Navigating the Human-Robot Interaction Landscape. Practical Guidelines for Privacy-Conscious Social Robots

Nicholas Callander
University of Glasgow
Glasgow, UK

Andrés A. Ramírez-Duque
University of Glasgow
Glasgow, UK

Mary Ellen Foster
University of Glasgow
Glasgow, UK

ABSTRACT

Social robots are a type of robotics that focuses on creating intelligent and embodied machines capable of interacting and communicating with humans in a socially acceptable manner. However, these robots' potential to capture user data, including emotions, biometrics, and behavioural habits, raises significant privacy concerns that could influence users' intention to use and trust social robots. Therefore, there is a pressing need to synthesize a privacy model that helps unravel the complex behavioural processes underlying current and future HRI technologies. This work aims to contribute to the growing body of privacy-friendly robot design by proposing comprehensive guidelines that enable the development of trustworthy and transparent social robots that respect user privacy. We have established a set of theoretical constructs to address people's concerns regarding privacy across four dimensions: physical, informational, psychological, and social. Finally, guidelines are provided in each construct to enhance transparency and trust through compliance with laws like GDPR.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; **HCI theory, concepts and models**; • **Computer systems organization → Robotic autonomy**; **Robotics**; • **Computing methodologies → Planning under uncertainty**; • **Security and privacy → Social aspects of security and privacy**.

KEYWORDS

Privacy, Trust, Transparency, Acceptability, Human-Robot Interaction

ACM Reference Format:
Nicholas Callander, Andrés A. Ramírez-Duque, and Mary Ellen Foster. 2024. Navigating the Human-Robot Interaction Landscape. Practical Guidelines for Privacy-Conscious Social Robots. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3610978.3640637

1 INTRODUCTION

Understanding the dynamics of interactions between users and technology and the reasons that motivate humans to interact with artificial agents has been a longstanding research field, with greater emphasis in the last ten years being placed upon human-robot interaction (HRI) especially. [2] From a human-centred perspective, the user's intentions when interacting with a robotics agent have been predominantly described through two frameworks. The first centred on pragmatic acceptance theories influenced by social factors [4], and the second centred on confidence-based models where there exists a high level of expectation of a positive outcome from an interaction with a robotics agent [11]. Within the HRI community, both perspectives have been established, widely adopted, and have influenced the development of new frameworks. While these perspectives have their merits, they are generally analysed in isolation and with the fast-moving trend towards embodied AI-powered conversational robots, there are mounting concerns regarding privacy and data protection. Therefore, there is an imminent need to synthesise a new perspective that ensures adequate consideration of privacy values while also helping unravel the complex psychological processes underlying current and future HRI technologies. This new perspective should foster greater acceptance and integration between humans and social robots. This work seeks to contribute to the growing body of work by proposing a set of comprehensive guidelines that aim to foster the development of social robots that are trustworthy, transparent, and respectful of user privacy (privacy-friendly). These guidelines will be developed through a literature review, drawing on insights from various technology-acceptance models and trust-based frameworks.

HRI '24 Companion, March 11–14, 2024, Boulder, CO, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0323-2/24/03.
https://doi.org/10.1145/3610978.3640637

This paper's primary contribution is to develop practical guidelines to help robotics designers create privacy-conscious social robots capable of earning users' trust. Furthermore, this research fills the previously neglected aspect of HRI, exploring the impact of embodiment on user privacy concerns and trust in social robots. The study promises valuable insights into user data privacy concerns and the complex and pivotal role trust plays in accepting social robots, all of which deepen our understanding of these concepts.

2 RELATED WORK: PRIVACY-CONSCIOUS ROBOTS

The rapid advancement in the field of social robotics has led to the development of social robots used in various domains, including healthcare [24], education [15], and homes [27]. Social robots are a branch of robotics that focuses on creating intelligent machines or robots that can interact and communicate with humans and other robots in a socially acceptable way [29]. These robots can perceive, understand, and respond to human emotions, behaviours, and social cues and learn to adapt to new situations [10]. However, their ability to capture user data, including emotions, biometrics, and behavioural habits, raises important privacy implications [7, 20]. Furthermore, social robots differ from our traditional provisions for technology due to their physical presence, which presents further

