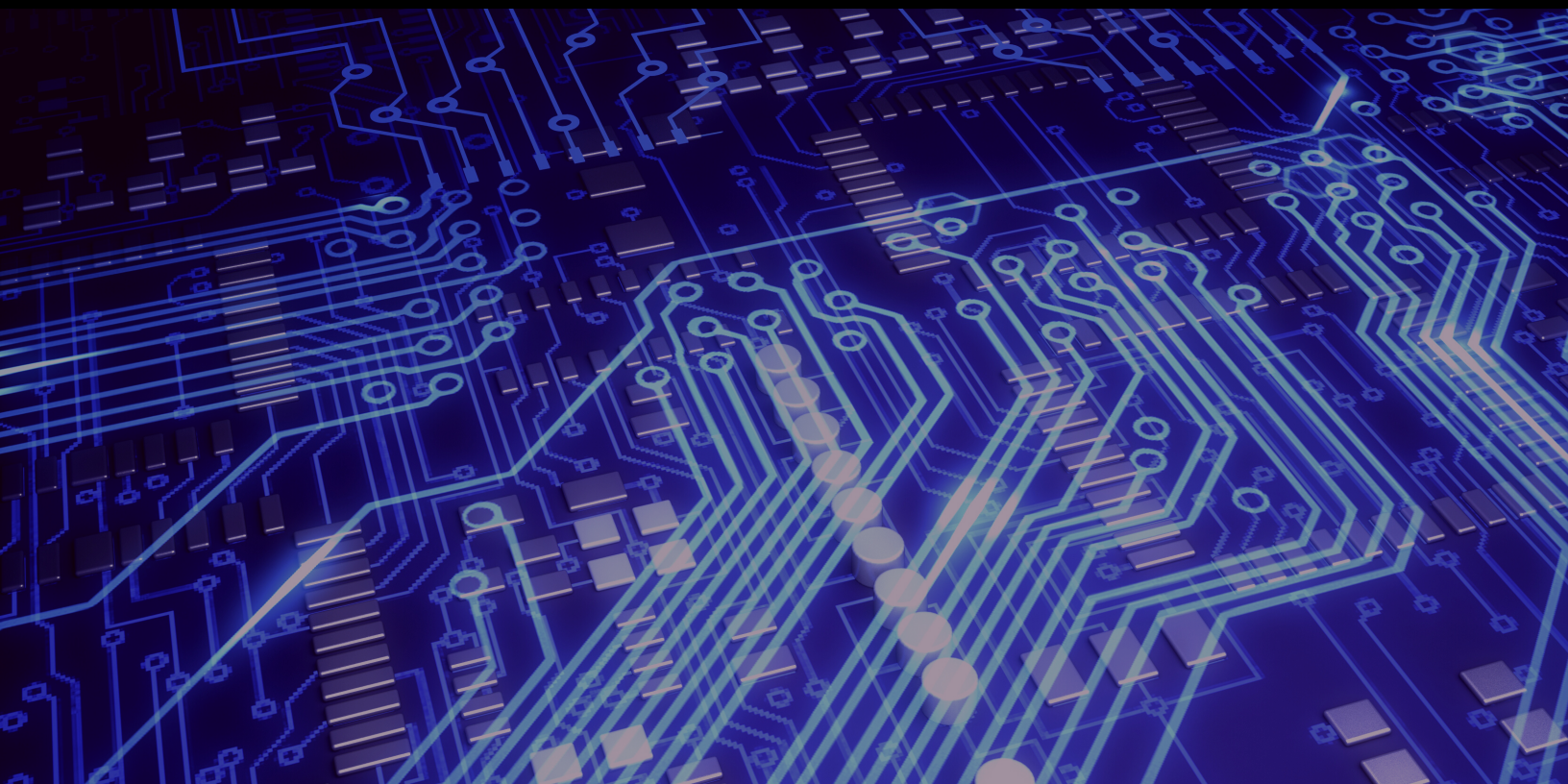
EAST-WEST CENTER

Humane Artificial Intelligence

Working Paper No.02 | July 2020

The Fragility of Human Rights Facing AI

Maria Stefania Cataleta



The Fragility of Human Rights Facing AI

Maria Stefania Cataleta[1], Université Côte d'Azur

Abstract: Machines do not have morality so they must be designed according to shared ethical rules. In this regard, affective computing, a branch of information technology that aims to transmit information on human feelings to machines, can improve the relationship between man and computer, the HCI (human computer interaction), because a system capable of perceiving the user's state of mind can better evaluate his intentions and his/her real will. In relation to the violation of human rights, it is necessary to develop ethical principles that can be negotiated on a computational basis and used in the face of unforeseen situations, to limit regulatory violations or to deal with unforeseeable situations with a morally significant impact.

As we delegate more responsibility to machines in regards to autonomous decision making, we must guarantee proper accountability in protecting human rights. One example of the use of AI for a presumed good that might also violate human rights is human monitoring. Currently, many governments are implementing new technologies, such as video surveillance and biometric tracking, to thwart illegal and threatening behavior, including acts of terrorism. These government activities make our lives more secure, and do work to hinder criminal activities. However, these same technologies actively monitor and track common citizens, which constitutes a violation of individual privacy and could result in future discrimination based on religious beliefs, health conditions, or even political opinions. Also, evolution within the nanotechnology sector raises further issues. In the face of technological and scientific progress, the concept of a *legal person* is pushed to its limits, for "*the scientific and technological world, artificial in conceptual nature, come to encroach upon the already defined legal dimension of a person, an artificial concept in itself*".[2]

[1] Ph.D., LL.M., Italian lawyer admitted on the List of counsels of the International Criminal Court and other international criminal jurisdiction, associate researcher at LADIE, Université Côte d'Azur, France. The present paper is an extract of the article "AI and Human Rights, an Unequal Struggle", Authors Cataleta Maria Stefania & Cataleta Anna, published on CIFILE Journal of International Law in 2020.

[2] *See* Rodotà, Stefano(2007), *Dal soggetto alla persona*, Editoriale Scientifica, at 42 quoted in Pariotti, *supra* note 15, at 184.

