MARIA STEFANIA CATALETA[*]

HUMAN ALGORITHMS AND ARTIFICIAL HUMAN MINDS: THE HUMANIZATION OF AI IN AN EUROPEAN PERSPECTIVE

SUMMARY: 1. The primacy of artificial brain over human brain? – 2. The consciusness, free will and morality of machines. – 3. Personal data and digital identity. – 4. The ethics and responsibility of algorithms. – 5. Conclusions.

1. *The primacy of artificial brain over human brain?*

John McCarthy, one of the fathers of Artificial Intelligence (AI), was convinced that computers could simulate many of the cognitive functions of human beings and thus the expression AI was invented, distinguishing it from simple automation [1].

Essentially, the result of an operation performed by an intelligent system is not distinguishable from an operation carried out by a human[2]. In other words, that linked to AI is a discipline that studies the design, the development and the realization of systems capable of simulating human ability, reasoning and behavior.

Moreover, it is arduous to commensurate human intelligence with artificial intelligence, neither the speed of calculation can be only an indicator of the fact that machines possess a superior intelligence. Human capacities, in other words, are not a good meter of judgment for AI, mainly if you can consider that there are activities of AI that are precluded from human intelligence and that notwithstanding express intelligence. A tsunami alert system is not comparable with human abilities, because it is based of minimum movements of ocean heights imperceptible by human senses, those incapable of perceiving the submarine geological upheavals. However, the study of the human mind is still concentrated on understanding how it is possible that neurons, masses of homogeneous cells, through their interconnections (the synapsis) can change electrical or

---

[*] Associate researcher at Université Côte d'Azur.
[1] According to the McCarthy's definition, "intelligent" is any system capable of performing actions that would be qualified as intelligent if they were accomplished by a human being. *See* J.-L. CHABERT et al., *A History of Algorithms: From the Pebble to the Microchip*, Springer, 2013, p. 2.
[2] European Commission for the Efficiency of Justice (CEPEJ), *European ethical Charter on the use of the Artificial Intelligence in judicial systems and their environment*, p. 69.

chemical signals and by virtue of this, perform the most varied activities.

Something similar happens in the research of artificial neural networks, where one can search to understand how to make interconnections converge toward the most acceptable solution in a reasonable time. Artificial neural networks, more than learning, seem more like imitators of strategies taken from a large number of examples. AI learns by doing by way of machine learning, of which one of the principal aspects is deep learning. Deep learning refers to the use of artificial neural networks with many inner layers, called "hidden layers". Deep learning is a system of learning and classification that, across networks of artificial digital neurons, allows a computer to acquire some capacities of the human brain. The artificial neural network seems to be closer to the human nervous system. In the MLP-Multi-Layers Perceptron networks there are hidden neural layers where every neuron in a level is linked to all neurons of the immediately preceding level and of the level right after. The real power of the algorithm is given by the capacity to train the neural network and to allow it to gain experience.

AI was born from deep learning; its name was actually coined in reference to "smart" calculations, similar to those of a powerful calculator, capabilities much less complex and intelligent than their current uses in recognizing the contents of images or understanding spoken language[3].

Many problems that at first sight would seem to require logic and reasoning can be resolved through machine learning. Self-learning of computer has an exponential tendency, they improve constantly and refine output by accumulating experiences. Machine learning is defined as the mother of all the algorithms in AI, with all its variables of self-learning. Firstly, there is supervised learning: data associated with information that interests us is given to the algorithm, on the basis that the algorithm will learn how to understand and how to behave (an example is the classification of potential clients according to the profile and history of other customers' purchases). Then, there is unsupervised-learning, where the algorithm is more complex because it needs to extract yet unknown information from the data, or further reinforce its own learning, in which the algorithm has a goal to reach and thus auto-defines a way in which it behaves which can change in the face of differing situations.

Scholar Jerry Kaplan believes that a computer can be more intelligent than a human being, even if limited in manner[4]. It is true that computers, in a wide range of intellectual duties, are superior to man, but this does not necessarily mean that will dominate us.

Furthermore, there is a question which scholars try to answer when investigating the primacy of machines over man: if they are equipped with a mind and thoughts. In this field, two theories face each other, that of "strong" AI, referring to machines which have a mind or in any case will end up having one someday, and that of "weak" AI, which considers that machine realities are only simple simulations and not a duplication of real intelligence. The conceptual crossroads is between the possibility that machines can be

---

[3] *See generally* L. ALEXANDRE, *La guerra delle intelligenze. Intelligenza artificiale contro intelligenza umana*, EDI, 2017, p. 10 et seq.; P. MIKALEF, M. GUPTA, *Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance,* in *Information & Management,* Volume 58, Issue 3, April 2021, pp. 1-20; R. SCHMIDT, A. ZIMMERMANN, M. MÖHRING, B. KELLER, *Value Creation in Connectionist Artificial Intelligence-A Research Agenda*, in *AMCIS*, 2020; S. RUSSEL, P. NORVIG, *Artificial Intelligence: A Modern Approach*, Pearson, London, 2016; P. WANG, *On defining artificial intelligence,* in *Journal Artificial General Intelligence*, 2019, pp. 1-37.

[4] *See* J. KAPLAN, *Intelligenza artificiale. Guida al futuro prossimo,* Second edition, Luiss University Press, 2018, p. 30 et seq.

truly intelligent or simply capable of acting like they are[5].

According to the distinction made by American philosopher John Searle, weak AI acts and thinks as if it had a brain, but it is not intelligent; it limits itself to emulating the human brain. To offer the best answer to a problem it investigates similar cases, it studies them, and chooses the most rational solution. Weak AI does not understand all the human cognitive processes but only deals with problem solving; it answers problems on the basis of known rules. By contrast, strong AI has cognitive capacities indistinguishable from human capacities, but according to Searle, we are still far from this reality.[6]

"Expert systems" are highlighted in this context: software that reproduces performance and the knowledge of experts in certain fields. The inferential engine is at the centre of these systems. In other words, it deals with an algorithm which, similarly to how the human mind works, starts from a proposition whose truth derives from the content of the first proposition, according to deductive or inductive logic.

