ROBOTS AND INTELLIGENT/AUTONOMOUS SYSTEMS: TECHNOLOGY, SOCIAL IMPACT AND OPEN ISSUES

Daniele Andresciani and Roberto Cingolani

Istituto Italiano di Tecnologia – Genova Italy

Contents

Part 1. Robotics: its development and technological implementation

The ancient Greeks used the term *Banausia* (from *banausos, "craftsman," "manual, mechanical work"*) to refer to manual work and mechanical arts in general, and it had a negative meaning: craftsmen, or whoever performed a manual work, were considered inferior to those engaged in an intellectual work.

Many centuries later, between the 1400s and the early 1700s, the European culture reassessed manual techniques. Some of the procedures used by technicians and craftsmen *to modify* nature turned out to be useful and beneficial to understand the natural environment. The defence of mechanical arts from the accusations of unworthiness and the refusal to make practical activities coincide with the concept of slavery led to a historical cultural turning point: the end of an elitist image of science and of the distinction between knowledge and skills. In the reassessment of science and mechanical arts, a major and original role was played by Francis Bacon. In fact, he wrote an important and lucid critical treatise on the experimental method and on the good or bad use of science and technology. In Novum Organon, he talked about the condition preliminary to all scientific works: that is, the removal of ***idóla,*** namely advance information or prejudices that pollute scientists' mind and their objectivity. Said ***idóla*** were divided into *tribus* (typical of everybody), *specus* (typical of the single individual), *fori* (related to controversies and verbal disputes) and *theatri* (due to philosophical, religious, cultural dogmatisms). In the same treatise, Bacon stigmatised the existence of two opposite anti-scientific behaviours which he described as being similar to the typical actions of *spiders* and *ants*: rationalist dogmatists, lacking contact with reality, are like spiders, that spin webs from themselves; empiricists, lacking theoretical foundations, are like ants, that simply accumulate and use rashly. True scientists combine both theory and experimentation, like bees, that take material from flowers but then have the ability to convert and digest it.

Finally, in De Sapientia Veterum, Bacon brilliantly used the myth of Daedalus to talk about the **constituent ambiguity** of technology. Daedalus built a device to enable Pasiphae to mate with a bull; this pernicious use of technology gave birth to the Minotaur, devourer of men. At that point, Daedalus made a good use of his intelligence and built a labyrinth in which to confine the Minotaur. The labyrinth was also provided with a safety system, Ariadne's thread, that allowed Theseus to find his way out. The metaphor is clear: science and technology can be used against or in favour of mankind; therefore, scientists must be responsible and forecast remedies and limitations of the possible negative outcomes of their discoveries. Despite Bacon's ideas are more than 400 years *old*, they are extraordinarily topical; in particular, the intuition concerning the ambiguity of technological progress is perfectly apt with the issues concerning the development of robotics and artificial intelligence (AI).

1.1 The different types of robots

It is not easy to define what a 'robot' is, considering the rapid and continuous development of robotics. The word ‘robot’ was introduced in 1920 by Karel Capek, a writer who went beyond the concept of ‘automaton.’ In fact, he introduced the idea of an artificial machine built by humans to perform precise functions related especially to work (in Czech *robota* means forced labour).

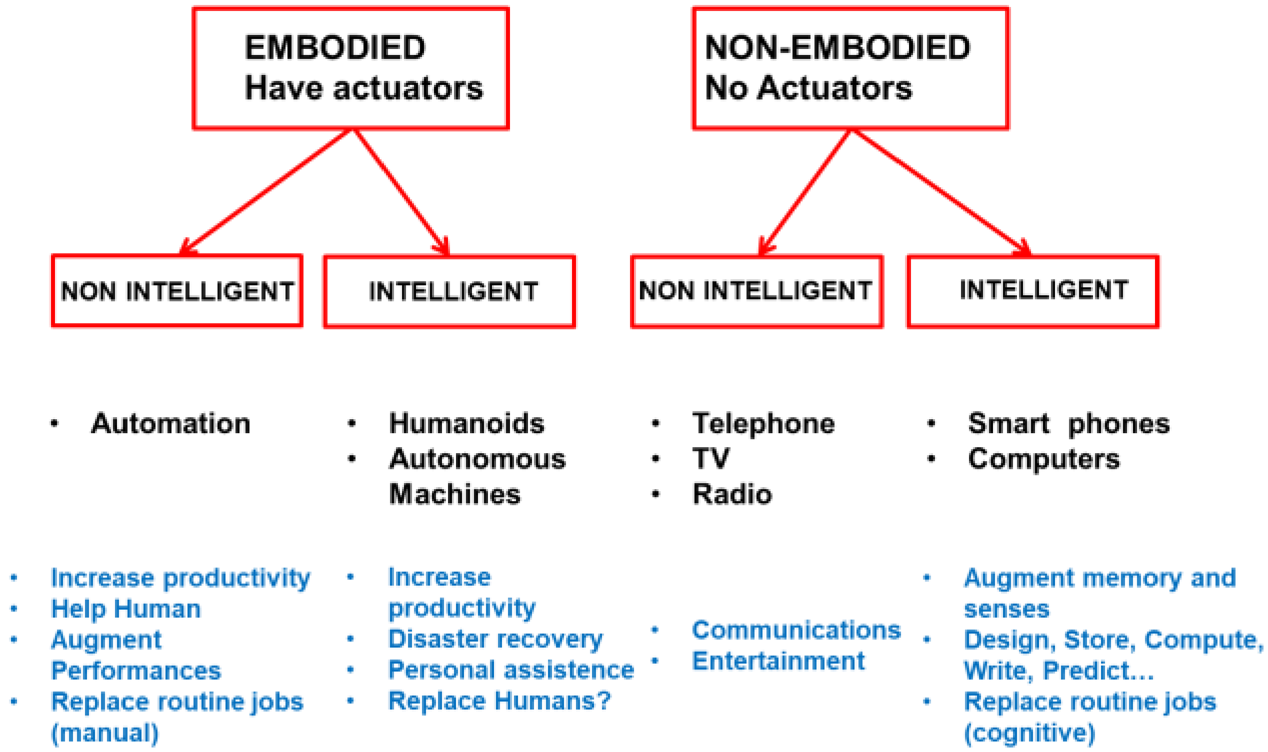
Over the last sixty years, robotics has progressed extraordinarily. Initially, its products consisted of mechanical, static, passive, repetitive and executive objects; today, robots are becoming autonomous and mobile realities capable of performing not only specific functions, but also general ones. They can be provided with learning and adaptation skills, and act autonomously, without the control of an operator.

The most advanced robots have cognitive abilities similar to those of primates. They are able to communicate through the recognition of words, and can have expressions in their outward appearance that imitate several human emotions.

Currently, there is much debate on the possibility to realise robots provided with an advanced artificial intelligence (AI) such to be able to develop decisional abilities and self-determining processes similar to those of humans. In actual fact, in the collective imagination and in the representation given by the mass media, literature, movies and TV series, robots are increasingly viewed as entities provided with a mechanical body that thinks and behaves like humans. In real life, it is not that simple. Indeed, it is necessary to better explain these concepts.

To begin with, it is important to highlight a simple and univocal classification of autonomous and intelligent machines with their different characteristics. Often robots, humanoids and artificial intelligence are considered all alike, but that is not so. First of all, it is necessary to make a distinction between two large types of machines: those provided with a body (embodied) and those without a body (non-embodied). Secondly, it is necessary to verify if machines, both the embodied and the non-embodied, are provided with some form of artificial intelligence (that is, if they are *stupid* or *intelligent*).

Embodied machines (provided with a body) are capable of moving and performing physical work. These machines are well-known: scrapers, automation systems and all the technologies that replace humans in physical work, or help humans increase their performances (for example strength, precision, speed of execution, etc).



Usually these machines are "*stupid*": they are programmed to work automatically, to perform demanding or repetitive activities with the aim to increase productivity and the performances of the operators using them. Since they are controlled by humans (or programmed by humans), they do not take decisions autonomously: their actions depend on programmes or on human operators. The impact of this robots on workforce, and particularly replacing humans in routine manual works (eg manufacturing) is a very debated issue.

In recent years, some embodied machines have been provided with artificial intelligence acquiring increasing cognitive and decisional abilities. They can be non-anthropomorphic machines (for example self-driving cars), or actual humanoids developed to interact with humans and support them in various environments, such as at work, at home or in hospitals.

The ability to make autonomous decisions, without the control of an operator, is a great technological challenge, which anyway gives rise to many questions from an ethical and regulatory viewpoint. In fact, although this type of intelligent robots is designed to replace humans in dangerous situations, or help them in case of need, it is important not to underestimate the problem of their impact on workforce, not only on routine works. It is also important not to neglect the issue related to the future cohabitation of the two species: humans and robots, both "thinking" but with totally different logics and functioning.