On the ongoing path of the social legitimization of technological progress, human rights represent a referential normative principle. This legitimization "*cannot be accepted only on the grounds of security or on the logic of economic efficiency"* and *"must always remain measured by the metre of democracy and of respect of people"*.[3]

Along with this technological development, dignity—the foundational notion for all human rights and of the natural equality of human beings—is called into question. Concern for dignity is most manifestly present in the United Nations Charter and Universal Declaration of Human Rights of 1948, in particular in the preamble, and in articles 1 and 2. It is the recognition of dignity that entitles all human beings to inalienable rights, ones which guarantee natural equality, protecting us from any form of discrimination.[4] Technology has the potential to place equality at risk.[5]

The issue of equality is linked to that of non-discrimination, two specular concepts which represent positive and negative articulations of one unique principle. Indeed, equality means to treat all cases equally, and non-discrimination serves to prohibit biased treatment for all reasonably motivated cases. All treaties on human rights affirm the principle of equality.[6] For certain serious discriminations, such as those concerning race, ethnic origin, sex, or religion, stringent tests are posed upon the State to justify their existence.

The conventions on human rights prohibit direct discrimination, which occurs when a person is treated in a disadvantageous way with respect to another who is in a similar situation; as well as indirect discrimination that occurs when a person who is formally treated like others,

[3] *See* Rodotà, Stefano (2005), *Tecnologia e diritti*, Il Mulino, at 26, quoted in Pariotti, *supra* note 15, at 184.
[4] *Id.*
[5] *See* Mittellstadt, Brent et al. (2017), "The Ethics of Algorithms: Mapping the Debate", 3 Big Data & Soc'y 2.
[6] Art. 2(1) Universal Declaration, Art. 2(1) ICCPR, Art. 2(2) ICSECR, Art. 1 American Convention of Human Rights, Art. 14 European Convention of Human Rights.

suffers a disadvantage from predefined equal treatment. In extant cases, treaties on human rights do not ask for discriminatory intent, as they currently also prohibit unintentional discrimination.[7]

It is true, as the UN Committee on human rights has observed, that not all differentiations in treatment constitute discrimination. When founded on reasonable and objective criteria, there may be a legitimate goal.[8] Where legitimate purpose is missing, appropriate justification will also be missing, resulting in illegitimate discrimination; at the same time, even in the presence of a legitimate purpose, it will be illegitimate discrimination if the purpose is pursued with disproportionate means. It follows that the only case of legitimate differentiation will happen when the legitimate purpose is pursued with proportionate means. It is a double test that has been accepted in the context of human rights and which is adopted by supervisory bodies such as the United Nations Human Rights Committee. As a rule, then, the burden to prove discrimination is on the victim, whereas the burden of proving the presence of a cause of justification lies on the State.[9]

Having discussed the principle of equality and non-discrimination, one can ascertain that "algorithmic prejudices" or *bias* also exist, which are capable of causing social discrimination. Indeed, the increase of available data and individual computing capacities of AI systems risks amplifying discrimination. Until risks of this type are delineated, it is crucial to develop an *ethics of data.*[10] Aimed at this, the European Union is preparing to publish a first draft of an Ethical

---

[7] Furthermore, the European Court of Human Rights, in a case concerning the use of languages in Belgian education, has maintained that the principle of equality is undermined if the differentiated treatment has no objective and reasonable justification and that the measure that accounts for the differentiation must pursue a legitimate purpose and present a ratio of reasonable proportionality between the means employed and the objective pursued, and "*Caso relativo a certi aspetti delle leggi sull'uso delle lingue nell'istruzione in Belgio"*, sent. del 23 luglio 1968; *see generally* Focarelli, Carlo (2013), *La persona umana nel diritto internazionale*, Il Mulino, at 224-227.

[8] General Comment n. 18, 10 November 1989, par. 13.

[9] *See, e.g.*, Focarelli, *supra* note 20.

[10] *See also* Giribaldi, Davide (2019), "Intelligenza artificiale, tutti i pregiudizi (bias) che la rendono pericolosa",

Code, grounded on the premise that AI must never damage the dignity, physical security, psychological security, or financial security of human beings, animals, or nature.

In December 2018, a group of experts drew up the "Draft Ethics Guidelines for Trustworthy AI".[11] With this document, the European Commission warned of the risks associated with AI, despite its considerable advantages, and recognized the need for an anthropocentric approach to AI. This is the only approach capable of guaranteeing the dignity and autonomy of people, who must always be given the power to supervise machines.[12] Even the Council of Europe recently warned against the risk of "social discrimination" provoked by algorithms.

As examples, we can consider the risks of facial recognition systems and the use of algorithms in judicial processes. Numerous studies ("Gender shades" by the researcher Joy Buolamwini from MIT included) sustain that facial recognition can jeopardize our freedoms. Indeed, research conducted on different facial recognition systems (such as IBM Watson, Microsoft Cognitive Services and Face++) has shown that some ethnicities are treated in a more imprecise way compared to others. Notably, identification accuracy for Caucasian men was 99%, but only 34% for women with dark complexion. This is because algorithms of these systems are based on subject-data inputs which are prevalently male and of light complexion. It is evident that mistakes in programming algorithms have been committed, and it they are not easy to correct. This is due to the quantity of data analyzed by the algorithms which grows

Agenda Digitale (accessed March 14, 2019, 17:00), https://www.agendadigitale.eu/cultura-digitale/intelligenza artificiale-tutti-i-pregiudizi-bias-che-la-rendono-pericolosa/ .

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

[12] *«AI is human-centric: AI should be developed, deployed and used with an "Ethical purpose" (...), grounded in and reflective of fundamental rights, societal values and the ethical principles of Beneficence (do good), Non-Maleficence (do not harm), Autonomy of humans, Justice, and Explicability», ibid.* p. 13.

exponentially, creating errors which are deeply buried inside the artificial neuronal layers.

Another study published in 2018 by the ACLU (American Civil Liberties Union)—an American association defending civil rights—used Rekognition to analyze photos of members of the U.S. Congress in a database of about 25,000 images and demonstrated that in 5 percent of cases, an inexistent correspondence emerged between Congress members and criminals. However, what makes matters worse is the fact that 39% of the false positives included members of dark skin. Similarly, recruitment software for Amazon job candidates favored hiring males instead of females.

Let us turn to the risks for the legal system.[13] Scholars have offered examples from the American legal system where AI is used for crime prevention. The programs are developed to calculate the probability that the accused would be a repeat offender, aiding judges in establishing appropriate sentences to avoid such risk.

