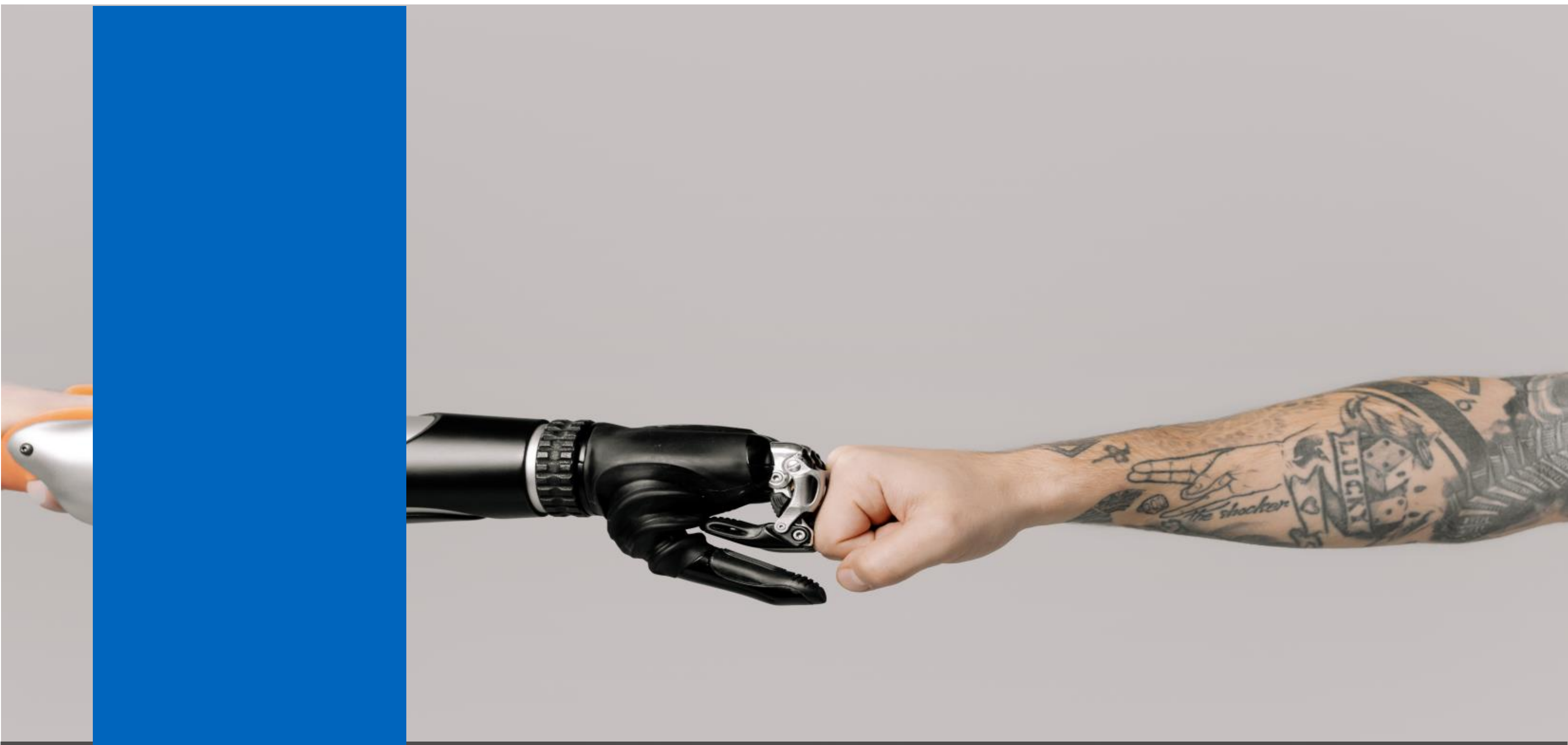
Technical University of Munich
Munich Center for Technology in Society
Institute for Ethics in Artificial Intelligence

Research Brief – April 2021



A Robotic New Hope: Opportunities, Challenges, and Ethical Considerations of Social Robots

by Auxane Boch, Laura Lucaj, Caitlin Corrigan

Robotics technologies are now increasingly used in a multitude of ways related to social interaction with humans. Considering their expanding employment in many areas of everyday life, it is now crucial to understand the implications of robotic technology for users and society. In this Brief, we introduce the latest in social robot technology designed for human-centered services and offer an assessment of the ethical considerations needed to accompany their design and governance.

In 2020, 3.13 million social and entertainment robots were sold worldwide (Statista & Vailshery, 2021). Indeed, robots are no longer confined to the manufacturing world. Robotics technologies are now increasingly used in a multitude of ways related to social interaction with humans, from teaching children, to protecting family homes, distributing medications in hospitals or serving as personal companions (Stroessner, 2020). Considering their increasing employment in many areas of everyday life, it is now crucial to understand the implications of robotic technology for users and society. In this Brief, we introduce the latest in social robot technology designed for human services and offer an assessment of the ethical considerations that need to accompany their design and governance.

