The Possible Relationships Between Law and Ethics in the Context of Artificial Intelligence Regulation

Maria Cristina Gaeta [*]
mariacristina.gaeta@unisob.na.it

Livia Aulino [*]
livia.aulino@gmail.com

Emiliano Troisi [*]
emilianotroisi@gmail.com

ABSTRACT

The latest academic discussion has focused on the potential and risks associated with technological systems. In this perspective, defining a set of legal rules could be the priority but this action appears extremely difficult at the European level and, therefore, in the last years, a set of ethical principles contained in many different documents has been published. The need to develop trustworthy and human-centric AI technologies is accomplished by creating these two types of rule sets: legal and ethical. The paper aims to critically analyse and compare these rule sets in order to understand their possible relationships in the regulation of legal problems, not only theoretically but also, where present, in some practical applications of AI, such as self-driving cars, smart toys, smart contracts and legal design. Indeed, the purpose is to identify how legal rules and ethical principles can interact for adequate regulation of AI, with particular regard to the fields of application that will be analysed.

1. The need for regulating Artificial Intelligence through the creation of two types of rule sets: legal and ethical.

The paper is based on the legal and ethical challenges arising from the development of new technologies, with particular regard to AI. This technological development is very fast, and it has rapidly accelerated following the COVID-19 pandemic, as AI technologies are used to fight the virus but also to allow humans to carry out the activities of daily life, albeit at distance (e.g. smart working, online meeting, and so on).

[*] University Suor Orsola Benincasa, Napoli, Italy.

As it will be underlined, the new technologies development has not only entailed benefits for the human being but also multiple risks ranging from a data breach to a cybersecurity breach, as well as a physical injury or psychical manipulation. For this reason, it is necessary to intervene by regulating AI to limit the legal and ethical risks for human beings related to the use of technological systems and, instead, to fortify their benefits.

The starting point of the paper is the relationship between law and ethics in the AI domain, where the interaction is of great relevance.

Legal positivists are used to study and apply binding law but, currently, the problem of identifying the applicable law regarding AI application has highlighted the need to resort to other sources of regulation, such as ethical principles or standards. This need is due to the lack of *ad hoc* binding legal regulation of AI technologies and the proliferation of ethical principles with the function of guiding individuals, companies, and institutions. Ethics does not provide binding legal rules, but has social and moral value and is of particular importance in relation to new technologies for two reasons: (i) firstly because of the impact of AI on human beings; (ii) secondly, because initially (before 2021) the European legislator was not ready to regulate the phenomenon through binding regulation able to really understand new technologies and being able to keep up with its development. During the last two years (2021 and 2022), instead, the European Union started to translate these ethical principles into legal rules, that now constitute regulatory proposals but which it is hoped will become binding rules in the near future. Ethical principles would remain to complement and supplement the law but would largely be transposed into these binding rules. In this way, the ethical principles implemented in European regulations would become binding rules.

The legal system is not incomplete, but it cannot regulate every single phenomenon, being possible to apply the principles of *analogia legis* or *analogia iuris*, or just the extensive interpretation of the existing legislation (according to art. 12 of the preliminary provisions to the Italian civil code). In this context, ethical principles can be a criterion for a complete interpretation of the law and should be compliant with hard law (*secundum legem*).

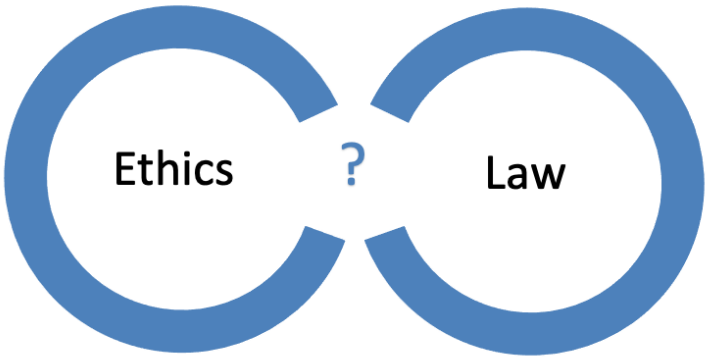


Figure 1: The difficult relationship between law and ethics

Concerning the AI regulation, as has been already anticipated, the introduction of ethical principles has proceeded that of the legal ones.

The International Institutions, the European Union, and its Member States have always been founded on the values of human dignity, equality, solidarity, freedoms, and respect for human rights,[1] and this approach is being applied also in relation to robotics and artificial intelligence. The legislators are now called to intervene to guarantee the protection of human beings in relation to new technologies also with binding rules. Indeed, the technology is not neutral and, in addition to entailing a series of advantages, it also implies serious risks for individuals who have a structural condition of vulnerability in relation to them.[2] Therefore, new technology, especially when equipped with AI, must be properly regulated, to put these new technologies at the service of humans.

In this light, at all levels of regulation, the awareness of the need for a concrete strategy and regulation of AI has matured. In particular, in May 2019, the Council of Europe adopted some recommendations aimed at maximizing the potential of AI systems and preventing a negative impact on human rights.[3] Its

[1] Universal Declaration of Human Rights of the United Nation (UDHR) of 10 December 1948; European Convention on Human Rights (ECHR) of 4 November 1950; Charter of Fundamental Rights of the European Union of 18 December 2000 (FREU); National Constitutional Charters.
[2] Gatt L. (2022), Legal anthropocentrism between nature and technology: the new vulnerability of human beings. *EJPLT* 1, 15 ff.
[3] *See* in particular Unboxing artificial intelligence: 10 steps to protect human rights, 2019.

primary interest is to verify the impact of AI systems on human rights in the public and private sectors, by imposing on Member States to guarantee human rights through accurate information, transparency, and independent and effective oversight of technology compliance, but without creating obstacles to the identification of liabilities and remedies in case of the violation of these human rights.

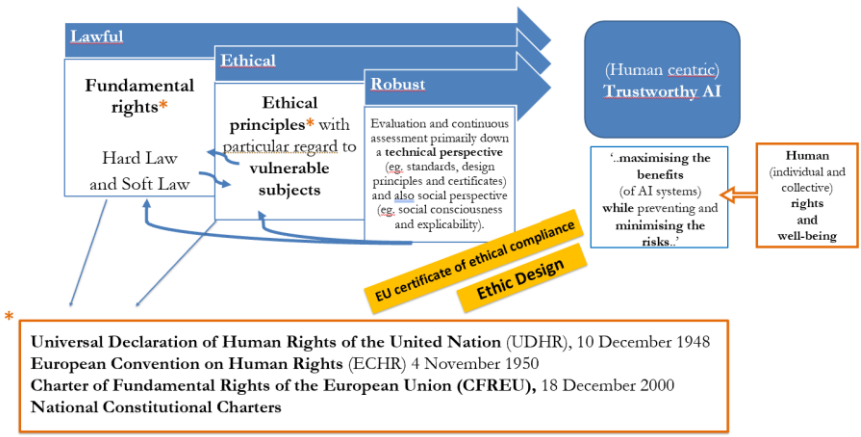


Figure 2: The characteristics that AI should have.

In the same direction, the Council of Europe appointed an *ad hoc* committee for Artificial Intelligence (CAHAI), in September 2019, that aims to assess the impact of AI on the individual and the society, as well as on existing soft law and hard law instruments that deal with AI. In this light, in December 2020 the CAHAI published a feasibility study that examines,[4] based on broad multi-stakeholder consultations, the potential elements of a legal framework for the development, design and application of AI, based on Council of Europe standards in the field of human rights, democracy and the rule of law.

Also, the EU initiatives have been developed over the last six years, from the European Parliament Resolution on civil law rules on Robotics,[5] until

[4] Ad Hoc Committee on Artificial Intelligence (CAHAI) feasibility study, 17.12. 2020.
[5] European Parliament resolution of 16 February 2017 on Civil Law Rules on Robotics (2015/2103(INL)

today. The protagonists of the debate are the European Parliament and the European Commission. While the former seemed certain from the outset of the need to intervene to provide specific rules for automation and AI, the European Commission initially considered completely effective the existing regulation also applied to new technologies and then evaluated the possibility of intervening to regulate this phenomenon (AI Act Regulation Proposal,[6] AI liability Directive Proposal,[7] Proposal for the revision of the PLD[8]).

2. The dialectic between legal rules and ethical principles in the field of Embedded Artificial Intelligence.

2.1. Artificial Intelligence embedded in Robotics.

In the 21st century, the impact of new technologies is enormous, as they have changed the way of life of human beings, from personal relationships to work activities. In this scenario characterised by the quick evolution of technologies, Internet development has played a central role enhanced by the extension of the network to the world of objects that become 'smart'[9] (Internet of Things – IoT or, more properly now, Internet of Everything - IoE).[10] Human beings are constantly monitored through the growing number of identification and tracking

[6] European Parliament and Council Proposal for a Regulation on AI (Artificial Intelligence Act), 3.11.2022.

[7] European Parliament and Council Proposal for AI Liability Directive, 28.09.2022 COM (2022) 496 final.

[8] European Commission Proposal for PLD Revision, 28.09.2022, COM (2022) 495 final.

[9] Smart things are tagged with a Radio Frequency Identification (RFID) tag with a single ID called Electronic Product Code (EPC). About RFID *see* Pallone E.K. (2016), "Internet of Things" e l'importanza del diritto alla privacy tra opportunità e rischi. *Ciberspazio e diritto*, 17(55), 174 ff.

[10] With regard to the introduction of the term Internet of Things Ashton K. (2009), That "Internet of Things" Thing. In the real world, things matter more than ideas. *RFID J*, 1; Haller S., Karnouskos S., Schiroh C. (2008), The Internet of Things in an enterprise context Future Internet, *Lecture Notes in 5468 Computer Science*, 1. About Internet of Things definition, instead, *see* Ziegler S. ed. (2019), Internet of Things Security and Data Protection, Springer; Noto La Diega G., Walden I. (2016), Contracting for the 'Internet of Things': Looking into the Nest. *Queen Mary School of Law Legal Studies Research Paper*, 219; Peppet S.R. (2014), Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent. *Texas Law Review*, 93, 85 ff.; Weber R.H. (2010), Internet of Things, New security and privacy challenges. *Computer law & security rep*, 23 ff.

technologies and, in some cases, their behaviour is already influenced by the (ab)use of smart devices.

Nowadays, included in this category are incredibly wide kinds of smart robots developed in different sectors which include autonomous vehicles in the smart mobility sectors, as well as smart toys in the recreational sector. These concrete applications will be analysed in the following subparagraph, with particular regard to legal and ethical risks and their regulation.

2.2. The potential of self-driving cars and their issues.

Among the protagonists of this ever-changing scenario, there are undoubtedly driverless cars.[11] The first evolution in the direction of car automation was based on the development of Advanced Driver Assistance Systems (ADASs). They are those technologies that collect data on the performance of the car and the space-time context of its circulation, informing the driver and reaching up to make suggestions to the driver or, even, taking partial or total control of the vehicle. Today, self-driving cars are classified on the base of automation levels, that at the higher one is equipped with AI. The most used classification is that of the SEA International standard J3016,[12] which has defined six different automation levels, based on vehicle automation and human-machine interface (HMI). More in detail, the degree of driver's intervention in driving activities decreases proportionally as the vehicle's automation increases.

[11] On self-driving cars legal regulation *see* Bertolini A., Riccaboni M. (2020), Grounding the case for a European approach to the regulation of automated driving: the technology-selection effect of liability rules. *Eur J Law Econ*, 1 ff.; Al Mureden E. (2019), "Autonomous cars e responsabilità civile tra disciplina vigente e prospettive de iure condendo. *Contr. impr.*, 2019, 3, 895 ff.; Ruffolo U., Al Mureden E. (2019), "Autonomous vehicles" e responsabilità nel nostro sistema ed in quello statunitense. *Giur. It.*, 7, 1657 ff.; Albanese A. (2019), La responsabilità civile per i danni da circolazione dei veicoli ad elevata automazione. Europa e diritto privato, 4, 995 ff.; Davola A., Pardolesi R. (2017), In viaggio col robot: verso nuovi orizzonti della r.c. auto ("driverless")?. *Danno e responsabilità*, 5, 616 ff.. Please, allow me to refer to Gaeta M.C. (2019), *Liability rules and self-driving cars: The evolution of tort law in the light of new technologies*. Editoriale Scientifica Italiana (ESI); Von Bodungen B., Caggiano I.A., Steege H., Gaeta M.C. (2023), *European regulation for self-driving cars*. Springer (forthcoming).

