Some Ethical Reflections on Relations between Human Beings and Social Robots

Anne Gerdes

University of Southern Denmark, Institute of Business Communication and Information Science, Campus Kolding, Engstien 1, 6000 Kolding, Denmark
{Gerdes@sitkom.sdu.dk)

Abstract. The purpose of this paper is to reflect upon ethical implications of human-robot interaction. Issues are discussed within two scenarios: (1) In focusing on robots with intelligent behavior, but without consciousness, attention is paid to obstacles for forming trustful relations. Here, it is concluded that human-robot interaction will lack the kind of commitment, which stems from the fact that life is interpersonal, implying that trust is a fundamental human condition. (2) In focusing on the possibility of developing intelligent robots with a mental life of their own, issues of our responsibility as creators of robots are discussed, as well as issues dealing with the kind of relationships we might have with such robots. Here, we are faced with a Good-like responsibility and ethical obligations towards a creature, who possible will develop a mind of its own, which might turn out to be radically different from the human mind.

Keywords: Human-robot interaction, trust, artificial intelligence, ethics

1 Introduction

With the recent upcoming of more and more human-like robots, which of course still are nothing but "stupid machines", we might expect that such surprisingly human-like geminoids in a near future will be able to simulate intelligent behavior, when acting within restricted contexts.

On an epistemological level we may still argue about the status of intelligence. But, in real life people will start to form relationships with robots, whether they are intelligent or not. The fact that they look a lot like us, combined with their growing ability to behave intelligent will cause new forms of friending and bonding in connection with human-robot interaction.

The purpose of this extended abstract is to reflect upon ethical implications of such relationships. In particular issues are discussed from the perspectives below:

(1) In focusing on the possibility of developing robots with intelligent behavior, but without consciousness, attention will be paid to obstacles for establishing trustful relations in connection with human-robot interaction.

(2) In focusing on the possibility of developing intelligent robots with a mental life of their own, issues of our responsibility as creators of robots will be discussed, as

well as issues dealing with the kind of relationships we might establish with such robots[1].

2 Perspectives towards Artificial Intelligence

In what follows, I will arrange my discussion with reference to two well-known perspectives towards strong artificial intelligence; since these two positions raise similar as well as different sorts of ethical issues regarding the character of human-robot interaction.

Within a behaviorist framework, we might be concerned with the idea of artificial intelligence from a perspective of performance. Consequently, it is not considered a meaningful project to maintain a distinction between real human intelligence and artificial intelligence, if last mentioned is indistinguishable from human intelligent behavior. The behaviorist perspective focuses on appearance, in holding a definition of intelligence in which intelligence equals intelligent behavior. This idea is encapsulated in the famous Turing test [2], which has not yet been passed by any machine.

On the other hand, the perspective of reductive materialism towards intelligence assumes that consciousness is a valid concept; we do have a mental life, but mental states can be explained for in terms of the laws of physics. Hence, from a position of reductive materialism, we might argue that we can account for intelligence, emotions and consciousness within a physicalist framework - for instance by reference to neurology and bio-chemical processes. As such, we are (nothing but) nice machines ourselves; or as phrased by Marvin Minsky: "The brain is just a computer made out of meat!" [3].

The ethical implications of these positions will be discussed from a phenomenological approach. Thus, one might ask what kind of ethical issues we are faced with if robots in the future come to *look* like us (sec. 2.1) or *be* like us (sec. 2.2)?

2.1 Ethical Issues in Relation to a Behaviorist Approach

In a behaviorist framework, what can be said to characterize the kind of relations involved in human-robot interactions? Here, we are dealing with a "look-alike setting", in which mental states are considered unnecessary. The robot's behavior is all that counts. Similarly, if we were to deal with our human existence from a behaviorist perspective, we should only be interested in accounting for human actions with reference to complex stimuli-response patterns. In holding a pure behaviorist point of view, presupposing a symmetric relation between human and robot, ethical issues regarding human-robot interactions might be addressed within a utilitarian

[1] There are of course relevant ethical related issues regarding agency and responsibility in a legal context, which I do not touch upon. These issues are discussed in an excellent paper by Ugo Pagallo [1]

framework, in which consequences of behavior could be accounted for ethically by measuring which behavior gave rise to the greatest amount of welfare.

