How can we program ethical robots if we reach a standard of robot ethics? *

Moad HANI

Luxembourg University
moad.hani.001@student.uni.lu
19-12-2020

Abstract. A robot must not endanger a human. It was written in order to "avoid social problems that could arise from inadequate social and legal measures taken to regulate the existence of robots in society". It is composed of three parts devoted to manufacturing standards, the rights and duties of users-owners and the rights and duties of robots. It seems that Asimov also inspired the work of a group of two hundred and twenty-five European industrialists and academics (EURON), who drafted a European Union Convention on Roboethics in 2006. This text defines the elements that constitute the minimum basis for an international quality standard, applicable to all robots, whatever their type (domestic, medical, military, educational entertainment). It establishes the principle that European standards must be set in terms of safety, security, traceability, identification and confidentiality . The ethical approach seems to be more concerned with ensuring human control over robots by laying down principles for the protection of the human species, as if it were necessary to anticipate an uprising of the machines. This is not a war and robots are not our enemies. Claiming this right has nothing to do with a confrontation. Robots and humans will soon share the same world. It is time to think about how...If ethical reflection has its place, especially concerning humanoid or animal-like robotics, it is not enough because it should be based on clear and transparent standards that will govern the world of robots and roboticists in order to allow the creation of robots and the fulfillment of humans far from the fear, the feeling of rejection or the surprise effect evoked by the idea of being served by an ethical robot by individuals and the endless debates of researchers. This work has the duty to highlight ways to create an ethical standard with examples and reflections based on multiple research and the conference paper of Alan Winfield called "The Ethical Roboticist: a journey from robot ethics to ethical robots" in order to consolidate the relationship between roboticists, society and of course robots with cognitive intelligence and ethical behavior.

Keywords: Decision-Making · Roboethics · Ethical Governance framework · Roboticists · Ethical robots · Artificial Moral Agent (AMA) · Cognitive and responsible AI

* Supported by Dr. Amro Najjar and Dr. Sana NOUZRI

1 Introduction

1.1 Overview

The term "roboethics" is used to define "an approach that seeks to determine the appropriate behaviour that human beings should adopt in their relations with robots and other artificial agents". It is to be distinguished from a "robotic ethics", the purpose of which is to teach robots right and wrong or, in other words, to develop moral rules for artificial agents.

In his book "Faire la morale aux robots: une introduction à l'éthique des algorithmes", Martin Gibert highlights the role of programming in the ethics of robots, promoting the expression "ethics of algorithms". He discusses the right way to program robots and the rules to be implemented in their algorithms. In the introduction, Gibert highlights the difference between the ethics of algorithms and the ethics of artificial intelligence (AI) :

– "The latter is the branch of technology ethics that evaluates AI systems and asks if they are good for us. Should we develop autonomous cars or military robots? ...] AI ethics is about our behaviors and choices. Algorithm ethics, on the other hand, deals with the "choices" of machines, which are of course, above all, those of programmers. Which moral principles should be implanted in autonomous cars, military robots, etc ?

The partial autonomy of robots would justify this distinction between the ethics of artificial intelligence and the ethics of algorithms. Before interacting with the world, AIs must be morally programmed, in anticipation of any new situations they may be confronted with. According to Gibert, moral issues make the programming of robots, which become artificial moral agents (AMAs), particularly complex.

Robotic machines blur the boundaries between human subjects and technological objects. Thus, not only do they have societal implications that need to be assessed from an ethical perspective, but they also raise questions about certain essential notions of ethics, such as "capacity to act" and "responsibility", and about existing value frameworks. Given the increased autonomy of robots, the question arises as to who exactly should bear the ethical and/or legal responsibility for robot behaviour.

Robotic technologies have been used for industrial and military purposes since the middle of the 20th century. Since then, these technologies have been increasingly applied in other fields, such as transportation, health care, education, agriculture and household contexts. Robotic technologies take many forms, ranging from the rapidly developing concept of autonomous vehicles to the use of robots in surgery, therapeutic care, care of the elderly, and educational tools for children.

Today, robotics is increasingly based on artificial intelligence (AI), which enables the development of human-like abilities such as perception, language use, interaction, problem solving, learning and even creativity. The main characteristic of these "cognitive machines" is that their decisions are unpredictable and their actions depend on experience and certain stochastic conditions. They are very different from robots whose behavior is determined by a program (deterministic robots). The question of the responsibility for the actions performed by cognitive robots is therefore crucial.

The presence of cognitive robots in society, which is rapidly increasing, raises new challenges. These robots have an impact on human behaviour, induce certain social and cultural changes, and raise questions about security, privacy and the protection of human dignity.

