Governing Artificial Intelligence

di **TULIO ROSEMBUJ**

SUMMARY: **1.** THE BASICS. – **2.** THE BIG DATA: FROM COGNITIVE TO SURVEILLANCE. – **3.** THE ALGOROTHMIC SOCIETY. – **4.** ALGORITHMIC PRINCIPLES. **5.** – ABOUT THE GOVERNANCE COMMON PRONCIPLES.

"I think we should be very careful about artificial intelligence.If I had to guess at what our biggest essential threat is, it´s probably that…I `m increasingly inclined to think there should be some regulatory oversight,maybe at the national and international level, just to make sure that we don't` do something very foolish."(*Elon Musk).*
(Aileen Graef, Elon Musk:We are "Summoning a Demon" with Artificial Intelligence, UPI,October 27,2014).

1. The basics. *Marvin Minsky* defines artificial intelligence as "the science of making machines do things that would require intelligence if done by (people)."

There is a machine, a device, an artifact doing things that represents tasks to be performed and goals to be obtained; which, at some grade, will be similar to human acts.

Beyond the metaphors, which may be more or less appropriate – the machine is not a human being-, the fact is that there are computational processes formulated as computer programs, capable of communication, representing knowledge and carrying out automatic analysis to reach conclusions. The computer program needs to get data, find patterns, and do predictions to achieve its programmed outcome.

Soshana Zuboff observed that information technology has a duality which she calls "computer mediated", and it means automation and information at the same time. The computer, in this sense, is not only the medium of information but it also produces information.

The algorithm is the core of this computer process and computer programs, which means a specific sequence of logical operations that provides step- by-step instructions for computers to act on data and take predetermined

automated decisions. The outcomes are based on data inputs and decisional parameters.

Artificial intelligence is a computer program – an algorithm- that makes and executes decisions in response to external circumstances, (LYNN M.LOPUCKI*).*

The algorithm could change in response to its output and improve with the experience, learning with data, or extracting patterns from data. Machine learning is a statistical process that starts with a bulk of data and tries to derive a rule or a procedure that explains the data or can predict future data.

The definition is clear: “Machine learning algorithms can figure out how to perform important tasks by generalizing from examples “. (PEDRO DOMINGOS, 2013*).*The machine is not able to self-program, but it could be prepared to generate and store associations and facts from the data. The generalization means the capability to associate in timely fashion based on limited data.The generalization implies some kinds of presumptions which can drive to the repetition of past mistakes(e.g. criminal activity) or effects which were not intended to assume(e.g., unfair discrimination). The interpretation and prediction rules show one of the main problems in artificial intelligence, due to the fact that outcomes could be influenced by the programmers and then the conclusions become arbitrary or disparate, contaminated by the “beliefs, fallibilities, and biases of the person who created them.”(L.BARRET, 2016*)*

One of the main characteristics in machine learning is the so called deep learning. Deep learning uses learning techniques combining layers of neural networks to identify the profiles of a data set that are necessary to make decisions. There are many layers between input and output data and the outputs from the previous layers are inputs for the next (artificial neural network). (JERRY KAPLAN, 2016*).*

Machine learning and deep learning algorithms are the last frontier of artificial intelligence and are used in Web search, spam filters, credit scoring, insurance risk, fraud detection, stock trading, drug design, employment evaluations, health records, hiring searches, housing, and many other

applications. In addition, artificial intelligence agents that have the physical support and interact with the environment a machine “capable of performing tasks by sensing its environment and/or interacting with external sources and adapting its behavior” (*EN ISO 8373)*

The emergence of such robots are occupying sensitive and perhaps dangerous spaces that were previously occupied by human beings. We now have robot killers (military drones), robot cars, health robots, home robots, which all share common features: they may have access through their sensors, to a large and uncontrolled volume of information; they are ordinarily connected to the cloud and become big consumer of personal data and operate on the people behavior.

The best example are the autonomous weapons that would select and engage targets without meaningful human control. This shows the absence of appropriate regulation on artificial intelligence in general and over its agents in particular. There are not international legally binding instruments or even national laws prohibiting the development, production and use of the so called killer robots.

M.R.Calo exposes the raise of robots and privacy concerns. First, in terms of direct surveillance on the persons; second, the access to historically protected spaces-home robots-; and, third, and more dangerous, the social meaning, implicating what he called “ setting privacy”, or, the description of how a person programs and interact with a robot in his or her intimacy.

