

Artificial and Biological Neurons: Interdisciplinary Issues and Future Perspectives. White Paper

Daive Bazzanella, Giorgia Binoletto, Monica Consolandi, Marta Fasan, Francesca Gennari, Federico Carmelo La Vattiata, Luca Rinaldi, David Roccaro, Clara Zaccaria*

ARTIFICIAL AND BIOLOGICAL NEURONS: INTERDISCIPLINARY ISSUES AND FUTURE PERSPECTIVES.
WHITE PAPER

ABSTRACT: Recent developments in the technological domain have increased the interactions between artificial and natural spheres, leading to a growing interest in the ethical, legal and philosophical implications of AI research. The present paper aims at creating an interdisciplinary discussion on issues raised by the use and the implementation of artificial intelligence algorithms, robotics, and applied solutions in the neuroscience and biotechnology field. Building on the findings of the webinar “Workshop neuroni artificial e biologici: etica e diritto”, this work explores the issues discussed in the workshop, it attempts to show both the existing challenges and opportunities and it seeks to propose ways forward to overcome some of the investigated problems.

* *Daive Bazzanella, Ph.D. Student, Department of Physics, University of Trento, davide.bazzanella@unitn.it. Davide Bazzanella has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 788793); Giorgia Binoletto, Ph.D. Candidate in Law, Science and Technology at the University of Bologna and Research Fellowship Holder at the University of Trento, Faculty of Law, giorgia.binoletto@unitn.it; Monica Consolandi, Ph.D. Candidate in Philosophical Issues in an International Perspective at the Vita-Salute San Raffaele University, Faculty of Philosophy, and Fondazione Bruno Kessler of Trento, mconsolandi@fbk.eu; Marta Fasan, Ph.D. Candidate in Comparative and European Legal Studies at the University of Trento, Faculty of Law, marta.fasan@unitn.it; Francesca Gennari, Ph.D. Student in Law Science and Technology- Joint Doctorate - Rights of the Internet of Everything (LAST-JD-RIoE), MSCA ITN EJD n.814177, Mykolas Romeris University- University of Bologna-University of Turin, frgennari@stud.mruni.eu; Federico Carmelo La Vattiata, Ph.D. Student in Comparative and European Legal Studies at the University of Trento, Faculty of Law, and Fondazione Bruno Kessler of Trento, federico.lavattiata@unitn.it; Luca Rinaldi, Ph.D. Student in Comparative and European Legal Studies at the University of Trento, Faculty of Law, and Fondazione Bruno Kessler of Trento, luca.rinaldi-1@unitn.it; David Roccaro, Ph.D. Candidate, Department of Law, University of Catania, david.roccaro@phd.unict.it; Clara Zaccaria, Ph.D. Student, Department of Physics, University of Trento, clara.zaccaria@unitn.it. Clara Zaccaria has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 788793).*

All authors contributed equally to the conceptualisation and supervision of this work. Davide Bazzanella wrote Section 1.1 and Section 3.1. Giorgia Binoletto wrote the Introduction and Section 3.2. Monica Consolandi wrote Section 1.2 and 2.3. Marta Fasan wrote Section 2, Section 3, Section 3.4 and reviewed and edited the final manuscript. Francesca Gennari wrote Section 3.3 and most specifically section 3.3.1 concerning private law liability, Section 3.3.2 and Section 3.3.3. Federico Carmelo La Vattiata wrote Section 2.4 about the question of AI entities’ criminal law subjectivity, Section 3.3.1 about the functions of criminal liability, and Section 3.3.4. Luca Rinaldi wrote Section 2.2 and 2.4. David Roccaro wrote Section 2.1, 2.3 and Conclusive remarks. Clara Zaccaria wrote Section 1.1 and Section 3.1.



KEYWORDS: AI systems; organoids; ethical and legal issues; trustworthiness; interdisciplinarity

SUMMARY: 1. Introduction – 1.1 An attempt to define AI and its applications: the scientific and technological perspective – 1.2 An attempt to define AI and its applications: the philosophical perspective – 2. Emerging challenges in AI – 2.1 The dialogue between natural and artificial dimensions: the human’s role in artificial decision-making process – 2.2 The challenges raised by artificial intelligence for the labour market, equality and data protection – 2.3 Ethical considerations about AI: freedom, autonomy, authenticity, and self-perception – 2.4 Artificial agents as legal entities – 3. Building a pathway for a trustworthy AI’s application – 3.1 AI and social awareness – 3.2 AI and interdisciplinary approach – 3.3 AI and liability: theoretical and practical considerations – 3.3.1 The functions of liability. A concise analysis of the private and criminal perspectives – 3.3.2 What changes when liability meets the AI? Practical examples and open issues – 3.3.3 Policy and legal challenges concerning AI and liability in the EU – 3.3.4 A regulation fit for AI-risk. An example of technological risk-management from the US – 3.4 AI and the promotion of a constitutionally oriented approach – 4. Conclusive remarks.

1. Introduction

On December 3 and 4, 2020 the “Workshop neuroni artificiali e biologici: etica e diritto” took place in a webinar form. This workshop has been organised by the ERC-AdG Backup Project and the BioLaw Project¹ of the University of Trento, and by Bruno Kessler Center in Information and Communication Technology (FBK-ICT Irst Center)². The event aimed at creating an interdisciplinary discussion on issues raised by the use and the implementation of artificial intelligence algorithms, robotics, and applied solutions in the neuroscience and biotechnology field. The interactions between artificial and natural spheres played an important role in the discussion and several perspectives have been presented.

Recent developments in the technological domain have led to an increased interest in the ethical, legal and philosophical implications of AI research. This paper explores the issues discussed in the workshop and it attempts to show both the existing challenges and opportunities. Even if unable to encompass the entire and complex concerns, this work explores problematic implications of AI and artificial neurons from an interdisciplinary perspective and it seeks to propose ways forward to overcome some of the investigated problems.

The paper has been organised in the following way. After this Introduction, Section 1.1 and 1.2 attempts to define AI and some applications from the scientific and technological perspective, and the philosophical perspective, respectively. Section 2 investigates the main challenges, while Section 3 discusses a pathway for a trustworthy use of AI. Conclusive remarks are provided in Section 4.

¹ See respectively the websites of the projects at <https://r1.unitn.it/back-up/>, <https://www.biodiritto.org/> (accessed on 22/02/2021).

² <https://ict.fbk.eu/> (accessed on 22/02/2021).

1.1 An attempt to define AI and its applications: the scientific and technological perspective

Artificial Intelligence (hereinafter: AI) is a rather vague term currently used to describe a variety of concepts and devices. One of the most used definition of AI reads «any device that perceives its environment and takes actions that maximize its chance of successfully achieving its goals»³, or, according to the definition given by the High-Level Expert Group on AI (EU Commission), «(...) 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, 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»⁴. More practically, as computers have been for a long time the preferred tool for building AIs, the predominant meaning of AI refers to algorithms studied by the branch of computer science and information technology which has the same name of AI.

The development of AI algorithms started as early as the second half of the 20th century and it encompasses many different techniques⁵. Recently, thanks to the exponential increase of computational power and the advent of Big Data, a specific approach, called Deep Learning, achieved a number of celebrated successes, among which we can name the attempts of IBM's Watson at Jeopardy⁶ and of Google's AlphaGo at Go⁷. The decision processes embedded in these AI algorithms are usually not funded on causality, rather than on a statistical correlation model. For this reason, even if Deep Learning devices can carry out very elaborated tasks with a fairly high level of autonomy and efficiency, their operation is the more obscure to us the more complex the task at hand is. Indeed, they are often regarded as *black boxes*, due to the incapability of fully explaining the logic and working mechanisms which turn specific inputs into the final outputs⁸.

³ D. POOLE, A. MACKWORTH, R. GOEBEL, *Computational Intelligence: A Logical Approach*, New York, 1998.

⁴ HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (EU COMMISSION), *A Definition of AI: Main Capabilities and Disciplines*, Brussels, 2019, 6, at <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines> (accessed on 22/02/2021).

⁵ On this point see J. MCCARTHY, M. L. MINSKY, N. ROCHESTER, C. E. SHANNON, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. August 31, 1955*, in *AI Magazine*, 4, 2006, 12-14; N. J. NILSSON, *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, New York, 2010, 53-54; S. J. RUSSEL, P. NORVIG, *Artificial Intelligence. A Modern Approach (4th ed.)*, New York, 2020, 17 ss.

⁶ J. MARKOFF, *Computer wins on 'Jeopardy!': Trivial, it's not*, in *New York Times*, 16 February 2011, at <https://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html> (accessed on 22/02/2021).

⁷ S. BOROWIEC, *AlphaGo seals 4-1 victory over Go grandmaster Lee Sedol*, in *The Guardian*, 15 March 2016, at <https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol> (accessed on 22/02/2021).

⁸ F. PASQUALE, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Cambridge-London, 2015, 3 ss.



Als have so far provided society with some interesting applications and many more are under development, e.g. autonomous driving, content creation (music, pictures), personal assistants, telemedicine, and basic research (protein folding)⁹. However, Als certainly do not lack problems which could seriously affect their users. In fact, for instance, they are often developed by major companies, as the process requires considerable resources. Hence there is the inherent risk that these companies alone will benefit from Als, aggravating the inequalities already existing in our society. The development could also be carried out using controversial methods¹⁰. Additionally, being statistical models, they carry with themselves a set of problems, mainly related to their reliability. Firstly, there is no guarantee that the device's answers or actions will be correct: Als decision processes could be hindered by systematic errors during their development, such as algorithmic biases, which could lead to discrimination and unfair decisions. From this it also originates the question of the responsibility of the mistakes of Als. Secondly, there is the problem of the emergent effects, that is the response of the system to an input (environment) for which has not been developed, which assume a rather dire connotation in respect to autonomous ethical decisions. Last but not least, the actual target of any AI research is reason for heated debate, because it triggers important social, economic, scientific, ethical and legal problems, as in the case of political profiling from the law enforcements and that of autonomous weapons¹¹.

The increasing technological successes in the development of these systems are not just confined in the field of AI devices, tools and applications, but they're already influencing other scientific subjects connected to AI improvement too, e.g. biotechnology or neurosciences.

