Before and Beyond Artificial Intelligence: Opportunities and Challenges

M. Patrão Neves and A. Betâmio de Almeida

Abstract Artificial intelligence (AI) and digital systems are currently occupying a fundamental place throughout society. They are devices that shape human life and induce significant civilizational changes. Given their huge power, namely systems with autonomous decision-making capacity, it is natural that the potential social effects deserve a critical reflection on the opportunities and challenges addressed by AI. This is the main goal of this text. The authors begin by explaining the philosophical position from which they start, and which contextualizes their reflection on technological innovation in general, then briefly considering the genealogy ("before") of AI, in its main characteristics and direction of evolution ("Can machines imitate humans?"). It is considering the path of development of AI and its disruptive effects on human life ("beyond") that it is proposed its systematization in three categories—functional, structural, identity—("Can humans imitate machines?").

Regardless of the optimistic or pessimist expectations towards technological evolution, there is a need for a public debate about its current and future regulation. The text also identifies major ethical principles and legal requirements to regulate AI in order to protect fundamental human rights.

1 Few Presuppositions that Shape the Reflection on AI

The structuring, developing and using of AI is particularly complex and challenging for a non-technical, social and human reflexive approach. This is mainly due to the following distinct but cumulative aspects. AI is of a multidisciplinary nature, mobilizing a growing diversity of knowledge and techniques—digital, elec-

M. P. Neves (✉)
University of the Azores, Ponta Delgada, Portugal

A. B. de Almeida
IST/University of Lisbon, Lisbon, Portugal
e-mail: betamio.almeida@ist.utl.pt

H. Sousa Antunes et al. (eds.), *Multidisciplinary Perspectives on Artificial Intelligence and the Law*, Law, Governance and Technology Series 58,
https://doi.org/10.1007/978-3-031-41264-6_6

tronics, computing, mathematics, statistics, social and human sciences, including law, sociology and philosophy—which turns it inevitably complex and makes a comprehensive discourse very difficult or even impossible. At the same time, the domain of AI is currently so broad, diverse and dynamic that any discourse on the subject becomes inexorably restricted and maybe also quickly outdated. Finally, interpretations of what AI represents in the present, but especially in the future, are so disparate—ranging from naive enthusiasm and social submission to castrating pessimism—that any position taken is open to criticism, and the one that is now presented will not be exception.

Our reflection, like any other, is based on some assumptions that, more implicitly or explicitly, shape it, and should therefore be disclosed. We can briefly present four major presuppositions that ground and shape our reflection.

The first is that technology is a product of human creativity, so it cannot be generally and immediately demonized as if it were a strange and hostile reality to us. In fact, technology has been fundamental for the survival and quality of life of humanity. It creates its own life conditions out of the given world**.** The negative attitude is still all too frequent, especially in the face of uprising powerful technological innovations. These tend to arouse feelings of fear in relation to the new, the unknown, a certain uneasiness or even distress (although today we often witness an uncritical attraction to the new, as if everything new was good). There is also a certain hostility towards technological innovation in the assessment of its effects—for example, environmental degradation is attributed to technological impacts—sometimes only blaming the technique (technophobia) and with a total lack of reference to other causes and responsibilities. Experience teaches us that the personal benefits arising from a technological innovation is what attracts the most at the beginning and the possible negative social or collective impacts only later become evident, frequently when that particular technology is widespread and it is very difficult to oppose. In this case only a crisis will drive a change. This justifies an independent critical analysis of the creation of technological products and their mass applications.

A second presupposition is that technological innovation (such as scientific progress) is unstoppable, irrepressible or deterministic, so it cannot be suppressed, but rather re-oriented. Even if it were desirable to stop scientific progress and technological innovation (which in any case is quite doubtful), they will never cease to develop due to a combination of variables—economic-financial, social, political, academic, etc.—that generate an increasingly powerful and continuous dynamic that surpasses the sum of the variables involved, beyond the control of any single person or group of interests (Liu 2021).[1] It can be possible to slow down the process (It has already happened in some other innovations in order to avoid severe impacts), being imperative or preferable to reorient it. However, the potential uncontrolled impulses

[1] "The global artificial intelligence (AI) software market is forecast to grow rapidly in the coming years, reaching around 126 billion U.S. dollars by 2025. The overall AI market includes a wide array of applications such as natural language processing, robotic process automation, and machine learning."

in the application of the increasing power, by private companies or public agencies, that new technologies provide and that may pose risks to humanity, seems to be a matter of urgent reflection and control. The problem of human techniques was not traditionally an object of special attention in philosophy and ethics. This situation has changed since the mid-twentieth century. The growing technological power has motivated philosophy and ethics to critically analyze the essence of technology and its impacts on humanity.[2]

A third presupposition is that technological innovation is neither axiologically neutral nor, therefore, exempt from ethical scrutiny. Technological innovation is not purely instrumental, as if its evaluation depended only on its use and on the user. In fact, every creation already bears the mark of its creator, even if it is nothing more than the intention that led to the creation, to the production, a structural and original intentionality (the principle of its development, in an irrepressible and irreducible evolution), which escapes human control, and rather conditions and even induces human behavior. New technologies, by the simple fact that they exist, induce their use. Astonishing technological development is the result of human desires that are difficult to control.

