPI	0.251*	2.328	✓
H3	TB	RB	-0.036(ns)	-0.376	✗
H5	PU	ITDPI	-0.231(ns)	-1.847	✗
H6	PENJ	ITDPI	0.232*	1.996	✓
H7	PEOU	ITDPI	-0.211(ns)	-1.751	✗
H8	PEOU	PU	0.425***	4.940	✓
H10	SI	ITDPI	0.236*	2.098	✓

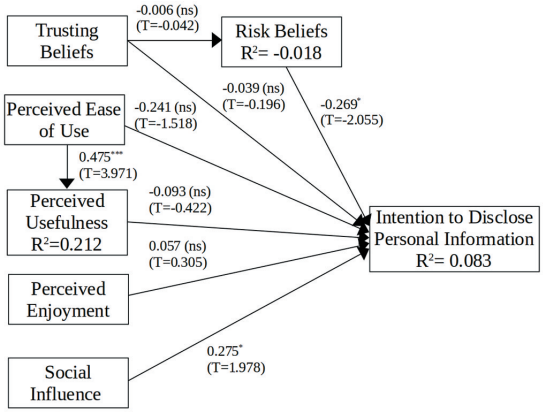

Fig. 5: Control model

Note: ns denotes no significance, * denotes $p < .05$, ** denotes $p < .01$, *** denotes $p < .001$

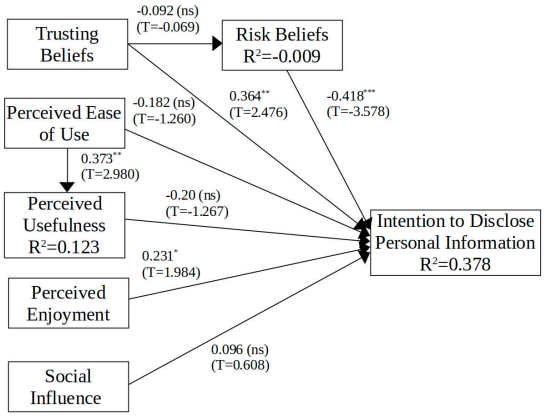

Fig. 6: Study model

Note: ns denotes no significance, * denotes $p < .05$, ** denotes $p < .01$, *** denotes $p < .001$

5.4 Results of the disclosure

In this section, we examine the participants' intentions to disclose personal information to the robot in both groups (WH9) and assess whether there are differences in the actual disclosure between participants who are interacting with a robot delivering privacy warnings and those who are interacting with a robot without privacy warnings (WH10). To achieve this, we utilized the questionnaire results to evaluate the ITDPI and computed measures such as *DI*, *DI_L*, *DI_M*, and *DI_H* to

assess the actual disclosure, as outlined in Section 4.2.

The average values of the measurements are depicted in Fig. 7. According to the results of the ANOVA single factor test, although the average value of the ITDPI for participants interacting with the robot equipped with privacy warnings (M_{study} = 4.11, 95% CI=[3.89, 4.34]) is slightly less than that for those interacting with the robot without privacy warnings ($M_{control}$ = 4.17, 95% CI=[3.84, 4.50]), this difference did not reach statistical significance ($F = 0.08$, $p = .78$). Hence, the findings suggest that participants who interact with the robot delivering privacy warnings do not exhibit a reduced intention to disclose their information compared to those who interact with the robot without privacy warnings, which contradicts WH9.

Regarding the *DI* values, there was statistical significance in terms of the *DI_L* between participants who interacted with the robot with privacy warning messages and those who interacted with the robot without privacy warning ($M_{study} = 0.89$, 95% $CI = [0.87, 0.91]$ *vs.* $M_{control} = 0.95$, 95% $CI = [0.94, 0.97]$, $F = 29.36$, $p < .001$). However, we did not find statistical significance in the *DI* ($M_{study} = 1.0$, 95% $CI = [1.0, 1.0]$ *vs.* $M_{control} = 0.995$, 95% $CI = [0.99, 1.0]$, $F = 1.47$, $p = .23$), *DI_M* ($M_{study} = 0.87$, 95% $CI = [0.85, 0.89]$ *vs.* $M_{control} = 0.87$, 95% $CI = [0.85, 0.89]$, $F = 0.004$, $p = .95$), and *DI_H* ($M_{study} = 0.59$, 95% $CI = [0.55, 0.63]$ *vs.* $M_{control} = 0.56$, 95% $CI = [0.52, 0.60]$, $F = 0.91$, $p = .34$). Overall, however, the results show that the (*DI_L*) for the study group is less than the (*DI_L*) for the control group, and the (*DI* and *DI_M*) for both groups are almost equal. It seems that the participants who interact with the robot with privacy warnings do not disclose their information less than those who interact with the robot without privacy warnings, suggesting that WH10 is not supported.