The robot is the avatar on earth, a player of virtual world in the real world. As we know an avatar is the physical representation of the player in the virtual world, used to co-inhabit simultaneously and interact with others players’ avatars. Now, we have a physical avatar in the real world interacting under human programs with others agents and adapting its behavior. We can say that the robot is an avatar in the real world. The risk and the threat is obvious.The avatar has no limits, moral, feelings, conscience, intentionality, other than the planned use of the algorithms, the time to use them and the purpose planned by humans.

2. The big data: from cognitive to sourveillance. Data is the foundation of artificial intelligence. It is the principal raw material of the algorithm, like cotton, wheat or fuel in the last century.

Data processing is the digital and virtual essence: without data there is no algorithm and without algorithm it is difficult to argue that there is artificial intelligence, digital goods or virtual goods. The value of data lies in its infinite reuse "the data's value is calculated on the basis of all the possible ways in which it could be used in the future and not merely on the basis of its present use"(V.MAYER-SCHONBERGER,K.CUKIER*)*

The recombination of data, its accumulation and its extension are its real value and therefore the impulse for its accumulation by the organizations like *Facebook, Twitter, Amazon, Visa, MasterCard, Bloomberg* and the like.

The initial data is susceptible to be eternal, reiterated and repeated continuously and systematically applied. Should that be personal data, she will lose track of her identity by the deprivation of personal rights.

D.J.Solove highlights privacy, because it is the protector of personal interests, above all, of society, before that of the individual. Privacy is not a fight for personal interest against social interest, but the protection of the individual on the basis of society's own values. "You cannot fight for an individual right against the more important social good. Questions of privacy involve a balance of social interests on both sides of the scale".

S.Rodota focuses its critical concern on the fact that data, by its spread use and reuse suggests that the identity of individuals "is built by others". The problem is not only the construction of profiles but the future use of algorithm, treated with "probability techniques which constructs a hypothetical distorting identity". Further to the reconstruction of new personal identities, these could not be the same as the initial ones.

The algorithm builds an identity through some kind of classification which is the result of programs created by humans that go beyond rationality and become manipulation or monitoring, without any transparence.

The algorithm provides generalization and prediction from the extraction, storage, and analysis of data, creating correlations, almost without evidence. Data mining is the process which takes place to define artificial intelligence as rational behavior. However, the identification of predictive algorithms with rational action cannot be the best solution because they are usually developed by somebody with a clear intention of profit or governance.

The rationale behind the algorithmic decision making is essentially opaque, secret, and few know how it works.

Big Data is the fuel that runs the algorithm success. “Collection and processing of data produces ever more data, which in turn, is used by algorithms to improve themselves. Based on Kant’s famous statement, algorithms without data are empty; data without algorithms are blind.”(J.M.BALKIN, 2015)

Y.Moulier Boutang coined the term *cognitive capitalism* and *S.Zuboff* qualified the process as *surveillance capitalism*. The surveillance, access and control becomes the core of the system, through the, collection, extraction, storage and analysis by big data. It is likely that surveillance provides a better description of what is actually happening. Many scholars use the concept of surveillance as the major threat of artificial intelligence, together with access and control.

Data flows from sources through very well-known ways, providing huge amounts of statistical raw material information on social, economic, financial, consumer, climate, genomic trends and human behavior. First of all, there was e-commerce, Internet transactions and the mere access to Internet. Second, the flows collected through artificial intelligence devices, sensors, agents placed in a wide range of objects and persons. Third, the flows from the banking system, credit rating agencies, credit cards, airlines, health care, insurance, telecommunications companies and government databases on climate,

genomic, or taxation. Fourth, the flows from public cameras on the streets, shops, public and private buildings and not least the smartphones. Finally, the individual attraction for entertainment, consumption, use or mere information of digital goods, books, films, music - or digital and personal needs (*Facebook).*

"Big Data is constituted by capturing small data from individuals computer mediated actions and utterances in their pursuit of effective life" (S.ZUBOFF, 2017*).* This last source it is very important. Personal data is the essential resource for the digital economy and it is "unpaid work" (P.COLLIN-N.COLIN, 2013*).* Personal data is the most valuable currency of the millennium. The unpaid work of users is probably the most important source of profits by large companies providing Internet content.

The free contributions from users to the digital economy illustrate the extraordinary increase in profits. Users without remuneration incorporate originality in the co-creation of the applications of the artificial intelligence software, fueling their exponential returns.

J.Lanier affirm that digital designs do not treat people as special, but as necessary elements in a larger machine of information. However, people are the only source and target of information; the unique meaning of the artificial intelligence. He proposes that the organizations have to pay people for the information collected if that information is found to be valuable. That includes financial technology schemes such as high frequency trading, search engines of social networks, insurance and intelligence agencies.

