



Human rights perspectives on the implementation of ethical frameworks for artificial intelligence.

J. Carlos Lara Gálvez *

Abstract

The promotion of ethical frameworks, guidelines, and principles for the formulation and implementation of artificial intelligence systems has provided a frame of reference for these expressions of technological development whose deployment implies high risks for individual and collective interests. However, there has been some criticism of the proliferation of non-mandatory guidelines, compared to other normative frameworks of global validity, such as international human rights law. In turn, there is limited understanding of the transposition of this regulatory scheme to advanced technological developments, particularly on its practical consequences and the ability to enforce compliance. This article constitutes an effort to operatively connect the ethical principles on artificial intelligence with the normative framework of the instruments of international human rights law, especially on a Latin American scale, and to lay the foundations to promote the practical implementation of ethical requirements that simultaneously recognizes, respects and promotes fundamental rights.

* Lawyer, Bachelor of Legal and Social Sciences from the University of Chile. Master of Laws (LL.M., Law and Technology) from the University of California, Berkeley. Director of Public Policy on Digital Rights - Latin America. He can be contacted at: jclg@riseup.net

Key words

Artificial intelligence
algorithms
automation
human rights
responsibility

Introduction

The past decade has witnessed a proliferation of frameworks, guidelines, and principles for "artificial intelligence" (AI) ethics focused, in particular, on the formulation, programming, or implementation of AI systems, but often also directed at state regulation of AI use. The trend appears, at least in part, as a reaction to the growth in the use of certain forms of AI by different industries and state structures, with the consequent documentation of negative impacts derived from the delegation of relevant decision-making aspects that have an appreciable impact on the population or part of it in automated systems.

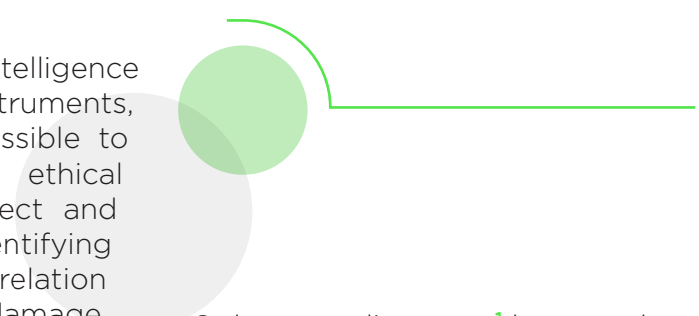
This proliferation of ethical frameworks is part of a debate that is not restricted to academic disciplines but extends logically to a discussion on the governance of algorithmic technology at the level of states, multilateral institutions, and of course, technology companies. In turn, this multiplicity of stakeholders is an expression of the search for solutions that involve legal, ethical and technological aspects of ideal coexistence for the adequate operation of technology.

In this sense, the aforementioned documents have allowed the creation of one or more frames of reference in the face of technological development possibilities involving high risks. However, these are still initiatives to systematize non-mandatory guidelines, ideally adopted but not imperative, in contrast to other action frameworks with deontic aspects (that is, norms) with globally recognized validity, such as the international human rights law. In other words, the ethical initiatives thus formulated stand, in general, separate from the regulatory framework that best identifies the areas of negative impact on the use of technology. In turn, there seems to be a limited understanding of the transposition of a broad regulatory framework to specific advanced technological developments, particularly on its practical consequences, and on the ability to enforce its compliance. In short, the coexistence of documentation on principles for artificial intelligence and frameworks that specifically allude to international human rights law are different political expressions to react as regards the risks assumed or identified by these technologies.

How are ethical principles on artificial intelligence connected to the normative framework of instruments, especially at the regional level? How is it possible to promote the practical implementation of ethical demands that simultaneously recognize, respect and promote fundamental rights? In addition to identifying the existence of ethical frameworks and their correlation with human rights, whether in terms of damage prevention or positive promotion, it is necessary to anchor specific obligations on the design and implementation of automated or automation-assisted decision systems.

The present document intends to abound in literature on the accountability of automated decisions based on the use of international human rights law, the current framework known as the Inter-American human rights system (IAHRS), in general, and the obligations and rights contained in the American Convention on Human Rights (ACoHR) and the Charter of the Organization of American States (COAS), in particular, with the purpose of filling some of the gaps in the application of ethical frameworks. The perspective adopted is the one pertaining to the Latin American context: mostly developing countries, where technology is not a significant part of national production or income, with consequent disadvantages when compared to countries where not only are automation technologies developed, but also where their governance frameworks are discussed, including ethical codes.

Given the IAHRS membership, Latin American states have, in addition to their own internal regulatory capacity, an assortment of tools related to their duties to prevent infringements of human rights, to establish safeguard, monitoring and accountability mechanisms, and to grant complaint mechanisms for their citizens. Furthermore, associated with the duties of prevention, protection and remedy of the states, there exists an expectation of respect for human rights by private companies. This expectation serves as a point of contact for the active participation of states by asserting the different levels of responsibility throughout the development and implementation chain of automation tools.



Only recent literature¹ has sought a closer relationship between automation governance and international human rights law beyond the identification of possible affected rights. However, as well as the technologies to which they refer, these efforts come from countries where there seems to be a greater capacity to influence technological development from the early stages. The search for operationalization for countries such as those Latin America also requires recognizing conditions of unequal development.

The second part of this document summarizes part of the findings in the literature on the impact of artificial intelligence on fundamental rights as a family of technologies. In order to do this, we will build on existing research which is analyzed in turn within the regulatory framework provided by the IAHR. Furthermore, we will add the presence of political and economic aspects that are also part of a thorough analysis from the point of view of international human rights law, including its collective dimensions which are related in turn to other perspectives on technology governance. For these purposes, we will use the term “artificial intelligence” loosely, without reference to a specific form of technology but to a family of technologies that depend on data processing to make predictions, thus including technologies better identified as technologies to make algorithmic or automated decisions.

¹ See, for example, McGregor et al. (2019), who has forged a framework for the application of accountability of international human rights law for algorithms.

