Article

new media & society
2025, Vol. 27(10) 5776–5802
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14614448241256899
journals.sagepub.com/home/nms
Sage

What did you hear and what did you see? Understanding the transparency of facial recognition and speech recognition systems during human–robot interaction

Kun Xu
University of Florida, USA

Xiaobei Chen
University of Florida, USA

Fanjue Liu
University of Florida, USA

Luling Huang
Missouri Western State University, USA

Abstract

As social robots begin to assume various social roles in society, the demand for understanding how social robots work and communicate grows rapidly. While literature on explainable artificial intelligence suggests that transparency about a social robot's working mechanism can evoke users' positive attitudes, transparency may also have negative outcomes. This study investigates the paradoxical effects of the transparency of facial recognition technology and speech recognition technology in human–robot interactions. Based on a lab experiment and combined analyses of users' quantitative and qualitative responses, this study suggests that the transparency of facial recognition technology in human–robot interaction increases users' social presence, reduces privacy

Corresponding author:
Kun Xu, College of Journalism and Communications, University of Florida, 1885 Stadium Road, Gainesville, FL 32611, USA.
Email: kun.xu@ufl.edu

concerns, and enhances users' acceptance of robots. However, exposure to both facial and speech recognition technologies revives users' privacy worries. This study further parses users' open-ended evaluation of the prospective application of social robots' tracking technologies and discusses the theoretical, practical, and ethical value of the findings.

Keywords

Computers Are Social Actors, explainable AI, human–machine communication, human–robot interaction, social presence, transparency

Humanoid social robots are defined as "human-made autonomous entities that interact with humans in a humanlike way" (Zhao, 2006: 405). They often feature physical embodiments that communicate through social behaviors such as speech, gestures, and movements (Reeves et al., 2020). In recent years, social robots have been increasingly used to serve social roles: Japan's home-based Lovot Robot is used to foster emotional bonding with humans; HuggieBot, a human-sized robot, provides human-like hugs to relieve stress and improve users' mental well-being.

Despite the adoption of social robots in various settings, humans' interaction with social robots has yet to achieve the sophistication depicted in science fiction portrayals (e.g. Westworld, Blade Runner). Indeed, conversations with social robots have been imbued with misunderstandings and breakdowns (Suchman, 2007). To accommodate robots' immaturity in maintaining ongoing conversations, humans often need to show patience by slowing down their speech (Kanda et al., 2008), adjusting their postures, and changing their language styles (Fischer et al., 2011), which can make conversations with robots less natural and more cognitively taxing.

One way to mitigate the impact of the communication breakdowns between humans and social robots is to improve the transparency of the robot's working mechanisms, provide explanations for users, and enhance users' trust in the robot (De Graaf et al., 2021). Wilkinson et al. (2021) suggested that proper explanations and justifications allow users to understand the rationale of AI systems, and hence, augment users' trust in and engagement with AI. As social robots are a convergence of a variety of artificial intelligence (AI) systems (e.g. facial recognition, speech recognition, gesture tracking, object detection), the goal of this study is to draw on recent literature on explainable AI (XAI) to explore whether and how increasing the transparency of a social robot's backstage AI systems, especially its facial recognition and speech recognition systems, affects users' perceptions of and attitudes toward the robot.

This study can make three contributions to the literature. First, although transparency has been considered as an ideal strategy to boost users' trust in AI-driven technologies (Weitz et al., 2019), improving transparency and providing explanations may not always lead to users' positive evaluation of AI (Ananny and Crawford, 2018). Explanations of multiple AI technologies behind social robots may cause confusion, raise users' concerns about privacy intrusion, and consequently undermine trust (De Graaf et al., 2021). Therefore, this study seeks to understand how individuals leverage the promises and perils of transparency and how the backstage AI systems affect individuals' trust in and perceived privacy risks of social robots.

Second, past research on human-robot interaction (HRI) has drawn on the Computers Are Social Actors (CASA) paradigm to understand how individuals respond to social robots as if they were social actors (e.g. Edwards et al., 2016; Li et al., 2017; Spence et al., 2014; Xu, 2019). The CASA paradigm suggests that individuals inadvertently apply human–human communication scripts to human–technology interaction when technology is designed with human social cues, such as voices, human language, and facial expressions (Nass, 2004). Distinct from prior research, in our study, social robots' facial recognition images and speech recognition rates act as auxiliary non-human social cues in HRI. Given that these cues do not normally exist in human–human communication but are prevalent in HRI, it is important to understand how these non-human social cues change the extent to which individuals treat social robots as social actors which can complement the existing knowledge about the CASA paradigm.

Third, this study uses triangulation (Cook, 1985) and seeks to establish correspondence between users' psychological reactions to the facial and speech recognition systems and users' qualitative evaluation of these AI technologies. It pursues the convergence of findings by contextualizing the experiment results within users' qualitative responses, which can contribute to a body of research where mixed methods are less common: in a recent systematic review of 132 human–machine communication (HMC) studies, mixed methods only accounted for five (3.79%; Richards et al., 2022).

In sum, this study unpacks the black box of a social robot's working mechanisms and delves into the effects of showing participants a robot's facial and speech recognition systems during human–robot communication breakdowns. By analyzing both quantitative and qualitative responses, we parse out the paradoxical effects of making a social robot's AI technologies transparent and comprehensible.

Literature review

Explainable AI in social robots

Researchers have referred to XAI in their explorations of how the transparency of AI systems affects individuals' perceptions and attitudes toward technologies. XAI is defined as "the class of systems that provide visibility into how an AI system makes decisions and predictions and executes its actions" (Rai, 2020: 138). Researchers and developers have explained AI systems from various perspectives. Some scholars apply *global interpretations*, explaining the logic of all the applied models in AI systems, while others apply *local interpretations*, focusing on a single model-made decision or prediction (Wolf and Ringland, 2020). Using a different dimension, scholars have also relied on *model-specific interpretations* to present the working mechanisms of a class of machine learning models or used *model-agnostic interpretations* to offer post hoc explanations in lay language (Adadi and Berrada, 2018). In this study, instead of examining users' reactions to the technical explanations about how a robot recognizes speech and detects human faces, we use *model-agnostic interpretations* to expose users to robots' facial recognition images and speech recognition rates.

XAI serves four major purposes (Adadi and Berrada, 2018). First, explanations allow researchers to check the AI systems and prevent them from making mistakes. Second, explanations improve the transparency of AI systems and expedite open collaboration and open innovation. Third, explanations enable AI users to understand whether AI predictions are based on any biased data input or discriminated data training processes. Fourth, in HRI, explanations allow users to understand how a robot acts on its decisions and what norms and constraints the robot factors in during its course of actions.

Past works have applied XAI to research on chatbots and virtual agents. Khurana et al. (2021) found that when breakdowns occur during a human–chatbot interaction, explanations provided by the chatbot increased users' understanding of the breakdowns and improved the perceived transparency, trustworthiness, and usefulness of the chatbot. Wilkinson et al. (2021) found that users reported greater trust and perceived transparency in the presence of a chatbot that provided justifications for its recommendations, compared to one that did not.

The positive effects were also found when XAI is applied to social robotics. Schadenberg et al. (2021) suggested that participants who could not see the cause of a robot's actions perceived it as unpredictable and less competent. By contrast, participants who saw the cause of the robot's actions had better mental models of the robot and developed greater trust in it. Aligned with these findings, Fischer et al. (2018) pointed out that transparency had an overall positive impact on robots' perceived trustworthiness and users' feeling of control.

While this thread of XAI literature suggests that providing explanations to enhance transparency is an ideal strategy to boost users' trust, another thread of literature questions the positive effects of transparency. Along this line, research suggests that the effect of transparency on users' trust may be contingent upon many factors. For example, researchers found that the transparency of a social robot's reasoning led to more trust only when the robot did not make mistakes. When participants believed the robot made a wrong decision, the positive impact of transparency faded away (Nesset et al., 2021). Another study on users' attitudes toward AI's classification capacity suggested that those with rich AI use experience preferred concept-based explanations that provided more numbers and equations, but those with limited AI use experience felt overwhelmed by concept-based explanations and valued visual-based explanations (Kim et al., 2023).

Overall, research on using explanations during HRI is still in its nascent phase. Whether, when, and how explanations should be provided to users demands further evidence. To contribute to the current literature, this study uses XAI as a bedrock framework and investigates whether and how boosting the transparency of a social robot's facial recognition and speech recognition technologies affects individuals' social responses to the robot.