Non-embodied technological products are not able to perform work or to make movements, and belong to technologies commonly called digital, ranging from telecommunications to artificial intelligence. Also for this type of products a distinction can be made similar to the previous one. Some non-embodied machines are "*stupid,*" such as TVs and radios; they have become objects of daily use for years now and humans are almost addicted to them. They process and transmit information, both audio and visual, and have opened the communication sector to modern society. Non-embodied machines have gradually become increasingly "intelligent": from smartphones to supercomputers, they can perform calculations at an extremely high speed, from several million operations per second in the case of smartphones to quadrillions of operations per second in the case of supercomputers. This has taken place in the last years thanks to an increasing miniaturisation of integrated circuits, allowing electronic devices to perform an increasing number of operations per second, the electric power consumed being equal, and to memorise an increasing amount of data in mass memories.

Over the last 50 years, the progress of electronic technologies has followed Moore's Law, on the basis of which (about) every 2 years the number of transistors on an integrated circuit doubles, passing from several thousand transistors in 1970 to about 20 billion transistors in 2016.[1] At the same time, the progress of manufacture has allowed to reduce the electric power consumed by transistors proportionally to the reduction of their size (according to the so-called Dennard Scaling rules). This has led to a constant and uninterrupted development of computational abilities in more recent years. The increase in the speed of calculation and in the ability to store data has allowed to realise electronic devices more and more sophisticated, determining the current digital revolution. Internet, research engines, mobile phones, social networks, Big Data, Industry 4.0, digital wholeness, forecasting models in the financial, social, medical, climatic fields are all direct or indirect consequences of the technological evolution which has characterised transistors.

Today computers can process an enormous amount of data, analysing them statistically very quickly and applying mathematical models that allow to forecast future situations and scenarios (in the economic, medical and climatic fields). At the same time, they can also imitate the cognitive processes of the human brain, and create what is commonly called "artificial intelligence." This is what gave origin to the research engines we all use, as well as to the artificial intelligence capable of beating humans at chess or outdoing them in other activities strongly computational.

[1] M.Roser, H.Ritchie, Technological Progress, Empirical view, https://ourworldindata.org/technological-progress

However, how can the computational power of a computer be compared with that of the human brain? The performance of computers is measured in FLOPs (Floating point operations per second), that is the amount of operations that a computer can perform in one second. Today's most powerful supercomputers are able to perform dozens of PetaFLOPs, that is dozens of quadrillions of operations per second.

In 2017 the top 5 computers in the world as to power of calculation were developed in China (Subway TaihuLight with 93 PetaFLOPs, and Tianhe-2 with 33.9 PetaFLOPs), Switzerland (Piz Daint with 19.6 PetaFLOPs), Japan (Gyoukou with 19.1 PetaFLOPs) and in the USA (Titan with 17.6 PetaFLOP). Their consumption of electric power is tremendous, and ranges from China's TaihuLight with 15.4 MegaWatt to the USA's Titan with 8.2 MegaWatt. It is interesting to notice that regardless of the records of the single machines, the USA is the country that holds the overall highest power of calculation in the world: currently, 46% of the global power of calculation is American, owing to a wide network of supercomputers disseminated on the national territory. Following there are China and Japan with 8%, and Germany with 7%.

Such a high power of calculation allows to process quadrillions of instructions per second, which is close to the calculation ability necessary to simulate with precision complex biological organisms. We have to remember that it is not possible to measure the power of calculation of the human brain in PetaFlops, essentially because the functioning of the brain is not based on digital electronic operations. However, the instrument of comparison that can be used is an empirical unit of measure called MIPS (Million Instructions per Second), that is the number of instructions per second that can be processed by a processor, be it biological or artificial. A capacity of 1,000 MIPS is sufficient to reproduce the complete functioning of a complex organism such as a lizard, while 1 billion MIPS is the minimum amount necessary to simulate humans. The development of increasingly powerful computers and software is reaching millions of MIPS; we are getting closer and closer to biological performances, but with an incomparably higher energy consumption.

The availability of increasingly powerful calculation machines is constantly extending the limitations of artificial intelligence. At the same time, it is allowing the development of increasingly performing embodied machines (provided with sight, touch and biomechanical abilities), making realistic the assumption that robots are characterised by performances increasingly closer to those of humans.

It is interesting to notice that until the *physical and intellectual* performances of machines developed separately, the latter did not constitute a source of apprehension for us. We never feared that machines such as computers could be faster than us in performing calculations, or that robots could be stronger, more rapid and more precise than us in carrying out physical work. On the contrary, a large part of our industrial progress has been based on the use of machines that, subdued to our purposes, have increased human performances in specific fields.

The fact that the physical and intellectual "powers" of machines have always been separate has made us feel safe: computers "*think faster than us, but they cannot move,"* robots "*are stronger than us, but they cannot think;*" for a long time, these paradigms have safeguarded the supremacy of humans over their technological creatures. Robotics and artificial intelligence are two worlds created from different realities and technologies (mechatronics and computer science, respectively). Over the years, they have produced "incomplete" technologies compared to humans, in the sense that their increasingly effective emulation concerned only a part of our potentialities: the physical ones (strength, duration, precision) or the cerebral ones (calculation, memory, logical process), and have become of common use without too many problems.

No one felt threatened by a computer capable of winning at chess, because said computer did not have a body and was not able to do anything else; no one feared a machine capable of raising tons with disarming easiness and precision, because it did not have cognitive abilities. The mutual contamination of the two technologies gave origin to the *other species*: intelligent machines capable of moving, acting and taking decisions autonomously, making us feel threatened as dominant species. In the collective imagination, if strong robots are also capable of thinking and computers that beat us at chess are also capable of running, humans are in danger because they can lose control over their artificial creatures.

Intelligent and autonomous machines (A/IS Autonomous Intelligent Systems) [2] represent a true technical, scientific and cultural revolution, and are perhaps the greatest consequence of nanotechnologies; they are seriously starting to measure up with our culture, customs and society, while raising doubts, anxieties and fears. We need to ask ourselves a question, though: are these fears grounded? According to our opinion, similarly to all new and unusual realities, this revolution-evolution is not to be feared, rather it is to be studied and understood.

1.2 Robots and artificial intelligence: possibilities and limitations of the integration between body and brain

[2] Today, intelligent autonomous systems are used more than robots, broadly speaking; we will use both terms in an almost equivalent manner.

Throughout the history of technology, researchers and scientists have worked hard with the aim to create robots increasingly similar to humans and an artificial intelligence increasingly similar to the human one. The expression "artificial intelligence" was used for the first time in 1956 by American mathematician John McCarthy; since then, as technical and scientific results progressed, scientists and philosophers have started to reflect passionately and critically on how much it is possible to talk about "intelligence" when referring to machines, and how much analogy there can be between machines and humans. The first essays on this theme started to be published in the 1960s, identifying with far-sightedness many obstacles that research would have then met in the following years, especially due to the strong difference between the cognitive results achieved in machines, and those typical of humans. In 1969, Marvin Minsky and Seymour Papert, in their volume *Perceptrons*, highlighted the limitations of the first artificial neural networks they had realised.[3] Given the authors' authoritativeness, the publication of *Perceptrons* stifled the enthusiasm that had risen on the topic, leading to a consistent lowering of the scientific and economic interest towards AI. Such situation lasted until the early 1990s.

In the 1990s, owing to the acceleration of electronic technologies and an increased performance of computers, researches in the field of artificial intelligence resumed with great determination: studies concentrated on *intelligent agents* as autonomous entities, intelligent software, and intelligent agents embodied in a physical system.[4] Encouraged by the results achieved with the new generation computers, scientists were fuelled by a renewed hope to be able to build intelligent robots, that is to insert artificial intelligence in a synthetic body to imitate humans.

The progress made in the field of sensors allowed to build very advanced vision and sensation systems; tactile sensors and hearing systems were realised and integrated in robots with increasingly improved biomechanical abilities. Control algorithms evolved quickly. Electronics and computers allowed to store and use the stimuli coming from the sensors enabling robots to move autonomously and even to start taking decisions, owing to the first embodied artificial intelligence. Many of those systems ended up being useful to solve important problems for mankind and the society,[5] but they also brought back to the attention an old issue: can humans become obsolete?

[3] M. Minsky, S. Papert, *Perceptrons: An Introduction to Computational Geometry*, MIT Press, 1969.

[4] In 1997, IBM's Deep Blue computer, capable of processing eleven billion operations per second, performed a historic feat beating at chess world champion Garry Kasparof. At the same time, it was evident that Deep Blue was not able to think. This unhappy statement of fact mitigated enthusiasm in the research of artificial intelligence.