The third part will examine the proposed frameworks, guidelines and ethical principles for artificial intelligence particularly focused on transparency and prevention of some of the risks, but which may be insufficient for the range of rights possibly affected and identified in the second part. This is the circumstance that makes it necessary for states to implement their obligations under international human rights law, including action for corporate responsibility. In this regard, we assert our support for the creation of binding international regulations on business and human rights.

In the fourth part, we will propose the application of IAHRs as an appropriate normative framework, from the substantive and procedural point of view, for the governance of artificial intelligence in Latin American countries while moving towards operationalization. In addition to having the ability to identify damage areas, IAHRs allows a range of action throughout the development chain of automation technologies, including the decision on their adoption or non-adoption, in a context of public policies aided by automation. On this point, we recommend updated legal frameworks to assert both state and private responsibility. By the way, given that the main responsible parties for the adoption of measures in this sense are the states, skepticism will remain regarding state action for the regulation of technologies.



2

Artificial intelligence and its impact on fundamental rights

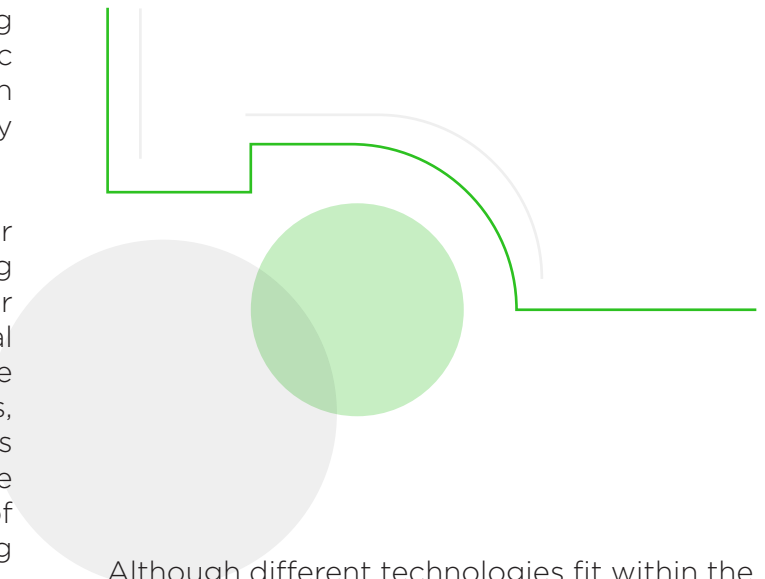
2.1. “Artificial Intelligence”

Extensive literature has existed for at least half a century seeking a definition for artificial intelligence. In very general terms, this is the technology that automatically detects patterns in data and makes predictions from said data (Daly et al., 2019) by simulating human cognitive processes, that is, “intelligence”. With less emphasis on technology, Muñoz León (2013) refers to AI as “the system that

interacts with its environment by selecting information through a certain epistemic apparatus, and that reacts to the information considered by the system itself as relevant by delivering consistent behavioral responses”.

The generic definitions account not only for the non-reference to a single way of applying knowledge through technique. In other words, we are not talking about artificial intelligence in reference to a single technology but to a family of technologies, which includes techniques such as computerized vision, natural language processing, virtual assistants, automation of robotic processes, and automated learning (Bughin et al., 2017).

A common feature of the most commonplace applications within this family of technologies is the processing of huge databases as a way of generating better predictions, either to classify or identify objects or to reproduce complex tasks. This information processing appears as an evolution of other concepts that have also been widely debated recently, such as “algorithmic decisions”, “automated decisions”, or “Big Data”. The key value of the use of AI is in its predictive power, as a component of intelligence (in general terms, as a characteristic assumed as human), as the means of obtaining useful information for decision-making, which at the time the decision is made is not available, but is concluded from existing data (Agrawal et al., 2018).



Although different technologies fit within the notion of “artificial intelligence”, the increased interest from academia and civil society stems fundamentally from its growth in the market as a result of a reported increase in the use of some forms of AI associated with growth in information gathering and processing. At the same time, a certain sense of urgency is manifested, especially in developing countries such as those in Latin America, to adopt technologies that are attributed with having an exaggerated potential to provide solutions to an increasing number of problems (Elish and Boyd, 2017; Broussard, 2018; Citron and Jurecic, 2018), with a confidence in machines that is close to magical thinking. Thus, an apparent need arises to take advantage of the potential of these technologies, with the corresponding preparation of the states in the adoption of AI as a probable development factor in the current global scenario² under the assumption - not questioned - that AI is useful for development.

In short, the discussion on the governance of artificial intelligence, as a family of technologies, requires observation that does not isolate it as an object but rather contextualizes it as a manifestation of the development of industries based on digital technologies, taking in turn that development within a broader context of economic development.

² See, for example, the UNESCO report, *Artificial Intelligence for Sustainable Development* (2019).

2.2. The impact of AI on human right

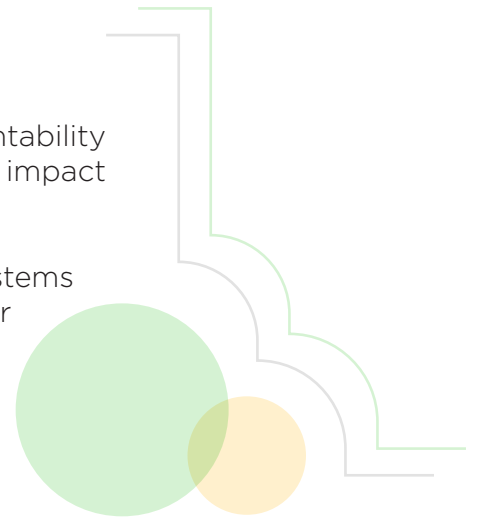
There is a vast body of literature that identifies both V questionable practices of companies using data through increasingly advanced systems to achieve their goals and the negative effects that this entails. As marketing technique and vocabulary vary, the risks of "Big Data" from using "algorithms" or "artificial intelligence" take the leading role in the discussion. And, in general, although algorithm applications with varying degrees of complexity are used in everyday aspects such as content management or social network recommendations, the impact is largely pronounced with regard to the use of systems that O'Neil says (2016) constitute weapons of mathematical destruction on account of their opacity, their scale and the damage they cause. Examples of the negative use of opaque computer systems in the financial system (Pasquale, 2015), in education (O'Neil, 2016), in the criminal prosecution system (O'Neil, 2016; Angwin et al., 2016), in the delivery of social benefits (Eubanks, 2018), and even in search results (Noble, 2018) have become popular. Incidentally, the examples continue to grow³. An explanation indicates that the nature of technologies, as a non-neutral construct, implies the incorporation of the cultures and values of those who design and develop them, from which artificial intelligence technologies are not exempt, including the Silicon Valley culture (Broussard, 2018), in intense juxtaposition to the cultural spheres in which these technologies are intended to be deployed today. This point is especially relevant to deal with the systemic aspects of negative impacts, as will be discussed later.

