Human-Social Robots Interaction: the blurred line between necessary anthropomorphization and manipulation.

Rachele Carli
UNIBO, UNILU
Bologna, Italy
rachele.carli2@unibo.it

Amro Najjar
LIST
Luxembourg City, Luxembourg
amro.najjar@list.it

Davide Calvaresi
HES-SO Valais/Wallis
Sierre, Switzerland
davide.calvaresi@hevs.ch

ABSTRACT

In the context of human-social robot interaction, it has been proven that an affable design and the ability to exhibit emotional and social skills are central to fostering acceptance and more efficient system performance. Nevertheless, these features may result in manipulative dynamics, able to impact the psychological sphere of the users, affecting their ability to make decisions and to exercise free, conscious will. This highlights the need to identify a legal framework that balances the interests at stake. To this end, the principle of human dignity is proposed here as a criterion to ensure (i) the protection of users' fundamental rights, and (ii) an effective and truly human-friendly technological development.

CCS CONCEPTS

• **Human-centered computing** → *Interaction design process and methods.*

KEYWORDS

HRI, anthropomorphization, manipulation, human-dignity

ACM Reference Format:
Rachele Carli, Amro Najjar, and Davide Calvaresi. XXXX. Human-Social Robots Interaction: the blurred line between necessary anthropomorphization and manipulation.. In *Proceedings of Conference Title (HAI'22).* ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Social robots are conceived to operate closely with humans through a long-term interaction based on collaboration and/or substitution in daily tasks [9]. Therefore, they are designed to display social skills, elicit a positive empathic response from the user, and appear physically pleasant. This is the result of a mere, albeit sometimes highly sophisticated, emulation [3]. Indeed, robots are still far from being sentient and endowed with emotional depth. These emulated mechanisms facilitate the interaction and the goals for which the machine was designed [27], but they also may have manipulative consequences. Interaction may induce users to create an emotional attachment and trust that, if not modulated and controlled, can be detrimental to their mental, physical and economic integrity.

HAI'22, Date, Place

ACM ISBN XXXXXXXXXXXXXXX…$15.00
https://doi.org/XXXXXXX.XXXXXXX

Hence, this paper aims to underline the necessity for such a dynamic to become the subject of legal analysis to identify the most appropriate regulatory instrument to ensure the dissemination of technologies that serve rather than use humans.

To this end, Section 2 presents the features of social robots designed to mirror human characteristics and how they may lead to manipulative effects and detrimental consequences for users. Section 3 proposes Human Dignity as a crucial principle for balancing responsible technological development with the inalienable protection of the human person. Section 4 concludes the paper.

2 SOCIAL ROBOTICS MIRRORING HUMANS

From the Imitation Game onward, researchers have developed increasingly sophisticated systems to replicate features familiar to humans, thus facilitating the interaction [20]. Human-inspired systems demand information processing, communication abilities, task execution, and movements – crucial for modern Human-Robot Interaction (HRI) – is called anthropomorphization [2]. It is the result of ages of studies on human cognitive, emotional, and perceptual structures to exploit such knowledge in encoding both embodied and non-embodied AI systems [13, 22]. Nonetheless, numerous current social robots are not developed with the "primary" ambition to look like humans or living beings in general. In those cases, the main aim is to design them to give the user the idea of being in the presence of "someone" and not of "something" [8]. The goal is to reproduce a human-like experience of sociability, companionship, and support.

Thus, some devices are equipped with a round, often tilted head, big eyes, child-like size, moving arms (if any), a human-like voice, and a proper name [18, 25]. These features convey identity to the robot, instilling tenderness [1], attachment, protection, and care in the users [14]. Indeed, social robots are often designed to appear doubtful in their actions, clumsy, or seeking support [16] to elicit collaboration, teamwork, and greater tolerance in case of technical inefficiency. For instance, Pepper is often immortalized by scratching his head or thinking out loud before answering. This is merely functional to fill the time needed to process information. However, it was deliberately chosen because people link this act to reflection and self-analysis, being more inclined to consider the final act as "pondered" [7]. An analogous effect is obtained by programming the machine with a "cheating function" [28], which contributes to conveying the idea of intentionality and substantial autonomy [21]. The strategies to advertise these devices to the public contribute to the same end [5]. The terminologies used to present robots, their abilities, and usefulness mostly draw on typically anthropomorphic elements or capabilities — necessary to make highly complex and technical concepts accessible to non-specialists.

