United Nations
Educational, Scientific and
Cultural Organization

World Commission on
the Ethics of Scientific Knowledge
and Technology (COMEST)

Distribution: limited

SHS/YES/COMEST-9EXT/16/3
Paris, 5 August 2016
Original: English

PRELIMINARY DRAFT REPORT OF COMEST ON ROBOTICS ETHICS

Within the framework of its work programme for 2016-2017, COMEST decided to address the topic of robotics ethics building on its previous reflection on ethical issues related to modern robotics, as well as the ethics of nanotechnologies and converging technologies.

At the 9th (Ordinary) Session of the COMEST in September 2015, the Committee established a Working Group to develop an initial reflection on this topic. The COMEST Working Group met in Paris in May 2016 to define the structure and content of its text. Based on the work completed so far, this document contains the preliminary draft report prepared by the COMEST Working Group.

As it stands, this preliminary draft report does not necessarily represent the final opinion of COMEST and it is subject to further discussion within the Commission in 2016 and 2017. This document also does not pretend to be exhaustive and does not necessarily represent the views of the Member States of UNESCO.

PRELIMINARY DRAFT REPORT OF COMEST ON ROBOTICS ETHICS

PRELIMINARY DRAFT REPORT OF COMEST ON ROBOTICS ETHICS

I. INTRODUCTION

1. Autonomous robots are becoming more and more visible in modern societies, regardless of their state of industrial development. Robots are in factories across the world, such as in China, Japan, the USA, and Europe. Drones-robots are used in Afghanistan and in Yemen to fight in the wars against terrorism. Robots can be found in stores replacing workers. Humanoid robots are beginning to be used in schools or for the care of elderly people in Japan and the USA. Robots are also part of therapeutic strategy for children suffering from autism. Furthermore, robots are being used for decontamination purposes or to defuse bombs.

2. The presence of robots at home, in the workplace, and more generally in society is more and more challenging because such presence impacts human behaviours and induces profound social and cultural changes (family and community relationship, workplace, role of the state, etc.). This report aims to raise awareness and promote public consideration and inclusive dialogue on ethical issues concerning the different use of autonomous robots in society.

3. COMEST first had a discussion on the ethical issues related to modern robotics at its 7th Extraordinary Session in July 2012. The discussion was followed by a two-day Pugwash workshop on the Ethics of Modern Robotics in Surveillance, Policing and Warfare held at the University of Birmingham, UK, from 20 to 22 March 2013. At the end of this workshop, a short report was written by John Finney (Pugwash) to summarise its main conclusions, which are included in this report. At the 8th Ordinary Session of COMEST (Bratislava, 20-30 May 2013), a Conference on Emerging Ethical Issues of Science and Technology was organized with a session devoted to autonomous robots. Finally, at its 9th Ordinary Session in September 2015, COMEST) decided to work specifically on robotics ethics and established a Working Group to develop a reflection on this topic. The Working Group met in Paris from 18 to 20 May 2016 and participated in an International Interdisciplinary Workshop on “Moral Machines: Developments and Relations. Nanotechnologies and Hybridity”, co-organized by COMEST and l’Université de Paris 8.

(Comment: Please note that this is the preliminary draft of the report, consolidating chapters that have been prepared by different members of the Working Group. This preliminary draft will still have to be collectively revised in 2016 and 2017.)

II. DESCRIPTIVE PART

II.1. What is a robot?

II.1.1. The complexity of defining a robot

4. Defining a ‘robot’ is a complex and possibly open-ended task due to the rapid developments in robotics. The word ‘robot’ (which replaced the earlier word ‘automaton’) is of Czech origin and was coined by Karel Čapek in his 1920 SF play *R.U.R.* (*Rossumovi Univerzální Roboti* [*Rossum’s Universal Robots*]). The term originates from the word ‘robota’, which means ‘work’ or ‘labor’ in Czech. However, focusing on the etymology of the word ‘robot’ is of little help when it comes to defining what is a robot. To say that a ‘robot’ is something created to do certain work is relatively uninformative because there are many things which fit this description but which do not count as robots (e.g. personal computers or cars).

5. The prevailing view of what is a robot – thanks to SF movies, TV series and literature – is that of a machine that looks, thinks and behaves like a human being. This anthropomorphic view of robots (androids, if designed to resemble human males, or gynoids, if designed to resemble human females) is only partially correct and does not necessarily correspond to the

definitions of robot that can be found in scholarly literature. In such definitions, there is indeed no necessity for a robot to be of a humanoid form. Here are two examples of such definitions:

a. Robot is ‘(1) a machine equipped with sensing instruments for detecting input signals or environmental conditions, but with reacting or guidance mechanisms that can perform sensing, calculations, and so on, and with stored programs for resultant actions; for example, a machine running itself; (2) a mechanical unit that can be programmed to perform some task of manipulation or locomotion under automatic control’ (Rosenberg, 1986: 161).

b. A robot is ‘a smart machine that does routine, repetitive, hazardous mechanical tasks, or performs other operations either under direct human command and control or on its own, using a computer with embedded software (which contains previously loaded commands and instructions) or with an advanced level of machine (artificial) intelligence (which bases decisions and actions on data gathered by the robot about its current environment)’ (Angelo, 2007: 309).

6. In his *Encyclopedia of Robotics*, Gibilisco (2003: 268-270) distinguishes five generations of robots according to their respective capabilities. The first generation of robots (before 1980) was mechanical, stationary, precise, fast, physically rugged, based on servomechanisms, but without external sensors and artificial intelligence; the second generation (1980-1990), thanks to the microcomputer control, was programmable, involved vision systems, as well as tactile, position and pressure sensors; the third generation (mid-1990s and after) became mobile and autonomous, able to recognize and synthesize speech, incorporated navigation systems or teleoperated, and artificial intelligence; the fourth and fifth generations are speculative robots of the future able, for example, to reproduce, acquire various human characteristics such as a sense of humour, etc.

7. One of the pioneers of industrial robotics, Joseph Engelberger, once said: ‘I can’t define a robot, but I know one when I see one.’ As it should become obvious from this report, such a confidence in one’s intuitive ability to distinguish robots from non-robots may be unwarranted because robots are already extremely diverse when it comes to their form, size, material and function.

II.1.1.i. History of robotics – fiction, imaginary and real

8. Artificially created intelligent beings are one of the ancient and persistent preoccupations of human mind and imagination. Such beings can be found in many human narratives: mythology, religion, literature and, especially, SF movies and TV series. It would be impossible to provide a survey of all or even the most important of such narratives, but some paradigmatic and frequent examples from robotics and roboethics literature may be mentioned.

9. According to a Greek myth, Hephaestus, the son of Zeus and Hera, created two female servants out of gold, who helped him in his workshop (as Hephaestus was crippled at birth, these ‘golden servants’ may be seen as fictional prototypes of personal assistant robots). Hephaestus is also considered a creator of the artificial “copper giant” Thalos, whose function was to control and defend the island of Crete (the parallel with military and police robots is obvious). According to another Greek myth, Pygmalion was a gifted sculptor who created an ivory statue of a woman, Galatea, of such an extreme beauty that he fell in love with her. Pygmalion treated Galatea as if it was a real woman – bringing her expensive gifts, clothing her and talking to her. Since it was merely a statue, Galatea, technically speaking, was obviously not a robot (not even as a predecessor of a robot). Nevertheless, the Pygmalion story is relevant for roboethics as it illustrates human tendency to create and form strong emotional ties to inanimate anthropomorphic objects.

10. Hebrew mythology contains a number of stories of the Golem, an anthropomorphic creature which, according to some versions of the story, can be made out of earth or clay and

brought to life through religious or magic rituals. In some versions of the story, the Golem is a faithful and beneficial servant, whereas in other versions it is dangerous and may turn against its creator. In Inuit mythology, there is also a similar legend of such an artificial being called Tupilaq, the only difference being that Tupilaq is made out of parts of animals and humans (often children). Tupilaq is also portrayed as potentially dangerous for its creator.

11. When it comes to literary depictions of robots or robot-like creatures, a *locus classicus* is Mary Wollstonecraft Shelley's novel *Frankenstein; or, The Modern Prometheus* (1818). This is a story of Victor Frankenstein, a scientist who creates an artificial living being using parts of human corpses (in the novel this being is called 'the Creature' or 'the Monster', but it is most commonly referred to as 'the Frankenstein'). Victor Frankenstein is horrified by the outcome of his experiment (especially the Creature's look) and abandons the Creature and all his research. However, the Creature rebels against his creator and resorts to threats and blackmail in order to force Frankenstein to create a female companion for him, invoking his right to happiness. The novel ends in violent and tragic deaths of almost all the main characters.

12. Karel Čapek's SF play *R.U.R.* (*Rossumovi Univerzální Roboti* [*Rossum's Universal Robots*]) published in 1920 occupies a unique place not only in the history of literature on robots, but in robotics as well. The word 'robot', namely, was used for the first time in the play, coined from the Czech word 'robota' (meaning something like 'work' or 'labor'). Čapek acknowledged that his brother Josef Čapek actually invented the word. The play depicts a factory founded by Old Rossum and his nephew Young Rossum that manufactures robots from organic matter (intelligent and often indistinguishable from humans), as a cheap work-force to be sold world-wide. The central event of the play is the rebellion of robots against their creators, which ends in the extinction of nearly all of humanity and the establishment of a new, robot world government (although the new robotic race displays some human characteristics, such as love and self-sacrifice).

13. The literary work of Isaac Asimov is also of special importance for robotics because the word "robotics" was used for the first time in his SF story 'Liar!' (1941). Even more famously, in his story 'Runaround' (1942), Asimov introduced his Three Laws of Robotics:

a. A robot may not injure a human being or, through inaction, allow a human being to come to harm;

b. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law;

c. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

14. In his book *Robots and Empire* (1985), Asimov introduced the Zeroth Law of Robotics: A robot may not harm humanity, or, by inaction, allow humanity to come to harm. Asimov's SF and popular science works were inspiration for many real-world roboticists and AI scientists like Joseph Engelberger and Marvin Minsky.

15. Some cultural differences seem to exist in attitudes towards artificial beings like robots. Whereas in Western culture such beings are typically portrayed as evil and as a possible threat to humans, this is not always so in some non-Western cultures. As Veruggio and Operto (2008: 1503) point out, in Japanese culture 'machines (and, in general, human products) are always beneficial and friendly to humanity.' According to Bar-Cohen and Hanson (2009: 146-147), Japanese are more receptive to humanlike robots because of the strong influence of Buddhism and Shintoism on their culture. Their point is that while Western religions (Christianity and Judaism) consider creation of such beings to be an interference with the role of the Creator, Buddhism and Shintoism have no problem in this respect due to their tradition of "animism", according to which both living and non-living things are endowed with a soul or spirit.

16. Robots (especially humanoid ones) owe a great deal of their popularity to SF movies and TV series. The movie industry developed an early interest in robots. Some of the earliest movies featuring robots are *The Clever Dummy* (by Ferris Hartman, 1917), *L'uomo mecanico*

(by André Deed, 1921, only partially preserved) and *Metropolis* (by Fritz Lang, 1927). Movie makers were always particularly interested in depicting potentially catastrophic effects of the use of robots and their rebellion against humans. For example, in *2001: A Space Odyssey* (by Stanley Kubrick, 1968), a computer that controls all functions of a spacecraft heading for Jupiter sets out to destroy its human crew. In *Colossus: The Forbin Project* (by Joseph Sargent, 1970), two supercomputers, originally built as the United States' and Soviet Union's defense systems in control of nuclear weapons, unite and seize power over the entire human race. In *Demon Seed* (by Donald Cammel, 1977), a computer that operates a scientist's household imprisons and artificially inseminates his wife (she subsequently gives a birth to a clone of hers). *The Terminator* (by James Cameron, 1984) depicts the future world as the dictatorship run by the artificial intelligence with the assistance of an army of robots.