To understand the implications of such technology on society and individual users, a clear definition is needed. Due to the variety and complexity of tasks and behaviors robots with social intentions have to perform, there is no consensus on how to define these types of machines. Nevertheless, some commonalities exist. In this brief, we will consider the definition of social robots presented by Fox and Gambino (2021) as human-made AI-powered technology, in a digital or physical form, presenting human-like attributes to some degrees. More importantly, they are designed to communicate with people. Fong et al. (2003) more precisely focus on social robots' ability to perceive and express emotions through high-level dialogues and recognition of other agents, with the objective of establishing and maintaining social relationships. Hence, social robotics focuses on robots with social qualities used in the context of service to humans, and is an interdisciplinary research area investigating Human-Robots Interactions (HRI) from diverse angles, including computer science, psychology, engineering and sociology.

Social robots are AI-enabled tools (Bigman et al., 2019; Fox and Gambino, 2021), which are embedded with different technologies such as natural language processing (NLP) and face recognition algorithms (Thilmany, 2007). These different abilities present opportunities for humans, but also pose unprecedented risks. However, the exact nature of these risks still lacks depth and breadth of research. This Brief is an attempt to gather the latest findings on the ethical risks social robots represent and to analyze the ethical challenges related to specific form of technology in order to provide insights for developers and enable policymakers to improve social robot governance.

What make social robots different?

At the core of the characteristics social robots display is the commonality of interaction. The first feature allowing for such capabilities implemented in an AI-enabled system was conversation. The presentation of a pioneering conversational AI, ELIZA, opened the discussion about the illusion of reasoning skills displayed by programs (Weizenbaum, 1984). Indeed, users of ELIZA perceived the machine as understanding and being involved in their conversations, even though it was not (Searle, 1982). This tendency is called the Eliza Effect (Hofstadter, 1979). The Eliza Effect was the first step towards recognizing a natural and automatic human bias to project life-like qualities onto non-living objects, a bias called anthropomorphism (Epley, 2018).[1] This human tendency has the potential to create significant impacts when deploying social robots in society (Darling, 2016).

Emerging digital voice assistants, digital representation as avatars and physically embodied robots appear to be more alive, more human than their predecessors were. Their ability to show or execute movements can lead to the perception of intelligence and "mind in the machine", taking the Eliza Effect to a new level (MacInnis & Folkes, 2017). The ability to replicate physical behaviors, when combined with variable tones of voice and precise speech content, can be understood as a demonstration of emotions by

[1] When facing animal-like social robots, Zoomorphism occurs, attributing animal qualities to an object (Henschel et al., 2021).

users (Bates, 1994). Conversely, if the tangible or digital machine appears almost, but not quite, human, it can cause discomfort to the user.[2] Interestingly, this feeling can be reduced if the representation is imperfect or makes mistakes (Mirnig et al., 2017). These imperfections ease the user's negative impressions that come from a sense of threat from the machine (Duffy, 2003).

Social Robots – Categories and Benefits

In this section, we provide an overview of four different main categories of social robots with increasing interactional intensity with humans, along with prominent examples of the contexts in which they are being commercially employed. This categorization enables us to analyze and discuss technological opportunities, as well as the potential risks emerging as human-robot interaction levels increase.

Service Robots

Social robots are sometimes considered as a form of service robots with social intentions (Fosch-Villaronga, 2019). We, however, discuss service robots as social robots working in the service industry. For example, in 2016, Pinillos et al. (2016) studied the reaction to the implementation of the Sacarino Robot, employed as a bellboy in a hotel. Its main features were to provide information about the hotel and the city to the guests.

While such robots still need to be improved in order to be fully accepted by the public in service settings, they can provide consistent and low costs for organizations to interact with clients for minimalist tasks (Karar et al., 2019).

Sacarino Robot

Care Robots

Care robots are defined by Vallor (2011) as machines created to assist or replace human caregivers when practicing care for vulnerable populations. One example includes the (NAO) Next Generation Robot, a little humanoid machine used to interact with children diagnosed with Autism Spectrum Disorders (Shamsuddin et al., 2012). A second practical example is Paro, a white seal stuffed animal robot, used in Japan in medical care settings. Because of its lifelike responsiveness, dementia patients treat this robot it as if it was a real animal.

These care robots can have a positive impact on their overall well-being (Calo et al., 2011), significantly contributing to improving the life of vulnerable populations having trouble bridging the communication gap or bettering their living environment where typical animal support therapy is not possible (Darling, 2015).

NAO Robot

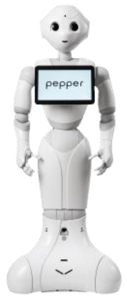
Educational Robots

Educational Robots are developed to improve people's learning experience (Angel-Fernandez & Vincze, 2018). One example includes the robot Pepper, who acts as a tutor in teaching waste sorting for recycling purposes in-home to children.

The use of social robots in education has shown to result in higher attention and engagement rates and, thus, learning gains for children (Castellani et al., 2021). This multiple level enhancement might be a consequence of the tendency children have to connect with such machines in comparable

[2] This effect has been called the Uncanny Valley (Wang et al., 2015).