[12] SAE J3016 'Recommended Practice: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles' (lastly amended in 2021), well know as 'SAE Levels of Driving Automation'. They are: (L0) no automation, (L1) driver assistance, (L2) partial automation, (L3) conditional automation, (L4) high automation e (L5) full automation.

The introduction of autonomous vehicles on the market entails considerable advantages, but the production of such cars with high technology brings serious legal and ethical issues that need to be solved, as current regulation does not always appear adequate.

More precisely, on the side of the legal issues, it is difficult to identify the tortfeasor in the case of damage due to highly (semi)automated driving. Furthermore, the difficulty in establishing clear and precise liability rules is reflected in the difficulty in evaluating how the insurance policies should be developed.[13] In addition, damages arising from a data breach as well as a cybersecurity breach of the ADAS, are very concrete.[14]

Regarding ethical issues, instead, there are several issues concerning both the experimentation phase of autonomous vehicles and the phase of placing them on the market.[15] As for the experimentation phase, it is important that the same takes place respecting human rights and protecting individuals who play the role of drivers of these (semi)autonomous vehicles tested on the road, putting their life, or at least their safety, at risk. About the application phase, the decision-making process that has raised most discussions concerns the operating contexts of the self-driving car, known as the 'Trolley problem'.[16] The phenomenon concerns a case in which, given the imminent collision of the autonomous vehicle and the loss of human life as a result, the AI system has to make a choice that involves the necessary sacrifice of some victims in favour of the most

[13] *See* European Parliament in-depth analysis, Artificial Intelligence in road transport, PE 654.212, January 2021; European Commission Report, The safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, 19 February 2020, COM 2020/64 final; European Parliament Study, A common EU approach to liability rules and insurance for connected and autonomous vehicles, PE 615.635, February 2018; European Parliament Report, Autonomous driving in European transport (2018/2089(INI)), 5 December 2018

[14] *See* European Union Agency For Cybersecurity (ENISA), Cybersecurity challenges in the uptake of artificial intelligence in autonomous driving, EU30568EN, 2021.

[15] *See* Horizon 2020 Commission Expert Group to Advise on Specific Ethical issues raised by driverless mobility Report, Ethics of Connected and Automated Vehicles: recommendations on road safety, privacy, fairness, explainability and responsibility, 2020, that is the UE report on self-driving cars more relevant ethical issues and individuate specific ethical principles to solve or, at least, mitigate these risks. On Trustworthy AI for the specific domain of Autonomous Vehicles, see also also Fernandez Llorca, D. and Gomez Gutierrez, E., Trustworthy Autonomous Vehicles, EUR 30942 EN, Publications Office of the European Union, Luxembourg, 2021.

[16] Jarvis Thomson J. (1985), The Trolley Problem. *Yale Law Journal*, 94 (6), 1395–1415. On the trolley problem applied to self-driving cars, the MIT has carried out research study known as moral machine: www.moralmachine.net

acceptable outcome from a social point of view. This scenario is the result of a theoretical analysis, because in concrete terms issues of this type have not yet occurred as there are no fully autonomous vehicles on the road and the driver must always be able to resume the control of the car, but it results very complex to establish if and how to program the driving algorithm to make it face these dilemmas in the future.

These risks have found initial mitigation or even remedy in some legal systems where existing legislation has been amended to also include (semi)autonomous vehicles or a specific legislation on autonomous driving has been introduced. This legislative evolution happened primarily in some of the United States of America,[17] but also in the United Kingdom[18] and some European Member States such as Germany[19]. Whereas in other legal systems, such as Italy,[20] there is still a tendency to apply existing legislation extensively. In these EU States, the Product Liability Directive, as implemented in the respective national legal systems, applies also to autonomous vehicles. When the revised version of PLD will come into force, together with the AI Act and the AI Liability Directive,[21] the regulations will also be implemented also in the national regulation.

The already existing regulations of autonomous vehicles as well as the proposed AI regulations are examples of the introduction of binding rules that partly incorporate the ethical principles of AI and Robotics. These binding legal regulations on autonomous vehicles that have come, or will come, into force provide for liability rules from which specific sanctioning regimes and insurance

[17] Concerning the USA regulation on self- driving cars on the website of the National Conferences of State Legislature is possible to verify the state of the art on the legislations in force and the bills: https://www.ncsl.org/transportation/autonomous-vehicles.

[18] In the United Kingdom, Automated and Electric Vehicles Act (AEVA) of 2018 regulated for the first time the phenomenon. The Act gained Royal Assent on the 19th of July 2018 and, then, came into force on the 21st of April 2021. UK Government published in November 2023 a draft Automated Vehicles Bill that contains amendments to the Automated and Electric Vehicles Act of 8 November 2023 (HL Bill 1) availabe at https://bills.parliament.uk/publications/52900/documents/3973

[19] In Germany, regulations concerning conditional vehicle automation were added to the Federal Road Traffic Act (StVG) in 2017, and subsequently legal provisions for highly automated vehicles were introduced in 2021.

[20] In Italy, for product liability is applicable the Italian Consumer Code, D.lgs. 6 September 2005, n 206, OJ 235, that currently implemented the PLD after the repealing of the d.P.R., 24th May 1988, n. 224 by the same consumer code.

[21] See note no. 6, 7 and 8.

schemes are derived. At the same time, ethical principles that have not been expressly transposed in the hard law regulation remain relevant in terms of their non-binding effectiveness in ethical norms, as well as on a social and moral domain, to guide members of society on levels other than legislation.

2.3. Smart Toys: characteristics and risks

Another field of application of AI embodied in robotics is that of smart toys[22] that are toy robots, sometimes equipped with AI, connected online in the IoT, also known as the Internet of Toys (IoToys). Smart toys are devices with the appearance of traditional children's toys, but they are capable of interacting with the children and, in general, with the surrounding environment, connecting through the Internet and using technological systems that range from sound systems (e.g. microphones) to visual systems (e.g. cameras), and the multiple types of sensors of movement or localization, with which they are equipped.

Considering the fast and constant evolution of AI, machine learning and big data[23] analytics, in all their applications, these connected toys will continue to increase and evolve quickly bringing with them advantages but also serious risks.

[22] About smart toys please allow me to refer to Gaeta M.C. (2020), 'Smart toys and minors' protection in the context of the Internet of Everything, 2 European Journal of Privacy Law &Technologies EJPLT, 2, 118 ff.

[23] A definition of big data is provided by De Mauro A., Greco M., Grimaldi M. (2016), A Formal definition of Big Data based on its essential features. *Library Review*, 63(3), 122 ff. Down a legal point of view with particular regard to data protection egal issues, *see* Council of Europe, Consultative committee of the convention for the protection of individuals with regard to automatic processing of personal data, Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data, 23 January 2017, T-PD(2017)01; Italian Competition Authority, Italian Authority for Communications Guarantees and Italian Data, Guidelines and policy recommendations for Big Data, July 2019. Literature on the topic: Mantelero, A. (2017), Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework. *Computer Law and Security Review*, 584 ff.; Mantelero, A. (2019), La privacy all'epoca dei Big Data, in Cuffaro, V., D'Orazio, R., Ricciuto, V. (eds) *I dati personali nel diritto europeo*, 1181 ff.; D'Acquisto, G., Naldi, M. (2018), *Big Data e Privacy By Design*, Giappichelli; Mantelero, A. (2019), La gestione del rischio nel GDPR: limiti e sfide nel contesto dei Big Data e delle applicazioni di Artificial Intelligence, in Mantelero, A., Poletti, D., (eds), Regolare la tecnologia: il Reg. UE 2016/679 e la protezione dei dati personali. Un dialogo fra Italia e Spagna. *Pisa University Press*, 289 ff.; Mantelero, A. (2012), Big Data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo. *Dir. Inf.*, 1, 135 ff..

Although smart toys are fun and sometimes even educational games or toys capable of monitoring children's vital functions, they are still tools that collect, process and communicate information and personal data, with consequent possible legal risks, especially for minors.[24] Indeed, about toys connection, there is a concrete risk of unlawful processing of minors' personal data, but also physical and psychological damages as a result of hacker attacks or malfunctioning of the smart toy.

Furthermore, ethical risks related to possible emotional bonds between the child and the toy can be configured. More precisely, as seen in the Hello Barbie[25] and Fisher-Price 'Smart Toy' examples, it could be possible to hack the IoToy and talk with the child, forcing him or her to fall into a trap. Furthermore, considering that smart toys are physical objects, with the aspect of an animal or a friend, they can easily influence vulnerable subjects such as children (i.e. minors).[26] Minors are particularly vulnerable subjects who can most easily be affected by a bad use of technologies that could lead to an alienation from real life, or, in the worst cases, could lead to make esteem acts that could even lead to the loss of life (psychical manipulation). Indeed, children do not yet have a well-developed decision-making capacity. In this context, the emotional bonds between children and smart toys could be a serious ethical issue. Therefore, it is necessary that minors, albeit digital natives, can understand technologies and their use, as well as distinguish what is human from what is not, without creating emotional relationships with these kinds of toys that are unable to return the feelings.

[24] About privacy and smart toys *see* Milosevic, T., Dias, P., Mifsud, C., Trultzsch-Wijnen, C.W. (2018), Media Representation of Children's Privacy in the Context of the Use of "Smart" Toys and Commercial Data Collection, *Media Studies*, 9, 26 ff..

[25] *See* Manta, I.D., Olson, I.S. (2015), Hello Barbie: First They Will Monitor You, Then They Will Discriminate Against You. Perfectly. *Alabama Law Review*, 67, 135 ff.. From a comparative point of view, *see* Fantinato, M., Hung P.C.K. (et al.) (2018), A preliminary study of Hello Barbie in Brazil and Argentina. *Sustainable Cities and Society*, 40 83 ff.. More recently, starting from 2017, a new Barbie model, Barbie hologram, was designed but no studies have yet been published on it.

[26] In the same direction, there are also examples of psychical manipulation through videogames. The so-called 'Blue Whale' and 'Jonathan Galindo' games are two examples. Even though are not smart toys, they are both online games where teenagers decide to participate voluntarily following an established procedure. The serious risk of those games is that the teenager could fall into a trap on the web, forced by someone to kill himself, perhaps to save the life of his family or to save himself.

A recent study by the European Joint Research Centre affirms that 'any robot with moving physical parts poses a risk, especially to vulnerable people such as children and the elderly'.[27] This is more and more true in the case of smart toys: in fact, they have very strong ethical implications, due to the possible emotional bonds between the child and the toy.

For these reasons, legal and ethical challenges arising from the development of new technologies impose to pay attention on the need for *ad hoc* rules for AI, as a specific regulation focused on smart toys is not required. It is undoubtable that the existent ethical principles integrate the legal regulation of these phenomena, but legal regulation is needed, and the EU Proposals of 2021 and 2022 go in this direction. In this regard, ethical principles can strengthen legal regulation, also endowing it with moral and social values, intervening before the law, when it appears backward and inadequate, but hard law regulation is of fundamental importance and the ethical principles can in part be implemented in binding regulation.

Although there is no specific hard law regulation of the Smart Toys phenomenon, the new EU regulatory proposals on AI, which include smart devices, appear to apply to smart toys as well. Furthermore, currently, there are regulations on some of the mentioned IoToy's legal risks that are also applicable to them (as well as to the already analysed autonomous vehicles). In the case of data protection regulation, for example, the so-called GDPR[28] also applies to data processing through Smart Toys.

3. AI and automated contracting: ethical concerns. An overview

The emergence of technologies characterised by the use of AI systems has ushered in a new season of debate on the main ethical, social and legal issues surrounding the use and consequences of the use of such technologies. This section - structured in this brief introduction and the following four subsections - focuses on the main ethical questions raised by the application of AI systems to the

[27] European Commission, Joint Research Centre, Technical Reports, Kaleidoscope on the Internet of Toys, 2017.

[28] Regulation of the European Parliament and of the council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), of 27 April 2016, Reg. UE/2016/679.

field of private law, and in particular to the automated (algorithm-driven) negotiation, formation and execution of contracts (i.e., *automated contracting*).

