Protecting User Data Through Privacy-Sensitive Robot Design

Dakota Sullivan
Department of Computer Sciences
University of Wisconsin–Madison
Madison, WI, USA
dsullivan8@wisc.edu

Bilge Mutlu
Department of Computer Sciences
University of Wisconsin–Madison
Madison, WI, USA
bilge@cs.wisc.edu

***Abstract*—While robots possess many capabilities that may positively influence human lives, their autonomous navigation and sensing capabilities pose threats to user privacy. These threats may be addressed at three key phases: *data collection*, *data retention*, and *data exposure*. In this work, we discuss our prior, current, and proposed robot design efforts to reduce privacy violations during human-robot interaction (HRI). At the data collection phase, we are currently exploring designs that enable robots to inhibit data collection by blocking their own sensors. At the data retention phase, we propose the exploration of privacy preferences to inform designs that grant users greater control over retained data. Finally, in the data exposure phase, we discuss our prior works developing a privacy controller for appropriate data exposure and generating task-planning strategies to limit unintentional data exposure. Through this work, we hope to protect user data and reduce the likelihood of harm to users.**

***Index Terms*—privacy by design; privacy-preserving; data privacy**

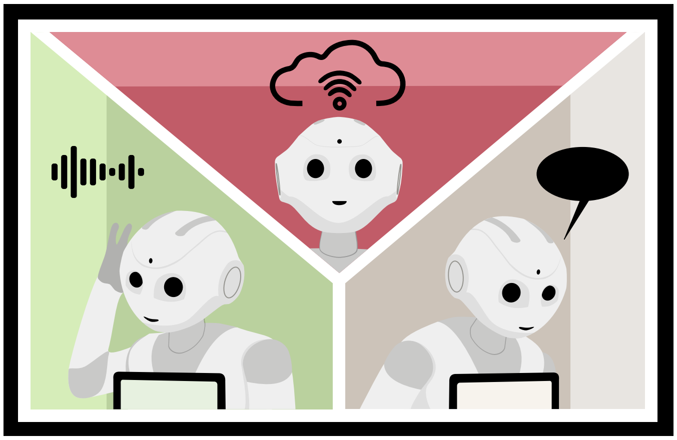

Fig. 1. A robot collecting audio data (left), processing and storing that data (top), and exposing that data to others (right).

I. Introduction

Robots have clear transformative potential to positively impact users, such as older adults [1] or patients and medical professionals in healthcare settings [2]. As robots become increasingly present in human spaces, however, they are in turn gaining increasing access to user data. Through a wide array of actuators and sensors, many robots are able to autonomously move within these spaces and collect continuous streams of data about their environments and nearby humans [3]–[7]. Ambient data collection is a well documented privacy concern in stationary technologies such as smart home [8]–[12] and surveillance [13], [14] devices, and a robot's ability to autonomously move significantly elevates the potential for related risks. Any data that is collected by the robot may subsequently be retained by companies [15], [16], exfiltrated through adversarial attacks [17], [18], disclosed intentionally by the robot to other users [19], or inferred based on the robot's behavior [20]. As a result, a user's privacy may be violated to varying degrees of severity.

With proper robot design, however, these risks may be mitigated. Existing work has attempted to address privacy concerns through design in a variety of ways. Some strategies have focused on limitations to data collection through sensor-redirection and avoidance [21]–[23], appropriate sensor selection [24], and blurred or low-resolution data capture [25], [26]. Other works have centered on limitations to unintentional data exposure during task execution [20], [27], [28]. While these strategies reduce some risks to user privacy, many gaps remain.

In this work, we discuss our prior, current and proposed efforts to design robots for greater privacy-sensitivity. Each of these works is intended to address privacy risks at the phases of data *collection*, *retention*, and *exposure* utilizing a research-through-design approach [29]–[31]. Our primary goal is to *develop privacy-sensitive robot designs that protect user data and mitigate risks to everyday users with whom the robot comes into contact*. Throughout our work, we seek to address the following questions:

- **RQ1**: How can a robot inhibit sensitive data collection in a manner that is apparent and intuitive to users?
- **RQ2**: How can users gain greater awareness and control of their data when it is retained by a robot?
- **RQ3**: How can inappropriate data exposure by a robot be mitigated?