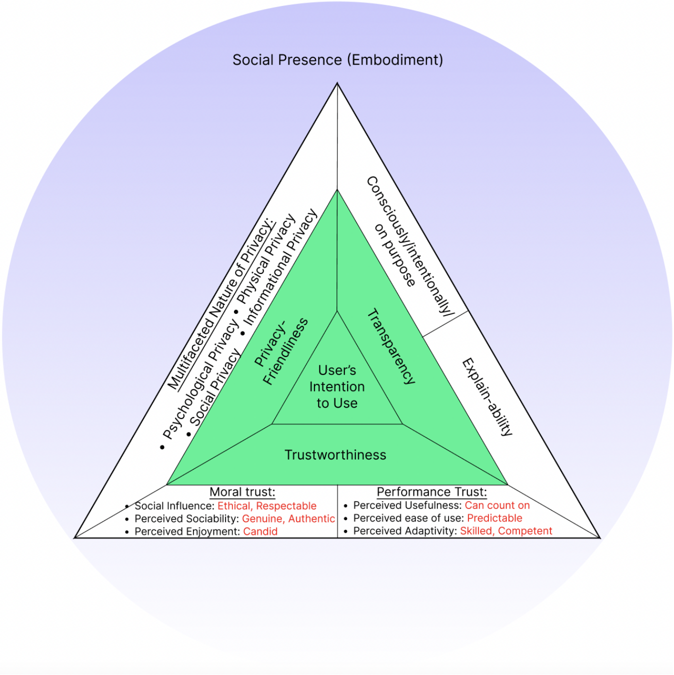
Social Presence (Embodiment)
Multifaceted Nature of Privacy:
• Psychological Privacy • Physical Privacy
• Social Privacy • Informational Privacy
Consciously/intentionally/ on purpose
Explain-ability
Privacy-Friendliness
Transparency
User's Intention to Use
Trustworthiness
Moral trust:
• Social Influence: Ethical, Respectable
• Perceived Sociability: Genuine, Authentic
• Perceived Enjoyment: Candid
Performance Trust:
• Perceived Usefulness: Can count on
• Perceived ease of use: Predictable
• Perceived Adaptivity: Skilled, Competent

Figure 1: Considered Factors in the Creation of Privacy-Conscious Social Robots

privacy concerns [18]. Trust plays a crucial role in HRI and can be established through factors such as reliability, predictability, transparency, and ethical behaviour [14]. Trust can be thought of as two distinct dimensions when interacting with an agent: performance and morale. Performance-based trust refers to an agent's reliability and capability, and moral-based trust refers to an agent's sincerity and integrity [16]. In the context of social robotics, trust refers to the confidence that a person has in a robot's ability to perform its intended functions and to behave in a manner that aligns with the person's expectations and values [16]. It is essential to address the ethical questions surrounding privacy, trust, and their respective impacts on HRI. While there has been extensive research on privacy and trust in the context of technology, much of it falls short when transferred to social robotics due to the defining factors of a social robot - embodiment and physical presence [7]. By understanding the impact of embodiment on privacy and trust, it becomes possible to develop ethical, responsible, and trustworthy social robots. The literature review involved a systematic search of various databases, including but not limited to IEEE Xplore, PubMed, and Google Scholar. Criteria used for inclusion or exclusion from the literature review had us focusing on studies that addressed human-robot interaction, privacy concerns, and trust in social robots - with these criteria also being used as keywords when searching for relevant literature.