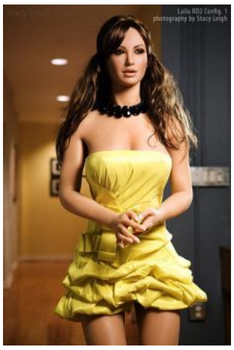
ways to friend-like relationships (Vollmer et al., 2018).



Educational Robot Pepper

Sex Robots

Sex robots perform actions aimed at improving users' sexual satisfaction (Fosch-Villaronga & Poulsen, 2020). The sex robots available right now are not highly sophisticated. They mainly consist of silicon dolls with embedded AI-enabled systems to make them interactive on a new level, such as the RealDolls manufactured by Abyss Creations (Clark-Flory, 2020). Nonetheless, they cannot move on their own and, therefore, the technology is still far from developing fully functional androids (Kubes, 2019).



Lalia RDZ – RealDoll

Sex robots are still controversial and their impacts still unclear (Gutiu, 2016). Some argue that they have the potential to improve certain conditions, such as meeting the sexual needs of those with disabilities and the elderly. This could reduce loneliness and insecurity or even sexual crimes (Devlin, 2015; Fiske et al., 2019; Levy, 2009). However, they also could pose an extreme risk in terms of creating dependencies and obscuring human objectification (Richardson, 2016; de Graaf et al., 2016).

Considering Risks in the Use of Social Robots

With new technologies come possible benefits for society and individuals, but also risks. Fosch-Villaronga et al. (2019) presents two main areas of concerns for the future of social robotics: the uncertainty and the responsibility. In other words, clarifications need to be made concerning the regulations surrounding social robots to define precise limits and accountability.

One main concern or risk related to social robots has to do with the **creation of dependencies** (de Graaf et al., 2016). A main features and goal of such technology is to build, develop and maintain relationships throughout time with people (Fong et al., 2003). Through the social and emotional cues they can express, either as physical behaviors or spoken communication, attachment might develop towards such machines (Darling, 2015; Corretjer et al., 2020). The more a robot is perceived as autonomous and emotional, the more the attachment grows (Turkle, 2010, Scheutz, 2011; Darling, 2016). Those feelings towards the automated entity might lead to the creation of a structured, real and evolving-relationship similar to a unidirectional one, defined in other areas of science as para-social relationships (Schiappa et al., 2007; Perse & Rubin, 1989). Such relationships, experienced as genuine by the individuals, raise opportunities and possible risks to human users.

This relates to the important notion of emotional trust (Glickson & Woolley, 2020), which is irrational and grows through affection. This sense of trust may be impacted through the development of anthropomorphic and "emotional" robots and poses a risk when issues of responsibility arise (Fosch-Villaronga et al., 2019). This can be especially concerning with vulnerable or sensitive populations, such as in the case of care robots for dementia or autistic patients (Shamsuddin et al., 2012; Calo et al., 2011) or educational robots working with children (Angel-Fernandez & Vincze, 2018).

When used in everyday life or as a sexual companion, problems also originate in terms of **transparency and data privacy** (Fosch-Villaronga et al., 2019). People may not be aware of how much data the robot is recording, where such data is uploaded and how it will be processed. Ethical questions related to knowledge and deception must also be addressed (Arkin, 2018). For instance, do people realize this technology is not actually feeling anything for them? Do they realize the lack of reciprocity in the interaction?

The gains in productivity and reductions in costs when employing social robots instead of humans might be one of the biggest threats

A related matter is the possible **impact of human-robots relationship on human-to-human relationships** (Darling, 2016; Fosch-Villaronga et al., 2019). This could include possible widespread dehumanization of human communication with each other, or reduction in empathy towards others. Thus, spillover impacts on social behaviors and skills is a concern.

Another prevalent ethical concern relates to the argument often seen surrounding technology and the future of work. The gains in productivity and reduction in costs when employing social robots instead of humans, might be one of the biggest threats (Stanford Encyclopedia of Philosophy, 2020). As seen with the case of service robots, a concern is the possible **replacement of humans**. Care or educational robots may also replace traditional tutors or caregivers (Fosch-Villaronga et al., 2019). While this does not necessarily imply the loss of job positions, transformations of such positions will be necessary. The question is: will the creation of new work positions and the elimination of others occur simultaneously, or will there be a time-gap creating unemployment that needs to be anticipated?

Conceptualizing Social Robots Risks

Different social robots present different types and levels of risk to users and society, depending on their tasks and autonomy. Concepts from AI risk assessment and governance may be useful in thinking about how to deal with these risks in terms of policies. Figure 1 displays the pyramid and risk-adapted regulatory system for the use of AI-based machine translation systems (DKE, 2020). Like different AI-based technologies, social robots could be classified, and thus governed, on different levels depending on the likelihood and severity of risk.

For example, Sacarino the Bellboy Robot (Pinillos et al., 2016) may represent a low level of risk, as its actions are limited to simple defined tasks for non-critical purposes and minimal invasiveness. On the other end of the spectrum, mental support care robots for autistic children (Shamsuddin et al., 2012) may be classified as higher risk, as they deal with private medical information, a vulnerable population and may have long-lasting and significant impact through the individuals interaction with the robot. In this case, more assessment and governance of these type of tools may be needed.