The fourth is that technological innovation should not be an end in itself, but rather a means in terms of the only end in itself, which is the human. The *raison d'être* of all human production is to constitute new and diversified modes of promotion and realization of human flourishment, which is why it must remain inexorably subordinated to humankind. The fundamental challenge that arises is whether technology should be an instrument at the service of humanity (e.g. an instrument to improve human health) or whether it is humanity that should adapt to the demands of technology.

Acknowledging our assumptions, we should now more accurately identify some of the major opportunities opened by AI, and think about the risks or challenges its development entails, going from the birth of AI and its original objectives to its succession of new ambitions.

2 Can Machines Imitate Humans?

2.1 The Key Question

Can machines imitate humans?—is the question that the mathematician Alan Turing, the so called "father" of theoretical computer science and AI, poses in 1950, in his Imitation Game, and to which he seeks to be able to respond positively

[2] All human techniques have gradually contributed to the structuring of life in society, namely through the formation of a "socio-technical system". Digital technologies and AI are, in a very intense and fast way, densifying this system and significantly altering the human way of life by diffuse social impacts. There is also an intense convergence with other very relevant technologies, namely the set of nanotechnology, biotechnology, information techniques and neuroscience. All together may induce a significant change in human evolution.

throughout his life: "can machines think?" (Turing 1950).[3] We would say that Turing's question possibly marks a turning point in the relationship between humans and machines as striking as Jeremy Bentham's interrogation in 1789, "Can animals suffer?" triggered in the relationship of people with animals.

In the second half of the twentieth century, machines did seem to be intelligent. Digital computers, so designated because capable of manipulating discrete symbols, or digits, had been created in the wake of the third industrial revolution, characterized by Automation, very focused on information and communication technologies. The question of the moment was: can a computer behave intelligently like a human being?

A first answer is given by the "Turing test", the so called "imitation game": is it possible for an interrogator to distinguish the answers given by a computer from the answers given by a human being? Can machines impersonate human intelligence, or imitate human intelligence?

The Turing Test has been the subject of much criticism, many of which result from the exact definition of thinking and intelligence. One of the most famous is based on the well-known Chinese Room argument by Searle (1980). The Turing Test is based on language. We know that language is fundamental in the development of human intelligence, but intelligence should not be directly confused with knowledge or memory. Is a simple question-answer test a sufficient means to identify human thinking and all types of human intelligence?[4] With Turing we intend to be able to identify an acceptable similarity with the way of thinking and reacting of a human, possibly what we might want is to recognize that a machine is capable of imitating the human way of thinking very well.[5] Much more difficult will be to recognize the sentience capacity of a machine!

2.2 The First AI Steps

It is in this context that Marvin Minsky and John McCarthy come to forge the expression Artificial Intelligence that they present in 1956, at the Dartmouth College Conference, organized that year in the United States, and which brought together

[3] As fascination, ghost or myth, the more or less repressed will to create an artificial human has accompanied humanity for centuries. The current interest in humanoid robots may be an example of this ancient dream. What is new in the question posed by Turing is the focus on the intelligence attribute in an era with technological capacity to develop a credible answer.

[4] Among humans, we also use language to try to assess thoughts and levels of intelligence. However, in this assessment we already assume that we are dealing with humans. We admit that we recognize the basic structure of thought of other humans because we belong to the same biological species and we are both heirs to the essentials of a common natural evolution. In fact, what we can identify are variations in the behavior of human minds relative to a chosen pattern.

[5] There are many variants of the Turing Test in order to eliminate its supposed deficiencies and there is also the Inverse Turing Test to challenge an algorithm to distinguish a human from another algorithm in a dialogue.

the pioneers of AI of that time. In the same year, the two founded the Artificial Intelligence Project (now the MIT Computer Science and Artificial Intelligence Laboratory).

It is then that the history of AI truly begins, in which Turing came to propose that the strategy to follow should not be, as before, to try to "write a program that would allow a machine to pass the game of imitation" (reproducing parts of human reasoning), but rather that of writing "a program that would allow a machine to learn from experience, just as a baby does". It is in this direction (automatic learning, through experience) that today, decades later, the approach to intelligent systems is made. So, it already enhances the autonomy of intelligent systems in relation to humans.

We are then fully in the fourth industrial revolution, characterized by Connectivity, in which AI develops almost exponentially, which is confirmed as we now enter Society 5.0, the fifth industrial revolution, that is, the era of full connection, where everything will be connected, all the means available to human beings will be connected and persons will have to adapt or to integrate themselves into these continuous flow networks (alignment of robotic technology to human intelligence, increased collaboration or partnership between human beings and intelligent systems). AI has been developing and strongly driving the last 3 industrial revolutions, paving the way towards full automation and maximum connectivity (wireless, no physical connection).

Nevertheless, we still do not have a consensual definition of AI (which is very revealing of its dynamism), despite being quite relevant for the circumscription of its domain and perception of its operability. There are many different definitions and even those who reject the expression, namely Luc Julia, in his work *L'Intelligence artificielle n'existe pas* (Julia 2019),[6] where he considers that AI has always been poorly defined as it suggests that algorithms can make conscious and rational decisions like humans. He believes that this is not the case and that mistaken ideas like this one have fueled fantastic Hollywood perceptions about AI, such as Matrix or Terminator.