This is, for now wishful thinking. In fact, the extraction of free data occurs in despite of any kind of reciprocity between the data miners and the miners. There is an enormous distance between the people and their co-creation of intangible capital in terms of sharing collective information technology and communication by the economics agents. There isn't any material reciprocity between them in the production of intangible digital or virtual assets. The logic of accumulation expulse the user, which is the main factor of revenue, from their personal data, characterizing "such assets as "stolen goods" or "contraband" as they were taken, not given…" (S.ZUBOFF, 2017*).*

3. The algorithmic society. Artificial intelligence departs from big data analysis by assessing data, constructing patterns and profiles, making predictions and generalizations and finally by applying the behavior of the person on her decisions. Data, big data and artificial intelligence close a magic circle to make profits from modified human behavior, without any particular effort of transparency, accountability or non-discrimination. Algorithmic decision-making is not neutral. ”Because human beings program predictive algorithms, their biases and values are embedded into the software instructions”. (D.KEATS CITRON,F.PASQUALE, 2014*).*

J.M.Balkin defines the Algorithmic Society as “a society organized around social and economic decision making by algorithms robots, and artificial intelligence agents; who not only make the decisions but, in some cases, also carry them out”. Data is the fuel that fuels the engines, devices and machines used by artificial intelligence. The issue at stake is governance of humans by humans using a particular technology, thus, we cannot segregate Big Data and Internet connection and artificial intelligence from Big Data.

Algorithms are human creations and there is a clear human responsibility for its use to influence other people. The human creator and inventor of the algorithm programs with data collection addressed to perform particular tasks. Programs are not capable of acquiring information and knowledge or to carry out any other purpose. Programs don’t take actions neither do things with our data. The author is a human being and therefore the author is responsible for the errors, any collateral damage and failures.

With the algorithm occurs something similar to financial derivatives before the financial crisis of 2008. During that time, *Bernard Buffet warned* against their potential power of destruction, which was later confirmed. The problem is not the algorithm, but the humans which create and use them as weapon of mass destruction.

The dark side of Big Data is very dangerous and it can damage lives at very critical moments, i.e. when going to college, borrowing money, getting

sentenced to jail or finding and holding a job. All of these life domains are increasingly controlled by secret models wielding arbitrary punishments."(CATHY O`NEIL, 2016*)*

Deregulation is the kingdom of the algorithm. If this situation doesn't evolve in a different direction, there will be a growing instrument of privacy vulnerability and perpetuation of social inequalities. Besides, there is a clear risk of systemic crisis, because the unknown begins to prevail over the known predictions and performances. It seems that we are just on the edge. The digital arrogance applied without care could be the source of troubling dynamics, something similar to environmental pollution. The negative externalities will be superior to the positive ones until the explosion of the systemic crisis.

4. Algorithmic principles. *J.M.Balkin* proposes three principles on artificial intelligence.

A) Information Fiduciaries.

The first principle is that those who use robots, artificial intelligence agents and algorithms have duties of good faith and trust toward their end users and clients. The basic duty of the *information fiduciaries* is that they are not permitted to induce trust in their end users and then turn around and use the information they collect in ways that betray that trust. There is an asymmetry between the Big Data agents and you, because they know a lot about you and you don't know a lot about them.

The principle relies on good faith, non-manipulation, and non-domination as obligations for information fiduciaries (*Google, Facebook, Twitter, and Apple).* Mistrust should not be a priority for artificial intelligence players, on the contrary, it is trust that builds relationships. For this reason, this principle has a broad breath as a collective and social meaning purpose, susceptible of inspiring regulation, sanctions or law.

B) Algorithmic operators have public duties toward general public.

The use of algorithms can harm not only the end user or the client, but many more. The motivation is that the public interference on business is morally justified when it is needed to prevent real or potential damages. Damage to others entails frustration, obstruction or defeat of an interest, all kind of things that improve o worsen our welfare. This principle is connected, as we will see, with the precautionary principle.

C) **The central public duty of algorithmic operators is not to be algorithmic nuisances. They may not externalize the social costs of their use of algorithms onto the general public.**

The principle deals with the negligent use of computational capacities that externalizes costs onto innocent others, or, a nuisance. The concept is related to the externalities produced by the environmental pollution. The companies must adopt methods that are justify from the standpoint of society as a whole.
J.M.Balkin mentions harms to reputation, by classification or risk assessment; discrimination; normalization or regimentation; manipulation; lack of due process, transparency and interpretability.