One of these subjects related to AI techniques and application concerns Human Enhancement (HE), here considered specifically in its connotation of artificial enhancement of human intelligence, which

⁹ See *ex multis* C. BROCK, *Where we're going, we don't need drivers: the legal issues and liability implications of automated vehicle technology*, in *Umc L. Rev.*, 83, 2015, 770-773; S. GRIGORESCU, B. TRASNEA, T. COCIAS, G. MACESANU, *A survey of deep learning techniques for autonomous driving*, in *Journal of Field Robotics*, 3, 2020, 362-386; J. CHUNG, *What Should We Do About Artificial Intelligence in Health Care?*, in *NYSBA Health Law Journal*, 3, 2017, 37; A. STROWEL, *Big data and data appropriation in the EU*, in T. APLIN (ed.), *Research Handbook on Intellectual Property and Digital Technologies*, Camberley, 2018; P. GUARDA, *"Ok Google, am I sick?": artificial intelligence, e-health, and data protection regulation*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2019, 359-375; E. LIM, *Meet my artificially-intelligent virtual self: creative avatars, machine learning, smart contracts and the copyright conundrum*, in *Journal of Intellectual Property Law & Practice*, 2020, 1-13. M. DURANTE, *Il potere computazionale. L'impatto delle ICT su diritto, società, sapere*, Sesto San Giovanni, 2019; R. F. SERVICE, *'The game has changed.'* *AI triumphs at protein folding*, 6521, 2020, 1144-1145; F. NOÉ, G. DE FABRITIIS, C. CLEMENTI, *Machine learning for protein folding and dynamics*, in *Current opinion in structural biology*, 60, 2020, 77-84; L. C. YANG, S. Y. CHOU, Y. H. YANG, *MidiNet: A convolutional generative adversarial network for symbolic-domain music generation*, 2017, arXiv preprint, at <https://arxiv.org/abs/1703.10847> (accessed on 22/02/2021); L. A. GATYS, A. S. ECKER, M. BETHGE, *Image style transfer using convolutional neural networks*, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 2414-2423; V. KEPUSKA, G. BOHOUTA, *Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home)*, in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, January 2018, 99-103.

¹⁰ <https://medium.com/@jyotimalhan/recaptcha-how-we-are-training-googles-ai-by-proving-i-am-not-a-robot-76651fbbe26a> (accessed on 22/02/2021).

¹¹ On these issues see, in general terms, the contributions collected in B. BRAUNSCHWEIG, M. GHALLAB (eds.), *Reflections on Artificial Intelligence for Humanity*, Cham, 2021.

is enabled by the convergence of technologies such as those provided by cognitive sciences, nanotechnology, biotechnology, and Brain-Machine Interfaces (BMI)¹². Clearly, the possibilities that such technologies envision are plenty, given by the combination of the capabilities of the human brain and those of modern machines. So, for instance, the progresses obtained in terms of Human Enhancement (HE) are quite promising in creating new communication and interaction opportunities for people affected by specific pathologies¹³.

If the desirables associated with HE are countless and of such great impact, the problems are surely not less numerous and not less important. Indeed, the acceptance from society of HE is conditioned by the answers to these problems, first and foremost the need to define which are the limits of these practices, although critical is the distinction of HE devices and practices between those that are historically already accepted and the novel ones which face resistance. Another critical differentiation is given by the purpose of the enhancement, e.g. therapeutic devices are accepted whilst plain enhancement techniques are seldom seen with hostility¹⁴.

Saving from the matter of collateral effects on the single person, the two most important problems for HE are the topic of authenticity and the matter of social equity: the former is the inquiry whether results obtained with and without HE are alike, the latter gathers all the concerns about the equity of the social impact that such technologies would produce.

A further subject associated with AI is a recent discovery in the field of biotechnology: the cerebral organoid. An organoid is a collection of cells which has been developed from stem cells and are organized into a simplified version of a full-size organ and which reproduces at least some of its functions¹⁵. They can develop from a multitude of different cells and are ergo divided in several types, such as, but not limited to, cardiac, lung, epithelial, and retinal organoids. A cerebral organoid is developed from cerebral cells and shares some mechanisms with the human brain. Therefore, a considerably advanced cerebral organoid - at present far from being realized - may exhibit reactions analogous to those proper of the human brain and, thus, it could earn the title of (artificial) intelligence.

Cerebral organoids, along with the other types of organoids, are used by researchers to study their mechanisms and functions in lieu of those of the full organ, in order to fight diseases and improve

¹² M. C. ROCCO, W. S. BAINBRIDGE, (eds.), *Converging Technologies for Improving Human Performance*, Dordrecht, 2004.

¹³ H. NAGANUMA, K. KIYOYAMA, T. TANAKA, *A 37 × 37 pixels artificial retina chip with edge enhancement function for 3-D stacked fully implantable retinal prosthesis*, in *2012 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Hsinchu, 2012, 212-215; X. ZHANG ET AL., *The combination of brain-computer interfaces and artificial intelligence: applications and challenges*, in *Annals of Translational Medicine*, 11, 2020, 1-9.

¹⁴ S. SONGHORIAN, *L'etica del potenziamento artificiale*, statement at the Workshop "Workshop neuroni artificiali e biologici: etica e diritto", 3-4 December 2020, University of Trento; G. VELTRI, *Incontro artificiale/naturale dalla prospettiva delle scienze sociali*, statement at the Workshop "Workshop neuroni artificiali e biologici: etica e diritto", 3-4 December 2020, University of Trento.

¹⁵ M. A. LANCASTER, J. A. KNOBLICH, *Organogenesis in a dish: modeling development and disease using organoid technologies*, in *Science*, 6194, 2014.



treatments. Given the relevance to medical and fundamental research, it is therefore of crucial importance to understand if, in which case or at all, the cerebral organoids should be considered to have a sort of intelligence and hence should be subject to rigorous oversight¹⁶.

In order for decisions to be made in regard to the intelligence of such concepts and devices and therefore to their regulation, it is very important to have a coherent definition of AI shared by all the academic disciplines, industry sectors, political and democratic institutions, and social strata.

1.2 An attempt to define AI and its applications: the philosophical and theoretical issues

Definitions are essential for us to be understood when we speak: we need to use a shared language to be clear and exchange information, opinions, and thoughts. Some attempts have been made throughout history to create a commonly shared language: consider Leibniz's *characteristica universalis*¹⁷, Frege's *Begriffsschrift*¹⁸, and the more recent *supervaluationism*¹⁹. The first two attempts were intended to structure the human language around precise definitions to avoid misunderstanding. Nonetheless, the struggle to create a shared language on a theoretical level does not exhaust the complexity of a "living" language spoken in a changing world, changing itself to follow the evolution of society: this is the supervaluationism goal, that aspires to create commonly shared definitions but always in updating and improving.

Since neither supervaluationism has become a rule - and probably never will -, we still have to deal with the language variety and its possible ambiguity, especially in a fast moving society such ours is. In particular, in recent years we are witnessing a surprisingly rapid progress in technologies, which is causing a change in our language, too: we have introduced in our daily vocabulary expressions such as Artificial Intelligence (AI), humanoid, enhancement. It seems that Brockman's prevision is actually happening, and more than ever we have to build a new, shared culture²⁰.

The complexity of the language derived from technological development is twofold: it shows its frailty when it comes to making interdisciplinary discussion; and it calls on ethical issues. As we experience in our daily life, technologies are cross-cutting: on a theoretical basis, it means that there is not only a right point of view to approach them but a series, meaning that the reflection upon technologies has to be interdisciplinary. This is the reason why it is hard to build an exhaustive definition of terms pertinent to the technological world: they have to include different perspectives on the same object. This

¹⁶ A. LAVAZZA, M. MASSIMINI, *Cerebral organoids: ethical issues and consciousness assessment*, in *Journal of Medical Ethics*, 9, 2018, 606-610; R. PROSSER SCULLY, *Miniature brains grown in the lab have human-like neural activity*, in *New Scientist*, 27 June 2019, at <https://www.newscientist.com/article/2207911-miniature-brains-grown-in-the-lab-have-human-like-neural-activity/> (accessed on 22/02/2021); I. SAMPLE, *Scientists 'may have crossed ethical line' in growing human brains*, in *The Guardian*, 21 October 2019, at <https://www.theguardian.com/science/2019/oct/21/scientists-may-have-crossed-ethical-line-in-growing-human-brains> (accessed on 22/02/2021); S. REARDON, *Can lab-grown brains become conscious?*, in *Nature*, 586, 2020, 658-661.

¹⁷ G. F. LEIBNIZ, *De arte combinatoria*, 1666.

¹⁸ G. FREGE, *Scrittura per concetti*, 1879.

¹⁹ A. C. VARZI, *Supervaluationism and Its Logics*, in *Mind*, 116, 2007, 633-676.

²⁰ J. BROCKMAN, *La terza Cultura. Oltre la rivoluzione scientifica*, trad. it. L. CARRA, Milan, 1995, 348, «[...] se cerco d'immaginare quale direzione prenderà la tecnologia nei primi anni del prossimo secolo, vedo che avrà luogo qualcosa d'incomprensibile [...]».

is not new: we know from Plato that the speakers' agreement on the subject of conversation is the essential basis for the possibility to speak about it²¹.

Thinking about AI, its definition is to be considered for its technical components, its legal limits, and its ethical possibilities. This last point recalls the ethical issues involved, showing that a complex definition is always related to the ethical thinking, and this case is no exception. Digging into this, we can point out selected interesting questions directly influencing the final definition of technological terms and interrelated between them.

First of all, is it possible to determine what is artificial without determining what is natural? It is not our intention to dig into a secular issue that is still unsolved (natural vs. artificial)²²: this thought-provoking question has the only function to help us in figuring the complexity underlying the apparently simple term *artificial*. Likewise, we need to know what a person is, i.e. a *human being*, before trying to define technological products that aim to imitate but differ it. This philosophical inquiry is far from being resolved, and possible answers reflect precise and arbitrary positions²³. Once we have defined a person, we will be able to better understand what a form of *enhancement* is, what we should define as organic and inorganic, what an organoid is intended not to be. In this context, the historical perspective also plays a key role: what is considered to enhance a human being really depends also on our habits and on the age we live. This kind of issue is well known in the pharmaceutical field, where is not unusual debating if a specific drug has to be considered therapeutic or enhancing²⁴. The same type of question could concern the classification of artificial neurons (both of organic and inorganic nature): when should we consider them a form of enhancement and when not? The choice is strictly connected to a biological perspective and the already mentioned definition of personhood or human. Moreover, the difference between therapy and enhancement has to be clarified, and it involves social and economic factors: considering a treatment a therapy or a form of enhancement has different consequences, both in terms of distribution and price. Defining the objects of discussion, it would be possible to choose more appropriate words to point them out: we coin new words and expressions everyday to describe a changing reality, and we need to choose the most proper one to avoid ambiguity and confusion²⁵. This very quick panoramic explains the interrelation of definitions like person, enhancement, organic and its constitutional interdisciplinary.

