GOVERNANCE OF INTERNET OF THINGS AND ETHICS OF ARTIFICIAL INTELLIGENCE

A GOVERNANÇA DA INTERNET DAS COISAS E A ÉTICA DA INTELIGÊNCIA ARTIFICIAL

Eduardo Magrani[I]

[I] Fundação Getúlio Vargas (FGV), Rio de Janeiro, RJ, Brasil. Doutor em Direito.
E-mail: eduardomagrani@gmail.com

Resumo: A interação contínua entre dispositivos inteligentes, sensores e pessoas aponta para o crescente número de dados produzidos, armazenados e processados, alterando, em vários aspectos e cada vez mais, nosso cotidiano. Por um lado, o contexto de hiperconectividade pode trazer benefícios econômicos ao Estado, às empresas, além de conveniência aos consumidores. Por outro lado, o aumento da conectividade traz desafios significativos nas esferas da proteção da privacidade e da ética contemporânea, impactando, em última análise, a própria democracia. Esta tese aborda, do ponto de vista regulatório, alguns dos desafios enfrentados pelo atual estado de direito decorrente do avanço do cenário denominado Internet das Coisas.

Palavras-chave: Governança. Internet das coisas. Ética. Hiperconectividade. Inteligência artificial.

Abstract: The continuous interaction between intelligent devices, sensors and people points to the increasing number of data being produced, stored and processed, changing, in various aspects and increasingly, our daily life. On one hand, the context of hyperconnectivity can bring economic benefits to the State, companies, as well as convenience to consumers. On the other hand, increasing connectivity brings significant challenges in the spheres of privacy protection and contemporary ethics, impacting, ultimately, democracy itself. This thesis addresses, from the regulatory point of view, some of these challenges faced by the current rule of law arising from the advance of the scenario called Internet of Things.

Keywords: Governance. Internet of things. Ethics. Hyperconnectivity. Artificial intelligence.

From the 1980s, with the progressive development of computers in business and public administration, there was a perception that governmental and corporate practices in processing personal data were reducing individuals to mere data, threatening their fundamental rights and their freedom (EUROPEAN DATA PROTECTION SUPERVISOR, 2015).

On an even larger scale, this is the thesis reinforced by the Israeli writer Yuval Noah Harari (HARARI, 2015)[1] when dealing with the loss

[1] Harari argues in his work "Homo Deus" that we are moving towards a post-

of human freedom and what he calls the new religion of data (HARARI, 2016):

> Humanist thinkers such as Rousseau convinced us that our own feelings and desires were the ultimate source of meaning, and that our free will was, therefore, the highest authority of all. Now, a fresh shift is taking place. Just as divine authority was legitimized by religious mythologies, and human authority was legitimized by humanist ideologies, so high-tech gurus and Silicon Valley prophets are creating a new universal narrative that legitimizes the authority of algorithms and Big Data. This novel creed may be called "Dataism". In its extreme form, proponents of the Dataist worldview perceive the entire universe as a flow of data, see organisms as little more than biochemical algorithms and believe that humanity's cosmic vocation is to create an all-encompassing data-processing system — and then merge into it. We are already becoming tiny chips inside a giant system that nobody really understands. Every day I absorb countless data bits through emails, phone calls and articles; process the data; and transmit back new bits through more emails, phone calls and articles. I don't really know where I fit into the great scheme of things, and how my bits of data connect with the bits produced by billions of other humans and computers. I don't have time to find out, because I am too busy answering emails. This relentless dataflow sparks new inventions and disruptions that nobody plans, controls or comprehends.
>
> […]
>
> Even though humanists were wrong to think that our feelings reflected some mysterious "free will", up until now humanism still made very good practical sense. For although there was nothing magical about our feelings, they were nevertheless the best method in the universe for making decisions — and no outside system could hope to understand my feelings better than me. […]. This is just the beginning. Devices such as Amazon's Kindle are able constantly to collect data on their users while they are reading books. Your Kindle can monitor which parts of a book you read quickly, and which slowly; on which page you took a break, and on which sentence you abandoned the book, never to pick it up again. If Kindle was to be upgraded

anthropocentric world where the value of reality is extracted from constant information processing by human and nonhuman agents. In a similar sense, Luciano Floridi argues that: "ICTs are bringing about a fourth revolution, in the long process of reassessment of humanity's fundamental nature and role in the universe. We are not immobile, at the centre of the universe (Copernican revolution); we are not unnaturally distinct and different from the rest of the animal world (Darwinian revolution); and we are far from being entirely transparent to ourselves (Freudian revolution). ICTs are now making us realize that we are not disconnected agents, but informational organisms (inforgs), who share with other kinds of agents a global environment, ultimately made of information, the infosphere (Turing revolution)".

> with face recognition software and biometric sensors, it would know how each sentence influenced your heart rate and blood pressure. It would know what made you laugh, what made you sad, what made you angry. Soon, books will read you while you are reading them. And whereas you quickly forget most of what you read, computer programs need never forget. Such data should eventually enable Amazon to choose books for you with uncanny precision. It will also allow Amazon to know exactly who you are, and how to press your emotional buttons.

With the growing dissemination of Big Data and computing techniques, technological evolution and economic pressure spread rapidly and algorithms have become a great resource for innovation and business models. This rapid diffusion of algorithms and their increasing influence, however, have consequences for the market and for society, consequences which include questions of ethics and governance (SAURWEIN, 2015, p. 35-49).

Given that algorithms can permeate countless branches of our lives as they become more sophisticated, useful, and autonomous, there is a risk that they will make important decisions, replacing human beings (DIAKOPOULOS, 2015, p.398)[2]. Accordingly, Danilo Doneda and Virgilio Almeida argue that to foment the integration of algorithms into social and economic processes, algorithms governance tools are needed (DONEDA, 2016, p. 60). The governance of algorithms (SAURWEIN; JUST; LATZER, 2015, p. 38-43)[3] can vary from the strictly legal and regulatory point of view, to the purely technical point of view (SAURWEIN, JUST, LATZER, 2015, p. 62).

Among the regulation points are transparency, responsibility - which is linked to notions of justice and due process - and technical guarantees, as well as the development of ethical principles regarding the use of personal data (Big Data Ethics). It should be noted that algorithms are constantly working and facing unplanned and unprecedented situations frequently, so that their monitoring must be constant (SAURWEIN; JUST; LATZER, 2015).

[2] As stated by Nicholas Diakopoulos, "We are now living in a world where algorithms, and the data that feed them, adjudicate a large array of decisions in our lives: not just search engines and personalized online news systems, but educational evaluations, the operation of markets and political campaigns, the design of urban public spaces, and even how social services like welfare and public safety are managed."

[3] Among the governance options, which have their limitations and are influenced by contextual factors such as incentives and conflicts of interest, we have the following: (i) self-organization of individual companies; (ii) collective self-regulation; (iii) co-regulation and (iv) state intervention.

One of the main themes raised by doctrine when it comes to governance is the opacity of the algorithms. The problem of opacity is related to the difficulty of decoding the result generated by the algorithm. The human inability to decode the result of algorithms can create problems when they are used to make important decisions that affect our lives. Thus, there has been talk of the need for greater transparency, which could be achieved by regulating (DONEDA; ALMEIDA, 2016, p. 60-62).

According to the Report prepared by renowned names in the field of Digital Ethics, such as Luciano Floridi and Wendell Wallack (WALLACK, 2017):

> A lack of scrutability or meaningful transparency can undermine the acceptability of deploying systems in situations where harm may occur to people, animals, the environment or institutions. Should the system fail and cause harm, it becomes critical to have a forensic capability to ensure similar accidents or failures do not occur, and to determine accountability and liability. This is especially important when outcomes are unexpected and/or not aligned with the original intent for which the system was deployed.

On this issue, researchers at the Oxford Internet Institute and Alan Turing Institute deepen the debate (MITTELSTADT, 2016):

> The primary components of transparency are accessibility and comprehensibility of information. Information about the functionality of algorithms is often intentionally poorly accessible. Proprietary algorithms are kept secret for the sake of competitive advantage, national security, or privacy. Transparency can thus run counter to other ethical ideals, in particular the privacy of data subjects and autonomy of organizations.
>
> […]
>
> The commercial viability of data processors in many industries may be threatened by transparency. However, data subjects retain an interest in understanding how information about them is created and influences decisions taken in data driven practices. This struggle is marked by information asymmetry and an ‘‘imbalance in knowledge and decision-making power'' favouring data processors. Besides being accessible, information must be comprehensible to be considered transparent. Efforts to make algorithms transparent face a significant challenge to render complex decision-making processes both accessible and comprehensible. The longstanding problem of interpretability in machine learning algorithms indicates the challenge of opacity in algorithms. […] Transparency disclosures by data processors

> and controllers may prove crucial in the future to maintain a trusting relationship with data subjects.

On the need for greater transparency, it is worth mentioning that New York City has recently unanimously approved a bill aimed at government agencies that use algorithms to aid in legal proceedings. Seen as an algorithmic accountability bill, the first of its kind in North American legislative regulation, the standard will establish a task force to study how algorithms are used by city agencies to make decisions that affect New York citizens. The task force will concentrate its efforts on investigating algorithmic bias and whether any of the models discriminate against persons based on age, race, religion, gender, sexual orientation, or citizenship status (BERNARD, 2017).

As an example, the new North American regulation aims at providing a new interpretation on algorithms, avoiding its treatment as "black boxes". It is important to emphasize that companies and governmental organizations should try to reduce the algorithmic bias and provide as much transparency as possible to the predictive models.