Furthermore, it has been shown that the *ab origine* collection of discriminatory data, such as the mapping of certain urban areas or the collection of data of potential criminals or victims, is able to consolidate prejudices, to the detriment of rights and fundamental liberties. It has been opportunely observed that to entrust a judgement on a crime to an algorithm based on the possibility that a future crime could occur is an obstruction of proper legal discipline.[14]

It is precisely in this area of predictive justice that there is a risk of massive violations of human rights through the use of such AI-based devices as so-called risk assessments tools (used in the United States) or computational tools that calculate the probability that a person will not show up for trial as scheduled or commit future crimes. These are mechanisms that examine a

[13] *See generally* Surden, Hurry (2019) "Artificial Intelligence and Law: An Overview", in *Georgia State University Law Review*, Vol. 35, at 1326-1335.
[14] *Id.*

large amount of data regarding socio-economic or family status and other factors, and identify patterns (i.e. recurrences) that purport to be more reliable than human judgements. Risk assessments are used mainly in North America at all stages of judiciary processes, from the preliminary stage of setting bail to the final stage of sentencing in the case of a conviction.[15]

A well-known tool is the "*Correctional Offender Management Profiling for Alternative Sanction*" (COMPAS), an algorithm that analyses the answers to a questionnaire of 137 items related to criminal involvement, relationships/lifestyles, personality/attitudes, family, and social exclusion.[16] The algorithm has been the subject of harsh criticism because it produces discrimination based on race, creating unequal treatment disadvantageous to individuals of color. Similarly, it creates bias related to the probability of committing crimes which affect individuals of color twice as often as individuals of lighter complexion. To eliminate the discriminatory effects of COMPAS, the Laura and John Arnold Foundation has created another tool, the Public Safety Assessment (PSA), which would eliminate the negative impact of information concerning gender, race or economic conditions. It is a tool that can assess, on the basis of nine risk factors, whether an individual will appear at trial and commit an offence if he or she is released before the trial.[17] It would reduce the risk of bias because the number of criminal convictions would have a greater influence than other assessments and criteria. PSA would be neutral in relation to race and provide its final assessment to the judge.

Risk assessment tools have also been used in the English judicial system, which uses the

[15] *See generally* Huq, Aziz Z. (2019), "Racial Equity in Algorithmic Criminal Justice", in *Duke Law Journal*, at 1043 ss

[16] Kehl, Danielle/Guo, Priscilla/Kessler, Samuel (2017), "Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiatives, in Berkman Klein Center for Internet & Society, Harvard Law School, at 9.

[17] See also Kleiberg, Jon/Lakkaraju, Himabindu/Leskovec Jure/Mullainathan, Sendhil (2018), "Human Decision and Machine Predictions", in The Quarterly Journal of Economics, at 237.

Harm Assessment Risk Tool (HARM) system for predictive assessments aimed at reducing the risk of recidivism. This tool has not been free from criticism in terms of violation of privacy, as it takes into account 34 variables, including those related to criminal records, age, gender and postcodes of residence of the individual.[18]

Thus, there have been numerous requests from scholars for the rapid development of an "ethic of data", precisely what Europe is undertaking through the European Commission body. Similarly, the United States, through the governmental agency of the Defense Advanced Research Project Agency (DARPA), which has the duty of developing new technologies for military use, has been developing tools to instill ethical norms in AI machines, through a $2 billion program.

In order to avoid other scandals such as Cambridge Analytica,[19] the future of AI largely depends on the ability to solve the problems that are inherent in the increase of data available to the machines and their calculation capacity. The big names of Silicon Valley are already working to reduce these risks linked to the prejudices that are hidden inside AI systems.[20] Yet, the problem remains that when fundamental rights come into play, it is difficult to entrust them to the decision of an algorithm because discretionary and ethical assessments typical of human elements are paramount.

---

[18] *See* Gialuz, Mitja (2019), "Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei 'risk assessment tools' tra Stati Uniti ed Europa", in *Diritto Penale Contemporaneo*, at 3-12; *see also* Barile, Fabio (2019), "Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine", in *Diritto Penale e Uomo*, at 16-19.

[19] The scandal shook the world of technology in 2018, with over 87 million items of personal and confidential data mysteriously passing from Facebook to Cambridge Analytica, a lobbying company founded by an American billionaire who, along with Steve Bannon, used that data to influence the American presidential vote in 2016. Cambridge Analytica took Facebook data and used them to create extremely detailed online profiles which were used to interact in a realistic way with the online community in order to manipulate public opinion in what is known as "Behavioural MicroTargeting".

[20] Giribaldi, *supra* note 11.

AI and the Protection of Human Rights in Europe

A study published in 2018 entitled "Algorithms and Human Rights-Study" on the human rights dimension of automated data processing techniques and possible *regulatory implications* was the basis for the European Ethical Charter regarding the use of AI in the judicial system, adopted by the Commission for the Efficiency of Justice (CEPEJ). The concern, in fact, was that the use of AI in this field would not violate the right to a judge and the right to a fair trial through the presumption of innocence, equality of arms and respect for the contradictory, but also the right of non-discrimination, given the use of sensitive data in predictive judgements of responsibility, such as racial or ethnic origin, political opinions, religious or political beliefs, socio-economic conditions, or data related to health or sexual orientation. In this sense, the right to a judge, in accordance with Article 5 of the European Convention on Human Rights takes on the meaning of the right to the physical presence of a judge, which therefore cannot be replaced by an algorithm.[21]

In the face of predictive techniques of analysis which are quite invasive, and the discriminatory risks connected to algorithmic choices, the problem of the ethical impact of AI

[21] In this regard, it should be noted that Article 15 of Directive 95/46/EC prohibits decisions based solely on automated processing, whereas Article 11 provides that *"Member States shall provide that a decision based solely on automated processing, including profiling, which produces adverse legal effects or significantly affects the data subject is prohibited unless it is authorised by Union law or by the law of the Member State to which the data subject is subject and provides adequate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention from the data subject."*, Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regards to the processing of personal data and on the free movement of such data (hereafter "Data Protection Directive"). This provision should be read in conjunction with Articles 5 and 6 of the C.E.D.U. on the right of access to the judge. However, automated decisions cannot be based on particular categories of personal data, such as those revealing racial or ethnic origin, political opinions, religious or philosophical beliefs or trade union membership, as well as genetic, biometric or health data, or data relating to the sexual life or sexual orientation of a person, unless there are adequate measures to safeguard rights, freedoms, and legitimate interests of the data subject. This, therefore, excludes for Europe a scenario similar to that of North America, since there is a set of rules, both at Council of Europe and European Union level, regulating the role of AI in decision-making processes, which remain in the hands of the individual to whom AI provides valuable but limited assistance.

arises. In this regard, there is a question of who exploits the role of data ethics, with all the possible superimpositions between ethical and legal aspects, wherein ethics are called to integrate the law. Ethical evaluations are requested when a use of AI that is compliant with the law intersects the ethical-social values of society. An example could be the management of smart cities through algorithms, where automation poses both ethical and legal questions. In both cases regulation is necessary.