Conforming to the Turing test, thought would be extended to the machines and this would have happened around the end of the XX century, whereas skeptics, like John Searle, sustain that machines cannot think at all, because it is an exclusively human activity, that computers are limited in simulating.[7] Nevertheless, the actual capacities of machines put their beliefs on the superiority of the human mind to the test.

Kaplan analyses the issue from the perspective of free will and of the capacity of machines (following anthropomorphic criteria) to make decisions, in a similar way to humans in their decision making, machines in their decision making would be capable of applying knowledge and be competent in assuming risks and modifying plans on the basis of added information by using analogies to solve concrete cases. No reason has been found to believe that man and machine follow different principles regarding decisional processing[8].

2. *The consciousness, free will and morality of machines*

Technological innovations need to be protected by some distortion. That is why AI applications demand regulation also on ethical implications, typical of humans, which derive from their use. Indeed, they have a certain level of "self-determination", affect many sectors, are often very delicate and also involve sensitive data[9].

According to Bertrand Russel, if you want to verify that consciousness is in the brain, you just need to shoot someone in the head and you will have the confirmation, because the shot will eliminate the consciousness for a while or forever. Our mind and consciousness are, in a certain sense, physical, but they do not coincide with our body.

---

[5] *Id.* pp. 104-105.

[6] J. R. SEARLE, *Minds, Brains, and Programs*, in *Behavioral and Brain Science,* Volume 3, Issue 3, September 1980, pp. 417-424.

[7] *Id.*

[8] J. KAPLAN, *supra* note 4, pp. 119-121; M.-S. CATALETA, A. CATALETA, *Artificial Intelligence and Human Rifights: un Unequal Struggle*, in *CIFILE Journal of International Law*, Volume 1, N. 2, 2020.

[9] S. RUSSEL, P. NORVIG, *Artificial Intelligence: a modern approach*, 2 ed., Upper Saddle River, NJ, Prentice Hall, 1995; *see also,* M. CRAGLIA et al, *Artificial intelligence: a European Perspective,* European Commission, Publication Office, Luxembourg, 2018, p. 63 et seq., https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/artificial-intelligence-european-perspective (access 26 August 2020); N.-A. SMUHA, *Beyond the individual: governing AI's societal harm,* in *Internet Policy Review*, 2021, https://doi.org/10.14763/2021.3.1574.

Consciousness and mind are in the brain and science does not find them elsewhere. Men are self-aware thanks to the human capacity for self-reflection. In this optic, machines are not self-aware, since they are not capable of self-reflection. The nature of consciousness will remain inexplicable and, until now, unattainable by means of technology. Consciousness is a limit for machines, that distinguishes them from humans.

However, machines and humans are strictly connected, like materiality and immateriality. You can find the relationship between materiality and immateriality in many things in the world. Material objects can have an immaterial value. In the virtual world, technological devices have a material value because they are objects with a cost and utility, but they also have an immaterial value. The most important immaterial value is represented by the amount of data that all the digital world encloses. Producing, storing and transferring material objects sometimes takes a huge cost and a lot of time, the same is not true for immaterial objects. Producing, storing and transferring data is simple and cheap. Every time we put data online, someone, who controls them, can store and use them to create individual and personalized offers and messages. This is, for example, what Cambridge Analytica does through their profiling techniques[10]. This can be done for beneficial or malicious purposes.

In some cases, the profiling techniques have been used to manipulate the will of the users, because the personal data of millions of people, such as sex, age, cultural or political preferences, can be illegally stored by Facebook or other platforms and analysed in order to create some micro-categories. Then these data, thanks to the algorithms, can be used to create many individualized propaganda messages so as to influence the will or the choices. This does not happen just for e-commerce, but also for civil and political rights. [11]

The control over our data is the antidote against personal data exploitation[12]. Human control mitigates the power of algorithms. Two different kinds of intelligence are in conflict. In the digital era, Regulation 679/16, the General Data Protection Regulation (GDPR) represents a limit to the interference of AI to our rights to privacy in Europe.

However, the protection of personal data is not the only problem concerning AI towards humans. Decisional algorithms governing our lives must possess some typical human qualities: they must be transparent, fair and respectful of the founding values of the society and users. Only in this way can AI be beneficial for society.

One can ask if AI, like human beings, can be moral, hence the fact that it can independently think and decide. At the same time, one can ask Kant's question if, similarly for man, an intelligent system can be considered a means or an end. Furthermore, one can ask if intelligent systems are a luxury good or a necessary good[13]. Finally, one can ask if the risks that AI provides can be afforded because benefits are greater or if such balancing between costs and advantages is unfair. It is common knowledge that computers do just what they are programmed to do, lacking free will, for this they would be deprived of moral sense.

---

[10] *See* the Federal Trade Commission Decision, December 18, 2019, File Number: 182 3107.

[11] X. DU, S. DUA, *Data mining and machine learning in cybersecurity*, Auerbach Publications, 2011; N. KÜHL, M. GOUTIER, R. HIRT, G. SATZGER, *Machine Learning in Artificial Intelligence: Towards a Common Understanding*, in *Arxiv*, Cornell University, 2020, https://arxiv.org/abs/2004.04686.

[12] S. QUINTARELLI, *Capitalismo Immateriale – Le tecnologie digitali ed il nuovo conflitto sociale*, Bollati Boringhieri, Turin, 2019.

[13] M.-R. LACHAT, *Artificial Intelligence and Ethics: An Exercise in the Moral Imagination*, in *AI Magazine*, Volume 7, Number 2, 1986, p. 71 et seq.