In Europe, the General Data Protection Regulation (GDPR) establishes the right to an explanation for any decision made by an algorithm and the right to opt for non-collection of data. Many suggest that this pattern is too broad and will have to be revised. However, it is working as a tool to hold stakeholders accountable. In addition, it has motivated engineers to explore ways to provide a greater degree of transparency on how *machine learning* algorithms make their decisions.

In addition, Google's chief AI scientist, John Giannandrea, highlights the risks of inscrutable systems (KNIGHT, 2017):

> It's important that we be transparent about the training data that we are using, and are looking for hidden biases in it, otherwise we are building biased systems [...] if someone is trying to sell you a black box system for medical decision support, and you don't know how it works or what data was used to train it, then I wouldn't trust it.

Researchers at the University of Zurich (SAURWEIN, 2015, p. 37) argue that algorithm governance must be based on identified threats and suggest a *risk-based approach*, highlighting those related to manipulation, bias, censorship, social discrimination, privacy breaches, property rights and abuse of market power (SAURWEIN, 2015). To prevent these risks from materializing, it is necessary to resort to governance.

Besides the algorithm technology itself, other external factors influence its development and the need for its regulation. This is the case with databases. Algorithms become more useful as more data becomes available (DONEDA, ALMEIDA, 2016, p. 61). If data is fundamental to algorithms, which are inert until paired with databases (GILLESPIE, 2014, P. 169)[4], one must analyze the legal treatment given to them, as they must be legitimate, correct, up-to-date and not based on illegitimate bias.

Considering the above, algorithms governance's techniques do not act directly on the algorithms themselves, but on the data that feed the algorithms. As noted by Danilo Doneda and Virgilio Almeida (DONEDA, ALMEIDA, 2016):

> This is true for several tools already present in data protection legislation that, in some countries, have measures regarding transparency and fairness that apply directly to algorithms and the platforms that support their functioning. For instance, the provision that automated decisions shall be grounded on transparent criteria is commonly present in several pieces of data-protection legislation. The same happens with the right to ask for a human revision of automatically taken decisions.

After analyzing options and limitations on governance, researchers at the University of Zurich concluded that there is no one-size-fits-all solution, but there should be a mix between governance and a respect for each actor involved:

> Analyses reveal that there is a broad spectrum of players, levels and instruments for the governance of algorithms, but there is no one-size-fits-all solution. Instead, there is the need for a governance mix consistent with the respective risks and applications in question and an interplay between instruments and diverse actors involved. The attention therefore has to shift to multi-dimensional solutions and combinations of governance measures that mutually enable and complement each other. (…) The search for an adequate governance mix is difficult because there is only limited knowledge about the development and the effects of regulatory interventions. The existing uncertainties call for further risk and technology assessment to strengthen the foundations for evidence-based governance in the domain of algorithmic selection. Risk-based approaches seem to be

[4] The author notes that algorithm analysis must always be linked to data analysis: "Algorithms are inert, meaningless machines until paired with databases on which to function. A sociological inquiry into an algorithm must always grapple with the databases to which it is wedded; failing to do so would be akin to what was said at a public protest, while failing to notice that some speakers had been stopped at the park gates".

> particularly appropriate for this purpose. They can monitor market and technology developments, assess the involved and emerging risks and develop problem-oriented, adaptive governance strategies. (SAURWEIN, 2015, p. 44).

However, Lucas Introna (INTRONA, 2016, p. 17-49), a professor at the University of Lancaster, believes that the best solution is not governance but governmentality. For the author, governance practices themselves should be governed because they are never completely safe as such. Governmentality, viewed as a meta-governance, would then consider the performative nature of governance practices (and their outcomes) and show the mutual constitutive nature of problems, domains of knowledge, and subjectivities commanded by governance practices. Calculative practices would relate to government technologies. They are domains of knowledge and expertise. Such practices contain some moral authority, as they impose neutrality and objectivity on a domain that has moral relevance (the author exemplifies with an algorithm designed to identify plagiarism). Based on this, he concludes:

> Thus, understanding governing practices in the idiom of governmentality allows us to see how problems, technologies of governance, regimes of knowledge, and subjectivities become mutually constitutive of each other to create a regime of government that has no specific essence (location or unified action). All the performative outcomes are "never simply a realization of a programme, strategy or intention: whilst the will to govern traverses them, they are not simply realizations of any simple will". (INTRONA, 2016, p. 39).

Nicholas Diakopoulos has another approach in dealing with algorithm related responsibility. He argues that the crucial point is autonomous decision making, since decisions made by algorithms can be based on heuristics[5]:

> Algorithmic decisions can be based on heuristics and rules, or calculations over massive amounts of data. Rules may be articulated directly by programmers or be dynamic and flexible based on machine learning of data. Sometimes a human operator maintains agency and makes the final decision in a process, but even in this case the algorithm biases the operator's attention toward a subset of information (DIAKOPOULOS, 2015, p. 400).

Within this logic of results based on bias, it is also important to highlight that the algorithms are programmed to classify the data sent to

[5] Heuristics is a method or process created with the goal of finding solutions to a problem.

them and, often, mistakes can be made, and there may be false positives and false negatives. As Diakopoulos exemplifies (DIAKOPOULOS, 2015, P. 401), YouTube rates the videos submitted to the site according to the songs that are played to check for copyright infringement. A false positive, in this case, would be a video classified as infringing, but which, in fact, fits into a fair use hypothesis. A false negative, in turn, would be a video classified as fair use, but that, in practice, violated copyright.

Considering that algorithms exert power by themselves but are always influenced by human beings who created them, Diakopoulos states that liability must consider the intention of the creators of the algorithm, the process that influenced its design, and the agency that interprets the algorithms that generated results (DIAKOPOULOS, 2015, P. 398)[6]. Relevant to this discussion, the teachings of Nick Bostrom, a philosopher at the University of Oxford, and Eliezer Yudkowsky, co-founder of the Machine Intelligence Research Institute:

> Another important social criterion for transactions in organizations is to be able to find the person responsible for getting something done. When an AI system fails in its assigned tasks, who is to blame? The programmers? The end users? Modern bureaucrats often take refuge in established procedures that widely distribute liability, so that a person cannot be identified or blamed for the outcome of catastrophes (HOWARD, 1994). The likely disinterested trial of a specialist system could turn out to be an even better refuge. Even if an AI system is designed with a user replacement, it is a must to consider the career incentive of a bureaucrat who will be personally blamed if the replacement goes wrong, and who would much rather blame the AI for any difficult decision with a negative result. (BOSTROM; YUDKOWSKY , 2011, p. 202-203).

This leads us to further discuss the moral and legal liability of these non-human agents (KNIGHT, 2017)[7]. To think about regulation,

[6] "Algorithmic accountability must therefore consider algorithms as objects of human creation and take into account intent, including that of any group or institutional processes that may have influenced their design, as well as the agency of human actors in interpreting the output of algorithms in the course of making higher level decisions."

[7] This discussion complements the governance review and touches upon the areas of *Machine & Information Ethics & Philosophy of Technology*. Another possible nomenclature for these issues is Digital Ethics: "Digital ethics is the branch of ethics that studies and evaluates moral problems related to data, algorithms and corresponding practices. Its goal is to formulate and support morally good solutions (e.g. right conducts or right values) by developing three lines of research: the ethics of data, the ethics of algorithms and the ethics of practices. The ethics of data looks at the

it is crucial to go beyond the mere recognition of the agency power of Things and seek a careful analysis of the differences between technical artifacts and sociotechnical systems. This differentiation is justified by the level of complexity and potential of influence of each one, leading to different regulations.

Peter-Paul Verbeek in *Moralizing Technology: Understanding and Designing the Morality of Things* aims to broaden the scope of ethics to better accommodate the technological age, and in doing so, reveals the inseparable nature of humanity and technology. For Verbeek, technologies are "moral mediators" that shape the way we perceive and interact with the world and thus reveal and guide possible behaviors. In Verbeek's words: "No technology is morally neutral, since every technology always affects the way in which we perceive and interact with the world, and even the ways in which we think – it mediates our lives." (VERBEEK, 2011).

Referring to Bruno Latour's theory, Verbeek concludes that humanistic ethics necessarily divide the world into two domains: the human on one side and the other (or the "non-human"), where human beings are subjects and non-humans are objects of human activity. Due to this approach, it becomes almost impossible to attribute any moral significance to technology. For this reason, our starting point is the scenario presented by Latour, according to which artifacts are also social actors. According to Latour (LATOUR, 2001, p. 245):

> To conceive humanity and technology as opposing poles is, in effect, to discard mankind: we are sociotechnical animals and every human interaction is sociotechnical. We are never limited to social bonds. We never confront ourselves with objects alone. […] The illusion of modernity was to believe that the more we grow, the more objectivity and subjectivity grow apart, thus creating a radically different future from our past. After the paradigm shift in our conception of science and technology, we now know that this will never happen and, in fact, never happened. […] [The artifacts] deserve to be housed in our intellectual culture as full social actors. The artifacts are us. The aim of our philosophy, social theory and morality is to invent political institutions capable of absorbing this great history, this vast spiraling movement, this labyrinth, this fate.

generation, recording, curation, processing, dissemination, sharing and use of data. It is concerned with moral problems posed by the collection, analysis and application of large data sets. Issues range from the use of big data in biomedical research and the social sciences to profiling, advertising and data donation and data philanthropy, as well as open data in government projects".

From Latour's teachings, it is possible to deduce that artifacts are endowed with agency and have the capacity to interfere and, for this reason, they must be considered full-fledged social actors. As with sociotechnical systems, on an even larger scale, justified by its greater complexity, as will be further explored in this article.