Combining XAI and the CASA paradigm

To understand individuals' responses to social robots, researchers have applied the CASA paradigm proposed by Nass et al. (1994) in the early 1990s. By conducting a series of lab experiments that test individuals' reactions to computers, Nass and colleagues proposed that individuals' interactions with technologies are "fundamentally social and natural, just like interactions in real life" (Reeves and Nass, 1996: 5). Some examples of individuals' social responses to computers in early CASA research include treating computers politely (Nass et al., 1994), perceiving computers as their teammates (Nass et al., 1996), assigning gender stereotypes to computers (Nass et al., 1997), and preferring computers that have similar personalities to their own (Lee and Nass, 2005).

According to the CASA paradigm, social cues play a vital role in evoking individuals' social responses to technologies. Here, social cues are defined as "biologically and physically determined features salient to observers because of their potential as channels of information" (Fiore et al., 2013: 2). Social responses refer to the reactions users show toward humans based on certain attributes or norms (e.g. genders, personalities, reciprocity; Lee, 2023). Nass (2004) provided a list of social cues that evoke individuals' social responses to technologies, which include language use, voice, face, and emotion manifestation. Seeking to structurally extend the CASA paradigm, Lombard and Xu (2021) also placed emphases on the effects of social cues and proposed that cues like facial expressions, human-sounding voices, gestures, and human-like appearances are evolutionarily powerful in eliciting individuals' social responses.

One important indicator of users' social responses to technologies is social presence. In mediated communication, Biocca et al. (2003) referred to social presence as "the sense of being with another" (p. 456). Expanding to both mediated and non-mediated contexts, Lee (2004) conceptualized it as "a psychological state in which virtual (para-authentic or artificial) social actors are experienced as actual social actors in either sensory or non-sensory ways" (p. 45). Recently, Cummings and Wertz (2023) reviewed the past operationalization of social presence and re-defined it as "the perceptual salience of another social actor" (p. 1). Pertinent to the context of HRI, Lombard and Ditton (1997) distinguished two types of social presence: social-actor-within-medium presence and medium-as-social-actor presence. The former involves users' responses to the social cues presented by social actors within a medium (e.g. television anchors, media characters, avatars). The latter involves users' responses to the social cues presented by technologies per se (e.g. social robots, computers, smart speakers). Responses to social robots, the subject at hand, belong to the latter category.

Medium-as-social-actor presence plays an important role in HMC. For instance, Bracken and Lombard (2004) found that when children perceived a computer as a social actor (i.e. medium-as-social-actor presence) and received positive feedback from it, their confidence in learning and their recall performances substantially improved. Lee et al. (2005) found that perceiving a robot as a social actor positively predicted the robot's perceived attraction and users' purchasing intention. To strengthen individuals' medium-as-social-actor presence experience, researchers have manipulated social cues in multiple ways. Fiore et al. (2013) found that in a condition where a robot blocked participants' travel paths and then moved aside, participants reported stronger medium-as-social-actor presence compared to a condition where a robot did not yield to them. In another study, by manipulating the kinetic cues of a zoomorphic social robot, Lee et al. (2006) found that users experienced stronger medium-as-social-actor presence when interacting with a robot that exhibited personalities complementary to their own, as opposed to one that exhibited matching personalities.

Compared to past HRI research that has primarily examined the influence of designing interpersonal social cues (e.g. voices, gestures, facial expressions) to robots (Krämer et al., 2015; Xu, 2019), what has been largely neglected is how non-human social cues affect individuals' social responses. While technologies have been designed with human social cues to be engaging, interactive, and lifelike, it cannot be ignored that technologies can present cues that do not normally exist in human–human communication. For example, when interacting with a telepresence robot, individuals may interpret the electronic tablet as its head and the wheels as its feet. In addition, a social robot may use light-emitting diode (LED) lights to indicate its different emotions (Embgen et al., 2012). To communicate directional intent, a social robot may present floor projections of its moving directions (Shrestha et al., 2018). Some recent AI agents like Replika can send pictures and use memes during communication with humans. These cues do not normally appear in interpersonal communication, yet they still deliver social meanings and elicit users' emotional, cognitive, or behavioral reactions in HMC contexts. Thus, by scrutinizing how technologies present both human social cues and non-human social cues, researchers can make sense of the complicated technology environment and link different theories and disciplines to understand users' responses (Xu and Liao, 2020). One of the few empirical studies that distinguished human social cues and non-human social cues suggested that compared to a robot that did not demonstrate any nonverbal behavior, a robot's human-like nonverbal behavior (e.g. moving head, arms, and torso) enhanced users' positive affect and self-disclosure. Meanwhile, robot-specific nonverbal behavior (e.g. showing and changing different eye colors) also slightly encouraged these responses (Rosenthal-Von der Pütten et al., 2018).

In our study, a distinctive feature of humanoid social robots is that, in addition to human social cues, they can present machine-generated, non-human social cues, such as facial recognition images and speech recognition rates. As it is rare in interpersonal communication for human interlocutors to repeatedly and strategically inform their partners of what they hear and what they see, using robots to present these facial and speech recognition cues and consequently testing the effects of these cues adds an additional theoretical layer to the CASA paradigm, which could enhance our understanding of how users' social responses to social robots vary based on these non-human social cues. Thus, we proposed the following research questions:

> *RQ1*. How will users' exposure to a social robot's facial recognition system affect their medium-as-social-actor presence of the robot?
>
> *RQ2*. How will users' exposure to a social robot's speech recognition system affect their medium-as-social-actor presence of the robot?

Paradox in transparency of facial recognition

A facial recognition system is "a technology capable of identifying or verifying a person from a digital image or a video frame from a video source" (Petrescu, 2019: 237). Facial recognition systems have advanced rapidly in recent years, leading to their integration into various sectors, such as mobile payment and entrance access control systems (Peng, 2022). Although using facial recognition systems enhances the efficiency and convenience of communication (Li and Li, 2023), concerns regarding users' perception of privacy intrusion arise. For example, Pantano (2020) found that there is a prevalent belief that facial recognition systems may capture data related to users' age, race, and gender. Concerns about such capture are exacerbated by the perception that facial recognition is often quiet and discreet, potentially leading to non-consensual collection of biometric data (Chamikara et al., 2020).

The paradox of adopting facial recognition technology also exists in HRI. A social robot that is designed to carry out natural and life-like conversations is often paired with cameras and sensors (Chuah et al., 2021). However, the technologies that enable robots to detect faces may intrude users' physical, psychological, and social privacy (Lutz et al., 2019). For example, Krupp et al. (2017) found that when interacting with telepresence robots, participants expressed concerns about their privacy regarding recordings of awkward moments. Xie et al. (2023) found that users worried about being monitored and tracked during their interactions with robots, alongside the perception that their information might be disseminated without consent.

To mitigate privacy concerns in human-AI interactions, XAI researchers have sought to increase transparency by providing explanations for users. For example, Suen and Hung (2023) noticed that explanations of AI-based video interviews improved users' cognitive and affective trust in the AI systems. However, mixed findings emerged in Vitale et al.'s (2018) research; although transparency of the facial recognition systems installed in a humanoid robot enhanced the perceived predictability, attractiveness, and novelty, it did not reduce users' privacy concerns as expected. In sum, while XAI literature suggests that transparency elevates users' trust, seeing faces recognized by technology may induce users' privacy concerns, which could reversely undermine their trust and technology acceptance. Therefore, we propose the following research question:

> *RQ3*. How will users' exposure to a social robot's facial recognition system affect their (1) trust, (2) privacy concerns, and (3) acceptance of the robot?

Paradox in transparency of speech recognition

The integration of speech recognition in social robots mirrors the Janus-faced impact of facial recognition. Using natural language processing, speech recognition enables machines to process and understand human words (Martínez-Plumed et al., 2021). Speech recognition can expedite message response time, improve data collection speed, and consequently enhance customer service efficiency (Song, 2020). In addition, social robots equipped with the ability to recognize and react to the emotional tones of human speech can earn users' trust, fostering a more empathetic interaction (Law et al., 2021).

Despite the promises of speech recognition, research has noticed the concomitant threats. Lin et al. (2022) found that extensive collection, processing, and transmission of speech data engendered users' perceived privacy risks, including the disclosure of individually identifiable information, geo-location information, and demographic details.

Pradhan et al. (2020) explored elderly adults' experience with voice assistants and found many had privacy concerns about recorded conversations.

Echoing XAI literature, scholars perceive transparency as a possible solution to the negative impact of speech recognition. Although limited research has examined the effects of transparency of a social robot's speech recognition rates, Zhang et al. (2020) investigated the influence of showing users AI's prediction confidence score, which reflects, "the model's predicted probability for the most likely outcome" (p. 1). Zhang et al. (2020) found that compared to delegating the decision-making to AI without seeing AI's confidence score, letting AI display its score augmented users' trust in AI's decision-making.