The technologies in question are the so-called Automated Decision-Making systems, i.e. - broadly - any process that enables, using technological means, decision-making without, or at least with irrelevant, human involvement. Such a definition, therefore, does not imply - being broader - but clearly includes the use of AI-based technologies[29], as more generally any computer technique that - relying on algorithms[30] - enables repetitive tasks to be performed with data without the need for constant human guidance.[31] Those systems, in other words, are capable of collecting data from a certain database or environment, interpreting it, and - in light of the goal to be achieved - *deciding* what the best action or decision is,[32] then acting accordingly in an almost automatic way,[33] potentially producing legal effects that are generally valid even with regard to natural persons.

[29] AI systems - according to one of the most widely accepted definitions - are distinguished as, essentially, rational systems capable of *'acting and thinking humanly'*. Cf. Somalvico, M., Amigoni, F., Schiaffonati, V., (2003) Intelligenza Artificiale, in Petruccioli, S. (ed.), *Storia della scienza*, IX, (615-624); VV. AA., (2016) Artificial Intelligence and life in 2030, *One hundred year study on Artificial Intelligence*, 5. See the definition developed by the High-Level Expert Group on Artificial Intelligence of the European Commission: AI HLEG, (2018) A definition of AI: Main capabilities and scientific disciplines; s. also, Russell, S., Norvig, P., (2009) *Artificial Intelligence: A Modern Approach*.

[30] Namely, a sequence of operations executable by a processor. Cf. Treccani online, entry: algoritmo, https://www.treccani.it/vocabolario/algoritmo/

[31] A broader definition of decision-making algorithm is given by the *State-of-the-Art Report on Algorithmic decision-making* by *Algo:aware*, a December 2018 study commissioned by DG Connect, European Commission.

[32] Possibly changing the environment (natural or virtual) in which it operates or otherwise proposing some output, the solution to a specific problem.

[33] This decision-making process is conducted by the machine - according to the AI technique implemented- by applying static reasoning schemes or by resorting to machine learning techniques (e.g., machine learning, deep learning, neural networks, decision trees and others). Without going into detail, it suffices to know - for present purposes - that in the latter case the machine, instead of executing pre-defined behavioral patterns, processes 'on its own' and dynamically - in application of self-learning and adaptive algorithms - the decision rule; in some cases resulting, therefore, also in being able to better respond and adapt to changes in the environment or refine, with use experience, the ability to generate an appropriate *output*. *Amplius*, s. Bathaee, Y., (2018) *The Artificial Intelligence Black Box And The Failure of Intent and Causation*, in *Harvard Journal of Law & Technology*, (31)2, 890. Cf. the definition developed by the AI HLEG: (Decembre 18, 2018) *A definition of AI: Main capabilities and scientific disciplines*.

The possibility of setting-up ADM mechanisms clearly opens up new scenarios with particular reference to private law negotiations; however, it also raises issues of no small importance: the decision (with legal effects) of these 'autonomous decision-agents' could be neither predictable nor verifiable; biased, harmful, discriminatory and detrimental to fundamental human rights; it could lead to market distortions; exclude persons or categories of persons from fair access to goods or services, perhaps unwittingly manipulated by (hidden) forms of unfair, aggressive and intrusive marketing.[34] Moreover, this situation would be further aggravated by the opacity that often characterizes these systems,[35] reflected in the (increased) difficulty of challenging the decision taken by automated means, thus threatening the legitimate claim of individuals to challenge unfair decisions affecting them.

3.1. Automated (autonomous) negotiation: a new challenge for ethics

These automated systems, in addition to processing all the information useful to carry out a negotiation activity (and thus to *decide*), can be also programmed in such a way as to perform, consequently, all the actions necessary to externally manifest a contractual determination, i.e. aimed at the conclusion of one or more contracts, in a completely automatic manner, interacting, in a network, with other computers or human counterparties.[36] These agreements are usually referred to as *algorithmic* contracts.[37]

Indeed, many of the contracts that are nowadays concluded in a digital environment involve ADM applications governed by AI in the formation and/or definition of the (contents of the) agreement. That is to say: AI algorithms enable

[34] Among others, see Casey, B., (2019) *Title 2.0: Discrimination Law in a Data-Driven Society*, in *J. L. & MOB.*, 36, https://doi.org/10.36635/jlm.2019.title; Crawford, K., (2013) *The Hidden Biases in Big Data*, *HARV. BUS. REV.*, https://perma.cc/E95C-TUQU

[35] See, on the point, AI HLEG (8 April, 2019) *Ethics Guidelines for Trustworthy AI*; Fondazione Leonardo (2019) *Statuto Etico e Giuridico dell'IA*, Further, s., Burrell, J., (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms, *Big Data & Society*, 1, (1-12).

[36] Among others, s. Bravo, F., (2007) *Contrattazione telematica e contrattazione cibernetica*, 187

[37] See, Scholz, L.H., (2017) Algorithmic Contracts, *Stan. Tech. L. Rev.*, 20; Bravo, F. (2007), cit.

complex inference engines to act like *Software-Agents*[38] with an assigned decision-making task: by analyzing current environment data (such as market trends, competitor pricing or customer profiles, etc.) they act like autonomous negotiators – therefore, without any human involvement – and so formulate a specific offer or decide whether to accept a proposal with certain conditions. These are legally relevant decisions (e.g., formulation or acceptance of a contractual proposal; negotiation of the terms of the agreement: subject of the contract, price, quantity, etc.; selection of the counterparty, etc.) capable of legally binding the party that uses the software (*user*) towards the counterparty (whether it is a human part or itself a different Software-Agent).[39]

Such agreements - which, however, are rapidly spreading to other contexts - are, for instance, already common in the financial market, through the widespread practice of the high-frequency trading (H.F.T.). This is a type of automated financial trading that relies on sophisticated computer tools that allow high-frequency, proprietary-algorithm-driven trading of financial instruments.[40] The strategy, characterized by quick trades and a high turnover rates (investment positions are taken even for just a few seconds) aims to profit on extremely small margins but on a large scale, involving a large number of daily transactions, so achieving significant revenues.[41]

HFT can distort the market; firstly, because automated high-speed trading gives an advantage to those who implement it over those who use traditional instruments. Furthermore, HFT can affect the volatility of securities and amplify abnormal price movements. Studies show how the proliferation of this practice in different markets could lead to a distancing of market prices from

[38] Cf. Sartor, G., (2002) Gli agenti software: nuovi soggetti del cyberdiritto*?*, in *Contratto e Impresa*, 2, 466. See also, Finocchiaro, G., (2002) La conclusione del contratto telematico mediante i software agents: un falso problema giuridico*?*, in *Contratto e Impresa*, 2, 501; Scholz, L.H. (2017), cit.

[39] Among others, see, Finocchiaro, G., Bomprezzi, C., (2020) A legal analysis of the use of blockchain technology for the formation of smart legal contracts, *MediaLaws*, 2; Benedetti, A. M., (2021) Contratto, Algoritmi e Diritto Civile Transnazionale: Cinque Questioni, Due Scenari, *Rivista Di Diritto Civile*; Rodanko, M., (2021) Smart contracts and traditional contracts: views of contract law, in Compagnucci, M.C. et al. (eds.), *Smart Contracts*; Bomprezzi, C., (2022) *Implications of Blochchain-based Smart Contracts on Contract law*, Nomos.

[40] Balp, G., Strampelli, G., (2018) Preserving Capital Markets Efficiency in the High-Frequency Trading Era, in *University of Illinois Journal of Law, Technology & Policy*, 1, 349.

[41] Aldridge, (2013) High-frequency trading: a practical guide to algorithmic strategies and trading systems, Vol. 604.

economic fundamentals, reducing the ability of the price of a security to represent the health of the company that issued it. Not to mention the destabilization that would be caused if an error in operation were to trigger abnormal transactions, with knock-on consequences given the propensity of automated systems to respond to market movements in a consequential manner.[42]

The problem, however, is much broader than that, and concerns the phenomenon of algorithmic contracting in general. In HFT, as more generally in the case of automated contracts, it is the algorithm that decides when to trade, the price and the counterparty, based on instructions given by the party on whose behalf it 'negotiates': these instructions (translated in software commands and embedded in the system) may even be very general and respond to the sole intention of operating to maximize profits.[43]

Apart from legal, fundamental ethical issues arise, which are rooted in two essential features of the functioning of algorithmic contracting models: 1) the user (individual or entity) on whose behalf the algorithm concludes the contract may not be able to foresee how it will carry out the instruction given to it or its concrete outcomes (which may also result to be different from those intended); 2) algorithms and artificial intelligence are not neutral: unintended unfair, harmful or discriminatory consequences may result not only from a biased decision or operational error, but also from the decision-making governance model embedded by the system.

3.2. The myth of algorithmic neutrality and the problem of contractual fairness

The initial myth of algorithmic neutrality has now been shattered;[44] intelligent machines are created, programmed and trained by human operators and therefore, like humans, are fallible; they can 'reason', or 'learn to reason' in a biased

[42] See, Approfondimenti. Mercati Finanziari. Il trading algoritmo e gli HFT, *www.consob.it*, online at: https://www.consob.it/web/investor-education/mercati-finanziari#trading (accessed: August 23, 2022). See also, Caivano, V., (2015) The impact of high-frequency trading on volatility. Evidence from the Italian market, *Quaderni di Finanza*; Fondazione Leonardo, (2019) *Statuto Etico e Giuridico dell'IA*, (73-74)

[43] Among others, s., Galiano, A., Leogrande, A., Massari, S. F., Massaro, A., (2019) I processi automatici di decisione: profili critici sui modelli di analisi e impatti nella relazione con i diritti individuali, *Rivista italiana di informatica e diritto,* 2, 42

[44] See, Airoldi, M., Gambetta, D., (2018) Sul mito della neutralità algoritmica, *The Lab's Quarterly*, XX, 4, 29

way.[45] The inferential rules embodied in the ADM algorithms - by which the system interprets knowledge (i.e., data) and operates in its target environment - may reflect cultural biases, even well-meaning and unintended ones.[46] In data-driven systems, such as machine learning-based ones, moreover, biases can also arise from the data collection: data sets used to train ML systems, or for their operation, may suffer from incompleteness or inclusion of unintended social bias, which are then reproduced and automatically reinforced by AI systems; from training, due to biases induced by human interpretation of the data, in the case of supervised-learning algorithms; or - when un-supervised - as a result of online learning and self-adaptation through user interaction[47]. These biases can lead to unfair decisions.[48]

Most importantly, as already mentioned, unfair or detrimental decisions can also result from the normal (or more correctly, for which it is programmed) functioning of the system. The algorithmic agent is tasked with solving a problem and the way it performs reflects the way it is programmed and its user's intended purposes. Performance, therefore, depends on the governance model orientation implemented, and theoretically tends to maximize user interest. Algorithms can make decisions, but the decision-making process requires a set of reference values. The fact that the algorithmic agent is used within a system-oriented, for example, towards profit maximization, can greatly affect the negotiation outcome.[49]

Paradigmatic is the example given by the practice of *Dynamic Pricing*, a widespread reality in the e-commerce area: online retailers often use algorithms to consider information about the market and the potential buyer to set prices at the highest price a given buying party (often, a consumer) is willing to pay. Indeed, these algorithms are programmed to automatically 'carve out' the

[45] Burrell, J., (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms, *Big Data & Society*, 1, (1-12), for whom the claim that algorithms classify information, and thus make decisions, in a more ″objective″ way cannot be taken literally given the degree of human judgment involved in the design of the algorithms themselves, particularly from the standpoint of defining clustering criteria, pre-classifying training data, and adjusting thresholds and decision-making parameters.

[46] On how classification systems can be and are concretely influenced, even with far-reaching consequences, by the ′point of view′ of those who construct them, see the work by Bowker, G.C., Star, S.L., (1999) *Sorting Things Out: Classification and Its Consequences*.

[47] think of common ranking algorithms.

[48] See, on the point AI HLEG (2019); Fondazione Leonardo, (2019) cit.