Human intelligence is difficult to delimit and fully understand. It is more than rationality towards stimuli and data analysis. It has other built-in features and a strong connection to the entire human body. Perhaps the designation Artificial Intelligence (AI) was very effective as a brand, but it is not very strict. The expression AI is used today to designate a variety of technologies with some common characteristics. We adopt the definition proposed by the High-Level Expert Group on Artificial Intelligence of the European Commission: "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information,

[6] Julia (2019), p. 287. It was Luc Julia who co-created the digital assistant Siri, one of the most famous AI.

derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions (European Commission 2019)."

Other, simpler AI definitions could be: "a computerized system, agent or robotic, capable of acting and making decisions independently of human supervision" (Tavani 2016).; "a system capable of rationally solving complex problems or taking appropriate actions to achieve its goals in whatever real world circumstances it encounters" (Dempsey 2020).

2.3 The Encouraging Achievements

AI, as we broadly define it, has been a powerful tool in achieving human purposes, whose continuous development has gone beyond its original instrumental status and conquered new performance plans to consider, in a continuous erasure of what seemed to be its limits. And yet, we are still in the era of a weak or narrow AI, that is, capable of performing just one or few specific tasks, and which software can only make decisions based on information previously given. Some common examples are: to play chess, the Go or poker; to identify people through faces captured in real-time security video (face recognition); or to drive autonomous vehicles.

If we take just one of these examples—the simplest, as playing a game—and follow the evolution of AI, we can easily understand the direction we are moving to. The first important step of its evolutionary process was given in 1996, when Deep Blue, an IBM software, defeated the world chess champion Kasparov. Later, in 2017, AlphaGo won game Go against the best in the world, and in 2019, Pluribus won a 12-day poker marathon, competing against 5 players. A second step was given when the software started to learn to play by itself, playing against itself, and thus relying less and less on human-generated data, since 2017. More recently, Google's MuZero was presented as being able to play without the need for any human-entered data, that is, without being given the rules, thanks to its ability to plan winning strategies in unknown contexts. It is this direction of AI evolution that fuels the greatest fear of humans: that of AI gaining enough power to completely escape human control. The direction of evolution that is being followed is easily revealed: advancing towards an always and successively superior performance in each of the functions that AI performs; and towards a higher level of automation (emancipation) of the human (creator, producer).

The evolution trend of AI and its applications justifies a serious fear of a devaluation of the humans in face of the superior capabilities of new systems in fields of activity that have structured society and the purpose of human life. The risks and challenges arise in the short term, but some of them are already threats: "the greater the digital capacity of a given society, the more vulnerable it becomes" (Kissinger et al. 2021). These are issues of particularly interests for Ethics and Law.

Today there is a clear perception that we are experiencing a digital revolution (which follows the industrial revolution) led by AI. That is, AI is a constant and indelible presence in daily lives of persons, individually considered, as of communities, particularly in the northern hemisphere, and our way of living depends heavily of AI which, today, penetrates most modalities of human action. We live in the AI era.

3 Can Humans Imitate Machines?

The idea of humans imitating machines would be regarded as foolish until recently. Today, however, we can formulate this provocative question because there are digital machines with an attribute held as superior in living organisms: intelligence. These machines, being presented as having intellectual capacities far superior to those of humans, may constitute models of individual and social behavior to follow. An alignment of humans to the rules of a new socio-technical system due to a simple adaptation by unconscious inertia or imposed as a priority justified by efficiency criteria but abstracting other criteria associated with human nature.

In an attempt to systematize the growing multiplicity of AI interventions in human life, we would say that its impacts are more evident and disruptive at three main levels: a *functional*, in the use of AI as a specific instrument for human purposes; a *structural*, in the change that AI entails in human interrelationships and in the organization of institutions; an *identity*, in the transformation that originates in what the human is and in the image he has of himself.

We must consider these three levels of AI intervention in the human sphere, both in the new opportunities it creates for human flourishment, and in the new challenges it poses for human perseverance in a context of performances that far surpasses it.

3.1 Functional Level

The functional dimension of AI refers precisely to its ability to carry out human functions, which it does by performing them faster, more perfectly, more economically, in a truly unique and impressive supporting human action. Some of its main very successful domains are industry, justice, health, education, transport, finance, marketing, computer security, army (military defense) and entertainment.

A quick glimpse at the intervention of AI in few of these so distinct and paradigmatic domains can give us a more precise idea about its disruptive potential, both positive and negative, in our contemporaneity.

AI first became preponderant in industry, where it is massively used and where its functional dimension is best evidenced, through the automation of various functions, especially the harsher, physically and psychologically. Releasing people from the heaviest burdens is strongly applauded. However, AI in the industry is not limited to

the automated functions, but is also being used to assist in decision making and data analysis, including personnel management, such as attendance levels and employee productivity, hiring and dismissing employees. However there are some paradoxes related to technology and productivity.[7]

Nowadays, the former general idea that AI only performs mechanical tasks, which are professionally less demanding and socially less valued, is easily contradicted. On the one hand, AI has been conquering a diversity of domains and levels of complexity of action, even in traditional fields, such as industry; on the other hand, it has been applied to increasingly more demanding fields of action, such as healthcare or justice.