17. Not all SF movies and TV series, of course, portray robots as the ultimate threat to humanity. Many of them present robots in neutral and even humane light, with all the human virtues and vices. Some prominent examples are the robotic duo R2-D2 and C3PO robots from *Star Wars* (by George Lucas, 1977); genetically engineered replicant Roy Batty in *Bladerunner* (by Ridley Scott, 1982) who saves the life of his pursuer and enemy (the 'bladerunner'); android Data who desires to acquire human traits (especially emotions) in *Star Trek* series; and a robot boy David in *AI: Artificial Intelligence* (by Steven Spielberg, 2001) who feels sorrow after being abandoned by his biological 'mother'.

18. When it comes to real world robots, their development was significantly slower and more modest than predicted by their literary and cinematographic portrayals. Their historical predecessors were ''automatons'', artefacts resembling real humans or animals able to perform relatively simple actions such as writing and playing musical instruments. According to Angelo (2007: 28-31), artisans in ancient China constructed various automata like the mechanical orchestra, Leonardo da Vinci made sketches for an automated medieval knight in the 15th century, Guinallo Toriano constructed a mechanical mandolin-playing lady in the 16th century, and highly sophisticated automata were constructed by Jacques Vaucanson (The Flute Player, The Tambourine Player and Digesting Duck) and Pierre Jaquet-Droz (The Writer, The Musician and The Draughtsman) in the 18th century. Automatons were typically intended for the purposes of entertainment.

19. As for the first robots in the modern sense of the word, most historical reviews (e.g. Stone 2005, Angelo 2007) mention 'Unimate' and 'Shakey'. 'Unimate' is considered to be the first industrial robot. It was designed in 1954 by Joseph Engelberger and George Devol for the General Motors assembly line in Trenton, New Jersey. 'Unimate' performed tasks that were dangerous for human workers, like transporting die castings and welding them onto auto bodies. The first 'Unimate', unlike its more sophisticated successors, had just one robotic arm and a program stored on a magnetic drum. 'Shakey' was the wheeled robot developed from 1966 to 1972 by Charles Rosen and his associates at the Artificial Intelligence Center in California. It was the first AI-based general purpose robot able to reason and optimize its commands. It received commands through a computer console and performed tasks such as travelling from room to room, opening and closing doors, moving objects, or switching lights on and off.

II.1.1.ii. Autonomy, interactivity, communication and mobility

20. There are certain features typically associated with contemporary robots which deserve special emphasis, not only because they are central for understanding what robots are, but also because these features, both individually and jointly, raise unique ethical concerns. These features are mobility, interactivity, communication and autonomy.

21. Robots need not be mobile; they can be stationary as it is the case with most industrial robots. However, mobility is essential for many types of robots because it allows them to perform tasks in place of humans in typically human environments (e.g. hospital, office or kitchen). Robot mobility can be realized in various technical ways and today there are robots that are able to walk (bipedal and multilegged robots), crawl, roll, wheel, fly and swim. Whereas

the range of possible harm caused by stationary robot is limited to those working or living in its proximity, mobile robots (especially if they have advanced autonomy and capacity to interact with their environment) may pose more serious threats. Here is an illustrative real-life example: in June 2016, ‘Promobot’ (advertising robot designed for interaction with humans in the street) escaped from the laboratory in the Russian town of Perm and ended up in the middle of a busy road.

22. The ability to interact with their environment is another important feature of robots. It is often claimed that this ability is what distinguishes robots from computers. Robots are equipped with sensors and actuators. Sensors enable a robot to gather relevant information from its environment (e.g. to recognize and distinguish different objects or persons and to determine their proximity). Sensors can include a range of devices (cameras, sonars, lasers, microphones etc.), enabling a robot to see, hear, touch and determine its position relative to its surrounding objects and boundaries. Actuators are mechanisms or devices which enable a robot to act upon its environment (e.g. ‘‘robot arms’’ or ‘‘grippers’’) and which may be mechanical, hydraulic, magnetic, pneumatic or electric. Robot interactivity (realized through sensors and actuators) is particularly sensitive from the ethical point of view. On the one hand, robots are able to actively intervene on their environment and thus perform tasks that could be harmful to human beings. On the other hand, robots may collect data with their sensors that may be used, either intentionally or unintentionally, so as to harm humans or their privacy (e.g. criminal activities, industrial espionage, yellow journalism).

23. Unlike early robots that required computer interfaces for communication, many robots today come equipped with sophisticated speech (voice) recognition and speech synthesis systems, enabling them to communicate with humans in natural languages. The need to develop robots able to reliably receive instructions or even be programmed using natural language (natural language programming) becomes particularly pressing as their presence increases in so many aspects of human life (e.g. robots as assistants to the elderly or nursing robots that need to be able to understand their user’s commands as precisely as possible, but also to explain their own actions). According to Wise (2005: 261), programming computers with human natural language is difficult because human language, on the one hand, contains many multiple meanings, intricate contexts and hidden assumptions, whereas computers, on the other hand, requires extremely explicit instructions. According to Gibilisco (2003: 295), despite the fact that speech recognition and speech synthesis in robots are far from perfect, this area of robotics is rapidly advancing. Two examples of this advancement are the humanoid robots Wakamaru (developed by Mitsubishi in Japan) and Nao (developed by Aldebaran Robotics in France) able to communicate with humans through speech and gestures (Bekey, 2012: 26-27).

24. According to Bekey (2012: 18), ‘the generally accepted idea of a robot depends critically on the notion that it exhibits some degree of autonomy, or can think for itself, making its own decisions to act upon the environment.’ Autonomy is defined as ‘the capacity to operate in the real-world environment without any form of external control, once the machine is activated and at least in some areas of operation, for extended periods of time’ (Bekey, 2012: 18). Earlier generations of robots were more or less sophisticated ‘automatons’ (machines performing relatively simple repetitive tasks, such as assembly robots). However, more recent generations of robots are becoming more autonomous (performing more complex tasks on their own, without depending thereby on human control or commands). ‘Autonomy’ could actually be considered central to a robot, because it is its autonomy that distinguishes it from mere tools or devices that need to be operated by human beings.

II.1.1.iii. Robots, artificial intelligence and algorithms

25. Advancements in robot mobility, communication, interactivity and autonomy became possible thanks to the development of ‘artificial intelligence’ (AI). According to Warwick (2012: 116), ‘an intelligent robot is merely the embodiment of an artificially intelligent entity – giving a body to AI.’ Nevertheless, whereas all robots have ‘bodies’ (which distinguishes them from

computers), not all of them are ‘intelligent’. Most industrial robots or ‘industrial manipulators’ cannot be considered intelligent as long as their behaviour is pre-programmed and adjusted to automatically perform a limited number of highly specific and repetitive tasks (e.g. welding or painting) in non-changing environments (structured world). In cases where there are changes to its environment, an industrial robot typically cannot adapt to these changes (most industrial robots do not even have sensors to perceive these changes) without the intervention of a human operator.

26. Robots that have some form of ‘artificial intelligence’ are generally able to perceive and represent changes in their environment and to plan their functioning accordingly. Artificial intelligence is crucial for robot autonomy because it enables them to perform complex tasks in changing (unstructured) environment (e.g. driving a car and adapting to the conditions on the road) without being teleoperated or controlled by human operator. According to Murphy (2000: 26), ‘while the rise of industrial manipulators and the engineering approach to robotics can in some measure be traced to the nuclear arms race, the rise of the AI approach can be said to start with the space race.’ Whereas teleoperated or automated robotic arms, grippers and vehicles were sufficient to protect human workers from radioactive materials, space robots (like space probes or planetary rovers) needed some form of ‘intelligence’ in order to be able to function autonomously in unexpected situations and novel environments (e.g. when radio communication is broken or when exploring previously uncharted parts of the moon or planet).

27. Since to define ‘intelligence’ itself is difficult, to define ‘artificial intelligence’ is even more difficult. ‘Artificial intelligence’ can generally refer to two interconnected things. It can refer to research area or ‘a cross-disciplinary approach to understanding, modeling, and replicating intelligence and cognitive processes by invoking various computational, mathematical, logical, mechanical, and even biological principles and devices’ (Frankish and Ramsey, 2014). In the words of Marvin Minsky, one of the founders of AI, it is ‘the science of making machines do things that would require intelligence if done by men’ (Copeland ,1993: 1). AI is one of the prime examples of interdisciplinary research area because it combines numerous and diverse disciplines like computer science, psychology, cognitive science, logic, mathematics, and philosophy.

28. The first theoretical proposal of the possibility of artificial intelligence was Alan Turing’s paper ‘Computing Machinery and Intelligence’ published in 1950 in the philosophical journal *Mind*. In this paper, Turing proposed what is now known as the ‘Turing Test’: ‘intelligence’ can be attributed to a machine or a computer programme if a human being, in a specific experimental setup (conversation in natural language via computer terminal), cannot distinguish the responses of the machine or the programme from the responses given by a human being. So far, despite many attempts and advances, no machine or computer program that would suceed in this ‘imitation game’ was developed (Franklin, 2014). It is generally agreed that AI research area (as well as the very name AI) was conceived during the so-called ‘Dartmouth Conference’ organized in 1956 at the Dartmouth College by the pioneers of the field such as John McCarthy, Marvin Minsky and others.

29. ‘Artificial intelligence’ may also refer to the final product of AI research: a machine or an artefact that embodies some form of ‘intelligence’, i.e. that is capable of ‘thinking’ or solving problems in a way similar to human thinking. Although it may seem that robotics is merely an application of knowledge created within the general AI research, the relationship between AI and robotics is more complex. According to Murphy (2000: 35-36), ‘robotics has played a pivotal role in advancing AI’ and it contributed to the development of all the major areas of AI:

a. knowledge representation,

b. understanding natural language,

c. learning,

d. planning and problem solving,

e. inference,

f. search,

g. vision.

In other words, a ‘coevolution’ of robotics and of AI seems to have occurred.

30. As an illustration of the enormous advancement of AI in robotics, consider the following comparison between two generations of AI-based robots (Husbands, 2014):

a. Developed from 1966 to 1972 by the Stanford Research Institute, ‘Shakey’ was one of the first AI-driven robots. It was mobile, able to perceive and represent its environment to some extent, as well as to perform simple tasks such as planning, route-finding and rearranging simple objects. However, it had to be controlled by a computer the size of a room and could not operate in real time (it sometimes needed hours to find its way to the other side of a room).

b. At the beginning of the 21st century, the MIT AI Lab developed a robot designed for social interaction called ‘Kismet’. Kismet was able to move its eyes, make facial expressions depending on its ‘mood’, communicate with humans via speech, and react to its collocutor’s mood and emotion. It was a pioneering work which led to the development of even more sophisticated robots for social interaction.

