Privacy and Personalization: Transparency, Acceptance, and the Ethics of Personalized Robots

Mariah L. Schrum
Georgia Institute of Technology
Atlanta, United States
mschrum3@gatech.edu

Matthew C. Gombolay
Georgia Institute of Technology
Atlanta, United States
matthew.gombolay@cc.gatech.edu

ABSTRACT

To effectively support humans, machines must be capable of recognizing individual desires, abilities, and characteristics and adapt to account for differences across individuals. However, personalization does not come without a cost. In many domains, for robots to effectively personalize their behavior, the robot must solicit often private and intimate information about an end-user so as to optimize the interaction. However, not all end-users may be comfortable sharing this information, especially if the end-user is not provided with insight into why the robot is requesting it. As HRI researchers, we have the responsibility of ensuring the robots we create do not infringe upon the privacy rights of end-users and that end-users are provided with the means to make informed decisions about the information they share with robots. While prior work has investigated willingness to share information in the context of consumerism, no prior work has investigated the impact of domain, type of requested information, or explanations on end-user's comfort and acceptance of a personalized robot. To gain a better understanding of these questions, we propose an experimental design in which we investigate the impact of domain, nature of personal information requested, and the role of explanations on robot transparency and end-user willingness to share information. Our goal of this study is to provide guidance for HRI researchers who are conducting work in personalization by examining the factors that may impact transparency and acceptance of personalized robots.

KEYWORDS

personalization, privacy, ethics, transparency, xAI, explainability

ACM Reference Format:
Mariah L. Schrum and Matthew C. Gombolay. 2023. Privacy and Personalization: Transparency, Acceptance, and the Ethics of Personalized Robots . In *Proceedings of (HRI '23).* ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The high degree of heterogeneity amongst humans means that each end-user is a unique system that robots must learn about and adapt to so as to optimize the human-robot relationship. There are many latent variables that govern human preferences, behavior, and decision making that robots must take into consideration. For example, an individual's personality and prior experiences may impact their level of trust in an autonomous system [1]. Biological differences stemming from both genetics and environment influence how an AI system should support a patient in a healthcare setting [8].

Prior work has investigated approaches for personalizing robot behavior [1, 4–6, 9, 10]. For example, Schrum et al. introduced Reciprocal MIND MELD, an approach for personalized robotic coaching [10]. This approach learns about the way in which an end-user in suboptimal in a given domain and provides robotic feedback to improve upon end-user's suboptimality. Basu et al. introduced an approach for personalizing the driving style of an autonomous vehicle (AV) to match the expectations of an end-user [1]. The authors found that the end-user's own driving style and their perception of their own driving style had an impact on the optimal driving style of the AV.

While these approaches have showed promise for personalization, they come with a caveat. Because of the nature of personalization, many personalized approaches require the robot to learn intimate details about the user. For example, in the work by Basu et al., aspects of the user's personality and their own driving style are important predictors for the optimal driving style [1]. In Schrum et al. robots require knowledge about sensitive, HIPAA information, including details about the patient's biology and how the patient responds to certain treatments, so as to optimize the treatment plan [9]. Many users may feel uncomfortable sharing this detailed personal information with the robot, especially if they are not aware of how the robot intends to use the information. Furthermore, prior work has indicated that learned private policies may still be vulnerable to adversarial attacks, suggesting that end-users are right to be wary of sharing private information with agents [7].

As HRI researchers, we should design systems that respect the privacy rights of end-users while also optimizing for performance. We must ensure that end-users are equipped with adequate information to make an informed decision about sharing personal information with a robot. To do so, we must first gain insight into the factors that affect understanding, acceptance, and willingness to share information with a robot. To gain insight into these factors, we investigate three research questions:

(1) What personal information are end-users willing to share with a robot?
(2) How does domain impact acceptance of a personalized robot?
(3) How can xAI approaches increase transparency and understanding of a personalized robot?

HRI '23, March 13–16, 2023, Stockholm, Sweden

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

2 METHODOLOGY

We investigate these questions in a human subject study in which we manipulate three independent variables: domain of interaction, personal information requested, and presence of explanations. We additionally propose to collect covariates of interest related to end-user characteristics and attitudes towards robots. Our goal is to investigate the causal relationship between these factors and end-user's attitudes towards sharing personal information with robots.