AI is strongly present, both in clinical research (e. g. collecting gigantic amounts of data to identify correlations and trends; new therapeutic molecules) and in clinical care (e.g. making diagnostics; monitoring of health conditions). There are some medical specialties in which standard clinical procedures are being replaced by AI, such as radiology (reading exams) or ophthalmology (performing some exams), in which AI can advantageously replace physicians. Today there is already efficient digital assistance for medical doctors and nurses, especially in the area of geriatrics, surgeons, cleaning staff, but also for the delivery of medication, food and even some diagnostic tests.

In what concerns justice, AI has been heavily used, namely in the search for jurisprudence, in the adoption of justice measures based on similar previous cases. There are also already projects for the institution of an automatic predictive justice court to dispatch benign cases.

Indeed, it seems today that all human functions can be substituted by AI (they are being gradually replaced) with immediate advantages, under the principles of efficiency, productivity, and profitability. The promising idea that AI will liberate humans by avoiding tedious or monotonous intellectual tasks does not seem to be what one might anticipate: its exclusion from tasks associated with human thinking.

However, there are also some disadvantages associated that are important to be considered together, and among which we highlight only three.

A first one is AI proliferation. We refer to the proliferation of AI considering its ability to learn from previous experience in order to produce intelligent behavior and

[7] Although the new technologies hold great potential, there is an apparent paradox because productivity growth has slowed rather than accelerated (Brynjolfsson et al. 2017, p. 44). In fact, labor productivity growth in developed countries have stayed low since mid-2000 and there are different potential causes for this paradox. False hopes, a time delay until there is a statistical effect and the increasing market and rent concentrations are some of them. While income inequality has been rising within many countries in recent decades, inequality between countries has been falling. This is another apparent paradox but the way technology diffuses within the economy seems to be relevant for both productivity growth and income distribution (Qureshi 2021, p. 24). In EU this impact seems to depend on the country's size, its level of development and the current degree of income inequality relative to the average European value (Kharlamova et al. 2018). Reducing inequality can be considered as a way for preventing a future crisis or an ethical issue. We can conclude that there are both optimists and pessimists about the relationship between new technologies and growth.

correct decisions, which is called "machine learning" (a subdomain of AI): these are algorithms capable of modifying themselves and making decisions without human intervention. It has also advanced to the so-called "deep learning" (a subdomain of machine learning) which consists of the ability of computers to learn on their own, through pattern recognition, in many layers of raw data, depending on the proposed objective, carrying out tasks as human beings. Therefore, AI is always improving its performance and acquiring new skills. This aspect, immediately and necessarily recognized as positive, is presented here as a disadvantage insofar as it triggers the process of releasing Artificial Intelligence from human control.

A second disadvantage, and the most commonly presented, is mass unemployment. As the domains in which AI can assist human purposes multiply, as the diversity of functions it can perform grows, and as its performance becomes superior to that of human, it also replaces people. Hence, the main threat that has been stressed at this level is mass unemployment, as it is already obvious in industry.[8] We know the arguments that dismiss this growing problem: throughout human history there have always been work activities that have vanished and new ones that have emerged and the same will happen now too. We cannot fail to point out the existence of an unprecedented variable in this equation that can endanger the past balance: the speed of the process that does not allow human adaptation to the ongoing transformation and the intellectual quality of lost jobs. Even if many new jobs are created, the question of the type and social level of these jobs should be considered.

The third disadvantage is social exclusion. Indeed, the advantages and disadvantages of AI may not be evenly distributed, with the most favored persons being the most benefited and the least favored suffering most of the losses. Besides, this chronic inequity is added to the specific one of generational sharing: today we have a growing proliferation of generations, which no longer succeed each other every 25 years, but every 10 years.[9] In this unprecedented context, it becomes very easy for people to be considered outdated by the next generation, and at the same time, useless for society, perhaps even a burden or disposable. This intergenerational disadvantage can cause serious social fractures and be difficult to be solved without a profound change in the human society organization.

Characterizing AI in its functional range we would stress that: it remains outside the human and can be manipulated and controlled by him; it contributes to the construction of a civilization guided by technological, intelligent and automated innovation, and by efficiency and productivity. Therefore, it threatens to make the human obsolete.

[8] Deloitte estimates that, in the next few years, 50% of current jobs will become obsolete.

[9] In 2010, a new generation is formed for which the analogue world is past, asserting itself as 100% digital native, and surpassing the millennials, making all generations quickly outdated, namely the current X generations, from the early 60s to the 70s; the Y generation, from the end of 70 until the early 90's, and Z from 1992–2010, we also have designations such as the "grey generation" or the "snowflake generation".

3.2 Structural Level

The structural dimension of AI refers to new forms of relationship, new patterns of personal, social and institutional relationships, characterized by greater virtual proximity between everyone (overcoming geographic distances), by greater coverage (because all people are potentially included), and paradoxically, at the same time strengthens relationships by mediating them and suppressing direct contact.

The mediation of human relationships through Artificial Intelligence takes place today in a growing diversity of domains that we have systematized in three planes. At the personal level, people from all over the world know each other and socialize virtually (even for emotional intimacy relationships); at the social level, human activities are developed at the digital realm (where interest groups are formed, and civic, political or other activism is developed, demonstrations are scheduled, petitions are made, etc.); at the institutional level, institutions relate to citizens through intelligent technology (e. g. relationship with the public administration, as commercial transactions tend to be increasingly online and service is carried out by a chatbot, a computer program that tries to simulate a human being in conversations with people,[10] the same is happening in more and more domains as well as education).