But a new law is not necessarily a solution. Existing laws are able to address different legal aspects inherent to these issues, such as that of civil responsibility to protect personal data. It is also true that it is necessary to consider that existing regulatory models were formed between the 1970s and 1980s and thus were created under vastly different social contexts, therefore highlighting their current unsuitability. Thus new regulation is necessary, especially with hard law, such as laws and conventions. Unfortunately, these types of legal changes are slow to come about, and this is in clear conflict with the very rapid evolution of technology. In the European environment, an example is the General Data Protection Regulation (EU), 2016/679, better known as GDPR, which was adopted on 27 April 2016, published on 4 May 2016 and passed as a law on 25 May of the same year, and in effect from 25 May 2018, but *in nuce* in 2011. One must consider that enactment takes many years. This is alarming considering the fact that over the course of a decade two entire technological generations can pass. The pace of regulatory change is too slow to keep up with that of technology.[22] It is evident that regulatory systems are always outdated in respect to technological progress, and there is no lack of criticism in academia regarding the GDPR and its failing to provide a clear remedy for the risks posed by

[22] *See, e.g.*, Mantelero, Alessandro (2019), "Come regolamentare l'intelligenza artificiale: le vie possibili", Agenda Digitale (accessed March 10, 2019, 17:00), https://www.agendadigitale.eu/cultura-digitale

the algorithmic "black box" which cannot be opened.[23]

AI and Digital Security: the Protection of Personal Data Online

Every second, billions of Internet users give big digital operators fantastic amounts of personal data, transmitted over social networks, equating to an annual market value of $1 trillion dollars. This confirms Metcalfe's Law, according to which the value of a network grows exponentially in relation to the number of users. In this case, the law also applies to the added value given to AI by every user, for example, of a social network like Facebook. From this immense social, economic and emotional heritage, the big digital operators create the world of AI. But how is this patrimony of data and the rights that are at its roots protected?

We are speaking, *in primis*, about personal data, because the protection of these data is one of the sectors most involved on a daily basis by the arrival of AI systems.[24] The operation of these systems, indeed, is based precisely on elaboration, analysis and treatment of big quantities of information, in particular personal data, data that travel on the net. [25] But, many risks lurk in this same space.

Online mass checks, data theft, phishing and malware are all risks to our digital security, a security which is interconnected with a number of rights. First and foremost, there is the right of privacy, which also implicates other rights such as that of expression or freedom of peaceful

[23] *See also* Edwards, Lilian & Veale Michael (2017), "Slave to the Algorithm? Why a Right to an Explanation' Is Probably Not the Remedy You Are Looking For", *16 Duke Law & Technology Review 18*, at 18 ss.
[24] According to art. 4(1) of the GDPR, it is defined as personal data *« any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person ».*
[25] *See generally* Anselmi, Niccolò & Olivi, Giangiacomo (2019), "Intelligenza artificiale e privacy, i 5 punti critici di una relazione pericolosa", Agenda Digitale (accessed March 14, 2019, 16:30), https://www.agendadigitale.eu/cultura-digitale

assembly and association. A number of rights are therefore called into question online. When we transmit information about our movements or habits through a mobile phone, our right to privacy is called into question. When we participate in online public debates and express our opinion, we exercise our freedom of expression. When we conduct online searches on a subject of our interest, the right to seek and receive information takes over. Finally, when we use an application to agree to participate in a public demonstration, we exercise our right to peaceful assembly. In all these cases, human rights are at stake: those minimum standards capable of preserving human dignity, which are all so interconnected, nonhierarchically ordered and interdependent imply that the violation of one jeopardizes the enjoyment of others. In the online space, proper and effective digital security ensures the protection of these rights.

Making a purchase, a reservation, or expressing a "like", are all actions that can provide, more or less consciously, information about ourselves. Today, you can communicate information through the internet in countless ways or have easy access to a large amount of data. Every time we put information on the net, these little segments of our lives are brought together to paint a picture of who we are, our tastes, our beliefs, our movements, and so on. The diffusion of personal information is not an end in itself, even now that the big firms of Silicon Valley have not limited themselves in just using our data to predict human behaviors, but have gone so far as actually attempting to modify them. The economic imperatives of giants like Amazon or Facebook erode democracy with their systems, reducing individual awareness, decision-making skills and Internet users' ability to react.

In the digital world, the right to privacy is protected by international treaties such as the Universal Declaration of Human Rights (art. 12) and the International Covenant on Civil and Political Rights (art. 17), in addition to regional instruments such as the Charter of Fundamental

Rights of the European Union (art. 7,). However, the right to our privacy is constantly undermined by the use of the Internet, which constantly is fed more and more information. This can be provided with our consent, but it can also be fraudulently extracted and used by criminal networks to extort us for money, by governments to carry out mass checks and surveillance or through more mundane ways, such as by companies modeling their advertisements according to our personal profiles.

While digital communication has, on the one hand, revolutionized the world of work and interpersonal relationships, it has, on the other hand, made our privacy more fragile, making it more permeable to violations. As we have said, interference in our right to privacy may involve the violation of other rights, such as freedom of expression, precisely because of the interdependence and interconnection among human rights; the violation of one right threatens all others. Freedom of expression is guaranteed, among other documents, by the International Covenant on Civil and Political Rights (art. 19.). No discrimination on grounds of nationality, gender, genetic characteristics, ethnic or social origin, religion, language, political opinions, property, disability, age, or sexual orientation is allowed in the exercise of that right, or any other status. Freedom of expression through the web can only be *constrained* by law and in such a way that its limitation is necessary and proportionate to a legitimate purpose, such as, for example, the protection of the national interest.

This right may be violated by mass controls—contrary to both the right to privacy and the freedom of expression—which act as Bentham's Panopticon, causing users to censor themselves for fear of being watched. In this sense, freedom of expression is restricted because it does not allow the user to express himself or herself freely on the web. Thus, freedom of expression is closely linked to the right to privacy in the digital world, because if you have the perception that

your privacy is preserved, you have a tendency to express yourself more freely and vice versa.

In Europe, limits to intrusiveness of digital evaluation on citizen's rights were posed by the GDPR. The Regulation reduces the freedom of companies in their management of data. Moreover, it is necessary to observe, as we have seen, the lack of a unified, communal regulatory body and the presence of different European CNIL (the national committee of informatics and liberties) in charge of regulating the global data-banks.

It should be stressed that there is a discrepancy in remedies between data protection and algorithm regulation, as the former involves individual rights—being human rights—and the latter involves the risks associated with algorithms, which affect groups of people.[26] The same discrepancy concerns other aspects. The regulatory authorities have requested that the reasons for collecting and processing data must be justified *a priori*, but deep learning is not readily subject to this type of regulation because it is a process of discovering correlations in sets of data that might otherwise not be apparent. Thus, restricting data collection is detrimental to those who run deep learning programs.