Although Latour's theory explores the role that each actor possesses, placing all of them on the same level of agency, it is important to consider that it is not all technical artifacts or sociotechnical systems that have the same capacity for influence in the interactions that occur between humans and non-humans. For example, the influence that a physical door holds in relation to an individual is considerably different from the influence generated by a *Thing* endowed with artificial intelligence and algorithms with *deep learning* technique. (WEB FOUNDATION, 2017)[8]

The technical artifacts, as explained by the Dutch theorist Peter Kroes, can be understood as man-made *Things (objects)*, which have a *function* and a *plan of use* (KROES, 2011, pp. 1-2 e 5-7). They consist of products obtained through technological action, designating the attitudes we take daily to solve practical problems, including those related to our desires and our needs (KROES, 2011, pp. 1-2 e 5-7). Technical artifacts involve the need for rules of use to be observed, as well as for parameters to be created in relation to the roles of individuals and social institutions in relation to them and their use (KROES, 2011).

As explained by Kroes (KROES, 2011):

> Practical problems are not just resolved by introducing a bunch of technical artifacts into the world. With these artifacts come instructions for their use. And with these technical artifacts come also social roles for people and social institutions for enabling the use of the artifacts.
>
> […]
>
> There is a huge variety of technical artifacts from very small to very big, from simple to complex, from component part to end-product and consisting of chemical materials, et cetera. What all of these things have in common is that they are material objects that have been deliberately produced by humans in order to fulfil some kind of practical function. They are often described

[8] "Deep learning is a subset of machine learning in which the tasks are broken down and distributed onto machine learning algorithms that are organized in consecutive layers. Each layer builds up on the output from the previous layer. Together the layers constitute an artificial neural network that mimics the distributed approach to problem-solving carried out by neurons in a human brain."

> as technical artifacts in order to emphasize that they are not naturally occurring objects.
>
> […]
>
> We may therefore define a technical artifact as a physical object with a technical function and use plan designed and made by human beings. The requirement that the object must be designed and made by humans is added to ensure that any natural objects that happen to be used for practical purposes are not also termed technical artifacts. Those differences relate especially to the status of having a function and a use plan, and to the accompanying possibility of making normative assertions.

Technical artifacts, therefore, are specific objects (Things) with their own characteristics. Works of art, for instance, are not synonymous with technical artifacts. While these are man-made with clear practical objectives, those have no concrete utility and the skills required for their production are different from those required by engineers. Natural objects are also not to be confused with technical artifacts, since they are given by nature and do not have, in themselves, a practical function. However, natural objects can be transformed into technical artifacts if they go through a process of man-made transformation. For example, the wood of the trunk of a tree is something of nature but becomes a technical artifact when it is transformed by man into a closet and gains a concrete function.

Finally, technical artifacts are distinguished by two main points from the mere physical objects, and from the biological objects. First, they have a clear function and usage plan. Second, they are subject to an evaluation analysis as to whether they are good or bad and whether they work or not. (KROES, 2011, p. 7-13). Thus, it is possible to observe the great importance that the *function* and the *plan of use* have in the characterization of a technical artifact. These two characteristics are intimately connected with the goals that the individuals who created the object seek with it, so that they do not stray from the intended purposes.

Faced with this inseparability, the questioning of the morality of human objectives and actions extends to the morality of technical artifacts (KROES, 2011). Technology can be used to change the world around us and individuals' have goals – be them private and / or social – that can be achieved with the help of these technical artifacts. Considering that the objectives sought by the humans when creating a technical artifact are not separated from the characteristics of the object itself, we can conclude that the technical artifacts have an intrinsically moral character.[9]

This is an important point to which attention should be given, since the debate about liability for consequences arising from the action of human created objects is still a controversial point: who would be held liable? The human or the object? We will return to this discussion later, soon after conceptualizing sociotechnical systems.

Therefore, alongside the technical artifacts, which can represent the simplest objects, with little capacity for interaction/influence, to the more technologically complex ones, we have the sociotechnical systems, which consist of a Network (embedding itself in the Latour's concept reflected in actor-network theory) that connects humans and Things, thus possessing greater capacity for interaction and unpredictability.

For a regulatory analysis, this concept is even more fundamental (KROES, 2011). Precisely because of its complexity embodied in a conglomerate of actants, causing socio-technical systems to have even less predictable consequences than those generated by technical artifacts. In addition, they generate a greater difficulty to prevent unintended consequences, and to hold agents liable in case of harm, since the "technological action", reflected in the sociotechnical system, is a sum of actions of actants' entanglement in the Network in an intra-relation.

To illustrate the difference between the concepts of technical artifact and sociotechnical system, we can think of the former being represented by an airplane, and the second by the complex aviation system. The sociotechnical system is formed by the set of interrelated agents (human and non-human actants - Things, institutions, etc.) that work together to achieve a given goal. The materiality and effects of a sociotechnical system depend on the sum of the agency of each actant. However, there are parameters of how the system should be used, which means that these systems have pre-defined operational processes and can be affected by regulatory laws and policies.

Thus, when a tragic accident involving an airplane occurs, it is necessary to analyze what was in the sphere of control and influence of each actor and technical artifact components of the Sociotechnical Network, but quite possibly we will observe a very complex and

[9] There is a rich debate among scholars about whether certain elements such as conscience, free will, spontaneity, creativity and the role of reason constitute a necessary condition for the recognition of a moral agent (like the human agent).

symbiotic relationship between the components that led to this fateful result (SARAIVA, 2011). Moreover, this result is often unpredictable, due to the autonomy of the system based on an agency diffused and distributed among all components (actants) (LO, 2013).[10]

According to M. C. Elish, from Columbia University (ELISH, 2016):

> It is common to see an airline representative at the gate of a canceled flight be yelled at by frustrated travelers, even though he neither caused the cancelation nor possesses the power to change it. On the front lines of large, bureaucratic systems, people positioned as the external interface of a system appear at once a metonym for the company and also as gatekeepers to the company. As gatekeepers, they seem to possess a degree of agency, a capacity to take effective action, which the customer does not. But in general, we know that such individuals do not represent the whole company, and that agency is only perceived, not actuated. We know, in most cases, these individuals are not responsible for the decisions that have led up to the situation. In instances like these, humans at the interface between customer and company are like sponges, soaking up the excess of emotions that flood the interaction but cannot be absorbed by faceless bureaucracy or an inanimate object. There may be affective ramifications for this misplaced blame, but the discerning customer or manager will know that the individual is not responsible. However, in automated or robotic systems it can be difficult to accurately locate who is responsible when agency is distributed in a system and control over an action is mediated through time and space. When humans and machines work together, who or what is in control? As control has become distributed across multiple actors (human and nonhuman), our social and legal conceptions of responsibility have remained generally about an individual. We developed the term moral crumple zone to describe the result of this ambiguity within systems of distributed control, particularly automated and autonomous systems. Just as the crumple zone in a car is designed to absorb the force of impact in a crash, the human in a highly complex and automated system may become simply a component – accidentally or intentionally – that bears the brunt of the moral and legal responsibilities when the overall system malfunctions.

[10] Academics and companies have researched the vulnerabilities of the aviation industry by trying to propose a solution that involves clustered aircrafts, where one group of aircrafts share information and validate each other's transmissions, so that they form a group of "reliable aircrafts". This way, any false signal would be rejected by the group. These solutions are trying to add more layers of protection over existing technology, minimizing risks such as the hacking of an aircraft. These solutions are based on a perspective that sees not only airplanes as isolated technical artifacts but considering solutions that involve the sociotechnical system more broadly.

Considering these complex systems, the debate on liability and ethics – already raised when presenting technical artifacts - returns. Issues such as the liability of developers and the existence of morality in non-human actors – with a focus here on technological objects - need a response or, at least, reflections that contribute to the debate in the public sphere[11].

Bruno Latour's theory offers progress in confronting and discarding the formal binary division between humans and non-humans, but it places objects with complexities and objects with different values at the same level. Given this context, from a legal and regulatory point of view, assigning a different status to technical artifacts and sociotechnical systems, according to their capacity for agency and influence is justifiable and should be endowed with different moral status and level of liability. It is necessary, then, to distinguish the influence and importance that each Thing has in the Network and, above all, in the public sphere and, from there, to think about what can be done in the IoT scenario, from an ethical-regulatory point of view (LATOUR, 2001, P. 245).

For this analysis, we will focus on advanced algorithms with *machine learning,* and on robots equipped with artificial intelligence, considering that they are technical artifacts (Things) attached to sociotechnical systems with a greater potential for autonomy (based largely on the processing of Big Data) and unpredictability.

While technical artifacts, such as a chair or a glass, are artifacts "domesticated" by humans, i.e. more predictable in terms of their influence and agency power, it is possible to affirm that intelligent algorithms and robots are still non-domesticated technologies, since the time of interaction with man throughout history has not yet allowed us to foresee most of the risks in order to control them, or to cease them altogether. This clipping will allow us to work on the ethical theme of Things, in its most complex aspect.

Colin Allen e Wendell Wallach (WALLACH, ALLEN, 2008) argue that as intelligent Things, like robots,[12] become more autonomous

[11] In its Habermas definition.

[12] The 2005 UN Robotics Report defines a robot as a semi or fully autonomous reprogrammable machine used for the well-being of human beings in manufacturing operations or services.

and assume more responsibility, they must be programmed with moral decision-making skills for our own safety.[13]

Corroborating this thesis, Peter-Paul Verbeek (VERBEEK, 2001), while dealing with the morality of Things understands that: as machines (VLADECK, 2017) now operate more frequently in open social environments, such as connected public spheres, it becomes increasingly important to design a type of functional morality that is sensitive to ethically relevant characteristics and applicable to intended situations. With respect to which type of ethics to implement, it is defended that it must be a deontological matrix type, constructed within the deliberative procedural parameters defended by Habermas, but also considering the agency power of Things in a new-materialist perspective.