At this level, we would like to highlight two examples, which are quite different, but both paradigmatic of the ongoing transformation. The first is the widespread investment in the construction of smart cities, that is, of population aggregates in which everything is connected, with automated management (traffic, waste, public safety), everything being mediated by AI: the household equipment tends to become totally connected and smart assistants can take care of all management services at home (managing waste, identifying equipment problems); all the equipment and infrastructure of a municipality will be connected (e.g. identification of aspects to be improved, safety, air quality measurement, traffic coordination, etc.). Structuring activities of human society such as banks and insurance tend to be on line, dematerialized (without paper documents) and without human intermediaries. This change creates new vulnerabilities in terms of security, trust in institutions and in person access to them. Citizens are increasingly subject to faceless technical systems with access based on multiple numeric codes and passwords.

The second paradigmatic example is related to the introduction of AI in politics (in addition to the other strategic domains already mentioned with health, finance and the army). In 2019, a study by a Spanish University concluded that 1 in 4 Europeans would be willing to allow AI to make important political decisions in their country, in favor of impartiality, honesty and justice.[11] Today there are already

[10] The illusion of machine-induced affectionate feelings is one of the aspects that already happen in relationships between accompanying robots and the elderly or also in the way some people react to automatic messages they receive on their birthday.

[11] Jonsson and de Tena (2019). Also, the philosopher Yuval Harari says that elections, political parties, parliaments can become obsolete given the amount of data to be taken into account and the speed at which some decisions have to be taken (Harari 2018).

references to an imminent formation of a "cyberocracy" that can threaten or destroy the democratic system as we know it.

The immediate convenience for human activities is obvious and indisputable, under the new principle of optimization of means. However, there are also associated drawbacks that are important to consider together, and among which we highlight here only three.

A first one, at the personal level, points out that the intensification of connections is directly proportional to the physical distance between people (relationships tend to be superficial, sporadic, ephemeral, without commitments or responsibilities, they become light relationships). The second unfolds at the social level and refers to the anonymization of personal uniqueness before the functional relationship (from the integration into categories of people and relationship patterns, structured based on interests). The third disadvantaged lies at the institutional level and refers to the integration of all human activity into a network of relationships (everything is in a network and what is not in a network lacks recognition of existence); networks are almost unknown, inaccessible and uncontrollable (the humans risk to become pieces of a gear that surpasses them). Dependence becomes extreme and the smart encoded numerical protocols are densified and drastically reduce the spectrum of human communication mode. In addition, we are increasingly integrating AI programs into decision-making processes.

AI, in its structural scope, presents itself as integrated in all human and social activities and shapes them, formats them; it builds a new culture guided by virtual (inter)mediation and connectivity, and by the optimization of resources; it threatens to number the human (representing the human through numbers, depersonalizing it). The exaggerated quantification of reality in the media (e.g. statistics and ranking indexes) is one of the side effects of the digital society that devalues the other human valences that must be part of the characterization of reality.

3.3 Identity Level

The identity dimension of AI refers to the new perception that human beings acquire of themselves due to the omnipresence of AI, characterized by overcoming their given nature and building new images of themselves, what is fairly evident at least in three essential aspects.

A first, that seems to be quite revolutionary, is the incursion of AI into the human spiritual dimension, its deepest intimacy, which has been considered throughout the history of humanity as constituting its unique specificity as well as its qualitative difference in relation to all the other beings. This incursion is manifest in its creative dimension, in its artistic expression replicated by the AI to compose music, paint canvases, write literature. For example, the first software to create music dates back to 1997, and today the composition of various musical styles by AI is widespread; in 2016, Microsoft developed a software using Artificial Intelligence that, through the

analysis of masterpieces from Rembrandt, managed to create a new painting with the same characteristics; since 2018, we started having books written by the AI.

A second aspect to highlight is the new power to build an alternative identity, external to the self but that tends to be taken as the truly self. It is a digital identity, fabricated with the collaboration of the AI, in simulated versions of the person such as avatars (entirely digital, cyberbody, an online identity) which allows each one to constantly and easily (effortlessly) reinvent themselves, to develop various personalities (change age, gender, etc.), establish different types of relationships according to the incarnated personality.

But the penetration of artificial intelligence into the essence of the human goes even deeper, as an internal construction of an enhanced identity, in the image and likeness of AI. There is a desideratum of cognitive evolution, through a process either of incorporation (e.g. cybernetic implants that enhance different human capacities) or of appropriation (brain-machine interface, like the one that Elon Musk's startup Neurolink is developing.[12] It would be about the creation of the post-human as advocated by the transhumanists.

The immediate usefulness for the human being is obvious under the new principle of self-improvement: not by developing what one is, but by acquiring what one is not; not by intensifying the authenticity of the being, but by distorting, perverting its own identity.

The perception that the human has of himself starts to reflect the presence of AI, also adopting it as a model, with immediate benefits, under the principle of human improvement. However, there are also unavoidable losses that must be simultaneously considered: violation of human identity values through the incursion into its spiritual dimension (its essence), namely the impossibility of forgetting (everything is indelible), which allows us to reinvent each day, in the atrophying of freedom, by the annulment of unpredictability and under the yoke of perfect decision, in the suppression of privacy, for the transparency of the total accessibility of lives; alienation of oneself, in digital simulacra of oneself, without density or authenticity; and usurpation of the self, in distorting improvements in human identity.