[49] Galiano, A., Leogrande, A., Massari, S. F., Massaro, A., cit.

commercial offer based on the profile of every single customer and the conditions offered by competitors; the aim: to make the product more attractive and at the same time maximize profit.[50]

While price discrimination is not necessarily unlawful or an unfair practice to the other party, it could lead to unfair results, i.e., to an unbalanced contract; especially when the counterparty is a natural person and perhaps a weaker party (e.g., a consumer), thus normally having a reduced bargaining power; a position that is further aggravated in the digital environment, where contracting with proprietary software agents makes the (already present) asymmetry of information and bargaining power, between businesses (relying on the algorithms) and the consumers, even more serious.[51]

The threat, posed by automatic contracting, to the fairness of the contract – the balance between parties – raises an ethical (more than legal) issue of the utmost importance, and still - at the time of writing - not adequately explored. It is due to a reflection, which will necessarily have to be mixed with, and feed into a legal analysis on the capacity of the contract law system to provide an adequate response, in other words, on the suitability of the remedies normally provided to resolve contractual inequality. This is in the light of the ethical conception of the contract, which has gradually made its way also into Italian legal culture (scholarly opinions but also case law) and which sees the contract as founded (also) around the concept of 'fair exchange', admitting in some cases

[50] See, Narahari, Y., et al., (2005) Dynamic Pricing Models for Electronic Business, *Sādhanā*, 30, 2 (231-256); Raju, C.V.L. et al., (2006) Learning dynamic prices in electronic retail markets with customer segmentation, *Ann Oper Res*, 143, (59-75); Ghose, A., et al., (2002) Dynamic pricing: a strategic advantage for electronic retailers, *Twenty-Third International Conference on Information Systems*, (305-315); Kung, M., et al., (2002) Pricing on the internet, Journal of Product & Brandmanagement, vol. 11, 5 (2002), 274-287; den Boer, A. V., (2013) Dynamic Pricing and Learning: Historical Origins, Current Research, and New Directions; Massaro, A., Galiano, A., Fanelli, G., et al., (2018) Web App for Dynamic Pricing Modeling in Automotive Applications and Data Mining Analytics, *International Journal of Computer Science and Information Technologies*, 9, 1, (4-9).

[51] For further about consumer smart contracts, see, Troisi, E., (2022) Smart contract: What (is in the) future for Consumer protection?, in Veiga, F., et al. (eds.) *Estudos Jurídicos sobre Inteligência Artificial e Tecnologias*, 185

corrective measures of substantive justice at least in the event of the gross disparity between the parties' loss and advantage.[52] Although ethics cannot replace or override the law, the doctrinal development and court application of the principles of fairness in contracts demonstrate how ethics can guide the interpretation of law within its own criteria. In doing so, ethics can assist in identifying potential solutions that are already within the scope of the law's capacity.

3.3. Ethical underpinnings and data protection: why transparent means trustworthy.

Algorithmic biases or inequitable governance models - as said - may lead to unintentional biases and indirect discrimination against certain groups of people, perpetuating or even exacerbating injustice, marginalization, and asymmetries. A situation further aggravated by the opacity that often characterizes these systems, which complicates or makes it impossible for the aggrieved party to have effective access to justice.[53]

[52] Among others, see: Alpa, G., (1994) L'equità, *Nuova giur. civ. comm.*; Alpa, G., (1997) *La protezione della parte debole di origine internazionale (con particolare riguardo al diritto uniforme)*, in Bonell, M.J. and F Bonelli, F., (eds.), Contratti commerciali internazionali e Principi UNIDROIT; Alpa, G., Patti, S. (eds.), (1997) *Le clausole vessatorie nei contratti con i consumatori*; Barcellona, M., (2002) La buona fede e il controllo giudiziale del contratto, in Mazzamuto, S. (ed.), *Il contratto e le tutele: prospettive di diritto europeo*; Bianca, C. M., (1983) *La nozione di buona fede quale regola di comportamento contrattuale*, in Riv. Dir. Civ.; Criscuolo, F., (1999) Equità e buona fede come fonti di integrazione del contratto. Potere di adeguamento delle prestazioni contrattuali da parte dell'arbitro (o del giudice) di equità, in *Riv. arbitrato*; Franzoni, M., (1999) *Buona fede ed equità tra le fonti di integrazione del contratto*, Contratto e Impresa; Galgano, F., (1993) *Sull'equitas delle prestazioni contrattuali*, Contratto e Impresa; Mengozzi, P., (2004) *Lo squilibrio delle posizioni contrattuali nel diritto italiano e nel diritto comunitario*; Rodotà, S., (2004) *Le fonti di integrazione del contratto*; Sacco, R., (1997) *L'abuso della libertà contrattuale*, Vv AA, Diritto privato, III. L'abuso del diritto. About the fair exchange principle in UK contract law, s. among others, Atiyah, P.S., (1990) *Contract and fair exchange*, Essays on Contract, https://doi.org/10.1093/acprof:oso/9780198254447.003.0011, accessed 2 Sept. 2022.

[53] A limit to effective judicial intervention or the implementation of redress measures sometimes also derives from the very features of the technology. an example is given by the case of *smart contracts*. the impossibility to alter - unilaterally and ex post - the chain of blocks of the distributed ledger - while it is a feature that makes the blockchain a reliable peer-to-peer exchange platform - is also its main limit: without the mutual consent of the parties (i.e. also resorting to agreed en-

The emergence of these problems and the growing concern, partly due to the fear that mistrust of new technological tools might limit their deployment in the market, led to the adoption of a series of Ethical Charters - public or even private - at all levels: international, European, national.[54]

coded mechanisms to bring the coded contract to an end or reverse its execution), the transactions, once processed and registered on the network, are irreversible, without residual space for the enforcement action of the judicial authority. This reduces the possibilities of judicial redress, limiting them, *de facto*, to compensatory damages - not necessarily effective in the specific case. However, it is not the only effectiveness, but the recourse to the judiciary itself, which by many commentators is called into question (see, among others, Cutts, T., (2019) Smart contracts and Consumers, *West Virginia Law Review*,122, 2; Cappiello, B., (2020) *Dallo smart contract computer code allo smart (legal) contract. I nuovi strumenti (para) giuridici alla luce della normativa nazionale e del diritto internazionale privato europeo: prospettive de jure condendo*, Diritto del Commercio Internazionale, 2, 477. Also, Cerrato, S.A., (2020) *Appunti su smart contract e diritto dei contratti*, Banca Borsa Titoli di Credito, 3, (370- 407). The smart contract - as developed through blockchain platforms (which are normally globally delocalized), and intended for (archiving and) execution on a distributed network by an indefinite number of randomly identified nodes – *eo ipso*, poses many problems of private international law when it comes to identify which national law governs the contract and which court has jurisdiction to hear the case (s. Cappiello, B., 2020); not to mention the difficulty, in many cases, of identifying (in order to sue) the very same parties of the electronic agreement, due to the secrecy - guaranteed by the asymmetric cryptography- that the blockchain ensures (s., Finocchiaro, G., (2018) Il contratto nell'era dell'intelligenza artificiale, Rivista Trimestrale di diritto e procedura civile, 2, 441). *Amplius*, see, Troisi, E., (2022), cit.

[54] A successful but already incomplete attempt to map the various Ethical Charters, Declarations of Principles, or Guidelines, classified by geo-political context and analyzed by content, is credited to Jobin, A., Ienca, M., Vayena, E., (2019) Artificial Intelligence: the global landscape of ethics guidelines, *Nat. Mach. Intell.* To name a few of the most relevant ones: Ad Hoc Expert Group, UNESCO, (2020) *First Draft Of The Recommendation On The Ethics Of Artificial Intelligence*, SHS/BIO/AHEG-AI/2020/4 REV.2; Council of Europe, (2020) *Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems*, CM/Rec(2020)1; more dated and limited to the Justice field: CEPEJ,(2018) *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*. In European Union context: AI HLEG, (2019) *Ethics Guidelines for Trustworthy AI*; in Italy is to be noted the paper edited by the Task force sull'Intelligenza Artificiale dell'Agenzia per l'Italia Digitale, *Libro Bianco sull'Intelligenza Artificiale al servizio del cittadino*, (2018). Famous among the privately authored declarations of principles is the one drafted within the *Future of Life Institute* and signed by some 1,800 researchers and nearly 4,000 other endorsers, some as famous as Stephen Hawking or digital market giants like Elon Musk and Jaan Tallinn: *The Asilomar AI Principles*, 2017, https://futureoflife.org/ai-principles/

Based on the idea that - as digital technology becomes an increasingly central part of all aspects of human life - people should be able to trust such technology[55] and it should be developed to serve humans, be ethical, and respect fundamental rights, a set of principles and requirements have been identified to which intelligent systems, their applications, producers, programmers, and users (i.e., all stakeholders) should adhere. Leverage, or one of the main levers, of this strategy for 'human-centric' and 'trustworthy' AI[56] - to use the words of the Expert Group on AI appointed by the EU Commission[57] - is precisely transparency.[58] Often, a prerequisite for ensuring that basic human rights and ethical principles are respected, protected, and promoted.[59]

If AI is to be developed at the service of human beings, humans must be able to make use of it consciously:[60] they have the right to always be aware of the fact that they are interacting with an AI system;[61] they must be able to understand its purpose, capabilities, and operating mode ('explicability') and the decisions, to the extent possible, must be explainable to those directly or indirectly affected by them[62] ('explainability') so that they can challenge its modalities and

[55] In this sense, European Commission, (February 19, 2020) *White Paper on Artificial Intelligence. A European approach to excellence and trust*, COM(2020) 65.

[56] See, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions, *Building Trust in Human-Centric Artificial Intelligence*, COM(2019) 168 final, 8 april 2019

[57] See the above-mentioned AI HLEG, (2019) *Ethics Guidelines for Trustworthy AI*

[58] See, AHEG-UNESCO, (2020) *First Draft of The Recommendation on the Ethics of Artificial Intelligence*, cit., III (2) §39, according to which "*transparency may contribute to trust from humans for AI systems*". See also the document compiled by the US National Institute of Standards and Technology (NIST): Vv. Aa., *Four Principles of Explainable Artificial Intelligence*, available online, in draft version, doi: https://doi.org/10.6028/NIST.IR.8312-draft

[59] Cf., AHEG-UNESCO, Ibid., III (2) §37. On the relationship between transparency and trust, see, Felzmann, H., Fosch-Villaronga, E., Lutz, C., Tamó-Larrieux, A., (2019) *Transparency you can trust: transparency requirements for artificial intelligence between legal norms and contextual concerns*, Big Data & Society, 1, (1-14); Schoeffera, J., Machowskia, Y., Kuehla, N., (2021) *A Study on Fairness and Trust Perceptions in Automated Decision Making*, online in arXiv: arXiv:2103.04757v1

[60] See, Vv. Aa., (2019) *Paper sui Principi etici*, in Fondazione Leonardo, *Statuto Etico e Giuridico dell'IA*; also, Longo, G. O., (2007) L'etica al tempo dei robot, Mondo digitale, 1, 3.

[61] See, AI HLEG, *Ethics Guidelines for Trustworthy AI*, cit, §78.

[62] Ibid., §53

content ('contestability'[63]), being able to resort to human intervention to that end.

Transparency also allows individuals control over their own data. AI software and hardware - as explained - can today be programmed in such a way as to autonomously carry out all the actions necessary for a contractual determination having a legal effect; even - referring to the example of *Dynamic Pricing* algorithms- 'tailor' a commercial offer to the interests of the possible buyer, selecting or defining, according to the counterparty's profile, the object of the contract and its characteristics or conditions. All this, 'feeding' on data. Also, and often, personal data. Indeed, between Internet-of-Things, Cookies, Big-Data Analytics, etc., every natural person, at all times, is concerned by or potentially subject to the processing of their data. Profiling and Data Analysis systems find correlations between these data, and from these, they extrapolate other data ('inferred' data); they predict behaviour, catch interests, and sense weaknesses. Based on this data, Automatic Decision-Making tools make decisions; negotiate, enter into obligations, and perform transactions. Sometimes they make mistakes.

The Ethical Charters already mentioned above, almost all agree that AI systems must guarantee the data protection of natural persons. This should include the information initially provided by the User, as well as the information generated about the User: the information that is inferred from the initially processed data (input) as a result of an automated process (of extrapolation, clustering, etc.).[64] For having a safe, reliable and ethical AI, development, deployment and use of ADM applications necessarily require the quality and integrity of the dataset to be assured, not only by testing the system processing and continuously assessing the impact of the algorithm in order to minimize misuse and negative impact, but also, above all, by allowing individual's control over their data. To this end, transparency is pivotal.[65]

[63] To explainability refers Section III (2) §40 of the AHEG-UNESCO, *First Draft of The Recommendation on the Ethics of Artificial Intelligence*, cit.

[64] See, Mittelstadt, B., Wachter, S., (2019) *A right to reasonable inferences: re-thinking data protection law in the age of Big Data and AI*, Columbia Business Law Review, 1.