[5] P.McCorduck, *Machines Who Think*, A.K.Peters Ltd, Natick 2004; S.Russell, S.Norvig, *Artificial Intelligence: a modern approach*, Pearson Education Ltd, New York 2016; P.Husbands, *Robotics*, in K.Frankish, W.M.Ramsey (eds), *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press, Cambridge 2014.

Indeed, Minsky wrote: "*Will robots inherit the earth? Yes, they will. But those robots will be our children!*".[6]

As mentioned, body and brain were studied and imitated separately for decades, giving origin to very powerful and precise machines, or "electronic brains" with extraordinary computational abilities; with the advent of intelligent machines (A/IS), mechatronic systems with sensorial abilities and biomechanical performances similar to those of the human body started to be connected with computers having computational powers similar or higher to those of the human brain. The integration of these two technological worlds in intelligent robots provided with artificial intelligence, constituted a very important step in the history of science and technology. It also characterised the beginning not only of a technical revolution, but also of an anthropological, legal and ethical one. However, an objective analysis of the potentialities and limitations of A/IS highlighted a fundamental limitation that had to be overcome: the extremely difficult reproduction of the inseparable relation between body and brain, typical of humans and of the more evolved living beings.

The distinction between body and brain can initially be valid for the couple robot-artificial intelligence, but certainly it is not applicable to living beings, and in particular to humans: humans' coordination between body and brain is totally different from the humanoids' coordination between body and artificial intelligence. In humans, body and brain are deeply interconnected and synergic; none of the two plays a dominant role: what matters is coordination, the harmonic and perfect synergy of their functions. The human muscle-skeletal structure evolved at the same pace of the cognitive system, with a process of mutual adaptation, mediated by life's biochemistry (hormones, metabolism, etc). The emotional state affects the body's physical response: anger increases our strength, fear makes us attentive and concentrated, and tenderness makes us particularly sensitive. The responsive fibres that constitute the muscles of the human body contract and relax owing to aware and unaware nervous stimuli, correlated to states of mind, decisions related to our movements or our needs. Our body is characterised by synergies that have developed over billions of years, exploiting biological mechanisms currently not reproducible in robots.

A long work devoted to integration has been carried out, but there is still much to do in order to make motorised actuators and computers both synergic and compatible, the former realising movements, and the latter commanding said movements. Intelligent machines are provided with electronic cards that calculate complex algorithms which generate digital signals; in turn, said

[6] Minsky, *Ibidem*.

signals control switches and amplifiers that put electric power in the robots' motors, through a process which is still burdensome and rudimental compared to the biological one. Humans, instead, do not have electric power and electron flow, but nervous impulses: packages of ions that move in water, the element that constitutes about 60% of an adult human body. Robots follow the laws of electricity, the human body those of biochemistry.

1.3 Further elements of comparison between humans and robots

After this brief description of the historic technological leap that gave origin to autonomous and "thinking" machines, we now want to provide some important data that will allow to make a comparison between the characteristics and performances of machines and those of humans. First of all, we have to make an ecological-constituent comparison: 99% of a human body - as all organic, biological and natural entities - is composed of 6 atoms: oxygen (65%), carbon (18%), hydrogen (10%), nitrogen (3%), calcium (2%) and phosphorus (1%). It is designed to grow and, at the end of its lifecycle, to dissociate itself from those 6 atoms. Artificial systems, instead, need from 30 to 50 atoms, they are designed to be assembled in the shortest time possible, and at the end of their lifecycle someone will have to disassemble them to recover the materials. A car is assembled in 4 hours, disassembled in 40 hours, and many of its parts deteriorate.

With regard to the relationship brain-body, humans have a system that has optimised itself in 3 billion years of evolution: a very long timeframe in which humans have developed an extraordinary ability to adapt and learn. The gap that technology has to fill in order to be able to compete with the results of human evolution is still enormous. Technology can only vaguely imitate the humans' system of stability and dynamic equilibrium: robots can be provided with equilibrium by using gyroscopes (such as those of mobile phones or airplanes), but the results are not at all comparable with the performances of the vestibular system of an athlete or an acrobat.

Furthermore, with regard to the relationship body-brain, it is difficult for computers connected to movement actuators and sensors to equal what takes place in humans. Owing to its long and complex evolution, the human brain works in synergy with the body: the same group of neurons that controls sight also supervises manipulation; the group that controls the tongue supervises the understanding of speech, and so on. Currently, it is impossible to transfer to machines the mind-implementation synergies typical of humans, because electronic intelligence and mechatronic bodies work with mechanisms that are different from the biological ones.

Besides, it is difficult to think of intelligent autonomous machines (A/IS) as being capable of communicating at a high level with humans using the body's semantic, interpreting correctly simple human actions and understanding humans' intentions. Non verbal language plays an important role in human communication: a wink, an annoyed expression can communicate more than a thousand words; it is an extremely intuitive and fast communication that we use continuously, but very difficult to teach with regard to machines.

Lastly, it is important to highlight a core aspect in the comparison between humans and A/IS: the latter need an enormously greater amount of energy (even million times greater) to process instructions (MIPS) (or equivalent PetaFLOP operations) necessary for a complex biological entity such as the human one to function. As mentioned, the computational ability of some machines can even reach hundreds of PetaFLOPs; however, in practice, they need an enormous electric power (dozens of Millions of Watts), they are as big as a room and use enormous cooling systems and an independent power station. The human brain is a "*ball*" that weighs about 1500 grams (slightly less than 3.5 pounds), fuelled by the metabolism of sugars with few dozens of Watts! Therefore, with the current technological standards, it is impossible to assume a self-moving system capable of thinking like humans, with the same mental and biomechanical abilities.

The Big Data Companies that manage big computers and large storage structures could offer a partial solution to the apparently unsolvable issue related to "intellectual abilities" and machines' high energy expenditure. With all likelihood, robots will remain "stupid" and with a limited individual computational ability (about a billion operations per second, comparable to those of a good computer that consumes a couple hundred Watts). This limited "intelligence" can be used to produce good movements, but the "cognitive" aspect has to be managed in a different way. The possibility would be to create a single brain to which all robots are connected, a sort of global repository of machines' intelligence, which could use the cloud to store all the information collected and the "things learned" from robots; each machine could upload its experiences and download those of others.

It is a fascinating scenario, in which humans' individual memory and intelligence oppose a single and shared intelligence for robots. These machines could act like a swarm, being provided with a single intelligence towards which all its individuals contribute: a species that has no equivalent in the biological world, and with which we might have to learn to cohabit some day.

It is difficult to say how much such scenario is realistic. There are still enormous technological difficulties to face: the extremely fast wireless technologies - necessary to dialogue in real time with the cloud, which hosts the single large intellect from which all robots draw on (remember the acting intellect of Avicenna) - are not available everywhere and require a very complex network infrastructure (for example a widespread 5G network). It is likely that companies that do not build

robots may be more enthusiast and interested in these technologies than those that do build them: the latter know that single robots will always be inferior to humans, while the former deem that a global artificial intelligence can be very performing.
Therefore, we can conclude this brief paragraph highlighting that, even with all the mentioned limitations, there is a possible regulatory and ethical problem: the moment in which there will be sufficiently sophisticated autonomous and intelligent systems (A/IS) and sufficiently powerful computers, what rules will govern the world's global intelligence, the global repository of intelligence that the cloud will become? And who will manage it? States or large companies?
And if it will ever be possible to create an A/IS with an extremely high intelligence not shared in the cloud, but residing individually in each machine (a very unlikely scenario at the moment), how will those entities be treated since at all effects they are not biological, but capable of understanding and with a will?
Intelligent machines may become capable of understanding and with a will, but for sure they will not share our biology. Will it be simpler to face the issue thinking that we are dealing with an alien race? Or should we try to adapt the rules that we have developed for ourselves? These are open issues on which we have the duty to reflect. Without radical pessimism or excessive optimism, but by interweaving history, philosophy and human sciences with the new horizons of technology.

Part 2. The coexistence of humans and intelligent machines

Even if the day when robots will be intelligent like humans seems distant (and according to our opinion, it may never arrive, at least until we use silicon), it is worth to consider the ethical, social and human issues that would arise if their presence in the society were widespread. We want to give some answers or at least try to provide some guidelines to reflect on the matter. In particular, these considerations are addressed to all those who are involved in the design, realisation and use of these new revolutionary technological products. The production of autonomous and intelligent machines (A/IS) capable of acting always in favour of humans and of the community is a common aim of experts in many and different disciplines: electronics and mechanical engineers, computer scientists, psychologists, neurologists, cognitive scientists, experts in artificial intelligence, logicians, mathematicians, philosophers, jurists, economists, designers and artists.
Going back to the basic theme of our debate, and in particular to the distinction between "stupid" and "intelligent" machines, it is important to highlight that the former are often extremely sophisticated human artefacts that do not raise particular or new problems from an ethical

viewpoint. In fact, the criteria and standards adopted are those already in use for technological products. Whereas, the situation is different when referring to "intelligent" machines. When robots develop the ability to decide and operate autonomously, to learn, to acquire experience with algorithmic decisional processes (although in the absence of emotiveness and spirituality), the ethical issues are totally new and relevant.