The network of definitions is thick and perfectly shows the importance of building a system of commonly shared definitions to make the interdisciplinary dialogue possible. Overturning the viewpoint, we should say that the interdisciplinary dialogue would make the creation of definitions possible, making accessible the complexity of technologies and their progress from different perspectives.

²¹ PLATO, *Sofista* (218c 1-5), «In ogni argomento bisogna invece sempre accordarsi sulla cosa stessa razionalmente piuttosto che sul solo nome separato da ogni ragionamento».

²² See one for all, the relatively recent book B. BENSUADE-VINCENT, W. R. NEWMAN (eds.), *The Artificial and the Natural. An Evolving Polarity*, Cambridge (MA), 2007.

²³ To deepen the argument, see E. T. OLSON, "Personal Identity", in E. N. ZALTA (ed.), *The Stanford Encyclopedia of Philosophy*, forthcoming.

²⁴ S. SONGHORIAN, *L'etica del potenziamento artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2021.

²⁵ This concept has been highlighted, among others, in L. JULIA, *L'Intelligence artificielle n'existe pas*, Paris, 2019, 122-123, in which the author suggests the use of the expression "Augmented Intelligence" instead of enhancement to avoid possible misunderstandings about the role of these technologies in relation to human intelligence.



2. Emerging challenges in AI

The technical features that characterise AI, and so the performance it provides, are gradually making this type of device more and more present in contemporary society, transforming the role and the impact these technologies can have in this context. Actually, AI has pervasively and disruptively become part of human life, both in the areas most closely linked to people's daily lives and in strategic sectors for the development and growth of individual countries and their societies. Such a similar pervasive presence of AI in the social dimension is pointed out by its application in, for example, information search and dissemination; document classification and analysis; commercial and financial transactions; home automation; the military; machine translation; industrial production; land, sea and air mobility; medical field; crime prevention; judicial sector; public administration; voice and facial recognition; encouraging better lifestyles; and in natural language understanding²⁶. Thus, the spread of AI is characterised by the variety of areas of application and by the diversity of functions it is able to perform.

The technical capabilities of AI, which enable these systems to perform functions once considered only humans, are giving rise to an increasingly presence of actions, interactions, choices and decisions, in which the "artificial" factor is assuming a leading role in these operations²⁷. As a result, there is a growing tendency to rely increasingly on AI solutions²⁸.

Such a similar phenomenon, as well as having an impact on the overall social structure, it also has significant effects from different perspectives, raising problematic social, ethical and legal issues.

2.1 The dialogue between natural and artificial dimensions: the human's role in the artificial decision-making process

We are in connection with artificial entities. This connection is evident for example when we change our usual behaviour by asking our voice assistant (and we rely on it) the fastest way to come back home, and when we follow the app suggestions where to dine or what movie to watch. We even allow the app to suggest words while we text a friend. Beyond the fideistic approach with which we experience this connection, we have to deal with a question: can we still distinguish the natural from the artificial? In this regard, this reminds the robot Andrew in Asimov's *The Bicentennial man*, who does not comprehend the «emotional antipathies» of the human beings who do not want to recognise him as a man. Andrew does wonder whether a common definition of the brain - including both his own and the human one - would be possible. The response he receives is even more remarkable: «Won't

²⁶ The mentioned fields of application of AI systems are exhaustively illustrated in G. SARTOR, F. LAGIOIA, *Le decisioni algoritmiche tra etica e diritto*, in U. RUFFOLO (ed.), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milan, 2020, 65; M. U. SCHERER, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, in *Harvard Journal of Law & Technology*, n. 2, 2016, 354 ss.; Y. N. HARARI, *Homo Deus. Breve storia del futuro*, Florence-Milan, 2018, 375 ss.

²⁷ On these aspects, M. ZANICHELLI, *Ecosistemi, opacità, autonomia: le sfide dell'intelligenza artificiale in alcune proposte recenti della Commissione Europea*, in A. D'ALOIA (ed.), *Intelligenza Artificiale e diritto. Come regolare un mondo nuovo*, Milan, 2020, 67-87.

²⁸ S. RODOTÀ, *Il diritto di avere diritti*, Rome-Bari, 2012, 401-402.

work, [...] your brain is manmade, the human brain is not. Your brain is constructed, theirs developed. To any human being who is intent on keeping up the *barrier* between himself and a robot, those differences are a steel wall a mile high and a mile thick»²⁹.

Thus, the question is exactly there: does that barrier actually still exist? Is it truly «a mile high and a mile thick»? In fact, a gradual blending of the biological and the artificial is emerging. Examples of this include research on the use of DNA in order to store data, and even the creation of *in vitro* mini-organisms. Nowadays, in hybrid systems the natural and the artificial merge. The boundary between *silicon* and *carbon*, which was used to be a sign of the clear distinction between life and artificiality, no longer seems to be such a clear-cut criterion³⁰. For instance, should I continue to assert that a clear distinction between the *organic* and the *inorganic* exists if the artificial-life organ grafted into my body - to replace or cooperate with one of my organs - is itself what allows me to survive?

Certainly, it is true, as Luciano Floridi affirms, that we are *enveloping* our environments around machines, so that they can comprehend precisely how to operate in complexity through their restricted intelligent abilities³¹. Nevertheless, it is also true that machines make decisions by processing a huge amount of data, linking up events, learning from the real world, forecasting what might happen and even making judgements that lead to changes. Through machine learning technologies, thanks to deep learning, big data and cloud computing, AI seems to be emancipating itself from human intervention and control, which seems to be no longer essential. It is as though the human being had succeeded in giving the life breath to machines and they now live a life of their own³². Although it would be worthwhile to dwell on the nature of these human-machine relations and especially on the possibility of defining these objects in subjective terms³³, what here arises is the fundamental issue regarding human control.

Human control can take place by means of Graphical User Interfaces (GUI) that allow to command the machines in a more intuitive way than through the programming code, for instance touching the smartphone with a finger. However, describing the issue in these terms is not useful when confronted with systems able to learn and decide autonomously. It is no longer a matter, in fact, of that kind of control that Douglas Engelbart in 1968 demonstrated to have during the so-called «The mother of all demos»³⁴. During this demo he showed through a mouse how to manage an interface to command a machine. Now, instead, the issue is to guarantee human-being participation in the decision-making process in order to avoid a total exclusion. For this reason, it is, for instance, more valuable to borrow

²⁹ I. ASIMOV, *The Bicentennial man*, in Id., *The Bicentennial man and other stories*, New York, 1976, 169.

³⁰ The topic was discussed in S. AMATO, *Tra silicio e carbonio: le machine saranno sempre stupide?*, statement at the Workshop “*Workshop neuroni artificiali e biologici: etica e diritto*”, 3-4 December 2020, University of Trento.

³¹ See L. FLORIDI, *La quarta rivoluzione. Come l’infosfera sta trasformando il mondo*, Milan, 2017, 165 ss; L. FLORIDI, *What the Near Future of Artificial Intelligence Could Be*, in *Philosophy & Technology*, 32, 2019, 1-15.

³² See R. BODEI, *Dominio e sottomissione*, Bologna, 2019, 297.

³³ F. PIZZETTI, *Embrioni, Organoidi e Robot: Soggetti giuridici?*, statement at the Workshop “*Workshop neuroni artificiali e biologici: etica e diritto*”, 3-4 December 2020, University of Trento.

³⁴ The Engelbart’s presentation is accessible at <https://www.doungengelbart.org/content/view/276/000/> (accessed on 22/02/2021).



concepts such as *meaningful human control*, which the European Parliament mentions in connection with the use of autonomous weapon systems³⁵.

In this sense, human control can take different forms³⁶: the first scenario sees human beings always having control of every part of the process (*human in the loop*). Another scenario is where human beings control the machine's activities successively without being able to interfere during the process (*human post-loop*). Finally, human beings can limit their role just to supervise the decision-making process (*human on the loop*). Obviously, another option is to abandon control over the automatic and autonomous operation of the AI (*human out of the loop*).

Merely reflecting on forms of control is not sufficient. In addition, it is necessary to consider to whom control should belong, how it should be exercised and what aims it should pursue. It is especially appropriate to reflect on how control can be exercised if the AI system functioning is opaque and not transparent (for instance, in black-box systems). Indeed, precisely for this reason, e.g. the «right not to be subject to a decision based solely on automated processing» provided for by art. 22 GDPR risks to be an empty shell that does not provide any enhanced control to human beings. Hence, besides providing for forms of effective human control, it is also necessary to tackle the risk of *getting used to* indiscriminate delegation to automated forms of decision-making³⁷. This risk must be countered by educating the younger generations (even if not exclusively) and by offering a range of possibilities to prevent artificial systems from assuming a dominant position.

2.2 The challenges raised by artificial intelligence for the labour market, equality and data protection

It has been a while since artificial intelligence has left the laboratories to become a technology more and more pervasive in our society. In the last decades, an increasing number of activities, previously realized entirely by humans, has seen its involvement: automation and robotics are already irreplaceable in many fields of industrial production, while automated decision systems, often based on *machine learning*, seem to have the potential to revolutionize many intellectual professions³⁸. This scenario opens unprecedented opportunities of development for the whole of humanity, but at the same time it raises pressing social issues, which cannot be ignored.

Firstly, it has to be highlighted that the advent of artificial intelligence could have serious repercussions on the labour market³⁹. Many technological innovations in history have been accused of decreasing

³⁵ See European Parliament resolution of 12 September 2018 on autonomous weapon systems (2018/2752(RSP)). See also EUROPEAN GROUP ON ETHICS IN SCIENCE AND NEW TECHNOLOGIES, *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*, 9 March 2018, at https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf (accessed on 22/02/2021); S. AMATO, *Biodiritto. Intelligenza artificiale e nuove tecnologie*, Turin, 2020, 107.

³⁶ See M. DURANTE, *Potere computazionale. L'impatto delle ICT su diritto, società, sapere*, Milan, 2019, 93.

³⁷ C. CASONATO, *Incontro tra artificiale/naturale: potenzialità e limiti*, statement at the Workshop "Workshop neuroni artificiali e biologici: etica e diritto", 3-4 December 2020, University of Trento.

³⁸ For an overview of recent development in AI see, among others, S. J. RUSSEL, P. NORVIG, *op. cit.*, 17-31.

