The EURON Roboethics Roadmap

Gianmarco Veruggio
Scuola di Robotica
Genova, Italy
gianmarco@veruggio.it

Abstract - **This paper deals with the results of the Euron Roboethics Atelier 2006 (Genoa, Italy, Feb.-March. 2006), comprising in the Roboethics Roadmap; and it offers a short overview of the ethical problems involved in the development of the next generation of the humanoid robots.**

Index Terms - Roboethics, Robotics, Ethics.

I. INTRODUCTION

The importance and urgency of Roboethics has been demonstrated by our recent history. Three of the front rank fields of science and technology: Nuclear Physics, Bioengineering, and Computer Science, have already been forced to face the consequences of their ethics and their research's applications because of pressure caused by dramatic and troubling events, or because of the concern of the general public. In many countries, public opinion, shocked by some of these effects, urged to either halt the whole research/applications, or to strictly control them.

Robotics is rapidly becoming one of the leading fields of science and technology: we can forecast that in the XXI century humanity will coexist with the first alien intelligence we have ever come into contact with - robots. It will be an event rich in ethical, social and economic problems. 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, in a few years, Robotics could also 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.

II. 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. This shows that 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.

A. *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 unifies the so called two cultures, Science and Humanities.

The effort to design Roboethics should take care of this specificity. This means that experts shall view 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.

B. *The Three Laws of Robotics*

In 1942, novelist Isaac Asimov formulated, in the novel Runaround, the Three Laws of Robotics:

1. A robot may not injure a human being, or through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the 1st or 2nd Law

Later on he added the 4th Law (known as Law Zero):

4. No robot may harm humanity or, through inaction, allow humanity to come to harm.

The theme 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. 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?
- Is it right that robots can exhibit a "personality"?
- Is it right that robot can express "emotion"?

C. *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)

III. THE BIRTH OF ROBOETHICS

The name Roboethics (coined in 2002 by the author) was officially proposed during the First International Symposium of Roboethics (Sanremo, Jan/Feb. 2004), and rapidly showed its potential, because naming things - according to the *Principle of Composition* - gives them reality.

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.

A. *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.

B. *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.

IV. 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.

A. *The Roboethics Atelier*

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

B. *The Roboethics Roadmap*

The ultimate purpose 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 studying and transdisciplinary research.

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 more than 50 scientists and technologists, in many fields of investigations from sciences and humanities.

This study hopefully is a useful tool in view of cultural, religious and ethical differences.

The Roboethics Roadmap should be considered the number 1 release, a preliminary and non exhaustive taxonomy of sensitive problems of the field.

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.

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.

C. *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).

D. *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.

V. THE ROBOETHICS TAXONOMY

In the period of a year, the Euron Roboethics Atelier has carried out a *tour d'horizon* of the Fields in Robotics: an overview of the state of the art in Robotics, and of the main ethical issues, driven by the most recent techno-scientific developments, which can only just be glimpsed.

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 encyclopedic approach, we have followed - with few modifications - the classification of EURON

Robotics Research Roadmap. 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 underlined that the present grid is not exhaustive; it is the first release of the Roboethics Roadmap, susceptible to be improved, and corrected.

Here below the chapter of the Roadmap devoted to humanoids, as emerged from the contributions of the Participants to the Atelier, and from a broad array of documentation.

VI. HUMANOIDS AND ROBOETHICS

One of the most ambitious aims of Robotics is to design an autonomous robot that could reach - and even surpass - human intelligence and performance in partially unknown, changing, and unpredictable environments.

"Essentially, it is expected that a robot will provide assistance in housework, for aged people and for entertainment to keep up the amenity of life and human environment in the next century. A type of human robot, a Humanoid is expected, to work together with human partners in our living environment, and it will share the same working space and will experience the same thinking and behaviour patterns as a human being. The robot will integrate information from sensors and show coordinated actions which realize a high level of communication with a human without any special training using multimedia such as speech, facial expression and body movement" (source, Waseda Humanoid Robotics Institute)

A. *Artificial Mind*

In the Roadmap, we limited ourselves to defining intelligence from an engineering point of view, that is, an operational intelligence – although we are aware of the fact that our terminology regarding robots' functions is often taken from the language used for human beings.

The Artificial Intelligence shall be able to lead the robot to fulfill the missions required by the end-users. To achieve this goal, over the past decades scientists are working on AI techniques in many fields.

From our point of view, one of the fundamental aspects of the robots is their capability to learn: to learn the characteristics of the surrounding environment, that is, a) the physical environment, but also b) the living beings who inhabit it. This means that robots working in a given environment have to recognize human beings from other objects. In addition to learning about the environment, robots have to learn about their own behavior, through a self reflective process. They have to learn from the experience, replicating somehow the natural processes of the evolution of intelligence in living beings (synthesis procedures, trying-and-error, learning by doing, and so on).

All these processes embodied in the robots produce a kind of intelligent machine endowed with the capability to express a certain degree of autonomy. It follows that a robot can behave , in some situations, in a way which is unpredictable for their human designers.

Basically, the increasing autonomy of the robots could give rise to unpredictable and non predictable behaviours.