3 TRUSTWORTHINESS, TRANSPARENCY AND PRIVACY-FRIENDLINESS: THEORETICAL CONSTRUCTS AND GUIDELINES

This work aims to enhance design guidelines for social robots by integrating the concept of privacy-conscious robotics and trust factors. It thoroughly explores how privacy considerations influence users' willingness to engage with social robots. The model, as seen in Figure 1, emphasises the consequences of three core concepts: "trustworthiness", "transparency", and "privacy-friendliness" on users' intention to use. These core concepts draw inspiration from established models such as E-UTUAT and MDMT [16] and insights from field experts such as Christoph Lutz. Trustworthiness, a central concept of our model, ensures that users can depend on social robots to behave appropriately and handle their personal data responsibly. Trust can be divided into moral and performance-based aspects, each of which can be aligned with specific facets of data protection and privacy concepts of our model. Constructs like "ethical" and "respectable" relate to moral trust, while "predictable", "skilled", and "competent" fall under performance-based trust. [16] Regarding privacy dimensions, we suggest that "physical" and "informational" privacy align with performance-based trust, as shortcomings in these areas are seen as under-performance. Conversely, when neglected, "psychological" and "social" privacy dimensions constitute a breach of moral-based trust. Regarding transparency, attributes like "explain-ability" and "conscientiousness" belong to the moral-based trust domain. Transparency refers to a social robot's ability to explain its processes and demonstrate a clear intention to fulfil its purpose. [3] Transparency and trust are closely intertwined, as understanding data collection and processing builds trust. Privacy-friendliness encompasses multiple dimensions of privacy, covering physical, informational, psychological, and social aspects. The above breakdown of privacy dimensions was initially framed by Leino-Kilpi et al. [17] and extended by [18]. These dimensions are distinct and categorise the different ways a social robot is capable of infringing upon a user's privacy. Designers should consider these dimensions when crafting interactions between social robots and users. Beyond these core concepts, the "social presence" concept in HRI is a pivotal factor. The tangible presence of a social robot within a user's environment has distinct effects compared to virtual avatars or chatbots. Addressing this embodiment aspect is crucial, as it significantly influences users' perceptions of a social robot's trustworthiness and their willingness to engage with it. Prioritising embodiment in design ensures the effective implementation of the core concepts, ultimately promoting wider acceptance and further development in social robotics. [5]

3.1 Physical Privacy (Personal Space)

Several strategies can be employed to enhance a social robot's trustworthiness regarding physical privacy. Firstly, user-controlled movement: users should be able to control when a robot moves between spaces for common use, such as a user's bedroom in their home. This control ensures the robot won't intrude at inopportune moments, such as when users are engaged in private activities like changing clothes. Secondly, limiting surveillance: social robots equipped with cameras should comfort users by letting them know they are not constantly monitored. Defaulting camera functions to "off" is a straightforward way to address this concern. More advanced techniques, like obfuscation, can also be employed to prevent capturing sensitive information. Thirdly, the adherence to Hall's proxemic zones [22]: These proxemic zones define personal space levels (intimate, personal, social, and public) and can enhance a robot's trustworthiness when used in the correct contexts. For example, a social robot in the receptionist role would not be suited to an intimate zone and would be better placed in a personal zone. These three strategies foster greater trust in HRI, ultimately enhancing the credibility of current and future social robots. Transparency can be achieved through a social robot clarifying and explaining its capabilities, intentions and purpose, as well as its interactive behaviours. [28] For example, medical robots can explain their hand sensors and pulse measurement process when touched by the user. This information gives users greater knowledge and comfort when interacting with social robots. It is vital, however, to avoid overwhelming the user with technical details. User-friendly training and reminders can be developed to offer straightforward explanations. This approach empowers users to make informed decisions, comprehend data collection and processing, and feel comfortable interacting with present and future social robots. Furthermore, a social robot must comply with UK GDPR's "Privacy by design and default" legislation to be classified as privacy-friendly. GDPR emphasises transparency in personal data handling, requiring designers to inform users clearly about data collection and usage, allowing users to monitor their data processing. Trust and empowerment can be built by granting users control over their data. Security is another critical component of GDPR guidelines, which can be incorporated by allowing the data controller to customise security features. For physical privacy, guidelines have already been discussed, such as defaulting to turning off cameras and entering specific areas only at user-designated times. Adhering to GDPR fosters user trust, enhances the social robot experience, and bolsters wider adoption. [25]