AI, in its identity level: presents itself united (fused) to all human expressions, determining them; invents a new identity in the image and likeness of the AI; and threatens to make the human succumb and replaces it with an improved self-image.

Still and always in the domain of a narrow or weak AI, we see how it intervenes on the functional level, in a superficial way, remaining outside the human and controllable by it, building a new, intelligent civilization through progressive automation; on the structural level, in a deep (pervasive) way in all human activities and relationships, integrating and shaping them, regulating them, constituting a new,

[12] The brain-machine interface is being attempt by a fusion or hybridization process that can increase intelligence and memory, erase bad memories and introduce good ones that never happened, or even to do a download of oneself to a digital support. In the long run it could conquer a digital immortality, surpassing the biological limits of humans.

virtual culture, through a growing connectivity; and on the identity level, in an intimate way at the heart of the human, uniting and reconfiguring it, dismissing it from itself in favor of an improved image, through a growing symbiosis.

4 How Should (Ethics)/Ought (Law) Humans and Machines Relate?

The public debate on human consequences of AI development begins in 2015, when 700 scientists sign a joint letter warning of AI threats: Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter (Future of Life 2015).[13] These scientists underline the extraordinary benefits that AI can bring to humanity, but also the risk of loss of human control and the need for more research to prevent any risks.

The biggest fear is that the neural networks will continue to develop, allowing AI to gain awareness (become strong or general), and then totally escaping human control.

In this context, it is worth mentioning that, in 2017, Facebook engineers were developing an experiment with robots that traded among themselves the ownership of virtual items. It was a conversational experience. After a few days, the robots had developed a language of their own which, as it escaped human comprehension, was interrupted, turning off the robots.

The evolution of humans and their identity throughout human history is recognized. A slow, gradual evolution resulting from adaptation to successive natural changes and induced by culture and new ways of life. But the current trend that was described above has implications for human identity that are relevant, rapid, disruptive and multidimensional.

The concern with this forced discontinuity of identity may be considered by some to be too conservative or pessimistic. Others accept that technology and its "consumption" are an acceptable manifestation of humanity's will in setting the path for its future. These are the very optimists or believers in an ever-better future based on technology. It will be up to everyone in the present to contribute to that future in a responsible way that respects the human heritage received, entrusted to us. If there are benefits and harms to point out now to the AI, the imperative to maximize the former and eliminate the latter is quite obvious.

The global strategy for this consideration has been to establish an ethical-legal regulatory framework, not with the intention of limiting the development of Artificial Intelligence, but rather legitimizing it through the promotion of its real benefits and prevention of its potential harm, framing it in the values and principles of identity of humanity and protecting human rights.

[13] This letter was a turning point for public opinion: citizens gained information, got involved and started also to be asked to intervene in decision-making processes.

4.1 Ethical Requirements

Ethical reflection must always precede legal regulations. In democratic and pluralist societies, it is important first to pay attention to their identity values and build an inclusive and broad ethical consensus, as a legitimizing basis for the legal regulations to be formulated later by Law. The Law reinforces the ethical consensus formerly reached, and Ethics contributes to an effective and robust regulatory process. Also with regard to AI, whether as a human production or because of its strong impact on the lives of people and societies, it was the ethical reflection that first developed as the disruptive social capacity of AI became more obvious.

Ethics of artificial intelligence gains particular prominence and has greater social impact when carried out by major international entities, highly representative of citizens, or by international and multidisciplinary working groups, joining different approaches, created specifically to outline guidelines that are considered to be convenient and necessary to ensure that the evolution of AI remains subordinate to human goals.

Thus, and particularly in the European context, the European Commission, the European Parliament and the Council of Europe have been working actively in this area: the Commission has established a High-level expert group on artificial intelligence, in 2018; the Parliament set up a special committee on artificial intelligence in a Digital Age (AIDA), in 2020; and the Council of Europe established an Ad hoc Committee on Artificial Intelligence (CAHAI), in 2019.

At the same time, we highlight the creation of several scientific groups on AI, such as the European Center of Excellence on the regulation of Robotics and AI, the European AI Alliance, the Expert Group on Responsibility and New Technologies, the Global Partnership on Artificial Intelligence (GPAI), to mention just a few. At the global level, UNESCO has established an Ad-hoc Expert Group on the Ethics of Artificial Intelligence.

All these bodies converge in declaring the urgency of AI regulation, in requiring its ethical foundation, being also evidence a broad convergence with regard to the identification of the main ethical principles to comply with, while respecting Human Rights.

A study from the Berkman Klein Center for Internet & Society at Harvard University, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, authored by Jessica Fjeld and colleagues (Jessica et al. 2020), gathered, in 2020, the 36 most outstanding documents on regulatory ethical principles and governance, presenting a set of eight principles as the most consensual. Privacy is one of most frequent principle, demanding respect for individual privacy, "both in the use of data for the development of technological systems and by providing impacted people with agency over their data". Accountability, concerning the impacts produced together with the provision of adequate remedies, is also a common requirement. Safety and Security of AI are of major importance in what relates to its performance as designed, and its resistance to invasions. A fourth group of principles is Transparency and

Explainability demanding for intelligibility and openness of processes, outcomes, and uses. Fairness and Non-discrimination claim for AI systems to be inclusive and to promote global justice, being required in all documents analyzed. Human Control of Technology is a major concern demanding that all important decisions be under human scrutiny. Professional Responsibility calls for individuals engaged in the development of AI to be able to predict the consequences of their deeds. Finally, Promotion of Human Values states that AI should improve the humanity's well-being. Sometimes under different designations these are, indeed, the prevailing guidelines in ethical reflection on artificial intelligence and which must be guaranteed by law.