"We might sum up this discussion by saying that algorithms (a) construct identity and reputation through (b) classification and risk assessment, creating the opportunity for (c) discrimination, normalization and manipulation, without (d) adequate transparency, monitoring, or due process.

The *Balkin* principles are important because they define the algorithmic risks around three main criteria: (i) precautionary duty, (ii) responsibility failure of the economic operators and (iii) disclosure of the secrecy about the means and purposes fulfilled by the artificial intelligence or "arbitrariness by algorithm". Although algorithmic predictions are capable of harming individuals when they have opportunities in life often in arbitrary and discriminatory ways, they remain secret". (D.KEATS CITRON & FRANK PASQUALE, 2014).

The new technology brings progress but also high risks and harm to others. There is an obligation to translate in values, principles and rules the behavior of economic operators when it is morally problematic.

The above is a good point of departure because it excludes some of the more dangerous misconceptions and misunderstandings on the matter, for example, the homunculus fallacy or the android fallacy. Artificial intelligence is human intelligence and therefore it must be subject to protection of personal rights, precautionary clauses and responsibility failures.

5. About the governance common principles. Each community has its own moral values which inform the social dimension and supply criteria for the rules. There are not any rules without a moral dimension. These principles can be described as superior common standards. The importance is that the common principles transpose the human values of fairness and justice and the direction, the general directive or trend, which drive the lawmaker and the law.

-**Privacy protection and civil liberties.**

"Knowledge is power. To scrutinize others while avoiding scrutiny one is one of the most important forms of power. Firms seek out intimate details of potential consumers and employees' lives, but give regulators as little information as they possible can about their own statistics and procedures…The decline in personal privacy might be worthwhile if it were matched by comparable levels of transparency from corporations and governments. But for the most part it is not."(F.PASQUALE, 2015).

The algorithmic decision making increases interference in everyday lives under a logic essentially opaque. The automated predictions, generalizations, patterns, are secret and make impossible challenge the decision, preventing the reaction against probable discrimination on employment, credit scores, insurance risk or, even, in law enforcement, sentencing.

The question is whether it is possible to have the advantages of the new information technology without sacrificing privacy and as *D.J. Solove* points out:"The protection of privacy in the information age is a question of social design".

The balance for now is against individual rights and personal liberties. Damages are extended and simultaneous, in collection, processing, dissemination, and invasion of information. Our identity "is built by others" and it is reused and converted in distortion by appropriation of the new identity.

"I classify as a privacy violation a problem I call *distortion*, which involves disseminating false or misleading information about a person."*(*D.J.SOLOVE, 2008*)* The main issue is the invasion on the decisions that a person can make. Distortion triggers invasion, caused by secret software programs through algorithms predictions. There is no neutral data nor neutral algorithms, sensitive information – gender, age, social status, religion, race, is the foundation for the predictions.

That means that caution and care on data mining and the algorithmic process is needed, because there should be a clear alignment with the person dignity in face of automated decisions. It is not only a problem of data protection, but also of human rights.

-Internalization of externalities.

The *algorithmic nuisance* (J.M.BALKIN, 2017*)* has the same sense of the *pigouvian* externalities.This is a well-known concept in environmental science.There are activities which don't reflect the real social costs of a resource use; so economic agents should have been forced to internalize them.In other cases, the social costs goes beyond the participants, on others,unrelated to the activity.The algorithmic nuisance harms others through the responsibility failure affecting reputation; disparate treatment and impact;generalization;manipulation;opacity.At a minimu, setting an identity which is artificial and different to the true human being identity.

Internalization opens the way to the precautionaire principle:when an activity raises threats of harm to human health, safety, or the environment, precautionairy measures should be taken even if some cause-and effect relationships are not scientifically and fully proved.There is no obligation to wait to the final scientific solution to prevent risks and harms, because that could be too late for the general prevention.There are founded suspicions on the effects of the algorithm society which incite to adopt regulations based in the beliefs or premonitions about its future consequences, without scientific certainty, at the moment. The scientific uncertainty shouldn`t be an excuse to the adoption of measures preventing the possibility of the worse-case scenarios.

Taxation could be a tool to balance the difference between social and private costs induced by the internalization of the agent`s undesiderable behavior.There is here an original idea of a tax on collection,storage,treatment and sale of data due to *Collin-Colin.*

- Bit property. Artificial intelligence ought to be property.