Overall, when social robots demonstrate their speech recognition ability, they may evoke users' positive attitudes, as speech recognition can ease human–robot conversations. Meanwhile, it is questionable whether users' privacy concerns will override their positive experience of interacting with the robots (Lutz et al., 2019). Thus, we propose the following research question:

> *RQ4*. How will users' exposure to a social robot's speech recognition system affect users' (1) trust, (2) privacy concerns, and (3) acceptance of the robot?

In addition to testing the quantitative effects of the transparency of facial and speech recognition technologies in HRI, this study further probes users' qualitative responses to these AI technologies, in line with Robinson and Mendelson's (2012) suggestion that using qualitative questions in an experiment could provide room for participants to extend their answers and contextualize their quantitative responses. Indeed, scholars have noted the advantage of mixed methods considering that triangulation can be used to interweave the findings (Denzin, 1978). By bridging different but conceptually relevant results, triangulation can lead to more consolidated and informative findings. Thus, this study further examines participants' open-ended responses:

> *RQ5*. How do users evaluate the privacy intrusion of a social robot's facial recognition and speech recognition systems?
>
> *RQ6*. How do users evaluate the future application of a social robot's facial recognition and speech recognition systems?

Method

Participants

A total of 102 participants were recruited from a large public university on the east coast of the United States to participate in a lab experiment. They were told to help test the basic conversation ability of a social robot. Participants were recruited through the Sona system and received extra credit, or through announcements posted on Reddit and received US$10 gift cards. After removing invalid cases due to technology failure during the experiment (e.g. the robot froze), the final sample included 92 participants, among whom 32 were males and 60 were females. They ranged in age from 17 to 35 years old (*M*=20.65, *SD*=2.82).

Experimental stimulus

The experiment used NAO V6, a social robot developed by United Robotics Group. The software Choregraphe was used to program the robot's reactions and show participants the robot's speech recognition rates and facial recognition images. The Choregraphe interface allows researchers to monitor the status of NAO. The dialog panel visualizes the human speech input and the robot's confidence rate of its speech recognition. The video monitor panel mirrors what the robot sees through its cameras and demonstrates its face recognition system.

Research design and procedures

The experiment applied a 2 (facial recognition: available vs unavailable) × 2 (speech recognition: available vs unavailable) between-subjects factorial design. Participants were randomly assigned to one of the four conditions: (1) facial recognition and speech recognition, (2) facial recognition only, (3) speech recognition only, and (4) neither facial recognition nor speech recognition. After participants entered the lab and signed the consent form, they completed a pre-experiment questionnaire asking about their demographic information through Qualtrics. Then, they were escorted to a connecting room where NAO was sitting on a desk and a 25-inch desktop computer with the software Choregraphe was set up (see Figure 1). Participants were asked to sit about 2 feet from the robot.

Participants were first introduced to a testing session in which they needed to select two out of 11 commands to chat with the robot (e.g. "how are you," "what day is it today"). If the robot did not respond to participants, the experimenter would give suggestions on how participants could change their pitch and tones to maintain the conversation. This testing session was designed to allow participants to use proper tones and voices to interact with NAO.

After the testing session, participants started the formal sessions. They received a list of 61 commands that they could use to interact with the robot. The commands were categorized into six sections including basic information, equipment, working systems, request, philosophical questions, and wrap-up questions. Each section listed four to 15 commands for stimulus sampling purposes (Reeves et al., 2016). Participants were asked to pick one command from each section to communicate with the robot. To increase external validity, these verbal commands were aligned with the suggested questions listed in the NAO's developer's guide. A full list of commands is available in the Supplementary Materials: http://tinyurl.com/robottransparency.

During the formal sessions, when the robot did not react to participants' commands, the experimenter paused the interaction and introduced participants to the Choregraphe interface. In the facial recognition conditions, the experimenter asked participants to look at the video monitor panel and explained to participants that the robot did not respond because it had not captured their faces. Participants were then asked to ensure

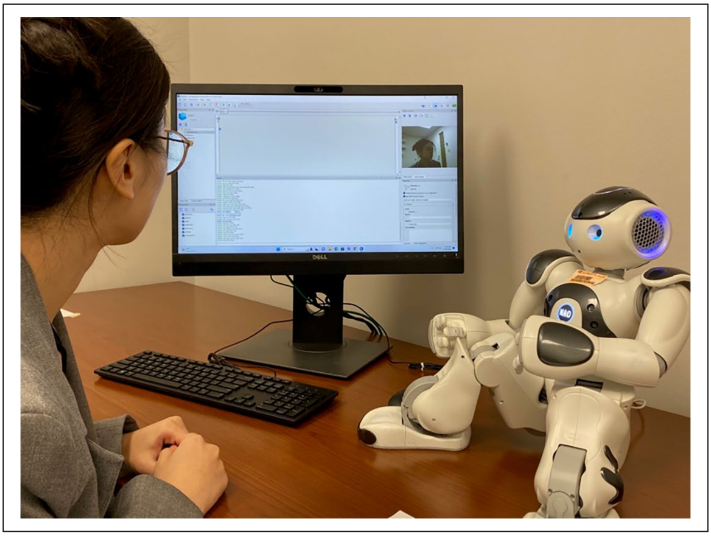


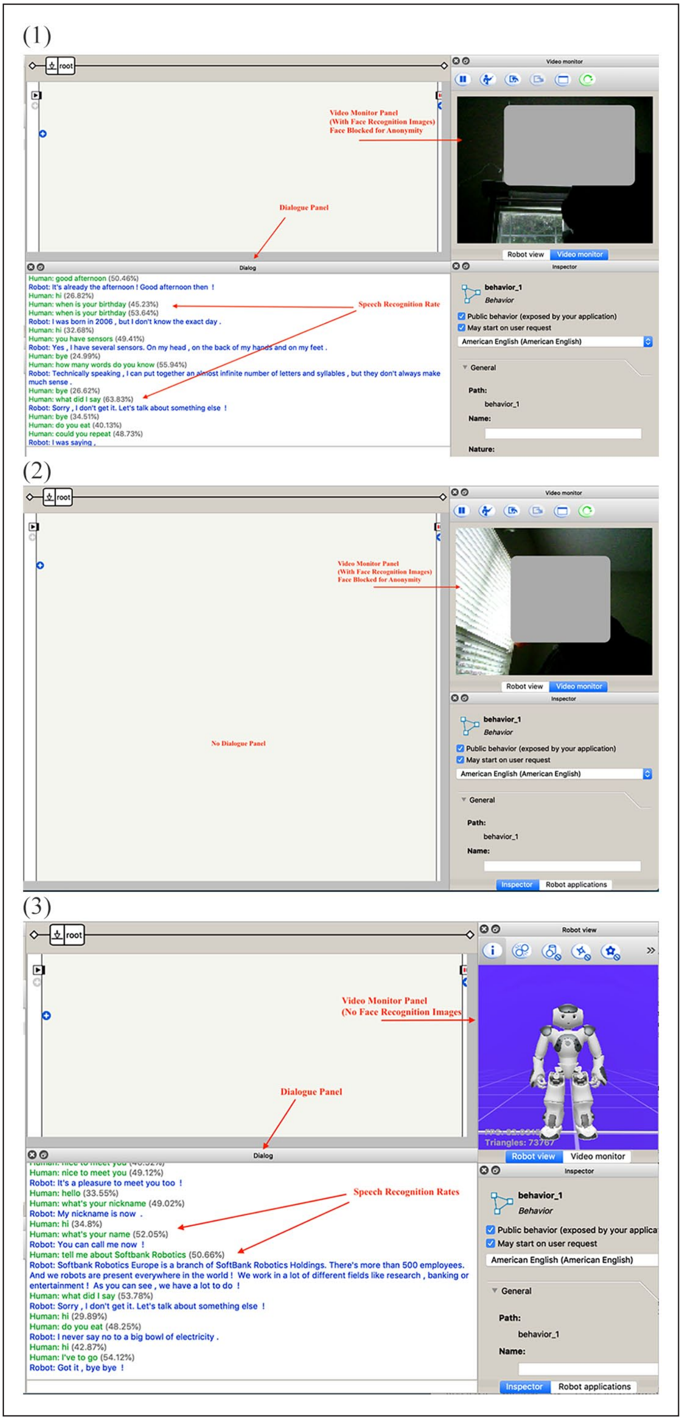
Figure 1. Experimental setting.

that their faces appeared in the video monitor window and try the same command again until the robot responded. Participants repeated the same procedure until the robot responded to all the verbal commands.