6 Discussion and Conclusion

In this paper, we explored how privacy warnings can impact people's intentions to disclose their personal information when interacting with a robot in public spaces and assessed their actual disclosure behavior. As prior research has not extensively covered this area, we proposed a model

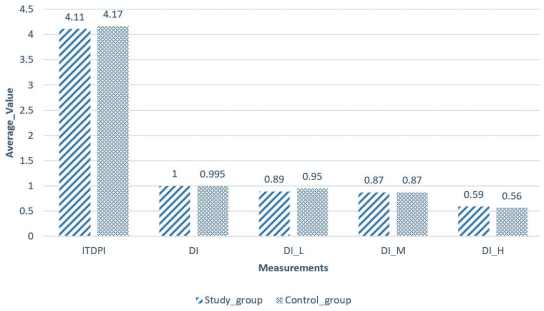

Fig. 7: The average value of the measurements

to estimate the users' intention to disclose personal information to the robot based on existing models [20, 23, 25] and a modified version of the UTAUT model [16, 66] (see Section 3). The robot was deployed to provide two scenarios to the participants: showing a recommended list of restaurants and taking a selfie with the robot. We conducted our experiment on 113 participants in two groups, one interacting with a robot with privacy warnings and another interacting with a robot without privacy warnings at a public science festival.

To address RQ1 and based on the proposed estimation model, the intention to disclose personal information serves as the dependent construct, determined by six constructs: trusting beliefs, risk beliefs, perceived usefulness, perceived enjoyment, perceived ease of use, and social influence. The experimental findings revealed variations in the impacts of the constructs on their respective dependent constructs, identifying those that attained statistical significance, the direction of their influence (positive or negative), and their weight values. In the general structural model (SEM), RB was identified as the most influential construct, exhibiting a negative value that confirmed hypothesis H1. This finding is in line with the studies [22, 23, 63]. RB displayed an intermediate weight, which was more relevant than that in [22, 63], but less relevant than that in [23]. Nonetheless, the studies [21, 24, 26, 70] did not achieve significant results for the same construct. TB is identified as the second most influential construct, confirming hypothesis H2. Its weight aligns with findings in [23], although it is lower than that reported in [63]. Nonetheless, TB did not reach statistical significance in [24, 70], and it achieved a negative weight as reported in [20]. Moreover,

the findings revealed that TB did not affect RB, thus H3 and H4 were rejected. Neither PEOU nor PU influenced ITDPI, resulting in the rejection of hypotheses H5 and H7, which aligns with findings from the studies [20, 25]. However, PU played a mediating role between PEOU and ITDPI, aligning with findings from [16, 66], and supporting hypothesis H8. In addition, PENJ, another influential construct with a positive weight, supports hypothesis H6, which aligns with the results of the studies [25, 72]. It should be noted that PENJ did not reach statistical significance in [20]. Another influential construct, SI, supports hypothesis H10, as confirmed by the results reported in studies [26, 70]. However, studies such as [21, 79] indicated that SI did not reach statistical significance.

To address RQ2, we examined the structural models of the study and control groups. The findings revealed a significantly greater R^2 value for ITDPI when the robot delivered privacy warnings during the interaction, suggesting a 29.5% improvement in predictive power compared to when no privacy warnings were presented. According to the structural models of the two groups and to address our working hypotheses, we found that participants who interacted with a robot equipped with privacy warnings perceived a greater level of risk than did those who interacted with a robot without privacy warnings, which aligns with findings from [36, 38]. However, due to the insufficient sample size, WH1 does not seem to be supported. In addition, we found that a robot delivering privacy warnings applied more TB, consistent with the studies [67, 73, 74] and was perceived as more enjoyable than a robot without privacy warnings, potentially supporting WH2 and WH5. Moreover, relationships involving PEOU and PU ITDPI, as well as TB with RB, did not reach statistical significance in either model, indicating a lack of support for WH3, WH4, and WH6. Contrary to our expectations outlined in WH7, when participants interacted with a robot without privacy warnings, PEOU had a greater influence on PU than in situations where privacy warnings were present, suggesting a lack of support for WH7. Furthermore, SI seemed to influence only those participants who interacted with a robot without privacy warnings, contradicting the findings of [34], and thus failing to support WH8. Therefore, concerning RQ2, it can be concluded that a robot equipped with privacy warnings led to a greater overall acceptance among participants than a robot without privacy warnings.