³⁹ C. WEBSTER, S. IVANOV, *Robotics, artificial intelligence, and the evolving nature of work*, in B. GEORGE, J. PAUL (eds.), *Digital Transformation in Business and Society Theory and Cases*, London, 127-143; C. CASONATO, *Costituzione e*

job opportunities, because of the automation of tasks undertaken by humans. This prevision has often proved to be incorrect, as in the past new technologies have caused a transformation – and not a reduction - of job opportunities⁴⁰. This will probably be the case of the spreading of artificial intelligence as well: new kinds of jobs will appear while many of the current ones will become obsolete. However, also this transformation appears to be capable of disruptive social effects, because it is foreseeable it will disproportionately affect different parts of our societies. Indeed, low-skilled and repetitive jobs will be automated and replaced by more qualified ones related to the interaction with the machines⁴¹. It will probably be needed a profound rethink of the educational system and the implementation of projects aimed at retraining already active workers to mitigate these risks and enjoy the full benefits of artificial intelligence.

Secondly, it has to be underlined that the inherently statistical nature of automated decision algorithms based on machine and deep learning could lead to disturbing consequences. These systems shape their behaviour on a set of training data, usually consisting of a body of past examples held by human actors. There is a concrete danger that they replicate bias and prejudices contained in human decisions, giving rise to various kinds of discrimination (i.e. based on ethnicity, gender, economic conditions, etc.) that risk to be very difficult to identify due to the opacity of many deep learning algorithms⁴². Measures have to be taken to manage this risk of *algorithmic discrimination*, such as the implementation of meaningful and effective human control on automated decisions and the development of procedures to guarantee that systems are trained on datasets conformed to the principles of diversity and inclusiveness.

Thirdly, algorithms' opacity raises the issue of defining in which fields, and within which limits, it can be socially acceptable to rely on the decisions of systems that preclude the reconstruction of meaningful reasoning to explain their outcomes⁴³. It is a complex topic with no unique solution and requiring a deep analysis to be conducted case-by-case. Taking as an example the judicial decision, its peculiarity in terms of the duty to state reasons imposes a distinction between *context of discovery* and *context of justification*. Probably in the former, dedicated to the study of the case and to the collection of information from different sources, the advice of a black-box algorithm can be considered together with the other elements useful for the decision. Instead, in the latter, a full explanation of the grounds

intelligenza artificiale: un'agenda per il prossimo futuro, in Vv. AA., *Liber amicorum per Pasquale Costanzo. Diritto costituzionale in trasformazione - vol. 1: Costituzionalismo, reti e intelligenza artificiale*, Genova, 2020, 380-381.

⁴⁰ D. H. AUTOR, *Why are there still so many jobs? The history and future of workplace automation*, in *Journal of Economic Perspectives*, 29, 2015, 3-30.

⁴¹ See in general A. LÓPEZ PELÁEZ (ed.), *The Robotics Divide. A New Frontier in the 21st Century?* London, 2014.

⁴² S. BAROCAS, A. D. SELBST, *Big Data's Disparate Impact*, in *California Law Review*, 3, 2016, 671-732. As an example of discrimination, can be mention the usage of the COMPAS algorithm in parole decisions in US court, discussed in the well-known *Loomis case*, see *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), in *Harvard Law Review*, 130, 2017, 1530-1537; G. CORBELLINI, *Opacità epistemiche e sfide etico-legali: perlustrazioni nella terra di nessuno fra intelligenze naturali e artificiali*, statement at the Workshop "Workshop neuroni artificiali e biologici: etica e diritto", 3-4 December 2020, University of Trento.

⁴³ See in general F. PASQUALE, *op. cit.*

of the decision should be provided and the advice of the system should never be considered *per se* as a valid justification of the judicial outcome⁴⁴.

Lastly, concerns in matters of rights to privacy and data protection have to be mentioned. The massive amount of personal data available in our new, digital societies and the involvement of artificial intelligence in data analytics permit profiling and data processing activities of unprecedented intrusiveness⁴⁵. In order to seize the opportunities of artificial intelligence and to mitigate related risks, legal frameworks should address these issues by establishing general principles (i.e. the work of EC High Level Expert Group on AI⁴⁶) and by providing rights and duties that allow a greater control over personal data.

2.3 Ethical considerations about AI: freedom, autonomy, authenticity and self-perception

The rise of AI technologies makes us face open ethical issues related to the scenario we are describing. Among the others, we chose three of them that in our opinion are strictly correlated to the overall discussion on artificial intelligence: freedom; autonomy; authenticity and self-perception.

The topic about freedom, in the sense of what Isaiah Berlin calls *negative freedom* as the absence of restrictions, i.e. independence from external constraints⁴⁷, is the one that emphasizes the necessity of reasoning about limits and potential of AI. The main questions that the topic of freedom calls on are: i) are there any limits to apply to the development and use of AI? Do we have the right to feel free about its growth and applications? In this regard it is useful to reflect on the possibility that the advance of AI in every field of individual and collective existence is not an inescapable necessity but derives from a choice that entails countless changes. ii) Do we experience any limit in terms of human freedom due to AI applications? Observing the problem from the point of view of personal data automated treatment is not a sufficient response. The issue is more complex and must lead us to wonder whether we can still speak of human freedom in a world characterised by blind reliance on continuous and uninterrupted calculation as a source of truth. Can we affirm in all conscience that there is still room for free human action when it is undermined by agents influencing behaviour, constraints given by technical structures, paths forced by AI determined preferences? These open issues remind us that it is necessary to set limits to AI not to be overcome; besides, we do have to consider its potential in concreteness: it is not exhaustive to think about pros and cons of AI when not realistically defined. It is required to reason about them in a rooted-in-real-world way.

On the other hand, if we delve into the conceptual terrain of what Berlin used to call *positive freedom*, we must reflect on the autonomy of machines, i.e. the possibility that machines can determine their

⁴⁴ A. SANTOSUOSSO, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*, Milan, 2020, 107 ss.

⁴⁵ See, among others, G. PASCUZZI, *Il diritto dell'era digitale*, Bologna, 2020, 77 ss.; J. E. MAI, *Big data privacy: the datafication of personal information*, in *The Information Society*, 3, 2016 192-199; S. RODOTÀ, *Privacy, libertà, dignità*, Conclusive Remarks at the XXVI International Conference on Privacy and Data Protection, 2004, at <https://bit.ly/31ZgTlx> (accessed on 22/02/2021).

⁴⁶ See for instance HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (EU COMMISSION), *Ethics guidelines for trustworthy AI*, 8 April 2019, at <https://bit.ly/2LVf4k6> (accessed on 22/02/2021).

⁴⁷ I. BERLIN, *Two concepts of liberty*, in ID., *Liberty*, Oxford, 2002.

own law. Should we let machines dictate their own *nomos* and acquire the freedom that human beings so ardently desire? Autonomy of AIs is a common fear in dystopias about the possible deviation of them⁴⁸: what would happen if machines learned to use their intelligence on their own? Even if it does not seem to be a realistic hypothesis, we have to face the fact that it is not possible even for programmers to fully explain the causality that links specific inputs to their corresponding outputs, and in the future AI could reach unexpected levels of autonomy through algorithms that operate beyond received instructions, learning and deciding autonomously. In this regard, the question of algorithmic reasoning opacity, the black-box effect about which the European Commission speaks in its White Paper⁴⁹, must be systematically addressed to put in place structural barriers that enable human beings to stay in the game, or *in the loop*. Trusting in the *doxa* that attributes a neutral attitude to an algorithm's action is definitely a leap in the dark. This raises an important question about agency: do we have to consider that machines could in the future have a legal status?⁵⁰ This question emphasizes two sides of the problem: i) we have to ask ourselves if we are ready to consider this possibility and if it is actually possible at all to apply the meaning of agency to an AI system; ii) we do not really know how far AI would eventually go: even if we are AI's inventors, it could become independent from us in an unpredicted and unprecedented way. The issue about autonomy forces us to reflect upon the limits we want to impose, and the ones we are ready to cross.

The interrelated topics of authenticity and self-perception are deeply human: how does AI have an effect on our integrity when applied to human beings? These two terms stand for two opposite movements: authenticity concerns the form, what is seen from the outside and considered by others in terms of intentions and being; self-perception represents the internal movement that a human being is expected to fulfil, looking for their existential foundation. Taking for example the use of artificial neurons, is a person's authenticity questioned if artificial neurons are adopted? Jean-Luc Nancy in his work about his heart transplant talks about an «intruder» in his own chest⁵¹: what if the «intruder» were artificial and "influencing" your thoughts, "enhancing" your abilities, "shaping" your actions? Can we still discuss an authentically human way of being, thinking and acting? Should we suppose there are differences between the therapeutic and the enhancing use⁵²? Moreover, how does this affect the persons' agency when it comes to establish mandatory criteria in a specific situation, e.g. a race? Asked these tricky questions, self-perception is still unsolved: how do I perceive myself knowing that I have in me artificial parts, influencing both my biological status and my thoughts? Does this have consequences in terms of behaviour, feeling, and emotions? We have reasons to think that this is a delicate problem: the case of a patient that did not recognize themselves during a Deep Brain Stimulation to

⁴⁸ See C. BRENTARI, *Macchine e motivi. Annotazioni filosofiche sulla sfera motivazionale dell'intelligenza artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2021.

⁴⁹ EUROPEAN COMMISSION, *White Paper on Artificial Intelligence: a European approach to excellence and trust*, 19 February 2020, 12, at https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (accessed on 22/02/2021).

⁵⁰ F. PIZZETTI, *op. cit.*

⁵¹ J.-L. NANCY, *Corpus*, New York, 2008, 161 ss.

⁵² L. GALVAGNI, *Naturale, artificiale e virtuale: il fenomeno complesso della vita*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2021.



relief the pain tells us that we are not reasoning on a merely theoretical level, and that it is a very actual problem instead⁵³.

The raised issues presuppose a fundamental one: a defined conception of human being on which the narrative about AI is based. Is a human being a living thing who thinks or that reacts to *stimuli*? If automated decision-making systems replace humans in all fields, what will happen to human faculty of judgement? And what kind of human being will emerge from the comparison with AI? Will humanity stand out just because it is capable of general intelligence rather than a specialised one? What about human imperfect uniqueness? Attempting to simulate the human brain, AI inevitably ends up modifying it, albeit indirectly, since the human being's representation of themselves changes. In today's society individuals are in continuous competition in every sphere of life: how will human beings be able to withstand competition with other intelligent entities that have an infinitely greater memory and certainly a greater speed of calculation? A possible consequence is that *imperfection, frailty, vulnerability, and diversity* risk to vanish from the horizons of reality. How can we imagine understanding and still accepting these human features when compared to AI, which is supposed to be impossible to fail?

