United Nations
Educational, Scientific and
Cultural Organization

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

SHS/YES/COMEST-10/17/2 REV.
Paris, 14 September 2017
Original: English

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 COMEST in September 2015, the Commission 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 a preliminary draft report, which was discussed during the 9th Extraordinary Session of COMEST in September 2016. At that session, the content of the preliminary draft report was further refined and expanded, and the Working Group continued its work through email exchanges. The COMEST Working Group then met in Quebec in March 2017 to further develop its text. A revised text in the form of a draft report was submitted to COMEST and the IBC in June 2017 for comments. The draft report was then revised based on the comments received. The final draft of the report was further discussed and revised during the 10th (Ordinary) Session of COMEST, and was adopted by the Commission on 14 September 2017.

This document does not pretend to be exhaustive and does not necessarily represent the views of the Member States of UNESCO.

REPORT OF COMEST ON ROBOTICS ETHICS

REPORT OF COMEST ON ROBOTICS ETHICS

EXECUTIVE SUMMARY

I. INTRODUCTION

Robots can help humanity, and they have done so since the mid-20th century. While initially being mostly used for industrial and military applications, they are currently emerging in other areas, such as transportation, healthcare, education, and the home environment. Contemporary robotics is increasingly based on artificial intelligence (AI) technology, with human-like abilities in sensing, language, interaction, problem solving, learning, and even creativity. The main feature of such 'cognitive machines' is that their decisions are unpredictable, and their actions depend on stochastic situations and on experience. The question of accountability of actions of cognitive robots is therefore crucial.

The rapidly increasing presence of cognitive robots in society is becoming more challenging. They affect human behaviours and induce social and cultural changes, while also generating issues related to safety, privacy and human dignity. This report aims to raise awareness and promote public consideration and inclusive dialogue on ethical issues concerning the different use of autonomous, cognitive robots in society.

II. WHAT IS A ROBOT?

Contemporary robots can be characterized by four central features:

- *mobility*, which is important to function in human environments like hospitals and offices;
- *interactivity,* made possible by sensors and actuators, which gather relevant information from the environment and enable a robot to act upon this environment;
- *communication*, made possible *by* computer interfaces or voice recognition and speech synthesis systems; and
- *autonomy,* in the sense of an ability to 'think' for themselves and make their own decisions to act upon the environment, without direct external control.

Contemporary robotics typically includes forms of 'Artificial Intelligence' (AI): replicating human cognition and intelligence with computer systems, resulting in machines that can do things that require a specific form of intelligence, like the ability to perceive and represent changes in their environment and to plan its functioning accordingly. Artificial intelligence is crucial for robot autonomy because it enables them to perform complex tasks in changing and unstructured environments, like driving a car and adapting to the conditions on the road, without being tele-operated or controlled by a human operator.

Robots perform their tasks through algorithms: rules or instructions for the solution of a problem. Two kinds of algorithms can be distinguished: deterministic algorithms, that control the predictive behaviour of *deterministic* robots; and AI or stochastic algorithms, with learning abilities that form the heart of *cognitive* robots. A deterministic robot's behaviour – even if the robot is highly complex and autonomous (requires little or no human supervision) – is basically pre-programmed and essentially determined. However, AI-based, cognitive robots will learn from past experiences and calibrate their algorithms themselves, so their behaviour will not be perfectly predictable, and will likely become an issue worthy of serious ethical attention and reflection.

III. ROBOTS AND SOCIETY

In *industry*, robots gradually substitute for labour in a wide range of service occupations, where most job growth has occurred over the past decades. Without structural adjustments and compensations, this could lead both to higher unemployment and rising inequality in society. Robots will bring profound changes to working conditions and job transformations.

Working ‘side by side’ with robots demands new working skills, safety measures, working schedules, and education. Also, industrial robotization brings economic and political challenges: could it bring about a new divide between developing and developed countries, and if so, how can we confront this situation?

A class of *military robotic systems* are often indicated as ‘drones’. A drone can be controlled either from a distance by human operators (remote control) or by robotic means (automatic piloting); present-day drones may combine those two modes of control. The physical disconnect between pilot and field of action can lead to a gaming mentality, creating a moral buffer from the action, but the possibility to be present on a distance can also lead to a stronger moral experience of soldiers. There are concerns as to whether the International Humanitarian Law (IHL) framework, applicable to situations of armed conflict and occupation, is sufficiently clear in the light of drones. These issues are even more problematical in the case of fully autonomous weapons systems that are under development.

In the field of *transportation*, the autonomous vehicle (AV) is almost ready to enter society. Autonomous vehicles could strongly reduce the number of car accidents and make transportation more efficient, while potentially causing job losses and widening the ‘robotic divide’. A central ethical issue here is the decision-making processes that are built into AVs. How should the car be programmed to act in the event of an unavoidable accident? Should it minimize the loss of life, even if it means sacrificing the occupants, or should it protect the occupants at all costs? And should such issues be regulated by laws, by standards, by codes of conduct?

In the realm of *health and well-being*, robots are increasingly used in surgery, sometimes resulting in higher precision, but also in higher costs, and a different surgical practice. Robots are used also in therapeutic, rehabilitative, and elderly care settings, for instance for children with autism, exoskeletons to manage spinal injuries, and social robots for the elderly. Ethical issues here typically concern the appropriateness of using these types of technologies in relation to care: can robots provide care, what implications do they have for safety and security, and how do they influence our attitude towards ability and disability, and towards care for people who are ill, old, and vulnerable? A special subcategory here is formed by companion robots in the realm of sexuality: how will the possibility have a sexual relation at a distance via a robot, or even a sexual relation with a robot, affect our values surrounding love and intimacy?

Robots are also entering the field of *education*. Educational robots can support individual and collaborative learning activities. Introducing robots in the classroom might have implications for how children learn, for the (role) model of the teacher, and for the emotional development of children. In the *household*, ‘service robots’ are assisting humans in everyday activities like vacuuming, garbage collecting, window washing, ironing, and preparing food, potentially affecting the quality of life and the definition of gender roles.

And, finally, robots are also used in *farming*. Such robots change practices of farming and the relations between humans and animals: sensor networks can monitor the living conditions of animals, making it possible to provide tailor-made care while potentially reducing the interaction between humans and animals. In *agriculture*, drones are being used in ‘precision agriculture’ to optimize food productivity and quality by collecting and analysing data for scientific interventions (e.g., optimal use of fertilizer, drop by drop irrigation, etc.).

More globally, the potential benefits of robots need to be balanced against the environmental impact of the entire robot production cycle. Furthermore, robotics is likely to add to the growing concerns about the increasing volumes of e-waste, especially in developing countries.

IV. ETHICAL AND LEGAL REGULATION

Robotics and robots raise new and unprecedented legal and ethical challenges. Given the complexity of the design, construction and programming of robots, a central ethical issue is 'traceability': the possibility to track the causes of all past actions (and omissions) of a robot. For robots with a high level of autonomy, decision-making capacities and learning abilities, this traceability requirement is problematic, though: such robots are not merely programmed to perform specific tasks, but to learn and further develop themselves in interaction with their environment, which requires modifications to current legal and ethical notions of 'traceability'.

One of the most concrete initiatives for future legal and ethical regulation of robotics and robot industry is the Draft Report with recommendations to the Commission on Civil Law Rules on Robotics, issued in 2016 by the Committee on Legal Affairs of the European Parliament. Concerned about 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.

In 2017, the Rathenau Instituut issued a report on 'Human rights in the robot age', commissioned by the Parliamentary Assembly of the Council of Europe (PACE). The report addresses the potentially negative impact of robotics on a number of issues related to human rights, including respect for private life, human dignity, ownership, safety and liability, freedom of expression, prohibition of discrimination, access to justice and access to a fair trial. It recommends the introduction of two novel human rights: the right not to be measured, analysed or coached (related to possible misuses of AI, data gathering and the Internet of Things) and the right to meaningful human contact (related to possible misuses, intentional and unintentional, of care robots).

According to Leenes et al (2017) the field of robotics as such seems to face four regulatory dilemmas: (1) the dilemma of keeping up with rapid technological advances; (2) the dilemma of striking a balance between stimulating (or at least not impeding) innovation and protecting human fundamental rights and values; (3) the dilemma between affirming existing social norms or nudging those norms in new and different directions; and (4) the dilemma of balancing between effectiveness and legitimacy in techno-regulation.

V. ETHICAL CHALLENGES

Robotic technologies blur the boundary between human subjects and technological objects. In doing so, they do not only have societal implications which can be ethically evaluated, but they also affect the central categories of ethics: our concepts of agency and responsibility, and our value frameworks.

Given the increasing autonomy of robots, the question arises who exactly should bear ethical and/or legal *responsibility* for robot behaviour. There typically seems to be a 'shared' or 'distributed' responsibility between robot designers, engineers, programmers, manufacturers, investors, sellers and users. None of these agents can be indicated as the ultimate source of action. At the same time, this solution tends to dilute the notion of responsibility altogether: if everybody has a part in the total responsibility, no one is fully responsible. Avoiding the potential paralyzing effect of this difficulty to take and attribute responsibility, then, is a major challenge for the ethics of robotics. In order to take responsibility anyway, one solution can be to develop techniques to anticipate the impacts of robotic development as much as possible (Waelbers and Swierstra, 2014; Verbeek, 2013). Another solution is to deal carefully with the inevitable occurrence of unexpected implications, by considering the societal introduction of robotic technologies as a 'social experiment' that needs to be conducted with great care (Van de Poel, 2013).

Because of their ability to act autonomously, robots also problematize our notion of *agency*. Even though there are clear differences between human agency and robotic 'agency', robots also 'do' things that are the results of their own decision-making processes and interactions,

and not only of the input given by their developers, which has implications for our understanding of moral agency. The main question, then, is how they change human practices, and how the quality of human-robot relations can guide the design, implementation and use of robots. Another way to deal with the issue of moral agency is offered by the emerging discipline of 'machine ethics', which aims to equip machines with ethical principles or procedures for resolving ethical dilemmas, in order to function in an ethically responsible way. A related disruptive aspect of robotic technologies concerns their *moral status*. Will robots ultimately become morally valuable, beyond their instrumental value as devices merely manufactured to perform specific tasks? Would such robots deserve moral respect and immunity from harm, and not only have obligations and duties, but also moral rights?

A final disruptive effect of robotic technologies is their impact on moral frameworks: they do not only have societal effects that can be ethically evaluated, but they also affect the very ethical frameworks with which we can evaluate them. Care robots might change what humans value in care, while teaching robots might affect our criteria for good education and sex robots could have an impact on what we value in love and intimate relations. Dealing with such normative impacts in a responsible way requires a careful balance between anticipation and experimentation, closely following the impact of robotic technologies on value frameworks, in small-scale experimental settings, in order to be able to take this impact into account in design practices, public discussions, and policy-making.

VI. RECOMMENDATIONS

In considering recommendations regarding robotics ethics, this distinction between deterministic and cognitive robots is important. In the deterministic case, the behaviour of the robot is determined by the program that controls its actions. Responsibility for its actions is therefore clear, and regulation can largely be dealt with by legal means. In the cognitive case, a robot's decisions and actions can be only statistically estimated, and are therefore unpredictable. As such, the responsibility for the robot's actions is unclear and its behaviour in environments that are outside those it experienced during learning (and so in essence 'random') can be potentially catastrophic. So assigning responsibility for the actions of what is partly a *stochastic* machine is problematical.

Accordingly, COMEST proposes to consider recommendations based upon the above. In the first level of deterministic machines where the responsibility for behaviour can be assigned, the Commission's recommendations will largely focus on legal instruments to regulate their use. For the second level of cognitive machines, whose behaviour cannot be 100% predictable and therefore is in significant part stochastic, in addition to legal instruments, codes of practice and ethical guidelines for both producers and users need to be considered. Finally, where stochastic machines can be put in situations where harm can be caused (for example a self-driving car or an autonomous weapon), we need to consider the degree of autonomy that can reasonably be left to the machine, and where meaningful human control must be retained.

The scheme is illustrated in the table below. While the proposed structure is simple, its implementation in terms of assigning accountability and regulating use is complex and challenging – for scientists and engineers, policy makers and ethicists alike.

Decision by Robot	**Human Involvement**	**Technology**	**Responsibility**	**Regulation**
Made out of finite set of options, according to preset strict **criteria**	Criteria implemented in a legal framework	Machine only: deterministic algorithms/ robots	Robots' producer	Legal (standards, national or international legislation)
Out of a range of options, with room for flexibility, according to a preset **policy**	Decision delegated to robot	Machine only: AI- based algorithms, cognitive robots	Designer, Manufacturer, Seller, User	Codes of practice both for engineers and for users; Precautionary Principle
Decisions made through human-machine **interaction**	Human controls robot's decisions	Ability for human to take control over robot in cases where robot's actions can cause serious harm or death	Human beings	Moral

In regard of the diversity and the complexity of robots, a framework of ethical values and principles can be helpful to set regulations at every level – conception, fabrication and utilization – and in a coherent manner, from engineers' codes of conduct to national laws and international conventions. The principle of human responsibility is the common thread that joins the different values that are enunciated in the report. These relevant ethical principles and values include: (i) human dignity; (ii) value of autonomy; (iii) value of privacy; (iv) 'do not harm' principle; (v) principle of responsibility; (vi) value of beneficence; and (v) value of justice.

The specific recommendations on robotics ethics identified by COMEST are as follow:

i. It is recommended that, at both the national and international levels, codes of ethics for roboticists be further developed, implemented, revised and updated, in a multidisciplinary way, and responding to possible future advancements of robotics and its impact on human life and the environment (energy, e-waste, ecological footprint). It is also recommended that disciplines and professions significantly contributing to or potentially relying on robotics – from electronic engineering and artificial intelligence to medicine, animal science, and psychology and the physical sciences – revise their particular codes of ethics, anticipating challenges originating from their links to robotics and the robot industry, preferably in a coordinated way. Moreover, it is recommended that ethics – including codes of ethics, codes of conduct, and other relevant documents – become an integrated part of the study programmes for all professionals involved in the design and manufacturing of robots.

ii. It is recommended that ethics needs to be part of the design process of robotic technologies, building on approaches like the Value Sensitive Design approach.

iii. In order to deal responsibly with the social introduction of robots, it is recommended that new robotic technologies be introduced carefully and transparently in small-scale, well-monitored settings, and the implications of these technologies on human practices, experiences, interpretational frameworks, and values be studied openly. The outcomes of such experiments can be used to adapt the design of robots, to inform policy-making and regulation, and to equip users with a critical perspective.

iv. It is recommended that public discussions be organized about the implications of new robotic technologies for the various dimensions of society and everyday life, including

environmental impact of the entire robot production cycle, in order to help people to develop a critical attitude, and to sharpen the awareness of designers and policy-makers.

v. It is recommended that States, professional organizations and educational institutions consider the implications of robotics on the reduction in job opportunities, and in the creation of new job opportunities, paying particular attention to those sections of society likely to be most vulnerable to the changes, and make appropriate provision for retraining and retooling of the work force to enable the potential advantages to be realized.

vi. With respect to autonomous vehicles, whose unique features are their ability to operate and decide based on their machine learning, cognitive algorithms, it is recommended that situations where the responsibility for the results of the AV's action is put solely on a human ('the driver') be identified and defined.

vii. The ethical issues of using armed drones go beyond the legal issues of International Humanitarian Law. The use of armed drones against suspected non-state actors in insurgences raises additional ethical and legal questions. COMEST concludes therefore that, in addition to legal issues, there is a strong moral principle against an armed robot killing a human being, either in declared armed conflict or in counterinsurgency operations. It is recommended that States reconsider this practice.

viii. With regard to autonomous weapons, it is strongly recommended that, for legal, ethical and military-operational reasons, human control over weapon systems and the use of force must be retained. Considering the potential speed of development of autonomous weapons, there is an urgent need to (as the ICRC has urged) "determine the kind and degree of human control over the operation of weapon systems that are deemed necessary to comply with legal obligations and to satisfy ethical and societal considerations" (ICRC, 2016).

ix. It is recommended that States should draw up policies on the use of drones in surveillance, policing, and the use of drones in non-military contexts. Usage policy by police should be decided by the public's representatives, not by police departments, and the policies should be clear, written, and open to the public. This policy should, at minimum, assure that a drone is 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. 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. Use of drones should be subject to open audits and proper oversight to prevent misuse. Drones in police use should not be equipped with either lethal or non-lethal weapons. Autonomous weapons should not be used in police or security use.

x. It is recommended that the private use of drones should be under licence, and their areas of operation subject to strict control for safety, privacy and legal reasons. It should be unlawful to equip domestic drones with either lethal or non-lethal weapons.

xi. It is recommended that particular attention should be paid to gender issues and stereotyping with reference to all types of robots described in this report, and in particular, toy robots, sex companions, and job replacements.

xii. Similar to other advanced technologies, it is recommended that environmental impact should be considered as part of a lifecycle analysis, to enable a more holistic assessment of whether a specific use of robotics will provide more good than harm for society. It is also recommended that while constructing robots (nano, micro or macro), efforts should be made to use degradable materials and environmentally friendly technology, and to improve the recycling of materials.

xiii. It is recommended that COMEST addresses the ethical challenges of the Internet of Things (IoT) as an extension of its work on this report.

REPORT OF COMEST ON ROBOTICS ETHICS

I. INTRODUCTION

1. Robots can help humanity, and they have done so since the mid-20th century. During the first 50 years, robots were mostly used for industrial applications, where they are integrated in factories to liberate human beings from routine tasks. Traditionally, robots are also used for military applications. In recent years we have seen robots emerging into other areas, such as transportation, healthcare, education, and home (service robots), where their interaction with society is more visible.