3.2 Psychological Privacy (Thoughts and Values)

The ability of social robots to learn and adapt to individual users' behaviour is crucial for gaining their trust and acceptance, particularly in the context of social therapy robots, where users may be emotionally vulnerable [16]. However, designers must consider trustworthiness and transparency to create a psychologically private robot. Users should be informed that dialogues or data will be recorded to personalise the robot's responses. Offering users control over their personal data (modify, delete, or update) enhances trust and fosters an open dialogue, particularly in therapy settings. Privacy concerns, such as disclosing information to non-whitelisted users, can be addressed by allowing the robot to depersonalise, i.e., revert to a default state. This instils user security and trust, ultimately boosting acceptance. Being conscious of psychological privacy in social robots enhances user experiences and makes them incredibly valuable tools for healthcare and personal assistance - widening their utility and, therefore, creating the opportunity for greater acceptance. To ensure a robot is privacy-friendly regarding psychological privacy, designers must carefully consider how sensitive data is stored. Storing conversations and interactions locally on the robot, rather than in the cloud, offers protection against network threats and unauthorised access. While cloud robotics offers benefits such as high performance and the ability for robot learning [1], it also poses privacy risks. Unauthorised access can lead to breaches, user surveillance, and even physical harm [26]. Striking a balance between local and cloud storage is crucial. A hybrid approach can store sensitive data locally and share non-sensitive data via cloud networks, enhancing robot capabilities while also safeguarding user privacy. Designers should carefully weigh the risks and benefits of storage methods to create capable and secure robots that respect user privacy and sensitive information.

3.3 Informational Privacy (Personal Information)

Articles 13 and 14 of the GDPR [6] are crucial for data controllers, especially in social robots. The provisions in these articles promote transparency and accountability, ensuring data subjects are provided with the necessary information to make informed decisions about their data. For social robots, it is essential for controllers to comply with Article 13. Controllers must provide clear, concise details on data processing purposes and duration, including the criteria used to determine the length of the storage period. Transparency on the data types processed and users' GDPR rights is equally crucial. This transparency builds trust and empowers users to make informed choices about their data. Educating users on data decisions is essential, not just informing them of their rights. This allows data controllers to foster greater user confidence that their data is being handled in compliance with GDPR. Adherence to Articles 13 and 14 can increase users' trust and acceptance, promoting responsible data handling in social robotics. Focusing on transparency and accountability allows data controllers to instil confidence and ensure responsible personal data management. For privacy-friendliness in this dimension, it is essential to follow GDPR guidelines [8] and Nissenbaum's "Contextual Integrity" theory [23] This involves collecting only necessary data aligned with user expectations, limiting sensors, using pseudonymisation, and ensuring transparency about data capture and storage methods, along with GDPR rights. Combining these principles and a contextual integrity framework creates effective, privacy-friendly social robots that respect users' informational privacy.

3.4 Social Privacy (Social Contacts and Influence)

Creating socially private, trustworthy, and transparent social robots requires addressing the privacy paradox, where a user's intention to protect their privacy conflicts with their online behaviour and is influenced by the privacy-utility trade-off and social factors [19]. To manage the privacy-utility trade-off, it is important to consider the context of the social robot deployment. Medical robots accessing health data may be deemed acceptable by users, while receptionist robots accessing the same data may seem intrusive to users. Trust is built by collecting only the necessary personal data for the context and using pseudonymisation when possible. Social influence also plays a vital role in robot acceptance. Users are more likely to use robots if they find them useful and receive positive recommendations from their social network [20]. Designers should, therefore, foster positive attitudes towards social robots and create a sense of community among users. By examining the context, instilling transparent data practices, and fostering positive social influence, it is possible to help users reconcile the privacy paradox and embrace the potential benefits of social robots. Ensuring privacy-friendliness here is intrinsically linked to the privacy-utility trade-off, as highlighted by GDPR Article 5(1)(c) [6] and Nissenbaum's "contextual integrity" theory [23]. Adhering to the principle of data minimisation and collecting only relevant data based on the specific deployment context can foster trust and acceptance in social robots. Furthermore, incorporating privacy-by-design principles [13] can help build users' trust and confidence.

3.5 Intention to Use

Intention to use is crucial in human-robot interaction, as it determines the user's motivation or reason to use a social robot [12]. As social robots have the potential to be deployed in various sectors [15, 21, 24], it is vital to align user needs with robot functionality to increase acceptance. Even a well-perceived robot may not be accepted if the robot doesn't land on the utility side of the privacy-utility trade-off. To ensure acceptance, understanding user intention is crucial. The model proposed by Giger et al. [9] provides a useful framework for identifying the determinants of behaviour influencing a user's decision to engage with a robot, such as attitudes, subjective norms, positive and negative anticipated emotions, and perceived behavioural control. Design can influence some of these determinants, e.g., a robot that provides a positive emotional experience and teaches users how to work with it can increase motivation and overcome perceived behavioural control issues. Through the privacy-friendly guidelines, the privacy-utility trade-off can be reconciled. Aligning the functionality of a social robot with user needs in different sectors can help drive acceptance. In summary, comprehending behaviour determinants and the privacy-utility trade-off is essential for effectively accepting social robot design.