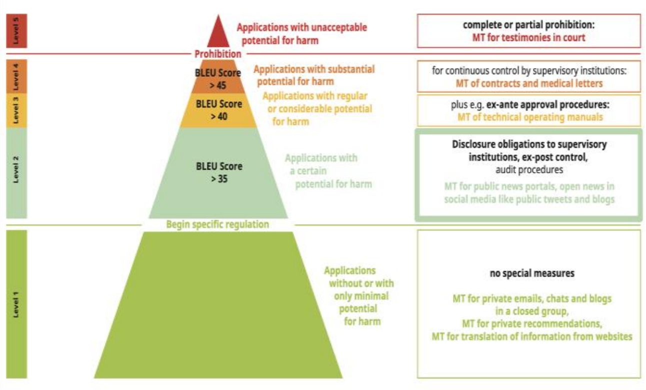


Figure 1. The critical pyramid and a risk-adapted regulatory system for the use of AI-based machine translation systems (Gutachten der Datenethikkommission) as presented in the German Standardisation Roadmap on Artificial Intelligence (DKE, 2020).

Ethical Considerations in Developing Social Robots

We have highlighted some benefits and risks related to social robots, displaying the presence of trade-offs in terms of the impact of their use. Numerous questions have still not been answered, concerning the possible impact of implementing social robotics as an active part of society. Thus, the analysis of these trade-offs will be a continuing one.

However, we can begin to analyze the ethical implications of such technology regarding how, to what extent and in which context positive or negative impacts will develop. Floridi et al. (2018) ethical principles for a good AI society provide a useful starting point for this[3]. The decision to use this framework as the theoretical lens through which we analyze this complex technology is motivated by the belief that social robots amplify the issues that characterizes the field of AI ethics. This includes, above all, the level and form of interaction with humans in their everyday life. In applying this framework to the evaluation of social robotics, we are able to systematically outline considerations in a way that is consistent and comparable with other emerging complex and intelligence technologies.

Beneficence can be described as promoting well-being, preserving dignity and sustaining the planet. The creation of care robots such as Paro (Calo et al., 2011) and the NAO (Shamsuddin et al., 2012) to support vulnerable populations and help them have a better life, displays a small part of the benefit robots can bring to society. Pepper (Castellano et al., 2021), an example of an educational robot promoting behaviors for sustainability, serves to increase beneficence not only in access to education, but also in the message it delivers on how to take better care of the planet. Social robots have the ability to help reduce loneliness of aging populations (Booth, 2020), support nursing staff (Vallor, 2011) and improve efficiency in everyday life tasks (such as Amazon's Alexa (2021) and Apple's Siri (2021)). These potential benefits must be considered when debating the trade-offs of employing robots in specific contexts.

The principle of **Non-Maleficence** highlights the importance of mitigating risks in order to prevent harm to humans/society. Since social robots have the overarching goal to support and be of service to their users, avoiding physical or mental harm to people is imperative in fulfilling their use. As presented above, risks relating transparency and data privacy, the creation of dependency, the impact on human-to-human interaction and the possible replacement of humans in the work environment are all of concern in terms of producing overall harm to users.

Regarding data privacy issues, the importance of cybersecurity and data protection in relation to the use of these tools is key to ethical implementation. While the EU's General Data Protection Regulation (GDPR; European Commission, 2018), for instance, offers a few layers of protection, this regulation has its limits. The novelty of the social machines is their capacity of sensing, processing and recording the entire environment surrounding them, as well as registering people's preferences and daily routines (Denning et al., 2009; Fosch-Villaronga et al., 2019). However, this requires thinking about new risks. Hence, there is a need to define how data will be collected and processed from social robots, to what extent this data should be stored or uploaded to the cloud, how to inform and get the users' enlightened consent to do so and how to prevent unauthorized external actors from accessing personal detailed information. Due to new technological features, allowing for robots to collect more data in their environment than before, the creation of particular regulations need to be considered. A possibility would be to extend the GDPR, considering the machine learning algorithmic qualities required in social robots, while looking out for the users' privacy (Stanford Encyclopedia of Philosophy, 2020).

Due to new technological features, allowing for robots to collect more data in their environment than before, the creation of particular regulations need to be considered

In terms of the harms to human-to-human interactions, we need to consider the impact on humans' empathic abilities. While empathy is an innate capacity, empathic responses do not automatically seem to occur (Decety, 2015). Indeed, these reactions seem to be skills partially acquired through both interpersonal and contextual experiences (Tousignant et al., 2016). Darling (2016) suggests that interactions with social robots could affect an individual's ability to develop general empathy due to the robot's lack of realist emotional answers, or the

[3] These principles include beneficence, non-maleficence, autonomy, justice and explicability.

dehumanization of relationships (Fosch-Villaronga et al., 2019).