It is not possible to trivialize the matter about the morality of certain algorithmic automated decisions, because the concept of morality is strictly linked to that of logic and reasoning. The emotional factor comes into play in morality, but unlike for logic and reasoning, the problem is to find an emotional capacity in autonomous decisional intelligent systems. The debate is interesting in order to give a human-centric character to machines and digital technologies *tout court*.

In the decisional process of a human judge many qualities are involved, such as: the knowledge of all relevant facts; empathy; the lack of prejudice; the absence of emotional involvement[14]. All these are features that are the assumption for a valid judgment from a moral point of view. Except for empathy, difficult but maybe not impossible for machines, the other features can be attached to an automated decisional process.

Can the lack of empathy, however, be sufficient enough to exclude the moral value of a decision when it has been taken by a human being? In other words, can an automated decision, by definition cold and aloof, be less valid than a human decision just because this last is equipped with empathy? In this case, the automated decision stands out from human decisions minimally. Can an individual trust a decision coming from algorithms in the same way or more than a decision coming from a human being?

The spread of AI systems raises concerns on how to mitigate the impact of machine intrusiveness, enabled with decisional capabilities, within our society. AI will create autonomous technologies with behavioural rules based on data analysis, which is based on statistics not morality. Furthermore, AI is increasingly able, often more than a human being, to do complex duties and to solve articulated problems by choosing between several possible alternatives. AI seems bound to substitute not only manual labour but also intelligent labour[15].

It is possible to point out that machine decision capabilities almost equal and duplicate human capability and this involves moral questions, because there are many ethical implications in automated decisions. Moral and ethical aspects are not antithetical and stranger for AI systems[16].

There are some principles to be respected in the decision-making process assisted by AI, so that it is transparent and has a positive impact on the individual and society. It is important to give information on how the risks to produce negative effects have been reduced and on the other hand, how the possible positive effects have been promoted. In this process, the protection of personal data is important, it must be imprinted to legality, transparency and equity principles. In this way the individual will be helped to understand in which moment, in the decision-making process, AI intervened, so as to understand it[17].

There are different ways to facilitate the understandability of the decisions coming from AI, which are mainly the following. Firstly, it is important to give the rational

---

[14] R. FIRTH, *Ethical absolutism and the ideal observer*, in *Philosophy and Phenomenological Research,* Volume 12, Number 3, 1952, pp. 317-345, in M.-R. LACHAT, *id.,* p. 77; G. AIRENTI, *The cognitive bases of anthropomorphism: from relatedness to empathy,* in *International Journal of Social Robotics,* 14 January 2015, pp. 117-127.

[15] D. BENNATO, *Quale futuro per l'umano, nel trionfo dell'intelligenza artificiale*, in *Cultura digitale*, 2019 (March 12, 2019, 10:30), https://www.agendadigitale.eu/cultura-digitale/quale-futuro-per-lumano-nel-trionfo-dellintelligenza-artificiale/.

[16] M.-S. CATALETA, *Artificial intelligence vs human intelligence*, in L. MIRAUT MARTIN/M. ZALUCKI (eds.), *Artificial intelligence and Human Rights*, Dykinson Eboook, AFM Krakow University, 2021, p. 107 et seq.

[17] G. NOTO LA DIEGA, *Against the Dehumanisation of Decision-Making*, in *Jipitec 3*, 2018; D. MAURI, *Algorithmic Target Construction' and the Challenges by International Human Rights Law*, in *Big Data and Public Law: new challenges beyond data protection*, University of Milan, Eurojus, 15-17 October 2018.

explanation, preferably not exclusively technical, for the reasons at the origin of such a decision; secondly, a responsibility explanation is necessary, that is the information on the individual responsible for implementation of the AI system involved, with the information on the individual to ask for the revision of the automated decision; a data explanation is also necessary, that is the explanation on which data was used and how; the equity explanation, so that there is some reassurance that the automated decisions do not produce bias, discriminations injustices for the individual; a safety and performance explanation is also demanded on the step faced by the decision taken by the AI system, in particular if it is as precise as possible, reliably and sure; furthermore, an impact explanation is demanded on the automated decisions' negative, positive or neutral effects towards the individual and society[18].

With reference to this procedure, there are two subcategories of explanations: the "*process-based explanations of AI systems*", with the aim to reassure that measures of equity and certainty have been adopted to avoid any risk of bias or automated discriminations in the data processing; and the "*outcome-based explanations of AI systems*", with the aim to give easy understanding of the results obtained and of the logical processes at the bases of an automated decision.[19] It is also important to ascertain the equity, safety and fairness adopted even when a human support has intervened in the decision coming from an AI system, explaining also when and how the human contributor has intervened.[20].

All the above-mentioned explanations generate reliability and safety, which are essential elements for an AI decision-support system, of which it is important to know all potential beneficial and harmful effects, for both the individual and society. Not explaining how AI decisional processes work even when a human support intervenes means to generate distrust and impedes the individual concerned from defending himself from any inappropriate use of his data[21].

3. *Personal data and digital identity*

Thus, information placed on the network, when processed by AI, produces data, which have a considerable commercial value the more they are shared. Every second, billions of internet users give web giants huge amount of personal data, transmitted over social networks, equating to an annual market value of one trillion dollars. According to the Metcalfe law, the value of a network grows exponentially in relation to the number of users.[22] The problem that arises concerns the protection of this patrimony of data, mainly

---

[18] S. MARANELLA, *La protezione dei dati personali contro un uso distopico dell'AI*, in R. GIORDANO, A. PANZAROLA, A. POLICE, S. PREZIOSI, M. PROTO, *Il diritto nell'era digitale. Persona, Mercato, Amministrazione Giustizia,* Giuffrè, Milano, 2022, p. 52-56.