The GDPR is expected to widen the legislative gap between Europe and the free reign of online web giants from America and China, especially seeing as there is no comparable digital giant of European-origin. Paradoxically, strict legislation on competition and privacy protection leads to European digital subordination. The ongoing legislative battle is over the proper level of privacy and technological freedom, and it would be an error to focus only on consumer protection. In the United States and China, priority is being given to protecting the interests of large digital industries. If consumer protection has primacy in Europe, it would suffocate AI operators, in effect stifling the emergence of meaningful and relevant European, technological

[26] *See* Edwards Lilian & Veale Michael, *supra* note 24, at 22.

startups. Despite Europeans having the strongest legislation in the world for consumer protection and privacy, its lack of digital industry rights makes it a *de facto* colony of the American and Chinese AI industries. To date, there are still no European *unicorns*: technology startups valued from at least one billion dollars, and are as rare as the mythological creatures for which they are named.[27]

However, according to the principles and provisions of GDPR, some criticisms concerning personal data protection arise. First, is the problem that in Europe, AI functions can be carried out only for specific purposes. This means attention is focused on protections against intended uses of AI. But, while this is necessary and appropriate when AI processes data for predetermined purposes, it is ineffective or inapplicable with respect to machine learning systems that have the ability to adapt and consequently change its behaviors. In this case, data can end up being used for purposes other than those set out in advance and beyond the purview of either data subjects or data controllers.

The second aspect to review concerns the legal basis in processing data. In addition to pre-established purposes, processing can only take place if there are adequate legal bases that make that processing lawful.[28] So, when data processing does not conform to the defined legal basis expressed in the contractual obligations between the data subject and the data controller, which happens when the AI system escapes proper human (or algorithmic) oversight, it is difficult to find an additional legal basis. It is important to underline Rule 22 of the GDPR which establishes the need for a human subject behind automated processes, in order to protect the data

[27] *See* Laurent, Alexandre (2017), *La guerra delle intelligenze. Intelligenza artificiale contro intelligenza umana*, EDI, at 22-25.
[28] Art. 6, GDPR.

subjects' rights and freedoms and legitimate interests.[29]

The third aspect which comes into focus is the issue of clarity of roles. And indeed, the functioning of AI systems presupposes the involvement of a large numbers of subjects (e.g. data subjects, data controllers, providers of ancillary services, third parties to whom data are disclosed for certain purposes and can become either data processors or even the new data controllers themselves etc.). On closer inspection, in the AI sector, it often happens that the privacy roles of each subject are not well defined. Moreover, the processing of multiple kinds of information by AI systems means that it is not uncommon for such systems to also obtain sensitive personal data, such as information about health condition or sex life attitudes, from the processing of non-personal data.[30]

The last aspect to be analyzed is that of controls and audits, since the GDPR foresees that appropriate audits should be carried out against those who process personal data. It must be recognized that it is not always possible to carry out controls on the functioning and processing of personal data placed on AI systems. That information is often inaccessible to the data subjects who freely give up their data which is in breach of the rights established by the GDPR that entitles the data subjects to receive information about the personal data processed or the transfer of such data to third parties and so on.[31]

These, then, are the main problems related to AI and data protection that, if they remain unaddressed in accordance with a strict reading of the ethical and normative canons determined by the GDPR and other sources, can turn into unfair, advantageous opportunities for the

[29] Art. 22, GDPR.
[30] *Id.* Art. 9 GDPR.
[31] *Id.* Artt. 16-22 GDPR.

operator.[32] Regarding the GDPR, recital n. 75 speaks to the risks to the rights and freedoms of physical persons (agents) that may result from the processing of personal data at the discretion of out-of-control AI systems or in the wrong hands and, therefore, liable to cause physical, material or immaterial damage. In particular, the recital warns against data processing which may involve discrimination, theft or misuse of identity, financial loss, damage to reputation, loss of confidentiality of personal data protected by professional secrecy, unauthorized decryption of pseudonymization, or any other significant economic or social damage. It also provides protection for data subjects at risk of being deprived of their rights and freedoms or being prevented from exercising control over personal data related to them, as well as for subjects whose personal data would likely be processed to reveal racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, health and sex-life data, or data related to criminal convictions and crimes or related security matters. It also recognizes that data processing can be used to create or use personal profiles by analyzing or forecasting behaviors relating to professional performance, economic situation, health, preferences or personal interests, reliability or behavior, location or travel, and that data processing risks are especially acute for vulnerable populations like children and the elderly.

Thus, in recital 76, it is stated that the probability and severity of the risk to the rights and freedoms of the data subject should be determined with regard to the nature, scope, context, and purpose of the processing. The risk should be considered on the basis of an objective assessment where it is to be established if the data processing operations involve a risk or a high risk. Recital n. 77 subsequently encourages adopting a code of conduct, approved certifications, guidelines provided by the Committee, which may also issue guidelines on processing operations that it

[32] *See* Anselmi & Olivi, *supra* note 26.

considers unlikely to pose a high risk to the rights and freedoms of physical persons and whose measures may be sufficient in such cases to ensure that these risks are addressed.[33]

It is clear that while in the past the aim was to protect the private citizen from State interference and abuse of power, protection against the misuse of personal information at the present calls into question the role of individuals who often voluntarily offer their own personal data to private companies in exchange for advantages. Internet users, in fact, make possible, willingly or otherwise, the reconstruction of their own individual profile through cookies, geo-tracking, and consent to the sale (or sometimes fraudulent acquisition) of their own data.

Scholars have pointed out that human rights lose their meaning when rights to privacy can be traded like any other commodity in exchange for money or other advantages.[34] The free sale of privacy ends up allowing totalitarian control by those who manage this information to learn about, pilot, and guide, through statistical analysis, the personal choices of the same users in exchange for utility. This is information given to the "public web record" on the precondition of democratic participation in online life. In this way, the logic underlying human rights would be reversed, as they would be invoked to protect individual choices as an expression of freedom, with the eventual result of being manipulated by corporate powers once they have acquired their personal information. The concept of "inalienable" human rights, which comes into play when a political authority is able to prohibit the sale, or even the free transfer of human rights, is

[33] For its part, recital 78 states that the protection of the rights and freedoms of natural persons with regard to the processing of personal data requires the adoption of appropriate technical and organisational measures. It also provides for the protection of the rights and freedoms of natural persons. For which *« the controller should adopt internal policies and implement measures which meet in particular the principles of data protection by design and data protection by default. Such measures could consist, inter alia, of minimising the processing of personal data, pseudonymising personal data as soon as possible, transparency with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing, enabling the controller to create and improve security features ».*

[34] *See generally* Focarelli, *supra* note 20, at 170-172.

therefore called into question, despite the implicit consent of the data subject.