2.1 From anthropomorphisation to (potential) manipulation

Following the analysis above of the main strategies of anthropomorphization, we can elicit two main remarks. First, HRI efficiency seems rooted chiefly in what the machine can obtain from the user. Second, when engaging with social robots, individuals seem to evaluate an effective performance much less than the possibility of a human-like experience of interaction – even if apparent. Therefore, anthropomorphization works at two levels: (i) by design aspects implemented in the machine, (ii) by the psycho-emotional bond via interaction [4]. Indeed, many of the dynamics described above appeal to the same areas of the brain that are activated by drugs or other dysfunctional practices (e.g., gambling), the *nucleus accumbens* [14, 24]. This can induce an actual form of addiction. Moreover, such an approach may induce disengagement with reality [30]. This can have unwanted consequences, depending on the nature of the subjects involved and on the tasks to be performed by the robot. Particularly, in fragile subjects it can induce an erroneous perception of human-human social interaction, thus encouraging isolation and loss of relational skills, which could result in the dehumanization [26]. Moreover, people can overestimate the actual capabilities of the AI system, resulting in over-trust episodes, which are detrimental to their safety [17]. Being part of the users' everyday life and needing to demonstrate a sufficient level of (apparent) autonomy, these devices also need to have access to a vast amount of data (necessary to track people's behavior and emotional states) without a clear understanding about by whom and how this information will be handled, analyzed, and stored [29]. Even the words we use to describe social robots, and with which they are programmed to talk present some critical aspects. Using an accessible and reassuring vocabulary might not fuel unfounded skepticism and fears and should not lead to the conscious and systematic distortion of reality. An equally effective but more intellectually (and ethically) honest way of discussing these technologies should be investigated.

3 HUMAN DIGNITY: BALANCING PRINCIPLE

Acceptance, frequency, and willingness to repeat the interaction are fundamental pillars of the HRI. To be successful in this sense, devices are designed to tap into the fallibility which characterizes human beings [31]. Such an approach, if not accurately balanced, may lead to forms of manipulation that produce a different effect, depending on the contingency and specificity of the user involved, the scenario of the interaction, and the performance required. For this reason, a change in perspective is suggested to put humankind back at the center of research and new technology development. In particular, we claim the need to identify an element able to balance (i) innovation strive with (ii) the protection of individuals' fundamental rights. In doing that, safeguarding users' psychological and decision-making integrity has to be considered an overriding aim. Such a criterion could be Human Dignity.

Although it is often accused of vagueness and lack of definition [12], it is a crucial, inherent, inalienable principle – both ethical and legal [10] – without which "we would be unable even to answer the simple question: what is wrong with slavery?" [19]. Indeed, it has already proven to be an effective legal instrument. In many countries, in the name of human dignity, decisive reforms have been implemented. Among others, the abolition of torture and the death penalty [23], the ban on working conditions considered degrading and dehumanizing, and the enactment of the Convention on the Rights of the Child [11, 15]. This, without the lack of definition representing an application obstacle [6].

Concerning social robotics, human dignity could be applied to evaluate the adequacy of a given technology, both when it is implemented and tested in the laboratory and to assess possible corrections when it is on the market. More specifically, it could represent an essential criterion to determine which degree of anthropomorphization is considered acceptable from a case-by-case perspective, taking into account the class of devices considered, the typology of end users (distinguish among experts, children, the elderly, disabled, general public, etc.), and the specific tasks that the robot needs to perform (see Figure 1).

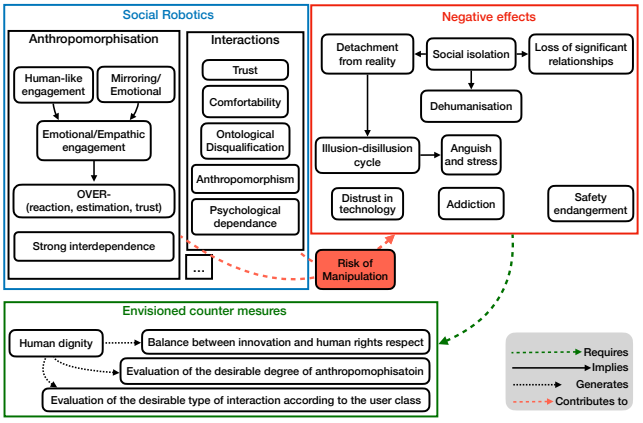

Figure 1: Principal conceptual components entanglement.

4 CONCLUSIONS AND FUTURE WORKS

We have underlined the HRI's benefits obtained from designing robots capable of simulating sociality, and empathy does not *per-se* justify possible manipulative effects on end users. This dynamic requires the utmost attention from technical experts and jurists. The aim should be to find a balance between the demands posed by technological development and those emerging from the need to protect individuals' integrity. To this end, we suggest using human dignity as the parameter to assess the characteristics different classes of devices should have to meet both the requirements from an objective and concrete perspective. Future research will focus on identifying precise procedures to make this process effective and more practical.

5 ACKNOWLEDGMENTS

This work has been par ally supported by the Chist-Era grant CHIST-ERA-19-XAI-005, and by *(i)* the Swiss National Science Foundation (G.A. 20CH21_195530), *(ii)* the Italian Ministry for Universities and Research, *(iii)* the Luxembourg National Research Fund (G.A. INTER/CHIST/19/14589586), *(iv)* the Scientific and Research Council of Turkey (TÜBİTAK, G.A. 120N680).

REFERENCES

[1] Thomas R Alley. 1983. Infantile head shape as an elicitor of adult protection. *Merrill-Palmer Quarterly (1982-)* (1983), 411–427.

[2] Paul Bartha. 2013. Analogy and analogical reasoning. (2013).