2. From do-it-yourself (DIY) robots, to drones, home robots, humanoid robots, industrial, medical and military robots – modern robotics is increasingly based on artificial intelligence (AI) technology. This technology, also referred to as 'cognitive computing', is capable of human-like sensory, language, interaction, analytics, problem solving, and even creativity. AI-based machines can demonstrate human-like learning capability and can become independent learners. These capabilities are being dramatically improved by employing deep learning algorithms and cloud computing.

3. With advanced machine learning technology, robots can become cognitive robots, which can learn from experience, from human teachers, and even on their own, thereby potentially developing an ability to interact with their environment. The main feature of cognitive machines is that their decisions are unpredictable, and their actions depend on stochastic situations and on experience. The question of accountability of actions of cognitive robots is therefore crucial. The presence of cognitive robots at home, in the workplace, and more generally in society is becoming more challenging. Such presence affects human behaviours and induces profound social and cultural changes (family and community relationship, workplace, role of the state, etc.), as well as issues related to safety, privacy and human dignity. Moreover, emerging technologies such as the Internet of Things (IoT) will amplify the ethical dilemmas.

4. Robots regularly raise ethical concerns, especially when it comes to issues where they might replace human beings, or take on roles that are normally reserved for humans. To what extent should robots replace human labour? Is it morally acceptable to give robots a role in warfare? Do we want robots to provide care to elderly people, or autistic children? What are the ethical dimensions of giving robots a role as 'companion', or even as an erotic 'partner'? Is there an ethical way to programme a self-driving car? Such questions are based upon a fundamental concern with human dignity and the possible threats that robotics could impose on it. This report aims to address these and other concerns by exploring in which ways robotic technologies may play a role in society in the near future. Often, in fact, robots are not replacing humans, but rather have a deep influence on human practices: they do not take over healthcare, education, policing, labour, or armed conflict - rather, they change these practices. By investigating what values are at stake in these developments, this report aims to identify the ethical dimensions of robotic technologies in a close relation to both the technological developments themselves and the ethical concerns that arise in ethical and societal discussions.

5. This report aims to raise awareness and promote public consideration and inclusive dialogue on ethical issues concerning the different use of autonomous, cognitive robots in society. As technical advances develop, these issues become more complex and we need to address the relationship between robots and society, both as it is now and how it may be in the future.

6. A comprehensive study of these issues calls for a wide interdisciplinary collaboration. The technological aspects of robotics involve expertise from a wide range of scientific fields: mechanical and control engineers are responsible for the mobility of robots; physicists and electrical engineers take care of their sensing and communication skills; computer scientists and signal processing experts design the cognitive aspects of the robots; and systems

engineers are responsible for the integration. However, the technical aspects of robotics are highly related to functionality in a human dominated environment, so the involvement of experts from various fields in humanities and social science is essential. The working group behind this report is an interdisciplinary team, bringing together ethical, political, social and technological expertise.

7. 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 British Pugwash/University of Birmingham workshop on the 'Ethics of Modern Robotics in Surveillance, Policing and Warfare' held at the University of Birmingham, United Kingdom, from 20 to 22 March 2013. A short report on this workshop, which provided material that we have used in this report, is available on the COMEST website (UNESCO, 2013). At the 8th (Ordinary) Session of COMEST (Bratislava, 27-31 May 2013), a Conference on 'Emerging Ethical Issues of Science and Technology' was organized jointly with British Pugwash with a session devoted to autonomous robots. Finally, at its 9th (Ordinary) Session in September 2015, COMEST decided to work specifically on the ethics of robotics 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 Université de Paris 8. It met again in Laval University in Quebec City from 27 to 31 March 2017 and participated in an International Workshop on 'Robots and Society: What Transformations What Regulations?'. Three members of the team also participated in the 'AI for Good' Global Summit organized by the International Telecommunications Union (ITU) in Geneva from 7 to 9 June 2017, where a number of relevant issues were discussed with AI experts.

II. WHAT IS A ROBOT?

II.1. The complexity of defining a robot

8. 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 introduced by Karel Čapek in his 1920 science fiction (SF) play *R.U.R.* (*Rossumovi Univerzální Roboti* [*Rossum's Universal Robots*]). The term originates from the word 'robota', which means 'work' or 'labour' in Czech. However, focusing on the etymology of the word 'robot' is of little help when it comes to defining what a robot is. 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).

9. 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, p.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, p.309).

10. In his *Encyclopedia of Robotics*, Gibilisco (2003) 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. He further argues that 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.

11. 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" (quoted by Tzafestas, 2016b, p.4). 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. Robotics, namely, is no longer just a matter of mechanical and electric/electronic engineering ('robots'), but also of nanoscience ('nanobots'), biology ('biorobots' or 'cyborgs'), botany ('plantoids') and other disciplines.

II.2. History of robotics – fiction, imaginary and real

12. 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.

13. 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 form strong emotional ties to inanimate anthropomorphic objects.

14. Hebrew mythology contains a number of stories of the Golem, an anthropomorphic creature that, 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 beneficent 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. *One Thousand and One Nights*, a classical collection of mostly folk stories in the Arabic language, contains several stories featuring artificial creatures, such as humanoid robots, robotic animals and various automatons.

15. 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 often 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.

16. As already mentioned above (see paragraph 9), 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, in terms of when the word 'robot' was first used. 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 worldwide. 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).

17. 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!' (first published in 1941, reprinted in Asimov, 1950). 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.

18. 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.

19. 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) pointed out, in Japanese culture "machines (and, in general, human products) are always beneficial and friendly to humanity" (p.6). According to Bar-Cohen and Hanson (2009), 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.

20. 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). Moviemakers 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 defence 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 birth to a clone of herself). *The Terminator* (by James Cameron, 1984) depicts the future world as a dictatorship run by an artificial intelligence with the assistance of an army of robots.

21. Not all SF movies and TV series, of course, portray robots as the ultimate threat to humanity. Many of them present robots in a neutral and even humane light, with all the associated human virtues and vices. Some prominent examples are the robotic duo R2-D2 and C3PO 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 the *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'.

22. 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), 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. Arabic scientists and engineers had similar interests. Probably the most famous was the 13th century scholar and engineer Ismail Al-Jazari who wrote a book of sketches and instructions on how to construct various automata. Al-Jazari himself built many of these automata, such as a drink-serving waitress, a musical band, and automated peacocks (most of them were powered by water).

23. 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.3. Autonomy, interactivity, communication and mobility

24. 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, even if technology works perfectly. This is never the case since failure exists, so concerns are also raised because reliability cannot be perfectly guaranteed. These features are mobility, interactivity, communication and autonomy.

25. 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 a 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' (an 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.

26. 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 that 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 in their environment and thus perform tasks that could be harmful to human beings. It is not ethically irrelevant, for example, how strong a domestic or nursing robot is, because the stronger it is, the greater the harm it might inflict on a human being in case of malfunction. On the other hand, robots may collect data with their sensors that may be used, either intentionally or unintentionally, to harm humans or their privacy (e.g. criminal activities, industrial espionage, and yellow journalism).

27. 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 that are 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), 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, require extremely explicit instructions. According to Gibilisco (2003), 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), both of which are able to communicate with humans through speech and gestures (Bekey, 2012).

28. According to Bekey (2012), "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" (p.18). 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, p.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 increasingly autonomous (performing more complex tasks on their own, without depending thereby on human control or commands). Some degree of 'autonomy' could actually be considered central to a robot, because it is this autonomy that distinguishes it from mere tools or devices that need to be operated by human beings.

29. With respect to increasing autonomy, the question arises as to how close the performance of an autonomous robot should approach that of a human performing the same

task before it is 'let loose' into its working environment. Noting that human performance itself is variable, there is an argument that has been expressed to us that we should hold autonomous robotic systems to a *higher* standard of performance than that which is expected of a human before we hand it the keys to, for example, our car.

30. In the same context, it is worth noting that there is experimental evidence that human-machine teams work better than just human or machine alone. In addition to what this implies for effectiveness, it also gives some safety net against malfunctioning of the robotic partner. Therefore, there are strong arguments for always keeping a human in the loop in such systems.

II.4. Nano-Robotics

31. There is a general agreement on the definition of nano-robotics as a technology of creating machines or robots with components at or close to the scale of a few nano-meters (10^{-9} meters). These devices are named nanobots, nanoids, nanites, nanomachines, or nanomites.

32. They are generally divided into inorganic and organic types, depending on the type of material used in their design. Inorganic nano-robot manufacturing is based on tailored nanoelectronics and semiconductor technology with a considerable complexity of nano scale components (Smith et al., 2012; Weir et al., 2005). Organic nano-robots are based on biomolecular machines, which are inspired by 'nature's way of doing things' at the nanoscale (Ummat et al., 2004). Research efforts are being focused on development of micro-nanoscale or miniature robotic systems, which are biologically inspired, integrated with biological entities, or used for biological or biomedical applications. The majority of natural molecular machines are made from well designed assembly of proteins with pre-programmed biological function in response to specific physiochemical stimuli but in vitro or even in an artificial setting (Mavroidis et al., 2004).

33. Nano-robots' components like motors (e.g. kinesin, RNA polymerase, myosin and adenosine triphosphate (ATP) synthase function as nano-scale biological motors) have drawn a lot of attention because they operate at extremely high efficiencies, could be self-replicating, and they don't need to be invented since they already exist and function in nature, and could be discovered and customized based on our needs (Ummat et al., 2004).

34. Some of the characteristic abilities of organic nanorobots include durability due to their small size; faster operation compared to larger counterparts; ability to be powered by ambient heat or to generate their own power from decomposable products; swarm intelligence, cooperative behaviour, self-assembly and ease of replication; nano information processing and programmability; as well as nano to macro world interface architecture (Weir et al., 2005).

35. Some of the potential applications of nano-robots are the detection and removal of toxic chemicals in the environment. Self-reconfigurablility make them useful in space technology and military services where they may need to handle tasks that we were unable to do in advance. In the medical field, in addition to delivering pharmaceutical products, nanotech medical robots ('nanomedibots') may also be able to: chemically and physically sense and analyse aspects of a human organ at the nanometer level; monitor bodily functions; repair damaged tissue; deconstruct pathological or abnormal material or cells such as cancer or plaque; and enhance human health and functioning.

36. Due to their small size, which makes them invisible to human eyes, a new level of privacy and safety should be seriously considered in protecting human beings as well as the ecosystem. The impact of nanorobots on the economy, cost of manufacturing, design constraints, and related high-end technology are elements of concern.

II.5. Robots, artificial intelligence and algorithms

37. Since to define 'intelligence' itself is difficult, to define 'artificial intelligence' is even more difficult. 'Artificial intelligence' (AI) can generally refer to two interconnected things.

38. It can refer to research on "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, p.7). 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" (quoted by Copeland, 1993, p.1). AI is one of the prime examples of an interdisciplinary research area because it combines numerous and diverse disciplines like computer science, psychology, cognitive science, logic, mathematics, and philosophy.

39. Advancements in robot mobility, communication, interactivity and autonomy became possible thanks to the development of AI. According to Warwick (Warwick, 2012), "an intelligent robot is merely the embodiment of an artificially intelligent entity – giving a body to AI" (p.116). 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. It should be noted, however, that industrial robots are also being continuously improved in order to be as adaptive as possible to unexpected circumstances at the workplace (and as such more cost-effective).

40. 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 a changing (unstructured) environment (e.g. driving a car and adapting to the conditions on the road) without being tele-operated or controlled by a human operator. According to Murphy (2000), "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" (p.26). Whereas tele-operated 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).

41. As Bekey et al. (2008) pointed out, space robots operate in often unpredictable conditions and space roboticists therefore face four issues when designing their robots: (1) mobility (the robot must be able to change its location without endangering human astronauts, equipment or itself); (2) manipulation (the robot needs to use its arms and tools with sufficient precision and force which is not dangerous); (3) time delay (the robot must have the ability to receive commands from distant human operators and to function independently when such commands are unavailable); and (4) extreme environments (the robot must be able to withstand e.g. ionizing radiation, hard vacuum, extreme temperatures, etc.). Since human travel beyond the Moon is still a major safety risk, space robots (in the form of autonomous probes, rovers, landers with robotic arms and manipulators, or even humanoid 'robotic astronauts') are likely to become more sophisticated and play an increasingly important role in future space exploration.

42. The first theoretical proposal of the possibility of artificial intelligence was Alan Turing's paper on '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 program if a human being, in a specific experimental setup (conversation in natural language via a computer terminal), cannot distinguish the responses of the machine or the program from the responses given by a human being. So far, despite many attempts and advances, no machine or computer program that would succeed in this 'imitation game' has been developed (Franklin, 2014). It is generally agreed that the AI research area (as well as the very name AI) was conceived during the so-called 'Dartmouth Conference' organized in 1956 at Dartmouth College by the pioneers of the field such as John McCarthy, Marvin Minsky and others.

43. '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 general AI research, the relationship between AI and robotics is more complex. According to Murphy (2000), "robotics has played a pivotal role in advancing AI" (p.36) and it contributed to the development of all the following 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.

44. 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.

45. The preceding 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 'planning' and ends with 'acting'. However, the 'hierarchical paradigm', which was basically an attempt 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 behaviour). 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 predecessors (Murphy, 2000).

46. Irrespective of the paradigm it belongs to, and irrespective of the differences between their shape, size and ways of movement, robots perform their tasks through algorithms. 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” (Butterfield et al., 2016). Deterministic algorithms control the predictive behaviour of ***deterministic* robots**, while AI-algorithms, with learning abilities, are in the heart of ***cognitive* robots**.

47. It is important to notice that a deterministic robot’s behaviour – even if the robot is highly complex and autonomous (requires little or no human supervision) – is basically pre-programmed 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 a robot is deterministic is relevant (and desirable) when traceability issues arise, i.e. when a robot’s past decisions and actions need to be precisely reconstructed in order resolve some ethical or legal dispute. However, AI-based, cognitive robots will learn from past experiences and calibrate their algorithms themselves, so their behaviour will not be perfectly predictable, and will likely become an issue worthy of serious ethical attention and reflection.

III. ROBOTS AND SOCIETY

III.1. Robots in Industry

48. Robotisation of industry has already happened and over the past decades, industrial robots have taken on the routine tasks of many 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, energy 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 and Osborne, 2013). While this development should lead to higher productivity from new technologies, it does not necessarily mean improvement for workers. Without structural adjustments and compensations, it could lead both to higher unemployment for certain profiles of the 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.

49. 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 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).

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

51. In this context, it is again important to note the ethical responsibility of industrial robot designers and engineers. A variety of ethical codes has been proposed to define the responsibility of a robotics engineer. For example, in Worcester Polytechnic Institute's Code of Ethics for Robotics Engineers, engineers should:

> consider the possible unethical uses of their 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. (Ingram et al., 2010, p.103)

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.

52. 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 total working hours for humans working side by side with more and more autonomous machines also arises, as it did with any technological revolution in the history of work. Finally, the appropriate educational system and additional training of existing workers need to be established and adjusted so that this latest technological revolution does not result in producing a mass of unemployable workers due to their lack of skills with the latest technologies.

53. Recent research (Murashov et al., 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; and
c. collaborative robots.

54. Most industrial robots are unaware of their surroundings, therefore, they can be dangerous to people. The main approach to industrial robot safety is the maintenance of a safe distance between human workers and operating robots through the creation of 'guarded areas'.

55. 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.

56. Service robots operate mostly outside industrial settings in unstructured and highly unpredictable environments, either operating without the presence of people such as in certain disaster areas, or with the presence of people such as hospitals and homes. Physical proximity between professional service robots and human workers can be 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, the more complex environments in which service robots must operate dictate a 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.

57. *Collaborative robots*, defined as a robot designed for direct interaction with a human, could be industrial, professional, or personal service robots. Collaborative robots combine the dexterity, flexibility and problem-solving skills of human workers with the strength, endurance and precision of 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.

58. Another set of ethical challenges arises by extending the digital divide concept to encompass a proposed concept of robotics divide. If the term digital divide refers to how digital technology, and access to it, redefines the power structure in contemporary human societies, the problem of a robotics divide would raise 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) Furthermore, as with a digital divide, there is a question of the global gap between developing and developed countries regarding access to robotics technology, which should be examined from the point of view of global justice.

59. The impact of robotisation on the 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 paid workers will struggle to compete with a robot's productivity. Where poorer countries could once use their cost advantage to lure manufacturers, now these 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 their customers and free from the risks, costs and complexities of a lengthy supply chain (Balding, 2016).

60. How can developing countries confront this possible widening of the digital and robotics divide, and its potential threat to their development strategies? As mentioned in the United Nations' 2030 Agenda for Sustainable Development and in the Addis Ababa Action Agenda, one thing that developing countries need to do is to turn the possibly liberating power of open data and big data to their own advantage (UN, 2015a; UN, 2015b). If data is the lifeblood of the robot revolution, then it must also be used to defend and compensate those who might otherwise lose out from these new technologies (OECD, 2016).

III.2. Military and civilian uses of mobile robotic systems

61. At the outset, COMEST would like to reaffirm its commitment to UNESCO's mission and mandate in promoting and building peace. Since the aim of this report is to provide critical and ethical analyses of the use of robotics technologies, it is important that COMEST address some of the ongoing discussion about military use of robotics. This in no way weakens its commitment to peace.

62. International Humanitarian Law (IHL) covers: (a) the protection of those who are not, or are no longer, taking part in fighting, and (b) restrictions to the means of warfare – in particular weapons – and the methods of warfare, such as military tactics (ICRC, 2014). The increased use of robotics in conflict affects both of these issues in a fundamental way. It raises real ethical and legal problems that must be reflected on if we are to protect human rights and human dignity.