31. The preceeding two examples testify to the paradigm shift in robotics during the past decades. Whereas early AI-based robots were programmed according to the ‘Hierarchical Paradigm’, the more recent ones are programmed according to the ‘Reactive Paradigm’ or the ‘Hybrid Deliberative/Reactive Paradigm’. In a nutshell, the ‘Hierarchical Paradigm’ implies that a robot performs each operation in a top-down fashion: it starts by ‘sensing’, then proceeds to ‘planing’ and ends with ‘acting’. However, the ‘hierarchical paradigm’, which was basically an atempt to imitate human thinking, faced several difficulties (especially the ‘frame problem’ – the problem of reconstructing the missing pieces of information and differentiating relevant from irrelevant information). This led to the formation of an alternative paradigm, the ‘Reactive Paradigm’, which was inspired in particular by biology and cognitive psychology. Its basic innovation was to reduce the ‘planning’ element (which consumed too much time and computational power) by linking various robot ‘actions’ directly to particular ‘sense’ data (imitating insect behavior). However, since ‘planning’ could not be avoided altogether, especially in general purpose robots, the ‘Hybrid Deliberative/Reactive Paradigm’ was developed, combining the best features of its two predecesors (Murphy, 2000: 2-12).

32. Irrespective of the paradigm it belongs to, and irrespective of the differences between their shape, size and ways of movement, robots perform their tasks thanks to algorithms. AI-driven robot behaviours are mediated by specific algorithms programmed into their ‘computer brains’ (as analogues of biological brains). Algorithm is a term best known in computer science. It is usually defined as ‘a prescribed set of well-defined rules or instructions for the solution of a problem, such as the performance of a calculation, in a finite number of steps’ (Oxford Dictionary of Computer Science, 2016). Algorithms, in other words, are a set of precise and unambiguous instructions ofhow a robot, given certain input or sense data, should perform some operation.

33. It is important to notice that algorithm is the ‘finite’ or ‘deterministic’ set of instructions of how a robot should perform its tasks. This implies that the robot’s behaviour – even if the robot is highly complex, intelligent and autonomous (requires little or no human supervision) – is basically preprogrammed and essentially determined (which may have implications for the moral status of robots; to be addressed later in this report). From the ethical point of view, the fact that robot algorithms are finite and deterministic is relevant (and desirable) when traceability issues arise, i.e. when robot’s past decisions and actions need to be precisely reconstructed in order resolve some ethical or legal dispute. However, assuming that future robots are likely to become even more sophisticated (perhaps to the point that they will be able to learn from past experiences and program themselves), the nature of their ‘algorithms’ will likely become an issue worthy of serious ethical attention and reflection.

II.1.2. Constructing a robot and related ethical challenges

II.1.2.i. Complexity of the construction process

34. The ethical challenges of constructing robots become particularly visible if one takes a closer look at robotics, as the science and technology dealing with design, building, programming and application of robots. Robotics is a truly multidisciplinary enterprise, because creating a robot involves contributions of numerous experts, from a wide array of disciplines like electrical/electronic engineering, mechanical engineering and computer science. Depending on the intended use of the robot, its type and complexity (e.g. stationary or mobile robots, flying or walking robots, individual or swarm robots), as well as the environment in which it will be used, collaboration with an even wider circle of experts may become necessary (e.g. artificial intelligence experts, cognitive scientists, space engineers, psychologists, industrial designers, artists). As it is the case with all complex technological products, creating a robot requires a number of steps: determining its function or task; conceiving of ways in which it will perform this task, designing its actuators and end-effectors, deciding which materials to use for its construction, creating and testing the prototype, designing and programing its software etc.

II.1.2.ii. Responsibilities of the actors

35. Bearing in mind the complexity of contemporary robots, as well as the complexity of their design and construction, the question arises as to who exactly should bear the responsibility (ethical and/or legal) in cases when robots' functioning or malfunctioning brings about harm to human beings. This question becomes even more pressing as robots are becoming more autonomous, more mobile and more present in so many areas of human life.

36. For example, in case an autonomous robotic car causes an accident with human casualties, with whom does the responsibility lie? The team of roboticists that developed the car? The manufacturer? The programmer? The seller? The person (or the company) who decided to buy and use such car? The robot itself? Similar ethical and legal dilemmas are possible with medical robots (e.g. when robot performs a surgery that results in death or harm to a patient), military and security robots (e.g. autonomous military robot killing civilians) or service robots (e.g. kitchen robot preparing meal or drink harmful for the user's health). Possible harm to humans does not have to be physical; it can be psychological (e.g. when personal robot breaches one's privacy; or when interaction with a robot, due to its excessively humanoid characteristics, triggers strong emotions and attachment, which is especially likely in children and elderly).

37. In such dilemmas, one implausible solution would be to resort to the idea of 'shared' or 'distributed' responsibility (between robot designer, engineer, programmer, manufacturer, investor, seller and user). The other solution would be to claim that roboticists are always responsible for any harm that ensues from using their products. A difficulty with this solution is the 'dual-use' issue: robots, like so many other technologies, can be used for both good and bad purposes. Robots may be used for purposes intended by its designers, but they may also be used for a variety of other purposes, especially if its 'behaviour' can be 'hacked' or 'reprogrammed' by their end-users. Avoiding these two extreme solutions is a difficult task which requires careful reflection, especially in view of the fact that robotics teams tend to be very large and that their individual members may be unaware of the final purpose of their segment of work.

II.1.2.iii. Legal and ethical regulation

38. When it comes to ethical reflection about robots and robotics, a lot of important theoretical work has been done and lively discussion is going on in the discipline called 'ethics of robotics', 'robot ethics' or 'roboethics'. 'The target of roboethics', as Veruggio and Operto (2008: 1501) explain, 'is not the robot and its artificial ethics, but the human ethics of the robots' designers, manufacturers, and users.' 'Roboethics' as the branch of applied ethics that deals with the way humans design, construct and use robots, should not be confused with 'machine ethics', as the discipline concerned with the question of how to program robots with ethical rules or codes (to be discussed later in this report).

39. Professions and disciplines that constitute or significantly contribute to robotics have well-developed codes of conduct, for example: IEEE Code of Ethics, ASME Code of Ethics, ACM Code of Ethics, Accreditation Board for Engineering and Technology Code of Ethics. However, robotics itself remains both ethically and legally under regulated, probably due to the fact that it is a relatively new and rapidly changing field of research whose impact on the real world is often difficult to anticipate. Although robotic research projects at universities and research organizations typically need approval from institutional review boards or ethical committees, there are no specific ethical guidelines as to how robotic research and projects, especially those that have direct bearing on humans, should proceed. One should also bear in mind that robotics industry is becoming highly lucrative business and codes of conduct are not likely to be implemented by commercial robotics labs and companies (moreover, they are likely to perceive any ethical guidelines as impediment to their research and development).

40. Despite the fact that there are no universally accepted codes of conduct for roboticists, some initiatives and proposals do exist and need to be mentioned.

41. According to Veruggio and Operto (2008: 1507), in 2007 the Government of the Republic of Korea established a working group (composed of robot developers, CEOs, psychologists, futurists, writers, government officials, users, lawyers and doctors) with the task of drafting a roboethics charter. The charter was planned to be presented at the *Robot Industry Policy Forum* and *Industrial Development Council*, after which specific rules and guidelines had to be developed. As Veruggio and Operto also report (2008: 1506-1507), in 2007 Japan's Ministry of Economy, Trade, and Industry initiated work on a massive set of guidelines regarding the deployment of robots. Particularly emphasized was the need for safety at every stage of robots planning, manufacturing, administration, repair, sales and use. It was recommended, for example, that all robots should be 'required to report back to a central database any and all injuries they cause to the people they are meant to be helping or protecting' (Veruggio and Operto, 2008: 1506). It is unclear whether the Korean charter or the Japanese guidelines were completed or adopted to this day.

42. According to '[a] code of ethics for robotics engineers' proposal, by Ingram et al. (2010), an ethical robotics engineer has the responsibility to keep in mind the well-being of the global, national and local communities, as well as the well-being of robotic engineers, customers, end-users and employers. Its first principle says: '[i]t is the responsibility of a robotics engineer to consider the possible unethical uses of the engineer's creations to the extent that it is practical and to limit the possibilities of unethical use. An ethical robotics engineer cannot prevent all potential hazards and undesired uses of the engineer's creations, but should do as much as possible to minimize them. This may include adding safety measures, making others aware of a danger, or refusing dangerous project altogether.'

43. 'A code of ethics for the human-robot interaction profession' by Riek and Howard (2014), contains the following 'prime directive': 'All HRI [human-robot interaction] research, development, and marketing should heed the overall principle of respect for human persons, including respect for human autonomy, respect for human bodily and mental integrity, and the affordance of all rights and protections ordinarily assumed in human-human interactions' (2014: 5). Among more specific principles Riek and Howard mention the necessary respect for human emotional needs and the right to privacy, the desirable predictability of robotic behaviour, opt-out mechanisms (kill switches), and limitations to the humanoid morphology of the robots.

44. Concerning the research ethics of robotics, Allistene, a consortium of French research institutes (CNRS, CEA, INRA, CDEFI, etc.) has made some practical propositions to enhance the responsibility of robotics researchers in ethical and legal issues. The document (2016) puts the emphasis on the necessity to anticipate the human-robot system. Researchers should also reflect on the limitation of the autonomy of robots and their capacity of imitating human social and affective interactions.

45. As to the legal regulation of robotics and robots, according to Asaro (2012), all potential harms are currently covered by civil laws on product liability, i.e. robots are treated in the same way as any other technological product (e.g. toys, cars or weapons). Harm that could ensue from the use of robots is dealt with in terms of product liability laws, e.g. 'negligence', 'failure to warn', 'failure to take proper care' etc. However, this application of product liability laws on robot industry is likely to become inadequate as commercially available robots will become more sophisticated and autonomous, blurring the demarcation line between responsibilities of robot manufacturers and responsibilities of robot users (Asaro, 2012).

46. In 2016, the Committee on Legal Affairs of the European Parliament issued a *Draft Report with recommendations to the Commission on Civil Law Rules on Robotics*. It proposes a European Parliament resolution and includes, in its Annex, a 'Charter on Robotics' consisting of a 'Code of Ethical Conduct for Robotics Engineers', a 'Code for Research Ethics Committees', a 'License for Designers' and a 'License for Users'. Concerned about the possible impact of robotics on human safety, privacy, integrity, dignity and autonomy, this proposal addresses a number of legal and ethical issues related to robotics and the use of robots, such as data and privacy protection, liability of new generation of robots and robot manufacturers, precautionary principle, testing robots in real-life scenarios, informed consent in robotics research involving humans, opt-out mechanisms (kill switches), creating new and unique insurance schemes for autonomous robots, impact of robotics on human employment and education.

47. A highly important concept for legal and ethical regulation of robotics and robot industry is 'traceability'. Traceability implies that, technically, one must be able to precisely determine (by analysing its program algorithms) the causes of all past actions (or omissions) of a robot. Traceability is ethically and legally important in the sense that 'a robot's decision paths must be re-constructible for the purposes of litigation and dispute resolution' (Riek and Howard, 2014: 6) or, in other words, we should be able 'to determine from its internal records which purposes it was actually in pursuit of when it caused a particular harm' (Asaro, 2012: 179). The Committee on Legal Affairs of the European Parliament, in its 2016 *Draft Report with recommendations to the Commission on Civil Law Rules on Robotics*, also emphasizes the importance of traceability: 'For the purposes of traceability and in order to facilitate the implementation of further recommendations, a system of registration of advanced robots should be introduced, based on the criteria established for the classification of robots' (p. 13). It is particularly emphasized in its sketch of the 'License for Designers': 'You should ensure that the robot's decision-making steps are amenable to reconstruction and traceability' (p. 18).

