Sustainable Anthropomorphism for the Improvement of Social Robotics

Investigating solutions to the issues of anthropomorphic technology applied to Social Robotics

Marco Petix
05/07/2020

Abstract. This essay aims to shed light on the discussions revolving around anthropomorphic and social robotics, and to hypothesize strategies to face the problems involved in these fields of research.

The issues concerning the tendency of users to personify social robots, and subsequently develop an emotional involvement in them, have been presented and discussed from practical and ethical standpoints. The ethical investigation concerning the exploitation of the anthropomorphic tendency highlighted the potential inefficiency of the deontological approach compared to a more flexible, but nonetheless cautious, one. The comparison between the dichotomous and the interactionist view on the emotional interaction between humans and social robots has instead highlighted the promising benefits of the latter regarding the ability to differentiate between empowering or detrimental human-robot relationships.

The speculation regarding efficient strategies to strengthen the boundary between deliberate "as-if" engagement and potentially dangerous misinterpretation was steered by the results of an analysis of the psychological and cognitive foundations of anthropomorphism. The confirmations regarding the importance of the user's familiarity with the robot and its functioning, as well as its social connection with other humans, has encouraged original proposals and references to already existing studies regarding features potentially able to mitigate the aforementioned problems.

Keywords. anthropomorphism, social robotics, human-robot interaction

1 - Introduction

The field of social robotics is dedicated to the study of social interactions between human users and robots. The relational nature of these machines is the fundament of many applications and research fields such as robot-education, robot-caretaking and robot-companionship.

In order to effectively interact with their users, social robots must be able to navigate the human's social environment. Therefore, means of human-like communication[1] are often accompanied by the implementation of other elements of human-like behaviour and appearance.

The anthropomorphic design is also used for its ability to enhance HRI by ensuring familiarity that is easily accessible, yet misleading. This, however, together with the affective nature of social robots, constitutes a context not devoid of ethically gray areas and potentially unpleasant consequences. Interaction with anthropomorphic machines presents significant disadvantages and risks that are emphasized by the influence on the social and emotional sphere of said artifacts.

Assuming a future increase in both the diffusion of social robots and the credibility of their anthropomorphic characteristics, it is important to develop ways to address some of the problems that could arise from the use of this type of technology.

The investigation covered by this essay hypothesizes the possibility of developing solutions capable of preventing or mitigating traditional issues of social robotics through a study of the mechanisms underlying anthropomorphism.

To better define the concerns that fuel this inquiry, **Section 2** presents an overview of the main issues commonly associated with interactions with social robots. These are mainly linked to the development of unidirectional emotional bonds between human users and their machines, and the potential abuse of this emotional dependence.

Section 3 instead provides a technical and ethical analysis of anthropomorphism and its impact on social robotics. It investigates the main factors behind the emergence of anthropomorphic projections in order to identify the best intervention points and thus better guide the development of significant solutions to the problems previously presented.

On the ethical side, the aim is to discuss the correctness of exploiting anthropomorphising technology and the evocative nature of social robots. A brief discussion is devoted to the clash between the deontological condemnation of said technology and the presumed inevitability of human-like machines. Subsequently, the legitimacy of the emotional interaction encouraged by social robots is discussed with a comparison between dichotomous and interactionist approaches.

Ultimately, **Section 4** uses the hypothesis formulated during the previous technical analysis to discuss strategies on how to mitigate the user instinct to anthropomorphize and thus potentially negate its unappealing consequences.

---

[1] Speech-based communication is not strictly necessary, most robo-pets and some therapeutic robots, like Aibo and Paro, use motions and animal-like sounds to communicate simple messages to their users

2 - Personification and emotional dependence

Robots tends to be personified by their users, this is one of the main foundations, and consequences, of anthropomorphism. If it's true that not every social robot possess human-like appearance[2], the fact remains that the purpose of these robots, or the means by which they perform their function, as supporting psychological therapies, is the development of a meaningful connection with the user.

The anthropomorphization of robots entails a wide range of problems including those related to the users expectations, a phenomenon certainly inflated by the influence of fictional narratives and media coverage on artificial life and intelligence. In the case of social robots, the main problems tend to be related to the emotional connection developed by the user during their interaction with the device.

Reports on the relationship between soldiers and military robots, like those dedicated to the defusing of explosive devices, identify as possible seed of this emotional bond the collaboration towards a common goal and belonging to the same social ingroup, be it a particular platoon or team. It appears that robotic pets and toys, such as Aibo or My Real Baby, instead, make use of the need for care and attention from their users. This dynamic, based on an almost maternal instinct, is usually used in interaction with both children and elders. In the former context, however, the results of said interaction can acquire a surprising self-therapeutic nature. [8][9]

Researches on Paro, the robotic seal, shed light on the cathartic function of the interaction between the robot and its older users. In this case the projection of typical human traits is associated with a projection of the user's own self. Elders affected by depression, and/or anxiety and frustration, project the same emotional states upon the seal. Caring to the robot, by stroking its fur or using reassuring phrases, implicitly turns into attending to their own inner afflictions[7].