II. Inhibiting Sensitive Data Collection

To address privacy risks at the data collection phase, we are currently exploring physical robot designs that enable a robot to block its own sensors. As compared to existing work that avoids sensitive data collection by redirecting sensors or moving a robot away from data [21]–[23], we intend to develop a robot prototype that is capable of blocking data collection altogether. These designs may involve additive mechanisms to temporarily cover sensors or utilize robot arms to block sensors in a manner that resembles a human blocking their eyes or ears. Ours is similar to an existing design that utilizes an ultrasound jamming device to prevent ambient smart speaker data collection [32]. Like this speaker design, our designs will allow a robot to remain in the presence of sensitive data rather than moving away from it. Additionally, blocking sensors will entirely inhibit data collection rather than redirecting collection toward other stimuli, in the case of visual sensors, or moving out of audible range, in the case of auditory sensors. Once we have developed our initial prototypes, we plan to evaluate our designs in a user study. This evaluation will involve the collection of data with unobstructed sensors, followed by collection with sensors covered. Participants may then review the collected data and report their perception of the robot's privacy-preserving capabilities (*e.g.,* trust perception). Through this work, we will begin to address **RQ1**.

III. Expanding User Control of Data Retention

While some data are unnecessary for a robot to properly function, and can therefore be avoided, other data are required to accurately navigate, successfully complete tasks, and adequately meet user preferences. In these cases, there is an inherent trade-off between effective functionality and the privacy risks that emerge from data retention by the robot. To better understand this trade-off and develop corresponding designs, we aim to (1) explore the privacy preferences of users in the presence of privacy threats and (2) generate designs that empower users to review and remove retained data in accordance with these preferences.

Exploring User Privacy Preferences – To understand user preferences in this context, we would like to conduct an exploratory user study in which participants interact with a robot that either has or does not have access to information about that participant. This information can be provided by the participant prior to the study session. Across multiple scenarios, the robot can illustrate the advantages (*i.e.,* meeting user preferences) and disadvantages (*i.e.,* the ability to expose information to the experimenter) of retained data. Conversely, a robot that does not possess this information can showcase corresponding advantages (*i.e.,* no capacity to violate user privacy) and disadvantages (*i.e.,* being unaware of user preferences). Through this study, we may better understand how users interpret and weigh the risks and benefits of data retention by a robot. Similar studies have been conducted to explore privacy preferences and expectations [33]–[36], privacy trade-offs [37], [38], and the privacy paradox (*i.e.,* the disconnect between user privacy concerns and behavior) [39]. To our knowledge, these studies do not, however, simulate the benefits and risks of retained data through real human-robot interactions. Through this study, we would also like to explore how users prefer to access the information a robot possesses. This process may take the form of verbal interactions with the robot or review of data through an accompanying interface.

Developing Data Management Prototypes – Based on the findings of our privacy preference exploration and existing usable privacy heuristics [40], we would also like to begin developing prototype designs that allow users to review and remove data retained by a robot. Depending on user preferences regarding the amount or type of retained data, or the method by which users can access these data (*e.g.,* verbal disclosure by the robot, direct access to video or audio recordings, or guided review on a robot's screen), our prototype designs may differ drastically. Once our initial designs are prepared, we can comparatively evaluate them with users (*e.g.,* perception of usability) and iteratively improve upon our designs. Regardless of the specific implementation, we aim to improve the transparency of retained data and grant users greater control over these data. Through these proposed works, we hope to address **RQ2**.

IV. Managing Data Exposure

Based on the information a robot is allowed to collect and retain, it may pose a negligible or substantial threat to user privacy. However, these threats may be contained if a robot is able to (1) appropriately navigate scenarios in which it is expected to intentionally expose or withhold information and (2) reduce unintentional data exposure implied by its actions.

Facilitating Appropriate Intentional Exposure – Within our prior work, we have developed a privacy controller that is able to determine the sensitivity level of a user's data disclosure based on several contextual factors (*i.e.,* the sentiment, topic, and details of the disclosure, along with details about those involved in the disclosure) [19]. We then evaluated the effectiveness and accuracy of the controller through a large online user study that showcased a robot exposing fictional user data based on its determined sensitivity level. Our findings demonstrated that a robot utilizing the controller was perceived as more privacy aware, trustworthy, and socially aware than a baseline alternative. These results showed great promise for the use of privacy controllers in human-robot interaction to reduce inappropriate data exposure. We would therefore like to build on these efforts in our future work by creating a more sophisticated controller that is able to capture greater nuance in complex data disclosure scenarios.

Limiting Unintentional Exposure – In our most recent work, we additionally explored unintentional data exposure that occurs when a robot's actions imply its overarching task-related goals. To reduce the likelihood of this type of exposure, we generated three deceptive task-planning strategies (*i.e., alternating*, *multitasking*, and *detour*) intended to obfuscate the robot's goals and protect any information those goals might imply about the users the robot served. Through our evaluation of these strategies, we determined that all three were able to reduce the likelihood of correct goal identification, however, some strategies appeared unlikely to convince an observer that the robot had alternative false goals. Through these prior works, we have begun to answer **RQ3**.