II.2. Robots in society

II.2.1. Areas where robots are introduced

<u>II.2.1.i. Healthcare robots in elderly care</u>

48. A social robot can be defined as an autonomous robot that interacts and communicates with humans or other autonomous physical agents by following social behaviour and rules attached to its role. Within this category, some researchers emphasise the combination of functional and affective dimension of social robots as the characteristic of robotic devices that can be called companion robots (Oost, Reed, 2010).

49. One area where the use of social robots and robots as companions is in a considerable rise is a healthcare, especially healthcare for elderly people. In the context of demographic projections and reality of rapidly aging population in some countries and the generally low social or gender driven status of caregivers, robots have been proposed as one form of assistive device that can help bridge the widening gap between the need and supply of healthcare services.

50. A healthcare robot is primarily intended to improve or protect the health and lifestyle of the human user. There are many different types and functions of healthcare robots, from surgical robots, robots that aid with rehabilitation therapy, help the disabled and cognitively

impaired maintain their independence, motivate people to exercise and lose weight, robots used for telemedicine, as well as robots that deliver meals, medication and laundry in hospitals. More broadly, healthcare robots can be categorized into those that provide physical assistance, those that provide companionship, and those that monitor health and safety. Some robotic systems cover more than one of these categories. While some are more directly involved in healthcare provision, others are designed more to provide positive effects through companionship. (Broadbent, Stafford and MacDonald, 2009)

51. For many healthcare researchers, the aged population presents a group who may particularly benefit from healthcare robots. They can be used in hospitals and care institutions. But perhaps more importantly, healthcare robots can help older people to live independently in their homes and extend the time of ‘aging in place’ instead of moving to institutional care.

52. If healthcare robots are to be successful, they need to be accepted by older people and their families. Acceptance is defined as the healthcare robot being willingly incorporated into the person’s life. For acceptance of robots to occur, there are three basic requirements: a motivation for using the robot, sufficient ease of use, and comfort with the robot physically, cognitively and emotionally. (Ibid.). Others elements should be considered: the cost and the maintenance of these robots, the impact on role of natural caregivers. The social and cultural acceptance of healthcare robots may also differ from a country to another.

II.2.1.ii. Robots as children’s toys and play companions

53. This category of robotic devices includes robots that actively interact in playing games with people (e.g. robotic pets, mechanical toy robots), but also more passive robots that simply keep a person company without necessarily responding to the person’s specific actions or commands.

54. Pearson and Borenstein (2014) recognize four main areas where ethics in designing children’s robots should be considered.

a. The importance of context-sensitive design: for human-robot interaction (HRI) to promote the welfare of children effectively, some degree of bonding between the robot and child will have to occur. It is necessary to determine which aesthetic features are more conducive to a robot’s ability to fulfil this function. A high priority is anticipating and avoiding design features likely to frighten children, such as an overwhelmingly large size or a ‘‘creepy’’ appearance. Also, whether a robot’s features will elicit a negative reaction from a child will depend in large part on the child’s developmental level and the social context in which the child’s interaction with the robot occurs. Both policy makers and consumers should be made aware of the true capabilities of companion robots and be provided with guidance about how to avoid negative consequences from their use.

b. Cultural factors: collective aesthetic preferences attributed to a particular culture may be considered in designing robots.

c. Humanlike appearance and the uncanny valley: a major design consideration is whether and to what degree robots should be humanlike. The uncanny valley hypothesis states there is a ‘‘valley’’ of great discomfort when we interact with a robot or other entity that looks human but lacks key attributes that we would normally expect to accompany a certain (human) appearance. On the other hand, research suggests the ethology or manifestation of uncanny valley-type experiences in adults may not necessarily translate to the case of children. Some studies also suggest that introverted humans prefer more mechanical-looking robots, whereas extroverts prefer a more human-looking robot. Considering the benefits of animal-assisted therapy, there are compelling reasons to continue creating at least some robots that look like non-human animals.

d. Gender: in literature, the relationship between gender stereotypes and decisions about the design and use of robots in reported as important but still under explored. A key issue to consider is whether robots should be gendered and exactly what it would mean for a robot to be gendered. Inquiry within the domain of robots and gender should force us to re-examine current social attitudes and practices. In general, design and use decisions that reinforce gender stereotypes are ethically questionable.

e. Children psychological development: there is few studies on the long term effects of children-robots interactions and affective dependency (imitation, developing autonomy and judgment, etc.). As in the case of elderly persons, children in research ethics are considered as more vulnerable groups, a precautionary approach may be an ethical requirement (Allistene, 2016).

II.2.1.iii. Robots in education

55. Educational robotics is the term used to describe the use of robotics as a learning tool. Typically, two goals are introduced as learning objectives (Eguchi, 2012). One goal is to use robots to make children interested in learning about the world of technology by incorporating classes and activities that focus on teaching children about robots. Another aim is the creation of new projects using robots as learning tools in order to engage children in activities while teaching concepts not easily taught with traditional approaches. Unlike the sophisticated robotics technology used in robotics labs, lower-cost educational robotics kits with less sophisticated sensors and controllers have made the technology available for use in classroom settings.

56. Typically described learning themes (Miller, Nourbakhsh and Siegwart, 2008) of educational robotics include:

a. Interest in science and engineering: the study of robotics can be sufficiently rewarding, intellectually and emotionally, as to generate general enthusiasm for the follow-on study of science and engineering. Some educational robotics programs have shown promise in increasing retention rates among female students, who are underrepresented in technology-focused fields.

b. Teamwork: an important result of team-based project-based robotic building experiences is that those with low self-esteem in their technology capability, teamwork skills, and communication skills can find measured improvement, leading to mitigation of fear towards working in teams on high-technology ventures.

c. Problem solving: the study of a complex robot system provides lessons for problem solving and decomposition skills with lifelong applicability to any complex system.

57. More generally, the role for educational robots is that of ‘the robot as a learning collaborator. In this case students are not designing robots, but rather conducting inquiry during which a high-functioning robot can serve as an all-season companion, aide, and even intellectual foil. Robots that have social capabilities are well suited in this regard, and this is an area where the potential for application greatly outweighs what has, so far, been achieved. There are two reasons why such a social role for robots in education holds particular promise. First, because of their novelty, robots can inspire dedication in learning. Second, students have few preconceived expectations for the behaviour of robots. In therapeutic cases, such as students with autism spectrum disorder (ASD), the robot can play an important role, as studies have shown that those students place disproportionate interest in mechanical systems.’ (Ibid.)

II.2.2. Psychological and sociological aspects of human-robot

58. Human-robot interactions lead to various psychological, sociological and cultural aspects that can be positive and negative. Some questions about social transformations induce by the use of robots on a large scale can be identify:

a. Does this phenomenon induce less human individual autonomy and more robots control at a social level?

b. Does the generalization of robots generate a standardization of behaviours and pattern of consumption therefore inducing an erosion of cultural diversity and way of life?

c. To what extent the service robots accentuate a consumer attitude rather than an individual producer/creator behaviour?

d. Does people want the presence of robots in their home (their oïkos, their family intimacy)? To what extend their private life will be analysed by robots and oriented by robot competency?

e. Does interaction with robots can be a substitute to family bond in health care for elderly or children' education? What are the consequences on the family organization at a social level?

f. What is the difference between interaction with robots and relation toward other people or animals?

g. To what extent caring expression of robot is fake or simulacra?

h. Do service robots at home needs to have a humanoid appearance? Does this induce gender bias? What are the consequences of the human attachment toward humanoid robot?

i. What kind of trust or reliability is possible to have toward robots at a social and individual level?

II.3. Robots in medicine and healthcare

II.3.1. Medical robots

59. In medicine, semi-autonomous robots have been used in surgery (prostate cancer) for the past 10 years. Their use is still subject to discussion for different reasons. Their defenders presented the comfort for the surgeon and some studies show shorter time in the hospital for the patient and lower blood lost. But studies do not show a significant difference in terms of efficiency between the usual surgery and the robot surgery (Kappor,2014). The major problem is the cost of surgery robots and their maintenance. So actually, the cost of a surgery by a robot is higher than one with an 'ordinary surgeon' (Gagnon et al., 2014). The multiplication of robots in surgery will have consequence on the allocation of resources in public health systems, on the number of surgeons and on their ability to practice surgery without robots.

II.3.2. Robots in healthcare

60. Robots are used also in therapeutic approach to children with autism. Nao is an example of a robot that is utilized on an experimental level to improve the learning of children with autism. Exoskeleton is used to help person with physical handicap (artificial legs with Ekso), but can be used in for other purposes (athletic performance of Oscar Pistorius in the London Olympics 2012, super-soldier, etc.). Exoskeleton, neurological implants, nanorobots, open the door to the transformation of human body and mind. The ethical question is to what extent the hybridity human-robot should be develop and for what purpose? Should it frame human being to be similar to a robot? Then which ideology (enhancement, submission, super-power movies hero's) drives this transformation?

II.4. Robots from warfare to surveillance and policing

II.4.1. Autonomous weapons and military ethics in general

61. Autonomous weapons can be defined as weapons that once launched would select and attack targets without further human intervention. Issues of distinction, proportionality, accountability and transparency are relevant both to remotely-piloted armed vehicles and with fully autonomous weapons systems. Taking the human fully out of the loop intensifies many of those issues and also raises additional ethical and legal concerns. Many of these issues relate to the ability of software to take the decisions that a human would normally take, as well as the ethical acceptability of machines taking such life and death decisions.

62. On the issue of distinction, discrimination systems are of necessity based on what is being looked for. If this is well defined, then it may be possible for sophisticated image recognition software to identify correctly an object, but outside that strict definition it will be unable to act reliably. For example, when a combatant acquires a new set of rights – e.g. as a Prisoner of War, or as an injured soldier – how can the software decide that this new set of rights has been acquired? Similarly, there is a whole set of ICRC (International Committee of Red Cross) guidelines relating to civilians directly participating in hostilities, guidelines which require human interpretation. How will software distinguish an offensive from a defensive system on the other side? Although advanced image recognition systems may in the future be able to identify clearly a man who is armed, will it be able to tell if that person is a combatant or a nearby civilian policeman? There is the issue of how to surrender to a remotely-piloted robot: in the autonomous system context, as there is no set requirement for a way to surrender, a response to an attempt to surrender cannot by definition be programmed, so also by definition recognising surrender requires human judgment.

63. The proportionality criterion is even more challenging for an autonomous system to fulfil. Can an autonomous system assess military necessity, an assessment that requires consideration of a range of issues, many of which are inherently unquantifiable? Can it determine and inflict the degree of harm that needs to be inflicted to counter a particular situation?

64. Proportionality assessment is more than just estimating the number of possible collateral deaths within the destruction area of the projectile. Can a programmed machine decide how many collateral deaths are justified? How can it estimate loss of life from various consequences such as destruction of infrastructure and loss of ability to provide, or consider in the equation the other social and psychological consequences? Assessing proportionality is very much a judgment call, and one which we expect to be taken by a ‘reasonable military commander’, requiring human contextual awareness. So it would seem an inevitable conclusion that it must involve human judgment and cannot ethically be left to a machine.