We believe these are questions not easy to respond, and in all probability essentially open. Even if we should presume that it is useful to keep them open, the breadth of issues should not lead to indifference and need to be addressed before AI continues its development. The overview of these three topics may be overwhelming; it would be helpful to face them step by step, always considering the powerful cooperation between different disciplines and approaches, in our opinion the only possible way to deal with the issues that these delicate themes bring with them.

2.4 Artificial agents as legal entities

From the legal perspective, one of the most challenging issues raised by the advent of artificial intelligence is the possibility of granting legal personality to artificial agents⁵⁴. The hypothesis of creating a *specific legal status for robots*, advanced in the *Civil rules on robotics* proposed by the European Parliament in 2017⁵⁵, has been widely discussed⁵⁶. It appears to us that there are no compelling reasons to be entirely hostile to the recognition of a form of diminished civil law agency to most advanced and autonomous robots, giving legal relevance to some of their actions in order to rationalize the collocation of risk and liability connected to their usage⁵⁷.

The question of AI entities' agency arises in the field of criminal law as well. However, we believe that the recognition of any form of AI's subjectivity in this sector faces insuperable obstacles. Although

⁵³ S. SONGHORIAN, *L'etica del potenziamento artificiale*, statement at the Workshop "Workshop neuroni artificiali e biologici: etica e diritto", 3-4 December 2020, University of Trento.

⁵⁴ See, among others, A. SANTOSUOSSO, *The human rights of nonhuman artificial entities: an oxymoron?* in *Jahrbuch für Wissenschaft und Ethik*, 19, 2015, 203-237; U. PAGALLO, *Robotrust and legal responsibility*, in *Knowledge, Technology and Policy*, 23, 2010, 367-379; G. TEUBNER, *Rights of Non-Humans? Electronic Agents and Animals as New Actors in Politics and Law*, in *Journal of Law and Society*, 4, 2006, 497-521.

⁵⁵ European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics, P8_TA(2017)0051.

⁵⁶ See, among others, the *Open Letter to the European Commission – Artificial Intelligence and Robotics* of 2018, at <http://www.robotics-openletter.eu/>, signed by various robotics and AI experts.

⁵⁷ F. PIZZETTI, *op. cit.*

some author has theorised possible forms of legal subjectivity (and liability) of AI systems⁵⁸, there are a number of arguments to rebut the aforementioned opinion, such as: AI systems are not actually *intelligent*, at least in the sense we mean human intelligence⁵⁹; also, criminal responsibility is based on the two crucial concepts of *wrongdoing* and *attribution*⁶⁰, which presuppose two requisites, *i.e.*, a human and voluntarily taken act or omission (the so-called *actus reus*) on the one hand, and some variety of fault (the so-called *mens rea*) on the other, that is the conduct needs to be *covered* by a guilty mind⁶¹. More generally, it is important to emphasize that the real peculiarity of human beings does not consist in legal subjectivity - constantly recognized also to companies and other non-human legal entities, to which could be added robots, with the mentioned limits – but in human dignity, which inner meaning is protecting human uniqueness, banning any equalization between human beings and other entities⁶². This is particularly relevant with regard to the recent developments in cerebral organoids, which could be defined as *artificial intelligence in the literal sense*⁶³. Organoids do not have the capacity to develop into complete human beings, unlike the embryos, so they are not entitled to the degree of dignity recognized to the latter by many national and supranational courts⁶⁴. Organoids could find some form of legal protection, beyond any recognition of legal personality or human dignity, if they evolved in entities complex enough to experience pain and suffering, as it happens for animals⁶⁵. This should be taken in account at the moment of regulating the usage of organoids in scientific research.

⁵⁸ See *ex multis* G. HALLEVY, *The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control*, in *Akron Intellectual Property Journal*, 2, Article 1, 2010, 186, at <http://ideaexchange.uakron.edu/akronintellectualproperty/vol4/iss2/1>.

⁵⁹ As John McCarthy *et al.* affirmed in 1955, the AI problem «is taken to be that of making a machine behave in ways that would be called intelligent if a human were so behaving» (see J. MCCARTHY, M.L. MINSKY, N. ROCHESTER, C.E. SHANNON, *op. cit.*, 12). Therefore, we can call such a behaviour *intelligent* in as much as a human behaves in that way, but it does not mean that the machine is intelligent (see L. FLORIDI, *Digital's Cleaving Power and Its Consequences*, in *Philos. Technol.*, 30, 2017, 123-129, at <https://doi.org/10.1007/s13347-017-0259-1>).

⁶⁰ As for the concepts of *wrongful act*, *wrongdoing*, and *attribution* see *amplius* the fundamental work by G. P. FLETCHER, *Rethinking Criminal Law*, Oxford – New York, 2000, 454-491.

⁶¹ G. P. FLETCHER, *op. cit.*, 475; T. C. KING, N. AGGARWAL, M. TADDEO, L. FLORIDI, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, in *Sci. Eng. Ethics*, 26, 2020, 90-91, at <https://doi.org/10.1007/s11948-018-00081-0> (accessed on 22/02/2021).

⁶² J. WEINRIB, *Human dignity and autonomy*, in *Max Planck Encyclopedia of Comparative Constitutional Law*, 2020, at <https://bit.ly/38VNFrz> (accessed on 22/02/2021).

⁶³ In reference to cerebral organoids, by “artificial intelligence in the literal sense” it is meant that they are *artificial* in the sense that they are biological tissue cultures lab-developed with human activity and they are *intelligence* in a figurative sense because they consist in miniature organs resembling some of the brain structures. Either way, it has to be highlighted that state-of-the-art cerebral organoids are very far from showing any meaningful form of what it is commonly meant by “intelligence”.

⁶⁴ See e.g. Corte cost., 18 February 1975, n. 27; Corte cost., 13 April 2016, n. 84; CJEU, 18 December 2014, n. C-364/13, *International Stem Cell Corporation*, ECLI:EU:C:2014:2451; ECHR, 28 May 2013, n. 46470/11, *Parrillo v. Italy*.

⁶⁵ A. LAVAZZA, F. PIZZETTI, *Human and cerebral organoids as a new legal and ethical challenge*, in *Journal of Law and the Biosciences*, 2020, 1-22; see also F. PIZZETTI, *op. cit.*



3. Building a pathway for a trustworthy AI's application

As it has been described so far in this contribution, the pervasive presence of AI technologies in daily life is an undeniable fact and it's representing a fundamental game-changer in the accomplishment of many functions and actions which used to be considered only human. Furthermore, the disruptive effects of this artificial and digital revolution, both in terms of benefits and risks, are not lacking relevant consequences from all the examined perspectives.

Indeed, the paradigm shift carried out through the development and the spread of AI systems entails a new way to consider the scientific, ethical, and legal categories which are traditionally involved dealing with the described problematic issues, wondering if they might effectively and adequately solve the AI most common emerging questions or if it's necessary to create new categories and tools in order to face AI deployment in the real world.

In such an uncertain scenario, where it's not already clear or evident how the experts and the institutions will deal with all the challenges posed by the AI advent, the first fundamental step to take is to identify which factors and tools could help define a new path to create a "trustworthy"⁶⁶ AI for the "Good Society"⁶⁷. With this aim, it's necessary to set up a proper framework, in which the risks and benefits associated with the use of AI are effectively and appropriately balanced, trying to guarantee both the maximization of benefits and the minimization of damages following from the use of intelligent technologies, assuring the protection and the empowerment of human freedoms, rights, and values and avoiding any unjustified fear which could limit the scientific and technological progress.

In this context, three specific elements could be useful to build up a set of rules, tools, policies and initiatives aimed at pursuing the set goals: the growth of the public social awareness on the real effects of AI technologies; the enhancement of an interdisciplinary approach in this field; the development of a proper regulation concerning liability issues; and the promotion of a constitutionally oriented AI in order to assure the protection of human values, freedoms, and rights.

3.1. AI and social awareness

As it was explored so far, the introduction of Artificial Intelligence technologies entails a lot of ethical, philosophical, social, and legal problems.

However, it is worth considering that many of these problems stem from incomprehension between the parties involved, hence an open and public debate should bring clarity to the concepts involved and, by doing so, it could also solve some of these problems. So, from this perspective, the growth of public social awareness has a fundamental role in the development of a trustworthy AI.

It is important to be aware of the public understanding of science, in order to face the social and economic effects that a new technology gives rise to. If the discussion remains closed within the specialist of the matter, some of these effects might not be considered or overlooked. Moreover, we are far

⁶⁶ HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (EU COMMISSION), *Ethics Guidelines for Trustworthy AI*, cit., 6 ss.

⁶⁷ C. CATH ET AL., *Artificial Intelligence and the 'Good Society': the US, EU, and UK approach*, in *Sci Eng Ethics*, 2, 2018, 505-528.

away from the idea of an *unthinking society*, which means that people will not easily accept scientists' and experts' discoveries and decisions, even though they could be right⁶⁸. It is then of main importance to create an open discussion between experts, companies, and the population in order to find a proper regulation in view of the introduction of the new technologies. Science must not repeat the errors done dealing with biotechnologies and GMO during the '80s: in that case the new social and economic relations that GMO technologies were introducing between farmer and the material suppliers were not properly considered⁶⁹.

This open approach has to be adopted especially in the case of AI technologies and its implications. Regarding HE, in particular in the case of internal enhancement, the main concerns are about the notion of authenticity and the matter of worthiness. Moreover, socially speaking, HE and AI algorithms bring with themselves the problem of access to the new technologies, i.e. their equity, and therefore the possible increase of social disparities. Finally, the scope of the AI technologies is also very important, as it can influence the acceptance from the population, as for example in the case of autonomous weapons.

The involvement of the society in the discussion from the beginning is of primary importance: this bottom-up approach can lead to the creation of a shared point of view, on which rules can be created. In this sense, the French normative experience in the field of bioethics is paradigmatic of such a social involvement in the law-making process⁷⁰.

The approach of taking into consideration new points of view, sensitivities, and attitudes could lead to the birth of a new ethic, more contextual, historical, inclusive, participatory, and creative. Contextual and historical because every population will react to new technologies depending on its own ethic, social, and economical system; inclusive and participatory because all the individual citizens can contribute with their own point of view; creative, because from the public consultation can emerge new aspects that would be otherwise ignored in a closed and specialistic debate⁷¹.

However, any debate will be useless and pointless if the people involved, from the general public to the specialists, don't share a clear definition of the concepts used in the dialogue: the understanding of the technologies - neural networks, organoids - rather than their potential and the consequences of their use. That's why it is of primary importance to involve as much as possible, besides the academics, also the general population and the companies in an educational effort that aims at laying the foundation for this shared discussion framework. Moreover, since these problems should be tackled from every side, it is essential for the positive outcome of this attempt to assume an interdisciplinary approach to Als.