These researches demonstrate the beneficial impact that interacting with a robot like Paro can have on its users. However, the ambiguity linked to this phenomenon remains in the unidirectionality of the emotional bonds developed by the human participants in the interaction. These, also, tend to consciously, or unconsciously[3], engage into the illusory notion of reciprocity and this state of emotional dependence opens them to many forms of manipulation and abuse[6].

A manifestation of the aforementioned manipulation could be the influence on the user's decision-making process regarding the choice of particular products or services. Like, for example, the incremented allure of a potentially expensive upgrade for the robot subject of the user's attention.

---

[2]In reality a large percentage of research on social robots is currently based on the interaction with therapeutic and companion robots that have a design inspired by animals like Aibo and Paro
[3] Similarly to how a temporary self-delusion based on a "as-if" ascription of reciprocity ends in plain delusion based on actual ascription

Manipulation stands as a non-trivial component of the human social environment, humans are able to employ manipulation on one another but it's assumed that these practices tend to be inhibited by feelings like respect or loyalty. Said social mechanisms, however, can't possibly influence emotionless machines like the current generation of social robots.

A similar problem can be presented for trust and confidentiality issues. This is particularly concerning given the greater ease of opening up with social robots regarding private matters. A tendency, probably induced by the perceived lack of judgment, manifested by some of the users involved in the research mentioned above.[9]

Finally, according to sociologists such as Sherry Turkle and her theories regarding the culture of simulation, the convenient simplicity of social relations between humans and robots could alter the value that future generations attribute to authenticity and to relationships between humans[7].

Assuming that a greater control over the anthropomorphic tendency of the user will play a fundamental role in the prevention and mitigation of the problems previously described, the following section is dedicated to the analysis of anthropomorphism and of its inner workings as a psychological phenomenon.

3 - Investigating anthropomorphism

3.1 - Technical Analysis

Many studies on the topic address anthropomorphism as an uncontrollable phenomenon deeply rooted in our brain.[3][10] Some even describe it as an adaptive trait with a non-trivial role within the evolution of human beings as a species. [2]

For the purposes of this essay, this analysis is limited to anthropomorphism as a phenomenon emerging from the interaction between human users and human-like machines. It is obvious, in this regard, how the capacity in which the machine appearance and/or behavior resemble their human counterparts plays a fundamental role in inducing a sense of similarity into the user. Their mere physical presence and apparent autonomy of motion constitute significant incentives for anthropomorphic projections.

The influence that the duration and the context of the interaction with a robot have on the development of anthropomorphic tendencies instead, has been the focus of a research from Lemaignan et al.[5]. They assume an evolution in the capacity in which the user is prone to anthropomorphize a robot over a long-term interaction. Different phases, identified by the authors, describe the user journey from pre-interaction to the construction of a reliable model regarding the behaviour of said machine.

According to this research, the instinct to anthropomorphize, initially modeled by the user's expectations, reaches a peak close to the beginning of his interaction with the robot.

Provided that the user has the opportunity to familiarize himself with it, and therefore to associate a certain degree of predictability with the robot's behavior, the instinct to anthropomorphize should slowly decrease and eventually stabilize.

The research of Epley et al.[10][4], instead, investigates which psychological factors are mainly responsible of encouraging the anthropomorphic tendency within the interaction between a human and a non-human entity. These, being partially unsurprising due to the already stated notions regarding familiarity and predictability, are:
- the unfamiliarity of the human being with the entity taking part in the interaction;
- his motivation to understand the entity's apparent behavior;
- his potential lack of social connection.

The context outlined by these factors suggests that a user interacting with an unfamiliar entity, a social robot in this case, would instinctively draw from his wider and more reliable knowledge of other humans to better interact with it, rationalize its behavior with greater ease and eventually compensate for his inherent sense of loneliness.

The conclusions of Lemaignan et al. regarding the reduction of anthropomorphism during a long-term interaction don't seem, at least initially, to coincide with the many concerns focused upon the use of robot-companions as robot-nannies and robot-caretakers. The possibility of a strong emotional attachment developed by the users towards theirs partners, and therefore the serious increase in anthropomorphic projections, however, seems quite reasonable if one takes into account the vulnerable psychological state of children and the elderly. Furthermore, the apparent discrepancy further underlines the fundamental importance of a coherent model regarding the robot's capabilities and behavior for the user.