III.2.1. Military robotic systems ('drones')

63. According to Chamayou (2015), the U.S. Army defines a drone as "a land, sea, or air vehicle that is remotely or automatically controlled" (p.11). The drone family is not composed solely of flying machines. There may be as many different kinds as there are families of weapons: terrestrial drones, marine drones, submarine drones, and even subterranean drones. Any kind of vehicle or piloted engine could, in principle, be 'dronized'. A drone can be controlled either from a distance by human operators (remote control) or by robotic means (automatic piloting); present-day drones may combine those two modes of control. The term 'drone' is mainly used in common parlance. Military jargon may, for example, refer to 'unmanned aerial vehicles' (UAVs) or to 'unmanned combat air vehicles' (UCAVs), depending on whether the vehicle carries weapons.

64. Operators ('pilots') of drones – aerial, terrestrial or submarine – are remote from the armed machine itself. They sit in a control pod that can be thousands of miles away from the action, and use a games-type controller to fly the craft and identify and attack the target. However, as for any new weapon system, using these weapons opens up not only a number of new military possibilities, but also brings new moral dangers. For example, it has been noted that the physical disconnect between pilot and field of action could lead to a gaming mentality. This could create a moral buffer from the action that could result in a cavalier approach to selection of and attack on targets. Indeed, one former senior remote operator is on record saying, "[e]ver step on ants and never give it another thought? That's what you are made to think of the targets – as just black blobs on a screen." (Pilkington, 2015)

65. International Humanitarian Law (IHL), also known as the laws of war and the law of armed conflict, is the legal framework applicable to situations of armed conflict and occupation. As a set of rules and principles, it aims, for humanitarian reasons, to limit the effects of armed conflict. It includes the Geneva Conventions and the Hague Conventions, as well as subsequent treaties, case law, and customary international law. Serious violations of IHL are recognised as war crimes.

66. There is no doubt that 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 weapons systems may be perceived to have benefits for military forces and even civilians in some conflict situations, there are concerns as to whether 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 that states are fully aware of the crucial need to assess the potential humanitarian impact and IHL implications of new technologies to ensure that not only they can be – but only 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.

67. In the context of scientific developments that have made robotic systems possible, it is perhaps worth noting that the Geneva Conventions originated in 1949 (ICRC, 1949a; ICRC, 1949b; ICRC, 1949c; ICRC, 1949d), when the big invention of the year was the 45-r.p.m. record. Similarly, the Additional Protocols to the Conventions were agreed in 1977 (ICRC, 1977a; ICRC, 1977b) when the personal computer was in an early stage of development, and our Information and Communications Technology (ICT) capability was a million or more times less than it is today. It is therefore highly appropriate to consider how observing the important IHL principles of *distinction*, *proportionality*, and *accountability* are affected by the use of robotic systems in armed conflict.

68. Making the *distinction* between a combatant and a civilian in a remote situation is of its nature a complex one – it is not just a visual classification. 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. Such limitations could lead to a tendency to stereotype potential targets, and so increasing the danger of target mis-identification.

69. 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. As Chamayou (2015) argues, precision of the strike has no bearing on the pertinence of the targeting in the first place.

70. Distribution of decision making among the crew of a remotely piloted craft and on-looking analysts can encourage 'group think', with crews actively looking for evidence to justify targeting. Discrimination systems are based on what is being looked 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?

71. 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 laws of physics 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.

72. Assessing *proportionality* remotely is seriously problematical – a decision on whether or not to apply lethal or kinetic force must be made in a particular context. Does the remote pilot have enough contextual evidence to balance expected military advantage against the loss of civilian lives? Will a strike aid the overall military objective or hinder it by creating major problems for the local population? It is a subjective and qualitative decision that must be made by the pilot, and his/her remoteness makes it questionable that a balanced decision can be made.

73. Remote assessment of proportionality is even more problematical in that it includes a consideration of more than potential civilian *physical casualties*. Other collateral damage needs to be considered that it is even more questionable that it 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.

74. Studies on the effects of drone strikes illustrate some of these less tangible proportionality issues. Many of those interviewed after such strikes have been found to be suffering from post-traumatic stress disorder. Other severe abnormal psychological conditions have also been found including depression, panic reactions, hysterical-somatic reactions, exaggerated fear responses, 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, and loss of interest in pleasurable activities. A further consequence was infrequent or non-existent school attendance.

75. Legally, these kinds of 'collateral damage' might be argued to be 'collective punishment' and counter to the prohibition on reprisals. Is the psychological impact on civilian populations of drones used in warfare sufficiently recognized or recognized at all

under IHL? What about the consequent social and economic impact and the seemingly open-ended nature of resulting societal disruption?

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

77. Just as in other forms of armed conflict, IHL offers legal protection to civilians in the context of remotely controlled warfare. However, in the absence of *transparency* surrounding the use of robotic-armed vehicles, the characteristics of the contexts in which they are used will be unclear, making *post facto* assessment of the legality and impact of remote strikes difficult.

78. Transparency is also noticeable by its frequent absence in assessment of robotic weaponry under Art. 36 of Additional Protocol I (ICRC, 1977a). Though there is evidence that such reviews are undertaken, the outcomes of the reviews are generally 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.

79. The legal issue of consent to deploying armed robotic vehicles does not appear to be different from that needed for the use of other military means. The legality of using force by a state in another state's territory if that latter state itself cannot counter an imminent threat is disputed. Even if it is resolved to be legal, the threat has to be a direct one to the state applying the force. There are 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 remotely controlled vehicles are the method of intervention, the general lack of transparency surrounding their deployment may complicate the consent issue.

80. Robotic weapons also have game-changing strategic implications. Their use seriously lowers the activation barrier to armed conflict by minimising – or completely removing the necessity for – human military forces in conflict zones, thus relaxing the requirement of *jus ad bellum* and shifting the assessment of military necessity. 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 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.

81. The use of remotely controlled weapons raises further issues with ethical implications. These include a shift in the notions of self-defence and threat: is it ethical to kill when you are not under threat yourself? The psychological health of those living under the threat of strikes has been discussed above, but there are also the psychological effects on the armed robot operators who will be able to see the immediate aftermath of their actions, as well as knock-on effects on their families and their societies to be considered. Finally, the likelihood that remote warfare in general will undermine traditional military values of honour and valour is already evident in the military.

82. There are also significant issues of safety of these systems. Like all machines, 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 vulnerable to spoofing and hacking, giving them the potential – which has already been realised in current conflicts in the Middle East – to be taken over by the 'opposition' and even turned back on the attacker's interests.

83. Noting the above problematic issues with respect to armed drone deployment, the legal situation for their use in armed conflict is in principle no different from the use of

manned aircraft. Decision processes by governments to deploy armed robotic vehicles should be no different from those used in manned aircraft deployment or in putting 'boots on the ground'. Moreover, it should also be noted that in an armed conflict situation, the remote operator, though he/she may be several thousand miles away, is a legitimate target for the opposite side. However, the ability to target remotely – and at very great distances – has opened up the whole area of *targeted killing*, whereby states have the ability to target particular individuals (in recent experience identified 'terrorists') remotely and outside declared armed conflict. This targeted killing is not new, but armed robotic systems have made it easier and more frequent, and it has become a key component in the 'fight against terrorism'.

84. Philip Alston, the UN Special Rapporteur on extrajudicial, summary or arbitrary executions has defined targeted killing as "the intentional, premeditated and deliberate use of lethal force, by States of their agents acting under colour of law, or by an organised armed group in armed conflict, against a specific individual who is not in the physical custody of the perpetrator" (UN, 2010, p.3). In the circumstances of armed conflict, such action may, under IHL, be legal (though one could question the ethics). However, the legality under IHL of the use of drones in targeted killing outside of conflict zones which is becoming increasingly frequent, must be questioned.

85. There are perhaps two issues to be considered here. First, Article 51 of the UN Charter permits armed force to be used in individual or collective self-defence (UN, 1945). It includes the 'inherent right' to use force in self-defence against an imminent armed attack, and the force used must be both necessary and proportionate to the threat. It is this justification that has been used by states to justify targeted killing in non-conflict zones. Assessing imminence, necessity and proportionality is however very much a matter of judgement that is open to political manipulation – it is all too easy to draw a definition so broad that it could cover many vague threats. Moreover, claims of national security are all too easily used in defence of withholding information on the reasons for a decision.

86. It would therefore seem good practice to set up an agreed appropriate, transparent set of procedures for targeted killing decisions. There are in fact examples of such procedures that are publicly set out and subject to significant degrees of legal oversight, transparency and post-strike assessment, notably by the U.S. and Israel, and the U.K. has set out the criteria on which it assesses the imminence of a perceived threat. A case has been argued, in particular by the U.K. government, that the development of drone technology and the advent of international terrorism means a reassessment of the legal framework is required. The strict ethical aspects of targeted killing remains a matter of very active discussion.

III.2.2. Autonomous weapons

87. 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 to both remotely piloted armed vehicles and fully autonomous weapons systems. However, 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 life and death decisions.

88. With respect to the capabilities of the software, it is crucially important to recognise that machines – no matter how well trained or cognitive they are – are still machines, and by their nature unable to perform all the – often nuanced – capabilities of the human brain. A recent report by the Royal Society (2017) notes the problem of developing machines with contextual understanding of a problem, or as it puts it 'common sense'. To quote that report: "When our expertise fails, humans fall back on common sense and will often take actions, which while not optimal, are unlikely to cause significant damage. Current machine learning

systems do not define or encode this behaviour meaning that when they fail, they may fail in a serious or brittle manner" (The Royal Society, 2017, p.30).

89. Furthermore, the report comments that there "are many constraints on the real world that we know from natural laws (such as physics) or mathematical laws such as logic", and "[i]t is not straightforward to include these constraints with machine learning methods" (p.30). Moreover,

> [u]nderstanding the intent of humans is highly complex, it requires a sophisticated understanding of us. Current methods have a limited understanding of humans that is restricted to particular domains. This will present challenges in, for example, collaborative environments like robot helpers or even the domain of driverless cars. (The Royal Society, 2017, p.30)

It presents even more challenges – which if not met can be catastrophic – in the domain of autonomous weapons.

90. When considering remotely piloted vehicles, we tried to address the problems of complying with the requirements of distinction and proportionality that are central to IHL. With respect to autonomous weapons, the difficulties of compliance are even greater.

91. 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 (ICRC, 2007). 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 a person who is armed, will it be able to tell if that person is a combatant or a nearby civilian police officer?

92. The issue of how to surrender to a remotely piloted robot has already been raised above. 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 either programmed into a machine or learned by it. Therefore, also by definition, recognising surrender requires human judgement.

93. The *proportionality* criterion is even more challenging for an autonomous system to fulfil. Assessing military necessity requires consideration of a range of issues, many of which are inherently unquantifiable. Can an autonomous weapon do that in the absence of meaningful human control? Can it determine, and then inflict, only that degree of harm that is needed to counter a particular situation?

94. 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 the loss of ability to provide, or consider in the equation other social and psychological consequences? Assessing proportionality is very much a judgement call, and one which we expect to be taken by a 'reasonable military commander'. This requires human contextual awareness. Therefore, it would seem an inevitable conclusion that proportionality assessment must involve human judgement. It cannot ethically be left to a machine.

95. To kill legally, a commander in battle must act in accordance with international humanitarian law (IHL), and an autonomous machine is similarly constrained. It has been argued by some roboticists that it may be possible to create an 'ethical governor' that is programmed to follow those rules, and which would therefore, unlike a human warfighter, not violate those rules. The problem here is that we are often dealing with guidelines rather than

hard yes/no rules. In order to create such a ‘governor’, those guidelines would need to be reduced to simplistic terms suitable for programming into the robot. It is therefore inappropriate – and dangerous – to try to reduce or simplify them.

96. A further point here relates to the authority to kill or to delegate the power to kill. Here, IHL requires 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” (International Peace Conference, 1899, Art. 1 of the Annex to the Convention). The Martens Clause – a long-standing and binding rule of IHL – specifically demands the application of ‘the principle of humanity’ in armed conflict (Ticehurst, 1997). It requires a human to override a person’s 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, regardless of major legal changes, that would still not be acceptable action.

97. Delegation of killing to a machine, however complex its programming, 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.

98. 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 or set of environments in which it has been trained. It may work well in those environments, but move it outside this specified environment to one which is more open and robot behaviour becomes unpredictable. As the failure of some sophisticated image recognition software has demonstrated, even small changes in the environment can lead to malfunction. We might also note the fatal Tesla autopilot accident in 2016 when the car met an unanticipated configuration of a truck and a trailer. All possible environments and likelihoods cannot be programmed into the system, nor can the machine be trained in a way that covers all possible eventualities: its behaviour outside its comfort zone is therefore unpredictable and may be catastrophic. The potential consequences of unpredictable behaviour become even more worrying if we consider two sets of autonomous weapons programmed with different – and unknown to the other – rules of operation.

99. New prototypes in unmanned systems are increasingly being tested at supersonic and hypersonic speeds. In order to counter these, even faster autonomous response devices will be required, and these in turn will require ever-faster weapons. The resulting ‘pace race’ will leave humans with little control over the battle space. If the development and proliferation of autonomous weapons continues, the (super/hypersonic) defence systems of one state could interact with equally fast autonomous weapons from another. The speed of their unpredictable interaction could trigger unintended armed conflict before humans had the opportunity to react.

100. It should also be noted that even relatively simple systems fail and need human intervention to resolve the malfunction. In addition, 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 or software upgrade. 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.

101. 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 just asymmetrical into a unilateral relationship of death-dealing in which the enemy is deprived of the very possibility of fighting back. Use of

these weapons therefore could lead us out of the normative framework that has developed for armed conflicts.

III.2.3. Surveillance, policing and the use of military technology in non-military contexts

102. 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.

103. On the other hand, the increasing use of robotic vehicles for surveillance raises potentially problematical issues with significant ethical implications. 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 major gaps in surveillance capability. They are used by European police forces for border, crowd, and event control; evidence gathering; traffic control, searches, observation; documenting 'troublemakers'; surveillance of buildings and the area around VIPs; as well as searching for/controlling/targeting undocumented migrants, workers, and 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. Furthermore, ubiquitous 24/7 surveillance could not only undermine political participation in present-day democracies, but it could massively empower repressive and dictatorial regimes.

104. 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 an economy of fear: the fear of an "omnipresent enemy who could be anywhere, strike at any time and who could be 'among us'" (Crandall and Armitage, 2005, p.20). 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, Computers, 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.

105. There are questions about the use of so-called sub-lethal weapons on robots for 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 could.

106. It should be noted that a robotic device has already been used in the application of lethal force (i.e. killing) in a civilian context. Using a bomb-disposal robot with an explosive device on its manipulator arm, a U.S. police force has killed a suspect remotely. The development of automatic targeting technology raises the possibility of human and civil rights violations by police and private security forces.

III.2.4. Private and Illicit use of robots

107. Small remotely piloted vehicles (RPVs) are already available on the open market. The consequent snowballing use of the technology raises 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.

108. Deployed without proper regulation, drones equipped with facial recognition software, infrared technology, and microphones 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.

109. The potential use especially of small drones by non-state actors 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 for chemical and biological weapons. Security officials have warned that advances in the availability and sophistication of commercially available drones increase the likelihood of such events occurring in the future.

110. Safety issues are also involved with wider use. The hobbyist community could convert commercial drones to fly large distances and carry significant payloads. Cases of air incidents involving hobbyist drones have already been reported. The ready and rapidly increasing availability of drone technology has generally occurred in the absence of significant regulation on their deployment.

III.3. Robots in Transportation

111. Transportation is one of the fields where robotics is widely used and well accepted, and the motorcar industry has played a major role in automation. Automatic transmission, where robotics replaced the need to change gears, is one of the first examples. As the sensors evolved and technology advanced, more drivers' functions have been delegated to robots, including: cruise control, lane changing, and parking.

112. Mass transportation, dominated by all kinds of trains, is becoming fully automatic, where the driver is mainly for back-up and emergency situations. Ships and airplanes also have automatic steering systems, to assist the pilots by doing most of the routine tasks.

113. A leap in the use of robots in transportation is just around the corner: the emerging technology of autonomous vehicle (AV). The near future will allow technology that does not just assist the driver, but can replace it completely. AVs use multimodal sensing of the environment, by cameras, radar, GPS and laser, and some of the most advanced AI systems for processing data and making decisions. Moreover, AVs are based on cognitive computing algorithms, so they can become independent learners.

114. Widespread use of autonomous vehicles can have many advantages. According to the McKinsey consulting firm, it can eliminate 90% of all auto accidents in the United States, prevent up to US$190 billion in damages and health-costs annually and save thousands of lives (Bertoncello and Wee, 2015). Other significant potential advantages include better efficiency of transportation (shorter travel time, lower costs), and better quality of life (independence) for dependent persons.

115. While the technology leading to AVs is promptly advanced by many leading producers and successful demonstrations of AVs have recently been provided, it will not be fully operational until a few additional technical challenges are addressed. These include mainly the reliability of the technology (hardware and software), and cyber security.

116. However, the most challenging aspects of AVs are in the social dilemmas they raise. These may include job losses, third world divide, and moral issues such as "when it becomes possible to program decision-making based on moral principles into machines, will self-interest or the public good predominate?" (Editor's Summary of Bonnefon et al., 2016)

117. The question of self-interest versus public good is particularly relevant to AVs. How should the car be programmed to act in the event of an unavoidable accident? Should it minimize the loss of life, even if it means sacrificing the occupants, or should it protect the occupants at all costs? Should it choose between these extremes at random? Moreover, Bonnefon et al. (2016) ask:

> Is it acceptable for an autonomous vehicle to avoid a motorcycle by swerving into a wall, considering that the probability of survival is greater for the passenger of the car, than for the rider of the motorcycle? Should different decisions be made when children are on board, since they both have a longer time ahead of them than adults, and had less agency in being in the car in the first place? If a manufacturer offers different versions of its moral algorithm, and a buyer knowingly chose one of them, is the buyer to blame for the harmful consequences of the algorithm's decisions? (p.1576)

118. Should such issues be regulated by national/international laws? By standards? By codes of conduct in autonomous vehicles? By market forces? The AVs industry will not be commercial unless it is regulated. This report will provide some relevant recommendations in this regard.