A common factor in the practical findings of damage situations produced by algorithms in general is the notion of a prediction, classification, identification or recognition which is wrong or unfair. The subsequent effect of the decisions made on the basis of that result (either by a human agent assisted by technology or by a system programmed to make automated decisions) logically constitutes a negative impact on relevant interests (education, social security, justice, employment) consistent with fundamental rights. A rapidly growing body of literature addresses those considerations from the perspective of international human rights law, ranging from broad considerations (Donahoe and Metzger, 2019), through an analytical perspective inherent in the human rights system (Wagner, 2018a;

³ In this regard, see the work of the AI Now Institute. AI Now Report (2017), available at: https://ainowinstitute.org/AI_Now_2017_Report.html, AI Now Report (2018), available at: <https://ainowinstitute.org/litigatingalgorithms.html>, and AI Now Report (2019), available at: https://ainowinstitute.org/AI_Now_2019_Report.pdf

Raso et al., 2018), or through a perspective that uses it as an accountability framework (McGregor et al., 2019) up to ambitious proposals for impact assessment frameworks (Mantelero, 2018).

There is consensus that the application of automated decision systems is capable of having an impact on a wide range of rights. Wagner (2018a) carries out an analysis of the most important rights affected by different applications of the AI family of technologies. Although his analysis is focused on the European Convention on Human Rights (ECHR), the study appropriately highlights rights that are also part of IACHR.



Taking the described study and contrasting it with the range of rights recognized in the IACHR, it can be said that the deployment and application of algorithmic and artificial intelligence systems, particularly in public functions, can especially affect:

- the right to a fair trial and due process (article 8 ACoHR);
- the right to privacy (articles 11.2 and 11.3 ACoHR), in particular in relation to the protection of personal data;
- the right to freedom of speech (article 13 ACoHR);
- the right of assembly and freedom of association (articles 15 and 16 CADH);
- the right to an effective remedy (article 25 ACoHR);
- the prohibition of discrimination and the right to equality under the law (articles 1.1 and 24 ACoHR);
- political rights related to elections (article 23 ACoHR); and
- economic, social and cultural rights (article 26 ACoHR, in relation to COAS).

In the study, Wagner (2018a) provides concrete examples where those rights are especially affected within the European context. However, from the perspectives of a possible risk analysis and an accountability and responsibility procedure, it is not entirely useful to measure a limited series of possible affected rights.

In contrast, the study by Raso et al. (2018) undertakes a perspective based on some prominent AI applications and the possible effects each of these may have on human rights, including possible positive effects. Without judging or endorsing the judgments on these analyzes, it is particularly

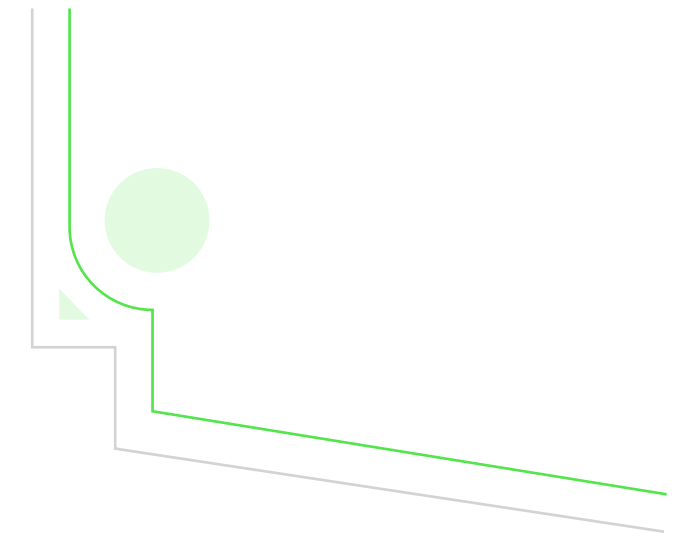
meaningful that the study not only devotes attention to extended forms of AI application such as credit rating and content moderation (also analyzed by Wagner [2019]), but also adds aspects to the analysis that are usually neglected in deeper examination, as occurs with economic, social and cultural rights. The study by Raso et al. (2018), even when undertaking a limited series of examples, is useful to highlight the convenience of a pre-development and implementation analysis of AI, which in its expression of human rights considerations complies with being a general, flexible and open process and at the same time informed by different areas of expertise.