65. To kill legally, an autonomous machine must be able to act in accordance with the rules of international humanitarian law (IHL). Much has been made of the possibility of creating an ‘ethical governor’ that is programmed to follow those rules. In order to be able to do so, those rules need to be reduced to simplistic terms suitable for programming into the robot. The problem here is that IHL consists of guidelines that try to establish processes of reasoning, and as such it would be inappropriate – and dangerous – to try to reduce or simplify them.

66. A further point here relates to the authority to kill or to delegate the power to kill, and here IHL require specifically human decision making. An underlying assumption of most legal and moral codes is that, when the decision to take life is at stake, the decision should be made by humans. For example, The Hague convention requires a combatant ‘to be commanded by a person’. The Martens Clause – a longstanding and binding rule of IHL – specifically demands the application of ‘the principle of humanity’ in armed conflict. It requires a human to override my right to life, and that right cannot be delegated to a machine. Even if it were technically feasible to programme in the rules of IHL, without major legal changes, that would still not be acceptable action.

67. Delegation of killing to a machine, however complex its programing, can be argued to deny moral agency that is a key feature of a person’s intrinsic worth – human dignity. Human dignity is fragile and should be respected and protected – it is fundamental to the ethical and legal bases in, for example, the UN Universal Declaration of Human Rights.

68. A key issue of robotics that relates to many of the above specific issues concerns their operation in ‘open’ as against ‘structured’ environments. A robot is designed for a specific purpose or set of purposes within a specified (‘structured’) environment, and can work well when the environment is adapted to suit them. Move them outside this specified environment to one which is more open and it becomes unpredictable. As the failure of some sophisticated image recognition software has demonstrated, even small changes in the environment can

lead to malfunction. All possible environments and likelihoods cannot be programmed into the system and its behaviour outside its comfort zone may be catastrophic. The potential consequences of unpredictable behaviour become even more worrying if we consider two sets of autonomous vehicles programmed with different – and unknown to the other – rules of operation.

69. In this context, it should also be noted that even relatively simple systems fail and need human intervention to resolve the malfunction. Also, complex systems can be made unpredictable by relatively simple means such as a bullet through a circuit board, or a Trojan virus inserted into the hardware during manufacture. The consequence in the battle space could be catastrophic, with unforeseen events snowballing rapidly and resulting in escalation without human intervention. The consequences of spoofing and takeover of autonomous machines would be even more problematical than for remotely-piloted vehicles.

70. Extending the ethical question beyond the conventional military ethics problems of distinction and proportionality, there is a fundamental reflection that autonomous weapons rule out combat and transform war from being possibly asymmetrical into a unilateral relationship of death-dealing in which the enemy is deprived of the very possibility of fighting back. With this, they possibly slip out of the normative framework initially designed for armed conflicts.

II.4.2. Military drones and ethical limits of legal approach

71. In the military vocabulary, a drone is defined as ‘a land, sea, or air vehicle that is remotely or automatically controlled’ (U.S. Army). The drone family is not composed solely of flying objects. There may be as many different kinds as there are families of weapons: terrestrial drones, marine drones, submarine drones, even subterranean drones. Provided there is no longer any human crew aboard, any kind of vehicle or piloted engine can be ‘dronized.’ A drone can be controlled either from a distance by human operators (remote control) or autonomously by robotic means (automatic piloting). In practice, present-day drones combine those two modes of control. The term ‘drone’ is mainly used in common parlance. Military jargon refers to ‘unmanned aerial vehicles’ (UAVs) or to ‘unmanned combat air vehicles’ (UCAVs), depending on whether the contraption carries weapons. (Chamayou, 2015)

72. Drones are not much different from conventional weapons at first glance. Pilots sit in control pod thousands of miles away from the action using a games controller to fly the craft and attack on the ground. However, as for any new weapon system it opens a number of new military possibilities and brings new moral dangers. It was noted that the distance between pilot and field of action could lead to a gaming mentality. This could create a moral buffer from the action that could lead to carelessness about the selection of targets.

73. There is no doubt that International Humanitarian Law (IHL) applies to new weaponry and to the employment in warfare of new technological developments. This is fundamental for the effective implementation of IHL and is also recognized in Art. 36 of Additional Protocol I which requires states to determine whether the use of any means or method of warfare would, in some or all circumstances, be prohibited by IHL. While developments of remotely-controlled and automated weapons systems can have benefits for military forces and even civilians in some situations, there are concerns as to whether the existing IHL rules are sufficiently clear in the light of the technology’s specific characteristics, and with regard to the foreseeable humanitarian impact that use of such technologies may have. It is therefore important to remind states of the crucial need to assess the potential humanitarian impact and IHL implications of new technologies to ensure that they can be – and are – used in a way to guarantee respect for IHL. There is also a need to ensure informed discussion and analyses of the legal, ethical and societal issues involved well before such weapons are developed.

74. Making the distinction between a combat and civilian in a remote situation is of its nature a complex one. The data the pilot has on which to make a decision is very different from that available to a commander in the field – a consequence of the shifting from embodied forms of information to database forms has a profound influence on situational awareness. This considerable increase in complexity of the data set on the basis of which a targeting decision

is made opens up the possibility of more error in identification. The technical quality of the image available to the pilot, together with the quality of the intelligence available, are both critical in informing a decision to fire. Yet flying an armed drone has been described as ‘flying a plane looking through a straw’, with obvious implications for limiting the spatial contextual information available when making a targeting decision.

75. The fact that your weapon enables you to destroy precisely whomever you wish does not mean that you are more capable of making out who is and who is not a legitimate target. The precision of the strike has no bearing on the pertinence of the targeting in the first place. (Chamayou, 2015)

76. Distribution of decision making among the crew of a remotely-piloted craft encourages ‘group think’, with crews actively looking for evidence to justify targeting. Discrimination systems are based on what we are looking for, and trying to do this remotely will tend to simplify characteristics used to identify a legitimate target. An increased focus on military age males is a consequence, leading to coarse stereotyping in decision making. A civilian directly participating in hostilities may be a legitimate target under IHL, but can a remote pilot make that distinction reliably?

77. Inherent time delays in signal transmission from operator to vehicle of between 1.5 and 4 seconds mean that a situation on the ground can change significantly between a decision to fire being made and the ‘trigger’ being pulled. Even in that short time, the situation on the ground can change, with the possible result of killing non-combatants. Though this time delay may be reducible through technical advances, basic physical limitations prevent it being reduced to zero. It is possible for a combatant to surrender in the field. It is unclear how he/she can surrender in the face of a remotely-piloted vehicle.

78. Assessing proportionality remotely is even more problematical, including as it does more than actual civilian casualties. Other collateral damage needs to be considered that it is questionable can be done remotely. Though there are claims that the precision of drone strikes can minimise collateral damage to property, it is difficult to see ways in which potential damage to the ability to provide and psychological damage can be assessed remotely.

79. Studies on the effects of drone strikes illustrate some of these proportionality issues. Many of those interviewed were suffering from post-traumatic stress disorder, and other severe abnormal psychological conditions were found including depression, panic reactions, hysterical-somatic reactions, exaggerated fear responses, depression and abnormal grief reactions. The impact on children was particularly concerning: they suffered from attachment disorders and exhibited a severe fear of noise, a lack of concentration, a loss of interest in pleasurable activities. A further consequence was infrequent or non-existent school attendance.

80. Legally, these kinds of ‘collateral damage’ might be argued to be ‘collective punishment’ and counter to the prohibition on reprisals. Is the psychological impact of drone warfare on civilian populations sufficiently recognized or recognized at all under IHL? What about the social and economic impact and the seemingly open-ended nature of such societal disruption?

81. Regarding accountability, where remote-controlled aircraft are used by regular military forces for the same tasks that conventional aircraft have traditionally been used for, the legal position under IHL appears to be relatively unproblematic – troops will have received IHL training and they would be part of a disciplined chain of military command.

82. In terms of drone warfare, IHL does offer legal protection to civilians. However, lack of transparency means the characteristics of the contexts in which armed drones are used can be contested and the post facto assessment of the legality and impact of drone strikes is difficult.

83. Transparency is generally noticeable by its absence in current deployment of armed drones. Most activity seems to be covert and information is generally not made available to the

public, nor in some cases to democratic governing bodies – with claims of national security being used in defence of the withholding of information. The rationality and reasons for legality of use tend not to be outlined by governments: this inevitably raises questions of who is making the decisions and why is there lack of transparency. But lack of transparency makes accountability difficult, and hinders investigations of abuse.

84. Transparency is also noticeable by its absence in assessment of robotic weaponry under IHL Art. 36 Though there is evidence that such reviews are undertaken, the outcomes of the reviews are not available. Although Art. 36 requires it to be shown that the weapon is capable of being used in a way that is consistent with IHL, how it will be used in a way that is consistent with IHL seems not to be transparent.

85. The legal issue of consent to deploying armed robotic vehicles does not appear to be different from that needed for use of other military means. It is legal to use force in another state's territory if that state itself cannot counter an imminent threat – but that threat has to be a direct one to the state applying the force. There may be issues concerning the nature of that consent. For example: does it have to be explicit to be legal? Who can give that consent? Can it be withdrawn? How can it be ascertained that the consent is real and not coerced? Although these issues hold whether or not drones are the method of intervention, the general lack of transparency surrounding their deployment may complicate the consent issue.

86. Robotic weapons also have potentially game-changing strategic implications. Their use seriously lowers the activation barrier to armed conflict, relaxing the requirement of jus ad bellum and shifting the assessment of military necessity. Targeted killing is not new, but armed robotic systems have made it easier and more frequent. We may thus be in real danger of a major strategic change in which we are able to conduct a 'riskless war' continuously and over the long term. We can begin to envisage a world in which low intensity war through targeted killing is the rule – a very different world in which the continuous possibility of targeted killing takes the place of a longer term development of stable co-existence. Once such a state is entered, it is difficult to see how it could be exited from.

87. A number of issues with ethical implications therefore include:

a. A shift in the notions of self-defence and threat. Is it ethical to kill when you are not under threat yourself?

b. Psychological effects on those living with the threat of strikes.

c. Psychological effects on armed robot operators, their families and their societies.

d. Undermining traditional military values of honour and valour.

88. Finally, there are significant issues of safety of these systems. Like all machines, remotely-controlled or not, they are subject to failure (internal or inflicted externally) which could have fatal or other damaging consequences. Unlike other systems, however, remotely-controlled systems are potentially open to spoofing and hacking, giving them the potential to be taken over by the 'opposition' and even turned back on the attacker's interests.

II.4.3. Surveillance, policing and continuity of using military technology in non-military contexts

89. There are many positive civilian uses of robotics in surveillance, for example in farming tasks, wildlife monitoring, finding lost children, in rescue missions after disasters, and environmental damage assessment in extreme environments. These would seem to raise few ethical issues.