The State thus now seems destined to succumb to the free will of the individual regarding the control of his or her own privacy and data rights. Yet, these rights, unfortunately, are in effect being transferred to centers of political and economic power (the creators of AI) which are then able to manipulate them in order to redirect their choices as individuals. There is no longer a Big Brother of the State that watches us, but rather individuals increasingly eager to be supervised.[35]

The European and International Alert System on the Protection of Human Rights and Fundamental Freedoms with Respect to the Insides of Algorithms

The Committee of Ministers of the European Union adopted, on 13 February 2019 at the 1337th meeting of the Ministers' Deputies, a declaration on the manipulative capabilities of algorithmic processes. The concern of the Committee of Ministers are the growing threats to the rights of human beings to form opinions and make decisions independent of algorithmic systems and other advanced digital technologies. The Committee affirms that attention must be paid particularly to the capacity of digital technologies to use both personal data and non-personal data to identify individual vulnerabilities, and it thus encourages member states to shoulder responsibilities for addressing this threat by such measures as: initiating informed and inclusive public debates with a focus on providing guidance to define the difference between permissible persuasion and unacceptable manipulation; taking appropriate and proportionate measures to ensure that effective legal guarantees are in place against such forms of illegitimate interference; and empowering users by promoting critical digital literacy skills, specifically, public awareness of the fact that algorithmic tools are widely used for commercial and political reasons, as well as the purposes of inflicting harm and with anti- or undemocratic intent.

[35] *Id.*

At present, machine-learning tools have the growing capacities, not only to predict choices, but also to influence emotions and thoughts and alter anticipated courses of action, often subliminally. Before you go to make a purchase, Alibaba already can predict what you will buy and in this sense you can be the beneficiary or victim of algorithms and their ability to capture information. In this way, Cambridge Analytica used information from Facebook to capture and then shape the voting intentions of American voters during the 2016 presidential campaign. *"There's no data like more data"* is the motto coined by the founder of Cambridge Analytica.

The Committee underlines the dangers for democratic societies that emanate from the possibility to employ such capacity to manipulate and control not only economic choices but also social and political behaviors (which has only recently become apparent). In this context, particular attention should be paid to the significant power that technological advancement confers to those, both public entities and private actors, who may use such algorithmic tools without adequate democratic oversight or control.

Fine grained, sub-conscious and personalized levels of algorithmic persuasion may have significant effects on the cognitive autonomy of individuals and their right to form opinions and make independent decisions. Such effects remain under-explored, but they should not be underestimated. Not only may they weaken the exercise and enjoyment of individual human rights, they may erode the foundational pillars of the Council of Europe since its central values of human rights, democracy and the rule of law are grounded on beliefs in the equality and dignity of all humans as independent moral agents.

On the international stage, the OECD has articulated five basic principles for regulating AI. This general agreement aimed at setting standards has been signed by 36 Member States, including the world's major economies, but excepting China and six nonmember states:

Argentina, Brazil, Colombia, Costa Rica, Peru and Romania. The first principle in the agreement is that AI must benefit both people and the planet by enabling inclusive growth, sustainable development, and shared welfare. The second principle states that AI systems must be designed with respect for law, human rights, democratic values and diversity, as well as including safeguards that allow human intervention. The third principle makes it clear that AI systems must be transparent and there must be a clear understanding of how they work. The fourth states that they must operate in a stable and secure manner throughout their existence and that the potential risks can be assessed continuously. Finally, the last principle requires that organizations and individuals developing, distributing or operating AI systems are responsible for their proper functioning in line with the above-mentioned principles.

According to some critics, the document, as ambitious and strongly desired as it is, also contains some inconsistencies. One of the most debated issues is the accountability of algorithms since many AI systems incorporate software that can learn autonomously or make decisions without human intervention. In this sense, it is often more difficult to open the black box of deep learning software to understand the ultimate reason for a decision. In this regard, the OECD principle of transparency could be interpreted as an obligation, which countries could include in their legislation, only to develop autonomous software that is always comprehensible to man. Alternatively, it could be interpreted as recommending a hybrid approach in which AI systems make recommendations, but not decisions—an approach in which humans have the last word and take responsibility for choices.

The point now is to get OECD principles translated from policy to business, to put the principles into practice. For this reason, starting in autumn 2019, the OECD website will publish a sector-based observatory of good corporate practices and solutions with the aim of developing

a set of rules for businesses to follow, in accord with an ethics of technological innovation in which AI remains secular, democratic and without preconceptions.

However, according to skeptics, it is too early to implement such rules because AI is still nascent. Indeed, this explains the current vagueness of laws written regarding AI. Granted the difficulty of imagining in advance the different applications of AI, it has been argued that it would be better to create rules once AI applications appear on the market and violate existing laws. Rules and sanctions should apply to those who use AI for illicit purposes, always bearing in mind that it is not the technology itself that should be culpable, but the person who misuses it and distorts it.

It is primarily up to each individual government to protect citizens with appropriate laws from the pitfalls of the web and the power of Palo Alto or that of the Zhongguancun high tech district in Beijing through ratifying binding international treaties, enforcing sanctions against offending States, including indiscriminate use, either public or private, of surveillance policies. There is also the non-negligible role of international human rights courts, which monitor fundamental rights,[36] and human rights organizations, which discipline governments, helping to enforce proper respect of universal rights. Finally, there are bodies such as the Special Rapporteur on privacy and freedom of expression, independent experts appointed by the United Nations to monitor compliance with human rights standards around the world, who submit reports to the Human Rights Council. In addition, independent watchdog groups at the national level can also help guarantee compliance in ways that are not negligible.

At the company level, the ethical principles agreed upon by the American Business

[36] By way of example, on 13 September 2018, the European Court of Human Rights ruled that UK laws allowing mass surveillance violate the right to privacy and freedom of expression.

Round-table are to be welcomed. Including such leaders of American capitalism as big telecommunications and digital giants like Apple, these principles speak on behalf of corporate accountability, not only to shareholders, but also to stakeholders and other members of society who are affected by corporate decisions on AI, including workers, the environment, society as a whole, and consumers. It is potentially a socially conscious value revolution that opposes neo-liberal ideology with the rise of predatory digital capitalism. One hopes that these initiatives will also arrive in Europe and that doing so will have positive repercussions on the respect for human rights, in practical terms, as the theoretical promises have made believe.