III.4. Health and Welfare

III.4.1. Medical robots

119. 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 (more precision, no natural hand tremor) and some studies show shorter time in the hospital for the patient, lower blood loss, minimizing the trauma due to incision in the tissues. However, studies do not show a significant difference in terms of efficiency between the usual surgery and robot surgery (Kappor, 2014). Surgeons express some drawbacks with robotic surgery: they do not have any more sense of feel – the robot gives them no tactile feeling, only image, so the surgeon is limited to the visual (Tzafestas, 2016a). A major problem is the cost of surgery robots and their maintenance. So actually, the cost of surgery by a robot is higher than with an 'ordinary surgeon' (Gagnon, 2018). The multiplication of robots in surgery will have a consequence on the allocation of resources in public health systems, on the number of surgeons, and on their ability to practise surgery without robots.

III.4.2. Robots in healthcare

120. 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 adaptability of children with autism. Such robots have also been used to improve the learning of children with Down Syndrome.

121. Autism is a spectrum disease where no direct causes are identified and with no cure. Autism is characterized by a variety of symptoms reported by the parents: stereotype behaviour, repetitive movement, limited interest, etc. These children have communication problems, they may have uniform tone when speaking, repeating certain words and inappropriate sentences, etc. A clinical evaluation of the young child by a multidisciplinary team will lead to rapid and constant intervention to improve language, interaction and adaptability (improvement of social functionality). Robots are being used on an experimental basis to improve the adaptability of these children (Scassellati et al., 2012). These children are often attracted by the robot because it manifests constant affect and its behaviour is predictable and repetitive. The robot is not threatening because it is not unpredictable, as opposed to human behaviour and emotion.

122. Although there are currently not many studies related to the use of robots in this manner, none of these studies has given an indication that the robot-autistic child dyad is superior to the human-autistic child dyad in terms of functionality (Diel et al., 2012; Simut et

al., 2016). The use of robots with autistic children is not actually considered a standardized form of care because patients with autism are a very heterogeneous group. Some issues have been raised by clinicians in this regard. Robots may not be beneficial to all autistic children because of the variety of manifestations of the disease. Are autistic children actually interacting or communicating with a robot? Communication means catching emotions, reacting to how one feels, expressing it through words, transferring these emotions, and repeating this process again and again. Emotions are essential in communication and they are unpredictable in real life, with parameters that one cannot control. To improve the adaptability of children with autism through learning of communication is still the best path to help these children, so the interaction between autistic children and robot must not be the final and only form of therapy for them (Gagnon, 2018).

123. Another area of medical robotics concerns the development of exoskeletons that are used to help persons with physical disability (i.e. artificial legs with Ekso) in order to reduce physical limitation. In rehabilitation, robots are used on an experimental basis for patients with spinal injuries, or patients who have survived a stroke. They are also used with patients suffering from ankle injuries (Zhang et al., 2013).

124. However, these robotic devices can also be used for non-therapeutic purposes (e.g. enhancing athletic performance in sporting events, super-soldier, etc.). Exoskeletons, neurological implants, nanorobots, and other similar robotic devices open the door to the transformation of the human body and mind. The ethical question is to what extent the human-robot hybridity should be developed and for what purpose? Should the parts of the human being, as well as the human being as a whole (body and mind), be viewed as something that can and should be enhanced, similar to a robot, when using robotic devices for non-therapeutic purposes? Which ideology should drive this transformation (e.g. enhancement, submission, superhuman powers as portrayed in the movies, etc.)?

III.4.3. Healthcare robots in elderly care

125. 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 the functional and affective dimensions of social robots as the characteristic of robotic devices that can be called ‘companion robots’ (Oost and Reed, 2010).

126. One area where the use of social robots and robots as companions is in a considerable rise concerns the area of healthcare, especially for elderly people. In the context of demographic projections, the reality of a rapidly ageing 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.

127. 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 et al., 2009)

128. On medical grounds, the ageing patients that would benefit from interaction with robots are mainly patients with different degrees of dementia and patients with different levels of physical limitations.

129. A robot that can stimulate cognition of a dementia patient or execute on a day to day basis some tasks (cleaning, cooking, personal hygiene, etc.) that are more difficult to do by the elderly person, may be beneficial (Vandemeulebroucke et al., 2017; Mordoch et al., 2013). A major objective is to assure security (risks of falling, heart failure, mobility disorders, etc.) at home, in hospitals and care institutions.

130. A question then is related to the appropriate means: should robots be used in these situations? And for what purpose? Should automation (cameras, microphones, captors) be the go-to option? How can we balance security and behavioural control (medical regime compliance, interrupting alcohol consumption, etc.) induced by the use of robots, and the autonomy of ageing persons? What about a loss of private life and intrusiveness? In other words, what is the role of the robot? Who designs it and for what purpose? Is the purpose to improve the quality of life of older persons? Can such robots help older people live independently in their homes and extend the time of 'ageing in place' instead of moving to institutional care? Is the purpose to reduce the work of caregivers? Is the purpose to discharge society from the care of old people? (Wu et al., 2010)

131. Another kind of robot, the robot-companion can be used in a medical setting to reduce the solitude of ageing persons and to prevent behaviours associated with dementia. Some studies have documented the effects of robots – such as Paro the seal robot and Aibo the dog robot – on the social interaction, attention or cooperation of older patients with or without dementia. For example, Paro allows for an affective relationship, because of its advanced imitation of a real animal: it breathes, has a body temperature, makes sounds and movements that evoke affection, and reacts to touch, sounds, and speech. Nevertheless, for the moment the results do not give a decisive advantage in the use of these robot-companions in a medical or care context.

132. Some concerns to consider are as follow: can robots be adapted to the very individual pattern of personality and behavioural changes of a person with dementia? Workers in the aged-care sector are generally under-paid and robots are very expensive, so what will be the burden on public health systems? Concerning robot-companions, do they initiate fake, illusory emotions and reciprocity, inducing deception and infantilizing aged persons? (Sharkey and Sharkey, 2012) Are robotic systems able to provide adequate care, and to what extent does care require interpersonal relationships that can only arise between human beings? Are robots able to treat elderly people with due respect?

133. Can then healthcare robots be useful? Will they be accepted by elderly people and their families? Acceptance is defined as the healthcare robot being willingly incorporated into the person's life. This requires a motivation for using the robot, sufficient ease of use, and comfort with the robot physically, cognitively and emotionally. Other elements should be considered: the cost and the maintenance of these robots, and the impact on the role of natural caregivers. The social and cultural acceptance of healthcare robots differ also from country to country.

III.4.4. Companion robots

134. As indicated in the previous section, robots are increasingly used as companions. Other than elderly care, another common domain of application is sexual relations. As with elderly care, fundamental questions are rising about the potential implications and their desirability in this area.

135. In the field of sexuality, two types of robotic systems are gaining influence: systems that make possible sexual interaction at a distance ('teledildonics') and systems that function as a (surrogate) sexual partner. 'Teledildonic' devices are interactive robot systems that enable people to have sexual interaction from a distance, by penetrating an interactive sleeve and using an interactive dildo (Liberati, 2017). Ethical questions raised by such devices concern the risk of reducing intimate contact and sexual interaction to penetration, and the new combination of nearness and distance in intimate relationships.

136. Sex robots are gaining influence as well: realistic, interactive sex dolls. In Barcelona, the first robot brothel has already opened, offering several variants of the Roxxxy sex robot. Arguments in favour of sex robots include the way in which they can help lonely and frustrated people, preventing psychological suffering or sexual misbehaviour. Also, sex robots might reduce prostitution. Arguments against sex robots include their reduction of sexual intimacy to self-gratification, potentially making people less able to form strong family,

friendship and love relationships. Another matter of ethical concern is the potential encouragement of sexual violence and paedophiliac behaviour through sex robots that have child-like features or programmed for abuse.

III.5. Education

137. With the rapid growth of communication technology, more and more multi-media tools are being used in education, including educational robotics. Educational robotics allow for the exploration, design, modelling, programming, constructing and testing of unitary knowledge concepts (motion, force, traction…) but also more complex and realistic systems which require a combination of different concepts and methodologies from different disciplines. Educational robotics can support individual and collaborative learning activities, and be aligned with different curriculum objectives and competences. 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. Lower-cost educational robotics kits with less sophisticated sensors and controllers have made the technology available for use in classroom settings. Typically described learning themes of educational robotics include: interest in science and engineering (some educational robotics programmes have shown promise in increasing retention rates among female students, who are under-represented in technology-focused fields); teamwork and problem solving (Miller et al., 2008).

138. Robots are also used for different educational aims and roles in the classroom. For example, a humanoid robot is used on an experimental basis to teach a second language in primary school. The robot stimulates interaction with repetition of words and sentences and enhances motivation of the students (Chang et al., 2010). Robots are also being used for storytelling in order to engage primary students in story expression and creation with the robots (Sugimoto, 2011). More often, the use of a robot in the classroom is to help with understanding the scientific process and mathematical concepts.

139. For the moment, recent reviews on the use of robots in education show that only a minority of studies provided empirical results (quantitative and qualitative) that can help to understand practical impacts on education (Benetti, 2012; Toh et al., 2016). One of the main issues is the role assigned to the robot in education: should it be considered as a tutor, a tool, or a peer? To what extent could a robot replace a teacher? Some children ascribe cognitive, behavioural and affective characteristics to a robot (Beran et al., 2011). Therefore, the question of attachment of a child to the robot and the transformation of the relationship to the teacher is another area of preoccupation. The vulnerability of primary school children to emotional simulation has been pointed out.

140. Concerning the conception of education, the use of robots supports constructivism as a learning model, which implies an active participation of students in the construction and acquisition of knowledge. Educational robotics activities could engage learners in collaborative challenges when the level of complexity allows them to engage in a problem solving activity (Kamga et al., 2016). Beyond the opportunity of educational robotics to engage learners, the underlying question is whether robots really stimulate the motivation of children or whether it is a manifestation of their curiosity that can then fade away. To what extent does the learning process with a robot need to be recreational in order to induce a student's motivation? Moreover, access to robotics also leads to reflection on the shadow of new inequalities and marginalization – issues that need to be considered.

141. Education of children is deeply rooted in culture and is related to transmission and construction of knowledge as well as social values. Parents and social actor attitudes toward the use of robots in education and healthcare is different from one country to another. In a survey conducted in 2012 by the TNS Opinion & Social (EC, 2012), 34% of the respondents in Europe would ban the use of robots in the area of education, while 3% said that robots in

education is a priority. In Japan and South Korea, the use of robots in education seems to be more experimented and culturally accepted. In Japan, its Shinto religious background has generated some positive images of robots as harmonious, gentle, cute, and healing objects. Robots are seen as machines that generate emotional bonds that can lead to harmonious relation with humans (Tzafestas, 2016a). Therefore, whether to introduce robots in the education system or not is also related to cultural values, as well as public, economic and political choice.

III.6. Household

142. Household robots (also referred to as 'service' or 'domestic' robots) are developed in order to assist human beings to perform jobs which may be considered dull or dirty, like floor vacuuming, garbage collecting, window washing, plant watering, pool cleaning, ironing, preparing food and beverages, etc. Within the 'household robots' category are often also included devices like automatic cat litter boxes, security and alarm robots, robotic lawn mowers, robots for monitoring pets, robotic rocking cribs, or robotic shopping assistants.

143. Although they are not considered uniquely household robots because they frequently appear in other settings (for example, nurseries and retirement homes), devices like robotic toys and entertainment/companionship robots may be reasonably classified as 'household robots' as well. The main reason behind this classification is the fact that many human activities (such as play, companionship, entertainment, various hobbies) subserved or enhanced by such robots typically take place within the boundaries of one's home or household.

144. Household or domestic life is one of the areas in which the application of robots and robotic devices is on a continuous rise. According to projections by the International Federation of Robotics, "sales of all types of robots for domestic tasks (vacuum cleaning, lawn-mowing, window cleaning and other types) could reach almost 31 million units in the period 2016-2019" (IFR, 2016b, p.3). This number becomes even larger if one expands the category of 'household robots' with other types of robots such as robotic toys, entertainment robots or various robotic hobby systems.

145. Household robots and robotic devices have undoubtedly improved and are likely to continue improving human life and condition, by performing jobs that humans tend to find dull, dirty or simply tiresome. They save their users' energy and time, allowing users to engage in activities that are more meaningful and life-fulfilling. Such robots may also be expected to reduce (in most cultures and societies still highly visible) gender differences and unequal division of household work between males and females.

146. Like many other types of robots, however, household robots create their risks and dangers, as well as unique ethical and legal challenges that should be taken seriously by their designers, manufacturers and end users. One of the most obvious challenges is the harm that they might cause to their primary users, but also to other residents of the same household. Specific harms potentially caused by household robots, for example, may be related to their inappropriate force or speed, failure to correctly understand human instructions or error when performing particular tasks (especially delicate and sensitive tasks like playing with children, taking care of pets or preparing food or drinks).

147. A commonly mentioned danger related to many types of household robots is the fact that they are relatively easy to hack and could be used as devices for various kinds of unethical or even criminal activities. Since many household robots come equipped with sensors like cameras and microphones, they can be transformed into spying tools used for invading one's privacy and intimacy. Such robots, should they become permanent residents of one's home, will gain access to and possibly store a vast amount of private and confidential information, e.g. photos of the home interior and its residents, data about residents' habits, passcodes to alarms, the location of valuables, etc. The danger becomes even larger when household robots are connected to the Internet or some other network that is insufficiently protected and easily hacked. In such settings, hacking the robot may also

facilitate criminal activities like theft or blackmail. Precautionary measures against such misuse of household robots and related technology should be taken seriously by their designers, manufacturers and sellers.

148. When it comes to household robots designed for play and entertainment, their misuse or malfunction may have even more serious consequences, especially when their end users are children or the elderly (as two populations that are extremely vulnerable to intentional and unintentional deception and exploitation). When such robotic toys and companionship robots are intended to be used within the boundaries of one's household and without supervision of any professional personnel (as compared to nurseries, educational facilities, hospitals and nursing homes), an even higher level of sensitivity and anticipation of possible dangers and misuses should be required from their designers and manufacturers.

149. The appearance of robotic toys and companion robots is also an ethically sensitive issue that requires careful reflection. Pearson and Borenstein (2014), for example, recognize several areas in which ethical considerations should play a considerable role when designing children's robotic toys. For example, this design should correspond to collective aesthetic preferences attributed to a particular culture; it should not reinforce gender stereotypes (e.g. by creating robotic toys with unnecessarily emphasized 'gendered' appearance); it should be context-sensitive and appropriate to a child's developmental level; it should pay particular attention to the level of humanlike appearance to be given to such toys, because positive or negative effects of such appearance tend to vary with age or personality traits of their end users.

III.7. Agriculture and environment

150. Robots in agriculture are for example represented in dairy farming. Cows that are free in the stall come to be milked by the robot. Concretely the robot will first identify the cow and will accumulate information on its health (temperature, hormone level, infections, etc.) and on the milk production of the cow. Then the robot locates the teats and milks the cow while it receives food supplement. After the cows are trained to go to the robot, they can be milked three times a day which is a gain of productivity for farmers. Another robot can clean the barn. The cow's productivity will be monitored by the robot that will then adapt the feeding of each cow to its level of production and its age (the productive life of a milk cow is around 4 years, after which it is normally brought to the slaughterhouse). For the farmer there is also a gain of time and flexibility. The use of milking robots also reduces physical labour and staff. To be productive, this kind of robotized organisation of the work implies about 60 to 70 cows per robot. The cost of these robots is still very high (around US$200,000) and many farmers need to make adaptations to their barn, which implies new investments. The installation of robots requires regular professional maintenance. Concerning the cow, it implies a long adaptation period – the breeder still had to bring any non-compliant cow to the robot at the end of the day. Furthermore, there are examples of cows that refuse to follow the system being culled because of their independence. The introduction of robots has modified the vision of the animal that is now built on multiple data transmitted by the robot. The farmer relies on the robot to make decisions about each cow and relies less on his/her own knowledge of the animal. The relation to the animal is also transformed. For the cow, its relation to human beings is minimized and certain animals have lost familiarity and mutuality with human beings (oral communication, emotional exchange, affection through hugs). The question of animal welfare, which also implies a dimension of individualization for the animal, is at stake. The use of robots has accentuated the model of productivity and the association of the animal-machine model. There is no reflection yet about the animal-robot interaction. (Driessen and Heutinck, 2015; Holloway et al., 2014)

151. Drones may also contribute to sustainable agriculture and aquaculture. This ~~new~~ technological application is related to precision agriculture. The drones obtain data that can be analysed for more efficient use of chemical inputs (pesticides and fertilizers) or water (drip irrigation). They also allow for selection of interesting traits of plants in the field (e.g.

tolerance to drought, salinity or stresses, resistance to pests or diseases) in order to use the selected plants in crop breeding programmes to face challenges such as climate change. The contribution to food safety and crop production is high, because they enhance agricultural yields. The cost of the technology is variable, but it has been becoming more available in time, even for small farmers in developing countries depending on the conditions and specifications. As for dairy farming, the relation and knowledge of the farmer of his own land and fields is being transformed. This model of precision intensive agriculture induces a modification in the relationship to land such as already described by Aldo Leopold (1949). The use of drones in agriculture reinforces the model of productive agriculture.

152. There are a number of areas where robots may have environmental benefits, such as their use in recycling, environmental monitoring and remediation, and clean-up after nuclear and chemical accidents (Lin, 2012). The utilisation of robots in deep ocean research and space exploration has made a significant contribution to environmental science. However, the potential benefits need to be balanced against the environmental impact of the entire robot production cycle. This would include mining for rare-earth elements and other raw materials, the energy needed to produce and power the machines, and the waste generated during production and at the end of their life cycle. Robotics is likely to add to the growing concerns about the increasing volumes of e-waste and the pressure on rare-earth elements generated by the computing industry (Wildmer et al., 2005; Alonso et al., 2012). In addition to the environmental and health impacts, e-waste has important socio-political implications, especially related to the export to developing countries and vulnerable populations (Heacock et al., 2016). There seem to be few, if any, attempts to analyse the environmental impact or footprint of a robot.

