Gianmarco Veruggio, Fiorella Operto:

Roboethics: a Bottom-up Interdisciplinary Discourse in the Field of Applied Ethics in Robotics

Abstract:

This paper deals with the birth of Roboethics. Roboethics is the ethics inspiring the design, development and employment of Intelligent Machines. Roboethics shares many 'sensitive areas' with Computer Ethics, Information Ethics and Bioethics. It investigates the social and ethical problems due to the effects of the Second and Third Industrial Revolutions in the Humans/Machines interaction's domain. Urged by the responsibilities involved in their professions, an increasing number of roboticists from all over the world have started - in cross-cultural collaboration with scholars of Humanities – to thoroughly develop the Roboethics, the applied ethics that should inspire the design, manufacturing and use of robots. The result is the Roboethics Roadmap.

Agenda

Authors:

Gianmarco Veruggio:

- CNR, National Research Council, Via De Marini, 6 -- 16149 Genova, Italy
- ☎ +39 (0) 10 64 75 625 , ✉ gianmarco@veruggio.it

Fiorella Operto:

- School of Robotics, C.P. 4124, piazza Monastero, 4, 16149 Genova, Italy
- ☎ +39 (0) 348 09 61 616, ✉ operto@scuoladirobotica.it, 💻 http://www.scuoladirobotica.it/

ISSN 1614-1687

Gianmarco Veruggio, Fiorella Operto:

Roboethics: a Bottom-up Interdisciplinary Discourse in the Field of Applied Ethics in Robotics

Introduction

Robotics is rapidly becoming one of the leading fields of science and technology. Figures released by IFIR/UNECE Report 2004 show the double digit increasing in many subsectors of Robotics as one of the most developing technological fie4ld. We can forecast that in the XXI century humanity will coexist with the first alien intelligence we have ever come into contact with - *robots*.

All these developments have important social, ethical, and economic effects. As for other technologies and applications of scientific discoveries, the public is already asking questions such as: “Could a robot do "good" and "evil”? “Could robots be dangerous for humankind?”.

Like Nuclear Physics, Chemistry or Bioengineering, soon also Robotics could be placed under scrutiny from an ethical standpoint by the public and Public Institutions (Governments, Ethics Committees, Supranational Institutions).

Feeling the responsibilities involved in their practices, an increasing number of roboticists from all over the world, in cross-cultural collaboration with scholars of Humanities, have started deep discussions aimed to lay down the Roboethics, the ethics that should inspire the design, manufacturing and use of robots.

Robotics and Ethics

Is Robotics a new science, or is it a branch or a field of application of Engineering? Actually Robotics is a discipline born from Mechanics, Physics/Mathematics, Automation and Control, Electronics, Computer Science, Cybernetics and Artificial Intelligence. Robotics is a unique combination of many scientific disciplines, whose fields of applications are broadening more and more, according to the scientific and technological achievements.

Specificity of Robotics

It is the first time in history that humanity is approaching the challenge to replicate an intelligent and autonomous entity. This compels the scientific community to examine closely the very concept of intelligence – in humans, animals, and of the mechanical – from a cybernetic standpoint.

In fact, complex concepts like autonomy, learning, consciousness, evaluation, free will, decision making, freedom, emotions, and many others shall be analysed, taking into account that the same concept shall not have, in humans, animals, and machines, the same semantic meaning.

From this standpoint, it can be seen as natural and necessary that Robotics drew on several other disciplines, like Logic, Linguistics, Neuroscience, Psychology, Biology, Physiology, Philosophy, Literature, Natural History, Anthropology, Art, Design.

Robotics de facto combines the so called two cultures, Science and Humanities.

The effort to design Roboethics should take into account this specificity. This means that experts shall consider Robotics as a whole - in spite of the current early stage which recalls a melting pot – so they can achieve the vision of the Robotics’ future.

From Myth to Science Fiction

The issue of the relationship between humankind and autonomous machines – or, automata - appeared early in world literature, developed firstly through legends and myths, more recently by scientific and moral essays. The topic of the rebellions of automata recurs in the classic European literature, as well as the misuse or the evil use of the product of ingenuity. It is not so in all the world cultures: for instance, the mythology of the Japanese cultures does not include such paradigm. On the contrary, machines (and, in general, human products) are always beneficial and friendly to humanity. This difference in seeing the machines is a subject we should take into account and analyse.

Some common questions:

- How far can we go in embodying ethics in a robot?
- Which kind of “ethics” is a robotics one?
- How contradictory is, on one side, the need to implement in robots an ethics, and, on the other, the development of robot’s autonomy?
- Although far-sighting and forewarning, could Asimov’s Three Laws become really the Ethics of Robots?