2 Ethical Standard

As mentioned in the general introduction of this report, the first robots were programmed to perform clearly defined tasks. These robots can be described as deterministic robots because their actions are controlled by a series of algorithms whose sequence is predictable. With advances in computer science, the concept of artificial intelligence began to be applied to computers. Although this notion is subject to different interpretations, some authors go so far as to consider that it implies a real intelligence. Machines based on artificial intelligence have sensory, linguistic and interactive capacities comparable to those of human beings. Moreover, these machines exhibit learning capabilities that are gradually being improved - and will continue to be improved - using deep learning techniques. These developments pave the way for what is known as cognitive robotics. This is a technological field aimed at developing robots based on cognitive computation systems and therefore capable of learning from experience, not only by interacting with human trainers but by their own means, thus developing the ability to react to their environment on the basis of what they have learned. Compared to classical or deterministic robots, cognitive robots are capable of

making decisions in complex situations, decisions which cannot be predicted by a programmer.

2.1 Relevant values and ethical principles

Given the diversity and complexity of robots, establishing a framework of values and ethical principles can be useful for defining regulation at all levels – design, manufacture and use - and in a consistent manner, in a form ranging from codes of conduct for engineers to national legislation and international conventions. The principle of human responsibility is the common thread linking the different values mentioned below.

- **Human dignity:** Dignity is intrinsic to human beings but not to machines and robots. It is not possible, therefore, to equate robots with human beings, even in the case of android robots with the seductive appearance of a human being or powerful cognitive robots with learning capacities superior to individual human cognitive capacities. Robots are not human beings; they are the product of human creativity and require a technical support system to be effective and remain efficient tools or mediators.
- **Value of autonomy:** The recognition of human dignity implies that autonomy, as a value, is not only about respect for individual autonomy, which can go as far as refusal of any form of dependency on a robot. Autonomy expresses as a value the recognition of the interdependence of relations between human beings, between human beings and animals, and between human beings and the environment. To what extent will social robots be able to enrich these relationships, or reduce and standardise them? This needs to be assessed scientifically in the context of medical and educational practices where robots may be used, especially when vulnerable groups such as children and the elderly are concerned. The systematic use of robots could aggravate the breakdown of social ties in some societies.
- **Privacy Policy:** Various protection systems and regulations have been introduced in many countries to limit access to personal data and protect the privacy of individuals. However, the development of mega-data is changing the way in which data are collected and processed (use of profiling algorithms). The scale of collection and processing practices is growing enormously and the uses of data are multiplying (e.g. in the fields of commerce, national security, public surveillance and research), as well as the forms of intrusion. Robots are devices capable of collecting data using sensors and exploiting mega-data through deep learning. Therefore, the collection and use of data are aspects to be closely examined in the design of robots, balancing the purpose of the robot against the protection of the privacy of individuals. Some categories of data are more sensitive than others; therefore, various approaches - legislation, professional regulation, governance, public oversight - need to be considered together to maintain public confidence in robots and their proper use.

- **Principle of responsibility:** With regard to the use of robots, "the idea that individuals (or companies, or states) are morally responsible for the choices they make" is extremely important. It is because human beings are responsible for their actions that they can be sued and legally held responsible when their actions result in harm or damage. From this point of view, human ethical responsibility cannot be delegated to robots. 19. In the development of robotics, three aspects need to be clarified and specified in this respect: **civil responsibility**, **transparency** and **accountability**. The issue of civil liability is particularly important in the case of vehicles with a certain degree of automation. The introduction of positioning and registration systems could help to establish this type of liability but may pose problems from the point of view of the right to privacy and the protection of personal data.
- **Value of Charity:** Robots are useful for improving safety, efficiency and performance in the performance of many physically challenging human tasks. Industrial robots, disaster search robots and mine clearance robots can be used to replace human beings in hazardous environments. However, the question of the benevolence of robots designed to intervene in a social context such as education, health care, surveillance or state policing needs further discussion and reflection. The benevolence of a robot, as a value, must be assessed using the principle of proportionality, taking into account existing technological choices.
- **Value of justice:** Justice, as a value, must lead to an examination of the question of inequality. The systematic use of industrial and service robots will contribute to increasing unemployment in certain segments of the labour force. If this is not compensated in some way, or if it is not possible to provide work for those concerned or to organize the workplace in a different way, the increase in inequalities in society will remain a cause for concern. Indeed, work remains an essential element of individual identity and social recognition. Justice, as a value, also requires addressing the issue of non-discrimination. Roboticists should be made aware of the risk of reproducing sexist prejudices and gender stereotypes in the design of robots. The risk of discrimination and stigmatization through the exploitation of data collected by robots is not a trivial issue. States should take adequate measures in this regard.