A cognitive machine capable of learning raises the problem of how to educate it: which "educational" strategies need to be put into action; how should it be supported; and how should it be punished if it violates rules. The fundamental question becomes "*How to punish a robot that makes a mistake?".* [7] Humans violate rules out of need, by mistake, or due to revenge or malice. In most cases, the motivation that leads to violating rules lies in psychological and existential alterations or in conditions of particular need or suffering. The re-educational path passes through a punishment that usually involves the reduction of a liberty or is expressed in the payment of a sanction. More in general, any human educational path is based on balancing rewards and punishments. Punishments are based on the fact that any human being, and even the more intelligent animals, fear to be deprived of something they consider important: be it freedom in case of imprisonment or a toy if punishing a child for an escapade. The fear of losing something important is part of human psychology, but it is also a consequence of the principle of self-preservation typical of all human beings. The fear of being punished makes us behave better, in order to avoid a worsening of the quality of life.

In machines, the violation of a rule can be triggered simply by the evaluation that said transgression is necessary in order to reduce negative collateral effects of a specific action. The violation is simply the result of an algorithm that, minimising risks and collateral damage, aims to achieve a goal with side conditions. The most classic case is that of a self-driving car that with malfunctioning brakes has to choose if to crash into a wall and sacrifice the passenger or if to continue along the established direction and sacrifice pedestrians. In a situation where the violation is equal, the machine will necessarily choose the solution that is likely to cause the minor damage, whatever it is. Different machines in the same situation and with the same side conditions will all reach the same conclusion. Humans, instead, in the same situation and conditions, do not all act in the same way. The decision is based on imponderable subjective elements that inevitably lead to different personal decisions (for example, if a relative is among the pedestrians, the final decision may be different, regardless of the overall risk evaluation).

The diversity in human behaviours derives from the non-algorithmic nature of our intelligence,

[7] W.Wallach C.Allen, Moral Machines, Teaching Robots Right from Wrong, OUP 2010

from the ever present component of irrationality, emotiveness and imponderability inherent our logical mechanisms. Human irrationality or non-rationality, resulting from the hormonal component of our species, generates creativity, imagination, feelings that in turn lead to varied and unexpected behaviours.

It would be very difficult to punish a machine that violates a rule. It would not be possible to deprive it of food and liberty, since in both cases such elements are not fundamental for its species. Nor would it be possible to deprive it of life, since taking its batteries out equals to nothing more than what we do when we turn a computer off. Machines may be able to understand and decide, but that is not enough to have a conscience and a consequent instinct of preservation, survival and preservation of the species. Whatever the code which may be drawn up to regulate the coexistence between humans and intelligent machines, it is necessary to keep into account the fact that machines follow laws that are different from biological ones.

Of course, this all depends on how fast the identity and personality of autonomous and intelligent systems will evolve (A/IS). In the event, although remote, that they actually become a species in themselves, it will be necessary to lay down a specific legal and ethical framework capable of inserting them in the society and of establishing their rights and duties. It will not necessarily have to be a subordinate relationship as to man - as in the case of animals under their owners' responsibility or of an artefact under the craftsmen's responsibility - but something new and deeply different from what we have formulated up to now.

The current debate on these issues is in increase, both in Europe and outside of our continent, in countries technologically more evolved. Each new robotic technological product raises issues that spur ethical and political reflections, for the protection of the common good and of the community. A sustainable progress of A/IS, the latter's impact on single individuals and on the different segments of the society, the dangers that can arise and the damage that can be caused are topics that require new cultural and ethical instruments, as well as new international regulations that have to be developed ad hoc. Indeed, something is moving also at political and institutional level.

In 2016, two important documents on robotics were drawn up: in October, the White House's Office For Science and Technology published *Artificial Intelligence, Automation, and the Economy*, while in May the European Parliament's Committee on Legal Affairs published the *Report with recommendations to the Commission on Civil Law Rules on Robotics*.

The White House's document[8] is centred on artificial intelligence: the core is not robotics, but Artificial Intelligence (AI) and, in particular, how to organise a Good AI Society: a good society in which artificial intelligence plays a dominant role.

Very likely influenced by the representatives of the Silicon Valley, the document is strongly optimistic: AI will help improve anything, and dangers will be restricted to cyber war and autonomous weapons; the ethical reflection will be restricted to the effort to make transparent all which concerns machines and relevant research.

Instead, the document drawn up by the legislative office of the European Parliament[9] is mainly centred on robotics: on the Good Robotics Society, and not on the Good AI Society.

In order to realise the Good Robotics Society it will be necessary to evaluate how many job positions would be lost by introducing a massive amount of robots in the society. Moreover, it will be necessary to introduce soft and hard laws regulating possible crimes in this field and their seriousness; to this regard, it will be necessary to create a legal agency for robotics and artificial intelligence: a legal framework.

The two approaches are antithetical, reflecting the Americans' "business-oriented" approach and the Europeans' "regulation-oriented" approach, and separating AI from robotics.

Lastly, it is important to remember that always in 2016, the study group of the *Convention of the Society for the Study of AI and Simulation of Behaviour* of the United Kingdom formulated five rules for managing intelligent machines, stating at least in part the principle according to which artificial intelligence and robotics go hand in hand.[10]

First rule: robots are multiuse instruments; they must not be designed with the sole or primary aim to kill or harm humans, unless national security is involved.

Second rule: humans are the responsible agents. Robots must be designed and realised in such a way to comply with the laws in force and with citizens' fundamental rights, liberty and privacy.

Third rule: robots are technological products. They must be designed using processes that guarantee users' safety.

Fourth rule: robots are artefacts, objects resulting from manufacture. They must not be realised in such a way to be able to deceive subjects who are psychologically weak or vulnerable. Their nature as machines must be transparent and clear.

[8] https://obamawhitehouse.archives.gov/blog/2016/12/20/artificial-intelligence-automation-and-economy

[9] http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+REPORT+A8-2017-0005+0+DOC+XML+V0//EN

[10] http://www.sheffieldrobotics.ac.uk/aisb-workshop-por/

Fifth rule: all legal liability as to machines must be assigned to a person.
This approach - more effective than the broad dissertations presented by the USA and Europe - is simple and pragmatic, as it provides definitions and defines criteria. However, the creation of a regulatory structure remains distant.

2.1. Roboethics

The mentioned documents drawn up in 2016 were formulated on the basis of a reflection started about fifteen years before at international level by the entire scientific committee. In the third part of this document, we will analyse more in depth the current ethical proposals for A/IS; here, we simply want to mention the origin and development of Roboethics around year 2000.
As highlighted, A/IS can be controlled remotely by humans: such machines are used in various types of jobs and human activities, and are an important aid for the safety, efficiency and productivity of certain processes. They necessarily require the presence of humans, who control and govern them and take decisions for them. Other types of more evolved A/IS can be programmed to perform specific tasks in complete autonomy. They are provided with AI, they are able to “think” on their own, and they do not need any input from humans. Scientists and scholars worldwide immediately realised that such type of machine raises safety, ethical and legal issues. Therefore, from the very outset they began a passionate cultural and anthropological debate which is still in full development (as confirmed by this conference).
All professionals of the sector have warned that the perspective to use robotic systems and AI as humans’ assistants in performing various tasks in the social field (industrial work, domestic work, the selection of information, problem solving) and in the medical field will raise new human and ethical issues that will require new paradigms, suitable to face new forms of interaction between humans and machines.
In the 2000s, on the basis of this felt and shared need, a new field of study was developed, Roboethics, addressed to scientists, philosophers, jurists, sociologists and anthropologists with the aim to involve them in an objective and shared debate and to lay down the ethical basis for designing, realising and using robots.[11]
Recently, Spyros G. Tzafestas suggested a visionary definition according to which roboethics is a “branch of applied ethics, that is a philosophical, systematic and informed reflection, which

[11] G. Verruggio “The Birth of Roboethics” ICRA 2005, IEEE International Conference on Robotics and Automation Workshop on Robo-Ethics, Barcelona, April 18, 2005. The birth of Roboethics can be identified with the “First International Symposium on Roboethics - The ethics, social, humanitarian and ecological aspects of Robotics” held in Sanremo in 2004, http://www.roboethics.org/sanremo2004/

studies both the positive and negative consequences of robots in the society with the aim to spur the moral design, development and use of robots, in particular of "intelligent" and "autonomous" robots.[12]

In the third part, we will see how the lively and interesting international scientific debate on the ethical use of A/IS, started by Roboethics in the 2000s, is still ongoing. Today the term Roboethics has been in part overcome and replaced by the general term "Ethics of A/IS." However, the questions raised from the very outset by Roboethics remain topical: can robots perform good and bad actions? Can robots be dangerous for mankind?[13] Also the ethical values defined in the "Roboethics Road Map" remain topical, such as, for example:[14]

- The respect of values related to human dignity and human rights
- The promotion of equity, justice and equality in accessing new technologies
- The correct evaluation of damage and benefits
- The protection of cultural diversity and legitimate pluralism
- The avoidance of discrimination and stigmatisation
- The support of solidarity and cooperation
- The respect of privacy and the need of an informed consent
- The assuming of personal responsibilities for the Biosphere

2.2. Innovation: public awareness and 'metabolization'

Insofar as machines allow humans to avoid humble, tiring or dangerous tasks, the replacement of the human activity with the robotic activity is desirable, and even praiseworthy from an ethical viewpoint. Moreover, the use of machines with an increased ability to make up for human deficits is to be considered highly positive, as in the case of assisting vulnerable people.

