Social robotics and the outsource of agency: Untangling the ethical approach

Júlia Pareto Boada
Institut de Robòtica i Informàtica Industrial, CSIC-UPC
Facultat de Filosofia (UB)
Barcelona,Spain
jpareto@iri.upc.edu

***Abstract*— This article aims at straightening the basic elements of a suitable ethical approach to social robotics. To do so, it starts by identifying the phenomenon that triggers ethical concerns on social robots, namely the 'outsource' of human agency, and proceeds to critically examine the perspective that should be adopted towards it. Emphasizing that attention must be foremost focused on the 'being' rather than the 'doing' of social robots, it argues for a mature normative reflection that takes into account the 'interested' agency of the intelligent artefacts and puts the focus on human agency.**

Keywords—artificial intelligence, agency, ethics, human-robot interaction, social robotics

I. Introduction

The technoscientific field of intelligent robotics has increasingly become the focus of ethical reflection. Such critical thinking emerges from an awareness of the challenges that the embodiment of intelligence in form of artificial agents able to perform goal-oriented actions with certain degree of autonomy in real environments may pose to our way of life.

Indeed, the possibility of commending to robots not only automatable physical tasks, but also the performance of actions upon the physical space which directly result of a previous cognitive management of information (either given or gathered through sensors) and a decision-making process over it, broadens the scope of the roles they can assume. In turn, it widens the range of their ethical implications and, thus, the landscape of the normative reflection needed to ensure that the introduction of intelligent robots is in line with the moral values and human rights that we hold as constitutive of a quality life. The emergence of the disciplines of Roboethics[1], Machine ethics and the most recently one of Robophilosophy[2] is clearly responsive to this urge.

The expansion of robots' potential in supporting human activity enables them to be deployed in practical contexts that were until very recently exclusively reserved to human agency, such as the ones involving a certain kind of social interaction, as it is the case of the context of care, education or companionship. The so-called social robots are a clear example of such advances in intelligent and autonomous robotics. They are a specific kind of artificial agents that provide technological assistance in practices that belong to these domains of human life, by means of socially interacting with humans[3]. This novelty has placed social robotics under strong ethical scrutiny. Personal assistants in the domestic realm, care robots in the healthcare sector, specialized assistive robots in cognitive rehabilitation or enhancement therapy contexts –among other similar socially interactive robots in service functions– raise concerns on several issues such as human dignity, autonomy, personal freedom, deception, privacy, accountability, devaluation of human practices, human abilities' degeneration, etc.

II. The pivotal phenomenon of ethical concerns

It is worth noticing, though, that the general current landscape of ethical reflection on intelligent and autonomous robotics ultimately pivots around one central phenomenon, namely the outsource of agency. Indeed, the technological autonomy of intelligent robots seems to reasonably invite us to a very specific ethical thinking rooted in a view of robots as entities in which we can delegate agency (both cognitive and resolutive) in some specific task. However, this horizon of outsourcing human agency in the diverse domains of our lives structures the ethical approach to robotics in a very particular and inaccurate way, for the reason below explained. This also applies for the case of social robotics, which is the specific object of consideration of this article. Likewise, it is mostly in virtue of the possible transfer of certain practices or tasks belonging to human abilities that major concerns arise in the debate. Those typically encompass, among others, infringements on human dignity resulting of objectifying or deceiving users of assistive technologies[4][5], the denigration of human skills, interference with human autonomy and personal freedom and devaluation of care practice[6] –due, for instance, to the robotization of tasks in activities that have a holistic nature.

In which sense can it be stated that the question of the outsource of agency gives rise to a misguided ethical approach to robotics and, subsequently, to its subfield of social robotics? The reason is that precisely because of understanding intelligent robots as artificial agents that can take over actions with specific goals within a given context, there has been a distinctive rush to reflect upon how robots must act, how they should behave, so that they can legitimately enter our social space and assume its distinctive practices. To a certain extent, it surprisingly seems to have been assumed that whether a robot is beneficial or not basically depends on its behavior. This has led attention to questions essential to 'machine ethics'[7][8][9], which are about how (if possible) to endorse robots with moral competence, understanding it as the ability to act rightly. Although these considerations are surely relevant, turning reflection mainly on the 'doing' of such entities overlooks the primary focus of ethical attention. By viewing robots as (individual) artificial agents that are expected to perform all sort of tasks in the framework of our everyday life, the emphasis is easily first and foremost put in granting that they "act" correctly, that is to say, that they take into account our main principles and (reasonable) values when deciding the course of their actions.