[65] According to the findings of the study conducted in 2019 by Jobin, A., et al. (s., Jobin, A., Ienca, M., Vayena, E., *Artificial Intelligence: the global landscape of ethics guidelines*, op. cit., 7 ff.) *'transparency'* is the most prevalent principle in the current literature, present in 73 of the 84 documents analyzed, albeit with different shades of meaning, summarized by the locutions: '*transparency, explainability, explicability, understandability, interpretability, communication, disclosure, showing*'.

The data sets, the input information as well as the inferred personal data (e.g., the profile or score assigned to the User; all data 'derived' from the analysis of his or her personal data), the logic and processes that determine the inference and the outcome of the algorithm, the rationale of that certain decision affecting the legal sphere of a natural person should be traceable, transparent, explainable. This is also to enable the individual affected by the decision to assess data accuracy, completeness, and relevance; otherwise, to contest the decision that is biased or based on partial or incorrect information. Instead, algorithm functioning and the reasons for its decision are rarely transparent or understandable, at least adequately: either by choice[66] - for reasons of competition, protection of know-how - or because of technological limitations: this is the case of those algorithms that are properly referred to as 'black-box', systems whose inferential mechanisms are not (completely) predictable *ex ante*[67] or which, in any case, do not always make it possible to explain why a model has generated a particular result or decision (and what combination of factors contributed to it).[68]

It is the essential need for effective protection of the human being against digital power[69] that requires a more central conception of transparency, entailing that the automatic decision is explained to the individual involved and

[66] The fortunate appellation of 'Black Box Society' is due to Frank Pasquale, who masterfully outlines its features with evocative expressions such as '*the Secret Judgments of Software*' and '*the Secrecy of Business and the Business of Secrecy*'; cf. Pasquale, F., (2015) *The Black Box Society. The Secret Algorithms That Control Money and Information*. Harvard University Press, 2015.

[67] But it is sometimes possible - it should be noted - to investigate its behavior by analyzing the responses the system produces in response to the stimuli it receives. So-called 'explanatory tools' [or *post-hoc* explanation techniques, c.f. Zhong, J., Negre, E., (2021) *AI: To interpret or to explain?*, INFORSID] are capable of *ex-post* reconstructing the functioning of certain 'opaque' decision-making models; in particular results of the examined model would be explained by finding the links between input data characteristics and results, or by constructing a simpler model to approximate the original model (Ibid., 6); the accuracy and reliability of these 'explanations' is challenged, for example, by Rudin, C., (2019) *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*, in *Nature Machine Intelligence*, 1(5), (206-215).

[68] Burrel, J., op. cit; s. also, Bathaee, Y., op. cit.; Castelvecchi, D., (2016) Can We Open the Black Box of AI, in *Nature*, 538(20).

[69] In addition to the general principle of fair processing, to which certainly can be said to be contrary the refusal, when unjustified, to provide information about the derived data referable to the data subject or about the rationale for a certain automatic decision.

that they have communication of (and thus control over) all data concerning him/her, including 'inferred' data.[70]

If the digital world has become a 'social', living space, within which a person's '*onlife*'[71] existence takes place and is delineated (i.e., his or her identity is determined) in a way that is often also legally relevant (and otherwise uncontrolled), then the protection of human dignity - anchor of all data protection legislation - individual freedom, and the fundamental right of each person to affect his or her individual and collective dimension, goes first and foremost through awareness, i.e. transparency, meant as the justification of digital power – namely, its explanation - and, only then, through an active power of control over one's own data - particularly automatically inferred data - and the possibility of contain (and contesting) the impact of technologies on and from them, which without a (prior) right to an explanation would be voided. Denying *this* transparency, which ultimately means not only the right to know the rationale for a certain automatic decision affecting a natural person but also the relevant (inferred) data behind it (i.e., who that person *is* online) is to somehow dehumanize the individual.[72]

3.4. Software agents and the problem of the contracts without agreement

In conclusion, there remains to consider – or at least pose – one last, fundamental ethical (even more than legal) question, which is – albeit treated last – of primary importance.

[70] For Messinetti, R., op. cit, p. 867, the legal-technological system must enable the person to understand the machine′s understanding of that person in the context of decision-making processes directed to affect his or her legal and vital sphere; it would therefore be the need to retain control over one′s personal identity (and its formation) that would justify an interpretation of the right of access as a right to understand the logic and justifying reasons for the automated decision (having as its object precisely that identity) and one′s personal data ″derived″ from the original inputs, which constitute the intermediate and/or conclusive outputs of the processing.

[71] The evocative term is from Luciano Floridi, who chose it to represent man′s experience in hyper-historical societies in which he ″no longer distinguishes between online or offline″ and even becomes increasingly ″unreasonable to ask whether one is online or offline.″ Cf. Floridi, L., (2014) *The Fourth Revolution. How the Infosphere is reshaping human reality*; ID. (ed.), (2015) *The Onlife Manifesto: Being Human in a Hyperconnected Era*.

[72] Such an understanding of transparency can be found, for example, in Waelen, R., (2022) Why AI Ethics Is a Critical Theory, Philosophy & Technology, 35,9. https://doi.org/10.1007/s13347-022-00507-5

As already mentioned, a computer can be programmed to conduct a negotiation (i.e. to choose whether, when and with whom to contract and on what terms) by interacting with a human counterparty or with one or more other software agents in a way that, depending on the complexity of the automatism and the number of variables involved, is not only independent of any subsequent will or behaviour whatsoever by the user, but its outcomes are also, often, essentially unpredictable.[73] As known, what makes a valid and enforceable contract – admittedly oversimplifying – is the parties' mutual agreement over the terms of the contract (i.e., the meeting of minds; the exchange of their conscious offer and acceptance) and the intention for that agreement to be legally binding between them. Software Agents, however, are not simply telematic tools; they do not merely express a definite or always predictable will of the parties in the negotiation: they actually define it, and not necessarily as expected. AI decision-making algorithms – due to the complexity of the programming, the environment or both – could easily end in a behaviour that is not foreseeable by their user and often not human-intelligible at all.[74] This is especially the case of the so-called black-box algorithms, whose logic – as noted previously – is un-decipherable, in opposition to that of clear-box ones.[75] Consequence of this is that the use of black box algorithms as negotiators risks so to introduce a gap between the intent of the party using the algorithm and what the software agent actually does on their behalf.[76] Turned into a question: if mutual consent is necessary to have a contract, meaning that the parties must mutually agree on quite specific ob-

[73] Scholz, L. H., *Algorithmic Contracts*, op. cit.; Caggiano, I. A., (2018) *Il contratto nel terzo millennio*, in Nuova giur. comm., 1152; Bravo, F., *Contrattazione telematica e contrattazione cibernetica*, op.cit., 201.

[74] Among others, s., Burrell, J., op. cit.

[75] Scholz, L. H., *Algorithmic Contracts*, op. cit.

[76] In smart contracts, for example, there is the possibility of generating successive, separate 'follow-on' contracts. In practice, when the parties have voluntarily entered into a primary contract, it may itself, as a software agent, stipulate additional, secondary contracts. The parties could in principle not be aware of the content of this follow-on contracts. See, Rao, M., Lezzi, L., Germani, A. R.,(2021) Blockchain e smart contracts: sfide e opportunità di un futuro già presente, *Diritto Mercato e Tecnologia*, online at: https://www.dimt.it/wp-content/uploads/2021/01/01_29-Rao_Lezzi_Germani-Blockchain.pdf. Accessed: 3 Sept. 2022

jects and terms in order to be bound, do we have a contract if these are not foreseeable at all before algorithms conclude (and maybe also perform[77]) the agreement? Do we (which is asking the same), if the instructions given to the algorithm are vague and the decision-making output is unpredictable? More importantly: if yes, in any case? Even when the recourse to the algorithm leads to results absolutely unwanted by its user?

Nevertheless, it is also true that the *User*, in programming some software to act as a negotiator on their behalf, and thus in giving it more or less precise instructions, does in fact express the will to regulate in a certain way the content of the agreement that will follow,[78] entrusting, well aware of the risks, the concrete definition of the agreement to a *bot* as if it were their legal agent.[79] Nor should it be forgotten that on the other side of the automated-contracting algorithm there is another party, which arguably has an interest in the enforceability of the agreement, whose position is therefore also worthy of protection. Indeed, it cannot be ignored the symmetrical need to recognise a certain protection for the other contracting party for their innocent reliance on the presumed conformity of the negotiating intent asserted by the software agent with the actual will of the party behind it.

The role of ethics is therefore to help strike the right balance between these opposing positions and, in doing so, to guide the definition of precise limits. Answering the questions posed above is in fact not only a matter for the legal experts. At the legal level, even amidst opposing theories, it is actually possible to rediscover broad principles that, in one sense or another, regulate the issue; but the inexorable ethical question remains: in case of a positive answer to the above questions - that is: yes, do we have a valid contract - is it fair to attribute unlimited responsibility to the party using the algorithm for un-wanted consequences and unpredictable damages to the other party or third parties This is not a meaning-less dilemma, and it is also ethics that should help guide law towards sustainable solutions or inspire regulation of these phenomena, in the

[77] Peculiarity of smart contracts is, as well known, the automatic performance. Once launched in the blockchain, the smart contract - after the agreement is finalized by counter-party acceptance - executes itself in accordance with the coded terms, regardless of the subsequent will of the debtor and without the parties, as well as third party, being able to arrest or condition it.

[78] See, Clarizia, R., (1985) *Informatica e conclusione del contratto*, 72 ff. and 91.

[79] Among others, s. Chopra, S., White, L., (2011) *A Legal Theory for Autonomous Artificial Agents*, (University of Michigan Press. Available at: https://www.jstor.org/stable/10.3998/mpub.356801

sense of preventing and reducing the social impact of the incidents that inevitably accompany rapid technological progress.[80]

4. AI, Data Protection and Legal Design: automated individual decision-making

The development of A.I. systems has involved numerous ethical, social, and legal issues about the personal data protection.

The operation of A.I. systems is based on a continuous exchange of information (personal data[81]) or a self-learning process (machine learning[82]).

The EU Regulation 2016/679 (GDPR) addressed the issue with particular attention to the automated processing of personal data, where Article 22 has established the right of the data subject not to be subject to a decision based exclusively on the automated processing of his or her data[83] and profiling[84] and, therefore, that it produces legal effects that affect him or that significantly affect his person[85].

However, this article provides for derogations. It does not apply in cases where the decision is: necessary for the conclusion or execution of a contract between the data subject and a data controller; authorized by the law of the European Union or the Member State to which the data controller is subject; it is based on the explicit consent of the data subject[86].

[80] See, Delgado, A. (2017) *Technoscience and Citizenship. Ethics and governance in the digital society*, Springer.; Cacciari, S., (2019) Scenari. Etica, Antropologia, Intelligenza Artificiale, Diritto dell'Informazione e dell'Informatica 2(6), 1175; about the role of ethics and its relationship with human rights in regulating AI, see, Santosuosso, A., (2020) *Intelligenza Artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritt*o, 30 ff.

[81] The definition of "personal data" is regulated by art. 4 of the GDPR.

[82] The Machine Learning is a branch of the AI, which allows software to learn information from data, automatically, to perform cognitive tasks without receiving any instruction. On the point: De Mauro A. (2019), Big Data Analytics. Analizzare e interpretare dati con il machine learning, Apogeo.

[83] Automated decision-making is not allowed without the consent of the data subject.

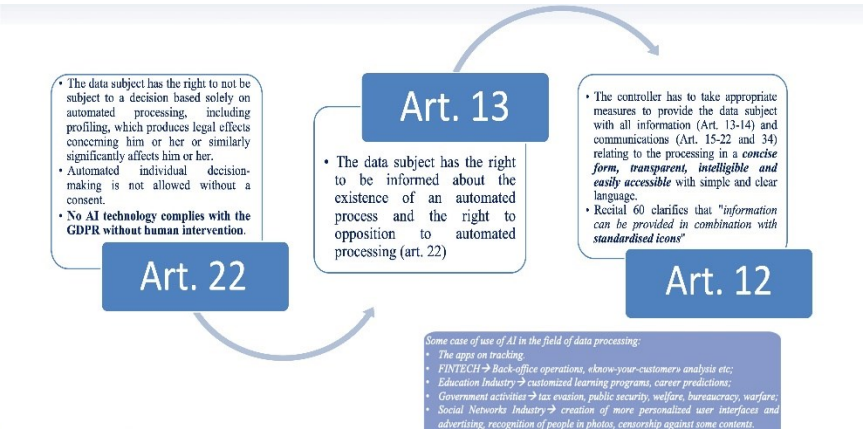
[84] The European Regulation defines "profiling" in Article 4, paragraph 4

[85] *See* Troisi E. (2019), "AI e GDPR: L'autometed decision making, la protezione dei dati e il diritto alla intellegibilità dell'algoritmo", *European Journal of Privacy Law &Technologies EJPLT*, 1, 41-59.