2.1 Conditions

Below we describe our independent variables. We aim to determine how each variable impacts the end-user's attitude and willingness to share personal information with the robot. The domain of interaction and personal information requested are within-subject variables. The xAI factor is a between-subjects variable.

Domain of Interaction: We explore three domains of interaction to determine how the situation and circumstances governing the interaction impact an end-user's willingness to divulge personal information to a robot. We investigate coaching, AV, and healthcare domains. While there are many other potential domains, we choose these three because prior work has indicated the importance of personalization in each of these domains, and these domains cover a diverse range of scenarios with varying levels of intimacy [1, 9, 10]. All participants in the study experience each domain.

- *Coaching Domain:* In the coaching domain, we construct a scenario in which the end-user is a novice table tennis player and is being coached by a robot. To personalize its instructions and feedback for the end-user, the robotic coach must learn about the way in which the human is suboptimal with regards to table tennis.
- *AV Domain:* In the AV domain, the AV aims to personalize its driving style to match the preference of the end-user. To select the optimal driving style, the AV must learn about the end-user's own driving style and personality.
- *Healthcare Domain:* In the healthcare domain, the robot is tasked with creating and deploying a personalized plan to treat the patient's disease. To determine the best health plan, the robot must learn about the biology of the patient, their disease manifestation, and their medical history.

Personal Information Requested: To personalize its behavior, a robot will need to access personal information about an end-user. The sensitivity of the information may vary depending on the domain of interaction. For example, healthcare information is considered private data and is protected by HIPAA. Yet this information may be crucial for a healthcare robot to have access to for making informed decisions related to patient care. In the three domains discussed above, each participant will experience the robot requesting each type of information listed below. Even though health information may not seem relevant in, for instance, a tutoring domain, we include each of these conditions in each domain to determine if it is the domain or the type of personal information requested that impacts the attitude of the end-user. All participants in the study experience each request condition.

- *Competence:* In this condition, the robot will request information with regards to the end-user's skill at the task.
- *Personality:* In this condition, the robot will request information about the end-user's personality.
- *Healthcare Information:* In this condition, the robot will request sensitive healthcare information.

xAI: To improve the human-robot relationship, robots should be transparent about why they are requesting information from end-users. This transparency will allow end-users to make informed decisions about whether or not they wish to share personal information with the robot. Without context as to why the robot is requesting specific information, end-users may be less willing to divulge information due to the uncertainty involved.

For example, in an AV domain, soliciting information about the personality of the end-user may appear superfluous. However, as shown in prior work, personality may be an important factor for determining the optimal driving style [2]. If the robot communicates *why* it requires the personal information and the consequences of not receiving it, end-users will be better equipped to decide for themselves if they wish to share personal information. However, in some situations, we hypothesize that providing an explanation may decrease end-user's willingness to share information as it may draw additional attention to the robot's request and in some cases the end-user may not agree with the robot's justification. We investigate these hypotheses by introducing two additional conditions: explanation, and no explanation. This factor is between-subjects to avoid confounds from the explanations.

- *Explanation:* In this condition, the robot offers an explanation as to why it is requesting the personal information. Additionally, the robot explains the consequences of not having the personal information.
- *No Explanation:* In this condition, no explanation as to why the robot is requesting the information is provided.

2.2 Metrics

To determine which variables impact an individual's willingness to share information, we ask participants on a single Likert item with response scale from one to ten how willing they are to share the requested information with the robot. One of our goals of this study is to investigate if providing explanations to end-users increases the transparency of the system and provides the end-user with a better ability to make an informed decision with regards to sharing personal information. Therefore, we measure a participant's understanding of how the robot works [12] and the robot's transparency [11]. Lastly, we measure participant's comfort with the robot via the ROSAS discomfort scale [3].

3 EXPECTED RESULTS AND IMPLICATIONS

We expect to find that the domain in which the robot is requesting information will have a large impact on end-user willingness to provide personal information. Additionally, we hypothesize that users will be more comfortable providing specific information if this information aligns with their expectations in the domain. Lastly, we expect that, in most circumstances, explanations will increase understanding, comfort, and intention to use.