However, for RQ3, the average ITDPI values for participants interacting with the robot equipped with privacy warnings were slightly lower than those for participants interacting with the robot without privacy warnings; this difference was not statistically significant. Thus, participants who were interacting with the robot with privacy warnings did not demonstrate a decreased intention to disclose their information compared to those interacting with the robot without privacy warnings, contrary to WH9. In addition, the disclosure results during the interactions with the robot revealed that the participants who were presented with privacy warnings disclosed low-privacy-level information to the robot less than those who interacted with the robot without privacy warnings. However, contrary to WH10, participants who interacted with the robot equipped with privacy warnings did not disclose their information less than did those who interacted with the robot without privacy warnings, which means that privacy warnings did not effectively reduce disclosure, which aligns with the findings from [45–48].

There are numerous reasons behind the participants' responses to privacy warning messages. Based on our observations, participants sometimes rapidly responded to questions and ignored warnings, consistent with findings from [45], possibly due to trust in the system as suggested by [48]. Alternatively, this behavior may occur because warnings do not accurately align with users' risk perceptions, as noted in studies by Egelman et al. [46, 47]. Moreover, participants might lack knowledge of what information could be misused, perceive the questions as not personally relevant, or exhibit optimistic bias, believing that negative events are less likely to affect them than others, as described in [50].

Overall, this study makes several significant contributions: 1) It addresses a gap in the existing literature by investigating the impact of privacy warnings on disclosure behavior within HRI. While prior studies have primarily focused on robot usage intentions and the willingness to share information, this research proposes a model within the frameworks of TAM and the UTAUT. 2) The findings reveal that the effectiveness of privacy warnings is contingent upon their design

and alignment with user expectations, emphasizing the complexity of this relationship. 3) This study suggests that while privacy warnings may enhance perceived risk, trust beliefs, and enjoyment, they do not necessarily lead to a reduction in actual disclosure behavior, aligning with previous research on the complex relationship between privacy concerns and user behavior [47–50].

We aim to explore these dynamics further in future studies, particularly focusing on the effects of designing different forms of privacy notifications on user trust and actual disclosure behaviors in varying contexts.

7 Limitations and Future Works

While this study provides valuable insights, several limitations should be acknowledged, which may guide future research. Key limitations include:

First, although the public science festival where the experiment was conducted was open to all with free admission and not specifically focused on any scientific field, a limitation arises from the nature of the sample. It was drawn from attendees who were present and willing to engage, potentially introducing some bias. Since participation was voluntary, individuals with particular interests, motivations, or openness to the event's activities may have been overrepresented. This limits the generalizability of the findings, as the sample might not fully represent the broader population. Moreover, those who chose not to engage could hold different perspectives that were not captured, contributing to further bias in the results.

Secondly, although the robot used in the experiment was equipped with advanced functionalities like speech recognition and dialogue management, we opted not to use them. While these features could have offered a more natural and interactive experience for participants, the outdoor setting of the experiment, with over 1,000 attendees, presented a challenge. The expected level of background noise in a public space was likely to interfere with the accuracy of speech recognition, influencing our decision to rely on touchscreen interactions instead.

Lastly, the duration of the experiment presented a limitation. The model used in the study was based on 7 constructs with a total of 27 items, requiring participants to answer 27 questions after interacting with the robot. This made the experiment lengthier—around 20 minutes per participant—which not only impacted the number of participants we could recruit but may have also affected participant engagement and interest as the session progressed.

In light of our findings and also mentioned limitations, our future research will focus on enhancing the design and effectiveness of privacy warnings to better understand their impact on user behavior and trust. However, while some research suggests that privacy warnings can reduce users' willingness to share personal data [17, 39, 40], conflicting evidence exists [45–48]. For instance, findings from this study indicate that privacy warnings may not always effectively reduce disclosure; in some cases, they can even lead to increased information sharing despite these warnings [49, 50]. A key finding from this study and prior research [40, 47] is that the format and design of privacy warnings play a critical role in their effectiveness—an aspect we aim to explore further in future work.

Research indicates that transparency in AI systems and HRI has complex effects on user trust and privacy concerns, which can enhance understanding and accountability. In our upcoming studies, we aim to revisit the design of privacy notices by developing varying levels of transparency and providing users with greater control over the data they share, including informing participants about the sensitivity of the information collected by the robot during interactions. This approach aligns with findings that underscore the significance of transparency in shaping user trust and acceptance of intelligent systems [80–82].