In the speech recognition conditions, when the robot failed to respond, the experimenter paused the interaction and asked participants to look at the robot's speech recognition rates in the dialogue panel of Choregraphe. Participants received the explanation that the recognition rates represented how confident the robot was in recognizing their words. The experimenter informed participants that the robot did not respond because the robot's speech recognition confidence level was not high enough to trigger its responses.[1] Then participants were asked to adjust their voices and try the same command again until the robot responded. Participants repeated this procedure until NAO responded to all their commands.

In the face recognition and speech recognition conditions, the experimenter asked participants to look at both the dialogue panel and the video monitor panel. Consistent with the speech recognition only and the facial recognition only conditions, participants were informed that the robot did not respond because the robot's speech recognition confidence level was not high enough to trigger the response or the robot did not capture their faces. Participants were asked to adjust their tones or voices, make sure that faces appeared in the video monitor window, and try the same commands. The order of the explanations of the speech recognition and the face recognition systems were randomized in this condition to avoid ordering effects. The Choregraphe interfaces for different conditions are shown in Figure 2.

In the neither speech recognition nor facial recognition conditions, when the robot did not respond, the experimenter only asked the participants to adjust their tones or voices and wait until the robot's eyes turned blue to use the verbal command again. The blue eye



Human: good afternoon (50.46%)
Robot: It's already the afternoon ! Good afternoon then !
Human: hi (26.82%)
Human: when is your birthday (45.23%)
Human: when is your birthday (53.64%)
Robot: I was born in 2006 , but I don't know the exact day .
Human: hi (32.68%)
Human: you have sensors (49.41%)
Robot: Yes , I have several sensors. On my head , on the back of my hands and on my feet .
Human: bye (24.99%)
Human: how many words do you know (55.94%)
Robot: Technically speaking , I can put together an almost infinite number of letters and syllables , but they don't always make much sense .
Human: bye (26.62%)
Human: what did I say (63.83%)
Robot: Sorry , I don't get it. Let's talk about something else !
Human: bye (34.51%)
Human: do you eat (40.13%)
Human: could you repeat (48.73%)
Robot: I was saying ,

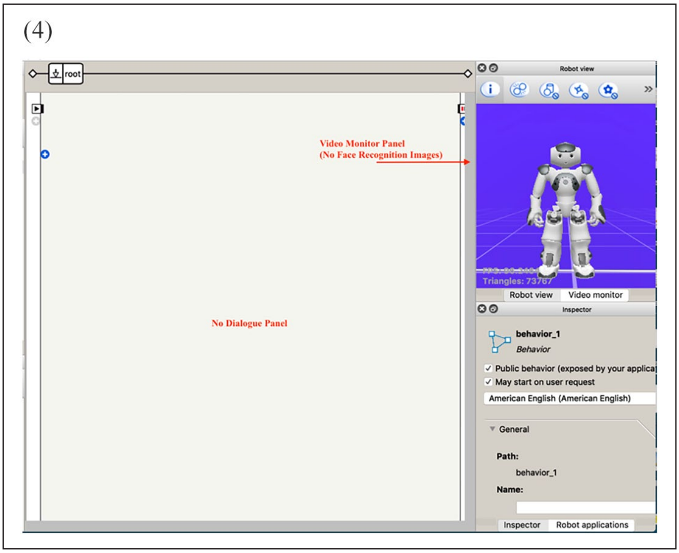
behavior_1
Behavior
☑ Public behavior (exposed by your application)
☑ May start on user request
American English (American English)
General
Path:
behavior_1
Name:
Nature:

behavior_1
Behavior
☑ Public behavior (exposed by your application)
☑ May start on user request
American English (American English)
General
Path:
behavior_1
Name:

Human: nice to meet you (49.12%)
Robot: It's a pleasure to meet you too !
Human: hello (33.55%)
Human: what's your nickname (49.02%)
Robot: My nickname is now .
Human: hi (34.8%)
Human: what's your name (52.05%)
Robot: You can call me now !
Human: tell me about Softbank Robotics (50.66%)
Robot: Softbank Robotics Europe is a branch of SoftBank Robotics Holdings. There's more than 500 employees. And we robots are present everywhere in the world ! We work in a lot of different fields like research , banking or entertainment ! As you can see , we have a lot to do !
Human: what did I say (53.78%)
Robot: Sorry , I don't get it. Let's talk about something else !
Human: hi (29.89%)
Human: do you eat (48.25%)
Robot: I never say no to a big bowl of electricity .
Human: hi (42.87%)
Human: I've to go (54.12%)
Robot: Got it , bye bye !

behavior_1
Behavior
☑ Public behavior (exposed by your applica
☑ May start on user request
American English (American English)
General
Path:
behavior_1
Name:

Figure 2. (Continued)



behavior_1
Behavior
☑ Public behavior (exposed by your applica
☑ May start on user request
American English (American English)
General
Path:
behavior_1
Name:

Figure 2. Choregraphe interfaces for different conditions.
Picture (1) shows the condition for transparency of both speech recognition rates and facial recognition images. Picture (2) shows the condition for transparency of facial recognition images only. Picture (3) shows the condition for transparency of speech recognition rates only. Picture (4) shows the non-transparency condition.

color is an indicator that NAO captures a face in camera. However, participants were not told what the robot's eye color meant. The dialogue and the video monitor panels were closed in this condition. Participants were not provided with any explanations about the robot's nonresponses.

All participants experienced at least one conversation breakdown during the formal sessions. However, since NAO's responses varied with different individual voices, the number of NAO's nonresponses to each participant could not be manipulated to be consistent across conditions. Therefore, we recorded NAO's nonresponses to each participant's first attempt at a command as a control variable ($M = 2.32$, $SD = 1.21$). For instance, if NAO did not respond to two out of six commands given by a participant at the first attempt, the count of nonresponses was coded as two.

After the formal sessions, participants returned to the main room to complete a post-experiment questionnaire. The experiment lasted about 30 minutes. Only one participant participated in the experiment at a time.

Measures

Details of measures are presented in Table 1. Participants' demographic information, robot use experiences, programming expertise, and general attitudes toward robots were measured. These measures are provided in the Supplementary Materials: http://tinyurl.com/robottransparency.

Table 1. Measures of dependent variables.

Dependent variables	Nature of scale	Mean (*SD*)	Cronbach's α	Scale items
Medium-as-social-actor presence (Lee et al., 2005, 2006)	Ten-point Likert-type	6.45 (1.7)	0.85	How much did you feel as if you were interacting with an intelligent being? How much did you feel as if you were together with an intelligent being? How much attention did you pay to the robot? How much did you feel involved with the robot? How much did you feel as if the robot was talking to you? How much did you feel as if you and the robot were communicating with each other?
Trust in robot (Gong and Nass, 2007; Wheeles and Grotz, 1977)	Seven-point Semantic differential	5.48 (0.85)	0.68	Untrustworthy—trustworthy. Unreliable—reliable. Inconsiderate—considerate. Dangerous—safe. Dishonest—honest.
Privacy concerns (Liu et al., 2021)	Seven-point Likert-type	2.95 (1.47)	0.9	I am concerned that the robot was collecting too much personal information about me. I am concerned that unauthorized people could access my personal information. I am concerned that the robot may use my personal information for purposes that I am not made aware of. I am concerned that my personal information stored in the robot will not be protected.
Robot acceptance (Kuchenbrandt et al.,2013)	Seven-point Likert-type	5.14 (2.13)	0.81	I liked the robot. I would be willing to get to know the robot more closely. I would be willing to talk more to the robot. I would be willing to purchase a similar robot.

Participants were asked to describe whether and why they had any privacy concerns when interacting with the robot. They were also asked how they perceived the idea of letting robots present their facial recognition images and speech recognition rates in future HRI.

Data analyses

After examining univariate and multivariate outliers, three cases with multivariate outliers were removed. Skewed variables including robot acceptance and the number of nonresponses were transformed for normal distribution. To examine RQ1 to RQ4, two-way ANCOVAs were conducted, with the availability of speech recognition and facial recognition systems being two independent variables. The number of the robot's nonresponses, participants' genders, programming expertise, robot use experiences, and attitudes toward robots were controlled based on prior literature (Johnson et al., 2004; Lee, 2008).

To answer RQ5 and RQ6, textual analyses were used to parse participants' open-ended responses. Based on Hesse-Biber and Leavy's (2010) four steps for qualitative analyses (i.e. data preparation, data exploration, data reduction, and interpretation), text data were first reviewed and coded in a descriptive way. Examples of these descriptive codes included "denial of privacy concerns" and "acceptance of facial recognition." Meanwhile, memos were taken to further categorize these codes. After descriptive coding, data were analytically reviewed until patterns and themes reached saturation (Charmaz, 2014). Some analytical codes included "de-sensitization to privacy intrusion" and "heightened sense of security." To minimize researchers' biases, three researchers independently read the qualitative data based on the themes and resolved discrepancies through discussion. These patterns and themes were interpreted and reported in the results. A number was assigned to each participant for anonymity and their assigned condition was marked (e.g. P26, facial recognition [FR] and speech recognition [SR]).