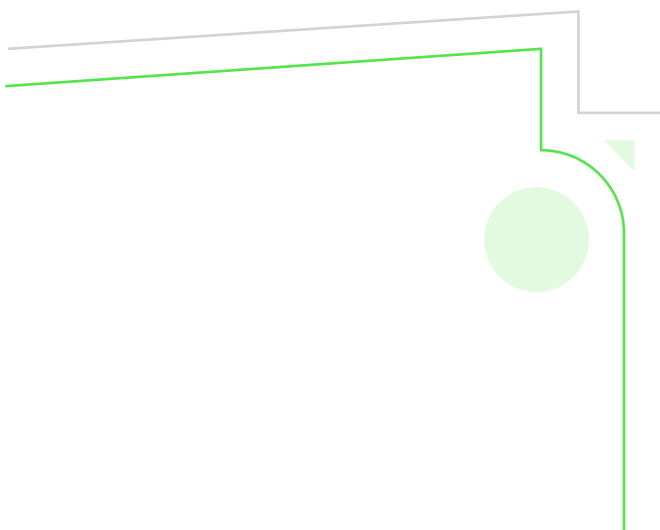
Furthermore, a list of rights affected by a technology or family of technologies, as a regulatory challenge, is not sufficient on its own to guide the adoption of preventive and remedial measures against harmful impacts. An impact analysis based on normative considerations tends to leave aside the study of economic, social and cultural conditions, with global aspects, that shape the existence and dimensions of these risks. This is a limitation that refers both to the fields of technology design and development, and to the fields of deployment and their application. The discussion on the governance of artificial intelligence, including the focus on human rights, not only neglects the economic conditions for technology development under an inherently unequal capitalist system for those who are subject to the application of technologies, but also runs the risk of ignoring the differences in the development stages between the latitudes where technological production is concentrated and where its impacts are intended to unfold. A holistic notion of human rights, including developmental aspects, economic, social, cultural and environmental rights, and of the collective aspects of these interests can be useful for economies in constant disadvantage, such as Latin American economies.

Paradoxically, much of the discussion regarding ethics and fairness in the use of AI algorithms is focused on notions of equality and rejection of arbitrary discrimination made or enhanced by technology, at the individual level, where there are collective inequality dimensions between the developed and undeveloped worlds, between the "global north" and the "global south". The political economy of artificial intelligence (as part of the economic development related to technology in general) is a factor that cannot be ignored when designing regulatory solutions to the negative impacts of its use.

Although international human rights law is not by itself -at least in its normative dimension- sufficient to face development disparities or profound questions about social justice, in our opinion it still constitutes a tool capable of controlling the application of technological developments that serve human development, insofar as the collective dimensions of human rights are assessed as well as the value of economic, social, cultural and environmental rights as part of a single and interconnected system of rights of equal value, and the political aspects that allow the full development of rights, such as (or may be) a democracy in the broadest sense.



A substantive part of the claim from the developing world refers no longer to the storage and processing capacity of information but to the control of the information itself, particularly information related to people that we call "personal data", and the ability of the collective to exercise effective control over that information beyond individual control. In other words, not only a personal data protection framework is key to effective governance over AI (Veale et al., 2018), but also the collective ability to claim control over the use of that information in an aggregate manner (Morozov, 2019), with the institutional and political infrastructure that enables this control. In addition to that normative aspect on data, other forms of control over the companies themselves that exercise their power through the design of AI systems constitute forms of state action that can serve individual and collective interests within a framework of full exercise of human rights, ideally not only at the national or intra-territorial level but also internationally.



3

**Ethical principles for
artificial intelligence**

As Taddeo and Floridi (2018) explain, technological development is guided and takes shape spurred in a variable manner by three normative forces: ethics, regulation, and technology itself, with ethics fulfilling functions both in the formulation of regulations and in the interpretation of normative statements while the three elements act in a complementary manner. From this perspective, it is not strange that the law by itself is not the response mechanism to technology challenges; nevertheless, it is not entirely clear the way in which the interaction between these different regulatory forces works in terms of different areas of application and different technologies.

Despite understanding the multiplicity of ways in which AI governance can be approached, it is in the area of ethics incorporation where the greatest development in recent years has occurred. Without underestimating its value, it should be examined against a normatively established structure of responsibility and rights, and in contrast to the latter, its real or possible effectiveness in AI development and application.



3.1. Principles, frameworks and ethical guidelines

Examining the discussion of ethics in algorithms, Mittelstadt et al. (2016) distinguished epistemic concerns (related to processed information) and normative concerns (related to the result of the algorithmic process) of the use of algorithms, in addition to the traceability problem that covers both dimensions. The study in turn recognizes the chain between data that reflects bias, leading to decisions that materialize discrimination, and the possible transformative effects on autonomy and privacy. For Selbst and Barocas (2018), concentrating on aspects such as the explicability and transparency of technologies can deviate from the value of intuition in the analysis of the normative aspects resulting from an algorithmic decision.

A systematic mapping and comparison exercise of 36 sets of principles and guidelines in force today reveals eight common themes addressed by these documents: privacy, accountability (responsibility), protection and security, transparency and explicability, equity and non-discrimination, human control of technology, professional responsibility, and promotion of humanistic values (Fjeld et al., 2020).

A different analysis (Jobin et al., 2019) summarizes the convergence of most of the 84 documents of ethical principles and frameworks in five items: transparency, justice and equity, non-maleficence, responsibility, and privacy. However, there is substantive divergence on how the principles are interpreted, their significance, the themes, areas or agents to which they refer, and their form of implementation.

These divergences, although not insurmountable if sufficient political agreement exists among stakeholders, reveal some significant gaps intrinsic to ethical frameworks that may lead to maintaining or amplifying the damage risks they seek to prevent. Firstly, several of the efforts to incorporate ethical aspects point to technological development in its different stages, which is the case of aspects such as explicability or data handling. Interestingly enough, other ethical aspects are left out, such as intrinsically destructive uses as in the case of automated weapons (Anderson and Waxman, 2013), or the effects on society regarding automated recommendations in digital social networks, or the environmental consequences of AI development. Research regarding the less immediate social consequences is necessary to shape ethical frameworks with a more complete vision of the risks that are sought to be prevented.

Secondly, the lack of diversity approach in the advanced AI industry is also a significant omission in such ethical frameworks, which ends up reinforcing cultural stereotypes precisely where it is urgent to cope with discrimination problems (Broussard, 2018; Crawford, 2016). In contrast with the abstraction of companies or institutions that undertake ethical commitments, the absence of responsibility or awareness in individual developers regarding the mediate and distant consequences of the systems they develop further alienates the operability of these ethical frameworks.