Given the strong dependence of both the context and the duration of the interaction with a social robot from the specific purpose of said machine, i consider the task of tampering with these parameters outside the scope of the general investigation provided by this essay. In this regard, strengthening both users' understanding and their social connection with other humans seem appropriate topics for the discussion to which Section 4 is dedicated.

Before continuing with the actual development of strategies for a safer and more sustainable exploitation of anthropomorphic and social technologies, I think as worthwhile to dedicate the rest of this section to a discussion about the ethical environment surrounding said technologies.

3.2 - Ethics Analysis

3.2.1 - Ethics of anthropomorphism, the deontological paradox

A tradition that we could trace back to Descartes' "The Passions of the Soul" evaluates the anthropomorphic tendency through a strictly dichotomous vision of authenticity versus simulation, the resulting definition of anthropomorphism is therefore that one of a cognitive bias. The ethical panorama concerning the exploitation of anthropomorphism, similarly to

that concerning the use of benevolent lies, is often the site of a debate between the deontological and consequentialist approaches.

The deontological approach has its roots in the Kantian theories regarding the inviolable and universal right to honesty and autonomy. The exploitation of anthropomorphism, taking advantage of a user's cognitive vulnerability, is therefore strongly condemned by the followers of this theory. Ultimately, however, the aforementioned condemnation perpetrated by the deontologists is far from stopping the development of anthropomorphic technology, as demonstrated by the current diffusion of companies interested in the development of said technology. It could be argued that the impact of the deontological approach upon the prevention of the actual dangers related to this technology may be inefficient. On the contrary, this rigid position could sever, or have already severed, the link between the ethical analysis and this specific domain of technological progress, thus precluding the possibility of an active participation in the research and development of this ever-growing technology.

The consequentialist approach is more flexible and permissive in this regard but it does not fail to consider the emotional or relational experiences among humans and anthropomorphic machines as a make-believe scenario, potentially useful, but nonetheless fictitious. However, is a challenge to the current robotic incarnation of Descartes' vision actually advisable?

3.2.2 - Dichotomous and interactionist approach to emotional robots

As mentioned above, the dichotomous vision inherited from Descartes highlights the lack of authenticity in the emotional expressions manifested by a robot. By actualising the French philosopher's theory, it's possible to infer that: machines are limited to the external expression of emotions by their lack of an internal dimension, the place of origin of emotions, according to Descartes, deeply dependent on the body. Therefore, this expressions of emotional states, manifested by the robot and perceived by its users, can only be mere simulations, inherently inferior reflections of the inner emotions exclusive of entities with a biological body.

This reasoning, however, may not be devoid of weaknesses. Damiano et al. in their paper discuss the theme of affective coordination between humans and robots from an interactionist standpoint[2].

According to the authors, the abilities of the robot:
- to induce into the user an emotional reaction through its affective expression;
- to react in an intelligent way to said reactions with new, still simulated but not pre-existing, expressions;

is enough to create an emotional dynamic where robot and user influence each other recursively. This form of coordination, or rather co-definition, should therefore qualify the relational machine as an affective agent and legitimate in some form its place as interlocutor within the interaction with the user.

Damiano et al. highlight the weaknesses of the reasoning based on the genuine/simulated division by comparing two examples of interaction with social robots. The first example

describes the use of the emotional bond developed by autistic children towards social robots in order to help them develop their social skills. The second example instead investigates the correctness of interacting with a sex robot through a predisposed "simulated rape" mode.

In the first example the illusion of reciprocal caring enforced by the interaction with the social robot has a tangible beneficial effect in its users. In the second example, however, the nature of agent/interlocutor of the sex robot, emerging from its affective coordination with the user, qualifies the simulated rape as a actual rape.

The aforementioned weakness of the dichotomous approach stands in its strong association of simulation with trickery. Said association in fact, is not able to perceive the ethical differences between the two examples and thus consider equally condemnable the empowerment of vulnerable users and the trivialization of a violent act.

Further elaborations on these concepts, however interesting, go beyond the scope of this essay. These arguments, however, demonstrated the potential legitimacy, and benefits, of the emotional interaction with a social robot, without however warding off the shadow of abuse and unsafe consequences. Therefore, strategies for the prevention and mitigation of anthropomorphic projections are still needed. These are discussed in the next section.

4 - Discussing strategies

4.1 - Nurturing the understanding of the user

The need for the user to understand the internal functioning of complex machines, like the current generation of social robots, seems to be wasting all the efforts made in prioritising user-friendliness in interfaces aimed at guaranteeing competence despite the transparency regarding the very same notions. Even when considering meeting this need as the cost of making interacting with this technology safer, the task of conveying these informations to everyday users is certainly no small feat.

The long-term interaction typical of many social robot applications, however, implicitly provides the means to deal with the complexity of this task and the notions involved in it, over a long period of time. Entrusting the responsibility of instructing the user to the social robot would mean the implementation of a meta-commentary through which the device could contextualize its behavior to the human[4] and, perhaps, outline the processes and factors behind its perception and decision making.