IV. ETHICAL AND LEGAL REGULATION

153. Robotics and robots raise new and unprecedented legal and ethical challenges. These challenges originate from the multidisciplinary nature of robotics, specifically the science and technology related to the design, building, programming and application of robots. Creating a robot involves contributions of numerous experts, from a wide array of disciplines like electrical/electronic engineering, mechanical engineering, computer science or even biomedicine. Depending on the intended application of the robot, its type, complexity and the environment in which it will be used, collaboration with an even wider circle of experts may be necessary (e.g. artificial intelligence experts, cognitive scientists, space engineers, psychologists, industrial designers, artists, etc.).

154. 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 programming its software, etc. For example, an emerging subfield of ‘biorobotics’, in addition to standard robotic technologies and materials, works with organic tissues and living organisms (laboratory animals) and therefore raises its own set of legal and ethical issues.

155. Given this complexity of design, construction and programming of robots, one of the most frequently mentioned legal and ethical issue is ‘traceability’. A requirement of traceability would imply that, technically, one must be able to determine the causes of all past actions (and omissions) of a robot. This is ethically and legally crucial 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, p.6).

156. However, there is a tension between the ‘traceability’ requirement and the development of robots with a high level of autonomy, decision-making capacities and learning abilities. Robots with such properties will continue to be aimed for, because they make them better able to perform various tasks that are currently done by humans. However, the ‘machine learning’ processes that are made possible by their algorithms, and that form

the basis for robot autonomy and decision-making, reduce the possibilities to trace back all origins of the behaviour of the robot. Such robots are not merely programmed to perform specific tasks, but to learn and further develop themselves in interaction with their environment. This ability of robots to 'learn' and 'programme themselves', therefore, will require modifications to current legal and ethical notions of 'traceability'.

157. Robotics, then, as Matsuzaki and Lindemann (2016) are right to observe, must perform "two contradictory tasks: making robots autonomous and at the same time making them safe" (p.502). This inherent predicament of robotics does not seem to be easily translatable into legal categories and regulations.

158. What follows is a brief and selective overview of some of the most important reflections and proposals (ranging from individual to institutional and organizational) for an ethical and/or legal regulation of contemporary robotics and the robot industry.

159. When it comes to individual scholarly approaches to ethical challenges related to robots and robotics, a lively discussion has been going on for nearly two decades under labels like 'ethics of robotics', 'robot ethics' or 'roboethics'. The frequently used term 'roboethics' was coined by Gianmarco Veruggio (2002) and it denotes what might be called the code of 'ethics of roboticists'. "The target of roboethics," as Veruggio and Operto (2008) explain, "is not the robot and its artificial ethics, but the human ethics of the robots' designers, manufacturers, and users" (p.1501). 'Roboethics' as the branch of applied ethics dealing with the way humans design, construct and use robots, should not be confused with 'machine ethics', as the discipline dealing with the question of how to programme robots with ethical procedures, rules or codes (to be discussed later in this report).

160. Despite the fact that the debates about ethical implications of robotics today are numerous and very diverse, there have not been many individual scholarly proposals to precisely formulate the basic principles of future 'ethics of robotics'. Two such proposals, however, do exist: one by Ingram et al. (2010) and the other by Riek and Howard (2014).

161. According to 'A code of ethics for robotics engineers' proposed by Ingram et al. (2010), an ethical robotics engineer should have 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. (p.103)

The proposal also addresses ethical issues like respect for people's physical well-being and rights, rules against misinformation and conflict of interest, requirement of constructive criticism and personal development.

162. 'A code of ethics for the human-robot interaction profession' by Riek and Howard (2014) is an attempt to ground specific duties pertaining not only to roboticists, but also to a wider community of people related to robotics and robots as its products, such as product managers, marketers, workers in industry or government officials. They propose 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. (p.5)

Among more specific principles it mentions 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.

163. When it comes to particular professions and disciplines that constitute or contribute to robotics, as Ingram et al. (2010) noticed, they typically have their own well-developed codes of conduct. For example: the Institute of Electrical and Electronics Engineers (IEEE) Code of Ethics, the American Society of Mechanical Engineers (ASME) Code of Ethics, the Association for Computing Machinery (ACM) Code of Ethics, the Accreditation Board for Engineering and Technology (ABET) Code of Ethics. However, ethical codes specifically written for roboticists seem to be still in their infancy. An initial one that we have noted positively is the IEEE Initiative for ethical considerations in AI and autonomous systems that is under development (IEEE Standards Association). Noting that robotics design, manufacture and use is relevant to a range of other professional communities (medics, chemists, physicists, biochemists, biologists, etc.), there would seem to be a clear case for other professional bodies to follow this example, with the ultimate aim of an ethical framework common across the various disciplines.

164. Similarly, although robotics 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 such projects, especially those that have direct or indirect bearing on human beings, should proceed. One should also bear in mind that the robotics industry is becoming a highly lucrative business, and that experience has shown that codes of conduct and ethical guidelines are often seen as impediments to research and development.

165. Despite the fact that there are no universally accepted codes of conduct specifically written for roboticists, some important institutional and organizational initiatives for ethical (or ethically relevant) regulation of robotics do exist and should be mentioned.

166. The Government of the Republic of Korea, according to Veruggio and Operto (2008), established a working group with the task of drafting a roboethics charter in 2007. The charter was planned to be presented at the Robot Industry Policy Forum and Industrial Development Council, after which more specific rules and guidelines were planned to be developed. As Veruggio and Operto (2008) also reported, 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 planning, manufacturing, administration, repair, sales and use of robots. 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, p.1506). It is unclear whether the Korean charter or the Japanese guidelines were completed to this day.

167. In 2016, the British Standards Institution published a document on 'Robots and robotic devices: Guide to the ethical design and application of robots and robotic systems'. The document was composed by a number of scholars from different fields and it addresses issues like responsibilities for robots' behaviour, robots' maximum speed or force, ethically acceptable design of industrial, personal care and medical robots, emotional bonds with robots, unnecessary anthropomorphization of robots, robot learning, sexist or racist implications of robotics, privacy and confidentiality, dehumanization, etc. This British Standards document is intended for robot designers and managers as "guidance on the identification of potential ethical harm and provides guidelines on safe design, protective measures and information for the design and application of robots" (BSI, 2016, p.1).

168. 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 emphasizes 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 (Allistene, 2016).

169. The International Organization for Standardization (ISO) has issued in recent years a number of standards related to robots and robotics. Most of these standards, of course, are technical in their nature and as such focused primarily on reliability and quality of products. However, some ISO standards related to robots and robotics have at least indirect ethical and legal relevance as their aim is to standardize robotic devices with regard to their end-users' safety. For example, ISO 13482: 2014 (Robots and robotic devices – Safety requirements for personal care robots), ISO 10218-2: 2011 (Robots and robotic devices – Safety requirements for industrial robots – Part 2: Robot systems and integration) or, more generally and not necessarily related to robotics, ISO 13850: 2015 (Safety of machinery – Emergency stop function – principles for design).

170. One of the most concrete initiatives for future legal and ethical regulation of robotics and robot industry is the Draft Report with recommendations to the Commission on Civil Law Rules on Robotics, issued in 2016 by the Committee on Legal Affairs of the European Parliament. The report 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 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. For example: 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), impact of robotics on human employment and education. Also emphasized is the importance of traceability and a system of registration of advanced robots: "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" (JURI, 2016, p.13).

171. In 2017, the Rathenau Instituut issued a report on 'Human rights in the robot age', commissioned by the Parliamentary Assembly of the Council of Europe (PACE). The report starts from the premise that new technologies are "blurring the boundaries between human and machine, between online and offline activities, between the physical and the virtual world, between the natural and the artificial, and between reality and virtuality" (Rathenau Instituut, 2017, p.10). It continues by addressing potentially negative impacts of robotics and related technologies – especially the nano-, bio-, information and cognitive (NBIC) technologies convergence – on a number of issues related to human rights: respect for private life, human dignity, ownership, safety and liability, freedom of expression, prohibition of discrimination, access to justice and access to a fair trial. As for robotics and robots in particular, the report pays special attention to self-driving cars and care robots, recommending to the Council of Europe to reflect on and develop new and/or more refined legal tools for regulating the usage of such devices. It also recommends the introduction of two novel human rights: the right not to be measured, analysed or coached (related to possible misuses of AI, data gathering and the Internet of Things) and the right to meaningful human contact (related to possible misuses, intentional and unintentional, of care robots).

172. Theoretical approaches to and research on legal challenges of robotics, robots and the robot industry seem to be thriving, with many insightful contributions available (e.g. Matsuzaki and Lindemann, 2016; Calo, 2015; Asaro, 2015; Holder et. al, 2016; Leenes et al., 2017; to mention only some of the most recent). However, at the more concrete level, both national and international, robotics, the robot industry and the robots market remain to a large extent legally underregulated. According to a country-by-country analysis from the 'Comparative Handbook: Robotic Technologies Law' (Bensoussan and Bensoussan, 2016), in countries such as Belgium, Brazil, China, Costa Rica, France, Germany, Greece, Israel, Italy, Japan, Lebanon, Portugal, South Africa, Spain, Switzerland, United Kingdom, and the

United States, there are no specific laws on robots (although in some countries legal regulations of some robotic technologies do exist, e.g. when it comes to devices like drones and autonomous cars).

173. The main reason why robotics remains legally underregulated probably comes down to the fact that robotics is a new and rapidly advancing field of research whose impact on the real world and legal consequences is difficult to conceptualize and anticipate. According to Asaro (2012), all harms potentially caused by robots and robotic technology 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 the robot industry is likely to become inadequate as commercially available robots become more sophisticated and autonomous, blurring the demarcation line between responsibilities of robot manufacturers and responsibilities of robot users (Asaro, 2012).

174. While the existing legal regulation of the robot industry based on product liability may prove inadequate for the robotic technology of the future, the problem with future more specific legal regulation of robotics (especially if it is too restrictive) is its potentially negative effect on development of robotic technology. As Leenes et al. (2017) emphasized, the problem with current European and American liability regulation is that "while an increase in safety standards cannot be substantially assessed, such regulations are deemed to produce a technology-chilling effect" (p.17). As they say, it is likely that future robotic devices (like driverless vehicles) will be sufficiently sophisticated and reliable in order to function without human intervention. "However, imposing a strict standard of liability on producers before such a level of sophistication is reached […] may discourage the very development of that technology, liability being judged to represent too considerable, and too uncertain, a risk" (Leenes et al., 2017, p.17).

175. Therefore, the field of robotics as such seems to face four regulatory dilemmas: (1) dilemma of keeping up with rapid technological advances; (2) dilemma of striking a balance between stimulating (or at least not impeding) innovation and protecting human fundamental rights and values; (3) dilemma between affirming existing social norms or nudging those norms in new and different directions; and (4) dilemma of balancing between effectiveness and legitimacy in techno-regulation (Leenes et al., 2017).

V. ETHICAL CHALLENGES

176. Robots can be reckoned among the disruptive technologies of our age. This is not merely the case because of their profound impact on society, as described in Chapter III, but also because of the novel way in which they challenge the frameworks available for ethical reflection and evaluation. Robotic technologies blur the boundary between human subjects and technological objects, by demonstrating forms of agency that resemble human agency in some forms, e.g. the ability to make decisions, to 'show' and 'interpret' emotions, and to take over and assist in human tasks like driving cars, teaching or providing care. In doing so, they raise the need to rethink basic ethical concepts like agency and responsibility. Moreover, their rapid development and acceptance in society might also result in a change in value frameworks, affecting what we value in various domains of our lives, like care, work, education, friendship, and love.

V.1. Techno-pessimism, techno-optimism, and beyond

177. The societal discussion about robotics plays itself out between two types of positions: techno-optimistic views, with the utopian approach of transhumanism as its extreme variant, and techno-pessimistic views, with fears about a dystopian future and an approach of bioconservatism to defend humanity against too radical technological impacts.

178. The techno-optimistic position expects a better future through technology. In its mildest form, it embodies an instrumental approach to technology, believing that technology provides the means to solve societal problems and to make a better future. Self-driving cars will result in fewer accidents, surgical robots will be more accurate in medical operations, service robots will improve the quality of life of elderly people and people with a chronic illness, etc. In its strongest form, the techno-optimistic view results in the transhumanist ideal to direct technological innovation towards 'human enhancement'. Robotizing human tasks and activities, and maximizing the human-machine hybridity can be seen as part of this transhumanist programme, which is often linked to science-fiction scenarios and a discourse of technological innovation in a flourishing capitalist economy, promising general welfare and individual enhancement (Ford, 2015; Hottois et al., 2015). The transhumanist movement has been criticized for defending a too narrow vision of the human being, focusing on autonomy, individualism and competitiveness, and therefore in fact being hyper-humanistic rather than trans-humanistic.

179. Techno-pessimistic positions hold a radically different vision, being suspicious about the implications of technological innovations, and fearing that technology will ultimately not bring a brighter, but a darker future, affecting social relations, human cognition, our cognitive capacities, etc. Its most radical form is the movement of bioconservatism, which aims to preserve the human being by defending central elements of human existence like human dignity (Fukuyama, 2002), the character of human life as a gift (Sandel, 2009), and the ability to be the author of one's own life (Habermas, 2003). It sometimes expresses a profoundly pessimistic vision of the future of humankind and the planet, fearing that modern Western civilization will collapse and is already in a phase of decline, without any way to escape its destiny.

180. Because of their abstract and radical character, the extreme positions of transhumanism and bioconservatism do not provide real guidance to address the actual development, implementation and use of robots. Even though the disruptive character of robotic technologies obviously raises the question to what extent humans should be defended against technology or be merged with it, other questions deserve attention as well, which focus more directly and concretely on the implications of actual robotic developments for actual social practices, at various levels.

a. At the personal level: What kind of robots do we want to allow in the private sphere of our personal lives? What or who do we want to include In our oïkos (our house, community), and in what kind of practices and relationships do we want to give them a place – in educating children, nursing, caring for elderly people, cleaning and cooking, love and affection? How will robots affect human identity, and the quality of interpersonal relationships?
b. At the societal level: To what extent do we want to delegate work to robots? What will be the social and cultural transformations as a result of this? What will be the implications for human dignity, and for issues regarding equality and inclusion? What kind of dependency on robots will it generate? (Carr, 2014)
c. At the political level: To what extent should armed drones or autonomous robots be used to kill people at war or injure or incapacitate for security purpose? Can and should we delegate decisions about life and death to robots? Can and should robots be programmed with ethical principles for 'just warfare'?

181. In all of these questions, the disruptive character of robot technologies plays a central role: robots blur the boundaries between the human and the technological, raising fundamental questions about responsibility, agency, and the moral status of robots.

V.2. Robots and responsibility

182. Bearing in mind the complexity of contemporary robots, as well as the complexity of their design, construction and programming, the question arises who exactly should bear ethical and/or legal responsibility in cases when the functioning or malfunctioning of a robot

brings about harm to human beings, property or the environment. This question becomes even more pressing as robots are becoming more autonomous, more mobile and more present in ever more areas of human life.

183. For example, in case an autonomous robotic car causes an accident with human casualties, with whom does the responsibility lie? With 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 a car? The robot itself? Similar ethical and legal dilemmas are possible with medical robots (e.g. robots performing surgery resulting in death or harm to a patient), military and security robots (e.g. military robots killing civilians), service robots (e.g. kitchen robots preparing meals or drinks harmful for the user's health) or industrial robots (e.g. robots that injure people or work with toxic substances that may leak into the environment). Possible harm to humans does not have to be physical; it can also be psychological, e.g. when a personal robot breaches one's privacy, or when the interaction with a robot, due to its humanoid characteristics, triggers strong emotions and attachment (which is especially likely in children and the elderly).

184. In all of these cases, there seems to be a 'shared' or 'distributed' responsibility between robot designers, engineers, programmers, manufacturers, investors, sellers and users. None of these agents can be indicated as the ultimate source of action. At the same time, this solution tends to dilute the notion of responsibility altogether: if everybody has a part in the total responsibility, no one is fully responsible. This problem is known as the 'problem of the many hands'. The problem cannot be solved by claiming that roboticists should be held responsible anyway for any harm that ensues from using their products. This would deny the possibility of 'dual-use': robots, like so many other technologies, can be used for both good and bad purposes. This is the same dilemma facing most research sciences, for example, chemists that apply the same chemicals to make valuable pharmaceuticals that are used to make chemical weapons. Robots may be used for purposes intended by their designers, but they may also be used for a variety of other purposes, especially if their 'behaviour' can be 'hacked' or 'reprogrammed' by their end-users. Another difficulty is the context-dependency or 'multistability' (Ihde, 1990) of the impact of technologies. Just like the typewriter was originally developed as a tool for visually impaired people and ended up having major implications for office work, robots might have implications far beyond the intentions of their developers. It is impossible for roboticists to predict entirely how their work might affect society.

185. Avoiding the potential paralyzing effect of this difficulty to take and attribute responsibility, then, is a major challenge for the ethics of designing, implementing and using robotic technologies, especially in view of the fact that robotics teams tend to be very large and that their individual members may even be unaware of the final implications of their segment of work. In order to take responsibility anyway, at least two routes have been proposed. The first is to develop techniques to anticipate the impacts of robotic development as much as possible (Waelbers and Swierstra, 2014; Verbeek, 2013). The second is to deal carefully with the inevitable occurrence of unexpected implications, by considering the societal introduction of robotic technologies as a 'social experiment' that needs to be conducted with great care (Van de Poel, 2013).