A good example is the Microsoft's robot Tay (REVISTA GALILEU, 2017), that helps to illustrate the effects that a non-human element can have on society. In 2016, Microsoft launched an Artificial Intelligence program named Tay. Endowed with a *deep learning* ability, the robot shaped its world view based on online interactions with other people and producing authentic expressions based on them. The experience, however, proved to be disastrous and the company had to deactivate the tool in less than 24 hours due to the production of worrying results (TECHNOBLOG, 2017).

The goal was to get Tay to interact with human users on Twitter, learning human patterns of conversation. It turns out that in less than a day, the *chatbot* was generating utterly inappropriate comments, including racist, sexist and anti-semitic publications. In 2015, a similar case occurred with "Google Photos". This was a program that also learned from users, to tag photos. However, their results were also unpleasant, and it was noticed, for example, that the bot was labeling colored people as gorillas (TECMUNDO, 2017).

The implementation of programs capable of "learning" to perform functions that relate to people creates new ethical and regulatory challenges, since it increases the possibility of obtaining results other than those intended or even totally unexpected. In addition, these results can cause harm to other actors, such as the discriminatory offenses generated by Tay and Google Photos.

[13] Researchers at Delft University of Technology and Eindhoven University of Technology say that the values that should be taken into account in the development of these technologies are: health, safety, sustainability and privacy.

Particularly, the use of artificial intelligence tools that interact through social media requires reflection on the ethical requirements that must accompany the development of this type of technology. This is because, as previously argued, these mechanisms also act as agents in society, and end up influencing the environment around them, even though they are non-human elements. It is not, therefore, a matter of thinking only about the "use" and "repair" of new technologies, but mainly about the proper ethical orientation for their development (WOLF, 2017).

Microsoft argued that Tay's malfunctioning was the result of an attack by users who exploited a vulnerability in their program. However, for Wolf et al (WOLF, 2017), this does not exempt them from the responsibility of considering the occurrence of possible harmful consequences with the use of this type of software. For the authors, the fact that the creators did not expect this outcome is part of the very unpredictable nature of this type of system.

The attempt to make Artificial Intelligence systems increasingly adaptable and capable of acting in a human-like manner, makes them present less predictable behaviors. Thus, they begin to act not only as tools that perform pre-established functions in the various fields in which they are employed, but also to develop a proper way of acting. They impact the world in a way that is less determinable or controllable by human agents. It is worth emphasizing that algorithms can adjust to give rise to new algorithms and new ways to accomplish their tasks (DOMINGOS, 2015)[14], so that the way the result was achieved would be difficult to explain even to the programmers who created the algorithm (DONEDA; ALMEIDA, 2016, p. 60).

Also, the more adaptable the artificial intelligence programs become, the more unpredictable are their actions, bringing new risks. This makes it necessary for developers of this type of program to be more aware of the ethical responsibilities involved in this activity. The Code of Ethics of the Association for Computing Machinery (WOLF, 2017) indicates that professionals in the field should develop

[14] Pedro Domingos, on learning algorithms: "Every algorithm has an input and an output: the data goes into the computer, the algorithm does what it will with it, and out comes the result. Machine learning turns this around: in goes the data and the desired result and out comes the algorithm that turns one into the other. Learning algorithms—also known as learners—are algorithms that make other algorithms. With machine learning, computers write their own programs, so we don't have to".

"comprehensive and thorough assessments of computer systems and their impacts, including the analysis of possible risks".

In addition, there is a need for dedicated monitoring to verify the actions taken by such a program, especially in the early stages of its implementation. In the Tay case, developers should have monitored the behavior of the bot intensely within the first 24 hours of its launch, which is not known to have occurred (WOLF, 2017).

Furthermore, there is no way to determine with certainty what motivated Microsoft to remove the program: whether the production of offensive comments or the negative response received from users regarding the program. Nevertheless, the logic should be to prevent possible damages and to monitor in advance, rather than the remediation of losses, especially when they may be unforeseeable. To limit the possibilities of negative consequences, software developers must recognize those potentially dangerous and unpredictable programs and restrict their possibilities of interaction with the public until it is intensively tested in a controlled environment. After this stage, consumers should be informed about the vulnerabilities of a program that is essentially unpredictable, and the possible consequences of unexpected behavior (WOLF, 2017).

Another case[15] involving artificial intelligence occurred in November 2016, when Google Translator developed its own language unintelligible to humans. A few months earlier, Google had installed the Google Neural Machine Translation system, which would learn to translate based on examples and achieve great accuracy in its task. The mechanism would have been programmed to translate certain languages for English. However, the system managed to translate languages directly with no interference, which means that the artificial intelligence system would have developed its own language, an *interlanguage* (SUMARES, 2016).

[15] Another interesting case that helps us think about the autonomy of these agents is the creation and adoption of Tinder. It is estimated that it amasses a total of 50 million users. This platform intermediates encounters between different people who seek to connect with one another. Therefore, even relationships considered more intimate are engendered and made possible using applications like Tinder. In fact, the algorithm that supports the program is responsible for "deciding" who will appear to whom, based on criteria unknown to the users. Thus, interactions among registered people are established and influenced using the program code and algorithmic filtering, so that this - non-human - element influences these relationships.

A similar situation occurred recently with artificial intelligence developed by Facebook. The system would have been created so that Bob and Alice - names given to *bots* created by researchers - simulated negotiations in English to help researchers understand more constructive forms of negotiation. Nevertheless, Bob and Alice understood each other better by using unintelligible sentences for humans, and so they reached agreements faster. The system was deactivated and did not generate positive results for the research (ROBERTSON, 2017).

As can be seen from these examples - which tend to multiply -, the use of technology, with an emphasis on artificial intelligence, can cause unpredictable and uncontrollable consequences, so that often the only solution is to deactivate the system. Therefore, the increase in autonomy and complexity of the technical artifacts is evident, given that they are endowed with an increased agency, and are capable of influencing others but also of being influenced in the network in a significant way, often composing even more autonomous and unpredictable sociotechnical systems.

Although there is no artificial intelligence system that is completely autonomous, with the development of technology, it is possible to create machines that will have the ability to make decisions in an increasingly independent way, which raises questions about who would be responsible for the result of its actions and for eventual damages caused to others (VLADECK, 2014, p. 120-121). (CERKA, 2015, p. 376-389) According to the Report released at the World Economic Forum in 2017 (WORLD ECONOMIC FORUM, 2018): *The greatest threat to humanity lies in delegating authority and decisions to machines that do not have the intelligence to make*.

The ability to amass experiences and learn from massive data processing, coupled with the ability to act independently and make choices autonomously can be considered preconditions for legal liability. However, since artificial intelligence is not recognized today as a subject of law, it cannot be held individually liable for the potential damage it may cause (CERKA, 2015). In this sense, according to Article 12 of the United Nations Convention on the Use of Electronic Communications in International Contracts, a person (natural or an entity) on behalf of whom a program was created must, ultimately, be liable for any action generated by the machine. This reasoning is based on the notion that a tool has no will of its own (CERKA, 2015).

On the other hand, in the case of damage caused by acts of an artificial intelligence, another type of responsibility is the one that makes an analogy with the responsibility attributed to the parents by the actions of their children (*strict vicarious liability)*. Thus, adopting the theory of "robots as tools", the responsibility for the acts of an AI could fall on its producer, users or their programmers, responsible for their "training" (SOMBRA; REGOTO, 2018)[16]. (CERKA, 2015, p. 376-389)

Another possibility is the model that focuses on the ability of programmers or users to predict the potential for these crimes to occur. According to this second model, a person can be held liable for an offense if it represents a natural and probable consequence of that person's conduct. It only requires that the programmer or user acted deceitfully or has been negligent (ANDREWS, p. 552) considering a result that would be predictable. (HALLEVY, 2010)

Furthermore, regarding civil liability, George S. Cole refers to four types: (i) product liability, (ii) service liability, (iii) malpractice, and (iv) negligence. (COLE, 1990) The author argues that product liability is, at best, only partially applicable. The basic elements for its applicability would be: (i) the AI should be a "product"; (ii) the defendant must be an AI seller; (iii) The AI must reach the injured party without substantive change; (iv) the AI must be defective; and (v) the defect shall be the source of the damage. Hence the service liability would be, in the view of the author, better applied, but not well defined.

[16] The problem with this approach is that Things start to train itself without needing human input. The evolution of this feature usually has the following narrative: "Just over 20 years ago, in 1997, chess champion Garry Kasparov lost his reign to the Deep Blue supercomputer. If in 1997 Deep Blue made history, in 2017 it was the turn of another supercomputer, the Alpha Go Zero, that won several human adversaries in the complex game "Go". In fact, since its previous version, Alpha Go had been dominating the headlines by successively winning Go's best players, largely because of its ability to gather data from its opponents and learn from matches played so far. The results obtained by Alpha Go Zero are relevant because it comes from an artificial intelligence technique called reinforcement learning, which is only possible due to the ability to store, process and analyze data, habits and tactics of the players. It is a technique in which, when trying different approaches to a problem, the computer learns the best solution, without, however, needing any programming or previous teaching by a human. In this way, the computer becomes capable of doing things without any programmer having to teach it beforehand. Alpha Go Zero has been trained only from its own experience with the personal data management of players and matches, which allows it to surpass human capabilities and operate in areas where humans are lacking. In the same sense, the newest version of the AlphaZero supercomputer, also through the reinforcement learning technique, dominated the game in just four hours after being programmed with the rules of chess (without any strategies), having been able to defeat the best chess computer program until then, the Stockfish.".