ISSN 1614-1687

- Is it right to talk about "consciousness", "emotions", "personality" of Robots?

What is a Robot?

Robotics scientists, researchers, and the general public have about robots different evaluations, which should taken into account in the Roboethics Roadmap.

Robots are nothing but machines. Many consider robots as mere machines - very sophisticated and helpful ones - but always machines. According to this view, robots do not have any hierarchically higher characteristics, nor will they be provided with consciousness, free will, or with the level of autonomy superior to that embodied by the designer. In this frame, Roboethics can be compared to an Engineering Applied Ethics.

Robots have ethical dimensions. In this view, an ethical dimension is intrinsic within robots. This derives from a conception according to which technology is not an addition to man but is, in fact, one of the ways in which mankind distinguishes itself from animals. So that, as language, and computers, but even more, humanoids robots are symbolic devices designed by humanity to improve its capacity of reproducing itself, and to act with charity and good. (J. M. Galvan)

Robots as moral agents. Artificial agents particularly but not only those in Cyberspace, extend the class of entities that can be involved in moral situations. For they can be conceived as moral patients (as entities that can be acted upon for good or evil) and also as moral agents (not necessarily exhibiting free will, mental states or responsibility, but as entities that can perform actions, again for good or evil). This complements the more traditional approach, common at least since Montaigne and Descartes, which considers whether or not (artificial) agents have mental states, feelings, emotions and so on. By focusing directly on 'mind-less morality' we are able to avoid that question and also many of the concerns of Artificial Intelligence. (L. Floridi)

Robots, evolution of a new specie. According to this point of view, not only will our robotics machines have autonomy and consciences, but humanity will create machines that exceed us in the moral as well as the intellectual dimensions. Robots, with their rational mind and unshaken morality, will be the new species: Our machines will be better than us, and we will be better for having created them. (J. Storrs Hall)

The Birth of Roboethics

The name Roboethics was officially proposed during the First International Symposium of Roboethics (Sanremo, Jan/Feb. 2004), and rapidly showed its potential. Philosophers, jurists, sociologists, anthropologist and moralists, together with robotic scientists, were called to contribute to lay the foundations of the Ethics in the designing, developing and employing robots.

Main positions on Roboethics

According to the anthropologist Daniela Cerqui, three main ethical positions emerged from the robotics community:

- **Not interested in ethics.** This is the attitude of those who consider that their actions are strictly technical, and do not think they have a social or a moral responsibility in their work.
- **Interested in short-term ethical questions.** This is the attitude of those who express their ethical concern in terms of "good" or "bad," and who refer to some cultural values and social conventions. This attitude includes respecting and helping humans in diverse areas, such as implementing laws or in helping elderly people.
- **Interested in long-term ethical concerns.** This is the attitude of those who express their ethical concern in terms of global, long-term questions: for instance, the "Digital divide" between South and North; or young and elderly. They are aware of the gap between industrialized and poor countries, and wonder whether the former should not change their way of developing robotics in order to be more useful to the latter.

Disciplines involved in Roboethics

The design of Roboethics requires the combined commitment of experts of several disciplines, who, working in transnational projects, committees, commissions, have to adjust laws and regulations to the problems resulting from the scientific and technological achievements in Robotics.

In all likelihood, we will witness the birth of new curricula studiorum and specialities, necessary to manage a subject so complex, juts as it happened with Forensic Medicine.

In particular, we mention the following fields as the main to be involved in Roboethics: Robotics, Computer Science, Artificial Intelligence, Philosophy, Ethics, Theology, Biology, Physiology, Cognitive Sciences, Neurosciences, Law, Sociology, Psychology, Industrial Design.

The EURON Roboethics Atelier

EURON is the European Robotics Research Network, aiming to promote excellence in robotics by creating resources and exchanging the knowledge we already have, and by looking to the future.

One major product of EURON is a robotics research roadmap designed to clarify opportunities for developing and employing advanced robot technology over the next 20 years. The document provides a comprehensive review of state of the art robotics and identifies the major obstacles to progress.

The main goals of the roadmapping activity are to identify the current driving forces, objectives, bottlenecks and key challenges for robotics research, so as to develop a focus and a draft timetable for robotics research in the next 20 years.

The Roboethics Atelier

In 2005, EURON funded the Roboethics Atelier Project, coordinated by Scuola di Robotica, with the aim of designing the first Roboethics Roadmap.

Once the profile of the Euron Roadmap project had been discussed and its frame identified, the selection of participants started. This was done on the basis of: a) their participation to previous activities on Techno/Roboethics, b) their cross-cultural attitude, c) their interest in applied ethics.