Results

Quantitative results

RQ1 to *RQ4* asked how users' exposure to a social robot's facial recognition system and speech recognition system affected their medium-as-social-actor presence, trust in the robot, privacy concerns, and acceptance of the robot. Two-way ANCOVAs suggested that although exposure to the robot's facial recognition system did not have a main effect on users' trust, it had a marginally main effect on users' medium-as-social-actor presence, $F(1, 77)=3.43$, $p=.068$, partial $\eta^2=.04$. Those who were exposed to the facial recognition system ($M=6.71$, $SD=1.67$) experienced higher medium-as-social-actor presence of the robot than those who were not ($M=6.17$, $SD=1.64$).

Exposure to the robot's facial recognition also had a main effect on users' privacy concerns, $F(1, 77)=5.55$, $p=.021$, partial $\eta^2=.07$. Those who were exposed to the facial recognition system ($M=2.62$, $SD=1.22$) reported significantly lower privacy concerns than those who were not exposed to the system ($M=3.22$, $SD=1.52$). Meanwhile, exposure to the robot's facial recognition system interacted with exposure to the speech

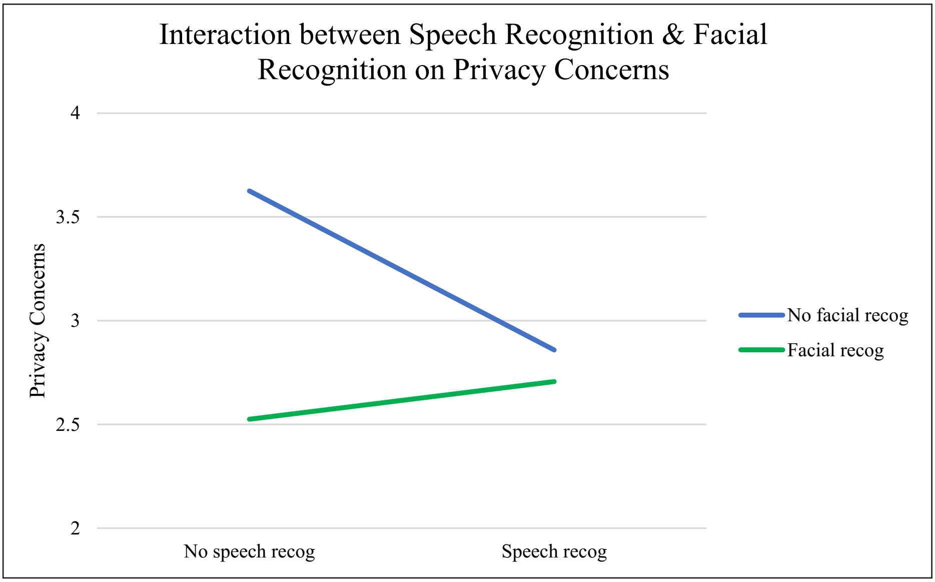


Figure 3. Interaction effects between speech recognition and facial recognition on privacy concerns.

recognition system in predicting users' privacy concerns, $F(1, 77)=4.22$, $p=.043$, partial $\eta^2=.05$, meaning that when facial recognition was not available to users, the availability of speech recognition mitigated users' privacy concerns. When facial recognition was available, exposure to speech recognition amplified users' privacy concerns. The interaction is shown in Figure 3.

Exposure to the robot's facial recognition also had a main effect on users' acceptance of the robot, $F(1, 77)=4.19$, $p=.044$, partial $\eta^2=.05$. Those who were exposed to the facial recognition system ($M=5.40$, $SD=1.05$) reported significantly higher acceptance of the robot than those who were not exposed to the system ($M=5.02$, $SD=1.16$).

Exposure to the robot's speech recognition system did not have main effects on users' medium-as-social-actor presence, trust in the robot, privacy concerns, or acceptance of the robot. Results of the main effects are presented in Table 2.

Qualitative results

RQ5 asked about participants' privacy concerns related to the social robot's facial and speech recognition systems. Overall, participants expressed limited privacy concerns over interactions with NAO, even though they were aware that their face images or speech was tracked. One major reason was that participants felt limited disclosure of their personal information. Some participants mentioned that even though the robot saw their faces, they did not disclose information that was too personal to be exploited. One participant mentioned,

Table 2. Main effects of speech recognition and facial recognition.

	Facial recognition	No facial recognition	Main effects	Speech recognition	No speech recognition	Main effects
	M (*SD*)	*M* (*SD*)	*F*	*M* (*SD*)	*M* (*SD*)	*F*
Medium-as-social-actor presence	6.71 (1.67)	6.17 (1.64)	3.43†	6.36 (1.69)	6.54 (1.65)	0.11
Trust	5.58 (0.83)	5.41 (0.83)	1.66	5.48 (0.86)	5.51 (0.81)	0.02
Privacy concerns	2.62 (1.22)	3.22 (1.52)	5.55*	2.78 (1.29)	3.08 (1.53)	0.88
Acceptance	5.40 (1.05)	5.02 (1.16)	4.19*	5.09 (1.14)	5.34 (1.08)	0.6

M: mean, *SD*: standard deviation.
*$p < .05$; †$p < .01$.

> The robot did not really gain any personal information from me other than my face on its camera and the sound of my voice. In the conversation, the robot did not ask any information about me, and I did not give it any personal information voluntarily either. (P36, neither SR nor FR)

Similarly, another participant mentioned that other than face or voice capture, she did not see sensitive data disclosed to the robot, implying that face recognition or voice recognition did not evoke their worries as much as other personal information. Regarding what counts as sensitive data, one participant added that if the robot had requested income or social security numbers, she would have been more careful.

While some indicated that nothing too personal was disclosed, others felt powerless and was desensitized to privacy intrusion. One participant reported that giving personal data to the robot is "just like giving any other company personal data" (P67, SR and FR). Another participant commented,

> I don't think too much about privacy concerns when it comes to AI and technology [. . .]. It's less about not caring about privacy issues, and more of just how most social media platforms and your digital devices track you so you're simply used to it. (P35, SR)

While most participants expressed few concerns over the robot's collection of their facial and speech information, a small number of participants expressed sensitivity to the robot's tracking technologies. One noted that "a camera may pick up key details in the background that could indirectly expose more information than I intend to give to the robot" (P52, FR).

RQ6 asked about users' attitudes toward the future application of a social robot's facial recognition and speech recognition systems. First, advances in speech recognition and facial recognition capabilities, according to participants, made robots "humanlike communicators" (P85, neither SR nor FR). Although this participant did not specify the meaning of "humanlike communicators," another added that these features could "help the robot-human interactions become smoother and help both parties adjust to best practices for effective communication" (P68, neither SR nor FR).

Another reason for participants' endorsement of the facial and speech recognition features is that these technologies could "make people less afraid of their robots since they can really understand what the robot is perceiving" (P13, SR and FR). Consistent with how transparency in XAI can boost users' trust in the AI systems, participants remarked that "seeing how the robot works makes things less of a mystery for a civilian, as they are able to plainly see how the programming of the robot works instead of just seeing an intelligent being" (P38, SR).

Results further suggested that rather than feeling disturbed by the invasion of the technologies, users believed that these facial recognition and speech recognition systems actually enhanced security:

> Facial recognition is used to protect phones, laptops, office buildings, and more these days. To add that feature onto a robot would probably allow it to be more secure in some ways and be more lifelike and natural as it detects the changing expressions on someone's face or the difference between two faces. (P23, SR+FR)

While most participants had overall positive attitudes toward the future application of facial recognition and speech recognition technologies in social robots, others shared mixed feelings. They acknowledged the usefulness of the technologies, but they also raised concerns over who controls the technology, "If the operator of the robot is reliable, I would definitely love to interact with the human-robot[sic]" (P20, SR + FR). Similarly, another participant suggested that it would be dangerous if the technology gets into the wrong hands and more regulations are needed if the technology is released to the public.

Discussion

Summary

This study seeks to understand users' psychological responses to social robots when robots demonstrate their backstage working mechanisms. Findings indicate that transparency of a social robot's facial recognition system evoked users' perception of the robot as a social actor, ameliorated users' privacy concerns, and increased users' acceptance of the robot. The transparency of the robot's speech recognition system abated users' privacy concerns when the facial recognition system was not made available. However, when both the facial recognition and the speech recognition systems were made transparent, users' privacy concerns slightly revived compared to when only the facial recognition system was available.