90. On the other hand, robotic vehicles are increasingly used for surveillance uses which raise serious issues. Other means of surveillance – e.g. CCTV, satellites – are much more limited in scope. 'Drones' and other robotic vehicles can not only see – they can also hear. They can explore space in three dimensions (reaching the parts other surveillance techniques cannot reach), are flexible, persistent and have 24/7 capability. They thus fill the major gaps in surveillance capability, and are currently used by European police forces for border, crowd and

event control; evidence gathering; traffic control, searches, observation; documenting ‘troublemakers’; surveillance of buildings and VIPs; searching for/controlling/targeting undocumented migrants, workers, demonstrators. The implications for privacy and data protection are major: not only is the data collection potential massive, but different kinds of information can be networked and used, and the need to classify personal data encourages stereotyping of people. And ubiquitous 24/7 surveillance could undermine political participation.

91. An increasing characteristic of civilian surveillance is a transfer of security technologies from the military realm to the civil. This military to civilian technology transfer has implications for the securitization of traditionally non-military problems and aspects of society, with further implications for human rights. This adaptation and proliferation of military technologies may lead to ‘militarisation’ of society, in the sense of a logic of ordering the world that runs on a productive economy of fear: the fear of an ‘omnipresent enemy who could be anywhere, strike at any time and who could be ‘among us’’. The construction of these surveillance assemblages makes possible not only prosecuting specific crimes and following concrete suspicions, but also the monitoring of a population systematically and thoroughly on an everyday basis. This is similar to the military concept of C4ISR (Command, Control, Communications, Intelligence, Surveillance and Reconnaissance) - a networking of all control systems to achieve a global overview in the war theatre. The effect on future generations of such a visible surveillance state being seen as normality has fundamental implications for our view of what future society might be.

92. There are questions about the use of so-called sub-lethal weapons on robots for use in the suppression of public dissent and for control of border regions. There is a concern that future generations of the establishment who inherit this technology will use them as tools of oppression. In applying lethal force to the population, robots will not disobey an authoritarian regime in the way that human soldiers or police would.

93. Also note that recently (Dallas protest shooting in July 2016) for the first time in history, US police have used a robot in a show of lethal force, using a bomb-disposal robot with an explosive device on its manipulator arm to kill a suspect. In light of this developments, clear rules and safeguards will have to be drawn, as proposed for example by American Civil Liberties Union:

a. Usage Limits: A drone should be deployed by law enforcement only with a warrant, in an emergency, or when there are specific and articulable grounds to believe that the drone will collect evidence relating to a specific criminal act.
b. Data Retention: Images should be retained only when there is reasonable suspicion that they contain evidence of a crime or are relevant to an ongoing investigation or trial.
c. Policy: Usage policy on drones should be decided by the public’s representatives, not by police departments, and the policies should be clear, written, and open to the public.
d. Abuse Prevention and Accountability: Use of domestic drones should be subject to open audits and proper oversight to prevent misuse.
e. Weapons: Domestic drones should not be equipped with lethal or non-lethal weapons.

II.4.4. Invading privacy and Illicit use of robots

94. Small remotely-piloted vehicles (RPVs) are already available on the open market, the consequent snowballing use of the technology raising a number of legal and privacy issues in their use e.g. by the media, by criminals (including drug transport), and for stalking and voyeurism.

95. Deployed without proper regulation, drones equipped with facial recognition software, infrared technology, and speakers capable of monitoring personal conversations would cause

unprecedented invasions of privacy rights. Interconnected drones could enable mass tracking of vehicles and people in wide areas. Tiny drones could go completely unnoticed while peering into the window of a home or place of worship.

96. The potential use especially of small drones by terrorists should be noted. There have been several documented instances of non-state groups employing, or seeking to employ, weaponised unmanned aerial vehicle technology. Certain terrorist organisations, and individuals claiming to be affiliated with these organisations, have demonstrated a willingness to employ drones as delivery vectors for improvised explosive devices, as well as chemical and biological weapons. Security officials have warned that advances in the availability and sophistication of commercially available drones increases the likelihood of such events occurring in the future.

97. Safety issues are also involved with wider use. The hobbyist community could covert commercial drones to fly large distances and carry significant payloads. However, the ready availability of drone technology will make restricting this technology or enforcing licensing schemes problematical. Cases of air incidents involving hobbyist drones have already been reported.

II.5. Robots and human labour

II.5.1. Economics of robotisation

98. Robotisation of industry has already happened and over the past decades, industrial robots have taken on the routine tasks of most operatives in manufacturing. However, the more advanced robots (mobile robotics) with enhanced sensors and manipulators and AI-improved algorithms (converging technologies) have now increasingly started to perform non-routine manual tasks. With improved sensors and mobility, robots are capable of producing goods with higher quality and reliability than human labour in large industrial sectors such as food industry, energetics and logistics. As robot costs decline and technological capabilities expand, robots can thus be expected to gradually substitute for labour in a wide range of low-wage service occupations, where most job growth has occurred over the past decades. This means that many low-wage manual jobs that have been previously protected from computerisation could diminish over time (Frey, Osborne, 2013). While this development should lead to higher productivity from new technologies, it doesn't necessary mean improvement for workers. Without structural adjustments and compensations, it could lead both to the higher unemployment for certain profiles of work force, and to the rising inequality in society if the profits from the higher productivity of new technologies flow mostly to the owners of the technology.

99. This situation was described by some authors as the entering into a new period in history, characterised by ‘the end of work’. Again, its interpretations fundamentally differ: for some, particularly the scientists, engineers, and employers, a world without human work will signal the beginning of a new era in history in which human beings are liberated, from a life of backbreaking toil and mindless repetitive tasks. For others, the workerless society conjures up the notion of a grim future of mass unemployment and global destitution, punctuated by increasing social unrest and upheaval (Rifkin 1995).

100. The second important ethical question of robot economy concerns financing and driving of research and development in robotics. Institutions like Foundation for Responsible Robotics (FRR) and its founder, AI pioneer prof. Noel Sharkey are drawing attention to the fact that despite large government and private investments in robotics research and development, the effects on society aren’t mentioned at all, while the ethical implications of robotics usually garner only few mentions in one policy documents.

101. In this context it is again important to note ethical responsibility of industrial robot designers and engineers. A variety of ethical codes have been proposed to define responsibility of a robotics engineer. For example, (in Worcester Polytechnic Institute’s Code of Ethics for Robotics Engineers), engineer should consider the possible unethical uses of his

creations to the extent that it is practical and to limit the possibilities of unethical use. An ethical robotics engineer cannot prevent all potential hazards and undesired uses of the engineer’s creations, but should do as much as possible to minimize them. This may include adding safety features, making others aware of a danger, or refusing dangerous projects altogether. A robotics engineer must also consider the consequences of a creation's interaction with its environment. Concerns about potential hazards or unethical behaviours of a creation must be disclosed, whether or not the robotics engineer is directly involved. If unethical use of a creation becomes apparent after it is released, a robotics engineer should do all that is feasible to fix it.

II.5.2. Robot divide

102. Another set of ethical challenges arises with extension of digital divide concept to proposed concept of robotics divide. If the term digital divide refers to how technology, and access to it, redefines the power structure in contemporary human societies, the problem of robotics divide would address the questions such as: Will robotics, which is characterized by converging technologies, rapid progress, and a progressive reduction in costs, be incorporated into our societies under the current market model? Will it create a new technological, social, and political divide? Will it transform power relations between individuals, groups, communities, and states in the same way as crucial military technology? Will it change our everyday social interactions by forming part of our daily lives in the same way it has changed industry where industrial robotics is already mature and fully established? (Peláez, 2014) Also, as with a digital divide, there is a question of global gap between developing and developed countries regarding access to robotics technology which should be examined from the point of view of global justice.

103. The impact of robotisation on globalized economy can be very asymmetrical. As robots get cheaper and better, the economic advantage of employing a low-skilled worker starts to fade. Even lowly payed workers will struggle to compete with a robot's productivity. Where poorer countries could once use their cost advantage to lure manufacturers, now all cost advantages are disappearing in the robotics age. A robot costs the same to employ whether in China, the U.S. or Madagascar. Some major western corporations are moving production back home to largely automated factories, closer to its customers and free from the risks, costs and complexities of a lengthy supply chain (Balding, 2016).

104. How can developing countries confront this possible widening of the digital and robot divide, and its potential threat to their development strategies? As mentioned in United Nations 2030 Agenda for Sustainable Development and in the Addis Ababa Action Agenda, one thing they need to do is turn the possibly liberating power of open data and big data to their own advantage. If data are the lifeblood of the robot revolution, then they must also be used to defend and compensate those who might otherwise lose out from these new technologies (OECD, 2016).

II.5.3. Changes in workplace and security of workers

105. Advancing robotisation brings profound changes to working conditions and job transformation. Working ‘side by side’ with robots demands new working skills and the need for new and adequate safety measures for workers. The question of working time for humans working side by side with more and more autonomous machines also arises, as it did with any technological revolution in the past history of work. Finally, education and the system for additional training of existing workers need to be established and adjusted so that the latest technological revolution doesn’t result in producing mass of unemployable workers due to their lack of skills with latest technologies.

106. Recent research (Murashov, Hearl and Howard, 2015) on occupational safety when working with robots recognizes three categories of robots in a workplace:

a. industrial robots;
b. professional and personal service robots
c. collaborative robots.

Most of industrial robots are unaware of their surroundings, therefore, they can be dangerous to people. The main approach to the industrial robot safety is maintenance of a safe distance between human workers and operating robots through the creation of ‘guarded areas.’ A number of guidance documents and international standards deal with workplace safety around industrial robots. However, continuing incidents resulting in harm and death to workers indicate that additional measures such as additional redundancies in safety measures, additional training and improvements in overall safety culture are necessary.

107. Service robots operate mostly outside industrial settings in unstructured and highly unpredictable environments without people such as disaster areas or with people present such as hospitals and homes. Physical proximity between professional service robots and human workers are much more common than between industrial robots and workers since they often share the same workspace. Therefore, worker isolation from the professional service robot is no longer an option as the main safety approach. Furthermore, more complex environments in which service robots must operate dictate much higher degree of autonomy and mobility afforded to service robots. This autonomous and mobile behaviour can result in dangerous situations for workers. Despite the proliferation of safety concerns involving service robot workers in the same workplace as human workers, no international standards have been developed to address human worker safety in maintaining or operating professional and personal service robots.

108. Collaborative robots defined as a robot designed for direct interaction with a human could be industrial, professional, or personal service robot. Collaborative robots combine the dexterity, flexibility and problem-solving skills of human workers with the strength, endurance and precision of a mechanical robots. A new field of collaborative robotics is managerial robotics. Instead of being relegated to mundane, repetitive, and precise job tasks, robots with their perfect memories, internet connectivity and high-powered computers for data analysis could be successful in particular management roles. Since robots are working alongside human workers, isolation as a safety measure is no longer an option, and other safety approaches (proximity sensors, appropriate materials, software tools) must be developed and implemented.

II.6. Moral machines

II.6.1. Programming moral codes for robots?

109. The problem with development and utilization of robots, as it was the case with many other technological innovations, are unforeseeable and unintended harms to human beings. The possible harm is related to the fact that many robots are designed to perform services to or work in close interaction with human beings. In such settings, the potential harm to humans arises due to inevitable errors in design, programming and production of robots. It is likely that malfunctioning of today’s sophisticated robots is capable of inflicting significant harm to a very large number of human beings (e.g. armed military robots or autonomous robotic cars getting out of control). The question is, therefore, not only if roboticists ought to respect certain ethical norms, but whether certain ethical norms need to be programmed into robots themselves. Such a need is apparent already if one focuses on personal robots and possible harm they could inflict on humans (e.g. robots for cooking, driving, fire protection, grocery shopping, bookkeeping, companionship, nursing).