Furthermore, we found that factors such as trusting beliefs and risk perceptions are crucial factors influencing individuals' intentions to disclose personal information, as enhancing trust or reducing perceived risks can significantly increase the willingness to share sensitive data. To streamline future experiments, we propose focusing on these key elements, which would not only shorten the duration of the study but also improve the participant experience and yield deeper insights into the dynamics of trust and privacy in HRI.

Additionally, we plan to explore incorporating various interaction modalities beyond the touchscreen to further enhance the effectiveness of privacy warnings and promote transparency, addressing limitations identified in the current study.

Therefore, building on our findings and the limitations identified in this study, our future research will concentrate on refining the design and effectiveness of privacy warnings, enhancing transparency in AI systems, and understanding the interplay of trust and risk perceptions to improve user engagement and data sharing behavior.

Declarations

- Funding

  This work was partially funded by the Research Council of Norway under the project SECUROPS (INT-NO/0875); by MCIN/ AEI /10.13039/501100011033, by the "European Union NextGenerationEU/PRTR" under the project ROB-IN (PLEC2021-007859) and by the European Union under the project ARISE (HORIZON-CL4-2023-DIGITAL-EMERGING-01101135959).

- Conflict of interest/Competing interests

  The authors declare that they have no conflicts of interest or competing interests.

- Ethics approval and consent to participate

  Our study was approved by the Ethics Committee of the Spanish National Research Council (CSIC) with the code number 111/2023. The bioethical aspects of the proposed research have been evaluated (research with the participation of human beings, the handling of their samples, and/or data that require protection), and, according to the terms defined in the project, the CSIC Ethics Committee declares that there are no objections that could constitute any impediment to its development. In addition, informed consent was obtained from all individual participants included in the study.

- Consent for publication

  The participants consented to the dissemination of the research results for publication in scientific journals and conferences.

- Data availability

  Although the gathered data are anonymized, in compliance with Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016, on the protection of personal data and the GDPR principles stated in ethical approval, the collected data cannot be made available.

- Materials availability

  The questionnaire, scenario questions, and a video of the interaction with the robot are available at https://www.iri.upc.edu/groups/perception/#SecuRoPS.

- Author contributions

  All the authors contributed to the study's conception and design. Material preparation, robot programming, and data collection were performed by Azra Aryania. Data analysis was performed by Ruben Huertas-Garcia and Santiago Forgas-Coll. The first draft of the manuscript was written by Azra Aryania, and all the authors commented on previous versions of the manuscript. All the authors have read and approved the final manuscript.

References

[1] Cobo Hurtado, L., Viñas, P.F., Zalama, E., Gómez-García-Bermejo, J., Delgado, J.M., Vielba García, B.: Development and Usability Validation of a Social Robot Platform for Physical and Cognitive Stimulation in Elder Care Facilities. Healthcare **9**(8), 1067 (2021) https://doi.org/10.3390/healthcare9081067

[2] Civit, A., Andriella, A., Barrue, C., Antonio, M., Boque, C., Alenyà, G.: Introducing social robots to assess frailty in older adults. In: ACM/IEEE International Conference on Human-Robot Interaction, pp. 342–346 (2024). https://doi.org/10.1145/3610978.3640660

[3] Barrue, C., Suárez, A., Inzitari, M., Ribera, A., Alenyà, G.: Nyam: the role of configurable engagement strategies in robotic-assisted feeding. In: ACM/IEEE International Conference on Human-Robot Interaction, pp. 228–232 (2024). https://doi.org/10.1145/3610978.3640691

[4] Konijn, E.A., Hoorn, J.F.: Robot tutor

and pupils' educational ability: Teaching the times tables. Computers & Education **157**, 103970 (2020) https://doi.org/10.1016/j.compedu.2020.103970

[5] Ceha, J., Law, E., Kulić, D., Oudeyer, P.-Y., Roy, D.: Identifying Functions and Behaviours of Social Robots for In-Class Learning Activities: Teachers' Perspective. International Journal of Social Robotics **14**(3), 747–761 (2022) https://doi.org/10.1007/s12369-021-00820-7

[6] Clabaugh, C., Mahajan, K., Jain, S., Pakkar, R., Becerra, D., Shi, Z., Deng, E., Lee, R., Ragusa, G., Matarić, M.: Long-Term Personalization of an In-Home Socially Assistive Robot for Children With Autism Spectrum Disorders. Frontiers in Robotics and AI **6**, 110 (2019) https://doi.org/10.3389/frobt.2019.00110

[7] Leung, A.Y.M., Zhao, I.Y., Lin, S., Lau, T.K.: Exploring the Presence of Humanoid Social Robots at Home and Capturing Human-Robot Interactions with Older Adults: Experiences from Four Case Studies. Healthcare **11**(1), 39 (2022) https://doi.org/10.3390/healthcare11010039