Interactions with the device would therefore acquire an educational nature and the "lessons" could be tackled with increasing levels of pertinence and complexity. Redundancy in the educational process enforced by multiple devices could be avoided by associating the users with particular levels of competence regarding the topics already addressed and by sharing these between each device with which the user would end up interactacting.

---

[4] And thus helping the user in rationalising the potentially unpredictable and unfamiliar conduct

While perhaps guided by good insights, this system appears relatively far-fetched when confronted with the need for user motivation and participation. Listening to the repeated sermons of a robot companion about artificial intelligence or complex robotics could easily result annoying to said user and thus reduce his acceptance and likeness of the robot itself.

4.2 - Dealing with the lack of social connection

Following the same reasoning just adopted would lead to an association with the same social robot of the role of promoter of social connections between humans. Well-known robots such as Paro have already achieved this effect passively with their application, as shared object, in elderly care centers. The presence of the robotic seal within the social environment of the elderly gave them something to discuss and establish a bond through, thus encouraging a renewed sense of community.

Another way to deal with this problem could be an active encouragement from the robot towards the user regarding the interaction with other humans. Assuming a context in which it would be possible to associate one device per user, the interaction between users could be masked and motivated as a need externalized by the robot itself. Taking advantage of the nurturing instinct elicited by pet-robots, the need for interactions with others of its robotic kind could be expressed by the latters, implicitly encouraging meetings between human users.

Finally, a recent research has studied the ability of reminders of significant social connections in increasing the perceived connection of the user and therefore reducing the instinct to anthropomorphism[1]. The robot could use the informations relating to the user's loved ones and acquaintances during its interaction with him in order to actively promote a virtual or physical meeting and implicitly reinforcing his memories of the connection. However, this does not take into account the privacy implications of said practice.

Conclusions

This essay examined how the interaction with a social robot can evolve into a state of emotional dependence between the device and its user, then vulnerable to manipulation and abuse. The beneficial aspects of said interaction were also recognized and in order to keep beneficing from them in a safe way, and to prepare to an assumed diffusion of social robots, the investigation was coupled with a speculation regarding possible solutions to these issues.

The ethical analysis regarding the deontological approach and its roots within the dichotomous view on simulated emotions highlighted the potential benefits hiding in the development in a new flexible form of ethical inquiry. Inspiration from more recent psychological and social theories[5] could also bring a new perspective into what may currently appear as a stagnant panorama.

---

[5] Like those regarding mirror neurons that inspired the notion of emotional co-definition between humans and robots

This increased permissiveness, however, doesn't comes without well accounted risks, as often claimed by Turkle and her assumption regarding the coming culture of simulation. If an overall condemnation is ineffective, the same applies to its total absence, the example previously presented about the sex robot being object of simulated rape is proof enough of this. Nonetheless, the application of ethical investigation on a single case/feature basis seems a more reasonable way to prevent dangerous ramifications without losing the potential benefits of this field of research.

The solutions discussed represent only a very rough idea on how to develop safer human-like robots. Another interesting, yet personally unexplored, answer to the issues related to the illusion of reciprocity in human-robot interaction is the development of social robots equipped with actual emotional models. Many research currently discuss the inherent benefits of empathic robots, some theories inspired by the idea of social intelligence even claim its potential in the development of an artificial intelligence closer to the human standard.

References

[1] Bartz, J. A., Tchalova, K., & Fenerci, C. (2016). Reminders of Social Connection Can Attenuate Anthropomorphism: A Replication and Extension of Epley, Akalis, Waytz, and Cacioppo (2008).

[2] Damiano L and Dumouchel P (2018) Anthropomorphism in Human–Robot Co-evolution.

[3] Duffy, Brian. (2003). Anthropomorphism and the social robot. Robotics and Autonomous Systems.

[4] Epley N, Waytz A, Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism.

[5] Lemaignan, Séverin & Fink, Julia & Dillenbourg, Pierre. (2014). The dynamics of anthropomorphism in robotics.

[6] Scheutz, Matthias. (2009). The Inherent Dangers of Unidirectional Emotional Bonds between Humans and Social Robots.

[7] Turkle, Sherry. (2007). Authenticity in the age of digital companions.

[8] Turkle, Sherry & Taggart, Will & Kidd, Cory & Dasté, Olivia. (2006). Relational artifacts with children and elders: The complexities of cybercompanionship.

[9] Turkle, Sherry. (2006). Robot as Rorschach: New Complicities for Companionship.

[10] Złotowski, J., Proudfoot, D., Yogeeswaran, K. *et al.* (2015) Anthropomorphism: Opportunities and Challenges in Human–Robot Interaction.