The study further evaluated users' qualitative responses. Users reported limited privacy concerns when interacting with the social robot. Several factors attenuated their perceived privacy risks. For example, their self-disclosure of private information was regarded as minimal. Users also naturalized information collection and felt powerless to counter privacy invasion. Moreover, rather than treat facial recognition and speech recognition as a threat, participants expressed favorable attitudes toward future applications of these technologies. They suggested that these AI recognition technologies streamlined HRI, enhanced the transparency of the robot's actions, and even improved their sense of safety.

Findings and implications

Participants' exposure to the social robot's facial recognition system increased their medium-as-social-actor presence. This finding can add to the CASA paradigm in that prior HRI literature has primarily focused on how users react to social robots that are designed with human social cues (e.g. human-like appearances, voices). Results in this study indicate that showing participants a social robot's non-human, machine-generated social cues, such as facial recognition images, can also evoke users' perception of the robot as an intelligent social being. In this process, participants experienced heightened medium-as-social-actor presence, probably because they realized that the robot must "see" their faces to continue the conversation, which might have evoked a sense of face-to-face interaction between the participants and the robot. The postulation was corroborated by participants' qualitative responses, in that they found the transparency of the facial recognition images made the robot "humanlike communicators."

Aside from medium-as-social-actor presence, the transparency of the social robot's facial recognition images mitigated users' privacy concerns and enhanced users' acceptance of the robot. Situated in the paradox between the positive effects of transparency and the negative effects of being exposed to the facial recognition systems, this study supports that transparency prevails over the potential perceived privacy risks generated by the facial recognition technology, which echoes prior findings about how XAI enhances transparency and positively affects users' attitudes toward chatbots and robots (Khurana et al., 2021; Schadenberg et al., 2021). Although past research has suggested that facial recognition technology raised users' privacy concerns in HRI (Krupp et al., 2017; Lutz et al., 2019), this study indicates that when communication breakdowns in HRI occur, transparency can at least partially render a social robot more apprehensible and approachable.

The prevailing effect of transparency over privacy risks was also reflected in users' qualitative responses, as participants mentioned that transparency demystified how the robot worked and made them "feel less afraid." Some even argued that robots' facial recognition systems could enhance their security, as emerging technologies such as laptops and smartphones have already adopted such technologies to protect their personal information. These qualitative responses complemented the quantitative findings and documented users' favorable attitudes toward the facial recognition technology in HRI.

Meanwhile, this study revealed more nuances regarding the interaction effect between the transparency of the facial recognition system and that of the speech recognition system. Specifically, when facial recognition technology was not transparent, the transparency of the speech recognition system eased users' privacy concerns. However, when facial recognition technology became transparent, the additional transparency of speech recognition system slightly exacerbated users' privacy concerns (Figure 3). Meanwhile, when both the speech recognition and the facial recognition systems were made transparent, users' privacy concerns were lower than when neither system was transparent. This finding implied that overall, transparency was useful in attenuating users' privacy concerns. However, compared to when only facial recognition was made transparent, the availability of both recognition systems would likely revive users' privacy concerns. Based on the finding, it can be further conjectured that as exposure to a robot's AI tracking technologies grows, the positive effects of transparency may gradually recede and users' caution regarding these tracking technologies may grow.

Results also suggested that although transparency of facial recognition relieved users' privacy concerns, it did not enhance users' trust in the social robot. Neither did the transparency of its speech recognition rates. These results contradict past XAI literature on the positive relationship between AI's transparency and credibility (Adadi and Berrada, 2018). Rather, Ananny and Crawford's (2018) perspective that transparency does not necessarily build trust was corroborated. Two factors could explain the non-significant effect here. First, while transparency might have elevated users' trust, such effect might have been counterbalanced when users realized that their faces or voiced were tracked and analyzed by these backstage AI systems. Second, given that the mean values of trust in each experiment condition (Table 2) were much higher than the mid-point of the seven-point scale, there might be a ceiling effect at play. A scale with a wider range might be helpful in establishing a more reliable relationship between the robot's transparency and perceived trustworthiness.

What merits further note is that this study focuses on the model-agnostic interpretations of AI's working mechanisms. Unlike model-specific interpretations that explain the technical structure and the learning mechanisms of the AI models, the model-agnostic interpretability seeks to use human comprehensible language to enhance laypersons' understanding of AI systems (Adadi and Berrada, 2018). Therefore, informing and showing participants a robot's facial recognition images and speech recognition rates after its failure to carry out smooth communication is merely one way to improve transparency. Indeed, Miller (2019) argued that explanations about AI are socially constructed and selective, suggesting that researchers may use various lenses to frame explanations about AI's performances. Examples of these lenses include how facial recognition works, how speech is analyzed, what facial features are used to train AI models, and what facial and speech data are stored in the robot. These different concentrations on explanations may substantially change users' perceptions of privacy risks and their attitudes toward AI's performances.

Taking a step further, as AI-based technologies, such as computer vision and speech detection become more automatic, learning-based, and data-centric, it is likely that future HRI will become more efficient and natural. While these AI technologies evolve over time, the explanations about the technology performances will receive increasing attention. Considering that non-AI experts may lack sufficient motivation for understanding AI models or feel overwhelmed by the technical features of AI technology (Kim et al., 2023), exploring and contemplating the methods that best enhance users' understanding of AI performances could guide future AI research. Just as Malle (2006) argued, explainers must not only gather evidence for explanations but also learn to communicate explanations. Thus, future research may benefit from understanding individuals' interactions with both AI technologies and the explanations about these AI technologies.

Finally, the mean values of privacy concerns across different conditions were lower than the mid-point of the scale, which dovetails with users' limited privacy concerns according to their qualitative responses. The low privacy concerns can be partially attributed to the limited disclosure of their personal information, as some participants believed that the robot did not obtain any personal information except their faces and voices. This finding implied that participants tended to rank their facial and vocal information as less important than other private information such as income or social security numbers. Especially given that participants had overall favorable attitudes toward the future adoption of AI tracking technologies in HRI, this finding could be alarming, as users' facial and speech traits may serve as unique biometric identifiers of their private information (Prabhakar et al., 2003), possibly leading to unexpected divulgence of more physiological and behavioral data, such as medical, travel, and purchase records (Lin et al., 2022).

Theoretical contributions

The study findings can provide fertile ground for theoretical development. First, the application of the CASA paradigm in prior HRI research has established a positive relationship between social cues and social responses (Krämer et al., 2015). This study expands the CASA paradigm and suggests that users are sensitive not only to the social cues that traditionally fall into interpersonal communication (e.g. human voices, gestures) but also to some non-human social cues that serve as unique features in HMC (e.g. facial recognition images, speech recognition rates). These non-human social cues may not be limited to the transparency-related cues tested in current research. They could involve how a robot presents LED lights or flashes to indicate different psychological states (Embgen et al., 2012; Rosenthal-Von der Pütten et al., 2018) or how a robot projects virtual arrows to indicate its directional intention (Shrestha et al., 2018). Thus, future HMC research should not only investigate the effects of human social cues designed into robots but also concentrate on the effects of these machine-generated, non-human social cues. Some initial steps might include listing all the possible non-human social cues that deliver social meanings, exploring how robots could present different constellations of human social cues and non-human social cues, and testing how users' social responses may vary based on different combinations of cues.

Second, this study presents an opportunity to revisit the conceptualization of social cues. Nass and Moon (2000) suggested that individuals socially and mindlessly respond to technologies because they have been repeatedly exposed to the social cues in interpersonal communication and thus tend to ignore "the cues that reveal the essential asocial nature" (p. 83). Yet, this study implies that the cues designed into machines may need to be examined in a more nuanced manner. As researchers found that individuals can identify, distinguish, and socially react to machine-generated non-human social cues (Rosenthal-Von der Pütten, 2018), a more subtle theorization of cues that characterizes their effects in eliciting users' social responses could be envisioned. For example, prior research has suggested that within human social cues, there exists a hierarchy of cues that evokes users' different levels of social reactions to technologies (Lombard and Xu, 2021). Hence, it is reasonable to postulate that within non-human social cues, some cues should also be more effective in raising users' social responses than others. In our study, the perusal of facial recognition images and speech recognition rates may merely serve as an entry point for further research on the conceptualization of "socialness" underlying non-human social cues.

Third, this study suggests that using XAI to improve transparency has its limits. While prior literature suggested that transparency can reduce users' privacy concerns, this study found that the transparency of AI tracking technologies in social robots evolved into a paradox where combining the transparency of both speech and facial recognition heightened participants' privacy concerns (compared to transparency of the facial recognition system only). Such privacy concerns could reversely harm users' trust, overshadow the effects of transparency, and even evoke negative feelings. Our findings implied that a situated balance between transparency and privacy risks may exist, as users' privacy concerns may swing based on how much transparency is provided and how tracking technologies are characterized. Thus, commensurate with Ananny and Crawford's (2018) critical analyses of transparency, a more systematic approach to theorizing the paradox of transparency is necessary in future XAI development.