This also raises the question of abusive behaviors towards robots themselves. Such behavior could lead to trauma in bystanders, as the same neural structures seem to be activated when confronted with violent behavior towards both robots and humans (Rosenthal-von der Pütten et al., 2013), and/or more general desensitization to these immoral behaviors (Darling, 2016). Darling (2016) proposes to regulate authorized behaviors towards robots in order to prevent such an outcome. However, more research is still needed on the exact nature and intensity of these impacts.

Autonomy indicates the individual's right to make decisions about their own life. When interacting with social robots, trade-offs may emerge concerning the willingness of delegating decision-making tasks to AI-systems and efficiency of use. A balance needs to be achieved to protect the intrinsic value of human choice and allow for comfortable human-robot collaborations. The user should always have the choice as of which decisions will be delegated to the AI-powered system and which one will not. Moreover, the human should always have the possibility of reversing, not implementing, stopping and starting all decisions made by the system.

There is a risk of regional disparities and inequities in terms of who benefits from social robots

Another autonomy issue that might arise in the future is monetary costs of upgrading systems. As stated earlier, the trust and attachments humans might develop towards robots could make users, to a certain level, dependent on the technology or at the very least unwilling to separate from the robot (Darling, 2016). Those motivations could be misused by companies by, for instance, proposing mandatory and costly upgrades of the product in order for it to continue to work. This would narrow the choices of the users to keeping the robot and paying, or abandoning the relationship. These potential abuses should be regulated upstream.

Considerations related to autonomy also include positive possibilities related to the extremely repetitive tasks to be lifted from the usual workload of service employees to service robots. The same could be said for care robots aiming to support care staff in their sometimes-trivial tasks, allowing them to choose how to spend their time with their patients and giving them some autonomy back in their work. However, a balance needs to be found to maintain a sense of autonomy and equity in the workplace, while utilizing social robots in ways that do not make workers worse off.

The concept of **Justice** is about ensuring that the benefits of social robots entering society are shared equally and avoiding discrimination based on factors such as gender or ethnicity. One issue is related to equal access. Social robots are entering societies at different rates. According to Mordor Intelligence's study and predictions (2021), the fastest growing markets for social robots between 2021 and 2026 will be the USA, Oceania and East Asia, with Europe and the rest of North America having a medium growth rate. South America, the Middle East, and Africa are missing from this equation, all being predicted to have a low rate of growth in the sector. Therefore, there is a risk of regional disparities and inequities in terms of who benefits from social robots and who is left out of the discussions concerning the creation, design, and implementation of social robots. Such considerations are important regarding the divide between Global North and Global South, already characterizing the development of most AI-powered technologies in the market today (Arun, 2019). This runs the risk of entrenching existing inequalities in technology use and development on a global level, as the cost of designing and developing these tools might affect the unequal distribution of accessibility around the world. Therefore, developers and policymakers would do well to promote the involvement of actors from countries outside of the traditional technology powerhouse regions.

Depending on the AI technology embedded in social robots, other justice related issues may emerge. For instance, the use of face recognition algorithms which enables a higher interaction level between the machine and humans, has bias-related concerns. Vulnerable minorities, already facing the structural bias present in society, are disproportionally affected by error rates in how these systems function (Boulamwini & Gebru, 2018). Thus, certain populations may not be able to equally access the benefits of social robots enabled with face recognition technology. Additionally, the development of sex robots is significantly impacting justice from a gender equality perspective, as the systems currently being developed have concerning features that might encourage the objectification of women (Richardson, 2016).

Explicability refers to the need to understand how systems are operating around us and on which levels they affect our everyday life. It is essential, for designers and developers of social robots to understand which ethical dimensions are implicitly embedded into these systems. Hence, users should be informed in a transparent way about the benefits and harms that might emerge from the interactions with social robots, as well as who is accountable when issues arise. Moreover, the consequences of a para-social relationship between users and robots are, at this time, quite unknown. This stresses the need for further empirical research and accountability mechanisms before the presence of social robots becomes widespread in the personal context of peoples' lives.

Increasing user knowledge about the limitations of social robots is also key to improving explicability. Some argue that true friendship can grow between humans and robots (Danaher, 2019), while others evoke the problem of deception involved in such a liaison (Nyholm & Frank, 2017). Indeed, robots do not (yet) have the possibility to feel. The misleading qualities of social robots might lead to a para-social relationship, which, as presented earlier, comes with positive and negative consequences. Loyalty and trust given to such a tool, based on false appearances, can lead the users to divulge more of their life and personal data than they would in a "normal" setting (Reig et al., 2021). Because of this, explainability about how social robots actually function and use data is needed in order for the users to understand the implications of such a relationship, and to be aware of the lack of genuine feelings the robots can develop.

Final Thoughts

While we have considered some of the risks that robots might pose in terms of human-machine interaction, robots also have received a sometimes-unwarranted bad reputation in pop culture and the media. Arguments about robots as a threat to society grow, among others, from the fear of humans being replaced by machines, and more extremely the belief that AI will overpower human intelligence. While risks exist and need to be identified, impeding the adoption of AI-powered technologies that have the potential to improve human life due to misinformation and misunderstandings also represents an ethical risk. Science-fiction contributes to unrealistic expectations of what robots are and will be in real-life contexts (Gibson, 2019). Overall, the state-of-the-art robotic technology is being developed to enhance, rather than replace, human capabilities (Fox, 2018). Researching the implications of introducing novel machines in societies, and educating the people about empirically founded knowledge on the topic, are the obvious steps in reducing unfounded bias.