Thirdly, in close connection with the above, the qualification of a certain culture is not merely a concern related to current economic circumstance but an expression of

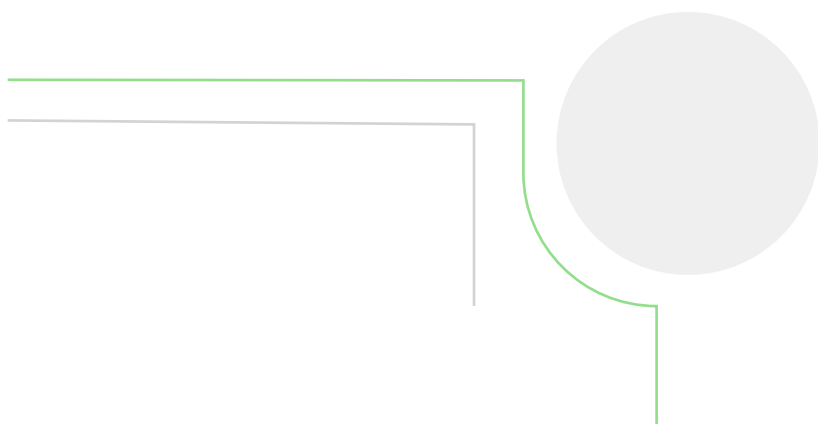
the power to exploit personal information at a massive level (Zuboff, 2019), which is in turn an extension of a power with roots in colonialism and capitalism (Couldry and Mejías, 2019). Although it is not our current purpose to discuss aspects of the global AI economy, the analysis of this relationship often involves blind spots regarding needs and contexts of application, perspectives of economic development, and in general, the collective dimensions in AI application in contrast with the emphasis on discrimination problems affecting individuals (even aggregates). Furthermore, as we will explain at a later stage, ethical frameworks, in view of their origin, tend to make geographic and political differences their own that already warrant questioning the technologies to which they refer.

3.2 Ethical guidelines and their relationship with human rights

Ethical guidelines and the law constitute, for Gasser and Almeida (2017), distinct levels of a three-tier model for AI governance. Under this model, a first basal layer is the one referring to the technical dimension of algorithms and data (including their governance), a second intermediate layer is the one referring to the ethical dimension, and a third layer is legal and social, including standards and rules, and having been built on the previous ones. This theoretical model illustrates a bond between ethics and law in which the latter creates more specific behavior patterns for technology. We believe, however, that the relationship is not necessarily built on abstraction levels of normative postulates, but rather on the operability that the law provides, especially through the institutional framework.

As regards the substantive normative content, for reasons other than its content limitations, it is necessary to assess the role of the law and, in particular, international human rights law to provide a governance framework for artificial intelligence.





Firstly, this is necessary in light of the origin of ethical frameworks. For Calo (2017), the ethical guidelines of the AI industry suggest to the regulatory state that self-regulation is sufficient to model the behavior of those who develop AI, without the need for specific laws to mitigate risks and eliminate abuse scenarios. As Hasselbalch (2019) claims, the different initiatives account for spaces for negotiation between different interest groups that try to lead the cultural definition of data ethics (and, therefore, its uses), within complex power relations exercised in cultural positioning. It seems a logical consequence of the fact that several of the same private companies in the AI industry have funded artificial intelligence ethics initiatives, thus postponing a serious debate on the legal regulation of AI (Nemitz, 2018). This also postpones concrete state and international actions to seek responsibility for harmful effects on human rights. For Wagner (2019), this situation risks representing a fallacious form of assuming effective ethical commitments while avoiding legal responsibilities⁴.

Secondly, it is necessary to assess the role of the law against ethical guidelines in view of the risks that both groups of rules seek to face. The discussion on ethics has not ignored the challenges that artificial intelligence presents for the law, including human rights. However, the ethical frameworks described fail to address the full spectrum of risks and governance challenges, and are not universally accepted either (Donahoe and Metzger, 2019), perhaps as a consequence of their proliferation as products of different stakeholders. For

Donahoe and Metzger (2019), a comprehensive, systematic and global governance framework is necessary, the foundation of which exists in international human rights law. Convenience is justified by four characteristics of the international human rights law framework: the person is placed at the center of the assessment, a wide range of concerns is covered, the roles and duties of governments and the private sector are defined, and it has the advantage of relying on global support (Donahoe and Metzger, 2019). Although it seems desirable to have this shared regulatory governance framework, its formulation would again seek the development of global guidelines without regional considerations, as occurs with the development needs and prospects of Latin American countries. Even outside the problem of technology governance, some of the aspects of international human rights law are already being questioned due to lack of cultural sensitivity. While waiting for a contemporary global consensus on the urgency for the incorporation of AI technologies, turning our gaze to our region is not only desirable and sensitive to our differences but also practical.