Our goal of this work is to shed light on issues regarding privacy and personalization in human-robot interaction. By gaining an understanding of end-users' willingness to sharing information

in specific domains, we can better understand how to design HRI system that respect the rights of the end-user and and refrain from requesting and relying on information that end-users do not wish to share. Additionally, by investigating the impact of explanations when querying end-users for personal information, we aim to provide guidance for how to maximize system transparency so that end-users can make informed decisions about information sharing.

REFERENCES

[1] Chandrayee Basu, Qian Yang, David Hungerman, Mukesh Sinahal, and Anca D. Draqan. 2017. Do You Want Your Autonomous Car to Drive Like You?. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction*. 417–425.

[2] Hanna Bellem, Barbara Thiel, Michael Schrauf, and Josef F. Krems. 2018. Comfort in automated driving: An analysis of preferences for different automated driving styles and their dependence on personality traits. *Transportation Research Part F: Traffic Psychology and Behaviour* 55 (2018), 90–100. https://doi.org/10.1016/j.trf.2018.02.036

[3] Colleen M. Carpinella, Alisa B. Wyman, Michael A. Perez, and Steven J. Stroessner. 2017. The Robotic Social Attributes Scale (RoSAS): Development and Validation. *ACM/IEEE International Conference on Human-Robot Interaction* Part F1271 (2017), 254–262. https://doi.org/10.1145/2909824.3020208

[4] Fredrick Ekman, Mikael Johansson, Lars-Ola Bligård, MariAnne Karlsson, and Helena Strömberg. 2019. Exploring automated vehicle driving styles as a source of trust information. *Transportation Research Part F: Traffic Psychology and Behaviour* (2019).

[5] Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. 2016. Affective Personalization of a Social Robot Tutor for Children's Second Language Skills. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (Phoenix, Arizona) *(AAAI'16)*. AAAI Press, 3951–3957.

[6] Daniel Leyzberg, Aditi Ramachandran, and Brian Scassellati. 2018. The Effect of Personalization in Longer-Term Robot Tutoring. *ACM Transactions on Human-Robot Interaction* 7 (12 2018). Issue 3. https://doi.org/10.1145/3283453

[7] Kritika Prakash, Fiza Husain, Praveen Paruchuri, and Sujit Gujar. 2022. How Private Is Your RL Policy? An Inverse RL Based Analysis Framework. *Proceedings of the AAAI Conference on Artificial Intelligence* 36, 7 (Jun. 2022), 8009–8016. https://doi.org/10.1609/aaai.v36i7.20772

[8] Lluis Quintana-Murci. 2016. Genetic and epigenetic variation of human populations: An adaptive tale. *Comptes Rendus Biologies* 339, 7 (2016), 278–283. https://doi.org/10.1016/j.crvi.2016.04.005 Trajectories of genetics, 150 years after Mendel / Trajectoire de la génétique, 150 après Mendel Guest Editors / Rédacteurs en chef invités : Bernard Dujon, Georges Pelletier.

[9] Mariah Schrum, Mark J Connolly, Eric Cole, Mihir Ghetiya, Robert Gross, and Matthew C. Gombolay. 2022. Meta-Active Learning in Probabilistically Safe Optimization. *IEEE Robotics and Automation Letters* 7, 4 (2022), 10713–10720.

[10] Mariah L Schrum, Erin Hedlund-Botti, and Matthew C Gomoblay. 2022. Reciprocal MIND MELD: Improving Learning From Demonstration via Personalized, Reciprocal Teaching. *Conference on Robot Learning* (2022).

[11] Graduate Theses and Jennifer Dapko. 2012. Perceived Firm Transparency: Scale and Model Development. https://digitalcommons.usf.edu/etd

[12] Giulia Troisi. 2021. Humanization builds trust the effect of human-like chatbots on the willingness to disclose personal information online. *Luiss University Department of Business and Management* (2021).

Received 20 Jan 2023; revised 1 March 2023; accepted 16 Feb 2023