[86] The Rules require the explicit consent of the data subject, which is confirmed by an express statement and not derived from conclusive conduct. *See* Gatt L., Montanari R., Caggiano I.A.

Article 22 should be read next Articles 12 and 13 of the same Regulation, which give the data subject the right to be immediately informed (in a concise, transparent, and easily accessible form (Art. 12), including the use of icons (recital 60)) about the existence of an automated process and the right to oppose biased and discriminatory algorithms (Art. 13).

The aim of all of this is to make users aware of what are the purposes of that processing and what data are necessary for the provision of the service, as well as the destination of all the additional data that are collected.



4.1. The legal design as a solution to the lack of understanding of the privacy policy

Despite these provisions, privacy information is often written in a difficult form. Failure to understand the legal document may constitute a risk - especially in the digital sphere - to the consumer or to the data subject, who usually are the weakest part of the contract. In this connection, from some experiments[87], emerged

(2017), "Consenso al trattamento dei dati personali e analisi giuridico-comportamentale. Spunti di riflessione sull'effettività della tutela dei dati personali", *Politica del diritto*, II, 345-350; Caggiano I.A. (2017), "Il consenso al trattamento dei dati personali", *Diritto mercato tecnologia*, 1-19.

[87] Studies have been carried out in Italy on the consent to the processing of personal data using instruments of measurement of behavioral analysis. *See*: Gatt L., Montanari R., Caggiano I.A. (2021), Privacy and Consent. A Legal and UX&HMI Approach for Data Protection, Suor Orsola University Press, Napoli.

that even those who claim to attach importance to the protection of privacy opt for the default of data settings. Therefore, the user does not perceive the risk of the provision of his consent.

In this regard, W.P. 29 specified that the invitation to accept the processing of data should be subject to strict criteria since the fundamental rights of the data subject are at stake. In this sense Article 7, 4 co., GDPR[88] dictates the rules for assessing whether consent has been freely given. This is to ensure that the processing of personal data for which consent is required does not turn directly or indirectly into a contractual service.

The risk of not understanding the information is even more relevant where art. 12 of the GDPR establishes the right of the data subject to receive the information, referred to in art. 13 and 14, and the communications related to the processing referred to in art. 15-22 and art. 34, in a concise, transparent, intelligible, and easily accessible form, with a simple and clear language, in the case of information specifically intended for minors.

Indeed, the W.P.29[89], about the transparency, has established an obligation to adapt the legal communication to the addressee, since it is not reasonable to address different recipients in the same way, also providing an in-depth analysis of the notion of consent and the methods of disclosure, publishing the relevant guidelines on the subject, in 2018 and 2020.

The W.P. 29 has clarified that the data controller must always use clear and simple language. This means that the message should be easily understood by an average person, not just a jurist/legal expert. The controller must ensure that the consent is provided based on information that allows the data subject to easily identify who the controller is and to understand what he is consenting to.

Therefore, the necessary information to give consent to the processing of personal data cannot be hidden within the framework of the general terms of contract/service[90].

[88] Article 7,4 of GDPR: « *When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract».*

[89] Article 29 Working Party: Guidelines on Transparency under Regulation 2016/679 (wp260rev.01), april 2018, in https://ec.europa.eu/newsroom/article29/items/622227.

[90] The article 7 of GDPR provides that in cases where consent is required in a paper contract but also covers other aspects, then the request for consent must be clearly distinguishable from the rest.

Furthermore, even when consent is to be given electronically, the request must be clear and concise[91]. About this consensus, the European Data Protection Board (EDPB) adopted guidelines in 2020[92].

The effective application of the principles of transparency, clarity and conciseness implies the need to simplify the privacy policy, to facilitate the data subject's understanding and thus enable him or her to exercise his or her right of objection in the event of automated processing of personal data.

The recent European Regulation, despite the attempt provided for in art. 12, has failed to solve this problem.

Indeed, the burden for the data controller to provide all the information provided for in art. 14 would seem incompatible with the requirement of clarity and transparency provided for in art. 12.

On the one hand, clarity is demanded and, on the other, completeness is required: the first principle is exemplification of a qualitative information suitable for informing; the second concerns above all the quantitative aspect of the information (and therefore how much information to provide).

Therefore, there is a need to simplify the privacy policy to facilitate the understanding of the information; allow the data subject to exercise the right of opposition in case of automated processing.

The possible remedy could be the "rewriting of legal clauses", also on the Internet and in general on the devices of I.A., according to the methodology of legal design.

Legal design is a methodology that fully incorporates the concept of "hybridization of knowledge" as it operates in multidisciplinary fields, such as law, design, technology. This methodology consists of behavioral techniques as well as information visualization, in addition to the principles of conciseness and transparency, immediacy of information and understanding of negotiating texts. More precisely, Legal Design is the application of Human Design-Centered to the world of law, to create clearer, understandable, and appealing legal services for the user[93].

[91] The recital 32 of the GDPR provides that if consent is requested electronically, the request may not appear in a paragraph within the general conditions of contract/service but must be separate and distinct.

[92] European Data Protection Board, Guidelines 05/2020 on consent under Regulation 2016/679, may 2020 at: https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-052020-consent-under-regulation-2016679_it.

[93] Hagan M. (2018), Law by Design, available at www.lawbydesign.co/en/home.

The aim, therefore, is to bring the legal world closer to people who do not have legal training or experience. However, this methodology is not exclusively concerned with the aesthetics of texts, and therefore it is not just a matter of changing the layout of the document, but it is necessary that the visual elements are functional for one purpose: to make the legal document more comprehensible and, therefore more efficient, placing the final user at the center of the design and delivery of services.

4.2. Information and legal design in automated processing

The need to inform and be informed, now, is an essential element of the current reality, as with the advent of the Internet, the contexts that require a capacity for choice and judgment have increased and people are not always able to respond in the most autonomous and efficient way.

The consumer, browsing the net, comes across a continuous online marketing activity made of banner ads, links to other platforms, sponsored posts or other, often inherent in tastes or previous choices.

These operations, however, necessarily involve the knowledge of information relating to natural persons, identified or identifiable, and can be qualified as processing activities of personal data[94].

Indeed, one of the most effective marketing procedures (cd. targeted advertising) is based on the reconstruction of the profile of the user who displays a site and then shows him the ads most relevant to his tastes[95].

The most obvious example of this advertising is represented by the automatic offers during the loading of a web page, without the user noticing. It is defined as behavioural advertising,[96] since the basis for its operation is the tracking of consumer activities.

[94] Art. 4 of Regulation (EU) 2016/679 defines the definition of treatment «*any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location or movements*».

[95] *See* Di Palma T. (2021), "Le necessità di contemperamento tra le finalità di marketing e la tutela del trattamento dei dati personali: Tecniche di marketing e adempimenti del titolare del trattamento", *Data Protection Law*, I, 40-56.

[96] The Opinion of the European Data Protection Board has given a definition of behavioral advertising; it is a profiling activity, consisting of an automatic processing technique using algorithms

Behavioural advertising realizes a system of 'aggregation in clusters' through the association of personal data of the users with the information elaborated through an algorithm, that can be of geographic nature (geotargeting), sociodemographic (socio-demographic targeting) of the population present in a relative place. However, such a system has consequences; in fact, the more detailed the cluster is, the more basic information can reveal the sensitive aspects of the life of the profiled subjects. Yet, the purpose of profiling is to allow an identification/ identifiability of the person.

Profiling[97] is mainly used in the field of marketing (cd. "Targeted marketing") just to get customer analysis. Well, on the level of discipline, we must reconnect to art. 22 of the GDPR, which establishes a general prohibition of subjecting an individual to automated decision-making processes, with some exceptions. Therefore, for profiling and automated decision processing required at least explicit consent; yet, in relation to the "sensitive data", art. 9 GDPR prohibits its profiling and treats relative protection by stating that the data subject has various rights.

Art. 35, about data protection impact assessment (DPIA), introduces the institute of impact assessment and data protection, therefore, the assessment also of the risk that the use of new technologies can represent for the rights and freedoms of natural persons.

This risk is to be understood as that negative impact on the freedoms and rights of data subjects and therefore not only as a right to the protection of personal data, but to the protection of a series of values (such as the freedom of expression of subjects).

The owner must develop a prior assessment of the consequences that that processing will have on the rights and freedoms of users, to avoid incurring the sanctions of the GDPR.

The European Parliament resolution of 16 February 2017 (updated in 2019) states that «the user is not authorized to collect or communicate personal information without the explicit consent of the data subject». In this perspective, the principle of privacy by design is even more intensified from an ethical-legal point of view.

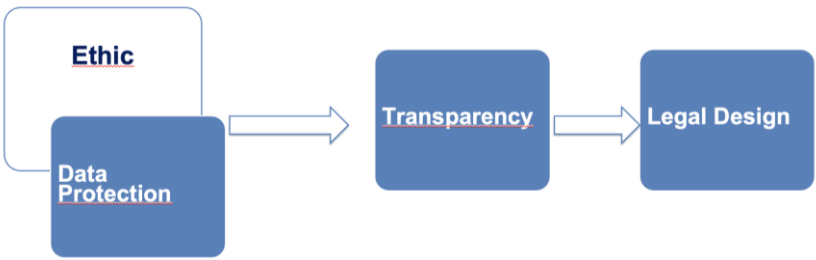
of multiple types of personal data relating to very large numbers of persons, to give each of them a profile, that is a predefined category and outlined through parameters that the controller considers necessary for its research, to achieve its purpose.

[97] Profiling is characterized by three elements: automated processing; performed on personal data; with the purpose of evaluating personal aspects of a natural person.

It emerges that the means of treatment must be designed according to ethics-legal values (therefore also in ethics by design and not only privacy by design).

In the field of the regulation of AI, ethics and law are brought together, and these influence one another. Ethical considerations ground legal norms, even the most fundamental ones, typically expressed through constitutional rights and international treatises.

Therefore, the response of law and ethics can only be based on intervention strategies to implement technological measures (software) that minimize the risks of behaviour harmful to certain categories of users.



The GDPR provides a particular exception, namely the consent of the data subject, to justify profiling or automated decision processing, requiring that the same consent is ″explicit″ (understood as an active and specific affirmative statement from the data subject and qualifies as a response of the data subject to the proposal to accept or refuse the data processing).

The activity that can cause the most concern, from the point of view of the protection of personal data, is that of web listening, a real service of ″listening to information″ provided by different companies. In fact, by verifying the frequency with which a given word or brand is mentioned in a conversation on social networks and websites, advertisers have a better chance to carry out their task of understanding what the user wants at that moment.

In this sense, the case of Facebook is exemplary[98]: the user, upon registration to the site "agrees" to grant all content to the social network.

In 2018, Antitrust had sanctioned Facebook for ten million euros for having put in place an incorrect commercial practice. In fact, social media did not adequately inform users at the time of registration, on how data were collected and processed. After the sanction, Facebook changed the information by entering a button at the time of registration «find out how we collect, use and share your data and how we use cookies». However, according to Antitrust, there is still no adequate reference to the activity of acquiring and using user data for commercial purposes.

The European Commission Antitrust - and on the same line also the British Antitrust "Competition and Markets Authority" - has opened a formal investigation on Facebook to assess whether there has been a violation of the European competition rules. Indeed, the concern lies in the fact that Facebook uses data obtained from competing providers in the context of their advertising on social networks to support Facebook marketplace.

Moreover, it cannot be considered that the social network is used by almost three billion people monthly and almost seven million companies advertise on the platform, which collects large amounts of data on the activities of users of its social network and beyond, allowing them to target specific customer groups.

Article. 12 of the GDPR may intervene, as a remedy, which, par. 1, provides that the data controller must provide information «concise, transparent, intelligible and easily accessible».

To fulfil this criterion, the following are necessary:

a. a clearly visible and suitably distinguishable indication in the text of the information indicating the activity of the controller.
b. The possibility for the data subject to express his or her consent (or to refuse the automated processing of data).

If the information allows the data subject to be fully aware of the data processing activities, he will be able to express an informed consent.