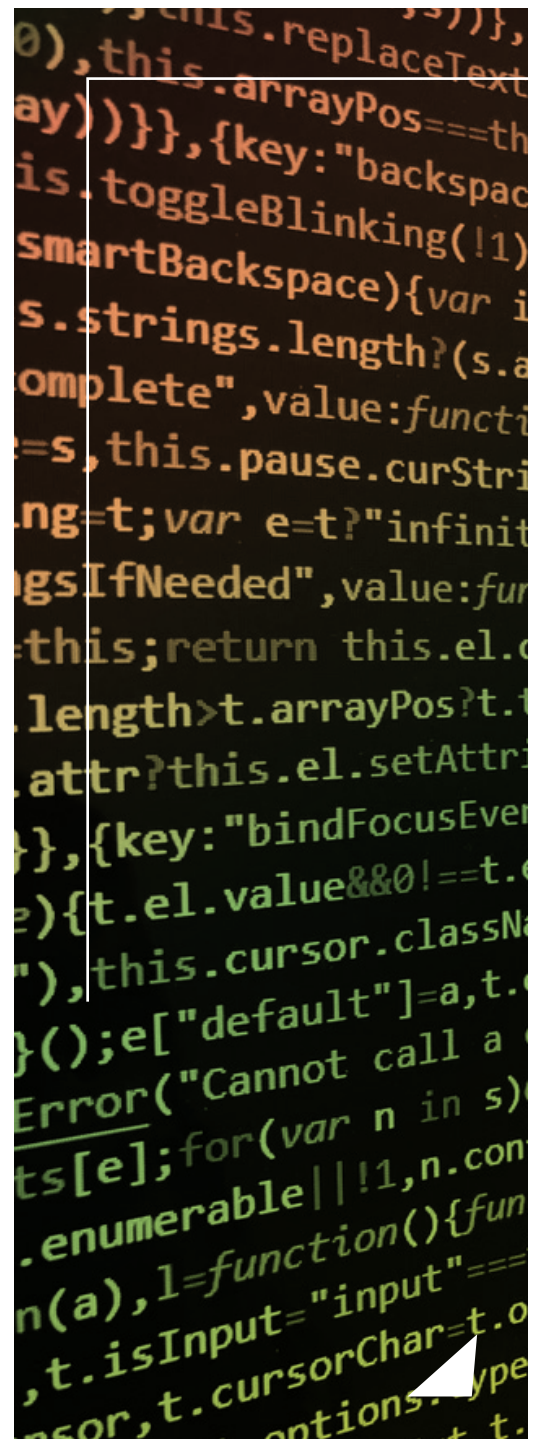
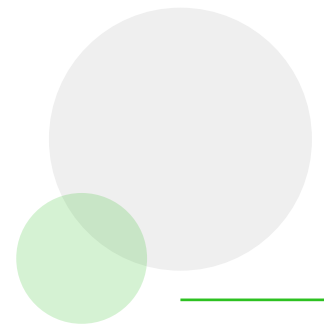
⁴ Indeed, regardless of the origin of the initiatives on ethical guidelines for AI, it is also necessary to recognize that the beginning of a discussion in the most recent context, from a regulatory perspective, also existed in the European context. See "Artificial Intelligence, Robotics, Privacy and Data Protection. Room document for the 38th International Conference of Data Protection and Privacy Commissioners", in the context of the 2016 ICDPPC in Marrakech, available at: https://edps.europa.eu/sites/edp/files/publication/16-10-19_marrakesh_ai_paper_en.pdf

Thirdly, it highlights the lack of mechanisms for the observance of ethical guidelines that are operated at least in part by people other than those who declare making the principles or guidelines their own. So far, these are voluntary commitments, lacking institutions and compliance procedures, and no legal backing to make them coercible at least with regards to other co-contractors. On the one hand, non-compliance with ethical standards does not necessarily lead to negative or unwanted effects for those who do not comply, and it certainly makes it difficult to attribute ethical violations where there are distributed design and development systems among many people. On the other hand, the real individual or collective impact derived from the infringement of ethical guidelines, if it occurs, does not find remedy or claim mechanisms in ethics itself. That is, because it focuses on the fulfillment of pre-established obligations, it does not ensure accountability for subsequent consequences, thus creating a responsibility gap (Liu and Zawieska, 2017).

Fourthly, in practical terms, there is no evidence of the usefulness of ethical frameworks in the development of artificial intelligence technologies. The commitment of private companies to ethical principles does not necessarily mean that their entire production line assumes different obligations in their daily operations. Empirically, there is no evidence that an ethical framework affects that operation. McNamara et al. (2018), measuring the influence of the code of ethics for software development of the Association for Computing Machinery (ACM), a 1972 document updated in 2018, found that there was no evidence of significant influence on the decisions made by developers of computer programs.

The latter may be related to the presence of codes of ethics as an added element to a chain that is capable of operating without them, a chain that finds those codes strange, distinct from production itself in the AI industry. As a framework for conduct without real consequences for individuals within complex business operations, a series of principles does not seem sufficient to raise awareness about social consequences or about a sense of responsibility. The absence of mechanisms for ethical observance, as happens in professional associations, deprives ethical principles of real normative value.

Fifthly, as an extension of the aforementioned statement, the effective influence of frameworks, principles or ethical guidelines must also be relativized in terms of the groups, institutions, and members of academia who pursue ethics in AI technologies. These groups are far from being agents of change with the ability to introduce modes of action in the AI industries.



Incidentally, these are not aspects that are unique to ethics in the use of data, algorithms or artificial intelligence, but are shared in the global economy with dissimilar participation in technology development and deployment with insufficient self-regulation and regulation levels.

Thus, even though it is a positive aspect that there are intentions and principles for ethics in algorithms and AI, this cannot serve to replace or avoid legal regulations, which inhibits the democratic process (Nemitz, 2018).

A huge aspect not covered by ethical principles is the one regarding inequalities between industrialized countries with strong participation in AI and developing countries, including those in Latin America. There are several relevant inequalities there, including the level of economic and technological development. The impact that the concentration of economic power on a global level has on development may even hamper the digital advance of developing countries⁵. As a consequence, unequal capital conditions entail the reproduction of unequal structures of opportunity and choice, and of collective economic self-determination. If AI serves to achieve development - and, consequently, if it is desirable for developing countries -, the discussion about its governance cannot be limited to the liberalism and individualism that dominate the debate, whose ideas ignore the dependence that hegemonic capital imposes on individuals and collectives by means of reduced choice structures in a broader neoliberal global order (Gurumurthy and Chami, 2019). Addressing economic exploitation by expanding individual and collective choices, overcoming bias and discrimination problems, is a challenge that exceeds current ethical frameworks while adding a complex dimension to a possible normative response.

4 Towards normative implementation of ethical governance

Adding and transposing ethical and human rights considerations into the normative and public policy arena can be done from a dual perspective of responsibility. These dimensions are those corresponding to prospective liability, aimed at preventing and mitigating risks and threats, and to historical or retrospective liability, aimed at fairly ensuring the attribution of responsibility in

⁵ For the United Nations Conference on Trade and Development, 'the emergence of monopolistic practices by leading companies and digital platforms in all global value chains, as well as the widening of the digital divide, makes it extremely difficult for developing countries, especially LDCs [least developed countries], to take on their own the leap towards digital industrialization' (UNCTAD, 2018).

the face of damage or interference regarding fundamental rights. Only if both dimensions are reclaimed, individuals and society may trust that efforts exist to prevent damages and institutional mechanisms to ensure reparation and mitigation of further damage (Yeung, 2019).