On the other hand, in arguing from a phenomenological position, we might maintain an asymmetric relation between the robot and ourselves, in which case the robot only looks like us, implying that even though interaction is smooth, the robot is simply a machine good at producing certain kinds of behavior, without any intentions behind it. Within this scenario, we can explore what is ethically at stake in human relationships, and discuss whether this can be carried over to human-robot interaction. Thus, it is generally acknowledged that trust is vital for the flourishing of human life, and a precondition of any cultural ordering[2]. Our fundamental human condition is rooted in the fact that life is interpersonal; we are mutually dependent on each other. Consequently, openness, in the sense of trusting, i.e., daring to risk ourselves in coming forward to meet the other, is a definitive feature of human co-existence and inherent in all communication [4]. When we place trust in others, it involves genuine risk-taking since we surrender ourselves to the other. But, in dealing with human-robot interaction, we are not faced with having to surrender ourselves to the social robot. Even though the robot act in a human-like way, and displays emotions, there is nothing at stake, and I know that this is the case about our relationship – the robot simulates and I invest without cost. This does not necessarily imply that I will be unable to respond emotionally to the robot. However, our interaction will be risk-free and without demand.

2.2 Ethical Issues from a Position of Reductive Materialism

In a physicalist framework, matters appear differently. Here, we assume that our mental states are programmable, which will enable us to develop a robot, who do not only simulate intelligence, but has a mind of its own, probably even different from the human mind. In this scenario, the above-mentioned phenomenological objections do not count, because now robots and human beings are on equal footing. If we for a moment leave aside the fact that should it ever turn out to be the case that physicalists came up with an artificial intelligence with a mind of its own, then, in general, the phenomenological position would suffer severe problems. However, for the sake of argument, I shall maintain a phenomenological perspective in the exploration of ethical issues.

Well aware that the robot might develop a mind radically different from ours, we would still have to address design issues in the first place, such as: should we set out to create a robot capable of feeling pain? Normally, we consider it morally wrong to cause somebody pain. Yet, we might argue that lack of ability to feel pain would reduce quality of life considerable for the robot, and maybe even make the robot unable to act emphatic towards others. Nevertheless, as human beings we use different kinds of enhancers to improve our life, so why not set out to design a robot in a state of permanent happiness? One objection could be that lack of challenge in life would probably make the robot unable to fulfill its potentials. But, we would not

[2] See for instance Løgstrup [4], Rawls [5: 433], Fukuyama [6: 126], also within the field of economics, it is commonly known that trust is in general regarded as a "critical commodity".

be able to take that for granted, since we might not recognize the kind of psychological developmental path the robot would follow. As such, the robot might evolve into a being entirely different from us, and demand ethical rights, which would be incomprehensible to us. Within this context of argument, we are faced with a God-like responsibility and ethical obligations towards a creature, who possible will turn out to be beyond our imagination.

3 Concluding Remarks

This paper has dealt with ethical implications related to human-robot interaction within two scenarios of artificial intelligence, of which the first is already on its way, whereas the second scenario is probably not realizable within the nearest future.

Thus, we are approaching a time in which human-like robots (capable of intelligent behavior within restricted contexts) will be able to provide us with reliable companionship. But, here we are dealing with risk-free relations without demands. Human-robot interaction will lack the kind of basic commitment, which stems from the fact that life is interpersonal. We live in a state of surrender to each other, implying that trust is a fundamental human condition, which we cannot escape. Placing trust in others thus involves genuine risk-taking, in the form of surrendering-ourselves-to-others. This is the fundamental nerve of all interpersonal interaction, which a human-robot relationship will not have.

In the second scenario, focus is on the possibility of developing robots with a mental life of their own. Here, we find ourselves faced with a God-like responsibility in deciding what kind of design we should implement. Furthermore, we might be unable to understand the robot, since it might turn out to develop a mind radically different from the human mind, and maybe even demand ethical rights of its own.

References

1. Pagallo, U.: The Human Master with a Modern Slave? Some Remarks on Robotics, Ethics, and the Law. In: Proceedings of the 11th International ETHICOMP Conference, pp.397--410. University of Rovira i Virgili, Tarragona (2010)
2. Turing, A.: Computing Machinery and Intelligence. Mind, 59, 433--460 (1950)
3. Minsky, M.: The Society of Mind. Simon and Schuster, New York (1988)
4. Løgstrup, K.E.: The Ethical Demand. University of Notre Dame Press, Notre Dame (1997)
5. Rawls, J.: A Theory of Justice - Revised Edition. Oxford University Press, Oxford (1999)
6. Fukuyama, F.: Our Posthuman Future - Consequences of the Biotechnology Revolution. Picador, New York (2003)