[19] S. CHESTERMAN, *Artificial Intelligence and the Problem of Autonomy*, in *Notre Dame Journal on Emerging Technologies*, Vol. 1, 2019, p. 3.

[20] V. *Explaining decisions made with AI*, The Alan Turing Institute, 20 May 2020, p. 20 ss.; W. SAMEK, K.-R. MÜLLER, *Explainable AI: interpreting, explaining and visualizing deep learning. Towards explainable artificial intelligence*, Springer, 2019 , pp. 5-22.

[21] S. MARANELLA, *supra* note 18.

[22] A. ODLYZKO, B. TILLY, *A refutation of Metcalfe's Law and a better estimate for the value of networks and networks interconnections*, Digital Technology Center, University of Minnesota, Minneapolis, 2005; I. GRAEF, J. PRÜFER, *Governance of data sharing: a law & economics proposal,* in *Research Policy*, Volume 50, Issue 9, November 2021; F. SCHACKL, N. LINK, H. HOEHLE, *Antecedents and consequences of data breaches: a systematic review*, in *Information & Management,* 2022.

personal data because these data are one of the topics most involved on a daily basis in the arrival of AI systems. While in the past, the aim was to protect the private citizen from State interference and abuse of power, protection against the misuse of personal information at the present calls into question the role of individuals who often offer their own personal data voluntarily to private companies in exchange for advantages. Internet users, in fact, make possible, willingly or otherwise, the reconstruction of their own individual profile through cookies, geotracking, and consent to the sale (or sometimes fraudulent acquisition) of their own data.

Scholars have pointed out that human rights lose their meaning, in the case of privacy, where their use can be traded like any other commodity in exchange for money or other advantages. The free sale of privacy ends up allowing totalitarian control by those who manage this information to learn about, to pilot and to guide, through statistical analysis, the personal choices of the same users in exchange for utility. This information is given to the "public web record" on the precondition of democratic participation in online life. In this way, the logic underlying human rights would be reversed, as they would be invoked to protect individual choices as an expression of freedom. However, were these companies to acquire the domain over this personal information, these rights would immediately end up being manipulated by power centres.

The algorithmic analysis and elaboration of big data make it possible to trace the identity (not exclusively the digital but also the personal identity) of each individual. The right to identity is the right to be himself, that is the entirety of opinions, habits, tastes, preferences, personal and intellectual characteristics, creeds and experiences. Not only is the identification of a person (that is his physical data, such as address, face or name) possible online but also the revelation of sexual or religious orientations and political ideas. Digital identity oversteps the physical dimension of an individual. Thanks to the huge amount of data, through profiling techniques, AI can not only know the preferences of an individual but can also orient them, foreseeing them[23]. In respect to revelation of his identity the individual is placed in a situation of greater vulnerability. Human control techniques over AI-based treatment are in question and in this regard, law struggles to keep up with technological progress[24].

At the European level, personal data are ruled by art. 4 of the Regulation 679/16 (GDPR-General Data Protection Regulation), which intends personal data and also information that indirectly can lead back to the identification of a person.[25] Definitely, the aim pursued by the GDPR is definitely to protect identity, also through anonymization, from abuse and theft, while, with the aim to limit the pervasiveness of AI with respect to biometric identification[26], the EU Proposal for a Regulation on AI was presented by the European Commission, which has been added to the Cybersecurity Act[27]. The European

---

[23] E.-C. RAFFIOTTA, M. BARONI, *Intelligenza artificiale, strumenti di identificazione e tutela dell'identità,* in *BioLaw Journal,* n. 1/2022, p. 168; E. PARISER, *The Filter Bubble: What The Internet is Hiding From You*, London, 2012.

[24] G. MOBILIO, *L'intelligenza artificiale e i rischi di una "disruption" della regolamentazione giuridica,* in *BioLaw Journal,* n. 2/2020; B. DEMBROW, *Investigating in human futures: how big tech and social media giants abuse privacy and manipulate consumerism,* in *University of Miami Business Law Review*, Volume 30, Issue 3, 2022, pp. 324-349.

[25] CGUE C-434/16, *Peter Nowak v Data Protection Commissioner*, 20 December 2017, where the Court stated that « *the written answers submitted by a candidate at a professional examination and any comments made by an examiner with respect to those answers constitute personal data ».*

[26] Recital art. 3 (33) of EU Proposal of Regulation on AI.

[27] This Regulation was published on 7 July 2019 and was enacted on 27 July 2019. it explanes the new European strategy for cybersecurity against cyber-attacks.

protection of identity online, even thanks to the creation of some supervisory authorities with sanctioning powers, is more human centric and stronger than international regulation and any other national regulation coming from extra-European countries.

A strong form of protection comes from supervisory authorities, as the Tik Tok case demonstrated, in 2021, following the death of an Italian minor aged 10 years, where the data protection supervisor authority (GPDP-*Garante per la Protezione dei Dati Personali*) immediately intervened against the social network, imposing a block in the use of the users' personal data when their exact age – which must be not under 14 years - is not certain[28]. After such a measure, Tik Tok started to use AI-based age verification instruments to guarantee the protection of minor users' identity[29], in conformity to art. 3 of the UN Convention on the Rights of the Child and art. 24, par. 2, of the Charter of Fundamental Rights of the EU, that prescribes that *"the child's best interest must be a primary consideration"*, but also in conformity to recital 38 of the GDPR, that gives special protection to minors' personal data.

Facial recognition is also a technique where the revelation of identity[30], through the use of biometric data, is in question. In Italy, it is banned until December 2023 whilst waiting for a specific regulation[31]. Through a minimal movement of our eyes, facial recognition can not only reveal physical data but also emotional information. According to art. 3 (34) of the Proposal of European Regulation on AI, an emotional recognition system refers to *"an AI system for the purpose of identifying or interfering with the emotions or intentions of natural persons on the basis of their biometric data"*. It is a system capable of threatening users' fundamental rights. For this reason, it is demanded that AI be explicable and trustworthy[32].