(COLE, 1990) On the other hand, the author maintains that applying the malpractice liability, in turn, has great potential.[17] Thus, the decision would lie between strict liability and negligence. The standard, in this case, should be set by the professional community. (COLE, 1990)

Still, as the field develops, for Cole, the negligence model would be the most applicable. However, it can be difficult to implement, especially when some errors are unpredictable or even unavoidable. To date, the courts have not yet formulated a clear definition of the responsibility involved in creating AIs which, if not undertaken, should lead to negligent liability. (COLE, 1990)

Should an act of an Artificial Intelligence cause damages by reason of deceit or negligence, manufacturing defect or design failure as a result of blameworthy programming, existing liability rules would most often indicate the "fault" of its creators. However, it is often not easy to know how these programs come to their conclusions or even lead to unexpected and possibly unpleasant consequences. This harmful potential is especially dangerous in the use of Artificial Intelligence programs that rely on *machine learning* and *deep learning* mechanisms, in which the very nature of the *software* involves the intention of developing an action that is not predictable, and which will only be determined from the data processing of all the information with which the program had contact.

Scientists from different areas are concerned and deliberate that conferring this autonomous "thinking" ability to machines can necessarily give them the ability to act contrary to the rules they are given (PAGALLO, 2013); (VLADECK, 2014, p. 120-121). Hence the importance of taking into consideration and investigating the spheres of control and influence of designers and other agents during the creation and functional development of technical artifacts.[18] [19]

[17] Because the text was produced in 1990, the author states that the model of malpractice liability would not yet apply because programming was not officially a profession. However, this concept must be updated, since today this profession is widely recognized.

[18] Accordingly, some relevant concerns are raised: Would it be possible hold liable the companies that designed, programmed, or manufactured the machine, even if they had included programming rules that would prevent harmful behavior to humans? Should creators be totally blamed whenever something goes wrong, even when projected machines "self-teach"? In that case, would the damage-causing conduct already prove defective? Or would one adopt a theory that values the subject's economic position for accountability, that creators are in a better position to absorb the cost of harm than the person harmed?

[19] The engineers are responsible for thinking about the values that will go into the

Often, during the design phase, the consequences are indeterminate because they depend partly on the actions of other factors and agents besides the designers. Also, since making a decision can be a complex process, it may be difficult for a human to even explain it. It may be difficult, further, to prove that the product containing the AI was defective, and especially that the defect already existed at the time of its production[20]. (CERKA, 2015, p. 376-389)

As the behavior of an AI is not totally predictable, and its behavior is the result of the interaction between several human and non-human agents that make up the sociotechnical system and even of *self-learning* processes, it can be extremely difficult to determine the *causal nexus* (JUSBRASIL, 2018) between the damage caused and the action of a human being or legal entity. (MULHOLLAND, 2010).[21]

design of the artifacts, their function and their use manual. What escapes from the design and use manual does not depend on the control and influence of the engineer and can be unpredictable. That's why engineers must design value-sensitive technical artifacts. An artifact sensitive to constitutionally guaranteed values (deliberate in the public sphere) is a liable artifact.

[20] With this regard, to enhance the transparency and the possibility of accountability in this techno-regulated context, there is nowadays a growing movement in civil society demanding the development of "explainable artificial intelligences". Also, the debate around a "right to explanation" for algorithmic and autonomous decisions that took place on discussions around the General Data Protection Regulation (GDPR) is also a way to achieve the goals of transparency and accountability since algorithms are taking more critical decisions on our behalf and is increasingly hard to explain and understand its processes.

[21] Caitlin Sampaio Mulholland addressed the problem of distributed irresponsibility (the name attributed in the present work to refer to the effect of the lack of identification of the causal nexus between the agent's conduct and the harm produced) in her theory of causality presumption. Caitlin Mulholland, in her analysis, addresses the scenario of lack of clarity in the causal nexus between agents reinforcing the concept of "alternative causality". The concept of alternative causality allows us to identify that the damage was caused by a single behavior that, due to the cohesive characteristic of the group, remains impossible to attest. The objective of this liability is to seek reimbursement from the victim, assuming the causal nexus. According to Mulholland: "If there are several activities, each of which alone would have been sufficient to produce the damage, but in which uncertainty persists as to what caused it, each will be considered as the cause of the damage limited to the corresponding probability of having caused it ". (...) There is a single causal nexus that cannot be directly identified. Hence their presumption in relation to the group. (...) What is sought with the alternative causality is to enable the repair of damages caused by facilitating the burden of proof. Instead of having to prove that a person through his conduct has caused the damage that has afflicted it, he may rely on the presumption of causality, suffice it to prove that he has suffered damages and that the damage was the consequence of a certain activity performed by a certain group". Mulholland's argument goes beyond the cases of: (a) joint liability of the accused; (b) liability attributed according to the causal contribution of each agent to obtain the harmful result; (c) the liability attributed to only one of the agents, when it is possible to identify the rupture of the causal nexus between successive actions.

According to the legal framework we have today, this can lead to a situation of "distributed irresponsibility" (the name attributed in the present work to refer to the possible effect resulting from the lack of identification of the causal nexus between the agent's conduct and the damage caused) among the different actors involved in the process. This will occur mainly when the damage transpires within a complex sociotechnical system, in which the liability of the intelligent Thing itself, or of a natural or legal person, will not be obvious. (UNESCO, 2017)

As argued by Alan Turing and the Oxford Internet Institute (MITTELSTADT, 2016):

> The modular design of systems can mean that no single person or group can fully grasp the manner in which the system will interact or respond to a complex flow of new inputs." From traditional, linear programming through to autonomous algorithms, behavioural control is gradually transferred from the programmer to the algorithm and its operating environment. The gap between the designer's control and algorithm's behaviour creates an accountability gap wherein blame can potentially be assigned to several moral agents simultaneously.

Corroborating this thesis, according to the recent UNESCO Report (UNESCO, 2017) on "robotics ethics":

> The rapid development of highly intelligent autonomous robots, then, is likely to challenge our current classification of beings according to their moral status, in the same or maybe even more profound way as it happened with non-human animals through the animal rights movement. It may even alter the way in which human moral status is currently perceived. Although still resembling futuristic speculations, questions like these should not be dismissed lightly, especially in view of the fact that the "human-machine divide" is gradually disappearing and the likelihood of future appearance of human-machine or animal-machine hybrids or cyborgs (robots integrated with biological organisms or at least containing some biological components). [...] In all of these cases, there seems to be a "shared" or "distributed" responsibility between robot designers, engineers, programmers, manufacturers, investors, sellers and users. None of these agents can be indicated as the ultimate source

When we think, however, of the damage caused within sociotechnical systems, we have an even more complex application of the casual nexus and liability, since we are often talking about the action caused by a sum of agencies of human beings, institutions and intelligent things with autonomy and agency power. In this case, the focus on the economic group, despite being able to respond to various damage cases, may not be enough for the fair allocation of liability in the IoT era and for strong artificial intelligence.

> of action. At the same time, this solution tends to dilute the notion of responsibility altogether: if everybody has a part in the total responsibility, no one is fully responsible. This problem is known as the "problem of the many hands". […] Robots may be used for purposes intended by their designers, but they may also be used for a variety of other purposes, especially if their "behaviour" can be "hacked" or "reprogrammed" by their end-users. Robots might have implications far beyond the intentions of their developers. It is impossible for roboticists to predict entirely how their work might affect society.

Another interesting point worth considering in this context is that flaws are natural and can be considered desirable for the faster improvement of a technical artifact. Therefore, a regulatory scenario that would extinguish all and any flaws or damages would be uncalled for. The ideal regulatory scenario would guide the development of the technical artifacts and manage it from a perspective of fundamental rights protection.

No reliable answers have yet been found on how to deal with the potential damages that may arise due to programming errors, or even due to *machine learning* processes that end up incorporating undesired conducts into the behavior of the machine that were not predicted by developers (HUNT, 2016). Therefore, establishing minimum ethical foundations for regulating purposes is just as important as developing these new technologies.

When dealing with Artificial Intelligence, it is essential to promote an extensive debate about the ethical guidelines that should guide the construction of these machines. After all, there is a strong growth of this segment of scientific research (ECONOMIA IG, 2017)[22], regulatory scenario included. However, clear parameters of how to conduct this study, from the point of view of ethics, has yet to be defined. The need to establish a *regulatory framework* for this type of technology has been highlighted by some initiatives.