[8] Aryania, A., Huertas-Garcia, R., Forgas-Coll, S., Angulo, C., Alenya, G.: How do people intend to disclose personal information to a social robot in public spaces?, Pasadena, CA, USA, pp. 1809–1814 (2024). https://doi.org/10.1109/RO-MAN60168.2024.10731370

[9] Tsai, Y.-L., Wadgaonkar, C., Chun, B., Knight, H.: How Service Robots Can Improve Workplace Experience: Camaraderie, Customization, and Humans-in-the-Loop. International Journal of Social Robotics **14**(7), 1605–1624 (2022) https://doi.org/10.1007/s12369-022-00898-7

[10] Ren, X., Guo, Z., Huang, A., Li, Y., Xu, X., Zhang, X.: Effects of Social Robotics in Promoting Physical Activity in the Shared Workspace. Sustainability **14**(7), 4006 (2022) https://doi.org/10.3390/su14074006

[11] Love, T., Andriella, A., Alenyà, G.: Towards explainable proactive robot interactions for groups of people in unstructured environments. In: ACM/IEEE International Conference on Human-Robot Interaction, pp. 697–701 (2024). https://doi.org/10.1145/3610978.3640734

[12] Fuentes-Moraleda, L., Lafuente-Ibañez, C., Fernandez Alvarez, N., Villace-Molinero, T.: Willingness to accept social robots in museums: an exploratory factor analysis according to visitor profile. Library Hi Tech **40**(4), 894–913 (2022) https://doi.org/10.1108/LHT-07-2020-0180

[13] Rossi, A., Staffa, M., Origlia, A., Maro, M., Rossi, S.: BRILLO: A Robotic Architecture for Personalised Long-lasting Interactions in a Bartending Domain. In: Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, pp. 426–429. ACM, Boulder CO USA (2021). https://doi.org/10.1145/3434074.3447206

[14] John, N.E., Rossi, A., Rossi, S.: Personalized Human-Robot Interaction with a Robot Bartender. In: Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization, pp. 155–159. ACM, Barcelona Spain (2022). https://doi.org/10.1145/3511047.3537686

[15] Thunberg, S., Ziemke, T.: Are People Ready for Social Robots in Public Spaces? In: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, pp. 482–484. ACM, Cambridge United Kingdom (2020). https://doi.org/10.1145/3371382.3378294

[16] Heerink, M., Kröse, B., Evers, V., Wielinga, B.: Assessing Acceptance of Assistive Social Agent Technology by Older Adults: the Almere Model. International Journal of Social Robotics **2**(4), 361–375 (2010) https://doi.org/10.1007/s12369-010-0068-5

[17] Lutz, C., Tamó-Larrieux, A.: The Robot Privacy Paradox: Understanding How Privacy Concerns Shape Intentions to Use Social Robots. Human-Machine Communication **1**, 87–111 (2020) https://doi.org/10.

30658/hmc.1.6

[18] Lutz, C., Schöttler, M., Hoffmann, C.P.: The privacy implications of social robots: Scoping review and expert interviews. Mobile Media & Communication **7**(3), 412–434 (2019) https://doi.org/10.1177/2050157919843961

[19] Rueben, M., Aroyo, A.M., Lutz, C., Schmolz, J., Van Cleynenbreugel, P., Corti, A., Agrawal, S., Smart, W.D.: Themes and Research Directions in Privacy-Sensitive Robotics. In: 2018 IEEE Workshop on Advanced Robotics and Its Social Impacts (ARSO), pp. 77–84. IEEE, Genova, Italy (2018). https://doi.org/10.1109/ARSO.2018.8625758

[20] McKnight, D.H., Lankton, N., Tripp, J.: Social Networking Information Disclosure and Continuance Intention: A Disconnect. In: 2011 44th Hawaii International Conference on System Sciences, pp. 1–10. IEEE, Kauai, HI (2011). https://doi.org/10.1109/HICSS.2011.379

[21] Xu, F., Michael, K., Chen, X.: Factors affecting privacy disclosure on social network sites: an integrated model. Electronic Commerce Research **13**(2), 151–168 (2013) https://doi.org/10.1007/s10660-013-9111-6

[22] Wang, T., Duong, T.D., Chen, C.C.: Intention to disclose personal information via mobile applications: A privacy calculus perspective. International Journal of Information Management **36**(4), 531–542 (2016) https://doi.org/10.1016/j.ijinfomgt.2016.03.003