References

[1] Y. Hu, L. Stegner, Y. Kotturi, C. Zhang, Y.-H. Peng, F. Huq, Y. Zhao, J. P. Bigham, and B. Mutlu, “”this really lets us see the entire world:” designing a conversational telepresence robot for homebound older adults,” in *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, ser. DIS ’24. New York, NY, USA: Association for Computing Machinery, 2024, p. 2450–2467. [Online]. Available: https://doi.org/10.1145/3643834.3660710

[2] A. A. Morgan, J. Abdi, M. A. Syed, G. E. Kohen, P. Barlow, and M. P. Vizcaychipi, “Robots in healthcare: a scoping review,” *Current robotics reports*, vol. 3, no. 4, pp. 271–280, 2022.

[3] J. Torresen, “A review of future and ethical perspectives of robotics and ai,” *Frontiers in Robotics and AI*, vol. 4, p. 75, 2018.

[4] U. Pagallo, M. Corrales, M. Fenwick, and N. Forgó, “The rise of robotics & ai: technological advances & normative dilemmas,” *Robotics, AI and the Future of Law*, pp. 1–13, 2018.

[5] C. Lutz, M. Schöttler, and C. P. Hoffmann, “The privacy implications of social robots: Scoping review and expert interviews,” *Mobile Media & Communication*, vol. 7, no. 3, pp. 412–434, 2019.

[6] M. Rueben, A. M. Aroyo, C. Lutz, J. Schmölz, P. Van Cleynenbreugel, A. Corti, S. Agrawal, and W. D. Smart, “Themes and research directions in privacy-sensitive robotics,” in *2018 IEEE workshop on advanced robotics and its social impacts (ARSO)*. IEEE, 2018, pp. 77–84.

[7] B. Cheatham, K. Javanmardian, and H. Samandari, “Confronting the risks of artificial intelligence,” *McKinsey Quarterly*, vol. 2, no. 38, pp. 1–9, 2019.

[8] J. Lau, B. Zimmerman, and F. Schaub, “Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers,” *Proceedings of the ACM on human-computer interaction*, vol. 2, no. CSCW, pp. 1–31, 2018.

[9] A. Sciuto, A. Saini, J. Forlizzi, and J. I. Hong, “”hey alexa, what’s up?”: A mixed-methods studies of in-home conversational agent usage,” in *Proceedings of the 2018 Designing Interactive Systems Conference*, ser. DIS ’18. New York, NY, USA: Association for Computing Machinery, 2018, p. 857–868. [Online]. Available: https://doi.org/10.1145/3196709.3196772

[10] E. Beneteau, A. Boone, Y. Wu, J. A. Kientz, J. Yip, and A. Hiniker, “Parenting with alexa: exploring the introduction of smart speakers on family dynamics,” in *Proceedings of the 2020 CHI conference on human factors in computing systems*, 2020, pp. 1–13.

[11] N. Malkin, S. Egelman, and D. Wagner, “Privacy controls for always-listening devices,” in *Proceedings of the New Security Paradigms Workshop*, 2019, pp. 78–91.

[12] N. Meng, D. Keküllüoğlu, and K. Vaniea, “Owning and sharing: Privacy perceptions of smart speaker users,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CSCW1, pp. 1–29, 2021.

[13] S. Shalawadi, C. Getschmann, N. van Berkel, and F. Echtler, “Manual, hybrid, and automatic privacy covers for smart home cameras,” in *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, 2024, pp. 3453–3470.

[14] M. Kashef, A. Visvizi, and O. Troisi, “Smart city as a smart service system: Human-computer interaction and smart city surveillance systems,” *Computers in Human Behavior*, vol. 124, p. 106923, 2021.

[15] A. Chatzimichali, R. Harrison, and D. Chrysostomou, “Toward privacy-sensitive human–robot interaction: Privacy terms and human–data interaction in the personal robot era,” *Paladyn, Journal of Behavioral Robotics*, vol. 12, no. 1, pp. 160–174, 2020.

[16] U. Pagallo, “Robots in the cloud with privacy: A new threat to data protection?” *Computer Law & Security Review*, vol. 29, no. 5, pp. 501–508, 2013.

[17] T. Denning, C. Matuszek, K. Koscher, J. R. Smith, and T. Kohno, “A spotlight on security and privacy risks with future household robots: attacks and lessons,” in *Proceedings of the 11th international conference on Ubiquitous computing*, 2009, pp. 105–114.

[18] F. J. R. Lera, C. F. Llamas, Á. M. Guerrero, and V. M. Olivera, “Cybersecurity of robotics and autonomous systems: Privacy and safety,” *Robotics-legal, ethical and socioeconomic impacts*, 2017.