However, paying attention to the labour market and therefore from an employment viewpoint, the robotic revolution will raise the problem of having to manage the progressive replacement of humans in various sectors with intelligent machines that work more efficiently, faster, in a more precise and economical manner.

Such phenomenon is not new in modern history. All new technologies have improved the efficiency of production processes. Besides, they have often reduced the humans' role, replacing them in fields in which they used to be protagonists.

---

[12] Spyros G. Tzafestas, *Roboethics. A navigating overview*, Springer 2016.

[13] G. Verruggio, quote.

[14] G. Verruggio EURON Roboethics Roadmap, July 2006, http://www.roboethics.org/atelier2006/docs/ROBOETHICS%20ROADMAP%20Rel2.1.1.pdf

This took place, for example, with the combustion engine that mechanised farming, transport and manufacture, or with printing machines that replaced amanuenses in writing. However, there is a difference between the robotic revolution and the previous technological turning points.
In the past, the evolutionary processes of machines required decades to take place: recent technologies, such as the automobile or the telephone, took various decades to change the organisation of the transport or telecommunication sectors. The slowness of the innovative process made it so that in most cases workers had time to reconvert to some other activity similar to what they had carried out up to that moment. The rhythm of progress and of the development of new technologies was slow, *inter-generational* and allowed workers, the production system and the society to adapt. The paradigm of farmers replaced by the threshing machine and reconverted to warehousemen in the same farm effectively describes the process that led to changes in work due to the introduction of new technologies. At the same time, it highlights a good system to cushion the social discomfort created.
Since every new technology generates new and unexpected jobs, the balance between jobs lost, jobs reconverted and new jobs acquired in the past has always been invariably positive in the medium-long term. However, in recent years, following the very fast evolution of the mentioned technologies aimed at electronic integration, the rhythm of the new technologies has undergone a striking acceleration. In a little less than ten years, commerce, the tertiary, industrial manufacture and telecommunications have been radically changed by the advent of digital technologies characterised by an *intra-generational* development.
We are witnessing a paradox according to which workers that ended their studies around twenty years old experience a change of jobs more than once (or they even disappear) during their working lives. The advent of new technologies that did not even exist when said workers used to study irrupts in their work environment disrupting it, generating in an extremely short timeframe further and multiple unsettling developments. In such scenario, against the jobs lost, new ones do not consolidate quickly, and the balance in the short and medium term risks to be invariably negative.
As much as the extremely fast progress of intelligent machines represents a great opportunity for global development, on the other hand it raises a problem of "*metabolization*" of the innovation for a society that is tarred for a much slower development. It is not simple to reconvert workers today, as they are asked a greater increase of their competences compared to the past (upskill of knowledge) which is impossible to guarantee, unless there is a system that provides constant refresher courses.

Society is like an interconnected neural system: such a metaphor helps us understand that in order for development to be sustainable, it must give citizens the time to metabolize innovation, possibly without slackening it. If workers undergo constant training, their ability to adapt and sometimes "*to reconvert*" can be accelerated whenever necessary. They need to be provided with precise and objective information, allowing them to adapt to the new scenarios in the best way possible. In the painful and extreme case they should lose their job, it is necessary to guarantee their capability and possibility to find another one.

Therefore, it is necessary to invest on informing citizens (raising public awareness on technology) in order to ease in any moment their reconversion and adaptation in front of fast and sometimes upsetting changes caused by technological innovation.

Emblematic are the episodes of industrial giants in the photography and mobile telephone sectors disappeared in a few years, causing an economic crisis and the loss of jobs in said fields and in the Countries of reference. Their mistake consisted in not forecasting, or anyway in underestimating, the range of technological epoch-making changes, such as the passing from films to solid sensors or from keys to touch screens.

The latest data collected by the World Economic Forum[15] through a survey among big companies on the main challenges of the future workforce strategies confirm our considerations. In fact, the surveys highlighted that the first priority is to invest in reskilling employees: for 65% of the companies the challenge will be to reconvert workers, requalifying them from a professional viewpoint. The second priority is to support mobility and job rotation for 39% of the companies. The process of continuous learning will have to be ethically guaranteed not only by the society, but also by big companies. Such challenge is inherently connected to the results of a second survey carried out by the World Economic Forum asking companies to identify the barriers to change with regard to the organisation of work in the industrial sector: for 51% of the companies the first barrier reported was the insufficient understanding of disruptive change, followed by resource constraints (50%) and pressure from shareholders for short term profitability (42%).

Innovation is taking place at such a high speed that the production world and its workers are unprepared both to identify changes and new opportunities and to provide workers inserted in the production system undergoing such a fast evolution with refresher courses and reconversion.

The ethnic and strategic resources of an advanced society must be able to forecast such dynamics before they take place, and therefore define a study a priori and a strategy that can protect

[15] World Economic Forum, The Future of Jobs Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution, January 2016, http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

citizens. A time lag between the fast, at times hectic, technological innovation and the slow individual and social assimilation is unavoidable, almost physiological. This automatically creates a gap which is difficult to fill with spontaneous mechanisms. It is necessary to formulate a strategy aimed at providing continuous information and training capable of mitigating said gap, involving citizens in the strategic choices and fostering their "flexibility" and "mobility". Such strategy should start from schools, that need to be able to intercept the signals of change. It should then continue and develop in universities and research centres, that have to prepare the innovators of the future as well as those of the present, creating an agreement between the public and private sectors for workers' continuous training.

Lastly, it is important not to forget the fundamental role played by dissemination and mass media to raise social awareness on the relevance of technology and its impact at social level. Part of the responsibility falls also on scientists and technologists: they have to plan and carry out their work according to ethical indications. At the same time, they must do all they can to make users aware of the social and ethical issues of robotics. This is the only way in which the society will be able to actively participate in the process devoted to creating a collective conscience capable of identifying and preventing a wrong use of modern technology. Users, suitably informed on the opportunities and limitations of technologies, will be able to participate in defining public policies and regulations.

In the specific case of A/IS, it is necessary to develop an inclusive and participatory strategy in citizens so as to avoid utopian hopes on the one hand, and irrational fears on the other. Emotional or ideologized behaviours can divert the attention from the real problems and, in the end, cause reactions of illusory enthusiasm or of generalised and uncritical refusal. The latter, though, can be very harmful if it hinders in a sterile manner the development of technology. Indeed, technology can truly be an important instrument for economic development and social progress, insofar as it helps humans without damaging them or it replaces them in a positive way.

2.3 Replacement of human workforce and new jobs

One of the most frequent questions on robotics concerns its impact on the labour market. The introduction in the society of robots that replace humans, and therefore reduce in some sectors the number of job positions, could lead to tensions and social crisis.[16] Such situations need to be somehow forecasted and analysed, assuming a development capable of respecting the right

[16] Bill Gates, the founder of Microsoft, in the global debate on the increasing presence of robots in factories at the detriment of humans losing their jobs, suggested to tax robots in order to create a solidarity fund for the unemployed.

balance between benefits (efficiency, saving, technological competitiveness in the international market) and risks or damage (decrease in employment, welfare problems, greater economic and social inequalities.

According to recent reports (Henry Siu and Nir Jaimovich, Third Way, WSJ.com), between 1988 and 2014, robotics has caused a decrease of -22% in cognitive-high routine jobs and of -25% in manual high routine jobs. On the other hand, there has been a consistent increase (+24%) in cognitive job positions with a low routine, that is jobs that require high experience and typically human intellectual and creative skills; whereas, a moderate increase (+10%) can be identified in operational jobs with a low routine such as, for example, manual jobs based on experience (craftsmen, plumbers, etc). This confirms that repetitive and foreseeable activities, without important creative and decisional processes, can be replaced more easily and effectively by machines.