⁶⁸ G. VELTRI, *op. cit.*

⁶⁹ The most problematic issues concerning the regulation of GMO products are generally highlighted in E. CATTANEO, *Ogni giorno. Tra scienza e politica*, Milan, 2016.

⁷⁰ This kind of social involvement has been institutionalised with the *états généraux de la bioéthique* (2009), which represent a type of public consultation instrument typical of the French system: the *débat public*. On this topic see S. PENASA, *La legge della scienza: nuovi paradigmi di disciplina dell'attività medico-scientifica. Uno studio comparato in materia di procreazione medicalmente assistita*, Naples, 2015, 152 ss.; J. F. BERAUD, *Il caso della Francia: la Commission National du Débat Public*, in A. VALASTRO (ed.), *Le regole della democrazia partecipativa. Itinerari per la costruzione di un modello di governo*, Naples, 2010, 387.

⁷¹ See L. GALVAGNI, *Naturale artificiale e virtuale: il fenomeno complesso della vita*, statement at the Workshop "Workshop neuroni artificiali e biologici: etica e diritto", 3-4 December 2020, University of Trento.

3.2 AI and the interdisciplinary approach

The second element which has a relevant and key role in the development of a trustworthy framework for the AI application is the enhancement of an interdisciplinary approach in this field.

As pointed out by Pascuzzi, inside universities and their communities traditional and academic disciplines (e.g. law, physics, mathematics) and categorisation of learning (e.g. natural science, social science, humanities) distinguish different approaches to knowledge⁷². Scholars are traditionally divided according to rigid disciplines of reference, and their research is carried out within these boundaries. Academic and professional careers are regulated in the same way⁷³. However, it should be stressed that reality is more complex than this categorisation. Real-world problems, including the issues arising in the Digital Age⁷⁴ with the use of AI, are truly complex and simultaneously involve different branches of knowledge. As demonstrated in the workshop during the speeches, AI raises concerns related to ethics, philosophy, law, social science, biology, medicine, computer science, etc., at the same time. Emerging problems should be addressed in dialog between disciplines since defining boundaries is nearly impossible in the new and disruptive scenarios that AI creates. Solutions to common problems (e.g. opacity, transparency, and discrimination) need comprehensive understanding, problem solving skills, plural perspectives, and studies that crosses disciplines. Thus, the method of research should be “inter-disciplinary”.

Interdisciplinary studies are emerging in recent years to investigate the impact of digital technologies. The process of integration of disciplines is always and inevitably tricky⁷⁵. Since experts and professionals are traditionally trained in the confines of a specific discipline, concepts, languages, and approaches to problems differ. Preconceived notions must not be maintained in the interdisciplinary work; at the same time, different experts must maintain the gained knowledge and roles, and achieve a better understanding of other disciplines involved without losing specificity, focus, and rigor. Studies should not simply use two or more disciplines in a multidisciplinary method, but they should integrate and synthesize plural ideas and approaches to seek advanced solutions. Primarily, common definitions on the objects of investigation (e.g. neuron or intelligence) should be found. On this common ground, solutions will be the result of a constructive dialogue and compromise from a comprehensive perspective. Research on AI issues should be carried out in collaborative research groups where lawyers, philosophers, experts on ethics of AI, scientists of different fields are involved. Moreover, a pivotal aspect for being interdisciplinary is training students, which in the future will become experts in a defined discipline, to use an interdisciplinary method and to achieve skills that cross that discipline of reference during their academic studies and education⁷⁶. For instance, in a faculty of law students should also be

⁷² G. PASCUZZI, *Quale formazione per la ricerca interdisciplinare?* in *BioLaw Journal – Rivista di BioDiritto*, 1, 2021.

⁷³ As regards Italy see the information available at <https://www.miur.gov.it/settori-concorsuali-e-settori-scientifico-disciplinari> (last accessed 01/17/2021).

⁷⁴ G. PASCUZZI, *Il diritto dell'era digitale*, cit., 21-24.

⁷⁵ G. D. BREWER, *The challenges of interdisciplinarity*, in *Policy sciences*, 4, 1999, 327-337; K. FUCHSMAN, H. STUART, *Rethinking integration in interdisciplinary studies*, in *Issues in Interdisciplinary Studies*, 27, 2009, 70-85.

⁷⁶ See e.g. W. H. NEWELL, W. J. GREEN, *Defining and Teaching Interdisciplinary Studies*, in *Improving College and University Teaching*, 1, 1982, 23-30; W. H. NEWELL, J. THOMPSON KLEIN, *Interdisciplinary Studies into the 21st century*, 2, 1996, 152-169.

trained on computer science and biological notions, and in a faculty of computer science courses on ethics and law should be provided⁷⁷. To achieve this goal, interdisciplinary degrees and laboratories on AI will increasingly be organised by universities. Furthermore, explicitly interdisciplinary Ph.D. programs are emerging. This is an important way forward to interdisciplinary research focused on AI since flexible and critical thinking is developed by young scholars from the beginning of their professional work. Being interdisciplinary by default is a challenge but may represent the future in artificial intelligence studies.

Beyond these training and research solutions, the importance of adopting the interdisciplinary approach from the earliest levels of education should be highlighted. The opportunity to create and invest in new models of education would make people from an early age used to improve the dialogue between different types of knowledge, to have a critical view of the AI's challenges, to recognise all the beneficial possibilities as well as the most problematic and dangerous aspects offered by these new artificial systems. As a result, it would also become easier to increase people's awareness when facing AI's applications.

3.3 AI and liability: theoretical and practical considerations

The Digital Revolution in general⁷⁸, and Artificial Intelligence (AI) in particular, have started changing not only our daily life but also what we think to be liable means. AI is at the moment far from being considered in a *strong* sense that can pass the Turing test⁷⁹. It would be more convenient to describe it as a set of decision-making techniques based on sophisticated algorithms and statistical methods⁸⁰.

3.3.1 The functions of liability. A concise analysis of the private and criminal perspectives.

The problem of AI's liability is still under study and at the moment legal scholarship is quite divided on the potential application of traditional legal remedies as such to new technologies in general⁸¹. As far as the private and continental law theory, liability has mainly a restorative-compensatory and protective function: its objective is to give people some form of compensation (in money or by re-creating again the situation *ex ante*) about a loss they have endured either for a breach of contract or

⁷⁷ See *ex multis* G. SARTOR, *L'informatica giuridica e le tecnologie dell'informazione, Corso di informatica giuridica*, Torino, 2016; C. CASONATO, *Introduzione al biodiritto*, Torino, 2012; M. DURANTE, U. PAGALLO, *Manuale di informatica giuridica e diritto delle nuove tecnologie*, Milano, 2012.

⁷⁸ L. FLORIDI, *The 4th Revolution*, Oxford, 2014.

⁷⁹ A. M. TURING, *Computing machinery and intelligence*, in *Mind*, 236, 1950, 433-460; S. FRANKLIN, *History, motivations, and core themes*, in K. FRANKISH, W. M. RAMSEY (eds.), *The Cambridge Handbook of Artificial Intelligence*, Cambridge, 2014, 17-18.

⁸⁰ This profile has been examined also in § 1.1 of the present paper.

⁸¹ See among others J. LUZAK, *Digital Age: time to say goodbye to traditional concepts*, in *Journal of European Consumer and Market Law*, 4, 2018, 133-135; G. GUERRA *La sicurezza degli artefatti robotici in prospettiva comparatistica. Dal cambiamento tecnologico all'adattamento giuridico*, Bologna, 2018.

because of a tort⁸². The definition of what is a tort and how to calculate the corresponding compensation (monetary or not) can vary greatly across the different EU Member States and for time and space constraints it will not be detailed here. Socially, liability also allows people to program their future actions and, even in continental law countries, private law liability can, under some limited and strict circumstances, have a sort of punitive function⁸³.

Criminal law liability has also some proper features which are detailed as follows. As a matter of fact, the law can pursue a number of objectives by means of holding a person criminally liable. In brief, in this regard two main philosophies may be mentioned, namely utilitarianism and retributivism.

As for utilitarianism, it aims at maximizing the net happiness of people. From this perspective, a number of theories of punishment can be enumerated.

In the first place, general prevention, also known as deterrence, can be quoted. Under this theory, the suffering imposed on the convicted person for the crime they have committed is supposed to deter others from committing crimes in future, if they do not want to suffer the same treatment.

In the second place, one can refer to the theory of special prevention, that does have three meanings:

- i. particular deterrence, i.e. the punishment aims to deter the criminal themselves (rather than to deter others) from committing further crimes in future, by giving them an unpleasant experience they will not want to endure again;
- ii. rehabilitation, that is the punishment, by making the convicted person suffer an appropriate treatment, aims at rehabilitating them and returning them to society so reformed that they will not want (or will not need) to commit further crimes in future; and
- iii. restraint, namely by means of punishment imposed on criminals (and so, isolating them), the society aims at protecting itself from persons considered dangerous because of their criminal deeds.

The second philosophy concerning the criminal liability is the so-called retributivism. For retributivists, the most important – sometimes, even the only – purpose of criminal law should be to punish the morally culpable. In other words, the society would impose the punishment (that is, the infliction of suffering) to criminals in order to obtain a sort of revenge for their conduct. For the supporters of this theory, when one commits a crime, they should receive commensurate punishment so as to restore the peace of mind and repress the criminal tendencies of others. Furthermore, they affirm that retributive punishment is needed in order for the law to be respected, to suppress “private justice”⁸⁴.

⁸² *Ex multis* C. CASTRONOVO, *La Responsabilità Civile*, Milano, 2018; L. NIVARRA, V. RICCIUTO, C. SCOGNAMIGLIO, *Diritto Privato*, Torino, 2018, 455-461.

⁸³ Generally, liability damages in Civil law countries do not have a punitive function as the punitive damages in Common law. However, as far as, the applicability of punitive damages in Civil law countries see *ex multis* the Italian Court of Cassation judgment –United Chambers, 7 February- 5 July 2017, n. 16601 which admitted, under specific circumstances, the applicability of punitive damages while applying a US judgment in the Italian territory.

⁸⁴ See S. L. EMANUEL, *Criminal Law*, New York, 2015, 2. See also *amplius* W. R. LAFAVE, A. W. SCOTT, JR., *Substantive Criminal Law*, Vol. 1, St. Paul (Minnesota, USA), 1986, 30-40; G. P. FLETCHER, *Rethinking Criminal Law*, Oxford-New York, 2000, 414-419.