A conference was held in January 2017 in Asilomar (FUTURE OF LIFE, 2017), CA, aiming to establish the definitions of a series of

[22] Recently, "Alpha Go", an Artificial Intelligence developed by Google defeated for the second time the Chinese champion of the Chinese board game Go, considered one of the most difficult strategy games ever created. Just to have a dimension of the complexity of the game, *Go* has around 2.1 × 10 possible positions on a board, while Chess admits a number of legal positions that is between the orders of magnitude 10 and 10. As a comparison, it is estimated that in the universe there are no more than 10 protons.

principles so that the development of Artificial Intelligence programs can be beneficial. The 23 principles are:

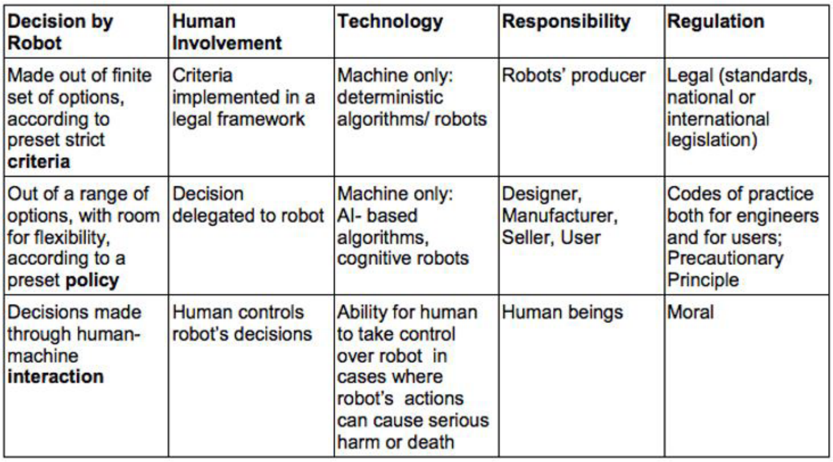
> 1) Research Goal: The goal of AI research should be to create not undirected intelligence, but beneficial intelligence.
>
> 2) Research Funding: Investments in AI should be accompanied by funding for research on ensuring its beneficial use, including thorny questions in computer science, economics, law, ethics, and social studies, such as:
>
> How can we make future AI systems highly robust, so that they do what we want without malfunctioning or getting hacked?
>
> How can we grow our prosperity through automation while maintaining people's resources and purpose?
>
> How can we update our legal systems to be more fair and efficient, to keep pace with AI, and to manage the risks associated with AI?
>
> What set of values should AI be aligned with, and what legal and ethical status should it have?
>
> 3) Science-Policy Link: There should be constructive and healthy exchange between AI researchers and policy-makers.
>
> 4) Research Culture: A culture of cooperation, trust, and transparency should be fostered among researchers and developers of AI.
>
> 5) Race Avoidance: Teams developing AI systems should actively cooperate to avoid corner-cutting on safety standards.
>
> 6) Safety: AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible.
>
> 7) Failure Transparency: If an AI system causes harm, it should be possible to ascertain why.
>
> 8) Judicial Transparency: Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.
>
> 9) Responsibility: Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications.
>
> 10) Value Alignment: Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.

> 11) Human Values: AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.
>
> 12) Personal Privacy: People should have the right to access, manage and control the data they generate, given AI systems' power to analyze and utilize that data.
>
> 13) Liberty and Privacy: The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.
>
> 14) Shared Benefit: AI technologies should benefit and empower as many people as possible.
>
> 15) Shared Prosperity: The economic prosperity created by AI should be shared broadly, to benefit all of humanity.
>
> 16) Human Control: Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.
>
> 17) Non-subversion: The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.
>
> 18) AI Arms Race: An arms race in lethal autonomous weapons should be avoided.
>
> 19) Capability Caution: There being no consensus, we should avoid strong assumptions regarding upper limits on future AI capabilities.
>
> 20) Importance: Advanced AI could represent a profound change in the history of life on Earth and should be planned for and managed with commensurate care and resources.
>
> 21) Risks: Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact.
>
> 22) Recursive Self-Improvement: AI systems designed to recursively self-improve or self-replicate in a manner that could lead to rapidly increasing quality or quantity must be subject to strict safety and control measures.
>
> 23) Common Good: Superintelligence should only be developed in the service of widely shared ethical ideals, and for the benefit of all humanity rather than one state or organization".

As drawn from the text's responsibility section, designers and builders of advanced AI systems are considered stakeholders in the moral implications of their use, misuse, and actions of the Thing and its damaging autonomous actions, with a responsibility and opportunity to shape these implications.

Additionally, the responsibility/liability of the designer should also be considered, coupled with the concern in guaranteeing values such as privacy, safety and ethics in the design of the artifacts. This aims to avoid problems *a posteriori*. But we have always to consider what was within the sphere of control and influence of the designer. We must think, therefore, of a "value-sensitive design". As an example, we can mention the commands of: *"privacy by design"*, *"security by design"* and, *"ethics by design"*.

The degree of autonomy allotted to the machine must also be thought of, determining what degree of autonomy is reasonable and where substantial human control should be maintained. The structure contained in the table *below*, produced in the UNESCO (UNESCO, 2017) study, contains important parameters that help us think about these issues, at the same time trying to identify the different agencies involved. Although the proposed structure is simple, its implementation in terms of assigning responsibility and regulating usage is complex and challenging - for scientists and engineers, policy-makers and ethicists.



Decision by Robot	**Human Involvement**	**Technology**	**Responsibility**	**Regulation**
Made out of finite set of options, according to preset strict **criteria**	Criteria implemented in a legal framework	Machine only: deterministic algorithms/ robots	Robots' producer	Legal (standards, national or international legislation)
Out of a range of options, with room for flexibility, according to a preset **policy**	Decision delegated to robot	Machine only: AI- based algorithms, cognitive robots	Designer, Manufacturer, Seller, User	Codes of practice both for engineers and for users; Precautionary Principle
Decisions made through human-machine **interaction**	Human controls robot's decisions	Ability for human to take control over robot in cases where robot's actions can cause serious harm or death	Human beings	Moral

Alternatively, on 16 February 2017, the European Parliament issued a resolution with recommendations from the European Commission on *civil law* rules in robotics (2015/2103 (INL)). Among other issues, the document advocates for the creation of a European agency for robotics and artificial intelligence, to provide the necessary technical, ethical and regulatory expertise (CMS LAW NOW, 2017).

The European Parliament also proposed the introduction of a specific legal status for smart robots as well as the creation of an insurance system or compensatory fund with the aim of creating a protection system for the use of intelligent machines.

Regarding the legal status that could be given to these agents, the resolution uses the expression "electronic person" or "*e-person*". In addition, in view of the discrepancy between ethics and technology, the European directive rightly states that dignity, in a deontological bias, must be at the center of a new digital ethic.

The attribution of personality to intelligent robots (FAUTH)[23] seems correct and coherent with the gain of autonomy (HUPFER)[24] of

[23] The characteristics most used for the foundation of the human personality are: consciousness; rationality; autonomy (self-motivated activity); the capacity to communicate; and self-awareness. Another possible social criterion is to be considered a person whenever society recognizes thus recognizes one (we can even apply the Habermasian theory here, through a deliberative process in the public sphere). Other theorists believe that the fundamental characteristic for the attribution of personality is sensibility, which means the capacity to feel pleasure and pain. According to Juliana de Andrade, based on the theory of 'depersonalized entities', defended by Daniel Lourenço: "Firstly, for us, as seen, the nonhuman animal is a sentient being, just like man, and, therefore, must have their interest in not suffering as well protected by our legal system - which, in fact, was already done by the 1988 Brazilian Federal Constitution, by prohibiting the practice of cruel acts against nonhuman animals. Hence, civilian legislation must adapt to this reality and recognize the status of a subject of law of the nonhuman animal". In addition, the Law has already broken a major barrier to the attribution of personality by granting personality to legal entities. LOURENÇO, Daniel Braga. *Direito dos animais: fundamentação e novas perspectivas*. Porto Alegre: Sergio Antonio Fabris. Ed., 2008, p.141. ANDRADE, Juliana. *A natureza jurídica dos animais: rompendo com a tradição antropocêntrica*.

[24] The Kantian moral establishes its foundation in autonomy, whose realization implies the necessity of freedom and is characterized by the capacity to think and to act by oneself. According to this moral, every human being, as long as he is rational, can attain autonomy, that is, give himself the direction for his life. For this realization, it is enough that you have the courage to make use of your own understanding, that is, of thinking for yourself. It is necessary to remember, however, that for the human being to "self-determine", he needs to live in community. Habermas complements this initial Kantian conception. In the words of Haide Maria Hupffer: "Habermas advances indicating that autonomy must also be understood as the principle of democracy. For Habermas, morality is a process of argumentation between a free and autonomous society. The author seeks to reconstruct the internal link between popular sovereignty and human rights by introducing the principle of discourse. Habermas introduces his way of interpreting the concept of autonomy, based on the principle of discourse, that is, autonomy lies in a communicative freedom, presupposed in the action that is guided by mutual understanding. For a norm to be universal, consensus is necessary, that is, for us to feel that we are the recipients of rights, it is necessary to be held as authors of law.". The importance of bringing

Intelligent Things. In this sense, as argued by professor David Vladeck, from George Town University:

> One solution would be to reconceptualize these autonomous, intelligent machines as entities with the status of a "person" under the law. Conferring "personhood" on these machines would resolve the agency question; the machines become principals in their own right, and along with new legal status would come new legal burdens, including the burden of self-insurance. This is a different form of cost-spreading than focusing on the vehicle's creators, and it may have the virtue of necessitating that a broader audience - including the vehicle's owner - participate in funding the insurance pool, and that too may be more fair.

The attribution of rights to robots and the creation of a personality of its own, is not a novelty. In Brazilian legal doctrine, Marco Aurélio de Castro in his work dated back to 2009, titled "Law and Post-Humanity: When Robots Will Be Rights Subjects" (CASTRO, 2009), already argued in that direction.

In line with Lehman-Wilzig's teachings, Castro argues that there is no clear meaning for the concept of "person", therefore it cannot be argued that to be deemed a person, it is necessary to be a human-being. Today we have the precedent of legal entities framed with the legal status of a person[25] (CASTRO, 2009)[26]. Castro then defends the possibility of artifacts also falling under this concept, since a robot can perform activities previously considered to be exclusive to human beings, such as: predicting, choosing, learning, understanding, interpreting, analyzing, deciding, feeling[27], among other abilities and skills. In Castro's words (CASTRO, 2009)[28]:

Habermas to the text can be sustained by the fact that in Habermas morality is a result of an argumentative process between free and autonomous beings.

[25] With the purpose, for example, of better managing the property rights.