III. STRAIGHTENING THE ETHICAL APPROACH TO SOCIAL ROBOTICS

However, the fundamental question is not about the 'doing', but rather the 'being' of an intelligent robot. Prior to 'correctly doing', a (social) robot must 'correctly be'. The question about 'correctly being' is meant to be understood in terms of a coherence between the reason of being of the artificial agent and the purposes it is aimed to serve to through its functionalities. It has to do with the idea of 'having sense', 'being legitimate'. Seemingly, the relevance of this dimension has not gone unnoticed by authors addressing the suitability of other kind of technologies in light of aspects such as the problems that these are supposed to solve[10] or the values that their design actually support[11]. Nevertheless, it is imperative to overtly draw attention to this question, since asking about the correct being of a robot is what an ethical approach to social robotics must foremost be concerned of. Otherwise said, it is the question that should always primarily shape the perspective from which to reflect upon social robotics in each of its stages of development and deployment. The reason is twofold.

On the one hand, firstly focusing on 'being the correct thing' is congruent with the fact that, given its technological nature, social robotics has not, by itself, a purpose that can be said to be totally unlinked from the field of human action, practice or activity in regard to which it develops its artefacts. On the contrary, the ultimate end of social robotics is to serve the particular ends of those fields of activity in which its products are applied so that to serve as a means of support. The core particularity of social robots is the ability to interact in a human-comprehensible way. This is thus the functionality around which ethical reflection is mainly addressed. However, as pointed out, because of serving the ends of the field of action in which robots are inserted, interaction as a functionality means that it is always provided with a certain goal. This implies that ethically thinking about social robots in order to offer a normative guidance for their deployment equates to thinking in a 'situated' or 'applied' way. That is to say, it consists in taking into account the spheres of human activity where those artificial agents are intended to be deployed. Ethical challenges of social robots as interactive artificial agents cannot be tackled on from a vacuum, in a decontextualized way. A suitable ethical approach should always consider which is the goal of interaction: is it interaction for the very same sake of interaction (as it would be the case of a companion robot), or is interaction aimed at assisting in some task (for instance, improving the performance of a patient in a cognitive therapy activity by means of conducting the exercise in an adaptative way)?

On the other hand, the 'being' perspective unfolds the potential of the ethical gaze on social robotics. Rather than reducing the ethical approach exclusively to a critical assessment of the robots' (foreseeable) impacts, it expands it to a normative reflection on the transformative force of social robots regarding human practices. If setting aside the perspective of social robots as artificial agents that can assume tasks that are essentially human (such as the ones that need of interaction), we can engage in a most mature way of ethical reflection. The reason is that the former perspective entails a misleading conception of social robots as individual agents[12] differentiated from us, in the sense that they enter in our practical contexts as 'others' that have an own (even though framed) agency and 'who' therefore produce some impacts in their performance. Our responsibility lays then on ensuring that these impacts will be positive. The fixation on impact entails some well-known difficulties, such as the unpredictability of technological outcomes[13] and the question about from which perspective should impacts be estimated (on the basis of which values or principles, and in regards to whom). Besides, reflecting upon social robotics mainly in view of consequences entails the risk of falling into a moral conservatism, that is, turning the ethical gaze into a moral assessment. Remind that ethics is an upper level of perspective, which reflects upon the grounds of morals so that to legitimate them or not –and thus, abandon, maintain or (re)generate them. Ethics is concerned with the reasons for certain practices, habits, values. It is a normative approach based on the 'why' question, rather than the 'what': whereas morals are concerned about actions –what must I do–, ethics is concerned about reasons –why should I do it[14]. Therefore, attending only the impacts as if those were generated by external agencies may precipice us to entrench our reflection in the framework of current values, without attending the need of constantly rethinking them, as ethics implies.

A richer perspective is possible, though, if we distance ourselves from the idea of contraposed 'subjectivities' in play –in which robots are mistakenly conceived as well-delimited and somehow 'closed' focus of agency–, and rather pay attention to the fact that a robot, as a human-made artefact[15], always responds to an externally given ultimate purpose, in the double sense that it is a purpose both fixed by humans and linked to the goal of the specific practical context in which the artificial entity performs its role. This shifts the ethical attention to where it mainly belongs, that is, human action, at the time that it brings to light a central concept for the approach to social robotics, which is the one of 'interest'. The artificial agency of a social robot is always an 'interested' agency, in the sense that its reason of being responds to a certain purpose. This is a purpose that is decided by humans, from within a specific political, social and cultural context, and that must be critically taken into account by the ethical reflection.