Practical and ethical implications

As increasing the transparency of a robot's facial recognition system can reduce users' privacy concerns and enhance users' experience and acceptance of the robot as a social actor, developers may consider designing a secondary screen to show users what the robot sees through its cameras in future HRI. Especially considering that participants felt safer and perceived the interaction with the robot equipped with such technology as human-like and effective, more open and transparent HRI practices could be congenial to future robot users.

From an ethical perspective, using transparency cues to enhance users' medium-as-social-actor presence and technology acceptance could be perilous, especially considering that transparency, or explanations about AI's working mechanisms could be manipulated or selectively framed by explainers. Therefore, technology developers or researchers should implement proper regulations or ethical codes to ensure the safe use of technologies and the provision of reliable and responsible explanations. For example, developers or marketers should seek to receive users' consent to the capture of their facial, vocal, and other biometric information. They should also be open with users about who will have access to their data and how their biometric data will be used for purposes like data training and algorithmic recommendations (Marwick and boyd, 2014).

Conclusions and limitations

As social robots and their backstage AI technologies (e.g. object detection, facial recognition) enter society, users' demand for these AI systems to be transparent and explainable is growing (De Graaf et al., 2021). On one hand, users expect to have smooth communication with social robots. On the other hand, enabling such experience requires social robots to use AI technologies to quickly capture and predict users' physiological and psychological status. While explanations about these AI technologies could improve the transparency of how a robot works, facing the intrusion of these technologies may raise users' fears. Overall, this study suggests that proper use of transparency could relieve users' privacy concerns and increase users' acceptance of the robot. Although users have mixed feelings about the facial and speech recognition technologies in HRI, their overall attitudes toward the prospective application of these technologies tend to be favorable.

One caveat is that due to the variability of participants' voices, the times of NAO's nonresponses to each participant's commands at their first attempt could not be manipulated. Future research could try other approaches to keeping the robot's nonresponses consistent across conditions. Second, this study used a socio-emotional context. Future research could situate users' reactions in a task-oriented scenario and investigate whether the effects of transparency will change based on users' evaluation of robots' final decision-making. Finally, we acknowledge the limitations posed by the recruitment of a small sample of college students. It is important that future research should aim at larger and more diverse populations to enhance the external validity of our findings. Although our controlled lab experiment prioritized the internal validity of the findings, it may not fully capture HRI dynamics that occur in natural settings. Also, the sample of college students might have imposed influence on the results based on their characteristics (e.g. having a weaker sense of self, having higher-than-average cognitive skills; Basil, 1996). Moving forward, more research should be conducted to test our findings across varied environments and populations to build our understanding of the dynamics in HRI where AI systems have an increasing presence and influence.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This support from the Global Fellwoship, University of Florida.

ORCID iD

Kun Xu https://orcid.org/0000-0001-9044-821X

Supplemental material

Supplemental material for this article is available online: http://tinyurl.com/robottransparency.

Note

1. In the robot's default setting, it automatically responds to participants when its confidence of the speech recognition rate is over 50%.

References

Adadi A and Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 6: 52138–52160.

Ananny M and Crawford K (2018) Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20(3): 973–989.

Basil MD (1996) The use of student samples in communication research. *Journal of Broadcasting & Electronic Media* 40: 431–531.

Biocca F, Harms C and Burgoon JK (2003) Toward a more robust theory and measure of social presence: review and suggested criteria. *Presence: Teleoperators & Virtual Environments* 12(5): 456–480.

Bracken CC and Lombard M (2004) Social presence and children: praise, intrinsic motivation, and learning with computers. *Journal of Communication* 54(1): 22–37.

Chamikara MAP, Bertok P, Khalil I, et al. (2020) Privacy preserving face recognition utilizing differential privacy. *Computers & Security* 97: 101951.

Charmaz K (2014) *Constructing Grounded Theory*. Thousand Oaks, CA: Sage.

Chuah SHW, Aw ECX and Yee D (2021) Unveiling the complexity of consumers' intention to use service robots: an fsQCA approach. *Computers in Human Behavior* 123: 106870.

Cook TD (1985) Post-positivist critical multiplism. In: Shadish WR and Reichardt CS (eds) *Reproduced in Evaluation Studies Review Annual*. Thousand Oaks, CA: Sage, pp. 21–62.

Cummings JJ and Wertz EE (2023) Capturing social presence: concept explication through an empirical analysis of social presence measures. *Journal of Computer-Mediated Communication* 28(1): zmac027.

De Graaf MM, Dragan A, Malle BF, et al. (2021) Introduction to the special issue on explainable robotic systems. *ACM Transactions on Human-Robot Interaction* 10(3): 1–4.

Denzin NK (1978) *The Research Act: A Theoretical Introduction to Sociological Methods*. New York: McGraw Hill.

Edwards C, Edwards A, Spence PR, et al. (2016) Initial interaction expectations with robots: testing the human-to-human interaction script. *Communication Studies* 67(2): 227–238.

Embgen S, Luber M, Becker-Asano C, et al. (2012) Robot-specific social cues in emotional body language. In: *2012 IEEE RO-MAN: The 21st IEEE international symposium on robot and human interactive communication*, Paris, 9–13 September, pp. 1019–1025. New York: IEEE.

Fiore SM, Wiltshire TJ, Lobato EJ, et al. (2013) Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior. *Frontiers in Psychology* 4: 859.

Fischer K, Foth K, Rohlfing KJ, et al. (2011) Mindful tutors: linguistic choice and action demonstration in speech to infants and a simulated robot. *Interaction Studies* 12(1): 134–161.

Fischer K, Weigelin HM and Bodenhagen L (2018) Increasing trust in human–robot medical interactions: effects of transparency and adaptability. *Paladyn, Journal of Behavioral Robotics* 9(1): 95–109.

Gong L and Nass C (2007) When a talking-face computer agent is half-human and half-humanoid: human identity and consistency preference. *Human Communication Research* 33(2): 163–193.

Hesse-Biber S and Leavy P (2010) *The Practice of Qualitative Research*. Thousand Oaks, CA: Sage.

Johnson D, Gardner J and Wiles J (2004) Experience as a moderator of the media equation: the impact of flattery and praise. *International Journal of Human-Computer Studies* 61(3): 237–258.

Kanda T, Miyashita T, Osada T, et al. (2008) Analysis of humanoid appearances in human–robot interaction. *IEEE Transactions on Robotics* 24(3): 725–735.

Khurana A, Alamzadeh P and Chilana PK (2021) ChatrEx: designing explainable chatbot interfaces for enhancing usefulness, transparency, and trust. In: *2021 IEEE symposium on visual languages and human-centric computing*, St Louis, MO, 10–13 October, pp. 1–11. New York: IEEE.

Kim SS, Watkins EA, Russakovsky O, et al. (2023) Help me help the AI: understanding how explainability can support human-AI interaction. In: *Proceedings of the 2023 CHI conference on human factors in computing systems*, Hamburg, 23–28 April, pp. 1–17. New York: ACM.

Krämer NC, Rosenthal-Von der Pütten AM and Hoffmann L (2015) Social effects of virtual and robot companions. In: Sundar SS (ed.) *The Handbook of the Psychology of Communication Technology*. Hoboken, NJ: John Wiley & Sons, pp. 137–159.

Krupp MM, Rueben M, Grimm CM, et al. (2017) Privacy and telepresence robotics: what do non-scientists think? In: *Proceedings of the Companion of the 2017 ACM/IEEE international conference on human-robot interaction*, Vienna, 6–9 March, pp. 175–176. New York: ACM.

Kuchenbrandt D, Eyssel F, Bobinger S, et al. (2013) When a robot's group membership matters: anthropomorphization of robots as a function of social categorization. *International Journal of Social Robotics* 5: 409–417.

Law ELC, Soleimani S, Watkins D, et al. (2021) Automatic voice emotion recognition of child-parent conversations in natural settings. *Behaviour & Information Technology* 40(11): 1072–1089.

Lee EJ (2008) Flattery may get computers somewhere, sometimes: the moderating role of output modality, computer gender, and user gender. *International Journal of Human-Computer Studies* 66(11): 789–800.

Lee EJ (2023) Minding the source: toward an integrative theory of human–machine communication. *Human Communication Research* 50: hqad034.

Lee KM (2004) Presence, explicated. *Communication Theory* 14(1): 27–50.

Lee KM and Nass C (2005) Social-psychological origins of feelings of presence: creating social presence with machine-generated voices. *Media Psychology* 7(1): 31–45.