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

[24] Harborth, D., Pape, S.: How Privacy Concerns, Trust and Risk Beliefs, and Privacy Literacy Influence Users' Intentions to Use Privacy-Enhancing Technologies: The Case of Tor. ACM SIGMIS Database: the DATABASE for Advances in Information Systems **51**(1), 51–69 (2020) https://doi.org/10.1145/3380799.3380805

[25] Song, C.S., Kim, Y.-K.: Predictors of consumers' willingness to share personal information with fashion sales robots. Journal of Retailing and Consumer Services **63**, 102727 (2021) https://doi.org/10.1016/j.jretconser.2021.102727

[26] Fan, A., Wu, Q., Yan, X., Lu, X., Ma, Y., Xiao, X.: Research on Influencing Factors of Personal Information Disclosure Intention of Social Media in China. Data and Information Management **5**(1), 195–207 (2021) https://doi.org/10.2478/dim-2020-0038

[27] Davis, F.D.: Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly **13**(3), 319 (1989) https://doi.org/10.2307/249008

[28] Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: Extrinsic and Intrinsic Motivation to Use Computers in the Workplace. Journal of applied social psychology **22**(14), 1111–1132 (1992)

[29] van der Heijden: User Acceptance of Hedonic Information Systems. MIS Quarterly **28**(4), 695 (2004) https://doi.org/10.2307/25148660

[30] Venkatesh, V., Davis, F.D.: A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science **46**(2), 186–204 (2000) https://doi.org/10.1287/mnsc.46.2.186.11926

[31] Venkatesh, V., Bala, H.: Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences **39**(2), 273–315 (2008) https://doi.org/10.1111/j.1540-5915.2008.00192.x

[32] Venkatesh, Morris, Davis, Davis: User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly **27**(3), 425 (2003) https://doi.org/10.2307/30036540

[33] Turja, T., Aaltonen, I., Taipale, S., Oksanen, A.: Robot acceptance model for care (RAM-care): A principled approach to the intention to use care robots. Information & Management **57**(5), 103220 (2020) https://doi.org/10.1016/j.im.2019.103220

[34] Wogalter, M.S., Mayhorn, C.B., Laughery, K.R.: Warnings and Hazard Communications. In: Handbook of Human Factors and Ergonomics, pp. 644–667. Wiley, New York (2021)

[35] Roussea, G.K., Wogalte, M.S.: Research on warning signs. In: Handbook of Warnings, pp. 147–158. LAWRENCE ERLBAUM ASSOCIATES PUBLISHERS, Mahwah New Jersey (2006)

[36] Larose, R., Rifon, N.J.: Promoting i -Safety: Effects of Privacy Warnings and Privacy Seals on Risk Assessment and Online Privacy Behavior. Journal of Consumer Affairs **41**(1), 127–149 (2007) https://doi.org/10.1111/j.1745-6606.2006.00071.x

[37] Braunstein, A., Granka, L., Staddon, J.: Indirect content privacy surveys: Measuring privacy without asking about it. SOUPS 2011 - Proceedings of the 7th Symposium on Usable Privacy and Security (2011) https://doi.org/10.1145/2078827.2078847 . ISBN: 9781450309110

[38] Carpenter, S., Zhu, F., Kolimi, S.: Reducing online identity disclosure using warnings. Applied Ergonomics **45**(5), 1337–1342 (2014) https://doi.org/10.1016/j.apergo.2013.10.005

[39] Carpenter, S., Zhu, F., Zeng, M., Shreeves, M.: Expert Sources in Warnings May Reduce the Extent of Identity Disclosure in Cyber Contexts. International Journal of Human–Computer Interaction **33**(3), 215–228 (2017) https://doi.org/10.1080/10447318.2016.1232909

[40] Carpenter, S., Shreeves, M., Brown, P., Zhu, F., Zeng, M.: Designing Warnings to Reduce Identity Disclosure. International Journal of Human–Computer Interaction **34**(11), 1077–1084 (2018) https://doi.org/10.1080/10447318.2017.1413792

[41] Feri, F., Giannetti, C., Jentzsch, N.: Disclosure of personal information under risk of privacy shocks. Journal of Economic Behavior & Organization **123**, 138–148 (2016) https://doi.org/10.1016/j.jebo.2015.12.001

[42] Epstein, R., Zankich, V.R.: The surprising power of a click requirement: How click requirements and warnings affect users' willingness to disclose personal information. PLOS ONE **17**(2), 0263097 (2022) https://doi.org/10.1371/journal.pone.0263097

[43] Krol, K., Preibusch, S.: Control versus Effort in Privacy Warnings for Webforms. In: Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society, pp. 13–23. ACM, Vienna Austria (2016). https://doi.org/10.1145/2994620.2994640