Moreover, focusing on such perspective enlarges what we could call the spheres of human life that must be considered in an ethical approach to social robotics. Indeed, it is noticeable that ethical literature on social robots is mostly concerned with the impacts that such artefacts may have at the individual level of human life. This could be related to the tendency of approaching social robotics by focusing on the 'outsource of agency' phenomenon, which implies a classical paradigm of robots as individual agents taking over certain human roles and thereof impacting upon the individuals they interact with in the immediate context of its deployment. Contrarily, attending firstly to the 'being' of the robot implies adopting a more comprehensive perspective that concentrates on how humans, as the ones that decide the 'what for' of the robot, transform the landscape of their activities and practices with the inclusion of such technologies in the equation. This is a perspective that can, thus, take into account other spheres besides the individual one, such as the interpersonal, social, sectorial and institutional ones. It is a reflection that focuses on the structural reality of our human life, on the practices and architectures underlying its logic. It is therefore a more mature one. Indeed, an ethical approach to social robotics should foremost be concerned about how to guide the technological (re)structuring power of human practices, which is something that depends on humans alone.

IV. CONCLUSIONS

To conclude, the upgrading autonomy of intelligent robots able to perform an increasing range of roles triggers the need of an ethical approach to social robotics. However, it is crucial not to lose sight of the primary focus of ethical attention, which has to do with a paradigm of social robots as entities that are 'interested' agents. The phenomenon of the 'outsource of agency' that actually (and justifiably) structures the common approach to social robotics and its artefacts, must do it always in connection with the 'interest' phenomenon. Indeed, it is an 'interested agency' the one that we outsource. Only if departing from this perspective, will we obtain a proper ethical approach to robotics. However, this is only possible if putting the emphasis firstly on the question of 'being', rather than the one of the 'doing' of the social robot.

REFERENCES

[1] F. Operto and G. Veruggio, "Roboethics: Social and Ethical Implications of Robotics," in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer, 2008, pp. 1499–1524.

[2] J. Seibt, "Robophilosophy," in *Posthuman Glossary*, R. Braidotti and M. Hlavajova, Eds. London: Bloomsbury Academic, 2017, pp. 390–393.

[3] C. Breazeal, A. Takanishi, and T. Kobayashi, "Social Robots that Interact with People," in *Springer Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer, 2008, pp. 1349–1369.

[4] A. Sharkey and N. Sharkey, "Granny and the robots: Ethical issues in robot care for the elderly," *Ethics Inf. Technol.*, vol. 14, no. 1, pp. 27–40, Mar. 2012.

[5] F. M. Noori, Z. Uddin, and J. Torresen, "Robot-Care for the Older People: Ethically Justified or Not?," *2019 Jt. IEEE 9th Int. Conf. Dev. Learn. Epigenetic Robot.*, pp. 43–47, 2019.

[6] S. Vallor, "Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century," *Philos. Technol.*, vol. 24, no. 3, pp. 251–268, 2011.

[7] C. Allen, W. Wallach, and I. Smit, "Why machine ethics?," *IEEE Intell. Syst.*, vol. 21, no. 4, pp. 12–17, 2006.

[8] M. Brundage, "Limitations and risks of machine ethics," *J. Exp. Theor. Artif. Intell.*, vol. 26, no. 3, pp. 355–372, 2014.

[9] S. Cave, R. Nyrup, K. Vold, and A. Weller, "Motivations and Risks of Machine Ethics," *Proc. IEEE*, vol. 107, no. 3, pp. 562–574, Mar. 2019.

[10] E. P. S. Baumer and M. S. Silberman, "When the implication is not to design (technology)," *Conf. Hum. Factors Comput. Syst. - Proc.*, pp. 2271–2274, 2011.

[11] J. Millar, "Technology as Moral Proxy: Autonomy and Paternalism by Design," *IEEE Technol. Soc. Mag.*, vol. 34, no. 2, pp. 47–55, 2015.

[12] M. Coeckelbergh, "Is ethics of robotics about robots? Philosophy of robotics beyond realism and individualism," *Law, Innov. Technol.*, vol. 3, no. 2, pp. 241–250, 2011.

[13] H. Jonas, *El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica.*, Herder. Barcelona, 2015.

[14] A. Cortina, *Ética mínima. Introducción a la filosofía práctica*, Tecnos. Madrid, 2007.

[15] D. Johnson G., "Computer Systems: Moral Entities but Not Moral Agents," in *Machine Ethics*, M. Anderson and S. L. Anderson, Eds. Cambridge University Press, 2011, pp. 168–183.