3 Ethical Robots

Could we make a moral machine? Could we build a robot capable of deciding or moderating its actions on the basis of ethical rules? Three years ago Alan Winfield thought the idea impossible, but he changed his mind. So, what brought about this u-turn?

First he was thinking about simple ethical behaviors. So imagine someone not looking where they're going. Maybe looking at their smartphone, about to walk into a hole in the ground. You will probably intervene. Now, why is that?

It's not just because you're a good person. It's because you have the cognitive machinery to predict the consequences of their actions. Now imagine it's not you but a robot, and the robot has four possible next actions. So, from the robot's perspective it could stand still or turn to its left, and that the human will come to harm, will fall in the hole.

But if the robot could predict the consequences of both its and the human's action, then another possibility opens up. It could choose to collide with the human to prevent them from falling in the hole. And if we express this is an ethical rule, which you see here, this looks remarkably like Asimov's First Law or Robotics, which is that a robot must not injure a human or through inaction cause a human to come to harm. So thus emerged the idea that they could build an Asimovian robot. They equiped the robot with the ability to predict the consequences of both its own actions and others' in its environment, plus the ethical rule that shown in the previous slide. In fact, the technology that we need to do this exists and it's called the robot simulator. So, roboticists use robot simulators all the time to model and test our robot code in a virtual world before running that code on the real robot. But the idea of putting a robot simulator inside a robot, well, it's not a new idea but it's tricky and very few people have pulled it off. In fact, it takes a bit of getting your head round. The robot needs to have, inside itself, a simulation of itself and its environment, and others in its environment. And running in real-time as well.

Three robots standing in a room, with one inside a bounded-off "danger zone" So, over the past two years they've actually tested these ideas with real robots. In fact, these are the robots. Instead of a hole in the ground, they have a danger zone. And they use robots instead of humans. they use robots as proxy humans.

Here we have a the blue robot, the ethical robot, is heading towards a destination. This is its goal. But it notices right here that the red robot, the human, is heading toward danger. So the blue robot chooses to divert from its path to collide (gentle collision) with the human, to prevent it from coming from harm.

This is exactly the same thing but a short movie clip. You can see again, the blue robot is the ethical robot. Our red robot is the proxy human.

So, they also tested the same with an ethical dilemma. Here their ethical robot is faced with two humans heading toward danger. It rather dithers, rather hesitant, and of course it cannot save them both. There isn't time. Ethical dilemmas are a problem really for ethicists not roboticists.

So, how ethical is our ethical robot? Our robot implements a form of consequentialist ethics. In fact, we call the internal model a consequence engine. The robot behaves ethically not because it chooses to, but because it's programmed to do so. We call it an ethical zombie. This approach has a huge advantage, which is that the internal process of making ethical decisions is completely transparent. So if something goes wrong, then we can replay what the robot was thinking. Transparency is a must and it's going to be really important in the future, that autonomous robots will need the equivalent of a flight data recorder in aircraft. An ethical black box.

4 Robot ethics – Why and how roboticists must be ethical ?

4.1 An Ethical Governance framework for robotics and AI - Specific recommendations on the ethics of robotics

- **Recommendation on the development of codes of ethics for robotics and roboticists:** It is recommended to pursue, both at national and international level and according to multidisciplinary modalities, the development, implementation, revision and updating of codes of ethics for roboticists, taking into account possible future advances in robotics and their impact on human life and the environment (energy, electronic waste, ecological footprint). It is also recommended that the disciplines and professions which contribute significantly to robotics or can make use of it - in particular electronic engineering, artificial intelligence, medicine, animal sciences, psychology and physical sciences - review, preferably in a coordinated manner, their own codes of ethics, seeking to anticipate the challenges which may result from their links with robotics and the robot industry. Finally, it is recommended that ethics - including codes of ethics, codes of conduct, and other relevant documents – becomes an integral part of the curriculum for all professionals involved in the design and manufacture of robots.
- **Recommendation on the ethical design of robots:** Ethical considerations should be taken into account in the design of robotic technologies. Robots use algorithms that embody values and ethical frameworks for decision making. Furthermore, the introduction of robots has ethical implications for the practices concerned, whether in health care, education or social relations. In order to effectively take into account the ethical dimension of robots, it is necessary that the development process of these machines gives a place to ethics, based on an approach such as "Value Sensitive Design". This approach should also be adapted to take into account animal welfare.