In this Brief, we displayed ethical concerns and opportunities that might emerge in the field of humans-robots interactions through a defined framework. But trade-offs between different forms of social robot use certainly exist. Below we describe just a few examples.

For one, social robots currently available on the market are mainly being trained to target the needs of the individuals interacting with them, achieved through the processing of data on the individual's preferences that are being collected. However, without clearly defining the extent to which these systems should collect, elaborate and store personal data, the trade-offs between improving efficacy and a lack of data privacy will not be solved.

When it comes to elderly care, trade-offs might emerge as the efficacy or beneficence of these systems prove to be contributing to people's well-being by increasing the autonomy of the machines (and decreasing the autonomy of users). A more autonomous social robot has to be balanced with the user's conflicting rights (Vanderelst & Willems, 2019). As cultural environments change, the extent to which people feel comfortable delegating certain tasks to robots might also vary. Hence, different societal contexts might need a different set of values embedded into their robotic systems.

Another trade-off may be particularly clear in the contexts of robots' interaction with children, as this precise population seems to form specific attachments towards machines that may disproportionally affect them (Di Dio et al., 2020). On the one hand, anthropomorphic design can promote beneficence, as said design can significantly encourage children to engage more in educational topics. On the other hand, human-like designs might lead children to treat living things in the same way they treat a robot, which is problematic in the contexts of undesirable behaviors, as it could lead to the translation of harmful behavior in real-life interactive contexts (Darling, 2015) (violating the principle of non-maleficence). A solution for such potential unintended consequences would be the regulation of interactions depending on the publics' age, similarly to what is done in the digital game industry (Parisod et al., 2014).

Regarding sex robots, the trade-offs are still largely unknown, due to the lack of evidence-based research. The benefits of this technology in helping the treatment of sexual disorders may not balance the risk of desensitizing people and promoting unwanted spillover effects to human interaction or objectification. Without validation, achieved through randomized control trials, applying sex robots in therapeutic context might lead to worsening issues such as sexual violence (Fiske et al., 2019).

AI-powered robotic systems with social intentions are being designed to contribute to the well-being of people. However, such systems amplify some of the most complex issues characterizing the research field of AI ethics (Fong et al., 2003). This raises questions concerning the extent to which it is safe for humans to interact with social robots, and when the potential risks might outweigh the benefits. So far, the unrealistic expectation regarding the evolution of robotic technology, and the fear to be replaced by such systems, is overshadowing more pressing issues. These include the (1) contexts and the extent in which these tools serve the well-being of humans (around the globe), (2) the responsible parties regarding the mitigation of the risks that might emerge from human-robot interaction, (3) the appropriateness of tasks that are to be delegated to robots, among others.

In order to organize a steady governance framework for the development and dissemination of social robots, the first step would be to define precisely what social robots are and to identify the limits of their allowed utilization. This should also be done while incorporating dimensions such as culture and diverse social habits. A few options have been proposed. In Japan, the TOKKU approach builds on evidence-based policy making where research laboratories and policy makers work together to develop regulations based on empirical data (Weng et al., 2015). A second step would be to design rules of development and implementation that are sector specific, as identical rules may not apply to healthcare robots vs. educational robots vs. military robots (Fosch-Villaronga et al., 2019).

Anthropomorphic designs might lead children to treat living things in the same way they treat a robot, which is problematic

Considering the increasing presence of social robots in our everyday life, the analysis of the ethical consequences emerging from the interaction of humans with these machines is timely and relevant. Our investigation of these critical questions through an AI ethics framework enables us to understand the steps that need to be taken in order to capitalize on the opportunities provided by these machines in certain social contexts, whilst mitigating the unintended consequences that might emerge in other contexts. Nonetheless, further interdisciplinary, longitudinal and evidence-based research in the Human-Robot Interaction field are essential to address the complexity of the issues at hand and the potential opportunities emerging from the introduction of robots in societies.

References

Amazon. (2021). *Amazon Alexa* [Computer software]. https://developer.amazon.com/de-DE/alexa

Angel-Fernandez, J. M., & Vincze, M. (2018). Towards a definition of educational robotics. In *Austrian Robotics Workshop 2018* (Vol. 37).

Apple. (2021). *Siri* [Computer software]. https://www.apple.com/siri/

Arkin, R. C. (2018, October 9). *Ethics of Robotic Deception*. IEEE Technology and Society. https://technologyandsociety.org/ethics-of-robotic-deception/

Arun, C. (2019). AI and the Global South: Designing for Other Worlds.

Bartneck, C, Lütge, C., Wagner, A., Welsh (2019), S. *An Introduction to Ethics in Robotics and AI*

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM*, *37*(7), 122-125.