The Chinese Threat: The Enjoyment of Human Rights in the Face of Invasive AI

China, the world's second largest economy, has an ambitious AI strategy to become the global leader by 2030, while already being on target to outperform the US in academic research in the field by this year (2019). In this sector, geopolitics play a leading role. The United States uses its natural products of capitalism, Wall Street and Silicon Valley, to lead the way in AI advancement. This contrasts China's approach which is evidenced by heavy public expenditures to finance public projects which do not seem to pay regard to fundamental rights of individual privacy. Yet, despite their differences in approach, both governments have been reluctant to pass legislation or regulations on the use of any AI technology.

This rivalry is similar to the Cold War. The United States feels that its AI technology is superior to that of China, showing a certain "complacency" about its own position. The USA falsely believes that China is capable of advancing its own AI technology only through Silicon Valley. Yet, as Kai Fu Lee opines in *AI Superpowers*, the United States is particularly susceptible to a *technological takeover* by China, a nation equipped with a population of nearly 1.5 billion, with over 600 million online agents to collect data from, and an authoritarian

government that imposes on itself no limits to privacy violation.

"See far, go further" is the motto of Hikvision, the Chinese company leader in the field of facial recognition, alongside Megvii, iFlyTek, Zhejiang, Dahua Technology, Meiya Pico and Yitu Technology. Hikvision works specifically with drones rigged with cameras with facial recognition technology. Facial recognition is a field in which AI is making important steps and the Chinese see facial recognition as a strategic advantage. This was one of the features of Google Glass presented in 2013, which turned people's faces into business cards, revealing their identity. Invasive applications of facial recognition can result in amoral or illicit uses, as would a system capable of creating personalized advertising based on facial recognition, turning a subject's face into a spam platform.

Up until now global companies have been cautious to exploit applications of such technology, though it does exist (Facebook has "Deep Face", Amazon "Rekognition", Apple "Face ID" for the iPhone). The Chinese, however, are among the world's greatest fans of facial recognition and have been intent on selling this technology around the world. China's facial recognition technology exists in many consumer products but this has recently been met with disapproval from certain nations that have banned their sale, notably the United States. Yet, China still sells these commercial products worldwide. Additionally, China sells its facial recognition technology to authoritarian governments who wish to track their own citizens. This Chinese tech is relatively inexpensive to acquire and works quite well, being employed furtively, without public detection or uproar.

The Chinese Government seems to want to include its 1.4 billion citizens in a database, possibly available to intelligence services, to control the population against possible disturbances. But, such a database can easily be misused by the government itself. The gigantic

Chinese archive would be vulnerable to being hacked by enemies of the State which would compromise the identities of Chinese citizens and their movements, and provide data-points which are geo-tagged and timestamped.

This mapping, with different modalities in relation to the different protection of privacy, already happens in the West. Smart cameras that can count the flow of passing people reveal each person's age, gender, idle time and reaction to a given context. These data are then categorized and can be searched, even on the smallest detail such as the color of one's clothes, hair, or shoes. A few hundred of these devices are enough to keep an entire metropolis under surveillance, with an average cost of 20 euros per device. Therefore, only modest outlays are required when compared to the immeasurable value of the service rendered. Moreover, the technology does not change between it use in the management of urban dynamics and the collection of data for commercial or security purposes.

In Los Angeles, thanks to sensors, commuter drive time was reduced by 15 percent by applying a different management of traffic lights. Transforming a city's life into digital information through sensors and AI has undoubted benefits as one can better manage traffic, waste collection, electricity and water, etc. There are 245 million security cameras active in the world, and according to the European Investment Bank the smart sensor market is worth $57 billion. But due to low-cost production, many of these systems are unreliable and run the risk of feeding the temptation to use these devices' data mining capabilities for illicit uses, such as those that would allow widespread mass control—as some would argue is already the case in China.

Predicated on the engineering power (electricity) of the 21st century, the AI industry rests on four fundamental factors: a great mass of data, aggressive entrepreneurs, specialized scientists and favorable policy. Starting from these four factors it is possible to determine, between China

and the US, which is likely to dominate AI globally, establishing a new bipolar world order wherein humans will coexist with AI. China's Zhongguancun is not yet the global competitive equal of Silicon Valley. For much of the digital revolution China was considered able only to copy American technology (just think that in 1975, when Microsoft was founded, China was experiencing a period of intellectual suppression due to the Cultural Revolution and that in 1992 only 0.2% of the Chinese population was connected to the Internet, compared to 30% in the United States). But today China has an immense amount of data, which it draws from the real world, gathering all the information of users—their daily habits, localizations and so on—and in this sense it has already far exceeded the use of AI in the United States. The AI balance thus arguably hangs in favor of China.

Chinese data gathering supremacy generates a vicious circle wherein more data produces better products which, in turn, will produce more users who, in turn, produce more data which, in turn, improve products and so on, exponentially.[37] If today, AI is the new electricity, Big Data is the oil that turns on the generator and China is the largest digital data producer. Its advantage is not only quantitative, since it has more data than Europe and the United States put together, but also qualitative, since it is not just the number of users at stake but also what they do online, which is thoroughly and constantly scrutinized. The Internet universe has pervaded the Chinese economy. But, this was possible due to the intervention of a leading actor: the Chinese government. This action introduced the concepts of mass entrepreneurship and mass technology, to which we also add mass data surveillance. China offers favorable taxation and investment incentives for private technology start-ups, as well as significant public-private partnership

[37] *See generally* Kai-Fu, Lee (2018), *AI Super-Powers. China, Silicon Valley and the New World Order*, Houghton Mifflin Harcourt, at 14-50.

opportunities with significant public funding. If a private enterprise fails, the State is ready to eat the loss in the face of having taken the risk. State incentives guide business choices, which inherently follow government agendas. The US technology market, on the other hand, seeks to remain independent, insisting on greater separation between the public and private spheres, all of which tilts the AI balance to China.

There are significant ethical issues related to AI, such as the kinds of choices that machines will make in certain circumstances, to which both superpowers, China and the US, respond differently on the basis of their own scale of values. For China, ethical issues are important problems to address, but not enough to hold back technological implementation. For example, the use of AI in medicine (such as the ability of machines to make precise diagnoses or the creation of bio-banks, access to which would be given to doctors to carry out more effective clinical research into diseases, wherein citizens would renounce the confidentiality of their own data in favor of the common good) or in public order (think of the predictive algorithms that manage to estimate in advance how and where criminal episodes may happen) could save hundreds of lives. Promoting social good is more than enough reason for the Chinese government to stimulate technological development. It is a techno-utilitarian approach, where technological development goes hand in hand with economic development.[38]

The United States and China are aware that leadership in AI represents a huge competitive advantage, not only in the development of autonomous weapons and defense systems but especially in economic terms. While the US leaves room for private industry to exploit citizens' data in pursuit of economic growth, China conceives AI as a control and management tool for and of citizens, and focuses on government applications such as the Social

[38] *Ib.* 102.