[98] *See* (2018) "Facebook, multa Antitrust da 10 milioni per uso dei dati degli utenti a fini commerciali" at: https://www.ilsole24ore.com/art/-facebook-multa-antitrust-10-milioni-uso-dati-utenti-fini-commerciali--AEL2lWvC?refresh_ce=1. Ibello G. (2020), "Facebook nel mirino dell'Antitrust" at: https://www.altalex.com/documents/news/2020/02/05/facebook-antitrust.

All this is because the consent (separate from the information) must be independent of other possible types of consent expressed (such as being the recipient of commercial and promotional communications).

Personalization, on the other hand, is a marketing strategy that consists in using both technology and the information that you have on your interlocutors to facilitate the interaction between companies and their consumers.

Most of the most advanced companies that invest in customization techniques have been awarded by major commercial successes.

At this point, one wonders if the choices of individuals can be influenced in some way. This theme is the subject of the study of behavioural economics: a hybrid of economics and social psychology that has redefined understanding. In this regard, psychologists D. Kahneman and A. Tversky have devised a Nobel Prize-winning theory that the human brain is composed of two decision-making processes: system 1 (fast process) that makes quick decisions based on emotions and instinct; system 2 (slow process) cerebral and logical, which usually requires more effort[99]. Decisions taken within the framework of behavioural economics are made with system 1, which is more easily influenced. Furthermore, nudge theory has revolutionized traditional economics and contributed to the development of behavioural economics. Nudge refers to a series of expedients that allow people's choices to be directed more rationally and above all more economically[100]. Yet, even if nudging is aimed at improving services and maximizing resources, there seems to be little freedom of choice for the consumer.

Therefore, nudge marketing aims to influence consumers' choices towards more 'favourable' options. The purpose of social networks is to encourage users to spend as much time on their platform.

Art. 25 GDPR has introduced the principles of privacy by design and privacy by default, which requires companies to provide, already at the beginning of a marketing project - then at a preliminary stage of processing - the tools and the correct settings to protect personal data.

[99] Maglia E. (2018), "Un'amicizia da Nobel: Kahneman e Tversky", at: https://www.economiacomportamentale.it/2018/12/07/unamicizia-nobel-kahneman-tversky-co-autorato-affezione/.

[100] A case of nudge is during the withdrawal of banknotes; it asked to choose whether to print the receipt, alongside the application to a terrestrial globe, somehow the user is influenced to make the most environmentally friendly choice.

The purpose is to design and develop a system or device, to support those principles, values, and security rules to make that system or device privacy-aware or privacy-friendly.

According to the principle of privacy by default, however, the owner must ensure that, by default, only the personal data necessary in relation to each specific purpose of the processing are processed and that the quantity of the data collected, and the duration of their storage do not exceed the minimum necessary for the purposes pursued.

This raises the question of whether companies are on an equal footing or are better off than consumers. This is because, as we have seen, companies often use "pushes" to influence consumers, claiming informational or motivational superiority over their customers.

Motivational psychologists believe that such an imbalance is not necessary and that the same purpose can be achieved by placing users in the condition of making informed choices.

This asymmetry creates a real paradox in the protection of data protection. Legal design is in line with the principles of privacy by design and by default, with art. 12 GDPR, with motivational psychology theories, and with Kahneman's theory of "slow thinking". It follows that what behavioural economics and psychology must focus on are conscious choices, which can only take place if legal information is designed with the legal design methodology.

5. Ethical and legal regulation for AI: principles and tools.

Although in part it is still possible to go for an extensive interpretation of the existing rules, the need to introduce a regulatory framework for AI cannot be ignored. This need has recently been recognised by the European institutions that are working on classifying (AI Act) and regulating AI (AI Liability Directive Proposal and, more in general on products, PLD revision proposal).

Ethics and law are two distinct domains that must be considered together for AI regulation, taking into account their convergence. In this regard, it appears possible to provide the following ethical (from 1 to 4) and legal principles (from 5 to 8) which are being transposed into EU hard law acts and could be able to regulate AI: 1) straightening of accountability and transparency, also thanks to the competent Authorities supervision powers; 2) trustworthy AI technologies based on non-discrimination, fairness and autonomy of the individuals; 3) human-centered AI (including customization), putting it at the service of the

humans, in the light of societal and environmental well-being; 4) robust AI, i.e. continuous assessment of AI, primarily under a technical perspective (e.g. standards, design principles and certificates), but also under a social perspective (e.g. social consciousness and explicability); 5) well-established liability rules related to any physical or psychical damage (personal injury, data breach, cybersecurity breach) for the producer; 6) precise insurance rules consistent with foreseen roles and related liabilities; 7) high fines for the damages occurred; 8) clear right to clime of the damaged party.

In order to guarantee the application of these effective ethical and legal principles for AI, it is important to specify that double action is needed: preventive and remedial protection.

First of all, it appears important to strengthen the *ex ante* protection to guarantee ethical and legal compliance by design, through adequate evaluation tools. This approach is being applied in relation to certain legal issues on a legislative initiative (DIPIA or ALTAI), but it would seem that there is a lack of a complete tool that takes into account all the most important ethical and legal issues, under an ecodynamic assessment[101].

On the other hand, the enforcement of the duties and the accountability of the AI technology programmers/developers/producers (hereinafter generically referred to as producers) should be implemented. This objective could be realised through the actions of the competent Authorities with their supervisory powers (preventive measures) and, if necessary, imposing high fines (remedial measures), following the way that appears to have been undertaken in recent years by the legislator. Also, the competent Courts have the power to impose fines and, depending on the case, can be sued alternatively or in addition to the competent Authorities. Furthermore, it is fundamental the power to orient and positively influence the producer through soft law (codes of conduct and guidelines). In this regard, also standards (technical regulation) play an important role as well as the techno-regulation by design to guarantee safety, privacy, cybersecurity and ethics by design measures.

In this scenario of regulation, the ethics-law dual-track approach does not guarantee certainty of law and legal situations. It makes the process of interpreting existing legislation even more complex and, consequently, makes it increasingly difficult to resolve concrete cases. For this reason, a convergence of

[101] For an initial analysis and application of the ecodynamic approach in the legal field, allow me to refer to Gaeta M.C.(2023), Gli strumenti privatistici a garanzia dell'ecosostenibilità nell'industria automobilistica 4.0. *Actualidad juridica iberoamericana* (forthcoming).

ethical choices, expressed through the identification of ethical principles, into legal ones appears necessary to have clear binding rules. In this way, the ethical principle transposed into positive law becomes binding, the other one can orient the individuals, companies and other institutions, as well as give a more suitable interpretation of the legal norms.

In conclusion, the law should keep up with technologies and evolve with them also implementing ethical principles (as EU proposals on AI do), in order to assure effective protection of the vulnerable subject in the digital habitat. In the case of AI applications, which have strong ethical implications, hard law (such as European Directives or Regulations), should be accompanied by soft law (including ethical codes) and standards, to better protect individuals, as vulnerable subjects, concerning the risks involved in new technologies. In this way, multi-level co-regulation mechanisms, fortified by certification tools based on technical standards and techno-regulation, will be able to guarantee adequate protection for the individual.

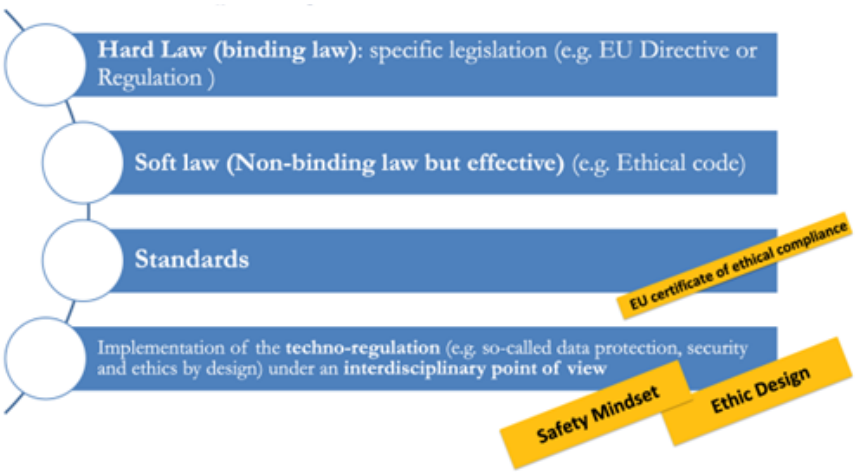


Figure 3: Type of regulatory framework for AI

ACKNOWLEDGMENTS

The author of paragraphs 1, 2 and 5 is Dr. Maria Cristina Gaeta, Ph.D., Lecturer in Private Law, Scientific Secretary of the Research Centre of European Private Law (ReCEPL) at Università degli Studi Suor Orsola Benincasa (Naples, Italy). The author of paragraph 3 is Dr. Emiliano Troisi, Ph.D, Junior Researcher of the Research Centre of European Private Law (ReCEPL) at Università degli Studi Suor Orsola Benincasa (Naples, Italy), Post. Doc. at Istituto Italiano per gli Studi Storici (Naples, Italy). The author of paragraph 4 is Dr. Livia Aulino, Ph.D., Senior Researcher of the Research Centre of European Private Law (ReCEPL) at Università degli Studi Suor Orsola Benincasa (Naples, Italy), Post. Doc. Research Fellow at University Federico II (Naples, Italy). The work has been coordinated by Professor Lucilla Gatt, Full Professor of Private Law and Director of the Research Centre of European Private Law (ReCEPL) and Professor Ilaria Amelia Caggiano, Full Professor of Private Law and Vice-director of the Research Centre of European Private Law (ReCEPL) at Università degli Studi Suor Orsola Benincasa (Naples, Italy).

REFERENCES

Airoldi, M., Gambetta, D., (2018) Sul mito della neutralità algoritmica, *The Lab's Quarterly*, XX, 4, 29.

Al Mureden E. (2019) Autonomous cars e responsabilità civile tra disciplina vigente e prospettive de iure condendo. *Contr. impr.*, 2019, 3, 895.

Albanese A. (2019) La responsabilità civile per i danni da circolazione dei veicoli ad elevata automazione. *Europa e diritto privato*, 4, 995.

Aldridge, (2013) High-frequency trading: a practical guide to algorithmic strategies and trading systems, Vol. 604.

Alpa, G., (1994) L'equità, *Nuova giur. civ. comm.*, 231.

Alpa, G., (1997) *La protezione della parte debole di origine internazionale (con particolare riguardo al diritto uniforme)*, in Bonell, M.J. and F Bonelli, F., (eds.), Contratti commerciali internazionali e Principi UNIDROIT.

Alpa, G., Patti, S. (eds.), (1997) Le clausole vessatorie nei contratti con i consumatori. Giuffrè.

Ashton K. (2009) That "Internet of Things" Thing. In the real world, things matter more than ideas. *RFID J*, 1.

Atiyah, P.S., (1990) Contract and fair exchange, Essays on Contract, https://doi.org/10.1093/acprof:oso/9780198254447.003.0011. [Accessed 2 Sept. 2022]

Balp, G., Strampelli, G., (2018) Preserving Capital Markets Efficiency in the High-Frequency Trading Era, in *University of Illinois Journal of Law, Technology & Policy*, 1, 349.

Barcellona, M., (2002) *La buona fede e il controllo giudiziale del contratto*, in Mazzamuto, S. (ed.), Il contratto e le tutele: prospettive di diritto europeo.

Bathaee, Y., (2018) The Artificial Intelligence Black Box And The Failure of Intent and Causation, in *Harvard Journal of Law & Technology*, (31)2, 890.

Benedetti, A. M., (2021) Contratto, Algoritmi e Diritto Civile Transnazionale: Cinque Questioni, Due Scenari, *Rivista Di Diritto Civile*, 414.

Bertolini A., Riccaboni M. (2020) Grounding the case for a European approach to the regulation of automated driving: the technology-selection effect of liability rules. *Eur J Law Econ*, 1.

Bianca, C. M., (1983) La nozione di buona fede quale regola di comportamento contrattuale, in *Riv. Dir. Civ.*, 205.

Bravo, F., *Contrattazione telematica e contrattazione cibernetica*, op.cit., 201.

Bomprezzi, C., (2022) Implications of Blochchain-based Smart Contracts on Contract law, Nomos.

Bravo, F., (2007) Contrattazione telematica e contrattazione cibernetica, 187.