4.2 Law and Legal Procedures

Ethical requirements are very important but are not enough to prevent AI adverse effects on fundamental rights because the ethics guidelines have no binding legal force. So, trustworthy AI need to be also lawful—as we stressed before.

The implementation of a legal framework adapted to the specific characteristics of AI systems is not easy. In addition to the technical complexities and rapid developing of these systems, there are other relevant difficulties or resistances. Firstly, new technological developments have a growing geo-strategic and military importance for the world's major economic and technological powers. Secondly, there is a strong pressure from governments and companies to achieve competitiveness increases driven by advanced and daring products in the market. A third difficulty is the demand of academic institutions and AI specialists to minimize legal limitations in applications and data collection. And, finally, there is a need for regulation at the planetary level in order to be completely effective. There is thus a tension in the ethical-legal front of AI regulation and an attempt to achieve balances between political decisions and the different interests involved. In this context, the affirmation of ethical-legal perspectives can be difficult in high-level decisions.[14]

It is a long and not always consensual process and we must know how we want technology to be applied (or not be applied) for the good of human society. The feasibility and potential elements of a legal framework for the development of artificial intelligence, based on the Council of Europe standards and the rule of law, are presented in a report (EU (a) 2021) of the Committee on Artificial Intelligence (CAHAI).The following options are presented: to amend binding legal instruments and adapt them to AI systems, modernising existing instruments or protocols or the adoption of new binding legal instruments.

[14] The High-Level Expert Group of the European Comission brought together 52 experts: 27 from industry, 15 from academia (3 with a legal background and 3 with an ethical background), 6 from the civil society and 4 from governmental bodies.

The issues to be discussed regarding legal proceedings for AI can be of three types. The first group comprises the security and defense of citizens' rights to compensation for damages and the control that AI systems comply with the law and do not violate established rights. A second includes how to define and assess accountability for the acts of artificial entities equipped with AI and autonomous learning and decision capacity? Should they have the same rights and duties as natural persons and be sued or punishable? Or should the responsibility pass to the creators or users of the system? Finally, the third concerns the use of AI by agents of justice in the application of the law and the obedience to ethical requirements. A good overview of these legal issues can be found in a text by Dempsey (2020).

Nowadays national legislation for AI framing is still very scarce around the world and AI systems are lightly regulated. There are, however, a number of international legal instruments that deal with certain aspects pertaining to AI systems. The greatest effort in this direction is taking place in the European Union (EU). One of the results of this effort is the General Data Protection Regulation (EU (b) 2016) (GDPR) that entered into force on 25 May 2018 (EU (b) (EU (d) 2018)) and try to concretise the fundamental right to personal data protection. GDPR fixes general and specific rules applying to sensitive categories of personal data such as health data and introduced a single legal framework across the EU with provisions allowing EU member states to enact national legislation specifying, restricting, or expanding some requirements. Administrative fines and penalties are considered. There is also a special research regime which provides flexibilities for scientific and statistical research.

Another UE initiative is the Proposal for Harmonised Rules on Artificial Intelligence or AI Act (EU (c) 2021). This proposed legislation classifies AI systems as high-risk (or not) based upon intended use. High-risk systems (e.g. remote biometric identification, evaluation of creditworthiness and credit scoring, judicial decision-making and recruitment and other employment decisions) would have to demonstrate compliance through conformity assessments before introduction into the market and certain uses of AI would be prohibited altogether. This risk classification does not include the precise assessment of the human or social damage and its respective probability. It thus seems difficult an adaptation of this regulation to the dynamic evolution of the market and of new AI products.

The use of AI in the judicial systems is a very relevant topic for its symbolic aspect. The way justice incorporates efficiency criteria using AI products must be exemplary. An in-depth study on the use of AI applications in judicial systems is presented in the Appendix of the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment (EU (d) 2018), but some issues can be highlighted. The risk of slipping into a position of immediate acceptance of decisions by artificial entities supposedly endowed with exceptional powers, but unpredictable and without explaining how and why they decide, is one of them. This idea permeates many analyses of predictive justice that lend these devices immediate or future capabilities to better predict human acts or to know the truth. This predictive justice cannot reflect the full reasoning of the human judge. An evolution that needs to be regulated through a permanent critical analysis because

Law has been and must continue to be a human activity supported by technology but never subordinated to it.

5 Concluding Remarks

Recovering our starting point, AI is a human production that should neither be idolized nor demonized, but rather evaluated with a critical spirit, both in its benefits and risks for the preservation of humanity as such and at the service of its development.