[44] Mamonov, S., Benbunan-Fich, R.: The impact of information security threat awareness on privacy-protective behaviors. Computers in Human Behavior **83**, 32–44 (2018) https://doi.org/10.1016/j.chb.2018.01.028

[45] Wu, M., Miller, R.C., Garfinkel, S.L.: Do security toolbars actually prevent phishing attacks? In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 601–610. ACM, Montréal Québec Canada (2006). https://doi.org/10.1145/1124772.1124863

[46] Egelman, S., Cranor, L.F., Hong, J.: You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1065–1074. ACM, Florence Italy (2008). https://doi.org/10.1145/1357054.1357219

[47] Egelman, S., Schechter, S.: The Importance of Being Earnest [In Security Warnings]. In: Financial Cryptography and Data Security: 17th International Conference, FC 2013, vol. 7859, pp. 52–59. Lecture Notes in Computer Science, Okinawa, Japan (2013). https://doi.org/10.1007/978-3-642-39884-1

[48] Krol, K., Moroz, M., Sasse, M.A.: Don't work. Can't work? Why it's time to rethink security warnings. In: 2012 7th International Conference on Risks and Security of Internet and Systems (CRiSIS), pp. 1–8. IEEE, Cork, Ireland (2012). https://doi.org/10.1109/CRISIS.2012.6378951

[49] Zhang, B., Wu, M., Kang, H., Go, E., Sundar, S.S.: Effects of security warnings and instant gratification cues on attitudes toward mobile websites. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 111–114. ACM, Toronto Ontario Canada (2014). https://doi.org/10.1145/2556288.2557347

[50] Junger, M., Montoya, L., Overink, F.-J.: Priming and warnings are not effective to prevent social engineering attacks. Computers in Human Behavior **66**, 75–87 (2017) https://doi.org/10.1016/j.chb.2016.09.012

[51] H. Smith, J., Dinev, T., Xu, H.: Information Privacy Research: An Interdisciplinary Review. MIS Quarterly **35**(4), 989 (2011) https://doi.org/10.2307/41409970

[52] Kokolakis, S.: Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. Computers & Security **64**, 122–134 (2017) https://doi.org/10.1016/j.cose.2015.07.002

[53] Martin, K.E.: Privacy Notices as Tabula Rasa: An Empirical Investigation into How Complying with a Privacy Notice is Related to Meeting Privacy Expectations Online. SSRN Electronic Journal (2014) https://doi.org/10.2139/ssrn.2518581

[54] Lee, M.K., Tang, K.P., Forlizzi, J., Kiesler, S.: Understanding users' perception of privacy in human-robot interaction. In: Proceedings of the 6th International Conference on Human-robot Interaction, pp. 181–182. ACM, Lausanne Switzerland (2011). https://doi.org/10.1145/1957656.1957721

[55] Sakai, K., Mitsuno, S., Ban, M., Yoshikawa, Y., Shimpo, F., Harata, S., Ishiguro, H.: Effect of Robot Notification on Acquiring Permission to Use Personal Information. In: International Conference on Human-Agent Interaction, pp. 124–132. ACM, Gothenburg Sweden (2023). https://doi.org/10.1145/3623809.3623896

[56] Yang, D., Chae, Y.-J., Kim, D., Lim, Y., Kim, D.H., Kim, C., Park, S.-K., Nam, C.: Effects of Social Behaviors of Robots in Privacy-Sensitive Situations. International Journal of Social Robotics **14**(2), 589–602 (2022) https://doi.org/10.1007/s12369-021-00809-2

[57] Fernandes, F.E., Guanci Yang, Do, H.M., Sheng, W.: Detection of privacy-sensitive situations for social robots in smart homes. In: 2016 IEEE International Conference on Automation Science and Engineering (CASE), pp. 727–732. IEEE, Fort Worth, TX (2016). https://doi.org/10.1109/COASE.2016.7743474

[58] Martelaro, N., Nneji, V.C., Ju, W., Hinds, P.: Tell me more designing HRI to encourage more trust, disclosure, and companionship. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 181–188. IEEE, Christchurch, New Zealand (2016). https://doi.org/10.1109/HRI.2016.7451750

[59] Saffarizadeh, K., Boodraj, M., Alashoor, T.: Conversational Assistants: Investigating Privacy Concerns, Trust, and Self-Disclosure. In: Thirty Eighth International Conference on Information Systems, South Korea (2017)

[60] Fishbein, M.: A theory of reasoned action: some applications and implications. Nebr Symp Motiv. **27**, 65–116 (1980)