110. Robots’ autonomy is likely grow to the extent that their ethical regulation will become necessary, by programming them with ethical codes specifically designed to prevent their harmful behaviour (e.g. endangering humans or the environment). The emerging ‘machine ethics’ discipline is already ‘concerned with giving machines ethical principles or a procedure for discovering a way to resolve the ethical dilemmas they might encounter, enabling them to function in an ethically responsible manner through their own ethical decision making’ (Anderson and Anderson 2011a: 1). This general research already has its specializations, e.g. ‘machine medical ethics’, focusing on ‘medical machines’ capable of performing ‘tasks that require interactive and emotional sensitivity, practical knowledge and a range of rules of

professional conduct, and general ethical insight, autonomy and responsibility' (Rysewyk, Pontier, 2015: v).

111. Two important questions need to be considered in this context: Is it possible to construct some kind of a 'artificial moral agents'? If it is, which moral code should they be programmed with? Most scholars working on machine ethics agree that robots are still far from becoming 'ethical agents' comparable to human beings. However, they also seem to agree that the capacity for moral agency comes in degrees and distinguish 'implicit' and 'explicit' ethical agents. According to Moor (2011), a machine is an 'implicit' ethical agent in so far as it has a software that prevents or constrains its unethical action. It is possible, in other words, to ethically program machines in a limited way in order to avoid some morally undesirable consequences of their performance of specific tasks (e.g. automated teller machines are programmed not to short-change customers, automated pilots are programmed not endanger passengers' safety). It is also argued by some philosophers (Savulescu, Maslen, 2015) that 'artificial ethical agents', thanks to their computing speed and lack of typically human moral imperfections (partiality, selfishness, emotional bias or prejudice, weakness of the will), could actually aid or even replace humans when it comes to difficult moral decision-making in specific and limited contexts (in the same way as human imperfect physical or mental labour was replaced by its robotic or AI counterpart).

112. A more intriguing question is whether robots can ever become 'explicit' ethical agents in the same, full-blown way as human beings. Is it possible to program machines or robots to 'act on principles' or to 'have virtues' or to 'provide moral reasons for their actions'? Can they have 'moral knowledge' and apply it in a range of different and possibly very complex moral dilemmas? According to Moor (2011:17), although interesting research in this area is happening (especially research on advanced deontic, epistemic and action logic), for the time being 'clear examples of machines acting as explicit ethical agents are elusive' and 'more research is needed before a robust explicit ethical agent can exist in a machine.' However, there are also some optimistic views. Whitby (2011) thus maintains that such 'explicit' ethical agents will not have to be designed in a principally different way than current chess-playing programs, especially in view of the fact that chess-playing programs do not have their every possible move preprogrammed, but actually rely on general *decision making procedures* (even *guesses* and *hunches* about best possible moves which can be refined and altered as the chess-match proceeds).

113. The fact remains, however, that 'programming a computer to be ethical is much more difficult than programming a computer to play world-champion chess', because 'chess is a simple domain with well-defined legal moves' whereas 'ethics operates in a complex domain with some ill-defined legal moves' (Moor, 2011: 19). In other words, making moral decisions and acting on those decisions in real-world and real-time would require not only knowledge of complex moral principles, but the ability to recognize and evaluate a vast array of facts concerning humans (and other sentient beings). To mention just one example: Human morally relevant preferences (like the preference to pursue various benefits and avoid various harms), vary not only across different groups of individuals (young/elderly, male/female, healthy/sick), but also within the same individual over time (that is, as the individual grows, matures and acquires new knowledge and experience). It does not seem probable that any machine that lacks emotions like empathy (which is of importance for assessing possible physical and psychological harms and benefits) could deal with this variation of morally relevant facts and preferences.

114. Using robots in most areas of life is generally considered acceptable as long as they perform their tasks better, and commit less errors, than humans would. The same logic could plausibly apply when it comes to using robots as 'implicit' ethical agents or artificial 'moral advisors', providing thus assistance to humans who remain the final decision-makers in complex moral dilemmas (e.g. in cases like natural disasters when limited resources need to be optimally distributed). Even this limited use of robots, however, is not without potential problems. It could have the undesirable effect, for example, that humans in such situations will

be less inclined to do the moral reasoning themselves and more inclined to delegate it too frequently and too excessively to their ‘artificial advisors’. A related danger is that, in time, various professionals (e.g. medical doctors or military commanders) might lose their moral sensitivity or the ability for critical moral reasoning, as it was the case with human calculating and writing abilities after the spread of pocket calculators and personal computers.

115. When it comes to creation of robots as ‘explicit’ ethical agents (i.e. agents capable of general moral decision-making and independent implementation of those decisions), despite the fact that their possibility is still speculative, a particular ethical attention is needed. The existence of such robots, hypothetically speaking, could be desirable in many situations that involve complex ethical problems, assuming they will ‘ethically outperform’ humans due to their lack of typically human moral weaknesses. However, as we have seen, it is highly questionable whether such ‘explicit’ ethical agents will ever become technically possible. It is unclear, moreover, at which point it could be said that a machine became an ‘explicit’ ethical agent. Since the Turing Test itself, as the theoretically best proposal for testing the existence of artificial intelligence, remains controversial, it is even more controversial to claim that some analogue test (‘Moral Turing Test’) could be devised to test the possible creation of ‘explicit’ ethical agents (discussed in Allen et al., 2000; Allen et al., 2011). Since any new piece of technology needs to be rigidly tested before being introduced into the market, it is not clear how any potential ‘explicit ethical agent’ could be tested and, actually, pass the test.

116. The problem of testing and verifying the existence of artificial and ‘explicit’ ethical agents is related to the second important question concerning such agents: If constructing ‘explicit’ ethical agents ever becomes possible, which ethical code should they be programmed with? Asimov’s Three Laws of Robotics are frequently considered as the standard answer to this question. However, it is generally accepted among ‘machine ethics’ scholars (and actually illustrated by Asimov’s stories) that such laws are too general, potentially contradictory and non-applicable in the real-world. Moreover, according to Abney (2012: 41) programming future robots with any top-down moral theory like deontology or consequentialism is implausible due to potential conflicts of duty and/or the inability of robots to calculate the long-term consequences of their actions (computational traceability problem). He proposes, therefore, developing a ‘robot virtue ethics’ focused on ‘the search for the virtues a good (properly functioning) robot would evince, given its appropriate roles’ (Abney, 2012: 38).

117. Among other proposals currently discussed are Kantianism (Powers 2011), virtue ethics (Abney, 2012), pluralistic theory of *prima facie* duties (Anderson and Anderson, 2011b), Buddhist ethics (Hughes, 2012) and divine-command ethics (Bringsjord and Taylor, 2012). However, there is obviously no consensus in sight as to which, if any, of these theories should be programmed into future ‘artificial ethical agents’, if they ever become technically possible. One of the main reasons for this theoretical impasse, of course, is the fact that there are still no definitive philosophical answers to the question of what morality is and what, if anything, does it mean to act morally.

II.6.2. Moral status of robots

118. An intriguing question regarding robots with enhanced autonomy and capacity for decision-making (possibly even moral decision-making) concerns their moral status. Will such robots become morally valuable in themselves – not just instrumentally valuable as mere devices manufactured to perform specific tasks? Would such robots deserve the same moral respect and immunity from harm (having certain moral rights) as it currently is the case with humans and some non-human animals?

119. How moral status is acquired and lost is a longstanding philosophical issue. Many philosophers believe that having moral status amounts to having certain psychological and/or biological properties. From the deontological point of view, to have moral status implies being a person, and being a person implies having rationality or the capacity for rational deliberation. In so far as they are able to solve many demanding cognitive tasks on their own, robots may be said to have some form of rationality. However, it is highly counterintuitive to call them

‘persons’ as long as they do not possess some additional qualities typically associated with human persons, such as the freedom of the will, intentionality, self-consciousness or a sense of personal identity. It should be mentioned in context, however, that the Committee on Legal Affairs of the European Parliament, in its 2016 *Draft Report with recommendations to the Commission on Civil Law Rules on Robotics*, already considers the possibility of ‘creating a specific legal status for robots, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons with specific rights and obligations, including that of making good any damage they may cause, and applying electronic personality to cases where robots make smart autonomous decisions or otherwise interact with third parties independently’ (p. 12).

120. From the utilitarian perspective, moral status does not depend on rationality, but on sentience or the capacity to experience pleasure and pain (broadly construed) and the accompanying emotions. According to this view, humans and many non-human animals have moral status, but robots do not because they are not sentient and lack emotions. According to Torrance (2011), genuine sentience can be ascribed only to organic beings, not to robots. Although sentient and/or emotional robots still do not exist, there is a growing research interest for ‘artificial’ or ‘synthetic’ emotions that might be programmed into future ‘sociable robots’ (Valverdu and Casacuberta, 2009). Depending on future advances in this research area, one should not exclude the possibility of future robots’ sentience, emotions and, accordingly, moral status.

121. In debates about moral status of robots (Floridi, 2011), useful analytical distinction is made between *moral agents* (beings capable to morally act on their own and to treat others in morally wrong or right way) and *moral patients* (beings incapable to morally act on their own, but capable of being acted upon in morally wrong or right way). For example, most humans are both moral agents and moral patients, whereas some humans (like people in coma) and some non-human animals are just moral patients. It is widely accepted, namely, that some non-human animals do have certain moral rights (like the right not to be tortured), despite the fact that they are unable to have any moral duties. The appearance of robots with enhanced autonomy and capacity for decision-making is likely to complicate this (already vague) classification. If robots as ‘explicit’ ethical agents ever become possible, they would clearly fall into the category of moral agents. However, it is less clear whether they would also fall into the category of moral patients, because it is unclear what could actually constitute a ‘moral wrong’ to be possibly inflicted upon a robot (for example, it would have no effect to morally praise or blame the emotionless robot or to threaten it with some kind of moral or legal sanction).

122. A possible third way of assigning moral status to robots (the way that does not focus on any particular psychological or biological property) is to adopt certain ‘relational perspective’, according to which robots would possess moral status in so far as they participate in unique, possibly existentially significant, relationships with human beings. However, this ‘relational’ solution faces the problem of also depending on human psychological tendency to anthropomorphize or ‘project’ human properties to inanimate objects and artefacts. Just as humans tend to develop strong and life-long attachment to their cars, boats or guitars, they are even more likely to develop such attachment to robots, especially those that possess a significant number of human-like traits. For example, there have been cases when human soldiers developed strong bonds with bomb-disposing robots, to the extent that they were crying when such robots were destroyed in the line of duty (Lin, 2012: 11). It remains unclear, in other words, what would be the crucial difference in moral status between artefacts like cars, boats or guitars, on one hand, and robots, on the other. In the absence of some fundamental (preferably non-subjective) criterion, the ‘relational’ solution to the problem of the moral status of robots does not seem too promising.

123. The rapid development of highly intelligent autonomous robots is likely to challenge our current classification of beings according to their moral status, in the same or maybe even more profound way as it happened with non-human animals through the animal rights movement. It may even alter the way in which human moral status is currently perceived.