Credit System (SCS) that will likely be mandatory for all Chinese citizens from 2020 and onwards. The purpose of the SCS is to track, evaluate, reward and punish citizen behavior, including their online conduct, taking into account parameters such as credit history, buying habits, online friends and public comments on social media, the ability to fulfill their contractual obligations, and so on. This is a move toward an all-pervasive AI.[39]

Americans, while open and permissive in the digital field, are not very open to the spread of pervasive innovations. In contrast, the Chinese public has largely embraced pervasive digitalization. It is now common in China, for example, to make purchases without money or credit card by visual identification or voice recognition alone. To regulate traffic, many Chinese cities have a myriad of sensors and cameras that store images, while in the US, one is less willing to accept these mass controls that restrict privacy for a presumptive public benefit.[40] The same applies to Europe, which has adopted the GDPR and, in general, is more concerned about monopolies, digital security and algorithmic biases. In sum, one might argue that in China the protection of privacy and human rights in general gives way to profit and utilitarianism. It should, however, be recognized that in 2017 China adopted a cyber-security law introducing sanctions for illegal data collection.[41]

A major concern globally is that AI systems can operate without human control and adapt to changing situations. For example, Google's search function is a self-improving system, as its machine learning algorithms constantly regulate and update the results of searches carried out by users. The possibility of systems getting out of control is quite real, especially if the system

[39] Pozzi, Cristina (2019), *Benvenuti nel 2050*, Egea, at 39.
[40] Although intelligent video surveillance systems are becoming increasingly popular in the USA. In addition, they are being distributed, for example, in schools to prevent armed attacks, but also to study student behaviour and to identify in advance those who might resort to violence. It is a sort of "predictive police" on a large scale.
[41] *Kai-Fu Lee*, *supra* note 38, 124-125.

design does not set operational boundaries related to the use of the system itself. The countermeasure is to set professional standards for the development and control of intelligent systems. Designers should accurately predict the operating scope of a system and provide ways to limit the risks related to possible overshoot of its intended operational scope. Intelligent systems should be able to independently monitor whether their own operation is within the limits set by their designers and enter "safe mode" or proceed to a self-monitoring shutdown, if those limits were to be exceeded. It is also possible to send an alert to a human supervisor as a security mechanism, in addition to the aforementioned provision, akin to the requirement of a badge of access by a State institute.

According to Laurent Alexandre (2017), around 2080, the world will be dominated by AI, which will tend to merge living beings and intelligence, and force humanity to defend the perseverance of the physical body to avoid its dissolution into the virtual world. What is certain is that AI must be educated and inculcated with ethical norms since the more autonomously AI operates, the more it will both pose and be called to solve moral dilemmas. In setting these norms, however, considerations of human rights should be both fundamental and paramount.

References

Anselmi, Niccolò & Olivi, Giangiacomo (2019), "Intelligenza artificiale e privacy, i 5 punti critici di una relazione pericolosa", Agenda Digitale

Barile, Fabio (2019), "Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine", in *Diritto Penale e Uomo*

Cagle, Matt & Ozer, Nicole (2018), "Amazon Teams Up With Government to Deploy Dangerous New Facial Recognition Technology", ACLU, July 15, 2020. Available at: https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazon-teams-government-deploy-dangerous-new

Edwards, Lilian & Veale Michael (2017), "Slave to the Algorithm? Why a Right to an Explanation' Is Probably Not the Remedy You Are Looking For", *16 Duke Law & Technology Review 18*

Focarelli, Carlo (2013), *La persona umana nel diritto internazionale*, Il Mulino –

Gialuz, Mitja (2019), "Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei 'risk assessment tools' tra Stati Uniti ed Europa", in *Diritto Penale Contemporaneo*

Giribaldi, Davide (2019), "Intelligenza artificiale, tutti i pregiudizi (bias) che larendono pericolosa", Agenda Digitale

Huq, Aziz Z. (2019), "Racial Equity in Algorithmic Criminal Justice", in *Duke Law Journal*Laurent Alexandre (2017), *La guerra delle intelligenze. Intelligenza artificiale contro intelligenza umana*, EDI

Kai-Fu, Lee (2018), *AI Super-Powers. China, Silicon Valley and the New World Order*, Houghton Mifflin HarcourtKaplan, Jerry (2018), *Intelligenza artificiale. Guida al futuro prossimo,* Second edition, Luiss University Press

Kehl, Danielle/Guo, Priscilla/Kessler, Samuel (2017), "*Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing, Responsive Communities Initiatives*, in *Berkman Klein Center for Internet & Society, Harvard Law School*

Kleiberg, Jon/Lakkaraju, Himabindu/Leskovec Jure/Mullainathan, Sendhil (2018), "Human Decision and Machine Predictions", in *The Quarterly Journal of Economics*Mittellstadt, Brent et al. (2017), "The Ethics of Algorithms: Mapping the Debate", 3Big Data & Soc'y 2Mantelero, Alessandro (2019), "Come regolamentare l'intelligenza artificiale: le vie possibili", Agenda Digitale Pariotti, Elena (2013), *I diritti umani: concetto, teoria, evoluzione*, CedamPozzi, Cristina (2019), *Benvenuti nel 2050*, Egea

Rodotà, Stefano(2007), *Dal soggetto alla persona*, Editoriale Scientifica

Rodotà, Stefano (2005), *Tecnologia e diritti*, Il Mulino

Surden, Hurry (2019) "Artificial Intelligence and Law: An Overview", in *Georgia State University Law Review*, Vol. 35

Humane Artificial Intelligence

Humane Artificial Intelligence is an intercultural, intergenerational, and interdisciplinary initiative of the East-West Center (EWC) to engage the societal challenges and opportunities that are emerging with advances in artificial intelligence, robotics, machine learning and big data. The purposes of the initiative are: to establish an ongoing forum for exploring differences in understanding the goal of aligning artificial intelligence with human values and societal wellbeing; to foster globally-shared commitments to equitable and humane artificial intelligence; and, to engage emerging perspectives on the evolving interplay of people, artificial intelligence, and related technologies.

The views expressed in this article are those of the authors and do not represent the views of the East-West Center.

EAST-WEST CENTER