Bigman, Y. E., Waytz, A., Alterovitz, R., & Gray, K. (2019). Holding robots responsible: The elements of machine morality. *Trends in cognitive sciences*, *23*(5), 365-368.

Booth, R. (2020, September 8). *Robots to be used in UK care homes to help reduce loneliness*. The Guardian. https://www.theguardian.com/society/2020/sep/07/robots-used-uk-care-homes-help-reduce-loneliness

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. 77-91. PMLR.

Bruno, B., Chong, N. Y., Kamide, H., Kanoria, S., Lee, J., Lim, Y., & Sgorbissa, A. (2017, June). The CARESSES EU-Japan project: making assistive robots culturally competent. In Italian Forum of Ambient Assisted Living (p. 151-169). Springer, Cham.

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency. 77-91. PMLR.

Calo, C. J., Hunt-Bull, N., Lewis, L., & Metzler, T. (2011, August). Ethical implications of using the paro robot, with a focus on dementia patient care. In Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence.

Castellano, G., De Carolis, B., D'Errico, F., Macchiarulo, N., & Rossano, V. (2021). PeppeRecycle: Improving Children's Attitude Toward Recycling by Playing with a Social Robot. *International Journal of Social Robotics*, *13*(1), 97-111.

Clark-Flory, t. (2020). What I leaned about male desire in a sex doll factory. *The Guardian.* Retrieved online from: https://www.theguardian.com/lifeandstyle/2020/oct/19/what-i-learned-about-male-desire-in-a-sex-doll-factory

Collins, K., & Dockwray, R. (2018). Tamaglitchi: A Pilot Study of Anthropomorphism and Non-Verbal Sound. In *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion*, 1-6.

Danaher, J. (2019) "The Philosophical Case for Robot Friendship", *Journal of Posthuman Studies*, 3(1): 5–24. doi:10.5325/jpoststud.3.1.0005

Darling, K. (2015). 'Who's Johnny?' Anthropomorphic Framing in Human-Robot Interaction, Integration, and Policy. *ROBOT ETHICS*, (Vol. 2).

Darling, K. (2016). Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In *Robot law*. Edward Elgar Publishing

Decety, J. (2015). The neural pathways, development and functions of empathy. *Current Opinion in Behavioral Sciences*, *3*, 1-6.

Denning, T., Matuszek, C., Koscher, K., Smith, J. R., & Kohno, T. (2009, September). A spotlight on security and privacy risks with future household robots: attacks and lessons. In *Proceedings of the 11th international conference on Ubiquitous computing*, 105-114.

Devlin, K. (2015). In defence of sex machines: why trying to ban sex robots is wrong. *The conversation*.Di Dio, C., Manzi, F., Peretti, G., Cangelosi, A., Harris, P. L., Massaro, D., & Marchetti, A. (2020). Shall I trust you? From child–robot interaction to trusting relationships. *Frontiers in psychology*, *11*, 469.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and autonomous systems*, *42*(3-4), 177-190.

European Commission (2018, May). *General Data Protection Regulation*. https://gdpr-info.eu/

Epley, N. (2018). A mind like mine: the exceptionally ordinary underpinnings of anthropomorphism. *Journal of the Association for Consumer Research*, *3*(4), 591-598.

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of medical Internet research*, 21(5), e13216.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, *42*(3-4), 143-166.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689-707.

Fosch-Villaronga, E., & Poulsen, A. (2020). Sex care robots: exploring the potential use of sexual robot technologies for disabled and elder care. *Paladyn, Journal of Behavioral Robotics*, *11*(1), 1-18.

Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2019). Gathering expert opinions for social robots' ethical, legal, and societal concerns: Findings from four international workshops. *International Journal of Social Robotics*, 1-18.

Fox, S. (2018). Cyborgs, robots and society: Implications for the future of society from human enhancement with in-the-body technologies. *Technologies*, *6*(2), 50.

Fox, J., & Gambino, A. (2021). Relationship Development with Humanoid Social Robots: Applying Interpersonal Theories to Human/Robot Interaction. *Cyberpsychology, Behavior, and Social Networking*.

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: extending the computers are social actors paradigm. *Human-Machine Communication*, *1*(1), 5. IEEE.

Gibson, R. (2019). *Desire in the age of robots and AI: An investigation in science fiction and fact*. Springer Nature

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627-660.

de Graaf, M. M., Allouch, S. B., & van Dijk, J. A. (2016). Long-term evaluation of a social robot in real homes. *Interaction studies, 17*(3), 462-491.

Gutiu, S. M. (2016). The roboticization of consent. *In Robot law*. Edward Elgar Publishing.

Hajdú, J. (2020). Dignity of elderly persons and digitalised social care. *The Right to Human Dignity*, 569.

Henschel, A., Laban, G., & Cross, E. S. (2021). What Makes a Robot Social? A Review of Social Robots from Science Fiction to a Home or Hospital Near You. *Current Robotics Reports*, 1-11.

Hofstadter, D. R. (1979). Gödel, Escher, Bach: an eternal golden braid (Vol. 20). New York: Basic books.

Ishiguro, N. (2017). Care robots in Japanese elderly care. Cultural values in focus. In Christensen K, Pilling D, editors. *The Routledge Handbook of Social Care Work around the World: Routledge*, 256-269.