[61] AJZEN, I.: The Theory of Planned Behavior. ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES **50**, 179–211 (1991)

[62] Ajzen, I.: The theory of planned behaviour: Reactions and reflections. Psychology & Health **26**(9), 1113–1127 (2011) https://doi.org/10.1080/08870446.2011.613995

[63] Dinev, T., Hart, P.: An Extended Privacy Calculus Model for E-Commerce Transactions. Information Systems Research **17**(1), 61–80 (2006) https://doi.org/10.1287/isre.1060.0080

[64] Culnan, M.J., Armstrong, P.K.: Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation. Organization Science **10**(1), 104–115 (1999) https://doi.org/10.1287/orsc.10.1.104

[65] Venkatesh, Thong, Xu: Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly **36**(1), 157 (2012) https://doi.org/10.2307/41410412

[66] Forgas-Coll, S., Huertas-Garcia, R., Andriella, A., Alenyà, G.: The effects of gender and personality of robot assistants on customers' acceptance of their service. Service Business **16**(2), 359–389 (2022) https://doi.org/10.1007/s11628-022-00492-x

[67] Wu, K.-W., Huang, S.Y., Yen, D.C., Popova, I.: The effect of online privacy policy on consumer privacy concern and trust. Computers in Human Behavior **28**(3), 889–897 (2012) https://doi.org/10.1016/j.chb.2011.12.008

[68] Kim, D.J., Ferrin, D.L., Rao, H.R.: A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. Decision Support Systems **44**(2), 544–564 (2008) https://doi.org/10.1016/j.dss.2007.07.001

[69] Van Slyke, C., Shim, J.T., Johnson, R., Jiang, J.: Concern for Information Privacy and Online Consumer Purchasing. Journal of the Association for Information Systems **7**(6), 415–444 (2006) https://doi.org/10.17705/1jais.00092

[70] Phonthanukitithaworn, C., Sellitto, C.: A Willingness to Disclose Personal Information for Monetary Reward: A Study of Fitness Tracker Users in Thailand. SAGE Open **12**(2), 215824402210973 (2022) https://doi.org/10.1177/21582440221097399

[71] Xu, H., Luo, X.R., Carroll, J.M., Rosson, M.B.: The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing. Decision Support Systems **51**(1), 42–52 (2011) https://doi.org/10.1016/j.dss.2010.11.017

[72] Krasnova, H., Kolesnikova, E., Guenther, O.: "It Won't Happen To Me!": Self-Disclosure in Online Social Networks. In: AMCIS 2009 Proceedings, p. 343 (2009)

[73] Liu, C., Marchewka, J.T., Lu, J., Yu, C.-S.: Beyond concern—a privacy-trust-behavioral intention model of electronic commerce. Information & Management **42**(2), 289–304 (2005) https://doi.org/10.1016/j.im.2004.01.003

[74] Pan, Y., Zinkhan, G.M.: Exploring the impact of online privacy disclosures on consumer trust. Journal of Retailing **82**(4), 331–338 (2006) https://doi.org/10.1016/j.jretai.2006.08.006

[75] What personal data is considered sensitive? https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en

[76] Fornell, C., Larcker, D.F.: Structural Equation Models with Unobservable Variables and Measurement Error: Algebra and Statistics. Journal of Marketing Research **18**(3), 382–388 (1981)

[77] Hair, J.F., Black, W.C., Babin, B.J.: Multivariate Data Analysis: A Global Perspective, Seventh edition edn. Global Edition. Pearson Education, US (2010)

[78] Bentler, P.M.: EQS 6 Structural Equations Program Manual. Technical report, Multivariate Software Encino, CA (2006)

[79] Lu, L., Cai, R., Gursoy, D.: Developing and

validating a service robot integration willingness scale. International Journal of Hospitality Management **80**, 36–51 (2019) https://doi.org/10.1016/j.ijhm.2019.01.005

[80] Felzmann, H., Villaronga, E.F., Lutz, C., Tamò-Larrieux, A.: Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. Big Data & Society **6**(1), 205395171986054 (2019) https://doi.org/10.1177/2053951719860542

[81] Chen, T.-W., Sundar, S.S.: 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, pp. 1–13. ACM, Montreal QC Canada (2018). https://doi.org/10.1145/3173574.3174111

[82] Nesset, B., Robb, D.A., Lopes, J., Hastie, H.: Transparency in HRI: Trust and Decision Making in the Face of Robot Errors. In: Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction, pp. 313–317. ACM, Boulder CO USA (2021). https://doi.org/10.1145/3434074.3447183