3.3.2 What changes when liability meets the AI? Practical examples and open issues.

Leaving aside the main problem of legal subjectivity which has been dealt with supra (see § 2.4), this paragraph will describe several practical examples⁸⁵ about how the new AI generated situations can question the fitness of legal liability rules in EU continental private law⁸⁶.

The main structural problem concerning liability both under private and criminal law perspectives is the scarce explainability of the functioning of automated decision-making algorithms, which is one of the main concrete applications of AI nowadays. The most effective algorithms have a black box mechanism and therefore only the input and output are known but the path connecting these two elements is often not readily knowable⁸⁷.

Hereafter we analyse two of the most famous examples concerning the unfair use of data, both against individuals. Firstly, there is the case of a person who wants to ask for a loan on the site of a bank that uses automated decision algorithms to calculate their eligibility for the loan. It has been shown that the results varied also in function of the race of the applicant, which was inferred from other data such as name and neighbourhood even if the applicant had assets and securities to sustain their application⁸⁸. An analogue case involved the British admission university tests in 2020⁸⁹. Due to the pandemic, the admission selection process used an algorithm that scored lower points to applicants from disadvantaged neighbourhoods. These episodes raise discrimination and data protection law problems but also private law liability problems. Is the damage in the cases mentioned just personal or also an economic one, or both? How to quantify it not knowing how the algorithms have worked with certainty? Furthermore, new objects and potentiated traditional objects (smart objects/IoT or robots) are being used more and more in our households and in healthcare. If we consider all the new robots for the house, or smart utilities management systems, or the healthcare robots and wearables as AI-powered objects *lato sensu*, what are the consequences on the product liability rules that have been in place (in

⁸⁵ This list does not aim to be complete but just to give the gist of the main problems concerning civil liability and AI.

⁸⁶ The questions about the criminal liability depending on the development and the use of AI-tools is a very controversial issue. In this regard, see *ex multis* F. BASILE, *Intelligenza artificiale e diritto penale: quattro percorsi di indagine*, in *Diritto penale e uomo*, 10, 2019, at https://dirittopenaleuomo.org/contributi_dpu/intelligenza-artificiale-e-diritto-penale-quattro-possibili-percorsi-di-indagine/ (accessed on 22/02/2021).

⁸⁷ *Ex multis* see L. HULSTAERT, *Black-box vs. White-box models*, in *Towards Data Science*, at <https://towardsdatascience.com/machine-learning-interpretability-techniques-662c723454f3> (accessed on 28/01/2021).

⁸⁸ Also, the kind of device and the kind of email used in the application could be the basis to infer elements about the solvability of the applicant. For more on this issue see A. KLEIN, *Credit Denial in the Age of AI (report)*, 2019, at <https://www.brookings.edu/research/credit-denial-in-the-age-of-ai/> (accessed on 28/01/2021). The US Federal Trade Commission (FTA) published some guidelines for any business using algorithm-based decision making in order to prevent discriminatory outcomes see <https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms> (accessed on 28/01/2021). The EU has a stronger protection with the General Data Protection Regulation (GDPR). There are both specific rights to have transparency concerning personal data processing (Article 13-14-15) and the possibility to opt out from an entirely automated decision-making process (Articles 21-22). See also in this regard the guidelines of the Council of Europe on AI and Discrimination, F. ZUIDERVEEN BORGESIU (CoE), *Discrimination, artificial intelligence and algorithmic decision making*, 2018, at <https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73> (accessed on 30/01/2021).

⁸⁹ *Ex multis* <https://www.bbc.com/news/uk-politics-53837722> (accessed on 28-01-2021).

the EU) for more than thirty years? Can the algorithms or the interface between software and the physical part of the object bear all the responsibility for a material damage to a person? Can this kind of liability be shared with the user who might have used the AI-powered object in an incorrect way? What will be the role of insurances on some products such as self-driving cars? Will the content of fault change? Will causality be evaluated differently? These are just some of the issues that are being discussed, at the moment, by legal scholars and policy makers and will likely provide the material for future legislation and regulation on the matter.

3.3.3 Policy and legal challenges concerning AI and Liability in the EU

The EU is trying to regulate the issue of AI civil liability under different points of view. Some important documents were published during the last year and a half on this issue⁹⁰. They all focused on four main aspects of contemporary AI decision making algorithms, such as: *i)* the opacity that the most efficient black box algorithms might have and its possible discriminatory outcomes; *ii)* whether the concepts of damage, causality and fault might be extended to either manufacturers-producers or software creators or both; *iii)* how insurances should tackle the damages that AI-powered objects might cause; *iv)* whether an AI system or powered objects can be held liable per se. Even though AI is still dealt differently compared to Robots and the Internet of Things (IoT), we might consider that the broader technological phenomenon of convergence will bring these three technologies closer together and that might help in creating at least some core rules of liability when dealing with machines with which we might have a meaningful agency relationship on a daily basis. Furthermore, in the near future, the EU products liability rules will most likely be changed and will affect most AI-powered consumers' objects⁹¹.

The EU does not have a clear competence to regulate enforcing aspects of AI liability yet. Given the sensitivity of this topic, EU Member States (MS) might also have to start by regulating on their own. For instance, Germany has just approved a law to regulate digital markets⁹², which rely heavily on AI algorithms, through competition law, thus preceding the much awaited EU New Tool for competition law proposal⁹³. Moreover, MS can autonomously investigate whether their tort or contractual liability

⁹⁰ EUROPEAN COMMISSION, *Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics*, COM (2020) 64 final; HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE (EU COMMISSION), *Guidelines on Trustworthy AI*, cit.; EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES-NEW TECHNOLOGIES FORMATION, *Liability for Artificial Intelligence and other emerging digital technologies*, 2019 at [https://bit.ly/3sxj1pi\(16/01/2021\)](https://bit.ly/3sxj1pi(16/01/2021)) (accessed on 22/02/2021).

⁹¹ See *ex multis* J. LUZAK, *A Broken Notion: Impact of Modern Technologies on Product Liability*, in *European Journal of Risk Regulation*, 3, 2020, 630-649.

⁹² B. FERRI, *The new German Competition Law tackling online platform supremacy – an attempt of balancing authoritative flexibility and legal certainty*, in *Medialaws. Law and Policy of Media in a Comparative Perspective*, 11 January 2021, at <http://www.medialaws.eu/the-new-german-competition-law-tackling-online-platform-supremacy-an-attempt-of-balancing-authoritative-flexibility-and-legal-certainty/> (accessed on 30/01/2021).

⁹³ Too read the opinions of the experts selected by the EU Commission about the new EU Competition Tool see https://ec.europa.eu/competition/consultations/2020_new_comp_tool/index_en.html (accessed on 22/02/2021).

are fit for AI. While doing that, MS might consider implementing ethical principles in processual remedies and tools for citizens who suffer material or immaterial damages from AI-powered systems or objects⁹⁴. This will connect with the GDPR principles of explainability and fairness as it is likely that the cause of damages most of the times will derive from an unfair or not transparent use of personal or inferred personal data. Nonetheless, explainability and transparency should be balanced with the principle of protection of trade and industrial secrets in IP law⁹⁵. Liability seems to have become a *fil rouge* connecting different civil law disciplines: not only traditional private law but also IP and competition law too. Automated decision-making algorithms could be the main authors of a work of art (for instance GAN neural networks or sophisticated data analytics techniques⁹⁶) or inventors. It can happen that a person would rather patent an innovation giving the AI system the inventor role also for liability reasons like in the Dabus-Thaler case⁹⁷. Ultimately, the evaluation of inventions and creativity (and the liability that this entails) must also take into account the infrastructure of AI and of digital markets: it must be avoided that just a few corporations own several digital markets and collude together⁹⁸. Breaching competition law thus becomes a cause for liability which is enforced by public interest subjects such as the European Commission DG Comp and the European Network of Competition authorities (ENC). Competition law infringements on digital markets are not just a competition distortion but could also be an obstacle to good innovation: if the market is dominated just by a few companies, innovative start-ups with less means and access to the market will have basically no incentive to invest in R&D and create innovative products and technological applications. This could be avoided by giving incentives to companies to be more transparent (e.g. in the creation of technological standards), by exercising either soft-law or regulatory checks ex-ante and ex-post, and by also considering the values of trust and accountability as part of a broader meaning of the term liability when dealing with AI.

3.3.4 A regulation fit for AI-risk. An example of technological risk-management from the US

Transparency, control, responsibility: these are the three cornerstones of what we mean as AI-risk management. As a matter of fact, the rise of AI in almost every side of our lives (from simple mobile-apps to advanced healthcare-software) implies the need for some serious consideration about the related benefits and risks.

⁹⁴ L. FLORIDI ET AL, *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds and Machines*, 4, 2018, 689-707.

⁹⁵ For some interesting considerations about why confidentiality is important in IP judicial proceedings (especially with regard to FRAND clause) see H. TSILIKAS, S. MAKRI, *Confidentiality and transparency in FRAND litigation in the EU*, in *Journal of Intellectual Property and Practice*, 3, 2020, 173-184.

⁹⁶ See on this point *The Next Rembrandt Project*, at <https://www.nextrembrandt.com/> or *This person does not exist*, at <https://thispersondoesnotexist.com/> (both accessed on 16/01/2021).

⁹⁷ G. NOTO LA DIEGA, *Brevetti e Intelligenza Artificiale*, statement at the Workshop “Workshop neuroni artificiali e biologici: etica e diritto”, 3-4 December 2020, University of Trento.

⁹⁸ M. BURREAU, A. DE STREEL, *Digital Conglomerates and EU Competition Policy*, in *SSRN Electronic Journal*, 2019, 44; N. PETIT, *Technology Giants, the M oligopoly Hypothesis and Holistic Competition: A Primer*, 2016, doi: 10.2139/ssrn.2856502.

From this point of view, we believe that the European Union should look at the United States, where an interesting model of assessment of technological risks is in force. If properly adapted to the European Union context and enhanced, it might be a useful reference to establish an AI-risk assessment paradigm at the supranational level.