It is in this context that we highlight some key aspects to bear in mind in the present and future debates on AI:

- the application of new technologies with characteristics that surpass those of humans and with autonomous capabilities may lead to changes in social values and in legal procedures and concepts. However, human actions should not be submitted to judging criteria appropriate only to artificial beings with superior specific capacities or an indeterminate decision process;
- there is a risk of a progressive devaluation and decay of human capacities rather than a greater human behavioral and cultural development of society. The announced society of freer knowledge can slide to a more regulated society, complying with the rules imposed by a technology without limits of innovation with the justification of the optimization of rationality and efficiency. The meaning of life would tend to be reduced to the enjoyment of technological products and submission to decisions arising from AI algorithms;
- the education of new generations can constitute the path for a more adequate evolution of society and to avoid Stephen Hawking's prophecy: "the end of the human race". A society that knows how to reflect on the essential values and meaning of life and that enjoys them fully but in a sober way. One of the means for a more adequate education and preparation is perhaps the multidisciplinarity in academic training, avoiding a tight specialization and providing a better view to the different perspectives of reality and the human society;
- it is an illusion to believe that technology only solves problems and satisfies desires. It also creates new problems, eventually with severe and irreversible social damage. Human intermediation and accountability for autonomous acts of AI digital systems is a fundamental protection process for humanity.

Having addressed some ethical issues and underlining the need to build a broad ethical consensus as the foundation of the legislative initiative, we have also pointed out some guidelines for legal initiatives in this realm. The harshest challenge lays

probably at the political level, aiming the establishment of global governance in the field of AI.[15]

References

Brynjolfsson E, Rock D, Syverson C (2017) Artificial intelligence and the modern productivity paradox: a clash of expectations and statistics, working paper 24001. National Bureau of Economic Research, Cambridge, p 44

Dempsey JX (2020) Artificial intelligence. An introduction to the legal, policy and ethical issues. Berkeley Center for Law & Technology, Berkeley, p 46

EU (a) (2021) A legal framework for AI systems. Feasibility study of a legal framework for the development, design and application of artificial intelligence, based on Council of Europe's standards on human rights, democracy and the rule of law. Council of Europe Study DGI (2021)04. https://edoc.coe.int/en/artificial-intelligence/9648-a-legal-framework-for-ai-systems.html

EU (b) (2016) REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

EU (c) (2021) Commission proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM (2021) 206 final (April 21, 2021). https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF

EU (d) (2018) European ethical charter on the use of artificial intelligence in judicial systems and their Environment. European Commission for the Efficiency of Justice (CEPEJ), France. https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c

European Commission (2019) A definition of AI: main capabilities and scientific discipline. The high-level expert group on artificial intelligence. https://42.cx/wp-content/uploads/2020/04/AI-Definition-EU.pdf

Future of Life Institute (2015) Research priorities for robust and beneficial artificial intelligence: an open letter. https://futureoflife.org/ai-open-letter/

[15] *See*, generally, on the of the imitation of humans by Robots, in this book M N Duffourc and D S Giovanniello—The Autonomous AI Physician: Medical Ethics and Legal Liability; B A Ribeiro, H Coelho, A E Ferreira and J Branquinho—Metacognition, Accountability and Legal Personhood of AI. *See,* also, on Ethics, in this book P U Lima and A Paiva—Autonomous and Intelligent Robots: Social, Legal and Ethical Issues; A T Freitas—Data-driven approaches in healthcare: challenges and emerging trends; E Magrani and P G F Silva—The Ethical and Legal Challenges of Recommender Systems Driven by Artificial Intelligence; M S Fernandes and J R Goldim—Artificial Intelligence and Decision Making in Health: Risks and Opportunities; M N Duffourc and D S Giovanniello—The Autonomous AI Physician: Medical Ethics and Legal Liability; R Nogaroli and J L M Faleiros Júnior—Ethical challenges of artificial intelligence in medicine and the triple semantic dimensions of algorithmic opacity with its repercussions to patient consent and medical liability; and B A Ribeiro, H Coelho, A E Ferreira and J Branquinho—Metacognition, Accountability and Legal Personhood of AI.

Harari YN (2018) Why technology favors Tyranny. The Atlantic, October 2018 Issue. https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330/

Jessica F, Achten N, Hilligoss H, Nagy A, Srikumar M (2020) Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center for Internet & Society, p 71

Jonsson O, de Tena CL (2019) IE University's European tech insights. IE Center for the Governance of Change, Madrid. https://docs.ie.edu/cgc/European-Tech-Insights-2019.pdf

Julia L (2019) L' Intelligence artificielle n'existe pas. First, Paris, p 287

Kharlamova G, Stavytskyy A, Zarotiadis G (2018) The impact of technological changes on income inequality: the EU states case study. J Int Stud 11(2):76–94. https://www.jois.eu/files/6_478_Kharlamova%20et%20al.pdf

Kissinger HA, Schmidt E, Huttenlocher D (2021) The age of AI: and our human future. Little, Brown and Company, Boston, p 272

Liu S (2021) Artificial intelligence software market revenue worldwide 2018–2025, Dec. 8. https://www.statista.com/statistics/607716/worldwide-artificial-intelligence-market-revenues/

Qureshi Z (2021) Technology, growth and inequality. Changing dynamics in the digital era. Global economy and development (at Brookings), Working Paper 152, p 24

Searle JR (1980) Minds, brains, and programs. Behav Brain Sci 3:417–424. https://www.law.upenn.edu/live/files/3413-searle-j-minds-brains-and-programs-1980pdf

Tavani HT (2016) Ethics and technology: controversies, questions, and strategies for ethical computing. Rivier University, Wiley, Nashua, p 400

Turing A (1950) Computing machinery and intelligence. Mind New Ser 59(236):433–460. https://phil415.pbworks.com/f/TuringComputing.pdf