[19] B. Tang, D. Sullivan, B. Cagiltay, V. Chandrasekaran, K. Fawaz, and B. Mutlu, “Confidant: A privacy controller for social robots,” in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2022, pp. 205–214.

[20] P. Masters and S. Sardina, “Deceptive path-planning.” in *IJCAI*, 2017, pp. 4368–4375.

[21] R. Shome, Z. Kingston, and L. E. Kavraki, “Robots as ai double agents: Privacy in motion planning,” in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 2861–2868.

[22] D. Yang, Y.-J. Chae, D. Kim, Y. Lim, D. H. Kim, C. Kim, S.-K. Park, and C. Nam, “Effects of social behaviors of robots in privacy-sensitive situations,” *International Journal of Social Robotics*, pp. 1–14, 2022.

[23] F. E. Fernandes, G. Yang, H. M. Do, and W. Sheng, “Detection of privacy-sensitive situations for social robots in smart homes,” in *2016 IEEE International Conference on Automation Science and Engineering (CASE)*. IEEE, 2016, pp. 727–732.

[24] S. Eick and A. I. Antón, “Enhancing privacy in robotics via judicious sensor selection,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 7156–7165.

[25] Y. Hu, S. M. Bejarano, and G. Hoffman, “Shadowsense: Detecting human touch in a social robot using shadow image classification,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 4, pp. 1–24, 2020.

[26] M. U. Kim, H. Lee, H. J. Yang, and M. S. Ryoo, “Privacy-preserving robot vision with anonymized faces by extreme low resolution,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 462–467.

[27] A. D. Dragan, R. M. Holladay, and S. S. Srinivasa, “An analysis of deceptive robot motion.” in *Robotics: science and systems*. Citeseer, 2014, p. 10.

[28] A. Price, R. F. Pereira, P. Masters, and M. Vered, “Domain-independent deceptive planning.” in *AAMAS*, 2023, pp. 95–103.

[29] M. Luria, M. Hoggenmüller, W.-Y. Lee, L. Hespanhol, M. Jung, and J. Forlizzi, “Research through design approaches in human-robot interaction,” in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 685–687.

[30] M. Luria, J. Zimmerman, and J. Forlizzi, “Championing research through design in hri,” *arXiv preprint arXiv:1908.07572*, 2019.

[31] J. Zimmerman, E. Stolterman, and J. Forlizzi, “An analysis and critique of research through design: towards a formalization of a research approach,” in *proceedings of the 8th ACM conference on designing interactive systems*, 2010, pp. 310–319.

[32] V. Chandrasekaran, S. Banerjee, B. Mutlu, and K. Fawaz, “{PowerCut} and obfuscator: An exploration of the design space for {Privacy-Preserving} interventions for smart speakers,” in *Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021)*, 2021, pp. 535–552.

[33] L. Levinson, C. Nippert-Eng, R. Gomez, and S. Sabanović, “Snitches get unplugged: Adolescents’ privacy concerns about robots in the home are relationally situated,” in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, 2024, pp. 423–432.

[34] B. Cagiltay, H.-R. Ho, J. E. Michaelis, and B. Mutlu, “Investigating family perceptions and design preferences for an in-home robot,” in *Proceedings of the interaction design and children conference*, 2020, pp. 229–242.

[35] S. Chatterjee, R. Chaudhuri, and D. Vrontis, “Usage intention of social robots for domestic purpose: from security, privacy, and legal perspectives,” *Information Systems Frontiers*, pp. 1–16, 2024.

[36] M. M. Krupp, M. Rueben, C. M. Grimm, and W. D. Smart, “A focus group study of privacy concerns about telepresence robots,” in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2017, pp. 1451–1458.

[37] D. J. Butler, J. Huang, F. Roesner, and M. Cakmak, “The privacy-utility tradeoff for remotely teleoperated robots,” in *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*, 2015, pp. 27–34.

[38] I. Leite and J. F. Lehman, “The robot who knew too much: Toward understanding the privacy/personalization trade-off in child-robot conversation,” in *Proceedings of the The 15th International Conference on Interaction Design and Children*, 2016, pp. 379–387.

[39] C. Lutz and A. Tamó-Larrieux, “The robot privacy paradox: Understanding how privacy concerns shape intentions to use social robots,” *Human-Machine Communication*, vol. 1, pp. 87–111, 2020.

[40] P. Jaferian, K. Hawkey, A. Sotirakopoulos, M. Velez-Rojas, and K. Beznosov, “Heuristics for evaluating it security management tools,” in *Proceedings of the Seventh Symposium on Usable Privacy and Security*, 2011, pp. 1–20.