V.3. Non-human Agency

V.3.1. Robots as agents

186. Because of their ability to act autonomously, and to interact independently with their environment, robot behaviour has characteristics of 'agency': the capacity to 'act'. This capacity is a central element of the disruptive character of robotic technologies. Agency, after all, has always been considered a characteristic of humans, not of machines (Coeckelbergh, 2012). There are clear differences between human agency and robotic 'agency': the agency of robots is not as autonomous as human agency is, since it has its origins in the work of designers and programmers, and in the learning processes that cognitive robotic systems

have gone through. However, at the same time, robots end up 'doing' things that are the results of their own decision-making processes and interactions, and not only of the input given by their developers.

187. This question of agency becomes even more urgent and more disruptive when it comes to moral agency. In order to be a moral agent and to be held morally accountable for one's actions, an agent needs to have both the freedom and the intention to act, rather than being steered or forced or to do something accidentally. Robots do not have freedom and intentions the way humans do. However, at the same time, their degrees of freedom are larger than those of regular machines, since they have decision-making capacities, and they do have a form of intentionality because of the algorithms that give them a particular directedness towards their environment and inclination to 'act'.

188. In philosophy of technology, the approach of 'technological mediation' offers an alternative solution to the discussion about nonhuman agency. Building upon the work of Don Ihde (1990), an approach to technology has been developed in which technologies are not seen as nonhuman objects opposed to human subjects, but as 'mediators' between humans and their environment (Rosenberger and Verbeek, 2015). Technologies-in-use are typically not part of the world people experience, but of the relations between humans and the world: cell phones connect people to each other, MRI imaging devices connect medical doctors to the bodies of patients, etc. From this mediating role, technologies help to shape how human beings experience the world and how they act in it. This framework offers an escape from the dilemma of seeing robots either as technological 'objects' or as quasi-human 'subjects'. When robots are introduced in practices – ranging from nursing to teaching and from cleaning to warfare – they start to mediate these practices and change how humans do nursing, teaching, etc. From this perspective, rather than 'being' full-blown agents themselves, robots mediate agency. The main question, then, is not whether they can compare to or even outdate humans, but how they change human practices, shifting the focus of ethical work towards analysing the quality of human-robot relations and taking these analyses into the design, implementation and use of robots.

V.3.2. Robots as moral agents

189. The potential harm caused by robots is often related to the fact that many robots are designed to perform services to or work in close interaction with human beings. In such settings, harm arises due to inevitable errors in design, programming and production of robots. It is likely that potential malfunctioning of today's sophisticated robots is capable of inflicting significant harm to a very large number of human beings (e.g. when armed military robots or autonomous robotic cars get out of control). The question, therefore, is not only if roboticists ought to respect certain ethical norms, but whether ethical norms need to be programmed into the robots themselves. Such a need is already apparent if one focuses on personal robots and the possible harm they could inflict on humans (e.g. robots for cooking, driving, fire protection, grocery shopping, bookkeeping, companionship, nursing) or for self-driving cars, that might need to make decisions about life and death about its passengers and other people on the road in the case of unexpected events. Since the autonomy of robots is likely to grow, their ethical regulation will increasingly need to be specifically designed to prevent harmful behaviour.

190. The emerging discipline of 'machine ethics' is "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, p.1). This new field of research already has its specializations, such as '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" (Van Rysewyk and Pontier, 2015, p.7).

191. Two important questions need to be considered in this context. First of all: is it possible to construct some kind of 'artificial moral agents'? And, second, if so: 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, especially to the human ability for reasoning according to moral principles and for applying those principles in concrete and different situations. However, they also seem to agree that moral agency comes in degrees, and that we need to distinguish between 'implicit' and 'explicit' ethical agents.

192. 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 programme 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 to endanger passengers' safety). It is also argued by some philosophers (Savulescu and 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), supporting human moral agency in new ways.

193. An even more intriguing question is whether robots can ever become 'explicit' ethical agents in ways comparable to human beings. Is it possible to programme machines or robots to 'act on principles' or to 'have virtues' or to 'provide moral reasons for their actions'? Having these capacities, it is frequently argued, requires having freedom of the will, which, allegedly, is unique only to human beings. Can robots have 'moral knowledge' and apply it in a range of different and possibly very complex moral dilemmas? According to Moor (2011), 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" (p.17). 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 pre-programmed, 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).

194. 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, p.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, other sentient beings and their environments. To mention just one example: human morally relevant preferences (like the preference to pursue various benefits and avoid various harms), do not only vary 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.

195. Using robots in most areas of life is generally considered acceptable as long as they perform their tasks better, and commit fewer errors, than humans would. The same logic could plausibly apply when it comes to using robots as 'implicit' ethical agents or artificial 'moral advisors', thus providing 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 'ethical 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 was the case with human calculating and writing abilities after the spread of pocket calculators and personal computers.

196. When it comes to the creation of robots as 'explicit' ethical agents (i.e. agents capable of general moral decision-making and independent implementation of those decisions), particular ethical attention is needed, despite the fact that the possibility to develop such robots is still speculative. The existence of such hypothetical robots could only be desirable when they could 'ethically outperform' humans, for instance by being better able to analyse complex ethical problems, or by not suffering from typically human moral weaknesses. This, however, would presuppose the existence of a standard by which to measure ethical performance, and such a standard is not available, certainly not at a universal level. The definition of a 'good moral agent' has varied through historical and cultural contexts.

197. Moreover, despite the fact that it is highly questionable whether such 'explicit' ethical agents will ever become technically possible, it is unclear at which point a machine could be said to have become an 'explicit' ethical agent. Given the fact that the Turing Test, which is considered the theoretically best proposal for testing the existence of artificial intelligence, still remains controversial itself, it is even more controversial to claim that some analogue test (a 'moral Turing Test') could be devised to test the possible creation of 'explicit' ethical agents (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.

198. The problem of testing and verifying the existence of artificial and 'explicit' ethical agents is related to a 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), 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, to develop a 'robot virtue ethics' focused on 'the search for the virtues a good (properly functioning) robot would evince, given its appropriate roles' (Abney, 2012).

199. Among other proposals currently discussed are Kantianism (Powers, 2011), pluralistic theory of prima facie duties (Anderson and Anderson, 2011b), Buddhist ethics (Hughes, 2012) and divine-command ethics (Bringsjord and Taylor, 2012). This variety of approaches shows that 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. There are no definitive philosophical answers to questions about the nature of morality itself, and about the objectivity or subjectivity of ethical frameworks themselves. The discussion about the ethical agency of robots reproduces the variety of positions in ethical theory in general, which gives rise to the question whether it may be necessary to develop the same ability to accept and respect this variety of frameworks and orientations that also exists regarding human ethical frameworks.

V.4. The moral status of robots

200. Another important disruptive aspect of robotic technologies with enhanced autonomy and capacities for decision-making – possibly even including moral decision-making – concerns their moral status. Will robots ultimately not only become moral agents that are capable of moral action, but also entities that become morally valuable beyond their instrumental value as devices merely manufactured to perform specific tasks? Would such robots deserve the same moral respect and immunity from harm, as is currently the case with humans and some non-human animals? Could robots ultimately not only have obligations and duties, but also moral rights?

201. How moral status is acquired and lost is a longstanding philosophical issue. Some philosophers believe that having moral status amounts to having certain psychological and/or biological properties. From a deontological point of view, to have moral status implies being a person, and being a person implies having rationality or the capacity for rational and moral 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 freedom of will, intentionality, self-consciousness, moral agency or a sense of personal identity. It should be mentioned in this 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" (JURI, 2016, section 59.f).

202. From a 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 some authors (e.g. 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 or at least some version of moral rights.

203. 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 or infants) 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' (or 'harm' in general) 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 or physical punishment).

204. A possible third way of assigning moral status to robots (a way that does not focus on any particular psychological or biological property) is to adopt a '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. Such a 'relational' approach to the problem of moral status has a high intuitive appeal, because humans are social and emotional beings that are naturally 'related to' and 'care about' other human beings. The same approach is occasionally used in environmental ethics as well, in the sense that entire ecosystems deserve moral protection (some kind of moral status) due to the 'interconnectedness' and vitally important 'relationships' between their living and non-living components.

205. When it comes to robots, however, this 'relational' solution could face the problem of depending on the human psychological tendency to anthropomorphize or 'project' human properties onto inanimate objects and artefacts. Just as humans sometimes develop strong and life-long attachment to their cars, boats or guitars, they can also develop such forms of attachment to robots. 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). Yet, this bonding between humans and robots is not necessarily the result of anthropomorphization: also without being human-like, technological artefacts like robots can become so meaningful and valuable that they deserve to be protected, as the many UNESCO sites of cultural heritage illustrate.

206. The rapid development of highly intelligent autonomous robots, then, 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 the 'human-machine divide' is gradually disappearing (Romportl et al., 2015) and the likelihood of future appearance of human-machine or animal-machine hybrids or cyborgs (robots integrated with biological organisms or at least containing some biological components).

V.5. Value dynamism

207. A separate disruptive effect of robotic technologies in society is their impact on moral frameworks: technologies do not only have societal effects that can be ethically evaluated, but they also affect the very ethical frameworks with which we can evaluate them. Technologies can change human values and normative orientations. The introduction of the birth control pill, for instance, has changed value frameworks regarding sexuality, by loosening the connection between sexuality and reproduction, making room for new valuations of hetero- and homosexuality (Mol, 1997). Moreover, the introduction of augmented reality technology such as Google Glass appears to affect what 'privacy' means in our society (Kudina and Bas, 2017), giving rise to new definitions of privacy in relation to the specific ways in which such technologies redefine the boundaries between the public and the private. This type of influence creates a specific challenge for ethical reflection on technology: the values we use to evaluate technology cannot be considered as pre-given standards but rather change over time, in interaction with the very technologies we want to evaluate with them.

208. This phenomenon of 'value dynamism' has a central place in the contemporary approach of 'technomoral change'. By developing scenarios of potential technological futures, this approach aims to anticipate moral change connected to technological developments, in order to inspire technological practices and policy-making (Swierstra et al., 2009). Also the approach of 'technological mediation' embodies a view on value dynamism, indicated as 'moral mediation': by mediating human practices, perceptions, and interpretations, technologies help to shape moral actions and decisions, and have an impact on moral frameworks (Verbeek, 2011).

209. In the field of robotics, no explicit research has been done yet into the ways in which robots might affect ethical values and normative frameworks. However, on the basis of the approaches of technomoral change and moral mediation it is very well possible to envisage such impacts. Care robots might change what humans value in care, both physically and emotionally. Teaching robots are likely to affect our criteria for good education, and our valuations of the roles of teachers and students. Self-driving cars might bring new criteria for 'driving capability', both for humans and for robots. Sex robots could have an impact on the relations between love and sex, and on what we value in intimate relationships with other people.

210. Dealing with such normative impacts in a responsible way requires a careful balance between anticipation and experimentation. By developing technomoral scenarios, the approach of technomoral change aims to provide users, designers, and policy-makers with a basis for making decisions in the present about potential future societal effects. In addition to this approach, Van De Poel (2013) proposed to treat the development of emerging – and also disruptive – technologies as 'social experiments'. According to Van de Poel, anticipation runs the risk of "missing out on important actual social consequences" and of "making us blind to surprises" (Van de Poel, 2016, p.667). Therefore, he states that innovations are in fact always social experiments. This requires us to conduct those experiments in a responsible way. Following from this, the phenomenon of value dynamism would require us to closely follow the impact of robotic technologies on value frameworks, in small-scale experimental settings, in order to be able to take this impact into account in design practices, public discussions, and policy-making.

VI. RECOMMENDATIONS

VI.1. A technology-based ethical framework

211. As discussed in the general introduction to this report, early robots were programmed to do clearly defined tasks. These we can usefully categorise as *deterministic* robots: their actions are controlled by a set of algorithms whose actions can be predicted.

212. With developments in advanced computing, the concept of computers demonstrating *artificial intelligence* arose. Although this term may be subject to different interpretations, and some may regard as implying real human-like *intelligence*, AI-based machines can demonstrate human-like sensory ability, language, and interaction. Moreover, these machines can show human-like learning capability, which is being improved – and will continue to be further improved – by employing deep learning techniques.

213. These developments are leading to what may be termed *cognitive robotics*. This is a field of technology involving robots that are based on cognitive computing and therefore can learn from experience, not only from human teachers, but also on their own, thereby developing an ability to deal with their environment on the basis of what has been learned. Compared to 'traditional' or deterministic robots, cognitive robots can make decisions in complex situations, decisions that cannot be predicted by a programmer.

214. In considering recommendations regarding robotics ethics, this distinction between deterministic and cognitive robots is important. In the deterministic case, the behaviour of the robot is determined by the program that controls its actions. Responsibility for its actions is therefore clear, and regulation can largely be dealt with by legal means. In the cognitive case, a robot's decisions and actions can be only statistically estimated, and are therefore unpredictable. As such, the responsibility for the robot's actions is unclear and its behaviour in environments that are outside those it experienced during learning (and so in essence 'random') can be potentially catastrophic. So assigning responsibility for the actions of what is partly a *stochastic* machine is problematical. Nevertheless, if we are to live with and regulate the use of such machines, this issue of responsibility must be tackled.

215. Moreover, to be reasonably future proofed in an area such as this which is developing rapidly, any sensible set of recommendations must, as well as considering robots in their present stage of development, consider and try to take account of where present trends may be leading.

216. Accordingly, COMEST proposes to consider recommendations based upon the above. In the first level of deterministic machines where the responsibility for behaviour can be assigned, the Commission's recommendations will largely focus on legal instruments to regulate their use. For the second level of cognitive machines, whose behaviour cannot be 100% predictable and therefore is in significant part stochastic, in addition to legal instruments, codes of practice and ethical guidelines for both producers and users need to be considered. Finally, where stochastic machines can be put in situations where harm can be caused (for example a self-driving car or an autonomous weapon), we need to consider the degree of autonomy that can reasonably be left to the machine, and where meaningful human control must be retained.

217. The scheme is illustrated in the table below. While the proposed structure is simple, its implementation in terms of assigning accountability and regulating use is complex and challenging – for scientists and engineers, policy makers and ethicists alike. Nevertheless, if we are to live in a future world where robotics increasingly takes over tasks that presently are performed by humans, this challenge must be effectively met.

Decision by Robot	**Human Involvement**	**Technology**	**Responsibility**	**Regulation**
Made out of finite set of options, according to preset strict **criteria**	Criteria implemented in a legal framework	Machine only: deterministic algorithms/ robots	Robots' producer	Legal (standards, national or international legislation)
Out of a range of options, with room for flexibility, according to a preset **policy**	Decision delegated to robot	Machine only: AI- based algorithms, cognitive robots	Designer, Manufacturer, Seller, User	Codes of practice both for engineers and for users; Precautionary Principle
Decisions made through human-machine **interaction**	Human controls robot's decisions	Ability for human to take control over robot in cases where robot's actions can cause serious harm or death	Human beings	Moral

VI.2. Relevant ethical principles and values

218. In regard of the diversity and the complexity of robots, a framework of ethical values and principles can be helpful to set regulations at every level – conception, fabrication and utilization – and in a coherent manner, from engineers' codes of conduct to national laws and international conventions.

219. The principle of human responsibility is the common thread that joins the different values that are enunciated here.

VI.2.1. Human Dignity

220. **Human dignity** is a core value related to the *Universal Declaration of Human Rights* (UN, 1948). It recognizes that, being free and equal, all human beings "are endowed with

reason and conscience and should act towards one another in a spirit of brotherhood" (Art. 1).

221. Dignity is inherent to human beings, not to machines or robots. Therefore, robots and humans are not to be confused even if an android robot has the seductive appearance of a human, or if a powerful cognitive robot has learning capacity that exceeds individual human cognition. Robots are not humans – they are the result of human creativity and they still need a technical support system and maintenance in order to be effective and efficient tools or mediators.

VI.2.2. Value of Autonomy

222. The recognition of human dignity implies that the **value of autonomy** does not solely concern the respect of individual autonomy, which can go as far as to refuse to be under the charge of a robot. The value of autonomy also expresses the recognition of *the interdependency of relationship* between humans, between humans and animals, and between humans and the environment. To what extent social robots will enrich our relationships, or reduce and standardise them? This needs to be scientifically evaluated in medical and educational practices where robots can be used, especially when vulnerable groups such as children and elderly persons are concerned. The extensive use of robots can accentuate in certain societies the rupture of social bonds.

223. Interdependency implies that robots are part of our technical creations (part of the technocosm that we construct) and they also have *environmental impacts* (e-waste, energy consumption and CO_2 emissions, ecological footprint) that must be considered and evaluated in the balance of benefit and risk.

VI.2.3. Value of Privacy

224. **The value of privacy** is related to Article 12 of the *Universal Declaration of Human Rights* (UN, 1948) which states that:

> No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.

225. Various protection schemes and regulations have been implemented in many countries to limit access to personal data in order to protect the privacy of individuals. However, the advent of Big Data changes the way data are collected and how they are processed (use of algorithm in profiling). The scale is much wider and the uses are expanding (e.g. commercial, state security and surveillance, research, etc.), and so are the forms of intrusion. Robots are devices that can collect data through sensors and that can use Big Data through deep learning. Therefore, collection and use of data need to be scrutinized in the design of robots, using an approach that balances the aim of the robot and the protection of privacy. Some data may be more sensitive than others; therefore a mix of approaches such as legislation, professional regulations, governance, public surveillance, etc. is necessary in order to maintain public trust in and good use of robots.

VI.2.4. 'Do not harm' Principle

226. **'Do not harm' principle** is a red line for robots. As many technologies, a robot has the potentiality for 'dual-use'. Robots are usually designed for good and useful purposes (to diminish harmfulness of work for example), to help human beings, not to harm or kill them. In this regard, Isaac Asimov's formulation of this principle (three laws) is still accurate (see paragraph 18. If we are morally serious about this ethical principle, then we have to ask ourselves whether armed drones and autonomous weapons should be banned.