We can reasonably assume that craftsmen will never disappear, because there will never be a robot capable of replacing said figure and of achieving the same ability to categorise the surrounding reality and objects.[17] On the other hand, such conclusion appears in line with the differences already mentioned between humans and machines. Humans are cognitive creatures that recognise objects on the basis of their function: they are able to distinguish a glass, a goblet and a jug on the basis of their use. Robots, instead, do not recognise objects on the basis of their function, but of their form: if they memorise a specific type of glass as a useful object to give in order to drink, they will carry out that task using only that specific glass and not other equivalent or similar ones. To teach a robot to reason on the basis of function and not of form requires a cognitive approach that can currently be adopted, but is very complex and costly. Hence the difficulty for robots to improvise as craftsmen, to invent, to create new solutions and perform non-routine intellectual or manual work.

It is likely that new professional profiles will exist in the future workforce: digital nurses; home care for elderly and very elderly people; body-part makers; nano-doctors; bio-computer scientists; igeo-biologists; digital architects (cloud controllers*)*; material architects (related to 3D printing, recycle, sustainable materials, water cycle, waste); energy managers; food technologists (traceability, analysis, packaging).

The mentioned workforce profiles still do not exist, or exist in an incipient manner; they will be created and developed together with a large number of job positions, owing to the spreading of

[17] “No machine can carry out the work of an extraordinary human," this is the phrase used in an advertisement showing a Stradivari violin.

the use of A/IS. They are professional roles that, according to the previous classification, require a computer, but are highly *cognitive* and lack *routine*. In the short term, however, the job positions that could disappear might be many more than those that will be created. Therefore, not only is it necessary to support workers with a programme of constant refresher courses, but also establish the priorities to adopt in the development of intelligent robotics, on the basis of a production system model that we want to have in our future.

Today, those who look at the future responsibly pose the problem of what kind of world we are leaving for the next generations; therefore, they do everything possible to contribute toward a sustainable development, in which humans do not simply consume resources, but also reintegrate them.

Also within the scope of A/IS, we should rightfully ask ourselves what type society we would like to have, for example, in the XXII century. Today, the spontaneous trend is to privilege a model in which *homo habens* is at the centre of society. Among the core objectives there is the constant increase of productivity and the growth of the GDP at local level through the unconditioned exploitation of natural resources, from water to lithium. In the future, in a world where robotics will have made gigantic steps and there will be intelligent machines a bit everywhere, different models may exist, among which those assumed by the supporters of post or trans humanism. The model we would like to assume could be called *homo sapiens 2.0*. *Homo sapiens 2.0* will be able to improve its performances owing to the help of robots: A/IS will optimise processes, substitute humans in burdensome, dangerous and exhausting tasks, but at the same time they will improve the sustainability of the production processes. Besides, in the model we are assuming, the GDP will grow at global level, not local. Robots will not be used trivially to increase the GDP indefinitely, but to reduce for example the water and carbon imprint of manufacture and its energy costs, spurring growth based on long-term sustainability more than on the increase of short-term productivity. [18]

2.4 Inequality: the 'robotic divide'

Providing citizens with training and information will be an essential element in order to guide the technical-scientific progress ethically and in favour of humans. Besides, it will help people adapt

[18] A part of the challenges lies in precision agriculture: currently millions of square kilometres of cultivated land are water-sprinkled with pesticides and fungicides with gigantic mechanical arms. This means that users ingest significant amounts of dangerous chemical compounds. The new robotics, provided with image reconstruction - that is, sufficient intelligence to recognise a sick plant - can be applied to GPS controlled robots that, walking in the fields and analysing one plant at a time, when finding a sick one, they take care of watering it. The saving of chemicals would be enormous, and also health would be more safeguarded. It is an example of an autonomous machine, provided with a body and artificial intelligence, used to improve a production process and public health.

more effectively to the swift changes in the labour market, avoiding them to remain excluded from it. However, the fast progress of robotics and its progressive insertion in the global market can create new forms of discrimination between those included (those inserted and at pace with the technological and robotic society) and those excluded (those unable to acquire the abilities necessary to be competitive their fields of work).

The rapid development of A/IS, as well as of all other technologies, can create an important gap between skilled and unskilled, qualified and unqualified, as we sometimes notice today the deep gap between digital natives and digital illiterates. As the presence of robots in the society consolidates and their contribution in improving the quality of life becomes more and more widespread, it will be necessary to make an economic and social evaluation to avoid the so-called '*robotic divide*': a strong inequality among individuals, social segments, countries, entire continents, due to the costs and difficulties to access new technological resources. It is a global problem: *robotization* can create a further gap among countries technologically evolved and underdeveloped countries. It is assumable that said gap may increase the already consistent energy divide and digital divide, due to which about 80% of the global energy is consumed by the eastern coast of the USA, by Europe and Japan, that represent 20% of the global population. The citizens in these areas of the world have a greater wellbeing and a longer expectation of life compared to the rest of the world. A global society in which intelligent machines play a primary role in improving sustainability could lead to a decrease in differences at global level, as well as a more fair distribution of water and energy resources and an improvement of the ecosystem.

Part 3. Ethics of Autonomous and Intelligent Systems (A/IS)

3.1 A code of ethics for machines and designers

As mentioned, robotics and the development of intelligent machines will lead to new social, economic and cultural scenarios that at the moment can only be imagined and forecasted. However, since this possible technological revolution is at its beginning, it is useful and important to make it grow and develop in the right direction: not against humans, but in favour of them.

Robotics and artificial intelligence are progressing in an increasingly fast and synergic manner, assisting, and sometimes replacing, humans in many ordinary activities. What many think will be an actual "robotic revolution," will cause changes in the civil society and in daily lives: in the field of games and entertainment, in the domestic, educational and industrial sectors; in transport, in the organisation of cities, in security, and in the maintenance of public order; in agriculture, in the production of energy, in the protection of the environment, and in the military field. Autonomous artefacts suitable for different uses - with different levels of autonomy in their behaviours and performance of tasks - will have to coexist with humans in an ethical and sustainable manner.

Many experts of the sector assume the need to formulate a code of ethics with which designers will have to comply when designing and programming intelligent machines, especially if autonomous; even machines will have to adapt to said principles, in order not to have damaging behaviours against human beings, the environment and themselves.

The forerunner of all codes of ethics valid for robots was the one laid down by Isaac Asimov[19] with his famous laws:

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

These laws are undoubtedly valid and brilliant, but they are especially an interesting and relevant cultural reference; to use them *sic et simpliciter* as moral rules of today's robotics would be like using Aristotle's Nicomachean ethics to regulate today's family and social life.

The programmers of future "machines provided with ethics" highlight that such rules are too general, potentially contradictory and do not explain the complex and real situations in which robots work with very different tasks. Moreover, should we ever manage to create robots capable of following moral rules, with which moral code will they comply given the presence of a widespread pluralistic ethics?

According to K. Abney,[20] it will be possible to ethically programme future robots only when it will be possible to programme and define with certainty their ability to calculate the long-lasting consequences of their actions. Abney holds possible the formulation of a code of ethics aimed at realising "good robots;" it will have to be based on programming pre-established and functioning virtuous acts, and keep into account the specific tasks of machines.

---

[19]Isaac Asimov (1920-1992), biochemist and author of science-fiction novels, was reflecting on these types of problems when he formulated the "Laws of robotics;" see among the many writings *Liar !*, 1941, *Runaround*, 1942; *I, robot*, 1950.

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

Although it is not assumable and realistic to talk about a "conscience" in machines, scientists, politicians and simple citizens feel the need to identify ethical guidelines for those who design and build intelligent machines.

The already mentioned "*Report with recommendations to the Commission on Civil Law Rules on Robotics*" established a guiding framework for designing, producing and using robots, based on principles of autonomy, beneficence, justice and non-maleficence. Said principles - introduced by Beauchamp and Childress in the field of medical bioethics[21] - are universally accepted as point of reference also in the technological sector.[22]

The principle of autonomy refers to the citizens' right to take informed and aware decisions on how to interact with the new technologies represented by A/IS.

The principle of charity recalls the requirement for A/IS to act in the interest of human beings.

The principle of lack of malice, according to the doctrine of '*primum non nocere,'* highlights the need not to injure humans.

Lastly, the principle of justice highlights the aim of a fair division of benefits associated to robotics and, in particular, on allowing robots that provide assistance and health care to be economically accessible to all.

Therefore, the general guiding principle established is that research activities in the field of robotics must comply with human fundamental rights. Moreover, their formulation, implementation, dissemination and use must be carried out in the interest of the single individual and of the society, and in the full respect of human dignity, both physical and psychological.