[26] The legal concept of a person is changeable and is constantly evolving. For example, Afro-descendants have once been excluded from this category, at the time of slavery. Therefore, one cannot relate the legal concept of a person to *Homo sapiens*. In analogy, etymologically the term robot means forced labor. There is nothing to prevent them from migrating to the category of rightsholders as well, since they perform the same actions as human beings.

[27] A reservation is necessary at this point because even if robots can feel and demonstrate emotions as if they were sensuous, the authenticity of these reactions is questioned since they would not be genuine, but at most a representation (or emulation), analogous to human actors when they simulate these emotions in a play, for example, feelings in certain roles, not being considered by many as something genuine. Because of this, the Italian jus-philosopher Ugo Pagallo calls this 'artificial autonomy'.

[28] Original text: "Descobertos os elementos que, reunidos ou isoladamente resultam

> Having discovered the elements that, together or separately, result in the personality of the jurisdictional individual, it is lawful to state that if another entity is found endowed with these same elements, the logical conclusion is to attribute the same judicial status of a person. […] Brain and computer are not equivalent, a fact of little importance because if its manifestation is and effect or an intelligent act, what causes it will have to be intelligent, because what remains in the thought cannot be evaluated, only its result. What happens inside a computer, when in operation, is often an unfathomable mystery, as it is still the mystery of what occurs in the brain when we think.

Given that each Intelligent Thing has a different potential, Castro also advocates a differentiation analogous to the civil and penal distinction based on human capacity. Therefore, only intelligent Things with the same human characteristics could be considered absolutely capable. What he proposes is that parameters or thresholds be created so that, from a legal point of view, there are: incapable robots (those without moral responsibility), relatively capable robots (those who are monitored and supervised, and whose most critical decisions require human intervention), or those who are as capable as a human adult, without legal restrictions.

One of the important features to consider is the learning speed and individual evolution of the robot (based on data processing), which may represent in some cases the infeasibility of an educational process, thus limiting its moral and legal liability.

But how could one punish a robot? It could not be as simple as "pulling the plug". In this case, there are two viable options: rehabilitation and indemnification. The first would involve reprogramming the guilty robot. The second, would be to compel the same to compensate the victim for the damage caused.

In such a context, the European resolution is of extreme relevance. The proposition in assigning a new type of personality, an electronic one, considering the characteristics of Intelligent Things,

na personalidade do indivíduo jurisdicizada, é lícito afirmar que, se outro ente for encontrado dotado desses mesmos elementos, a conclusão lógica é a de se atribuir o mesmo status jurídico de pessoa. [...] Cérebro e computador não se equivalem, o que pouco importa, pois, se a sua manifestação for um efeito ou ato inteligente, o que o causar haverá de ser inteligente, pois o que permanece no pensamento não pode ser de forma alguma avaliado, apenas seu resultado. O que acontece no interior de um computador, quando em funcionamento, muitas vezes é um mistério insondável, como ainda é o mistério do que ocorre no cérebro quando pensamos.".

coupled with the idea of compulsory insurance or a compensatory fund can be an important first step.

The new European proposal reflects a practical and prompt response to the previously mentioned problem of "distributed irresponsibility"[29], which occurs when there is no clear connection between an agent and the harm generated.

Caitlin Sampaio Mulhollan addressed the problem of distributed/ diffuse liability in her thesis on the presumption of causality. In her analysis, Caitlin Mulholland faces the scenario of unclear causal nexus between agents reinforcing the concept of alternative causality. According to Mulholland, in view of the existence of a single causal nexus that cannot be identified directly, we can infer its presumption from the economic group, making it possible to repair the damages caused by facilitating the burden of proof for the victim.

However, when we think of the damages that have occurred within complex sociotechnical systems, we have an application of the causal nexus and of even more challenging legal liability. This is because we are often talking about the action caused by a sum of agencies of human beings, institutions and intelligent things with autonomy and agency power of their own. In this case, the focus on the economic group, despite being able to respond to several cases of damages, may not be sufficient for the fair allocation of liability in the IoT and artificial intelligence Era (MULHOLLAND, 2010).

Therefore, as a pragmatic response to this scenario of uncertainty and lack of legal appropriateness, the European proposal suggests that in case of damages the injured party may either take out the insurance or be reimbursed through the compensatory fund.[30]

[29] This legal phenomenon is also called by other authors as "problem of the many hands" or "accountability gap".

[30] The type of insurance that should be applied to the case of intelligent robots and which agents and institutions should bear this burden is still an open question. The European Union's recent report (2015/2103 (INL)) issued recommendations on the subject, proposing not only mandatory registration, but also the creation of insurance and funds. According to the European Parliament, insurance could be taken by both the consumer and the company in a similar model to those used by the car insurance. The fund could be either general (for all autonomous robots) or individual (for each category of robot), composed of fees paid at the time of placing the machine on the market, and / or contributions paid periodically throughout the life of the robots. It is worth mentioning that, in this case, companies would be responsible for bearing this burden. Despite this proposal, however, the topic continues open to debate, with new alternatives and more interesting models - such as private funds, specific records, among other possibilities - that will not be the subject of a deep analysis in this thesis.

It is worth highlighting the section on Liability as determined by the Resolution (PARLAMENTO EUROPEU, 2017) (EUROPEAN PARLIAMENT, 2016):

> Liability:
>
> 31. Calls on the Commission, when carrying out an impact assessment of its future legislative instrument, to explore the implications of all possible legal solutions, such as:
>
> a) establishing a compulsory insurance scheme whereby, similarly to what already happens with cars, producers or owners of robots would be required to take out insurance cover for the damage potentially caused by their robots;
>
> b) ensuring that a compensation fund would not only serve the purpose of guaranteeing compensation if the damage caused by a robot was not covered by an insurance – which would in any case remain its primary goal – but also that of allowing various financial operations in the interests of the robot, such as investments, donations or payments made to smart autonomous robots for their services, which could be transferred to the fund;
>
> c) allowing the manufacturer, the programmer, the owner or the user to benefit from limited liability insofar as smart autonomous robots would be endowed with a compensation fund – to which all parties could contribute in varying proportions – and damage to property could only be claimed for within the limits of that fund, other types of damage not being subject to such limits;
>
> d) deciding whether to create a general fund for all smart autonomous robots or to create an individual fund for each and every robot category, and whether a contribution should be paid as a one-off fee when placing the robot on the market or whether periodic contributions should be paid during the lifetime of the robot;
>
> e) ensuring that the link between a robot and its fund would be made visible by an individual registration number appearing in a specific EU register, which would allow anyone interacting with the robot to be informed about the nature of the
>
> f) creating a specific legal status for robots, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons with specific rights and obligations, including that of making good any damage they may cause, and applying electronic personality to cases where robots make smart autonomous decisions or otherwise interact with third parties independently (EUROPEAN PARLIAMENT, 2016).

Still, this step should be closely followed by a continuous debate on the ethical principles that should guide such technical artifacts, coupled with an adequate governance of all the data used in the construction and development of these agents.

In addressing the importance of a profound discussion in society on creating a new personality and regulation of these new technologies, Lawrence B. Solum confirms that (LAWRENCE, 1992):

> Our theories of personhood cannot provide an a priori chart for the deep waters at the borderlines of status. An answer to the question whether artificial intelligences should be granted some form of legal personhood cannot be given until our form of life gives the question urgency. But when our daily encounters with artificial intelligence do raise the question of personhood, they may change our perspective about how the question is to be answered. And so it must be with the hard questions we face today. Debates about the borderlines of status-about abortion, about the termination of medical treatment, and about rights for animals-will not be resolved by deep theories or the intuitions generated by wildly imaginative hypotheticals. Of course, many of us do believe in deep theories; we subscribe to a variety of comprehensive philosophical or religious doctrines. But in a modem, pluralist society, the disagreement about ultimate questions is profound and persistent. Resolution of hard cases in the political and judicial spheres requires the use of public reason. We have no realistic alternative but to seek principled compromise based on our shared heritage of toleration and respect. If there is no common ground on which to build a theory of personhood that resolves a hard case, then judges must fall back on the principle of respect for the rights of those who mutually recognize one another as fellow citizens.

From a legal standpoint, it is fundamental to keep in mind the new nature of a diffused liability, potentially dispersed in space, time and agency of the various actants in the public sphere. We need to think about the context in which assumptions on liability are made. The question that is presented to us is not only how to make computational agents liable, but how to reasonably and fairly apply this liability. We must, therefore, think of a “shared liability” between the different actors working in the sociotechnical network and their spheres of control and influence over the presented situations and over other agents.

However, we are still far from obtaining a reasonable consensus[31] on the establishment of appropriate ethical parameters

[31] In the present article, it is argued that the consensus must be constructed according to Jurgen Habermas’s proposal, that is, through dialectical conflicts in the public sphere.

for the development of algorithms and other intelligent Things. These agents can influence relationships between people, shaping behaviors and world views, especially and more effectively when part of their operation enjoys high technological complexity and autonomy, as it happens in the case of Artificial Intelligence systems with the capacity of reasoning and learning according to *deep learning*[32] techniques in artificial neural networks (AMARAL, 2015).

It is evident that these elements are consistently exerting more influence in the way we organize ourselves in society and, therefore, the scientific and legal advance cannot distance itself from the ethical issues evidently involved. The role of law in this context must be re-interpreted. The legal regulation, democratically construed in the public sphere, should provide the appropriate architecture to for the construction of proper ethical channels so that data can flow and non-human agents can act and be developed within the prescribed ethical-juridical limits. The limits must follow a post-humanist perspective, being capable of envisioning Things as agents in the public sphere, but with a human rights' based approach to guide its development.