4 CONCLUSIONS

This work proposes practical guidelines to assist robotics designers in creating privacy-conscious social robots capable of earning users' trust. Their ability to capture various types of user data creates the potential for significant privacy risks. Our model centres on three key concepts: trustworthiness, transparency, and privacy-friendliness. Trust is essential in social robotics and requires the robot to behave according to the user's expectations and values. Guidelines are provided in each concept to bolster trust through greater transparency and adherence to laws in GDPR. These guidelines should help to increase intention to use, especially given they are provided whilst considering the other most central concept - embodiment. The physical presence of a social robot means previous recommendations for fostering trust in technology will not always translate directly, thereby creating the demand for the proposed model. This work is a valuable addition to the growing body of literature on HRI. Given the increasing use of robots in various applications, the proposed model is timely and relevant. By following the provided guidelines from each stage of the model, designers and developers can create social robots that are transparent, trustworthy, and privacy-friendly, ultimately improving human-robot interactions and bolstering intention to use. This has the potential to facilitate the widespread integration of robots into various aspects of our daily lives, such as healthcare, education, and the home, among others - offering a promising direction for future research in this area.

REFERENCES

[1] Soumalya Bhattacharyya. 2022. What is Cloud Robotics? Importance and Challenges | Analytics Steps. https://www.analyticssteps.com/blogs/what-cloud-robotics-importance-and-challenges

[2] Andrea Bonarini. 2020. Communication in Human-Robot Interaction. *Current Robotics Reports* 1, 4 (Dec. 2020), 279–285. https://doi.org/10.1007/s43154-020-00026-1

[3] Filippo Cantucci and Rino Falcone. 2020. Towards trustworthiness and transparency in social human-robot interaction. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*. 1–6. https://doi.org/10.1109/ICHMS49158.2020.9209397

[4] Maartje M. A. de Graaf and Somaya Ben Allouch. 2013. Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems* 61, 12 (Dec. 2013), 1476–1486. https://doi.org/10.1016/j.robot.2013.07.007

[5] Eric Deng, Bilge Mutlu, and Maja Mataric. 2019. Embodiment in Socially Interactive Robots. *Foundations and Trends in Robotics* 7, 4 (2019), 251–356. https://doi.org/10.1561/2300000056 arXiv:1912.00312 [cs].

[6] European Parliament and Council of the European Union. 2016. General Data Protection Regulation (GDPR) – Official Legal Text. https://gdpr-info.eu/

[7] Eduard Fosch Villaronga, Heike Felzmann, Robin Pierce, Silvia Conca, Aviva Groot, Aida Ponce Del Castillo, and Scott Robbins. 2018. *Nothing Comes between My Robot and Me: Privacy and Human-Robot Interaction in Robotised Healthcare.* https://doi.org/10.5040/9781509926237.ch-004

[8] GDPR. 2016. Recital 78 - Appropriate Technical and Organisational Measures. https://gdpr-info.eu/recitals/no-78/

[9] Jean-Christophe Giger, Daniel Moura, Nuno Almeida, and Nuno Piçarra. 2017. Attitudes towards Social Robots: The Role of Belief in Human Nature Uniqueness, Religiousness and Taste for Science Fiction.