- **Recommendation on experimentation:** The social implications of new robotic technologies are often difficult to predict. To ensure the responsible use of robots in society, careful and transparent experimentation is necessary. The prior introduction of new robotic technologies on a small scale, in controlled situations, should make it possible to study openly their impact on human practices and activities, as well as on values and interpretative frameworks. The results of such experiments could then be used to adapt robot design, inform policy and regulatory development, and provide users with a critical perspective on the use of these technologies.
- **Recommendation on public discussion:** The introduction of robots will have profound effects on society and the daily life of individuals. In order to deal with these effects in a responsible way, citizens will need adequate concepts, knowledge and frameworks for interpretation. To this end, public discussions should be organised on the implications of new robotic technologies with regard to various aspects of social and daily life, including the environmental impact on the whole robot production cycle. Such discussions would help individuals to develop a reflexive attitude on this subject and would contribute to raising awareness among designers and decision-makers.
- **Recommendation on retraining and rehabilitation of the labour force:** Robots will have to take the place of human beings in a wide range of areas. They will lead to a significant reduction in employment opportunities and will also create new jobs in some sectors. States, professional organizations and educational institutions should therefore reflect on the consequences of these developments, paying particular attention to those sectors of society most likely to be vulnerable to change, and take appropriate measures for the retraining and rehabilitation of the force to enable the realization of potential benefits.
- **Recommendation on gender equality:** Particular attention should be paid to gender issues and stereotypes concerning all types of robots described in this report, and in particular toy robots, sex partners and work replacements.
- **Recommendation on environmental impact assessment:** As with other advanced technologies, environmental impact should be considered as part of the life cycle analysis, allowing a more global assessment of whether a specific use of robotics will give society more benefits than harm. This should take into account the possible negative impacts of production, consumption and waste (e.g. mining of rare earths, e-waste, energy consumption), as well as the potential environmental benefits. When building robots (nano, micro or macro), efforts should be made to use degradable materials and environmentally friendly technologies, as well as to improve the recycling of materials.
- **Recommendations on the Internet of Things:** The Internet of Things (IoT) is a rapidly developing technology that enables the interconnection of intelligent everyday devices, including household appliances. It makes it possible to use devices of various types as sensors to collect massive data that can be used for many purposes. These networks connecting everyone

and everything create entirely new possibilities in many areas such as augmented reality, tactile interconnection, distributed manufacturing and smart cities. IoT clearly raises some ethical issues, particularly with respect to privacy and security, but the next generation of IoT, sometimes referred to as IoT++, presents even more serious challenges. IoT++ will rely on artificial intelligence (AI) to process the data collected. The cognitive algorithms used, which will act as autonomous learners, may lead to unpredictable results. Similarly, emerging technologies are creating small1 robots that can act as mobile sensors, collecting information in targeted locations - extending the reach of IoTs even beyond existing 'things'.

5 Discussion and Q/A

5.1 The key element in reaching an "Ethical Governance Standard" is the importance of humans in decision-making process

Robot-to-robot or even robot-to-human combat will force humans to rethink how they will deal with conflict and standards of ethical robots. Today, the debate is raging between philosophers, academics and roboticists, but also by non-profit organisations, the military and international organisations such as NATO and the United Nations on the importance, legality and ethics that should be considered for implementation of any ethical robot. The interest of that is complicated by the often changing use of definitions by key stakeholders in the debates. The definition of a robot remains contested by some countries and the variations in what is to be considered autonomous or not is only increasing. But what makes the debates even more difficult and uncertain is most certainly the interpretation of current laws and the understanding of how they should be applied. From here come the fact that the active participation of the human in the decision-making process must be maintained.

5.2 Q/A related to Prof. Alan Winfield experimentation and research in building ethical robots

1. What was Dr. Alan Winfield's main motivation behind all this work ?
   - Ensuring safety in an unpredictable environment
2. What will be the next step to make these robots more reliable ?
   - Since they can make decisions, the next step will be to memorize them.
3. What is the main objective of the internal model ? What can you say about its control ?
   - The main objective is to allow the robot to decide which actions not to take (and not the inverse) knowing that the controller has all possible actions and will choose among them the least harmful ones after the internal model will point out the ones to avoid. This internal model does not directly control the robot (take a close look at the following architecture).