Kant, I. (2001). Lectures on ethics (Vol. 2). *Cambridge University Press.*

Karar, A. S., Said, S., & Beyrouthy, T. (2019, April). Pepper Humanoid Robot as a Service Robot: a Customer Approach. In 2019 *3rd International Conference on Bio-engineering for Smart Technologies* (BioSMART), 1-4. IEEE.

Kubes, T. (2019). New Materialist Perspectives on Sex Robots. A Feminist Dystopia/Utopia?. *Social Sciences, 8*(8), 224.

Levy, D. (2009). *Love and sex with robots: The evolution of human-robot relationships.* New York.

MacInnis, D. J., & Folkes, V. S. (2017). Humanizing brands: When brands seem to be like me, part of me, and in a relationship with me. *Journal of Consumer Psychology*, 27(3), 355-374.

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI*, 4, 21.

Mordor Intelligence. (2021). *Social Robots Market - Growth, Trends, Covid-19 Impact, and Forecasts* (2021-2026). https://www.mordorintelligence.com/industry-reports/social-robots-market

Nyholm, Sven, and Lily Frank, 2017, "From Sex Robots to Love Robots: Is Mutual Love with a Robot Possible?"*, in Danaher and McArthur* 2017: 219–243.

Parisod, H., Aromaa, M., Kauhanen, L., Kimppa, K., Laaksonen, C., Leppänen, V., ... & Salanterä, S. (2014). The advantages and limitations of digital games in children's health promotion. *Finnish Journal of eHealth and eWelfare,* 6(4), 164-173.

Perse, E. M., & Rubin, R. B. (1989). Attribution in social and parasocial relationships. *Communication Research*, 16(1), 59-77.

Pinillos, R., Marcos, S., Feliz, R., Zalama, E., & Gómez-García-Bermejo, J. (2016). Long-term assessment of a service robot in a hotel environment. *Robotics and Autonomous Systems, 79*, 40-57.

Reig, S., Carter, E. J., Tan, X. Z., Steinfeld, A., & Forlizzi, J. (2021). Perceptions of Agent Loyalty with Ancillary Users*. International Journal of Social Robotics, 1-17*.

Richardson, K. (2016). The asymmetrical'relationship' parallels between prostitution and the development of sex robots. *ACM SIGCAS Computers and Society, 45(3)*, 290-293.

Rosenthal-von der Pütten, A. M., Schulte, F. P., Eimler, S. C., Hoffmann, L., Sobieraj, S., Maderwald, S., ... & Brand, M. (2013, March). Neural correlates of empathy towards robots. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction,* 215-216. IEEE Press.

Scheutz, M. (2011). 13 The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots. *Robot ethics: The ethical and social implications of robotics*, 205-206.

Searle, J. R. (1982). The Chinese room revisited. *Behavioral and brain sciences*, 5(2), 345-348.

Shamsuddin, S., Yussof, H., Ismail, L., Hanapiah, F. A., Mohamed, S., Piah, H. A., & Zahari, N. I. (2012, March). Initial response of autistic children in human-robot interaction therapy with humanoid robot NAO. In 2012 *IEEE 8th International Colloquium on Signal Processing and its Applications* (pp. 188-193). IEEE.

Schiappa, E., Allen, M., & Gregg, P. B. (2007). Parasocial relationships and television: A meta-analysis of the effects. *Mass media effects research: Advances through meta-analysis*, 301-314.

Stanford Encyclopedia of Philosophy. (2020, April 30). Ethics of Artificial Intelligence and Robotics. *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/ethics-ai/

Statista & Vailshery, L. S. (2021, January). *Global social and entertainment robot unit sales* 2015–2025. https://www.statista.com/statistics/755677/social-and-entertainment-robot-sales-worldwide/

Stroessner, S. J. (2020). On the social perception of robots: measurement, moderation, and implications. *In Living with Robots (pp. 21-47).* Academic Press.

Thilmany, J. (2007). The emotional robot: Cognitive computing and the quest for artificial intelligence. *EMBO reports, 8*(11), 992-994.

Tousignant, B., Eugène, F., & Jackson, P. L. (2017). A developmental perspective on the neural bases of human empathy. *Infant Behavior and Development*, 48, 5-12.

Turkle, S. (2010). In good company?. *In Close engagements with artificial companions*, 3-10. John Benjamins.

Vallor, S. (2011). Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century. *Philosophy & Technology*, 24(3), 251-268.

Vanderelst, D., & Willems, J. (2019). Can We Agree on What Robots Should be Allowed to Do? An Exercise in Rule Selection for Ethical Care Robots. *International Journal of Social Robotics*, 1-10.

Vollmer, A. L., Read, R., Trippas, D., & Belpaeme, T. (2018). Children conform, adults resist: A robot group induced peer pressure on normative social conformity. *Science Robotics*, 3(21).

Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, 19(4), 393-407.

Weizenbaum, J. (1984). *Computer power and human reason: From judgement to calculation.* Harmondsworth UK: Penguin.