The Committee of Ministers of European Union adopted, on 13 February 2019 at the 1337th meeting of the Ministers' Deputies, the Declaration on the manipulative capabilities of algorithmic processes. The Committee affirms that attention must be paid particularly to the capacity of digital technologies to use personal data and non-personal data to identify individual vulnerabilities and thus encourages member-States to assume

---

[28] GPDP, *Measure of 22 January 2021 [9524194]*, available online on: https://www.garanterivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9524194. This Measure was published the following day by media, GPDP, *Tik Tok: dopo il caso della bimba di Palermo, il Garante privacy dispone il blocco dei social*, available on: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9524224.

[29] D. TREHAN, *The inescapable AI algorithm: Tik Tok,* in *Towards Data Science*, 2020; N. MIYAKE, H. ISHIGURO, K. DAUTENHAHN, T. NOMURA, *Robots with children: practices for human-robot symbiosis, at the 6th* ACM/IEEE International Conference on Human-Robot Interacion, Lausanne, Switzerland, 2011.

[30] *See* the Indian Supreme Court judgment, *Justice K. S. Puttaswamy (Retd.) and Anr. Vs Union of India And Ors*, 24 August 2017.

[31] Law n. 205 of 03 December 2021, that passed into law the law decree n. 139 of 8 October 2021, laying down urgent provisions for access to cultural, sprotif and recreational activities, as well as for the organization of public administrations and in the field of personal data protection. *See* C.-M. BISHOP, *Pattern recognition and machine learning*, Springer, New York, 2006.

[32] E.-C. RAFFIOTTA, M. BARONI, *supra* note 22, p. 179; *see also,* M. TADDEO/T. MCCUTCHEON, L. FLORIDI, *Trusting Artificial Intelligence in Cybersecurity is a Double-Edged Sword,* in *Nature Machine Intelligence*, n. 1, 2019; H.Y. LIU, K. ZAWIESKA, *From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence,* in *Ethics Information Technology*, Volume 22, Issue 4, Dec. 2017, pp. 321-333; A. JOBIN, M. IENCA, E. VAYENA, *The global landscape of AI ethics guidlines,* in *Nature Machine Intelligence*, 2019, pp. 389-399; M. RYAN, B.-C. STAHL, *Artificial intelligence ethics guidlines for developers and users: clarifying their content and normative implications,* in *Journal of Information, Communication and Ethics in Society*, 3 March 2021, pp. 61-86; S. HAN, E. KELLY, S. NIKOU, E.-O. SVEE, *Aligning artificial intelligence with human values: reflections from a phenomenological perspective,* in *AI and Society*, 20 July 2021, https://doi.org/10.1007/s00146-021-01247-4.

their responsibilities in order to address this threat by adopting a number of measures, such as: initiating informed and inclusive public debates with a focus on providing guidance to define the difference between permissible persuasion and unacceptable manipulation; taking appropriate and proportionate measures to ensure that effective legal guarantees are in place against such forms of illegitimate interference and empowering users by promoting critical digital literacy skills, specifically, public awareness on the fact that algorithmic tools are widely used for commercial purposes and for political reasons, as well as for the wills of anti- or undemocratic processes, warfare, or direct harm.

4. *The ethics and responsibility of algorithms*

According to some scholars, it is not possible to talk of intelligence in relation to algorithms, *i.e.* not in a human sense. After 60 years from its eruption in the technological world, AI in the sense of the pure reproduction of human intelligence is still a dream. At the present, AI is still very far from being intelligent in a human sense. In reality, we are using the same name incorrectly to express two different things, this name is "intelligence", which would encompass both human and artificial capabilities. For these reasons, today, we can speak of General Artificial Intelligence (AGI) to mean the capacity of machines to reproduce human intelligence, whereas the present artificial intelligence would mean automated learning. Intelligent algorithms in a human sense does not exist yet, but the terminology creates confusion. Current AI consists of a bunch of computational systems that learn automatically through machine learning and take information from a huge amount of data. Also, neural layers typical of deep learning would not have anything of neural in a human sense.

Moreover, technological progress is evident. Today, machine learning algorithms can discover the presence of many cancers and in games can find several strategies to win in games. Evolution in AI concerns every aspect of human life where data play a crucial role. The future will be the creation of algorithms capable of learning automatically from data even where there is no previous human supervision on those data. Many risks can be glimpsed: algorithms can automatically reflect all the possible discriminations present in the data from which they learn; furthermore, it is always possible a dangerous use of autonomous new technologies like in modern warfare. Facing this uncontrolled evolution, the role of ethics is crucial, that is how AI applications are regulated. Uniformity in international regulation is necessary[33].

While the competition in this sector is between the USA and China, that thanks to their primacy also establish rules, Europe is trying to find its own space[34]. An example is the *European Laboratory for Learning and Intelligent Systems* (ELLIS).

Due to the fact that humanizing AI also involve introducing ethical principles into machines, the USA Department of Defense developed the Responsible AI Guidelines, which guarantee that ethical considerations are integrated into the design, training and organization of AI, with the aim to define a process that is scalable, responsible and reproducible. This document ensures also that the decisions of AI tools are coherent with human values.

The problem of ethical automated decisions has grabbed the attention and aroused

---

[33] M.-S. CATALETA, *The proposal of an EU Regulation on AI in a dystopian global scenario*, in *Cambridge Journal of International Law*, December 22, 2021.