The last step in the process involved a series of discussions via e-mail which led to the definition of the Programme. Participants were asked to prepare a major contribution on their area of expertise, and on a few more on topics they were interested to discuss, even outside their realm of expertise. The organizers promoted the cross-cultural and transdisciplinary contributions.

The Roboethics Roadmap

The Roboethics Roadmap outlines the multiple pathways for research and exploration in the field and indicates how they might be developed. The roadmap embodies the contributions of many scientists and technologists, in several fields of investigations from sciences and humanities. This study hopefully is a useful tool in view of cultural, religious and ethical differences.

Let's see firstly what the Roboethics Roadmap cannot be:

- It is not a Survey, nor a State-of-the-Art of the disciplines involved. This Roadmap does not aim to offer an exhaustive picture of the State-of-the-Art in Robotics, nor a guideline of ethics in science and technology. The reason is that: a) Robotics is a new science still in the defining stage. It is in its blossoming phase, taking different roads according to the dominant field of science undertaken (field Robotics, Humanoids, Biorobotics, and so on). Almost every day we are confronted with new developments, fields of applications and synergies with other sectors; b) Public and private professional associations and networks such as IFR-International Federation of Robotics, IEEE Robotics and Automation Society, EUROP - European Robotics Platform, Star Publishing House, have undertaken projects to map the State-of-the-Art in Robotics.
- It is not a list of Questions & Answers. Actually, there are no easy answers, and the complex fields require careful consideration.
- It is not a Declaration of Principles. The Euron Roboethics Atelier, and the sideline discussion undertaken, cannot be regarded as the institutional committee of scientists and experts entitled to draw a Declaration of Principles on Roboethics.

The ultimate purpose of the Euron Roboethics Atelier, and of the Roboethics Roadmap is to provide a systematic assessment of the ethical issues involved in the Robotics R&D; to increase the understanding of the problems at stake, and to promote further study and transdisciplinary research [9].

Scope: Near Future Urgency

In terms of scope, we have taken into consideration – from the point of view of the ethical issue connected to Robotics – a temporal range of a decade, in whose frame we could reasonably locate and infer – on the basis of the current state-of-the-Art in Robotics – certain foreseeable developments in the field.

For this reason, we consider premature – and have only hinted at – problems inherent in the possible emergence of human functions in the robot: like

consciousness, free will, self-consciousness, sense of dignity, emotions, and so on. Consequently, this is why we have not examined problems –debated in literature – like the need not to consider robot as our slaves, or the need to guarantee them the same respect, rights and dignity we owe to human workers.

Target: Human Centred Ethics

Likewise, and for the same reasons, the target of this Roadmap is not the robot and its the artificial ethics, but the human ethics of the robots' designers, manufacturers and users.

Although informed about the issues presented in some papers on the need and possibility to attribute moral values to robots' decisions, and about the chance that in the future robots might be moral entities like – if not more than– human beings, we have chosen, in the first release of he Roboethics Roadmap, to examine the ethical issues of the human beings involved in the design, manufacturing, and use of the robots.

We have felt that problems like those connected to the application of robotics within the military and the possible use of military robots against some populations not provided with this sophisticated technology, as well as problems of terrorism in robotics and problems connected with biorobotics, implantations and augmentation, were urging and serious enough to deserve a focused and tailor-made investigation..

It is absolutely clear that without a deep rooting of Roboethics in society, the premises for the implementation of an artificial ethics in the robots' control systems will be missing.

Methodology: Open Work

The Roboethics Roadmap is an Open Work, a Directory of Topics & Issues, susceptible to further development and improvement which will be defined by events in our technoscientific-ethical future. We are convinced that the different components of society working in Robotics, and the stakeholders in Robotics should intervene in the process of building a Roboethics Roadmap, in a grassroots science experimental case: the Parliaments, Academic Institutions, Research Labs, Public ethics committees, Professional Orders, Industry, Educational systems, the mass-media.

Ethical Issues in an ICT society

Roboethics shares many 'sensitive areas' with Computer Ethics and Information Ethics. But, before that, we have to take into account the global ethical problems derived from the Second a Third Industrial Revolutions, in the field of the relationship between Humans and Machines:

- Dual-use technology (every technology can be used and misused);
- Anthropomorphization of the Machines;
- Humanisation of the Human/Machine relationship (cognitive and affective bonds toward machines);
- Technology Addiction;
- Digital Divide, socio-technological Gap (per ages, social layer, per world areas);
- Fair access to technological resources;
- Effects of technology on the global distribution of wealth and power;
- Environmental impact of technology.