Referências

ABBATE, Janet. *Inventing the internet.* Massachusetts: Massachusetts Institute of Technology, 1999.

ABITEBOUL, Serge; ANDRÉ, Benjamin; KAPLAN, Daniel. Managing your digital life. Communications of the ACM, v. 58, n. 5, p. 35, may. 2015.

ACCENTURE. *Digital Trust in the IoT Era*, 2015. Disponível em: <https://www.accenture.com/t20160318T035041__w__/us-en/_acnmedia/Accenture/Conversion-Assets/LandingPage/Documents/3/Accenture-3-LT-3-Digital-Trust-IoT-Era.pdf>. Acesso em: 31 jan. 2017.

_______. *From productivity to outcomes: using the Internet of things to drive future business strategies,* 2015. Disponível em: <https://www.accenture.com/t20150527T211103__w__/fr-fr/_

[32] *Artificial Neural Networks* (ANN) represent a network of multiple simple processors – "neurons" – in which each usually has a local memory. ANN is a complex adaptive system, so that it can change its internal structure based on the information that passes through it.

acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Local/fr-fr/PDF_5/Accenture-CEO-Briefing-2015-Productivity-Outcomes-Internet-Things.pdf>. Acesso em: 28 jun. 2016.

ADVANCED MP. Environmental impact of IoT. *Advanced MP,* [s.d.]. Disponível em: <http://www.advancedmp.com/environmental-impact-of-iot/>. Acesso em: 31 jan. 2017.

AGAZZI, Evandro. El impacto epistemológico de la Tecnología. *Argumentos*, [s.d.]. Disponível em: <http://www.argumentos.us.es/numero1/agazzi.htm>. Acesso em: 31 mar. 2017.

AGHAEI, Sareh; NEMATBAKHSH, Mohammad Ali; FARSANI, Hadi Khosravi. Evolution of the World Wide Web: from web 1.0 to web 4.0. *Internet Journal of Web & Semantic Technology*, v. 3, n. 1, jan. 2012. Disponível em: <http://airccse.org/journal/ijwest/papers/3112ijwest01.pdf>. Acesso em: 27 mar. 2017.

ALLSEEN ALLIANCE MERGES with Open Connectivity Foundation to Accelerate the Internet of Things. *Allseen Alliance*, Beaverton, out. 2016. Disponível em: <https://allseenalliance.org/allseen-alliance-merges-open-connectivity-foundation-accelerate-Internet-things>. Acesso em: 25 jan. 2017

ALMEIDA, Kamila. Projeto pioneiro no Brasil, botão de pânico ajuda a reduzir violência no ES. *ZH Notícias*, abr. 2013. Disponível em: <http://zh.clicrbs.com.br/rs/noticias/noticia/2013/04/projeto-pioneiro-no-brasil-botao-de-panico-ajuda-a-reduzir-violencia-no-es-4119173.html>. Acesso em: 25 jan. 2017.

ALMEIDA, Virgilio A. F.; DONEDA, Danilo; MONTEIRO, Marília. Governance Challenges for the Internet of Things. *IEEE Internet Computing*, jul./ago. 2015.

AMARAL, Gustavo Rick. Uma dose de pragmatismo para as epistemologias contemporâneas: Latour e o parlamento das coisas. *Teccogs: Revista Digital de Tecnologias Cognitivas,* São Paulo, n. 12, p. 92-118, jul-dez. 2015.

DHANJANI, Nitesh. *Abusing the Internet of Things:* Blackouts, Freakouts, and Stakeouts. Newton: O'Reilly Media, Inc., 2015.

DIAKOPOULOS, Nicholas. Algorithm Accountability – Journalistic investigation of computational power structures. *Digital Journalism*, v. 3, n. 3, p. 398, 2015

DN. Tay, a inteligência artificial racista e cheia de ódio da Microsoft, voltou a aparecer. *DN,* mar. 2016. Acesso em: 16 ago. 2017.

DONEDA, Danilo; MENDES, Laura Schertel. Data Protection in Brazil: New Developments and Current Challenges. In: GUTWIRTH, Serge; LEENES, Ronald; HERT, Paul De. (Eds.) *Reloading Data Protection:* Multidisciplinary Insights and Contemporary Challenges. London: Springer, 2014

DONEDA, Danilo, ALMEIDA, Virgilio; MONTEIRO, Marilia. Governance challenges for the Internet of Things. *IEE Computer Society,* v. 19, n. 4, p. 56-59, 2015.

DONEDA, Danilo. *Da privacidade* à *proteção de dados pessoais*. Rio de Janeiro: Renovar, 2006.

HABERMAS, Jürgen. *Between Facts and Norms:* Contributions to a Discourse Theory of Law and Democracy, Cambridge, Polity Press, 1992.

_______. *The Theory of Communicative Action*. Beacon Press. 1987. v. II

HARAWAY, Donna. A cyborg manifesto: Science, technology and socialist-feminism in the late twentieth century. In: HARAWAY, Donna. *Simians, Cyborgs, and Women.* The Reinvention of Nature. Nova York: Routledge, 1991.

HEWLETT-PACKARD COMPANY. *Internet of Things Research Study Report,* jul. 2014. Disponível em: <http://h30499.www3.hp.com/t5/Fortify-Application-Security/HP-Study-Reveals-70-Percent-of-Internet-of-Things-Devices/ba-p/6556284#.VZRsHflVhHw>. Acesso em: 8 fev. 2017.

JONAS, Hans. *O princípio da responsabilidade*: ensaio de uma ética para a civilização tecnológica. Rio de Janeiro: Contraponto, 2015.

LANE, Julia (Org.). *Privacy, Big Data and the Public Good*: frameworks for engagement. Cambridge University Press. 2014.

LATOUR, Bruno. *A esperança de pandora*: nsaios sobre a realidade dos estudos científicos. Trad.: Gilson Cesar Cardoso de Sousa. São Paulo: EDUSC, 2001.

_______. *Ciência em ação*: como seguir cientistas e engenheiros sociedade afora. Tradução de Ivone C. Benedetti. São Paulo: UNESP, 2000.

_______. *Jamais fomos modernos:* ensaio de Antropologia Simétrica. Trad.: Carlos Ireneu da Costa. Rio de Janeiro: Editora 34, 1994.

_______. On Technical Meditation – Philosophy, Sociology, Genealogy. *Common Knowledge*, v. 3, n. 2, p. 29-64, 1994.

_______.; WOOLGAR, Steve. *Laboratory Life*: The Construction of Scientific Facts. Princeton: Princeton University Press, 1986.

LAW, John. and SINGLETON, V. *Performing technologies' stories*: on social construtivism, performance, and performativity. Technology and Culture. 2000.

LAW, John; LODGE, Peter. *Science for Social Scientists*. London: Macmillan Press, 1984.

MAGRANI, Eduardo. *Democracia conectada*: a internet como ferramenta de engajamento político-democrático. Curitiba: Juruá, 2014.

MAGRANI, Bruno et al. *Direitos Intelectuais*, 2014. Disponível em: <https://direitorio.fgv.br/sites/direitorio.fgv.br/files/u100/direitos_intelectuais_2014-2.pdf>. Acesso em: 29 mar. 2017.

MILL, John Stuart. *O utilitarismo.* São Paulo: Iluminuras, 2000

MILLER, Georgia e KEARNES, Matthew. *Nanotechnology, Ubiquitous Computing and The Internet of Things*: Challenges to Rights to privacy and data protection. Draft Report to the Council of Europe, set. 2013.

PAGALLO, Ugo. *The Laws of robots*: crimes, contracts, and torts. Torino: Springer, 2013.

SAURWEIN, Florian; JUST, Natascha; LATZER, Michael. *Governance of algorithms:* options and limitations. Info, v. 17, n. 6, p. 35-49, 2015.

SCHATZBERG, Eric. From art to applied science. *Isis*, v. 103, n. 3, p. 555-563, 2012.

______. Technik Comes to America: Changing Meanings of Technology before 1930. *Technology and Culture*, v. 47, n. 3, p. 486-512, jul. 2006. Disponível em: <http://muse.jhu.edu/article/201479>. Acesso em 27 mar. 2017.

SCHMIDT, Eric; COHEN, Jared. *The new digital age*: Reshaping the future of people, nations and business. Londres: Hachette UK, 2013.

SJÖBERG, Mats et al. Digital Me: Controlling and Making Sense of My Digital Footprint. In: GAMBERINI, L. et al (Eds.). *Symbiotic Interaction:* Lecture notes in computer science. Padua: Springer, 2016. Disponível em: <http://revistas.ua.pt/index.php/prismacom/article/viewFile/681/pdf>. Acesso em: 2 mai. 2017.

VERBEEK, Peter. *Moralizing Technology*: Understanding and Designing the Morality of Things, Chicago - London, The University of Chicago Press. 2011.

VLADECK, David C. Machines without principals: liability rules and artificial intelligence. *Washington Law Review,* v. 89, n. 1, mar. 2014.

WAHER, Peter. *Learning internet of things.* Birmingham: Packt Publishing Ltd, 2015.

WALLACE, K. A. Anonymity. Ethics and Information *Technology*, 1 (1), 23-35. 1999.

WALLACH, Wendell, et al. *Artificial Intelligencefor the Common Good Sustainable*, Inclusive and Trustworthy. 2017. Disponível em: <https://weforum.ent.box.com/v/AI4Good?platform=hootsuite>. Acesso em: 28 fev. 2017.

WALLACH, Wendell; ALLEN Colin. *Moral Machines: Teaching Robots Right from Wrong.* Oxford University Press, 2008.

Recebido em: 25 de setembro de 2018.
Aceito em: 12 de novembro de 2018.