[34] M.-S.CATALETA, *Diritti umani e algoritmi*, Nuova Editrice Universitaria, Rome, 2021, p. 77 et seq.

fears from scholars and technicians[35]. The biggest fear concerns systems capable of rivalling, if not, overcoming human capacities. The risk is that in the not-too-distant future a self-aware AGI can achieve super-human capabilities[36]. Even if the world is still far from AGI, autonomous systems already exist, such as drones, war robots and autonomous cars. At the present, the question is if in the future an AI system is able to manage any situation with no human support. Such a question is now more evident in the weapons industry, where many ethical and legal perplexities have arisen following the use of certain lethal autonomous weapons (LAWS).[37] Such autonomy equally poses the problem of how to incorporate ethical principles and considerations in these systems, in order to avoid ethical dilemmas, whereby different choices conflict with human rights.[38]

Thus, these intelligent systems must behave ethically. A project which has worked in this direction is the *Moral Machine project* of the *Massachusetts Institute of Technology* (MIT), that focused on the ethical dilemmas of automated cars in cases of accident in which human beings are involved and where several choices can enter in conflict (such as to save many lives, to save only the passengers' life or to distinguish on the base of the gender or age and so on).

One can distinguish between three dimensions of ethics: *Consequentialist ethics* (an agent is ethical if and only if he balances the consequences of every choice and chooses the option that is more moral); *Deontological ethics* (an agent is ethical only if he respects obligations, duties and rights linked to a certain situation: in other words, an agent gifted with deontological ethics acts in conformity to shared social norms); *Virtue ethics* (an agent is ethical only if he acts and thinks in accordance with moral values, such as courage, justice and so on)[39]. All these different kinds of ethical criteria must be incorporated into algorithms. In December 2018, a group of experts drew up the "Draft Ethics Guidelines for Trustworthy AI"[40]. With this document, the European Commission warned of the risks associated with AI, despite its considerable advantages, and recognised the need for an anthropocentric approach to AI. This is the only approach capable of guaranteeing the dignity and autonomy of people, who must always be given the power to supervise machines (*AI is human centric: AI should be developed, deployed and used with an "Ethical purpose" (...), grounded in and reflective of fundamental rights, societal values and the ethical principles of Beneficence (do good), Non-Maleficence (do not harm), Autonomy of humans, Justice, and Explicability*[41].

Another problem, besides that of ethics, is that of responsibility for damages caused by AI. When the damage has been done, can the intelligent system be chargeable for it? The most obvious answer is that the configurability of algorithms is arduous, not

---

[35] M. ANDERSON, S. L. ANDERSON, *GenEth: A General Ethical Dilemma Analyzer*, in *Journal of Behavioral Robotics*, Paladyn, Volume 9, Issue 1, 2018, pp. 253-261; R.C. ARKIN, *Ethics and autonomous systems: Perils and promises*, in *IEEE Xplore*, Volume 104, Issue 10, 2016.

[36] H. YU, Z. SHEN, C. MIAO, C. LEUN/V.-R. LESSER, Q. YANG, *Building Ethics into Artificial Intelligence*, arXiv:1812.02953v1 [cs.AI], 7 December 2018, p.1; N. BOSTROM, *Superintelligence: Paths, Dangers, Strategies*, Brilliance Publishing, 2015.

[37] G. RAMUNNO, *Etica dei sistemi d'arma autonomi. I robot che uccidono e il diritto umanitario*, in *Osservatorio di Politica Internazionale*, n. 22, November 2022.

[38] A. FENWICK, G. MOLNAR, *The importance of humanizing AI: using a behavioral lens to bridge the gaps between humans and machines*, in *Discover Artificial Intelligence*, Springer Link, Article number: 14, 22 August 2022.

[39] H. YU, Z. SHEN, C. MIAO, C. LEUN, V.-R. LESSER, Q. YANG, *supra* note 36.

[40] The European Commission's High-Level Expert Group on Artificial Intelligence, *Draft Ethics Guidelines for Trustworthy AI*, in http://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai.

[41] *Id.* p. 13.

necessarily robotic subjects, as criminal agents, which is why *machina delinquere (et puniri) non potest*[42]. Furthermore, it would be possible to recognize the vicarious responsibility of the users or programmer/builder. This is because machines lack the capacity of self determination or will? This means that machines do not show guilt. This is the peacefully dominant doctrine.

Nevertheless, the automated decisional capabilities of intelligent systems put in crisis this hired, because behind these decisional processes there is no more a man but the almost complete autonomy of certain algorithms due to the machine learning or cloud computing technologies or black box algorithms. A clear example of this can be represented by completely autonomous self-driving cars. In such cases, if one excludes the programmer's responsibility, it is possible to admit a lack of criminal protection for the possible damages[43].

Automated decisional processes can also concern not only robotics, such as combat drones not remotely guided by humans, but they can also involve algorithms that process personal data and that, through distorted and uncontrolled use, can cause offence to the person; in illicit uses of personal data by intelligent systems it is possible to find a gap in regulation and protection.

According to the distinction made by American philosopher John Searle, weak AI acts and thinks as if it had a brain, but it is not intelligent; it limits itself to emulating the human brain. To offer the best answer to a problem it investigates similar cases. It studies them and chooses the most rational solution. Weak AI does not understand all the human cognitive processes but only deals with problem solving. That is, it answers problems on the basis of known rules. By contrast, strong AI has cognitive capacities indistinguishable from human capacities, but according to Searle, we are still away from this reality.

According to another suggestive doctrine headed by Gabriel Hallevy, demands of social regulation and the overcoming of an anthropocentric prejudice against machines or artificial intelligent subjects could propel, like for corporate liability, for accountability of machines, given also their capacity for self-determination, self-representing the reality, willing and deciding, that could attribute them concepts as *actus reus*, *mens rea* and applicability of penalty[44].