Burrell, J., (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms, *Big Data & Society*, 1, (1-12)

Cacciari, S., (2019) Scenari. Etica, Antropologia, Intelligenza Artificiale, *Diritto dell'Informazione e dell'Informatica* 2(6), 1175.

Caggiano I.A. (2017), Il consenso al trattamento dei dati personali, *Diritto mercato tecnologia*, 1-19.

Caivano, V., (2015) The impact of high-frequency trading on volatility. Evidence from the Italian market, *Quaderni di Finanza*; Fondazione Leonardo, (2019) *Statuto Etico e Giuridico dell'IA*, (73-74).

Cappiello, B., (2020) Dallo smart contract computer code allo smart (legal) contract. I nuovi strumenti (para) giuridici alla luce della normativa nazionale e del diritto internazionale privato europeo: prospettive de jure condendo, *Diritto del Commercio Internazionale*, 2, 477.

Casey, B., (2019) Title 2.0: Discrimination Law in a Data-Driven Society, in *J. L. & MOB.*, 36, https://doi.org/10.36635/jlm.2019.title; Crawford, K., (2013) *The Hidden Biases in Big Data*, *HARV. BUS. REV.*, https://perma.cc/E95C-TUQU.

Castelvecchi, D., (2016) Can We Open the Black Box of AI, in *Nature*, 538(20).

Cerrato, S.A., (2020) *Appunti su smart contract e diritto dei contratti*, Banca Borsa Titoli di Credito, 3, (370- 407).

Chopra, S., White, L., (2011) *A Legal Theory for Autonomous Artificial Agents*, (University of Michigan Press. Available at: https://www.jstor.org/stable/10.3998/mpub.356801

Clarizia, R., (1985) Informatica e conclusione del contratto. Giuffrè.

Criscuolo, F., (1999) Equità e buona fede come fonti di integrazione del contratto. Potere di adeguamento delle prestazioni contrattuali da parte dell'arbitro (o del giudice) di equità, in *Riv. Arbitrato*, 71.

Cutts, T., (2019) Smart contracts and Consumers, *West Virginia Law Review*,122, 2.

D'Acquisto, G., Naldi, M. (2018) *Big Data e Privacy By Design.*

Davola A., Pardolesi R. (2017) In viaggio col robot: verso nuovi orizzonti della r.c. auto ("driverless")?. *Danno e responsabilità*, 5, 616.

De Mauro A. (2019) Big Data Analytics. Analizzare e interpretare dati con il machine learning, Apogeo.

De Mauro A., Greco M., Grimaldi M. (2016) A Formal definition of Big Data based on its essential features. *Library Review*, 63(3), 122.

Delgado, A. (2017) Technoscience and Citizenship. Ethics and governance in the digital society, Springer.

Den Boer, A. V., (2013) Dynamic Pricing and Learning: Historical Origins, Current Research, and New Directions.

Di Palma T. (2021) Le necessità di contemperamento tra le finalità di marketing e la tutela del trattamento dei dati personali: Tecniche di marketing e adempimenti del titolare del trattamento. *Data Protection Law*, I, 40-56.

Fantinato, M., Hung P.C.K. (et al.) (2018) A preliminary study of Hello Barbie in Brazil and Argentina. *Sustainable Cities and Society*, 40, 83.

Felzmann, H., Fosch-Villaronga, E., Lutz, C., Tamó-Larrieux, A., (2019) *Transparency you can trust: transparency requirements for artificial intelligence between legal norms and contextual concerns*, Big Data & Society, 1, (1-14).

Finocchiaro, G., Bomprezzi, C., (2020) A legal analysis of the use of blockchain technology for the formation of smart legal contracts, *MediaLaws*, 2.

Finocchiaro, G., (2002) La conclusione del contratto telematico mediante i software agents: un falso problema giuridico*?*, in *Contratto e Impresa*, 2, 501.

Floridi, L., (2014) *The Fourth Revolution. How the Infosphere is reshaping human reality*.

FLoridi L. (ed.) (2015) *The Onlife Manifesto: Being Human in a Hyperconnected Era*.

Franzoni, M., (1999) Buona fede ed equità tra le fonti di integrazione del contratto, *Contratto e Impresa*, 83.

Gaeta M.C. (2019) Liability rules and self-driving cars: The evolution of tort law in the light of new technologies. Editoriale Scientifica Italiana (ESI).

Gaeta M.C. (2020) 'Smart toys and minors' protection in the context of the Internet of Everything. *EJPLT*, 2, 118.

Gaeta M.C.(2023) Gli strumenti privatistici a garanzia dell'ecosostenibilità nell'industria automobilistica 4.0. *Actualidad juridica iberoamericana* (forthcoming).

Galgano, F., (1993) Sull'equitas delle prestazioni contrattuali, *Contratto e Impresa*, 419.

Galiano, A., Leogrande, A., Massari, S. F., Massaro, A., (2019) I processi automatici di decisione: profili critici sui modelli di analisi e impatti nella relazione con i diritti individuali, *Rivista italiana di informatica e diritto,* 2, 42

Gatt L. (2022) Legal anthropocentrism between nature and technology: the new vulnerability of human beings. *EJPLT* 1, 15.

Gatt L., Montanari R., Caggiano I.A. (2017) Consenso al trattamento dei dati personali e analisi giuridico-comportamentale. Spunti di riflessione sull'effettività della tutela dei dati personali, *Politica del diritto,* II, 345-350.

Gatt L., Montanari R., Caggiano I.A. (2021) Privacy and Consent. A Legal and UX&HMI Approach for Data Protection, Suor Orsola University Press, Napoli.

Ghose, A., et al., (2002) Dynamic pricing: a strategic advantage for electronic retailers, *Twenty-Third International Conference on Information Systems*, (305-315).

Hagan M. (2018) Law by Design, available at www.lawbydesign.co/en/home.

Haller S., Karnouskos S., Schiroh C. (2008) The Internet of Things in an enterprise context Future Internet, *Lecture Notes in 5468 Computer Science*, 1.

Ibello G. (2020) Facebook nel mirino dell'Antitrust at: https://www.altalex.com/documents/news/2020/02/05/facebook-antitrust.

Jarvis Thomson J. (1985) The Trolley Problem. *Yale Law Journal*, 94 (6), 1395–1415.

Jobin, A., Ienca, M., Vayena, E., (2019) Artificial Intelligence: the global landscape of ethics guidelines, *Nat. Mach. Intell.*, DOI:10.1038/s42256-019-0088-2.

Kung, M., et al., (2002) Pricing on the internet, Journal of Product & Brandmanagement, vol. 11, 5 (2002), 274-287

Longo, G. O., (2007) L'etica al tempo dei robot, Mondo digitale, 1, 3.

Maglia E. (2018) Un'amicizia da Nobel: Kahneman e Tversky, at: https://www.economiacomportamentale.it/2018/12/07/unamicizia-nobel-kahneman-tversky-co-autorato-affezione/.

Manta, I.D., Olson, I.S. (2015) Hello Barbie: First They Will Monitor You, Then They Will Discriminate Against You. Perfectly. *Alabama Law Review*, 67, 135.

Mantelero, A. (2012) Big Data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo. *Dir. Inf.*, 1, 135.

Mantelero, A. (2017) Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework. *Computer Law and Security Review*, 584.

Mantelero, A. (2019) La privacy all'epoca dei Big Data, in Cuffaro, V., D'Orazio, R., Ricciuto, V. (eds) *I dati personali nel diritto europeo*, 1181.

Mantelero, A., Poletti, D., (eds), Regolare la tecnologia: il Reg. UE 2016/679 e la protezione dei dati personali. Un dialogo fra Italia e Spagna. Pisa University Press.

Massaro, A., Galiano, A., Fanelli, G., et al., (2018) Web App for Dynamic Pricing Modeling in Automotive Applications and Data Mining Analytics, *International Journal of Computer Science and Information Technologies*, 9, 1, (4-9).

Mengozzi, P., (2004) Lo squilibrio delle posizioni contrattuali nel diritto italiano e nel diritto comunitario.

Milosevic, T., Dias, P., Mifsud, C., Trültzsch-Wijnen, C.W. (2018) Media Representation of Children's Privacy in the Context of the Use of "Smart" Toys and Commercial Data Collection, *Media Studies*, 9, 26.

Mittelstadt, B., Wachter, S., (2019) A right to reasonable inferences: re-thinking data protection law in the age of Big Data and AI, *Columbia Business Law Review*, 1.

Narahari, Y., et al., (2005) Dynamic Pricing Models for Electronic Business, *Sādhanā*, 30, 2 (231-256).

Noto La Diega G., Walden I. (2016) Contracting for the 'Internet of Things': Looking into the Nest. *Queen Mary School of Law Legal Studies Research Paper*, 219.

Pallone E.K. (2016) "Internet of Things" e l'importanza del diritto alla privacy tra opportunità e rischi. *Ciberspazio e diritto*, 17(55), 174.

Pasquale, F., (2015) The Black Box Society. The Secret Algorithms That Control Money and Information. Harvard University Press, 2015.

Peppet S.R. (2014) Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent. *Texas Law Review*, 93, 85.

Raju, C.V.L. et al., (2006) Learning dynamic prices in electronic retail markets with customer segmentation. *Ann Oper Res*, 143, (59-75).

Rao, M., Lezzi, L., Germani, A. R.,(2021) Blockchain e smart contracts: sfide e opportunità di un futuro già presente, *Diritto Mercato e Tecnologia*, online at: https://www.dimt.it/wp-content/uploads/2021/01/01_29-Rao_Lezzi_Germani-Blockchain.pdf. [Accessed: 3 Sept. 2022]

Rodanko, M., (2021) Smart contracts and traditional contracts: views of contract law, in Compagnucci, M.C. et al. (eds.), *Smart Contracts*.

Rodotà, S., (2004) Le fonti di integrazione del contratto. Giuffrè.

Rudin, C., (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, in *Nature Machine Intelligence*, 1(5), (206-215).

Ruffolo U., Al Mureden E. (2019) "Autonomous vehicles" e responsabilità nel nostro sistema ed in quello statunitense. *Giur. It.,* 7, 1657.

Sacco, R., (1997) *L'abuso della libertà contrattuale*, Vv AA, Diritto privato, III. L'abuso del diritto, 217.

Santosuosso, A., (2020) *Intelligenza Artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritt*o, 30.

Sartor, G., (2002) Gli agenti software: nuovi soggetti del cyberdiritto*?,* in *Contratto e Impresa*, 2, 466.

Schoeffera, J., Machowskia, Y., Kuehla, N., (2021) *A Study on Fairness and Trust Perceptions in Automated Decision Making*, online in arXiv: arXiv:2103.04757v1

Scholz, L. H., *Algorithmic Contracts*, op. cit.; Caggiano, I. A., (2018) *Il contratto nel terzo millennio*, in Nuova giur. comm., 1152.

Somalvico, M., Amigoni, F., Schiaffonati, V., (2003) Intelligenza Artificiale, in Petruccioli, S. (ed.), *Storia della scienza*, IX, (615-624).

Troisi E. (2019) "AI e GDPR: L'automated decision making, la protezione dei dati e il diritto alla intellegibilità dell'algoritmo", *European Journal of Privacy Law & Technologies EJPLT*, 1, 41-59.

Troisi, E., (2022) Smart contract: What (is in the) future for Consumer protection?, in Veiga, F., et al. (eds.) *Estudos Jurídicos sobre Inteligência Artificial e Tecnologias*, 185

Von Bodungen B., Caggiano I.A., Steege H., Gaeta M.C. (2023) *European regulation for self-driving cars*. Springer (forthcoming).

VV. AA., (2016) Artificial Intelligence and life in 2030, *One hundred year study on Artificial Intelligence*, 5.

Vv. Aa., (2019) *Paper sui Principi etici*, in Fondazione Leonardo, *Statuto Etico e Giuridico dell'IA*.

Vv. Aa., Four Principles of Explainable Artificial Intelligence, available online, in draft version, doi: https://doi.org/10.6028/NIST.IR.8312-draft

Waelen, R., (2022) Why AI Ethics Is a Critical Theory, Philosophy & Technology, 35,9. https://doi.org/10.1007/s13347-022-00507-5

Weber R.H. (2010) Internet of Things, New security and privacy challenges. *Computer law & security rep*, 23 ff.

Zhong, J., Negre, E., (2021) *AI: To interpret or to explain?*, INFORSID.

Ziegler S. ed. (2019) *Internet of Things Security and Data Protection*, Springer.