[10] Frank Hegel, Claudia Muhl, Britta Wrede, and Gerhard Sagerer. 2009. Understanding social robots | IEEE conference publication - IEEE xplore. https://ieeexplore.ieee.org/document/4782510/

[11] Aike C. Horstmann and Nicole C. Krämer. 2020. Expectations vs. actual behavior of a social robot: An experimental investigation of the effects of a social robot's interaction skill level and its expected future role on people's evaluations. *PLoS ONE* 15, 8 (Aug. 2020), e0238133. https://doi.org/10.1371/journal.pone.0238133

[12] H. Huang and J. Rust. 2018. Understanding The Intention To Work With Social Robots | Science Trends. https://sciencetrends.com/understanding-the-intention-to-work-with-social-robots/

[13] Information Commissioner's Office. 2022. Data protection by design and default. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-by-design-and-default/

[14] Bing Cai Kok and Harold Soh. 2020. Trust in Robots: Challenges and Opportunities. *Current Robotics Reports* 1, 4 (Dec 2020), 297–309. https://doi.org/10.1007/s43154-020-00029-y

[15] Elly Konijn, Matthijs Smakman, and Rianne van den Berghe. 2020. Use of Robots in Education. (Sep 2020), 1–8. https://doi.org/10.1002/9781119011071.iemp0318

[16] Allison Langer, Ronit Feingold-Polak, Oliver Mueller, Philipp Kellmeyer, and Shelly Levy-Tzedek. 2019. Trust in socially assistive robots: Considerations for use in rehabilitation. *Neuroscience and Biobehavioral Reviews* 104 (2019), 231–239. https://doi.org/10.1016/j.neubiorev.2019.07.014

[17] H. Leino-Kilpi, M. Välimäki, T. Dassen, M. Gasull, C. Lemonidou, A. Scott, and M. Arndt. 2001. Privacy: a review of the literature. *International Journal of Nursing Studies* 38, 6 (2001), 663–671. https://doi.org/10.1016/S0020-7489(00)00111-5

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

[19] Christoph Lutz and Aurelia Tamò Larrieux. 2020. The Robot Privacy Paradox: Understanding How Privacy Concerns Shape Intentions to Use Social Robots. *Human-Machine Communication* 1 (Feb 2020), 87–111. https://doi.org/10.30658/hmc.1.6

[20] Christoph Lutz and Aurelia Tamò-Larrieux. 2021. Do Privacy Concerns About Social Robots Affect Use Intentions? Evidence From an Experimental Vignette Study. *Frontiers in Robotics and AI* 8 (2021). https://www.frontiersin.org/articles/10.3389/frobt.2021.627958

[21] Lohse Manja, Frank Hegel, and Britta Wrede. 2008. Domestic Applications for Social Robots - an online survey on the influence of appearance and capabilities. *Journal of Physical Agents* 2 (Jan 2008). https://doi.org/10.14198/JoPha.2008.2.2.04

[22] Nicolai Marquardt and Saul Greenberg. 2012. Informing the Design of Proxemic Interactions. *IEEE Pervasive Computing* 11 (Feb 2012), 14–23. https://doi.org/10.1109/MPRV.2012.15

[23] Helen Nissenbaum. 2004. Privacy as Contextual Integrity. *Washington Law Review* 79, 1 (Feb 2004), 119.

[24] Research Outreach. 2020. Social Robots – a New Perspective in Healthcare. *Research Outreach* (Jun 2020). https://researchoutreach.org/articles/social-robots-new-perspective-healthcare/

[25] Hossein Rahnama and Alex "Sandy" Pentland. 2022. The New Rules of Data Privacy. *Harvard Business Review* (Feb. 2022). https://hbr.org/2022/02/the-new-rules-of-data-privacy

[26] Matthew Rueben, Alexander Mois Aroyo, Christoph Lutz, Johannes Schmölz, Pieter Van Cleynenbreugel, Andrea Corti, Siddharth Agrawal, and William D. Smart. 2018. Themes and Research Directions in Privacy-Sensitive Robotics. In *2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*. 77–84. https://doi.org/10.1109/ARSO.2018.8625758

[27] Thomas B Sheridan. 2020. A review of recent research in social robotics. *Current Opinion in Psychology* 36 (Dec 2020), 7–12. https://doi.org/10.1016/j.copsyc.2020.01.003

[28] Robert H. Wortham. 2020. *Transparency for Robots and Autonomous Systems: Fundamentals, technologies and applications*. Institution of Engineering and Technology, London. https://doi.org/10.1049/PBCE130E

[29] Karim Youssef, Sherif Said, Samer Alkork, and Taha Beyrouthy. 2022. A Survey on Recent Advances in Social Robotics. *Robotics* 11, 44 (Aug 2022), 75. https://doi.org/10.3390/robotics11040075