The artificial intelligence "are artifacts: they are product of human labor…From the normative and empirical premises, it would seem to follow that the makers of AIs are entitle to own them."(L.SOLUM, 1992*).*

The first corollary is thay always there is an owner,programmer, controller,who has the property or commands the algorithm.So the responsibility might always be attributable, if that is the issue, to a human being.The liability for harm caused by ALs belong to their owners.

The second corollary is that artificial intelligence artifacts must be qualified as intangibles assets, included the robots, oriented for profits or military and security purposes, as information or ,more, as surveillance assets(*Zuboff).*The definition will secure the fair accountancy and taxation rules of surveillance or information intangibles.

Bit property accentuates the equivalence between property and information assets, that must be organized in broad terms as an open protocol of communication,information and data storage .Property is a procedure rather

than a set of defined attributes.Property is a procedure to communicate information regarding and identity to a resource. The flow of information moves like “waves in a lake, from their place of origin” (J.FAIRFIELD, 2015*).*

The bit property seems accurate to assess if artificial intelligence property is secure and safe, susceptible to human control and aligned with human interests.In other words,it is the foundation to establish a public certification system in a neutral environment, to determine full ,partial or inexistent liability for designers, programmers,manufacturers.Besides, will be possible to establish rules for precertification research and testing of artificial intelligence.”At some point,the legal system will have to decide what to do when those machines cause harm and whether direct regulation would be a desiderable way to reduce such harm.”*(*M.U.SCHERER, 2016).

There is a huge legal challenge for the public,private, commercial,accountancy, employment and taxation ,local and international rules. Artificial intelligence is an **unknown unknown**, forming part of ”the ones we don`t know we don`t know” (E.HEMINGWAY*).*

We need national and international rules to define the level playing field of artificial intelligence between the organizations for profit and public entities who create the algorithms and those who are governed by them. This means an internationally legally binding instrument, through the United Nations *International Telecommunication Union* or, through *soft law*, as Recommendations, Guidelines, Standards, Good Practices and national laws and policies as partial contributions previously to a general regulation. The question it is under which principles and standards.

References

M. MINSKY**,** Semantic Information Processing,M.Minsky ed.,MIT Press,1969.
S.ZUBOFF**,** Big Other:surveillance capitalism and the prospects of an information civilization,Journal of Information Technology, 2015,30.
L.M.LUPOCKI**,** Algorithmic Entities, 95 Washington University Law Review,ssrn.com/abstract=2954173
P.DOMINGOS**,** A Few Useful Things to Know About Machine Learning,Department of Computer Science and Engineering,University of Washington, 2013.

L.BARRET, Deconstructing Data Mining:Protection Privacy and Civil Liberties in Automated Decision-Making,Geo.L.Tech Rev. 1253, 2016.
J.KAPLAN, Artificial Intelligence,Oxford University Press, 2016.
M.R.CALO, Robots and Privacy,2010, ssrn.com/abstract=1599189
V.MAYER-SCHONBERGER,K.CUKIER, Big Data, Garzanti, 2013.
D.J.SOLOVE, "I' ve Got Nothing to Hide" and Other Misunderstandings of Privacy, 2008, ssrn.com/abstract=998565; The Digital Person,2004 New York University.
S.RODOTA**,** Internet ha bisogno di nuove regole, La Reppubblica, 27-11-2014.
J.M.BALKIN**,** The Three Laws of Robotics in the Age of Bid Data, 78 Ohio St. L.J.,2017.
Y.MOULIER BOUTANG, Vers un capitalism cognitive, Paris, 2001, A.Corsani, P.Dieuaide, C.Azais.
P.COLLIN-N.COLIN**,** Mission d' expertise sur la fiscalitè de l' economie numerique,janvier, 2013, Paris.
J.LANIER, Who owns the future?, 2013,New York.
CATHY O`NEIL*,* Weapons of Math Destruction:How Big Data Increases Inequality and Threatens Democracy, Crown, 2016.
D.KEATS CITRON & FRANK PASQUALE, The Scored Society:Due Process for Automated Predictions,Washington Law Review, 2014,89:1
F.PASQUALE, The Black Box Society,The Secret Algorithms That Control Money and Information,Harvard University Press, 2015.
L.SOLUM, Legal Personhood for Artificial Intelligence, North Carolina Law Review, 1992.
J.FAIRFIELD, "BitProperty",Washington and Lee University School of Law,2014-17,October 2014,California Law Review, 2015.
M.SCHERER**,** Regulating Artificial Intelligence Systems:Risks,Challenges,Competencies,and Strategies,Harvard Journal of Law & Technology,29,2, 2016.