The governance of algorithmic systems in its aspect of state intervention offers at least four public policy options. These options are: the creation of awareness through education, monitoring and reporting; responsibility for algorithmic decisions in the public sector; regulatory oversight and legal liability; and global governance coordination (Koene et al., 2019). While all of these options acknowledge a diverse set of relevant stakeholders with different roles and responsibilities, the participation of states is crucial both in their normative roles at the national level and in their roles of national representation in the international context, including intergovernmental forums and human rights treaty bodies.

4.1. The role of the law

Placing emphasis on the development of a human rights framework for AI fulfills at least three functions. First, it allows for the possibility of confronting the absence of effectiveness or forced observance of ethical frameworks. Second, it allows the possibility of confronting the lack of participation in the formulation of AI technologies and the formulation of the ethical frameworks that refer to it from the capacity of Latin American states. Third, it opens up the possibility of a homegrown development of technologies with a more solid anchorage in the consequences of human rights. The latter also requires taking responsibility for development conditions that seem to exceed the scope of human rights; nevertheless, there are several channels of cooperation that may already be taken, at least regarding AI, which are closely related to the problems surveyed by the discussion on ethics (World Wide Web Foundation, 2017).

Legal regulation also constitutes a normative expression of the protection role of the relevant interests of its citizenships carried out by the state. As Nemitz (2018) claims, legal regulation recognizes that there is a normative dimension to algorithms as guides to human action, and said dimension can be assessed in a manner akin to legal review in light of higher standards, such as the Constitution and the instruments of international human rights law. In addition, a regulation constitutes a substantive advantage over ethics and self-regulation, in view of the concentration of power in a handful of companies: democratic legitimacy and

enforcement mechanisms, establishing a level ground with a focus on innovation and incentives so that it may be addressed to the public interest (Nemitz, 2018).

This does not imply getting rid of ethical frameworks. Despite the still-unresolved shortcomings, the literature optimistically suggests that ethical frameworks are useful in dealing with AI risks. This role can be assessed even in Latin America: ethics is rightly attributed "a decisive role for the development of inclusive and fair systems, limiting the biases that can be found both in data and in model building" (Mantegna, 2019). Prevention methodologies, even if they cannot clearly consider the full spectrum of opportunities, risks and possible consequences, can identify preferable alternatives and risk mitigation strategies, thus lowering the opportunity cost of not making decisions or choices due to lack of clarity or for fear of a negative reaction (Taddeo and Floridi, 2018). However, it is important to move beyond ethical frameworks or limited interpretations of fairness, accountability and transparency, addressing regulatory aspects and the data-driven business model in a holistic way (Cath, 2018).

Thus, these ethical frameworks can represent a guide as practical guidelines for the chain of AI design, development and implementation, especially when used within the public apparatus. In terms of preventing and remedying human rights violations, these are mechanisms for prospective and retrospective responsibilities (Yeung, 2019). The content of this transposition to a regulatory framework can be formulated in terms of obligations at different levels and stages, being the requirement to carry out impact assessments on human rights prior to the implementation of certain technologies a specific example of a proposal favorable to said transposition.

4.2. Reinforcement of human rights at the international level

It seems pointless to carry out an analysis of AI governance from a human rights perspective by neglecting power relations, particularly in an area where this concentration is so evident. The relationship between AI, democracy and human rights requires a holistic view of the reality of technology and business models, including the accumulation of technological, economic and political power in a few companies, for AI development in commercial services



(Nemitz, 2018). It is not simply about the operation of the human rights framework to face additional risks derived from this economic structure, but to understand it as an aspect that also requires a change in AI governance proposals that takes responsibility for the deepening development and control gaps from the perspective of constitutionalism (Gurumurthy and Chami, 2019) and fundamental rights.

As Yeung (2019) expounds, the power and sophistication of advanced digital network technologies can be taken as a basis to understand that these are radically different technologies from others, especially due to their consequences on social structures. This makes it necessary to strengthen the human rights discourse and instruments, assessing the collective responsibility of attending to the sociotechnical foundations of moral and democratic freedom, which are necessary for the demand and materialization of human rights and fundamental freedoms (Yeung, 2019). Human rights must be developed to protect humanity against the results of robotics and AI, strengthening the notion of being human and of human value as something intrinsic (Liu and Zawieska, 2017).

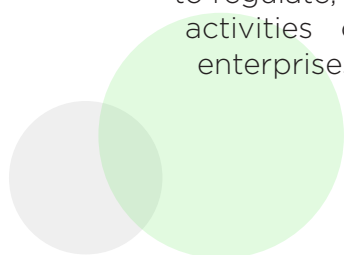
Part of this reinforcement consists in the respect for the set of international norms in force on human rights being shown by private agents, such as the dominant companies to which we refer. For Donahoe and Metzger (2019), a sufficient framework is provided by the sum of the United Nations Universal Declaration of Human Rights and the Guiding Principles on Business and Human Rights, both for governance and for evaluating the impact of AI decisions on people and society. According to the Guiding Principles, the State's obligation to protect human rights is compounded by the responsibility to respect and protect human rights, and to remedy their effects in relation to their products, services and operations. Under the Guiding Principles, companies should carry out human rights impact assessments, consult with stakeholders, and publicize their due diligence processes. We share with Donahoe and Metzger (2019) the notion of the usefulness of these standards and principles for the

task of delimiting responsibilities between governments and companies on the effects of technologies. However, there are intrinsic limitations to this approach due to the absence of normative force for its observance.

Therefore, reinforcement also implies advancing in the recognition (outlined above) of the concrete considerations on fundamental rights, adding to the analysis instruments such as the ACoHR, the COAS, the International Covenant on Civil and Political Rights (ICCPR) and the International Covenant on Economic, Social and Cultural Rights just as a starting point and without ignoring other instruments, especially those that aim at the prohibition of discrimination and the protection of vulnerable individuals and groups. The nature of state action in complex technological systems deserves to be the object of substantive and procedural materialization at the national level, as we will mention later. At the same time, the states still must provide mechanism assurance for both prospective and retrospective liability, at the international level, for which the Guiding Principles are insufficient. Other more recent initiatives, such as UNESCO's⁶ ROAM-X framework, and the aforementioned report of the Council of Europe prepared by Yeung (2019), point in a more concrete direction of state action.