It would be more arduous to attribute the same typical functions of penalty, whether retributive or preventive, in case of ascertained responsibility. Beyond doctrinal speculations, the lack of protection from the criminal law concerning the illicit consequences of the algorithmic automated decisions is indisputable, in cases in which it is impossible to configure the user or the programmer/builder responsibility.

During the works of the Conference "Governing the Game Changer-Impacts of

---

[42] A. CAPPELLINI, *Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Diritto penale e intelligenza artificiale*, 2019, p. 291 et seq.; *see also* D. LIMA, *Could AI Agents Be Held Criminally Liable? Artificial Intelligence and the Challenges for Criminal Law*, in *South Carolina Law Review*, 2018, p. 677; S. GLESS E. SILVERMAN T. WEIGEND, *If Robots Cause Harm, Who is to Blame? Self-driving Cars and Criminal Liability*, in *New Criminal Law Review*, 2016, p. 412 et seq.

[43] A. CAPPELLINI, *id.*, p. 293 et seq. and of the same author *Profili penalistici delle self-driving cars*, in *Diritto penale contemporaneo*, n.2, 2019, p. 325 et seq.; *see also* S. BECK, *Google, Cars, Software Agents, Autonomous Weapons Systems – New Challenges for Criminal Law*, in E. HILGENDORF, U. SEIDE (eds.), *Robotics, Autonomics, and the Law*, Baden-Baden, 2017 p. 229 et seq.

[44] G. HALLEVY, *Liability for Crimes Involving Artificial Intelligence Systems,* Springer, 2016 and of the same author, *The Criminal Liability of Artificial Intelligence Entities – from Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal*, 2010, p. 171 et seq.; *see also* A. CAPPELLINI, *supra* note 42, p. 300.

artificial intelligence development on human rights, democracy and the rule of law", held in Helsinki in 2019 and organized by the Council of Europe and by the Finnish Presidency of the Committee of Ministers, there was lively discussion on the impact of the development of AI on human rights and the necessity of major research along with trust and transparency in this field.

During the conference it has been underlined that AI should be developed in a way to put the human beings at the centre of advantages for people and society. The need to establish efficient mechanisms of supervision and structures of democratic vigilance in relation to the design, the development and the implementation of AI and the necessity to acquire the public consciousness of potential risks and advantages of AI were the main points highlighted.

Moreover, there are the questions of the Council of Europe, an international body with a broader horizon than that of the EU, which puts into question the adoption of the guidelines based, among others, on Convention 108 and on the Ethical Charter on the use of AI in judiciary systems, while, at the same time, taking into account the European Convention for the protection of human rights and fundamental liberties.

Regarding the ethics of the data, the European Commission has established seven requirements for ethical AI which industry, research institutes and public authorities must respect. These are: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, non-discrimination and fairness; societal and environmental well-being; accountability[45].

On 21 April 2021, the European Commission issued a draft proposal of the AI Act to the European Parliament. In this proposal, it established a framework for determining whether an AI application poses a significant risk which would subject it to additional obligations, such as post-market monitoring, auditing requirements and conformity assessment, with the aim to create an EU-wide framework.

The OSCE Policy Manual on AI and Freedom of Expression of January 2022[46] stress concepts such as inclusive growth, sustainable development and well-being, equity and human parameters safeguard, transparency and understandability, robustness, safety, reliability and accountability. The aim is to avoid a technocentric approach to AI applications, where the human role is marginalized, even if the ultimate finality of these applications is to grow human well being. AI must be programmed in conformity to the respect for human rights and democratic values and must be reliable and transparent from a technical point of view. In the field of justice, the Ethical Charter on the use of AI adopted by the European Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe in December 2018 goes in that direction, providing a framework of principles that can guide policy makers, legislators and justice professionals when they grapple with the rapid development of AI in national judicial processes[47].

There is a huge amount of soft law instruments but an international binding treaty is

---

[45] "*Will your algorithm pass the test? Create AI humans can trust. Europe to lead human-centric AI: we invite the industry, research institutes and public authorities to test ethics for trustworthy AI drafted by a group of experts. The guidelines highlight the necessity for AI to respect all applicable laws, and they purpose seven key requirements for AI development. These include among others: human oversight, transparency, privacy and fairness*", https://europe.eu/!Rh69By.

[46] Organization for Security and Co-operation in Europe, *Spotlight on Artificial Intelligence and Freedom of Expression: a Policy Manual*, 20 January 2022.

[47] CEPEJ European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment, December 2018.

still lacking as well as a proper international digital rights code with the provision of new rights related to the human digital dimension, such as the right to explanation in automated decisions and the right not to be subjected to an entirely automated treatment[48]. These new rights need to be not only established but also effectively enforced, because at the moment they are devoid of protection. Automated decisions are independent from human intervention but nevertheless engrave on human fundamental rights. Scholars propose to recognise legal subjectivity for AI systems like for human beings, to attribute rights and responsibilities to AI systems when they divert from the inputs received by the programmers or users.[49] In this way, AI will be considered responsible like humans and subjected to sanctions such as the cancellation and deactivation of the system for a period of time so to re-educate it or economic sanctions should be established, such as sanctions against companies[50]. A robot owner of subjective legal capacities should go side by side with the human being, reducing the programmer's and user's responsibility[51].

Case law has established the general administrative nature of the "algorithmic rule", created by man; as a general administrative rule, algorithmic rule must be transparent and public, reasonable and proportional. It must provide the possibility that the judge can analyze, according to a human evaluation, automated decisions to verify that the automated process is correct and appropriate[52].

A point of balance must be found between human intelligence and artificial intelligence, between man and machine so as to create an anthropocentric ecosystem, thus preventing technological progress from putting man in a subsidiary position with respect to his technological and algorithmic creation[53]. A supervisory authority with administrative, legal and judicial powers could be created which should intervene according to the principles of dignity, non discrimination, transparence, accountability, freedom of expression, privacy, protection of personal data, and so on[54].