3.2 An ethical framework shared by the entire scientific community

The valid ethical framework of reference in the technological field that we are analysing is the one promoted by IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (A/IS) formulated by the Institute of Electrical and Electronics Engineers, Inc. (IEEE). Said Institute is one of the largest technical and professional organisations currently existent, with more than 420,000 members in over 176 countries worldwide.

In guiding the ethics of the current digital world, IEEE Global Initiative keeps into account the different cultural traditions developed in the East and West in more than twenty centuries, while

---

[21] T.L.Beauchamp, J.F.Childress, *Principles of Biomedical Ethics*, OUP 2012

[22] Cf I. van de Poel, *An Ethical Framework for Evaluating Experimental Technology*, Science and Engineering Ethics, vol. 22, no. 3, pp. 667-686, 2016

studying morality in amoral systems: the question whether the decisions taken by technological products, which in themselves are amoral, can have moral consequences.

IEEE Global Initiative gathers more than 250 experts, considered worldwide leaders in the study of autonomous and intelligent systems in the academic, industrial, philosophical and political world. They are working together to find consent and shared positions with regard to ethics in their fields of study. Therefore, they are providing practical guidance to all those who design and develop A/IS, with the aim to help them give the right priority to ethical principles and maintain technology at the service of mankind.

As a result of said collaboration, in 2016 an interesting programmatic document was published: the first version of the Ethical Aligned Design; currently, a new work based on a shared research is being produced, that will lead to a second version within 2019.[23] Moreover, eleven IEEE P7000™ Standards Working Groups have been established, that will provide guidance and orientation in the research work that will be conducted in the upcoming years.[24]

The first two parts of this document were concentrated on explaining how robots and intelligent autonomous systems were created to reduce and sometimes replace the intervention and the presence of humans in several daily life fields. We also mentioned the issues related to their impact on the society and single individuals, the relevant highly positive confirmations, but also the concerns and alarms for possible threats to privacy, loss of competences, discriminations, impact on the economy, long-term effects on common wellbeing.

Given their role so closely connected to the vital aspects of the society, the greatest benefits can be achieved only if these new technologies are in line with the values and ethical principles of our

---

[23] *The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2.* IEEE, 2017. http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html.

[24] IEEE P7000™ - Model Process for Addressing Ethical Concerns During System Design
IEEE P7001™ - Transparency of Autonomous Systems
IEEE P7002™ - Data Privacy Process
IEEE P7003™ - Algorithmic Bias Considerations
IEEE P7004™ - Standard on Child and Student Data Governance
IEEE P7005™ - Standard for Transparent Employer Data Governance
IEEE P7006™ - Standard for Personal Data Artificial Intelligence (AI) Agent
IEEE P7007™ - Ontological Standard for Ethically Driven Robotics and Automation Systems
IEEE P7008™ - Standard for Ethically Driven Nudging for Robotic, Intelligent, and Automation Systems
IEEE P7009™ - Standard for Fail-Safe Design of Autonomous and Semi-Autonomous Systems
IEEE P7010™ - Wellbeing Metrics Standard for Ethical Artificial Intelligence and Autonomous Systems

communities. It is therefore necessary to establish an ethical framework of reference capable of guiding and supporting dialogue and debate concerning the consequences of their use.
According to EAD, the design, development and realisation of A/IS must be guided by several general principles, such as:

- Human rights
- Wellbeing
- Accountability
- Transparency
- Awareness of possible abuse

Let's briefly analyse these principles.

Human rights

After the horrors of the Second World War, in 1948 the General Assembly of the United Nations approved the Universal Declaration of Human Rights (UDHR), composed of thirty articles relating to two main categories: political and civil rights, and social and cultural rights. However, said Declaration did not have value of law. Therefore, in 1966, two treaties were drawn up: the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights (ICESCR), which in 1976 became laws. In the end, the two Covenants together with the previous Declaration constitute the International Bill of Human Rights, the universal framework of reference for human rights.
In the following years, further documents were produced on specific themes, such as racial and gender discrimination, or torture; we hold useful to mention, for our topic, the UN Guiding Principles for Business and Human Rights of 2011, which requires all enterprises to respect human rights, and therefore also to those involving A/IS. Indeed, they are of interest for the topic we are covering.
EAD, referring to human rights, concentrates especially on the protection of personal data and on the control of their access and use by third parties. People have the right to limit the access to their personal data, in particular digital data, and to give each time their informed consent. Indeed, it is necessary to have mechanisms that defend citizens' identity, as well as procedures that make citizens aware of the consequences deriving from the storing and sale of their personal data.

Wellbeing

The Gross Domestic Product (GDP) is a well-known meter of evaluation of the economic wellbeing of a country, and is often at the basis of the political choices adopted by governments to improve their country's conditions. However, the GDP keeps into account only monetary criteria: investments, individual and collective consumptions, importations and exportations.

To evaluate the wellbeing of a country there are other factors that need to be kept into account, such as for example environmental sustainability and social inclusion.

In November 2007, the European Commission and Parliament, the Club of Rome, WWF and OECD (Organisation for Economic Cooperation and Development) organised a conference to launch a political and public debate on the topic. Many experts and representatives of the international political world deem necessary to introduce other parameters to measure a nation's prosperity: the average citizen's health conditions, weather conditions, exhaustion of resources, pollution, biodiversity. There is the perception of an urgent need for alternative instruments capable of measuring wellbeing, as well as the need to establish indicators of social progress and environmental sustainability.

In October 2008, following the dramatic global financial crisis, the General Assembly of the United Nations set up the Commission of Experts of the President of the UN General Assembly on Reforms of the International Monetary and Financial System chaired by Nobel prize-winning Joseph E.Stiglitz.

The Commission's work produced the "Stiglitz Report: Reforming the International Monetary and Financial Systems in the Wake of the Global Crisis,"[25] which provides guidelines for a long lasting, democratic, fair, solid and sustainable development.

Lastly, we would like to mention OECD's Better Life Initiative[26] which every year publishes a "World Happiness Report," providing a list of the worldwide countries' wellbeing, on the basis of several indicators: housing, income, jobs, community, education, environment, governance, health, life satisfaction, safety, work-life balance.

EAD applies the principle of well-being to A/IS, according to which their design and use must give priority to personal and social wellbeing. Autonomous and intelligent systems must be accessible and bring benefits to all populations, wherever they are, also through a universal and easy access from an economic viewpoint to communication networks and the Internet. Said systems must improve individual and collective wellbeing, give an important contribution to the solution of

[25] http://www.library.fa.ru/files/Stiglitz-Report.pdf
[26] http://www.oecd.org/statistics/better-life-initiative.htm

humanitarian problems and issues related to the development of all populations, and lastly lead institutions to structure themselves in a more human-centric sense.

In order to make sure that A/IS truly bring actual benefits to mankind, suitable indicators are necessary. Evaluation criteria such as success, profit, safety at work and fiscal health are important. However, they do not manage to fully express all the factors necessary for individual and collective wellbeing, for example psychological, social and environmental factors.

It is necessary to develop suitable metrics that keep those elements into account as well, allowing a more complete evaluation of the benefits and damage caused by the new technologies. These new metrics, besides considering and evaluating what realised up to now, can open new and stimulating paths for technological progress.[27]

Accountability

The civil and criminal liability of designers, creators and users of A/IS needs to be clarified. The synergies between intelligent systems and robotic technologies have given life to systems with characteristics that simulate the humans ones: autonomy, ability to perform intellectual tasks and anthropomorphic physical aspects. This gives rise to the issue related to the legal status of A/IS, which interweaves with broader legal themes, in particular those related to liability should A/IS cause any damage.

EAD has highlighted the following priorities:

- Autonomous and intelligent systems must be subject to suitable regimes implementing property laws
- Governments and industries must identify decisions and operations that cannot be delegated to A/IS
- It is necessary to adopt rules and standardised procedures that assure humans' control over those decisions
- It is necessary to establish how to assign the legal liability of the damage caused

Transparency

Autonomous and intelligent systems must act in a transparent manner. They are provided with algorithms and systems for analysing data allowing them to "learn," to improve their performances, and to take autonomous decisions that have an impact on the society and on the

[27] Cf I. van de Poel, *An Ethical Framework for Evaluating Experimental Technology*, Science and Engineering Ethics, vol. 22, no. 3, pp. 667-686, 2016

single individual. Therefore, their design, realisation and functioning must be transparent, participative and truthful from an ethical and legal viewpoint.