[3] Andrea Bertolini and Rachele Carli. 2022. Human-Robot Interaction and User Manipulation. In *International Conference on Persuasive Technology*. Springer, 43–57.

[4] Herbert Blumer. 1986. *Symbolic interactionism: Perspective and method*. Univ of California Press.

[5] Jeffrey A Brooks, Holly Shablack, Maria Gendron, Ajay B Satpute, Michael H Parrish, and Kristen A Lindquist. 2017. The role of language in the experience and perception of emotion: A neuroimaging meta-analysis. *Social Cognitive and Affective Neuroscience* 12, 2 (2017), 169–183.

[6] Rachele Carli. 2021. Social robotics and deception: beyond the ethical approach. *Proceedings of BNAIC/BeneLearn 2021* (2021).

[7] Antonio Chella and Arianna Pipitone. 2020. A cognitive architecture for inner speech. *Cognitive Systems Research* 59 (2020), 287–292.

[8] Luisa Damiano and Paul Dumouchel. 2018. Anthropomorphism in human–robot co-evolution. *Frontiers in psychology* 9 (2018), 468.

[9] Kerstin Dautenhahn. 2007. Socially intelligent robots: dimensions of human–robot interaction. *Philosophical transactions of the royal society B: Biological sciences* 362, 1480 (2007), 679–704.

[10] Klaus Dicke. 2001. The founding function of human dignity in the Universal Declaration of Human Rights. In *The concept of human dignity in human rights discourse*. Brill Nijhoff, 111–120.

[11] Horst Dreier. 2011. Die „guten Sitten "zwischen Normativität und Faktizität. In *Gedächtnisschrift für Theo Mayer-Maly*. Springer, 141–158.

[12] Muriel Fabre-Magnan. 2007. La dignité en droit: un axiome. *Revue interdisciplinaire d'études juridiques* 58, 1 (2007), 1–30.

[13] Jean-Marc Fellous and Joseph E Ledoux. 2005. Toward Basic Principles for Emotional Processing: What the Fearful Brain Tells the Robot. (2005).

[14] Melanie L Glocker, Daniel D Langleben, Kosha Ruparel, James W Loughead, Ruben C Gur, and Norbert Sachser. 2009. Baby schema in infant faces induces cuteness perception and motivation for caretaking in adults. *Ethology* 115, 3 (2009), 257–263.

[15] Manuel Gros. 2013. Il principio di precauzione dinnanzi al giudice amministrativo francese. *Il principio di precauzione dinnanzi al giudice amministrativo francese* (2013), 709–758.

[16] Andrea L Guzman. 2015. *Imagining the voice in the machine: The ontology of digital social agents*. Ph. D. Dissertation. University of Illinois at Chicago.

[17] Yaniv Hanoch, Francesco Arvizzigno, Daniel Hernandez García, Sue Denham, Tony Belpaeme, and Michaela Gummerum. 2021. The robot made me do it: Human–robot interaction and risk-taking behavior. *Cyberpsychology, Behavior, and Social Networking* 24, 5 (2021), 337–342.

[18] Carl Gustav Jung. 2014. *Il libro rosso: liber novus*. Bollati Boringhieri.

[19] Leszek Kolakowski. 2002. What is left of Socialism. *First Things: A Monthly Journal of Religion and Public Life* (2002), 42–47.

[20] Cherie Lacey and Catherine Caudwell. 2019. Cuteness as a 'dark pattern'in home robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 374–381.

[21] Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan. 2006. Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of communication* 56, 4 (2006), 754–772.

[22] Simone Natale et al. 2021. *Deceitful media: Artificial intelligence and social life after the Turing test*. Oxford University Press, USA.

[23] Conor O'Mahony. 2012. There is no such thing as a right to dignity. *International Journal of Constitutional Law* 10, 2 (2012), 551–574.

[24] Paolo Riva, Simona Sacchi, and Marco Brambilla. 2015. Humanizing machines: Anthropomorphization of slot machines increases gambling. *Journal of Experimental Psychology: Applied* 21, 4 (2015), 313.

[25] Mark Seltzer. 2014. *Bodies and Machines (Routledge Revivals)*. Routledge.

[26] Amanda Sharkey. 2014. Robots and human dignity: a consideration of the effects of robot care on the dignity of older people. *Ethics and Information Technology* 16, 1 (2014), 63–75.

[27] Jaeeun Shim and Ronald C Arkin. 2016. Other-oriented Robot Deception: How can a robot's deceptive feedback help humans in HRI?. In *International Conference on Social Robotics*. Springer, 222–232.

[28] Elaine Short, Justin Hart, Michelle Vu, and Brian Scassellati. 2010. No fair!! an interaction with a cheating robot. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 219–226.

[29] Tom Sorell and Heather Draper. 2017. Second thoughts about privacy, safety and deception. *Connection Science* 29, 3 (2017), 217–222.

[30] Robert Sparrow and Linda Sparrow. 2006. In the hands of machines? The future of aged care. *Minds and Machines* 16, 2 (2006), 141–161.

[31] Joseph Weizenbaum and Gunna Wendt. 2015. *Islands in the cyberstream: Seeking havens of reason in a programmed society*. Litwin Books.