5. *Conclusions*

As an emergency remedy against possible distorted and harmful applications by intelligent systems, the use of an "emergency button", that one can activate to stop or transform a system, must always be possible, whilst avoiding that escaping all controls it causes damages or neutralizes the interruptive intervention, recognizing it as harmful for its

---

[48] Artt. 22 and 71 of the GDPR.

[49] G. SARTOR, *Cognitive automata and the law: Electronic contracting and the intentionality of software agents*, in *Artificial Intelligence and Law*, 2009, 17, 4, pp. 253 ss.; U. PAGALLO, *The Laws of Robots*, Springer, New York, 2013.

[50] G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities – From Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal*, 2016, 4, 2, pp. 171 et seq.

[51] G. TEUBNER, *Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten,* in *Archiv für die civilistische Praxis*, 2018, p. 218.

[52] Italian Council of State, sez. VI, 8 April 2019, n. 2270.

[53] A. PAJNO, M. BASSINI, G. DE GREGORIO, M. MACCHIA, F-P. PATTI, O. POLLICINO, S. QUATTROCOLO, D. SIMEOLI, P. SIRENA, *AI: profili giuridici. Intelligenza Artificiale: criticità emergenti e sfide per il giurista*, in *BioLaw Journal,* n. 3/2019; R. BURKHARDT, N. HOHN, C. WIGLEY, *Leading your organization to responsible AI,* QuantumBlack, 2 May 2019, https://www.mckinsey.com/business-functions/quantumblack/our-insights/leading-your-organization-to-responsible-ai (accessed 14 June 2022); L. AMOORE, R. RALEY, *Securing with algorithms,* in *Sage Journals, Secur Dialogue*, Volume 48, Issue 1, 2017, pp. 3-10; A. SALLES, K. EVERS, M. FARISCO, *Anthropomorphism in AI,* in *AJOB Neurosci*, 2020, pp. 88-95.

[54] G. HALLEVY, *supra* note 50*,* pp. 230-233.

existence[55]. It would be a so-called "big red button" to be activated by the human subject. It is impossible to apply a too strict precautionary criterion, because this would invalidate the possibility for the individual to benefit from the technological progress, whereas it is most reasonable a balancing between the undeniable advantage deriving from an intelligent application and a risk margin tolerable for the user.

Information on what is legally or morally right or wrong has to be incorporated in the intelligent system, even through social norms, in a way that automated decisions are compliant to those[56]. It is also important to add examples of human decisional processes where ethical dilemmas have been resolved and where illicit actions have been avoided. It is evident that much work needs to be done to create algorithms capable of ethical representations, like human beings do[57].

The Resolution of the European Parliament of 16 February 2017 lays down recommendations for the European Commission concerning norms of the civil rights regarding robotics[58]. The Parliament indicated as a general theme that in AI and robotic technology, human capacity should be integrated into technology rather than substituted. This is essential if society wants AI accelerate scientific progress, by also giving an ethical and legal dimension also to data that AI acquires, such as personal data, for which an ethical and legal use must be provided. The proposal of an EU Regulation on AI goes in the same direction."Symbiotic autonomy" is, in this sense, an instrument that allows human beings to help machines and *vice versa*, in order to overcome all the limits connected to ethical dilemmas[59]. A future in which humans and robots are complementary to each other is the rosiest.

The proposal of European Regulation of AI and the Liability Rules for AI of September 2022 offer therefore to guarantee a continuity path along with the GDPR. This development aims to better face the risks that have occurred in the technology sector due to the pandemic. This European legal proposal will not be able to have a deep impact even on a global level, as the case with GDPR, by ensuring the AI systems used in the EU will be safe, transparent, ethical, impartial and under human control and indeed. Anything considered a real threat to the European citizens will be banned.

To avoid the violation of human rights by intelligent systems, it is desirable that ethical principles will be developed and negotiated on a computational basis and used in the face of unforeseen situations, to limit regulatory violations or to deal with unforeseeable situations with a morally significant impact. Because machines do not have morality, it is expected that they will be designed according to shared ethical rules. Affecting computing, that is a branch of information technology that aims at the transmission of human feelings to machines, which will improve the relationship between man and computers, the HCI (human-computer interaction) because maybe a system

---

[55] T. ARNOLD, M. SCHEUTZ, *The Big Red Button Is Too Late: An Alternative Model for the Ethical Evaluation of AI Systems*, in *Ethics and Information Technology,* 2018, quoted in V. DIGNUM, *Ethics in Artificial Intelligence: Introduction to the Special Issue*, in *Ethics and Information Technology*, February 13, 2018.

[56] M.-P. SINGH, *Norms as a basis for governing sociotechnical systems*, in *ACM Transactions on Intelligent Systems and Technology,* Volume 5, Issue 1, December 2013, pp. 1-23.

[57] F. BERREBY., G. BOURGNE, J.G. GANASCIA, *A Declarative Modular Framework for Representing and Applying Ethical Principles*, in *AAMAS*, 2017, pp. 96-104; Y. BAR-COHEN/D. HANSON, *The coming robot revolution: expectation and fears about emerging intelligent, humanlike machines*, Springer, New York, 2016.

[58] European Parliament resolution of 16 February 2017 with recommendations to the Commission for civil law standards on robotics (2015/2103(INL).

[59] M. VELOSO, *Embrace a robot-human world*, in *Comment*, Vol. 521, May 28, 2015, p. 418.

capable of perceiving the user's state of mind can better evaluate his intention and his real will. In the future, this development could lead us to consider autonomous intelligent systems as a new form of life created by man, that is non-biological, deserving of rights, setting off a movement for recognition and self-determination. The intelligent machines of the future could be, like the animals of today, the target of ethical consideration, where equal rights are demanded in accordance with legal and ethical rules