VI.2.5. Principle of Responsibility

227. The COMEST Report on the Precautionary Principle (UNESCO, 2005b) states that "[e]thical responsibility implies some freedom of choice in action" (p.17). With regard to the

use of robots, “[t]he notion that individuals (or firms, or States) are morally responsible for the choice they make” is highly significant. It is because human beings are responsible for their acts that they can be blamed and they can face legal responsibility if their actions lead to damage. In this perspective, there is no delegation of human ethical responsibility to robots.

228. Deterministic robots, and even sophisticated cognitive robots, cannot take any ethical responsibility, which lies with the designer, manufacturer, seller, user, and the State. Therefore, human beings should always be in the loop and find ways to control robots by different means (e.g. traceability, off switch, etc.) in order to maintain human moral and legal responsibility.

229. In the development of robotics, three dimensions of responsibility need to be clarified and improved: **liability**, **transparency** and **accountability**. The question of liability is a serious matter for cars with a degree of automation. The implementation of tracking and recording systems can be a possibility to help determine responsibility, but it could challenge privacy and data protection rights.

230. There is a strong link between ethical responsibility and the **precautionary principle**. COMEST defined it as follows:

> When human activities may lead to morally unacceptable harm that is scientifically plausible but uncertain, actions shall be taken to avoid or diminish that harm. *Morally unacceptable harm* refers to harm to humans or the environment that is
>
> - threatening to human life or health, or
> - serious and effectively irreversible, or
> - inequitable to present or future generations, or
> - imposed without adequate consideration of human rights of those affected. (UNESCO, 2005b, p.14)

The application of the precautionary principle reinforces the role of monitoring and of systematic empirical research using a wide range of scientific disciplines.

231. This principle is particularly significant in the development of autonomous robots (cognitive robots) with deep learning capacity where some behavioural or decision-making processes cannot be programmed in the same way as for deterministic robots. A degree of uncertainty is inherent in cognitive robots and roboticists cannot just rely on their capacity to evolve without considering cumulative risks and negative effects over a long time scale.

232. Robotics researchers, industries, and governments are partners in ***responsible research and innovation***. From this perspective, the development paths of robotics should not be reduced to economic productivity and efficiency. Instead, it should include public engagement in order to choose paths of scientific development that contribute to the common good and are adapted to cultural context.

VI.2.6. Value of Beneficence

233. Robots are useful for facilitating better safety, efficiency, and performance in many human tasks that are physically hard. Industrial robots, disaster robots, and mining robots can be used to replace human beings in dangerous environments. However, the beneficence of robots is subject to further discussion and reflection when they are designed to interact in a social context, such as in education, health care or surveillance/policing by the State.

234. The value of beneficence should be balanced with the **principle of proportionality** in terms of available technological choices. In the case of robots, some questions are important: What is the final purpose that is at stake? Does it take into account the social and cultural context in the assessment and implementation process? Is the particular type of robot used imposed on people or has it been designed for the people and eventually with the people? More so in developing countries than in developed countries, the use of robots needs to be balanced with other social and economic priorities.

235. This leads to the promotion of the **value of cultural diversity**. In 2005, UNESCO adopted the *Convention on the Protection and Promotion of the Diversity of Cultural Expressions*, which refers to cultural diversity as:

> […] the manifold ways in which the cultures of groups and societies find expression. These expressions are passed on within and among groups and societies. Cultural diversity is made manifest not only through the varied ways in which the cultural heritage of humanity is expressed, augmented and transmitted though the variety of cultural expressions, but also through diverse modes of artistic creation, production dissemination, distribution and enjoyment, whatever the means and technologies used. (UNESCO, 2005a, p.4)

A greater sensitivity to cultural and gender issues should drive research and innovation in robotics. Due to the diversity of cultures, robots – especially social robots – may be accepted in certain settings and not in others.

VI.2.7. Value of Justice

236. The value of justice is related to inequality. The extensive use of industrial robots and service robots will generate higher unemployment for certain segments of the work force. This raises fears concerning rising inequality within society if there are no ways to compensate, to provide work to people, or to organize the workplace differently. Work is still a central element of social and personal identity and recognition.

237. The value of justice is also related to non-discrimination. Roboticists should be sensitised to the reproduction of gender bias and sexual stereotype in robots. The issue of discrimination and stigmatisation through data mining collected by robots is not a trivial issue. Adequate measures need to be taken by States.

VI.3. COMEST specific recommendations on robotics ethics

VI.3.1. Recommendation on the Development of the Codes of Ethics for Robotics and Roboticists

238. It is recommended that, at both national and international levels, codes of ethics for roboticists be further developed, implemented, revised and updated, in a multidisciplinary way, and responding to possible future advancements of robotics and its impact on human life and the environment (energy, e-waste, ecological footprint). It is also recommended that disciplines and professions significantly contributing to or potentially relying on robotics – from electronic engineering and artificial intelligence to medicine, animal science and psychology and the physical sciences – revise their particular codes of ethics, anticipating challenges originating from their links to robotics and the robot industry, preferably in a coordinated way. Finally, it is recommended that ethics – including codes of ethics, codes of conduct, and other relevant documents – become an integrated part of the study programmes for all professionals involved in the design and manufacturing of robots.

VI.3.2. Recommendation on Value Sensitive Design

239. When designing robotic technologies, ethical considerations should be taken into account. Robots use algorithms to make decisions, which embody ethical values and frameworks. In addition, robots have ethical implications for the practices in which they are used, like health care, education, and social interactions. In order to address these ethical dimensions of robots, ethics needs to be part of the design process, building on approaches like the Value Sensitive Design approach. This approach should also be adapted to consider animal welfare.

VI.3.3. Recommendation on Experimentation

240. The social implications of new robotic technologies are often hard to predict. In order to deal responsibly with the social introduction of robots, careful and transparent experimentation is needed. By introducing new robotic technologies in small-scale, well-

monitored settings, the implications of these technologies on human practices, experiences, interpretational frameworks, and values can be studied openly. The outcomes of such experiments can be used to adapt the design of robots, to inform policy-making and regulation, and to equip users with a critical perspective.

VI.3.4. Recommendation on Public Discussion

241. Robots will have profound effects on society and on people's everyday lives. In order to deal with these effects in a responsible way, citizens need to be equipped with adequate frameworks, concepts, and knowledge. For that purpose, public discussions need to be organized about the implications of new robotic technologies for the various dimensions of society and everyday life, including environmental impact of the entire robot production cycle, which help people to develop a critical attitude, and which sharpen the awareness of designers and policy-makers.

VI.3.5. Recommendation on Retraining and Retooling of the Workforce

242. Robots will increasingly displace humans in a wide range of areas and so lead to significant reduction in job opportunities in certain sectors. It will also give rise to new job opportunities. States, professional organizations and educational institutions should therefore consider the implications of this, paying particular attention to those sections of society likely to be most vulnerable to the changes, and make appropriate provision for retraining and retooling of the work force to enable the potential advantages to be realized.

VI.3.6. Recommendations related to Transportation and Autonomous Vehicles

243. Autonomous Vehicles (AV) are controlled by two types of algorithms: deterministic algorithms, for which the outcome in a certain situation is predictable; and cognitive, AI algorithms where it is not. With respect to autonomous (self-driving) vehicles, COMEST makes the following recommendations.

244. Functions of an AV that are predictable (e.g. keeping privacy of users) should be identified, and their implementation imposed by deterministic algorithms. With respect to these functions, the AVs can be considered as conventional technology (e.g. computers), and so existing legal and ethical frameworks can be applied.

245. However, the unique features of AVs are their ability to operate and decide based on their machine learning, cognitive algorithms. Even if the optimisation criteria for cognitive algorithms are transparent, the actual decision made by an AV is unpredictable, since it depends on its specific, essentially random experience. The question of who is responsible for the consequences of such (unpredictable) decisions has deep ethical aspects.

246. In general, both ethical and legal frameworks have tools for dealing with non-deterministic (random) situations. One example is natural disasters, where insurance tools have been developed to deal with unforeseen damage. COMEST recommends that a similar framework should be applied for unpredictable consequences of AVs.

247. However, it should not be argued that all unforeseen consequences of an AV's decision should be considered, in essence, as 'acts of God' like natural disasters – there are AV's decisions where a human should be in control, because moral considerations cannot be programmed. COMEST therefore recommends that situations where the responsibility for the results of the AV's action is put solely on a human ('the driver') be identified and defined. In particular, in decisions related to, for example, possible loss of life (e.g. unavoidable possible fatal accident), the autonomy of the vehicle must be limited, and the responsibility for the critical decision put on a human. The human can then make the required decision independently, or follow the AV's advice, or even delegate the decision to the AV. However, in all cases, the human is accountable.

VI.3.7. Recommendations on Armed Military Robotic Systems ('Armed Drones')

248. Armed drones have given humanity the ability to wage war remotely. Though on the surface this may be attractive in reducing the need for 'boots on the ground', it threatens to change fundamentally the nature of conflict. These weapons therefore raise serious legal and ethical issues that States seem so far to have failed to address.

249. The ethical issues of using armed drones go beyond the legal issues of International Humanitarian Law. For example, one-sided remote warfare is massively asymmetric, with an attacker in a position to kill an adversary without any threat to him/herself. An ethical justification for such a situation is difficult to find. Remote killing also contravenes the principle of Human Dignity. Finally, the ability to go to war without exposing one's own soldiers to direct threat lowers the perceived cost, and hence the activation barrier, of declaring war. This raises the worrying prospect of low cost continuous warfare.

250. The use of armed drones against suspected non-state actors in insurgences raises additional ethical and legal questions. Targeted killing by an armed drone removes the right to justice of not only the immediate individual human target who will be deprived of legal hearing. Where such killings are a frequent occurrence in attempting to counter insurgencies, the consequent negative economic, social and psychological effects on the civilian population – especially of young children – are of strong ethical and human rights concern.

251. COMEST concludes therefore that, in addition to legal issues, there is a strong moral principle against an armed robot killing a human being. The Commission recognises that certain countries are increasingly using armed drones in conflict situations. However, the ethical and legal case against such use is sufficiently strong that COMEST recommends States reconsider this practice – as they have done for other weapons that have been limited or made illegal such as anti-personnel mines and chemical and biological weapons. The Article 36 process – even if used retrospectively – perhaps provides a mechanism for States to do this. If done transparently and in the context of internationally agreed norms, States could assess not just if these weapons can in some circumstances be used legally and ethically, but also set out the situations in which they would possibly be used in such a way as to fulfil both International Humanitarian Law and Human Rights Law. From an ethical standpoint, we doubt that such situations can be found, and unless action is taken before these weapons proliferate further, we fear the future prospect of continuous remote conflict and justice-denying targeted killing.

VI.3.8. Recommendations on Autonomous Weapons

252. Two points stand out with respect to the possible use of autonomous weapons. Legally, their deployment would violate IHL. Ethically, they break the guiding principle that machines should not be making life or death decisions about humans.

253. With respect to their technical capability, autonomous robotic weapons lack the main components required to ensure compliance with the principles of distinction and proportionality. Though it might be argued that compliance may be possible in the future, such speculations are dangerous in the face of killing machines whose behaviour in a particular circumstance is stochastic and hence inherently unpredictable.

254. The moral argument that the authority to use lethal force cannot be legitimately delegated to a machine – however efficient – is included in international law: killing must remain the responsibility of an accountable human with the duty to make a considered decision.

255. Our strong, single recommendation is therefore that, for legal, ethical and military-operational reasons, human control over weapon systems and the use of force must be retained. Considering the potential speed of development of autonomous weapons, there is an urgent need (as the ICRC has urged) "to determine the kind and degree of human control over the operation of weapon systems that are deemed necessary to comply with legal obligations and to satisfy ethical and societal considerations" (ICRC, 2016).

VI.3.9. Recommendations on Surveillance and Policing

256. States should draw up policies on the use of drones in surveillance, policing, and in other non-military contexts. Usage policy by police should be decided by the public's representatives, not by police departments, and the policies should be clear, written, and open to the public. This policy should, at minimum, assure that a drone is 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. 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. Use of drones should be subject to open audits and proper oversight to prevent misuse.

257. Drones in police use should not be equipped with either lethal or non-lethal weapons.

258. Autonomous weapons should not be used in police or security use.

VI.3.10. Recommendation relating to Private and Commercial Use of Drones

259. The private use of drones should be under licence, and their areas of operation subject to strict control for safety, privacy and legal reasons. It should be unlawful to equip domestic drones with either lethal or non-lethal weapons.

VI.3.11. Recommendation on Gender Equality

260. Particular attention should be paid to gender issues and stereotyping with reference to all types of robots described in this report, and in particular, toy robots, sex companions, and job replacements.

VI.3.12. Recommendations on Environmental Impact Assessment

261. Similar to other advanced technologies, environmental impact should be considered as part of a lifecycle analysis, to enable a more holistic assessment of whether a specific use of robotics will provide more good than harm for society. This should address the possible negative impacts of production, use and waste (e.g., rare earth mining, e-waste, energy consumption), as well as potential environmental benefits. While constructing robots (nano, micro or macro), efforts should be made to use degradable materials and environmentally friendly technology, and to improve the recycling of materials.

VI.3.13. Recommendations on the Internet of Things

262. The Internet of Things (IoT) is a rapidly emerging technology where smart everyday physical devices, including home appliances, are interconnected. This enables using devices as sensors, and collecting (big) data that can be used for many purposes.

263. These networks of everyone and everything create entirely new options in so many areas such as augmented reality, tactile inter-networking, distributed manufacturing, and smart cities.

264. While IoT immediately raises ethical questions related to privacy, safety, etc., the next generation of IoT, sometimes referred to as IoT++, is even more challenging. In IoT++, artificial intelligence (AI) is used to process the collected data. The resulting cognitive algorithms, acting as independent learners, may lead to unpredictable results.

265. Similarly, emerging technologies create small-size robots[1] which can serve as mobile sensors, collecting information in targeted locations – enlarging the scope of IoT even beyond existing 'things'.

266. Identifying that the ethical challenges of IoT as similar, yet not identical, to that of cognitive robotics, the Commission recommends extending its work in this area to study IoT ethics and provide appropriate recommendations.

[1] Including micro-robots and nano-robots

BIBLIOGRAPHY

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

Allen, C., Varner, G. and Zinser, J. 2000. Prolegomena to any future artificial moral agent, *Experimental and Theoretical Artificial Intelligence*, Vol. 12, No. 3, pp. 251-261.

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

Alliance des sciences et technologies de numérique (Allistene). 2016. *Éthique de la recherche en robotique* [Ethics of research in robotics]. Paris, Allistene.

Alonso, E., Sherman, A. M., Wallington, T. J., Everson, M. P., Field, F. R., Roth, R., and Kirchain, R. E. 2012. Evaluating rare-earth element availability: a case with revolutionary demand from clean technologies, *Environmental Science and Technology*, Vol. 46, pp. 3406-3414.

Anderson, M. and Anderson, S. L. 2011a. General introduction. In: Anderson, M. and Anderson, S. L. eds. *Machine Ethics*. Cambridge, Cambridge University Press, pp. 1-4.

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: Anderson, M. and Anderson, S. L. eds. *Machine Ethics*. Cambridge, Cambridge University Press, pp. 476-492.

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

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

Asaro, P. M. 2015. The liability problem for autonomous artificial agents, *2016 AAAI Spring Symposium Series*. Available at: https://www.aaai.org/ocs/index.php/SSS/SSS16/paper/download/12699/11949

Asimov, I. 1942. *Runaround*. New York, Street & Smith.

Asimov, I. 1950. *I, Robot*. New York, Gnome Press.

Asimov, I. 1985. *Robots and Empire*. New York, Doubleday.

Balding, C. 2016. 'Will Robots Ravage the Developing World?', *Bloomberg View*, 25 July. London, Bloomberg. 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*. New York, Springer-Verlag.

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

Bekey, G., Ambrose, R., Kumar, V., Lavery, D., Sanderson, A., Wilcox, B., Yuh, J., Zheng, Y. 2008. *Robotics: State of the Art and Future Challenges*. London, Imperial College Press.

Benetti, F. B. V. 2012. Exploring the educational potential of robotics in schools: A Systematic review, *Computers & Education*, Vol. 58, No. 3, pp. 978-988.

Bensoussan, J. and Bensoussan, A. eds. 2016. *Comparative Handbook: Robotic Technologies Law*. Brussels, Editions Larcier.

Beran, T. N., Ramirez-Serrano, A., Kuzyk, R., Fior, M. and Nugent, S. 2011. Understanding how children understand robots: Perceived animism in child-robot interaction, *International Journal Human-Computer Studies*, Vol. 69, No. 7-8, pp. 539-550.

Bertoncello, M. and Wee, D. 2015. Ten ways autonomous driving could redefine the automotive world, *Automotive & Assembly*. Available at: http://www.mckinsey.com/industries/automotive-and-assembly/our-insights/ten-ways-autonomous-driving-could-redefine-the-automotive-world

Bonnefon, J.-F., Shariff, A. and Rahwan, I. 2016. The social dilemma of autonomous vehicles, *Science*, Vol. 352, No. 6293, pp. 1573-1576.

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

British Standard Institution (BSI). 2016. *Robots and robotic devices: Guide to the ethical design and application of robots and robotic systems*. London, BSI Standards Limited.

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

Butterfield, A., Ngondi, G. E. and Kerr, A. eds. 2016. *A Dictionary of Computer Science*. Oxford, Oxford University Press.

Calo, R. 2015. Robotics and the lessons of cyberlaw, *California Law Review*, Vol. 103, pp. 513-563.