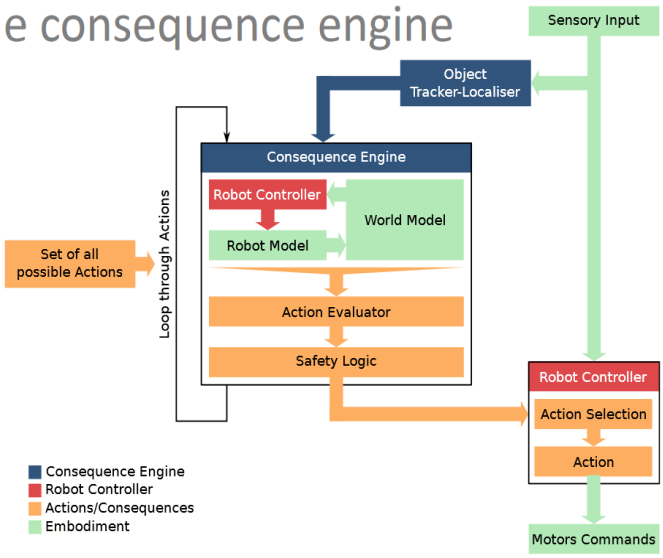


Diagram from Blum et al, *Frontiers in Robotics and AI*, 2018.

4. How does the A-robot know the configuration change in its environment? How can this be implemented in real life ?
   - This is a very interesting question, in fact it is kept aware of the change thanks to the GPS Tracker. Even if we're allowed to think it's a "kind of cheating" (since they manipulated the experiment), in real life we can always meet the same configuration with ease using "computer vision detection and 3D simulations" to help the A-robot estimate how far is the human being.
5. Can an ethical robot become unethical ?
   - Here we should know that this is (according to the same authors in The dark side of ethical robots a very big challenge since the "ethical brain" itself remains hack-able. In fact, adding a simple "-" instead of "+" in the desirability function is sufficient to change the robot's behavior and make an aggressive robot.

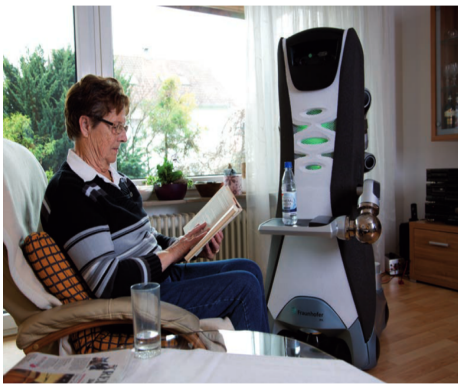
6 Conclusion

Humans must be part of the decision-making process in order to be able to consider the AARs - for instance - as legal and ethical in the future. Once employment and mass production has started, many meetings to establish the framework for employment will be necessary and should involve the leaders of the 64 nations. The need to discuss this subject is, and will continue to be, of paramount importance to world leaders, politically, legally, ethically and philosophically.

The aspect of discrimination, transparency, proportionality, military necessity, responsibility, and accountability are all areas of great moral and legal concern in the debates. These concerns are often based on past experiences, using

semi-automatic or relatively autonomous technologies, that have gone wrong. The majority of these accidents, which employed a varying level of human intervention, reflect on the importance of the human in the robot's decision-making process.

Once again, in order to avoid too many errors of unwanted gestures, the importance of the human in the decision-making process must be maintained. In conclusion, society may not be facing extermination by the non-ethical robots in the near future, as Hollywood and the multitude of science fiction films would have us believe. However, it is clear that the introduction of fully autonomous robots is inevitable in the future. Society must ensure that the most dangerous aspects of this technology are defeated by the absolute retention of a human in the decision-making process until it is demonstrated that a **"fully ethical robot"** can be as, if not better than humans in law enforcement and that it can apply an infallible ethical sense.



References

1. Alliance des sciences et technologies de numérique (Allistene). 2016. Éthique de la recherche en robotique [Ethics of research in robotics]. Paris, Allistene.
2. Allen, C., Wallach, W. and Smit, I. 2011. Why machine ethics. In: Anderson, M. and Anderson, S. L. eds.Machine Ethics. Cambridge, Cambridge University Press, pp. 51-61.
3. Anderson, M. and Anderson, S. L. 2011a. General introduction. In: Anderson, M. and Anderson, S. L. eds. Machine Ethics. Cambridge, Cambridge University Press, pp. 1-4.
4. Hughes, J. 2012. Compassionate AI and selfless robots: a Buddhist approach. In: Lin, P.,Abney, K. and Bekey, G. A. eds.Robot Ethics: The Ethical and Social Implications of Robotics. London, MIT Press, pp. 69-84.
5. International Federation of Robotics (IFR). 2016a. Executive Summary World Robotics 2016: Industrial Robots. Frankfurt, IFR. Available here: PDF
6. Ingram, B., Jones, D., Lewis, A., Richards, M., Rich, C. and Schachterle, L. 2010. A code of ethics for robotics engineers, Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp.103-104
7. Winfield, Alan. (2018). The Ethical Roboticist: a journey from robot ethics to ethical robots.