The aforementioned US model is based on the central role of federal agencies, which operate within the Executive Branch, e.g. the Food and Drug Administration (FDA), the Occupational Safety and Health Administration (OSHA), and the Environmental Protection Agency (EPA). Several characteristics are typical of the various agencies, yet a few common elements can be identified⁹⁹.

i. The agencies do have the power (originating from the law, e.g. acts of the Congress) to provide for the rules that must be respected by the organisations operating in a market's sector (e.g. pharmaceutical companies), with the aim of reducing the risks related to a certain dangerous activity within a risk-margin that is considered socially adequate.

ii. The adoption of the aforementioned precautionary rules is democratically legitimated (e.g. the Congress approves them, or – in the case of the pharmaceutical market – the President of the United States appoints the FDA's Commissioner with the advice and consent of the Senate¹⁰⁰).

iii. In the light of Section 10 of the Administrative Procedure Act of 1946 (APA), the federal courts of appeal do have jurisdiction on the rationality of the agencies' decisions, i.e. they may be invalidated for example where arbitrary or capricious, or not supported by substantial evidence.

iv. A rigid and preventive law enforcement system is provided: e.g. agencies may make inspections of the companies and address the Judiciary Branch to obtain an injunction to the companies themselves when there are grounds to believe that they do not comply with the abovementioned precautionary rules.

v. An education and compliance system is provided, with the aim of "educating" the companies, or "re-educating" them where some violation is found out: e.g. companies can ask for the agency's consultation assistance in arranging an adequate set of cares able to reduce the risk of harmful events.

vi. Besides, criminal law provisions seek to assure the law enforcement by means of the threat of sanctions, where serious violations of the aforementioned rules are committed.

The EU should take inspiration from such a framework. In this regard, there might be two scenarios: either to establish a new EU agency specifically committed to manage the risks related to the various applications of AI, or to set up new offices within the existing agencies dedicated to the same task.

⁹⁹ See *ex multis* H. L. PITT, K. A. GROSKAUFMANIS, *Minimizing Corporate Civil and Criminal Liability: A Second Look at Corporate Code of Conducts*, in *Geo. Law Journal*, 1990, 1559; S. JASANOFF, *Science at the Bar: Law, Science, and Technology in America*, Cambridge (MA) – London, 1995, 69-92; S. A. SHAPIRO, R. L. GLICKSMAN, *Risk Regulation at Risk: Restoring a Pragmatic Approach*, Stanford, 2003; F. STELLA, *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime*, Milano, 2003, 390-391 and 579-586; F. CENTONZE, *La normalità dei disastri tecnologici: Il problema del congedo dal diritto penale*, Milano, 2004, 400-410.

¹⁰⁰ FDA Act, 1988, 21 U.S.C.A. § 393.



3.4. AI and the promotion of a constitutionally oriented approach

The last relevant factor in the creation of a reference framework for a trustworthy AI concerns the dimension of the human rights, freedoms and values involved by the application and the deployment of the examined artificial systems.

As it has been so precisely described in this contribution, the development of this new kind of promising technologies could have problematic consequences on many different aspects of human daily life, moving from the possible risk of discriminatory actions, going through the questions regarding liability and personality profiles of artificial systems, to the eventuality that the use of AI may adversely affect people's right to self-determination.

From this point of view, it is evident that there is a real risk that an improper employment of AI technologies could entail a serious violation of the fundamental rights and freedom, as recognised by the contemporary legal systems. And this possibility would undermine all the benefits and the new opportunities that, also in terms of protection and promotion of the fundamental rights and freedoms¹⁰¹, are strictly connected to the AI development.

With the aim of avoiding such a similar situation, it is necessary and fundamental to support the adoption of a constitutionally oriented approach in the development of these systems and, especially, in providing for a possible regulatory framework in this field. This type of proposal is based on two main reasons.

Firstly, the risks associated with an unfair use of AI systems are such as to draw attention to the original vocation of constitutionalism, that is the real and concrete limitation of powers in function of an effective guarantee and protection of rights¹⁰².

Secondly, the affirmed constitutional principles, as recognised in most of the contemporary Constitutional Charters, may disclose the proper pathway to promote the application of these technologies and to protect people from their reckless use¹⁰³.

Following this constitutionally oriented approach, it may be easier to identify the right balancing point in setting up an appropriate legal framework for the application and the spread of a trustworthy AI in the contemporary society, achieving the intended goal of maximising the benefits and minimising the risks linked to this technology.

¹⁰¹ For example, the use of AI systems in the medical field, if properly used, could be an important tool to ensure and guarantee greater protection of the right to health, promoting, at the same time, a more human model of medicine. This thesis is exhaustively expressed in E. J. TOPOL, *Deep Medicine. How Artificial Intelligence can make healthcare human again*, New York, 2019.

¹⁰² These aspects are effectively underlined in C. CASONATO, *AI and Constitutionalism: The Challenges Ahead*, in B. BRAUNSCHWEIG, M. GHALLAB (Eds.), *Reflections on Artificial Intelligence for Humanity*, Cham, 2021, 127-149; A. SIMONCINI, *Sovranità e potere nell'era digitale*, in T. E. FROSINI, O. POLLICINO, E. APA, M. BASSINI (eds.), *Diritti e libertà in Internet*, Milan-Florence, 2017, 19-38; O. POLLICINO, *Forum: Law and Artificial Intelligence. L'impatto dell'AI sul diritto e sui diritti*, in *BioLaw Journal – Rivista di BioDiritto*, n.1, 2020, 491-492; A. SIMONCINI, *Forum: Law and Artificial Intelligence. L'impatto dell'AI sul diritto e sui diritti*, in *BioLaw Journal – Rivista di BioDiritto*, n.1, 2020, 498-501; O. POLLICINO, *L' "autunno caldo" della Corte di giustizia in tema di tutela dei diritti fondamentali in rete e le sfide del costituzionalismo alle prese con i nuovi poteri privati in ambito digitale*, in *federalismi.it*, 19, 2019, 12.

¹⁰³ Cfr. C. CASONATO, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, cit., 377-390.

In more specific terms, a suitable instrument to properly balance the constitutional values, affected by the concrete applications of the AI systems, could be the provision of an assessment mechanism which determines the level of AI's impact on fundamental rights and freedoms¹⁰⁴. Indeed, such a system would make it possible to provide for increasingly stronger legal guarantees as the effects produced by these artificial systems are such as to seriously affect the people concerned by the use of this technology. This kind of impact-assessment tool could, for example, be relevant in understanding how to regulate the use of AI systems, ensuring their application when they are able to guarantee a more effective protection and a more actual implementation of the rights and of the constitutional safeguards as affirmed in the contemporary legal systems. At the same time, this impact-assessment system should provide for stricter mechanisms and safeguards if the AI's application could seriously undermine the aforementioned rights and principles or it could offer a lower level of protection than the usually provided one.

4. Conclusive remarks

At the end of this document on an extremely complex subject, many existing challenges and opportunities have emerged. Even though there are many benefits, the emphasis has been placed mainly on the challenges and problems that these issues present because it is in those spaces that the awareness and action of scholars and society as a whole must occur.

Thus, in Section 1.1 and 1.2, an attempt was made to give a definition of AI, despite the vagueness of the term. Indeed, many are the meanings attributed to it, but a shared definition is essential to addressing the problems that arise from AI, including, but not exclusively, those of regulation. It is also for this reason that it is useful to focus on the practical applications of AI. Therefore, from a technological perspective, an attempt has been made to expose some applications in the field of Human Enhancement and to observe the frontiers opened by cerebral organoids. In addition, from a philosophical point of view, the need for a shared definition has emerged. Technology in general poses difficulties in the use of language because it gives rise to purely interdisciplinary problems and raises ethical issues (e.g. what is a human being, what is enhancement). Common definitions make interdisciplinary dialogue possible, and, at the same time, interdisciplinary dialogue constitutes a fertile ground for common definitions.

Section 2 was dedicated to investigating the main challenges. Firstly, it was possible to show in 2.1 how the barrier between the natural and the artificial has thinned, rendering it difficult to clearly distinguish between the two dimensions, which now appear to be in close communication. These developments certainly require a different approach in human control. Therefore, a *human in the loop* approach, which is effective and does not run the risk of becoming accustomed to machine delegation, is necessary. Secondly, the social issues related to the diffusion of AI in society were addressed in 2.2: repercussions on the labour market, the risk of bias and prejudices in automated decisions and therefore of algorithmic discrimination, the problem of algorithmic opacity and finally, the risks related to privacy

¹⁰⁴ Such a similar mechanism is provided by the Canadian Directive on Automated Decision-Making issued by the Treasury Board of Canada Secretariat, which has taken effects on the 1st April 2019. The full text of this Directive is accessible at <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592> (accessed on 22/02/2021).

and data protection. Thirdly, 2.3 considered some essentially open ethical issues related to AI, specifically: the questions concerning the limits in the development of AI and the resulting limits on human freedom; the question of what degree of autonomy to guarantee to systems which are able to decide without human control, and whether we are willing to attribute agency to an AI system; the questions of authenticity and self-perception of the human being who has undergone artificial modifications. In this respect, it is necessary to understand what idea of human beings underlies the development of AI. Lastly, legal issues related to the recognition of legal personhood to artificial agents both in civil and criminal law were investigated in 2.4.

Finally, in Section 3, a pathway for a trustworthy use of AI was discussed and analysed, and solutions were offered, considering that a framework is needed to balance the risks and benefits of AI so as not to restrict human freedom or limit technological progress in an unjustified way. First of all, 3.1 encouraged an attempt to make the whole of society, and not just a few insiders, aware of the risks and benefits of AI, e.g. regarding Human Enhancement, by ensuring a bottom-up approach. In order to do this, as expressed in 3.2, we need shared categories and definitions so as to adopt an approach that is as interdisciplinary as possible, avoiding rigid categorisation of disciplines that do not allow us to grasp reality in all its complexity. For this reason, it is necessary to abandon preconceived notions without losing the advantages of specialised knowledge, aiming at an interdisciplinary education. In 3.3 the problem of liability was discussed from a theoretical and practical perspective, focusing on: the functions of liability in private and criminal law; the examination of some practical examples concerning the unfair use of data; and changes at the European Union level. Finally, since the risk is that of an improper use of AI which is detrimental to human rights, freedom and values protected by contemporary legal systems, a constitutionally oriented approach to AI should be promoted, as suggested in 3.4. In particular, in order to identify the right balancing point for appropriate legal regulation, an impact-assessment mechanism of AI technologies to fundamental rights and freedom could be provided.

In conclusion, in the near future it will be inevitable to think about the main challenges that AI will generate and that have been envisaged in this paper. Thus, the social and legal issues arising from the widespread and pervasive use of AI in the society will have to be addressed with the essential aim of ensuring human control over artificial systems. In addition, the ethical issues that remain essentially open should not be underestimated. This will require the development of shared definitions that also take account of the practical applications of AI, the use of an interdisciplinary approach to AI, and the involvement of society as a whole. Reflection on these lines of research, also by means of a constitutionally oriented approach that guarantees the protection of fundamental rights and freedom, will make it possible to gradually build a pathway for a trustworthy use of AI.