Although still resembling futuristic speculations, questions like these should not be dismissed lightly, especially in view of the fact that ‘human-machine divide’ is gradually disappearing (Romportl, Zackova and Kelemen, 2015) and the likelihood of future appearance of human-machine hybrids or cyborgs (robots integrated with biological organisms).

II.6.3. Techno-optimism and bioconservatism

124. Robotizing human tasks and activities, maximizing the human-machine hybridity are part of a broad movement of techno-optimism that includes transhumanism and post-humanism ideas, scenario of science-fiction, which is related to technological innovations discourse in a flourishing capitalist economy that promise the general welfare and individual enhancement (Ford, 2015; Hottois et al., 2015). This movement proposed a vision of human as if he/she is a machine, without creativity and as a standardize model.

125. To this utopia, another movement of bioconservatism are suspicious about all technologies and neo-liberal economy. It expresses a profound pessimistic vision of the future of humankind and the planet. Western modern civilization will collapse and is already in a phase of decline. There is no way to escape our destiny, there is no redemption or Saviour. What is left now is wars with drones and robots as soldier, ruins, terrorism as an expression of a conflicting humanity.

126. These two movements - a wonderful land and a nightmare -, doesn’t provide any moral criteria and a worldview to address the actual development of robots. Instead some important questions deserve to be ask in order to orient the technological development of robots:

a. What kind of robots to we want to allow to get into our intimacy? In our oïkos (our house, community) *what* or *who* do we want to integrate and in which kind of relationship (educating children, caring for elderly, cleaning and cooking robots, home automation)?

b. In our society to which extend to we want to delegate our work to robot? What will be the social and cultural transformations? Will it respect human dignity? Will it generate more inequality and exclusion? What kind of dependency to robots will it generate (Carr, 2014)?

c. At a national and international level, to what extend autonomous robots should be used to kill people at war or for security purpose?

d. More generally, is it possible to identity when robots threaten human autonomy and the quality of human relationship?

II. NORMATIVE PART

(Comment: This part to be discussed during and completed after the 9th Extraordinary Session of COMEST in September 2016.)

BIBLIOGRAPHY

Abney, K. 2012. “Robotics, ethical theory, and metaethics: a guide for the perplexed.”, in P. Lin, K. Abney and G. A. Bekey (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, London, pp. 35-52.

Allen, C., Varner, G. and Zinser, J. 2000. “Prolegomena to any future artificial moral agent”. *Experimental and Theoretical Artificial Intelligence,* 12(3): 251-261.

Allen, C., Wallach, W. and Smit, I. 2011. “Why machine ethics”, in M. Anderson and S. L. Anderson (eds.), *Machine Ethics*, Cambridge University Press, Cambridge, pp. 51-61.

Allistene. 2016. *Éthique de la recherché en robotique*, Rapport no1., Commission de réflexion sur l’éthique de la recherche en sciences et technologies du numérique d’Allistene.

Anderson, M. and Anderson, S. L. 2011a. “General introduction”, in M. Anderson and S. L. Anderson (eds.), *Machine Ethics*, Cambridge University Press, Cambridge, pp. 1-4.

Anderson, M. and Anderson, S. L. 2011b. “A *prima facie* duty approach to machine ethics: machine learning of features of ethical dilemmas, *prima facie* duties, and decision principles through a dialogue with ethicists”, in M. Anderson and S. L. Anderson (eds.), *Machine Ethics*. Cambridge University Press, Cambridge, pp. 476-492.

Angelo, J. A. 2007. *Robotics: A Reference Guide to the New Technology*, Greenwood Press, Westport.

Audouze J., Chapouthier G., Laming D., and Oudeyer J-Y. (dir.). 2015. *Mondes mosaiques*, CNRS Editions, Paris.

Asaro, P. M. 2012. “A body to kick, but still no soul to damn: legal perspectives on robotics”, in P. Lin, K. Abney and G. A. Bekey (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, London, pp. 169-186.

Balding, C. 2016/ *Will Robots Ravage the Developing World?* Available at: http://www.bloomberg.com/view/articles/2016-07-25/will-robots-ravage-the-developing-world

Bar-Cohen, Y. and Hanson, D. 2009. *The Coming Robot Revolution: Expectations and Fears About Emerging Intelligent, Humanlike Machines*. Springer.

Bekey, G. A. 2012. “Current trends in robotics: technology and ethics”, in P. Lin, K. Abney and G. A. Bekey (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics*, MIT Press, London, pp. 17-34.

Brynjolfsson E., McAfee A. 2014. *The second Machine. Work, Progress and prosperity in a tume of brilliant technologies,* WW Norton.

Bringsjord, S. and Taylor, J. 2012. “The divine-command approach to robot ethics”, in P. Lin, K. Abney and G. A. Bekey (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics*. MIT Press, London, pp. 85-108.

Broadbent, E., Stafford, R., MacDonald, B. 2009. “Acceptance of Healthcare Robots for the Older Population: Review and Future Directions”, *International Journal of Social Robotics,* 1: 319–330, Springer Netherlands.

Carr, N. 2014. *The Glass Cage: Automation and Us*. W.W. Norton and Company, New York.

Chamayou, G. 2015. *A Theory of the Drone*. The New Press, New York.

Chapouthier G., Kaplan F. 2011. *L'homme, l'animal et la machine, perpétuelles redéfinitions,* Paris, CNRS Editions, Paris,

Copeland, J. 1993. *Artificial Intelligence: Philosophical Introduction*. Wiley-Blackwell.

Eguchi, A. 2012. “Educational Robotics Theories and Practice: Tips for how to do it Right”, in Barker S.B. (Ed.) *Robots in K-12 Education*, IGI Global, Hershey PA.

Floridi, L. 2011. “On the morality of artificial agents”, in M. Anderson and S. L. Anderson (eds.), *Machine Ethics*, Cambridge University Press, Cambridge, pp. 184-212.

Ford Martin. 2015. *Rise of the Robots: Technlogy and the Threat of Jobless Future*, Basic Books, New York.

Franklin, S. 2014. “History, motivations, and core themes”, in K. Frankish and W. M. Ramsey (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, Cambridge.

Frey, C.B, Osborne, M. 2013. *The Future of Employment*. University of Oxford, Oxford UK.

Gagnon, LO,Goldenberg L., Lynch K., and coll. 2014. “Comparision of open and robotic-assisted prostatectomy”. *Canadian Urology Association Journal,* 8: 92-7.

Gibilisco, S. 2003. *Concise Encyclopedia of Robotics*. McGraw-Hill, New York.

Hottois G., Missa J-N., Perbal L (eds). 2015. Enclyclopédie du trans/posthumanisme, Paris,Vrin.

Hughes, J. (2012), “Compassionate AI and selfless robots: a Buddhist approach” in P. Lin, K. Abney and G. A. Bekey (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics*. MIT Press, London, pp. 69-84.

Husbands, P. 2014. “Robotics”, in K. Frankish and W. M. Ramsey (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, Cambridge.

Ingram, B., Jones, D., Lewis, A., Richards, M., Rich, C. and Schachterle, L. 2010. “A code of ethics for robotics engineers” in *Procedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

Kappor Anil. 2014. “L’invasion robotique au Canada”. *Canadian Urology Association Journal,* 8: 5-6.

Lecourt, D. 2003. *Humain, post-humain*. PUF, Paris.

Lin, P. 2012. “Introduction to robot ethics”, I, P. Lin, K. Abney and G. A. Bekey (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics*. MIT Press, London, pp. 3-16.

Moor, J. 2011. “The nature, importance, and difficulty of machine ethics”, in M. Anderson and S. L. Anderson (eds.), *Machine Ethics*. Cambridge University Press, Cambridge, pp. 13-20.

Murashov, V., Hearl, F., and Howard, J. 2015. *A Robot May Not Injure a Worker: Working safely with robots*. Available at: http://blogs.cdc.gov/niosh-science-blog/2015/11/20/working-with-robots/

Murphy, R. R. 2000. *Introduction to AI Robotics*, The MIT Press, Cambridge, Mass.

OECD Insights. 2016. *The rise of the robots – friend or foe for developing countries?* Available at: http://oecdinsights.org/2016/03/02/the-rise-of-the-robots-friend-or-foe-for-developing-countries/

Oxford Dictionary of Computer Science. 2016. Oxford University Press, Oxford.

Oost, E., Reed, D. 2010. “Towards a Sociological Understanding of Robots as Companions”, in Lamers, Verbeek (Eds.), *Human-Robot Personal Relationships*, Springer, Berlin Heidelberg.

Pearson,Y., Borenstein, J. 2014. *Creating ‘‘companions’’ for children: the ethics of designing aesthetic features for robots*, AI & Society, 29:23–31, Springer, London.

Peláez, L. 2014. *The Robotics Divide. A New Frontier in the 21st Century?,* Springer-Verlag, London.

Powers. T. M. 2011. “Prospects for a Kantian machine”, in M. Anderson and S. L. Anderson (eds.), *Machine Ethics*. Cambridge University Press, Cambridge, pp. 464-475.

Riek, L. D. and Howard, D. 2014. “A code of ethics for the human-robot interaction profession”, in Proceedings of *We Robot 2014*. Available at: http://robots.law.miami.edu/2014/wp-content/uploads/2014/03/a-code-of-ethics-for-the-human-robot-interaction-profession-riek-howard.pdf

Rifkin, J. 1995. *The End of Work.* Putnam, New York

Romportl, J., Zackova, E. and Kelemen, J. (eds.). 2015. *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide*, Springer.

Rosenberg, J. M. 1986. *Dictionary of Artificial Intelligence and Robotics*, John Wiley & Sons, New York.

Savulescu, J. and Maslen, H. 2015. “Moral enhancement and artificial intelligence: moral AI?”, in J. Romportl, E. Zackova and J. Kelemen (eds.), *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide*, Springer, pp. 79-96.

Siciliano, B, Khatib, O. (Eds.). 2008. *Springer Handbook of Robotics*, Springer-Verlag, Berlin, Heidelberg

Stone, W. L. 2005. “The history of robotics” in T. R. Kurfess (ed.), *Robotics and Automation Handbook*, CRC Press, Boca Raton.

Torrance, S. 2011. “Machine ethics and the idea of a more-than-human moral world” in M. Anderson and S. L. Anderson (eds.), *Machine Ethics*. Cambridge University Press, Cambridge, pp. 115-137.

Valverdu J. and Casacuberta D. (eds.). 2009. *Handbook of Research on Synthetics Emotions and Sociable Robots: New Applications in Affective Computing and Artificial Intelligence*. IGI Publishing, Hershey, PA.

Van Rysewyk, S. P. and Pontier, M. 2015. “Preface”, in S. P. van Rysewyk and M. Pontier (eds.), *Machine Medical Ethics*, Springer, pp. v-x.

Veruggio, G. and Operto, F. 2008. “Roboethics: social and ethical implications of robotics”, in B. Siciliano and O. Khatib (eds.), *Springer Handbook of Robotics*. Springer, pp. 1499-1524.

Wallach, W. and Allen, C. 2009. *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press, Oxford.

Warwick, K. 2012. *Artificial Intelligence: The Basics*. Routledge, New York.

Whitby, B. 2011. “On computable morality: an examination of machines as moral advisors”. in M. Anderson and S. L. Anderson (eds.), *Machine Ethics*. Cambridge University Press, Cambridge, pp. 138-150.

Wise, E. 2005. *Robotics Demystified*. McGraw-Hill, New York.