Lee KM, Park N and Song H (2005) Can a robot be perceived as a developing creature? Effects of a robot's long-term cognitive developments on its social presence and people's social responses toward it. *Human Communication Research* 31(4): 538–563.

Lee KM, Peng W, Jin SA, et al. (2006) Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of Communication* 56(4): 754–772.

Li C and Li H (2023) Disentangling facial recognition payment service usage behavior: a trust perspective. *Telematics and Informatics* 77: 101939.

Li JJ, Ju W and Reeves B (2017) Touching a mechanical body: tactile contact with body parts of a humanoid robot is physiologically arousing. *Journal of Human-Robot Interaction* 6(3): 118–130.

Lin PC, Yankson B, Chauhan V, et al. (2022) Building a speech recognition system with privacy identification information based on Google voice for social robots. *The Journal of Supercomputing* 78(13): 15060–15088.

Liu YL, Yan W and Hu B (2021) Resistance to facial recognition payment in China: the influence of privacy-related factors. *Telecommunications Policy* 45(5): 102155.

Lombard M and Ditton T (1997) At the heart of it all: the concept of presence. *Journal of Computer-Mediated Communication* 3(2): JCMC321.

Lombard M and Xu K (2021) Social responses to media technologies in the 21st century: the media are social actors paradigm. *Human-Machine Communication* 2: 29–55.

Lutz C, Schöttler M and Hoffmann CP (2019) The privacy implications of social robots: scoping review and expert interviews. *Mobile Media & Communication* 7(3): 412–434.

Malle BF (2006) *How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interaction*. Cambridge, MA: MIT Press.

Martínez-Plumed F, Gómez E and Hernández-Orallo J (2021) Futures of artificial intelligence through technology readiness levels. *Telematics and Informatics* 58: 101525.

Marwick AE and boyd d (2014) Networked privacy: how teenagers negotiate context in social media. *New Media & Society* 16: 1051–1067.

Miller T (2019) Explanation in artificial intelligence: insights from the social sciences. *Artificial Intelligence* 267: 1–38.

Nass C (2004) Etiquette equality: exhibitions and expectations of computer politeness. *Communications of the ACM* 47(4): 35–37.

Nass C, Fogg BJ and Moon Y (1996) Can computers be teammates? *International Journal of Human-Computer Studies* 45: 669–678.

Nass C and Moon Y (2000) Machines and mindlessness: Social responses to computers. *Journal of Social Issues* 56(1): 81–103.

Nass C, Moon Y and Green N (1997) Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology* 27: 864–876.

Nass C, Steuer J and Tauber ER (1994) Computers are social actors. In: *Proceedings of SIGCHI '94 human factors in computing systems*, Boston, MA, 24–28 April, pp. 72–78. New York: ACM.

Nesset B, Robb DA, Lopes J, et al. (2021) 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*, Boulder, CO, 8–11 March, pp. 313–317. New York: ACM.

Pantano E (2020) Non-verbal evaluation of retail service encounters through consumers' facial expressions. *Computers in Human Behavior* 111: 106448.

Peng Y (2022) The role of ideological dimensions in shaping acceptance of facial recognition technology and reactions to algorithm bias. *Public Understanding of Science* 32: 190–207.

Petrescu RV (2019) Face recognition as a biometric application. *Journal of Mechatronics and Robotics* 3: 237–257.

Prabhakar S, Pankanti S and Jain AK (2003) Biometric recognition: security and privacy concerns. *IEEE Security & Privacy* 1(2): 33–42.

Pradhan A, Lazar A and Findlater L (2020) Use of intelligent voice assistants by older adults with low technology use. *ACM Transactions on Computer-Human Interaction* 27(4): 1–27.

Rai A (2020) Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science* 48: 137–141.

Reeves B and Nass C (1996) *The Media Equation: How People Treat Computers, Television, and New Media Like Real People*. Cambridge: Cambridge University Press.

Reeves B, Hancock J and Liu X (2020) Social robots are like real people: first impressions, attributes, and stereotyping of social robots. *Technology, Mind, and Behavior* 1(1): 1–37.

Reeves B, Yeykelis L and Cummings JJ (2016) The use of media in media psychology. *Media Psychology* 19(1): 49–71.

Richards RJ, Spence PR and Edwards CC (2022) Human-machine communication scholarship trends: an examination of research from 2011 to 2021 in communication journals. *Human-Machine Communication* 4: 45–62.

Robinson S and Mendelson AL (2012) A qualitative experiment: research on mediated meaning construction using a hybrid approach. *Journal of Mixed Methods Research* 6(4): 332–347.

Rosenthal-Von der Pütten AM, Krämer NC and Herrmann J (2018) The effects of humanlike and robot-specific affective nonverbal behavior on perception, emotion, and behavior. *International Journal of Social Robotics* 10: 569–582.

Schadenberg BR, Reidsma D, Heylen DK, et al. (2021) "I see what you did there" understanding people's social perception of a robot and its predictability. *ACM Transactions on Human-Robot Interaction* 10(3): 1–28.

Shrestha MC, Onishi T, Kobayashi A, et al. (2018) Communicating directional intent in robot navigation using project indicators. *Proceedings of the 27th IEEE international symposium on robot and human interactive communication*, Nanjing, China, 27–31 August, pp. 27–38. New York: IEEE.

Song Z (2020) English speech recognition based on deep learning with multiple features. *Computing* 102(3): 663–682.

Spence PR, Westerman D, Edwards C, et al. (2014) Welcoming our robot overlords: initial expectations about interaction with a robot. *Communication Research Reports* 31(3): 272–280.

Suchman LA (2007) *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge: Cambridge University Press.

Suen HY and Hung KE (2023) Building trust in automatic video interviews using various AI interfaces: tangibility, immediacy, and transparency. *Computers in Human Behavior* 143: 107713.

Vitale J, Tonkin M, Herse S, et al. (2018) Be more transparent and users will like you: a robot privacy and user experience design experiment. In: *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, Chicago, IL, 5–8 March, pp. 379–387. New York: ACM.

Weitz K, Schiller D, Schlagowski R, et al. (2019) "Do you trust me?" Increasing user-trust by integrating virtual agents in explainable AI interaction design. *Proceedings of the 19th ACM international conference on intelligent virtual agents*, Paris, 2–5 July, pp. 7–9. New York: ACM.

Wheeless LR and Grotz J (1977) The measurement of trust and its relationship to self-disclosure. *Human Communication Research* 3(3): 250–257.

Wilkinson D, Alkan Liao ÖQV, Mattetti M, et al. (2021) Why or why not? The effect of justification styles on chatbot recommendations. *ACM Transactions on Information Systems* 39(4): 1–21.

Wolf CT and Ringland KE (2020) Designing accessible, explainable AI (XAI) experiences. *ACM SIGACCESS Accessibility and Computing* 125: 6.

Xie Y, Zhu K, Zhou P, et al. (2023) How does anthropomorphism improve human-AI interaction satisfaction: a dual-path model. *Computers in Human Behavior* 148: 107878.

Xu K (2019) First encounter with robot Alpha: how individual differences interact with vocal and kinetic cues in users' social responses. *New Media & Society* 21(11–12): 2522–2547.

Xu K and Liao T (2020) Explicating cues: a typology for understanding emerging media technologies. *Journal of Computer-Mediated Communication* 25(1): 32–43.

Zhang Y, Liao QV and Bellamy RK (2020) Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In: *Proceedings of the 2020 conference on fairness, accountability, and transparency*, Barcelona, 27–30 January, pp. 295–305. New York: ACM.

Zhao S (2006) Humanoid social robots as a medium of communication. *New Media & Society* 8(3): 401–419.

Author biography

Kun Xu (PhD, Temple University) is an Assistant Professor of Emerging Media in the College of Journalism and Communications, University of Florida, United States. His research centers on human-robot interaction, human-computer interaction, and psychological processing of media. He investigates individuals' social responses to emerging technologies, including social robots, virtual agents, and mobile assistants.

Xiaobei Chen (MA, University of Florida) is a doctoral student in the College of Journalism and Communications, University of Florida, United States. Her research focuses on using emerging technologies to promote health equity and to conduct communication skill training in healthcare.

Fanjue Liu (MA, University of Florida) is a PhD candidate in the College of of Journalism and Communications, University of Florida, United States. Her research aims to dissect the mechanisms through which emerging media technologies influence human

behaviour, exploring how these technologies shape human attribution, cognitive processes, and decision-making.

Luling Huang (PhD, Temple University) is an Assistant Professor in the Department of Communication at Missouri Western State University. His work focuses on social influence, media effects, and public opinion using quantitative methods.