In particular, EAD suggests the following objectives:

- The parties involved, their lawyers and courts must have access to all the data and information generated and used by A/IS, and held by governments and authorities
- The logic and rules inherent those systems must be accessible to who has a function of control, and are to be subjected to rigorous risk evaluation tests
- A/IS must generate control memories, accessible to third parties, that register the facts accomplished and the decisions taken and must have legal consequences
- Users must know who sponsors and establishes, with their investments, the ethical decisions of those systems

Awareness of a possible abuse

It is necessary to minimise the consequences of a bad use of A/IS through strategies that raise awareness on possible risks, and guarantee the promotion and protection of safety, privacy, intellectual property rights, human rights, cybersecurity. It is necessary to lead users and communities to understand the possible consequences of the social and individual impact deriving from the use of these new technologies.

In order for A/IS to serve the common good in the best way possible, EAD has suggested the following aspects:

- To support, promote and implement legal rules and regulations recognised universally
- To create working groups with competences in technologies connected with A/IS
- To have a suitable leadership in the research conducted on said systems and in their development
- To establish rules that ensure public security and liability
- To educate users and citizens on the possible impact of the new technologies

3.3 Final considerations

On the basis of what laid down by the *IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, we have highlighted several general principles that can effectively guide the design, realisation and use of intelligent machines. Lastly, drawing inspiration from the same initiative, we would like to emphasise several themes of particular relevance for the present and for the immediate future.

The choice of the ethical values inherent autonomous systems

If machines become part of humans' communities as autonomous or almost autonomous agents, they will have to comply with the ethical rules and regulations of the community to which they belong. This will have to be kept in mind when machines are "taught" several ethical values, in other words when said values are technically inserted in their decisional processes. The ethical context of a robot working in communities with different culture and/or religion maybe very different. Even within a homogeneous community from a social and ethical viewpoint, there are different ethical requirements for a machine, depending if it interacts with children or adults. Therefore, it is necessary to identify ethical references suitable to the environment in which the system acts and to the type of operations it performs.

Guiding methodologies for an ethical research and design

It is necessary to develop A/IS which increase and extend humans' liberty and wellbeing. Design methodologies based on ethical and human values place human progress at the centre of the development processes of the new technologies. Machines must be at the service of humans and not vice versa. Developers of A/IS must be able to use methodologies based on fundamental values, create sustainable systems that can be correctly evaluated in terms of social costs and of the economic benefits they truly offer to the different enterprises and organisations.

Re-contextualisation of military A/IS

Autonomous systems designed to cause physical damage have special and unusual ethical consequences compared to traditional weapons or non-armed systems. It is necessary to guarantee at least the following requirements:

- To ensure they are controlled by humans
- To design them in such a way that they have tracing systems that guarantee their control and assignment of liabilities with regard to their use
- Their learning and adaptation systems must be able to explain their reasoning and decisions to human operators in a transparent and comprehensible manner
- It is necessary to train the human operators responsible for their use and they must be clearly identifiable

- It is necessary to ensure that human operators can forecast the behaviour of their autonomous functions
- It is necessary to make sure that the creators of military A/IS are aware of the consequences of their work
- It is necessary to develop professional codes of ethics that duly guide the development of military A/IS

Safety and benefits of the AGI (Artificial General Intelligence) and of the ASI (Artificial Super Intelligence)

Similarly to other current powerful new technologies, the development and use of intelligent systems capable of learning and improving themselves entail considerable risks, especially in the event of a bad use or wrong design. According to several current theories, as the system's experience and learning increases, several unexpected or unintentional behaviours can assume increasing dangerousness and become difficult to correct. Not all reasoning structures of the AGI and ASI can be in line with the common good and humans' interests; it is necessary to take care of establishing how the different mental architectures work, as they make progress and become more expert.

Affective Computing

Affectivity is a key aspect of intelligence; emotions such as joy, fear, anger are often at the basis of the motivations of our daily behaviours. It is necessary to make sure that A/IS are used to help humans with the widest range of situations possible, and that they include all dimensions of the humans' lives. When they are used to participate in the citizen's daily life and improve it, they are not to injure the citizen, changing the latter's emotional experience into something negative. The rudimentary versions of synthetic emotional systems currently in use highlight the great impact that they can have on users and on the large public in the social and political world.

Mixed reality

Mixed or hybrid reality is obtained when mixing real and virtual worlds to obtain new and unusual environmental conditions; real and digital objects coexist in said reality and interact in real time. Mixed Reality (MR) is becoming increasingly present at work, in education, in social life and commercial transactions. It is foreseeable that it will change the current and classic concepts of identity and reality. The possibility to modify in real time the parameters that regulate mixed reality raises ethical problems connected to individuals' right to control their multiform identity, both real and digital.

It will be necessary to pay particular attention also to the use of increasingly sophisticated immersive technologies, that use sensors more and more invisible and integrated in the human body.

Robotic addiction

In many fields of life and human customs it is possible to develop an addiction to objects or instruments that when used produce emotions and comfort. This can lead them to become an essential part of our lives, to the point that we can no longer live without them.

With social addiction to robots we mean the type of addiction to intelligent machines that we can easily assume possible through their dissemination and their becoming part of the life of many individuals and families.

It is thought that, in the short term, the evolution of robotics may produce a phenomenon similar to the revolution of computers: in just a few years, in fact, we have become addicted to computer technology, in all its aspects related to networks, streaming, games online, smartphones and social networks. It is easy to assume that in a near future a similar phenomenon will occur producing addiction to robots, as they will be inserted more and more massively in our daily lives.

Also due to this addiction, according to some, the increase of robotic technology in the fields of education, entertainment, art, healthcare instead of strengthening humans and their abilities, could increase their vulnerability, due to several "pathologies" typical of these situations: the difficulty to distinguish what is real from what is virtual, what is natural from what is artificial, what is true from what is likely.

In the case of anthropomorphic robots for social use, we have to make a further consideration: despite how clearly users are able to make a distinction between robots and human beings, the interaction with robots that have human features can produce emotions, attachment, addiction, in particular with regard to persons in situations of frailty: elderly people, people with disabilities, children with difficulties or not fully developed in their cognitive faculties.

There are studies that analyse the aesthetic impact of the design of robots at emotional level, on the basis of age, cultural conditions, people's character, etc. It is necessary to reflect together on how to use the results of these studies to protect users the best way possible. "Opt out" mechanisms can be forecasted capable of intervening before triggering a process that leads users to an excessive addiction to machines (such as "alarming" systems when the exposure to certain

technologies is excessive) or to think of limiting the perfection of robots' humanoid morphology, so as to reduce the affective impact, leaving unaltered the functional one.
To this regard it is necessary to consider the problem of "deception": robots act like human beings and imitate their behaviours; sometimes they simulate feelings that they do not really feel. Such situation can constitute a form of deception and illusion (robotic deception) in people that are not aware of this aspect, causing harm.

On the basis of these final considerations and in the light of the analysis of the developments of robotics in the social, medical, military and legal fields, it is possible to suggest several recommendations for the future:

1. To provide citizens with critical information on the developments, potentialities and limitations of A/IS, so that they may acquire critical awareness and avoid emotional reactions of excessive enthusiasm or repulsion, affected by science-fiction scenarios and non-realistic situations.

2. To promote an interdisciplinary analysis of the impact of A/IS on the society (with particular attention to work) and the study of strategies for the non-replacement of humans and the enhancement of human work in the robotics era.

3. To strengthen the training system inserting programmes and curricula necessary for the development of technological skills indispensible in the era of this robotics revolution.

4. To promote the analysis of the psychological and emotional impact of A/IS on humans and the formulation of strategies aimed at avoiding forms of robotics addiction.

5. To balance the developments taking place in robotics, avoiding discrimination between those included (those inserted in the technological and robotics society) and those excluded (due to the inability to acquire the necessary skills); to promote modalities for assisting who is in a condition of "technological vulnerability" (elderly people or people with cognitive disabilities).

6. To formulate code of ethics for designers/builders of A/IS and establish ethical committees for robotics research, as well as to foster interdisciplinary research among experts in the scientific, ethical and legal fields on themes raised by the fast technological innovation.

7. To introduce the study of ethics in engineering and computer courses, to spur from the very outset at university evaluation skills and moral responsibility in the field of the new technologies.

8. To adjust to the new robotics technologies the international laws and treaties that regulate and limit the use of weapons, in the respect of human rights and dignity. In particular, to promote a shared reflection on the use of A/IS which, lacking humans' remote control, may kill or cause serious harm.

9. To ensure the respect of privacy and the right to intimacy in the production and use of A/IS which can spy on citizens' lives and invasively control them.

10. To safeguard, with regard to the legal liability of A/IS, citizens, users and enterprises, keeping into account how much autonomy and learning abilities machines have, and therefore how much control they are given by programmers or owners. To aim toward the formulation of a common legislation at international level with the objective to ensure coherence and legal certainties.