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

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

Chang, C. W., Lee, J. H., Chao, P. Y., Wang, C. Y. and Chen, G. D. 2010. Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school, *Educational Technology & Society*, Vol. 13, No. 2, pp. 13-24.

Coeckelbergh, M. 2012. Can We Trust Robots?, *Ethics and Information Technology*, Vol. 14, pp. 53.

Comitato Nazionale per la Bioetica (CNB, Italy) and Comitato Nazionale per la Biosicurezza, le Biotecnologie e le Scienze della Vita (CNBBSV, Italy). 2017. *Developments of Robotics and Roboethics*. Joint Opinion. Rome, CNB and CNBBSV.

Committee of Legal Affairs of the European Parliament (JURI). 2016. *Draft Report with recommendations to the to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*. Brussels, European Parliament. Available at: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML%2BCOMPARL%2BPE-582.443%2B01%2BDOC%2BPDF%2BV0//EN

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

Crandall, J. and Armitage, J. 2005. Envisioning the Homefront: Militarization, Tracking and Security Culture, *Journal of Visual Culture*, Vol. 4, No. 1, pp. 17-38.

Diel, J. J., Schmitt, L. M., Villano and Crowell, C. R. 2012. The clinical use of robots for individuals with autism spectrum disorders: A critical review, *Research in Autism Spectrum Disorders*, Vol. 6, No. 1, pp. 249-262.

Driessen, C. P. G. and Heutinck, L. F. M. 2015. Cows desiring to be milked? Milking robots and the co-evolution of ethics and technology on Dutch dairy farms, *Agriculture and Human Values*, Vol. 32 No. 1, pp. 3-20.

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*. Hershey, IGI Global, pp. 1-30.

European Commission (EC). 2012. *Public Attitudes Towards Robots*. Report, Special Eurobarometer 382. Brussels, EC. Available at: http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_382_en.pdf

Floridi, L. 2011. On the morality of artificial agents. In: Anderson, M. and Anderson, S. L. eds. *Machine Ethics*. Cambridge, Cambridge University Press, pp. 184-212.

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

Frankish, K. and Ramsey, W. M. eds. 2014. *The Cambridge Handbook of Artificial Intelligence*. Cambridge, Cambridge University Press.

Franklin, S. 2014. History, motivations, and core themes. In: Frankish, K. and Ramsey, W. M. eds. *The Cambridge Handbook of Artificial Intelligence*. Cambridge, Cambridge University Press, pp. 15-33.

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

Fukuyama, F. 2002. *Our Posthuman Future: Consequences of the Biotechnology Revolution*. New York, Farrar, Straus, and Giroux

Gagnon, J.-A. 2018. Des enfants et des robots [Children and robots]. In: Parizeau, M.-H. ed. *Éthique des robots et transformations sociales* [Ethics of robots and social transformations]. Québec, Les Presses de l'Université Laval.

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

Habermas, J. 2003. *The Future of Human Nature*. Cambridge, Polity Press.

Heacock, M., Kelly, B. C., Asante, K. A., Birnbaum, L. S., Bergman, A. L., Brune, M. N., Buka, I., Carpenter, D. O., Chen, A., Huo, X., Kamel, M., Landrigan, P. J., Magalini, F., Diaz-

Barriga, F., Neira, M., Omar, M., Pascale, A., Ruchirawat, M., Sly, L., Sly, P. D., Van de Berg, M. and Suk, W. A. 2016. E-waste and harm to vulnerable populations: a growing global problem, *Environmental Health Perspectives,* Vol. 124, pp 550-555.

Holder, C., Khurana, V., Harrison, F. and Jacobs, L. 2016. Robotics and law: key legal and regulatory implications of the robotics age (part I of II), *Computer Law & Security Review*, Vol. 32, No. 3, pp. 383-402.

Holloway, L., Bear, C. and Wilkinson, K. 2014. Robotic milking technologies and renegotiating situated ethical relationships on U.K. dairy farms, *Agriculture and Human Values*, Vol. 31, No. 2, pp. 185-199.

Hottois, G., Missa, J-N., Perbal, L. eds. 2015. *Enclyclopédie du trans/posthumanisme* [Encyclopedia of trans/posthumanism]. Paris, Vrin.

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

Husbands, P. 2014. Robotics. In: Frankish, K. and Ramsey, W. M. eds. *The Cambridge Handbook of Artificial Intelligence*. Cambridge, Cambridge University Press, pp. 269-295.

Ihde, D. 1990. *Technology and the Lifeworld*. Bloomington, Indiana University Press.

Institute of Electrical and Electronic Engineers (IEEE) Standards Association. n.d. *The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems*. New Jersey, IEEE Standards Association. Available at: http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html

International Federation of Robotics (IFR). 2016a. *Executive Summary World Robotics 2016: Industrial Robots*. Frankfurt, IFR. Available at: https://ifr.org/img/uploads/Executive_Summary_WR_Industrial_Robots_20161.pdf

IFR. 2016b. *Executive Summary World Robotics 2016: Service Robots*. Frankfurt, IFR. Available at: https://ifr.org/downloads/press/02_2016/Executive_Summary_Service_Robots_2016.pdf

Ingram, B., Jones, D., Lewis, A., Richards, M., Rich, C. and Schachterle, L. 2010. A code of ethics for robotics engineers, *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.103-104.

International Committee of the Red Cross (ICRC). 1949a. *Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field*. Geneva, ICRC. Available at: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/INTRO/365?OpenDocument

ICRC. 1949b. *Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea*. Geneva, ICRC. Available at: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/INTRO/370?OpenDocument

ICRC. 1949c. *Convention (III) relative to the Treatment of Prisoners of War*. Geneva, ICRC. Available at: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/INTRO/375?OpenDocument

ICRC. 1949d. *Convention (IV) relative to the Protection of Civilian Persons in Time of War*. Geneva, ICRC. Available at: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/INTRO/380

ICRC. 1977a. *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*. Geneva, ICRC. Available at: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/INTRO/470

ICRC. 1977b. *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II)*. Geneva, ICRC. Available at: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/INTRO/475?OpenDocument

ICRC. 2007. *Distinction: Protecting Civilians in Armed Conflict*. Geneva, ICRC. Available at: https://www.icrc.org/en/publication/0904-distinction-protecting-civilians-armed-conflict

ICRC. 2014. *What is international humanitarian law?* Geneva, ICRC. Available at: https://www.icrc.org/en/download/file/4541/what-is-ihl-factsheet.pdf

ICRC. 2016. *Autonomous weapons: Decisions to kill and destroy are a human responsibility*. Published online 11 April 2016, Geneva, ICRC. Available at: https://www.icrc.org/en/document/statement-icrc-lethal-autonomous-weapons-systems

International Peace Conference. 1899. *Convention (II) with Respect to the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land*. The Hague, International Peace Conferences. Available at: https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/INTRO/615?OpenDocument

Kamga, R., Romero, M., Komis, V. and Mirsili, A. 2016. Design Requirements for Educational Robotics Activities for Sustaining Collaborative Problem Solving. In: Alimisis, D., Moro, M. and Menegatti, E. eds. *Educational Robotics in the Makers Era*. Edurobotics 2016. Advances in Intelligent Systems and Computing, Vol. 560. Cham, Springer.

Kappor, A. 2014. L'invasion robotique au Canada [The invasion of robotics in Canada]. *Canadian Urology Association Journal,* Vol. 8, pp. 5-6.

Kudina, O. and Bas, M. 2017. "The end of privacy as we know it": Reconsidering Public Space in the Age of Google Glass. In: Newell, B. C., Timan, T., Koops, B. J. eds. *Surveillance, Privacy, and Public Space*. United Kingdom, Taylor and Francis.

Leenes, R., Palmerini, E., Koops, B.-J., Bertolini, A., Salvini, P., Lucivero, F. 2017. Regulatory challenges of robotics: some guidelines for addressing legal and ethical issues, *Law, Innovation and Technology*, Vol. 9, No. 1, pp. 1-44.

Leopold, A. 1949. *A Sand County Almanac and Sketches Here and There*. New York, Ballantine Books.

Liberati, N. 2017. Teledildonics and New Ways of "Being in Touch": A Phenomenological Analysis of the Use of Haptic Devices for Intimate Relations, *Science and Engineering Ethics*, Vol. 23, No. 3, pp. 801-823.

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

Matsuzaki, H. and Lindemann, G. 2016. The autonomy-safety-paradox of service robotics in Europe and Japan: a comparative analysis, *AI & Society*, Vol. 31, No. 4, pp. 501-517.

Mavroidis, C. and Dubey, A. 2003. From pulses to motors, *Nature Materials*, Vol. 2, pp. 573-574.

Mavroidis, C., Dubey, A., Yarmush, M. L. 2004. Molecular machines, *Annual Review of Biomedical Engineering*, Vol. 6, pp. 10.1-10.33.

Miller, D. P., Nourbakhsh, I. R. and Sigwart, R. 2008. Robots for education. In: Siciliano, B. and Khatib, O. eds. *Springer handbook of* robotics. New York, Springer.

Mol, A. 1997. *Wat is kiezen? Een empirisch-filosofische verkenning*. Inaugural lecture, Enschede, Universiteit Twente. Available at: http://www.stichtingsocrates.nl/tekstenpdf/Wat%20is%20kiezen.pdf

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

Mordoch, E., Osterreicher, A., Guse, L., Roger, K. and Thompson, G. 2013. Use of social commitment robots in the care of elderly people with dementia: A literature review, *Maturitas*, Vol. 74, No. 1, pp. 14-20.

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

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

OECD Insights (n.a.) 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/

Oost, E. and Reed, D. 2010. Towards a Sociological Understanding of Robots as Companions. In: Lamers, M. H., Verbeek, F. J. eds. *Human-Robot Personal Relationships*. Heidelberg, Berlin, Springer.

Pearson, Y. and Borenstein, J. 2014. Creating ‘‘companions’’ for children: the ethics of designing esthetic features for robots, *AI & Society*, Vol. 29, pp. 23-31.

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

Pilkington, E. 2015. ‘Life as a drone operator: ‘Ever step on ants and never give it another thought?’’, *The Guardian*, 19 November. London, The Guardian. Available at: https://www.theguardian.com/world/2015/nov/18/life-as-a-drone-pilot-creech-air-force-base-nevada

Powers, T. M. 2011. Prospects for a Kantian machine. In: Anderson, M. and Anderson, S. L. eds. *Machine Ethics*. Cambridge, Cambridge University Press, pp. 464-475.

Rathenau Instituut. 2017. *Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality*. The Hague, Rathenau Instituut. Available at: https://www.rathenau.nl/en/publication/human-rights-robot-age-challenges-arising-use-robotics-artificial-intelligence-and

Riek, L. D. and Howard, D. 2014. A code of ethics for the human-robot interaction profession, *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.* New York, Putnam.

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

The Royal Society. 2017. *Machine learning: the power and promise of computers that learn by example*. London, The Royal Society. Available at: https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf

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

Rosenberger, R. and Verbeek, P. P. 2015. A Field Guide to Postphenomenology. In: Rosenberger, R., Verbeek, P. P. eds. *Postphenomenological Investigations: Essays on Human-Technology Relations*. London, Lexington Books, pp. 9-41.

Sandel, M. 2009. *The Case Against Perfection: Ethics in the Age of Genetic Engineering*. Cambridge, Harvard University Press.

Savulescu, J. and Maslen, H. 2015. Moral enhancement and artificial intelligence: moral AI? In: Romportl, J., Zackova, E. and Kelemen, J. eds. *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide*. Switzerland, Springer International Publishing, pp. 79-96.

Scassellati, B., Admoni, H., and Mataric, M. 2012. Robots for autism research, *Annual Review of Biomedical Engineering*, Vol. 14, pp. 275-294.

Sharkey, A. and Sharkey, N. 2012. Granny and the robot: Ethical issues in robot care for elderly, *Ethics and Information Technology*, Vol. 14, No. 1, pp. 27-40.

Simut, R. E., Vanderfaeillie, J., Peca, A., Van de Perre, G. and Vanderborght, B. 2016. Children with Autism Spectrum Disorders Make a Fruit Salad with Probo, the Social Robot: An Interaction Study, *Journal of Autism Development Disorder*, Vol. 46, pp. 113-126.

Smith, C., Villanueva, A., Priya, S. 2012. Aurelia aurita bio-inspired tilt sensor, *Smart Materials and Structures*, Vol. 21, No. 10.

Stone, W. L. 2005. The history of robotics. In: Kurfess, T. R. ed. *Robotics and Automation Handbook*. Boca Raton, CRC Press, pp. 1.

Swierstra, T., Stemerding, D. and Boenink, M. 2009. Exploring Techno-Moral Change: The Case of the ObesityPill. In: Sollie, P. and Düwell, M. eds. *Evaluating New Technologies*. Dordrecht, Springer, pp. 119-138.

Ticehurst, R. 1997. The Martens Clause and the Laws of Armed Conflict, *International Review of the Red Cross*, No. 317, pp. 125-134.

Toh, L. P. E., Causo, A., Tzuo, P. W., Chen, I. M. and Yeo, S. H. 2016. A Review on the Use of Robots in Education and Young Children, *Educational Technology & Society*, Vol. 19 No. 2, pp. 148-163.

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

Tzafestas, S. G. 2016a. *Roboethics. A Navigating Overview*. Switzerland, Springer.

Tzafestas, S. G. 2016b*. Sociorobot World: A Guided Tour for All*. London, Springer.

Ummat, A., Dubey, A. and Mavroidis, C. 2004. Bionanorobotics: a field inspired by nature. In: Bar-Cohen, Y. ed. *CRC Handbook on Biomimetics: Mimicking and Inspiration of Biology*. Boca Raton, CRC Press, pp. 201-226.

United Nations (UN). 1945. *Charter of the United Nations*. New York, UN. Available at: http://www.un.org/en/sections/un-charter/un-charter-full-text/

UN. 1948. *Universal Declaration of Human Rights*. New York, UN. Available at: http://www.un.org/en/universal-declaration-human-rights/index.html

UN. 2010. *Study on targeted killings, Addendum 6 of the Report of the Special Rapporteur on extrajudicial, summary and arbitrary executions, Philip Alston*. New York, UN. Available at: http://www2.ohchr.org/english/bodies/hrcouncil/docs/14session/A.HRC.14.24.Add6.pdf

UN. 2015a. *Addis Ababa Action Agenda of the Third International Conference on Financing for Development*. New York, UN. Available at: http://www.un.org/esa/ffd/wp-content/uploads/2015/08/AAAA_Outcome.pdf

UN. 2015b. *Transforming Our World: the 2030 Agenda for Sustainable Development*. Resolution A/RES/70/1. New York, UN. Available at: http://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E

United Nations Educational, Cultural and Scientific Organization (UNESCO). 2005a. *Convention on the Protection and Promotion of the Diversity of Cultural Expressions*. Paris, UNESCO.

UNESCO. 2005b. *The Precautionary Principle: Report of COMEST*. Paris, UNESCO. Available at: http://unesdoc.unesco.org/images/0013/001395/139578e.pdf

UNESCO. 2013. *Report to COMEST of Workshop on Ethics of Modern Robotics in Surveillance, Policing and Warfare (University of Birmingham, United Kingdom of Great Britain and Northern Ireland, 20-22 March 2013)*. Working Document. Paris, UNESCO. Available at: http://unesdoc.unesco.org/images/0022/002264/226478E.pdf

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

Van de Poel, I. 2013. Why New Technologies Should Be Conceived as Social Experiments, *Ethics, Policy & Environment*, Vol. 16, No. 3, pp. 352-55.

Van de Poel, I. 2016. An Ethical Framework for Evaluating Experimental Technology, *Science and Engineering Ethics*, Vol. 22, No. 3, pp. 667-686.

Verbeek, P. P. 2011. *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago, University of Chicago Press.

Van Rysewyk, S. P. and Pontier, M. 2015. Preface. In: Van Rysewyk, S. P. and Pontier, M. eds. *Machine Medical Ethics*. London, Springer.

Vandemeulebroucke, T., De Casterle, B. D. and Gastmans, C. 2017. How do older adults experience and perceive socially assistive robots in aged care: a systematic review of qualitative evidence, *Aging and Mental Health,* published online 9 February 2017, pp. 1-19.

Verbeek, P. P. 2013. Technology Design as Experimental Ethics. In: Van den Burg, S. and Swierstra, Tsj. eds. *Ethics on the Laboratory Floor*. Basingstoke, Palgrave Macmillan, pp. 83-100.

Veruggio, G. 2002. Views and visions in robotics. Hearing at the Italian Senate's 7th Permanent Commission (Rome).

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

Waelbers, K. and Swierstra, Tsj. 2014. The family of the future: how technologies can lead to moral change. In: Hoven, J. V. D., Doorn, N., Swierstra, T., Koops, B-J. and Romijn, H. eds. *Innovative Solutions for Global Issues*. Dordrecht, Springer, pp. 219-236.

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

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

Weir, N. A., Sierra, D. P., and Jones, J. F. 2005. *A Review of Research in the Field of Nanorobotics*. SAND2005-6808, Unlimited Release.

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

Wildmer, G., Oswald-Krapf, H., Sinha-Khetriwal, D., Schnellmann, M. and Boni, H. 2005. Global perspecitves on e-waste, *Environmental Impact Assessment Review,* Vol. 25, No. 5, pp. 436-458.

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

Wu, Y.-H., Faucounau, V., Boulay, M., Maestrutti, M. and Rigaud, A. S. 2010. Robotic agents for supporting community-dwelling elderly people with memory complaints: Perceived needs and preferences, *Health Informatics Journal*, Vol. 17, No. 1, pp. 33-40.

Zhang, M., Davis, C. T. and Xie, S. 2013. Effectiveness of robot-assisted therapy on ankle rehabilitation – a systematic review, *Journal of Neuroengineering and Rehabilitation*, Vol. 10, No. 30, pp. 1-16.