From the Computer and Information Ethics we borrow the known Codes of Ethics called PAPA, acronym of: privacy, accuracy, intellectual property and access.

- Privacy: What information about one's self or one's associations must a person reveal to others, under what conditions and with what safeguards? What things can people keep to themselves and not be forced to reveal to others?
- Accuracy: Who is responsible for the authenticity, fidelity and accuracy of information? Similarly, who is to be held accountable for errors in information and how is the injured party to be made whole?
- Property: Who owns information? What are the just and fair prices for its exchange? Who owns the channels, especially the airways, through which information is transmitted? How should access to this scarce resource be allocated?
- Accessibility: What information does a person or an organization have a right or a privilege to obtain, under what conditions and with what safeguards?

Questions raised on the range of application of sensitive technologies, and on the uncertainty of performance of these are raised in connection to neuro-robotics:

- Under what conditions should we decide that deployment is acceptable?

- At what point in the development of the technology is an increase in deployment acceptable?
- How do we weigh the associated risks against the possible benefits?
- What the rate of the ethics of functional compensation or repair vs. enhancement? This issue is especially notable regarding the problem of augmentation: In some cases a technology is regarded as a way of compensating for some function that is lacking compared to the majority of humans; in other cases, the same technology might be considered an enhancement over and above that which the majority of humans have. Are there cases where such enhancement should be considered unethical?
- Are there cases where a particular technology itself should be considered unacceptable even though it has potential for compensation as well as enhancement?

The question of identifying cause, and assigning responsibility, should some harm result from the deployment of robotic technology. (Wagner, J.J, David M. Cannon, D.M., Van der Loos).

The precautionary principle

Problems of the delegation and accountability to and within technology are daily life problems of every one of us. Today, we give responsibility for crucial aspects of our security, health, life saving, and so on to machines.

Professional are advised to apply, in performing sensitive technologies the precautionary principle:

"When an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause-and-effect relationships are not fully established scientifically."

From the precautionary principle derive some other rules such as: non-instrumentalisation, non-discrimination, informed consent and equity, sense of reciprocity, data protection.

The aim of this roadmap is to open a debate on the ethical basis which should inspire the design and development of robots, to avoid to be forced to become conscious of the ethical basis under the pressure of grievous events. We believe that precaution should not produce paralysis of science and technology.

The Roboethics Taxonomy

A taxonomy of Robotics is not a simple task, simply because the field is in a full bloom. A classification of Robotics is a work in progress, done simultaneously with the development of the discipline itself.

Aware of the classifications produced by the main Robotics organizations, which differ from one another on the basis of the approach – technological/applicational -, we have preferred, in the case of the Roboethics Roadmap, to collect the many Robotics fields from a typological standpoint, according to shared homogeneity of the problems of interface towards the society.

Instead of an encyclopaedic approach, we have followed - with few modifications - the classification of EURON Robotics Research Roadmap [8]. For every field, we have tried to analyze the current situation rather than the imaginable. Thus, we have decided to give priority to issues in applied ethics rather than to theoretical generality. It should be underscored that the Roboethics Roadmap is not exhaustive, and that, by way of discussions and comparing and collating, certainly it can be improved.

The robotics classification is matched with a discussions of the sensitive issues emerging from the application of that specific field, by *Pro's* and *Con's*, and by *Recommendations*.

References

[1] Asimov, I, Runaround, Astounding Science Fiction, March 1942. Republished in Robot Visions by Isaac Asimov, Penguin, 1991

[2] Asimov, I., I Robot, Doubleday, 1950

[3] Capurro, R. (2000), Ethical Challenges of the Information Society in the 21st Century, "International Information & Library Review" 32, 257-276

[4] Floridi, L., Information Ethics: On the Philosophical Foundation of Computer Ethics, Ethicomp98, The Fourth International Conference on Ethical Issues of Information Technology, Erasmus University, The Netherlands, 25/27 March 1998

[5] Floridi, L., Sanders, J. W., On the Morality of Artificial Agents, Information Ethics Groups, University .

[6] Galvan, J.M., On Technoethics, in «IEEE-RAS Magazine» 10 (2003/4) 58-63.

[7] Gips, J., Towards the Ethical Robot, published in Android Epistemology, K. Ford, C. Glymour and Hayes, P., MIT Press, 1995 (http://www.cs.bc.edu/~gips/EthicalRobot.pdf)

[8] EURON Research Roadmap (http://wwwiaim.ira.uka.de/euron/cwiki.php)

[9] ROBOETHICS ROADMAP (http://www.roboethics.org/roadmap)