So, without necessarily imagining some Sci-Fi scenarios where robot are provided with consciousness, free will and emotions, in a few years we are going to cohabit with robots endowed with self knowledge and autonomy – in the engineering meaning of these words.

B. *Artificial Body*

Humanoids are robots whose body structure resembles the human one. They answer to an old dream of humanity, and certainly do not spring only from rational, engineering or utilitarian motivations, but also from psycho-anthropological ones.

Humanoids are the expression of one of the demands of our European culture, that is that humankind were the creator of some mechanical being of the shape of a human. In the Japanese culture, it is the demand to carefully replicate nature in all its forms.

It is a very difficult and demanding enterprise, a project of the level of the Mission to the Moon. But, precisely for its characteristic of being one of humanity's' dreams, the investments are high and the progress speed very quick.

It has been forecasted that in a not so distant future we will cohabit with humanoids whose shape will be so similar to that of human beings that it will render it possible to get mixed up – in certain situations - with the latter. Humanoids will assists human operators in human environments, will replace human beings, and will cooperate with human beings in many ways.

Given the high cost and the delicacy of the humanoids, they will probably be employed in tasks and in environments where the human shape would really be needed, that is, in all these situations where the human-robot interaction is primary, compared to any other mission - human-robot interactions in health care; children/disable people/elderly assistance; baby sitting; office clerks, museum guides; entertainers, sexual robots, and so on. Or, they will be employed as testimonials for commercial products.

In the frame of this Roadmap, there was no need to closely examine the technological aspects of humanoids (actuators, artificial muscles; robot path planning; visual aspect and the realization of emotion in humanoid robots; expressions of verbal and non verbal information in robots; environment and human recognition; of human faces; human-machine communication interface; and so on). Many of these technologies comes from bio-robotics; and many, born in the humanoids labs, are and will be applied to bio-robotics.

C. *Benefits*

- Intelligent machines can assist humans to perform very difficult tasks, and behave like true and reliable companions in many ways.

- Humanoids are robots so adaptable and flexible that will be rapidly used in many situations and circumstances.
- Their shape, and the sophisticated human-robot interaction, will be very useful for those situation where a human shape is needed.
- Faced with an aging population, the Japanese society see humanoids robots as one way to enable people to continue to lead an active and productive life in their old age, without being a burden to other people.
- The researches carried out in humanoids laboratories over the world will have as a side effects the development of platform to study human body, for training, haptic test and trainings, with extraordinary outcomes on health care, education, edutainment, and so on.

D. *Problems*

- Reliability of the internal evaluation systems of the robots.
- Unpredictability of robots' behaviour.
- Traceability of evaluation/actions procedures.
- Identification of robots.
- Safety. Wrong action can lead to dangerous situation for living beings and the environment.
- Security. In the case the autonomy of the robot were controlled by ill-intentioned people, who can modify the robot's behavior in a dangerous and fraudulent course.

Because humanoids sum up almost all the characteristics of the whole spectrum of robots, their use implies the emergence of nearly all the problems we are examining below. In particular, their introduction in human environments, workplaces, homes, schools, hospitals, public places, offices, and so on, will deeply and dramatically modify our society.

We have forecasted problems connected to:

- Replacement of human beings (economic problems; human unemployment; reliability; dependability; and so on)
- Psychological problems (deviations in human emotions, problems of attachment, disorganization in children, fears, panic, confusion between real and artificial, feeling of subordination towards robots).
- Well before evolving to become conscious agents, humanoids can be an extraordinary tool used to control human beings.

E. *Recommendations*

Activate working groups inside Standards Committees to study the possibility to define international technical/legal rules for commercial robots regarding:

- Safety. We should provide for systems for the control of robots' autonomy. Operators should be able to limit robots' autonomy when the correct robot's behaviour is not guaranteed.
- Security: H/W and S/W keys to avoid inappropriate or illegal use of the robot
- Traceability: as in the case of sensitive systems, we should provide for systems like the aircraft's black box, to be able to register and document robot's behaviours.
- Identifiability: as cars and other vehicles, also robots should have identification numbers and serial numbers.
- Privacy: H/W and S/W systems to encrypt and password-protect sensitive data needed to the robot to perform its tasks or acquired during its activity.

VII. Conclusions

Following the Atelier, the next planned steps are:

The dissemination and Internet discussion of the Release 1.# of the Roadmap by a larger community of interested scientists who could/did not participate to the Atelier, but who, for good reasons, expressed their commitment to send comments and review. (More than 50 robotics scientists and experts in Computer Ethics and Science&Ethics from Europe, USA and Japan expressed their interest).

The promotion of a cross-cultural updates for engineering scientists who wish to monitor the medium and long effects of applied robotics technologies.

The promotion among robotic scientists of the spirit of the Fukuoka World Robot Declaration (2004), that is summarized by the following points:

1. Next-generation robots will be partners that coexist with human beings;
2. Next-generation robots will assist human beings both physically and psychologically;
3. Next-generation robots will contribute to the realisation of a safe and peaceful society.

Acknowledgment

The author thanks Fiorella Operto for her invaluable contribution to the Roboethics endeavor.

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)