⁶ In the report "Steering AI for knowledge societies", available at: <https://unesdoc.unesco.org/ark:/4822d/pf0000372132>

As we have mentioned, this need for responsibility that should be asserted at the international level is a nascent imperative of the accumulated (and growing) power of some of these companies, both for their insistence on accumulating information, for the absence of transparency regarding the technology they use, and due to the systematic failure of national legal frameworks with the permission, resignation or indifference of national governments. Much of this concern can be addressed through a new international treaty, as has been discussed since the creation by the United Nations Human Rights Council of an Intergovernmental Working Group for the formulation of a “legally binding instrument to regulate, in the framework of international human rights law, the activities of transnational corporations and other business enterprises”⁷, with the latest draft being released in July of 2019⁸.



4.3. Regulatory frameworks and implementation

At the Latin American level, several governments of countries in the region have published documentation or at least have made announcements to strategically guide the development and application of artificial intelligence⁹. The experience of digital agendas also shows that national priorities tend to change significantly with each change of government, reducing the probability of maintaining sustained state policies on technology and of coordinating collaboration with neighboring countries.

In this regard, and despite the apparent lack of materialization or assessment of its results, it seems appropriate to agree on a vision at the national or regional level that integrates technological development with economic and human development, according to the aspirations of each nation while understanding their peculiarities and the regional and global context. That is, assuming the role of politics in the field of technological development instead of separating it from said politics. For Cath et al. (2017), by comparatively analyzing documentation on the vision of the future of AI in the United States, the European Union and the United Kingdom, a global political vision and a long-term strategy for the development of a “good society of AI” is required. The question arises whether there are visions of that nature, and what legitimacy would they have, in Latin American countries.

⁷ Human Rights Council, Resolution 26/9 on the preparation of a legally binding international instrument for transnational corporations and other companies with regard to human rights (A/HRC/RES/26/9), 14th July 2014. Available at: http://ap.ohchr.org/documents/dpage_easpx?si=A/HRC/RES/26/9

⁸ Draft is available at: https://www.ohchr.org/Documents/HRBodies/HR-Council/WG-TransCorp/OEIGWG_RevisedDraft_LBI.pdf

⁹ See in this regard Aguerre (2020).

Regardless of the political agreement on a vision of society, the local implementation of rules that give normative content to the need for prospective and retrospective responsibility for AI systems is a complex task. State regulation, if it exists, is part of the public policy sphere in processes that cannot be theorized under a single model given the variations in diverse political and socioeconomic contexts (Evans and Cvitanovic, 2018). However, there are stages that are not necessarily linear or successive but emerge as opportunities for intervention, such as: problem formulation, policy formulation, decision-making on instruments and resources, policy implementation, and the monitoring and assessment of its performance (Evans and Cvitanovic, 2018). Such instances appear both with regard to comprehensive technology policies and the implementation measures they contain.

Implementation in the regulatory field, by the way, is a challenging task. The Council of Europe, in mid-2019, formed an Ad-Hoc Committee on Artificial Intelligence (CAHAI), the purpose of which is to examine the feasibility and potential elements of a legal framework for the development, design and application of artificial intelligence, on the basis of broad multi-sectoral consultations, based on the Council of Europe standards on human rights, democracy and the rule of law¹⁰. CAHAI's mandate extends until the end of 2021, with the duty to produce a feasibility study in the terms described. Although this result is uncertain, being a process initiated within the structure and function of the Council of Europe, it is worthy of attention as an example that other countries, including those in Latin America, may use to explore regulatory alternatives.

Procedural aspects and prospective responsibility

According to Wagner (2018b), regarding requirements for technology design and development within ethics or human rights, the following criteria must be met at the very least:

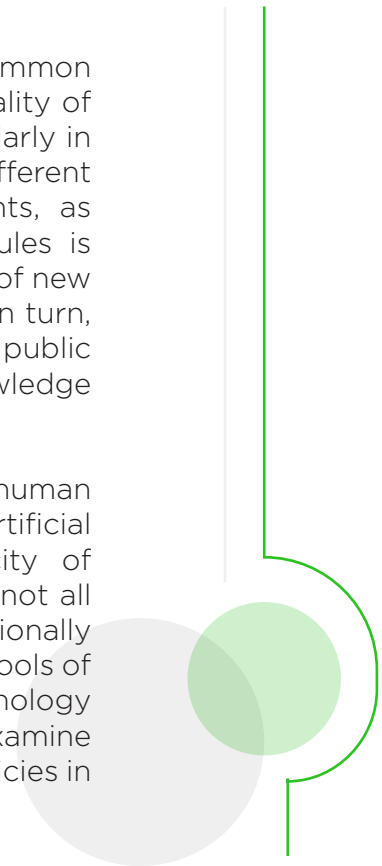
1. External participation: early involvement with relevant stakeholders;
2. Independent external oversight mechanisms (not necessarily public);
3. Transparent justification procedures on why decisions were made;
4. List of non-arbitrary standards where the selection of certain values, ethical principles and rights over others may be plausibly justified;
5. Ethical principles that do not replace human or fundamental rights;
6. Clear statement on the relationship between commitments and existing legal or regulatory frameworks, especially when in conflict.

¹⁰ *Terms of Reference for the Ad-Hoc Committee on Artificial Intelligence (CAHAI), approved by the Committee of Ministers of the Council of Europe in a decision adopted on September 11th, 2019 at the meeting of the Delegates of Ministers No. 1353.*

Such aspects seem valuable also in Latin America. It is a key, common characteristic in our region that, despite efforts to improve the quality of regulation in general terms, there are gaps to be overcome particularly in relation to the absence of significant participation from different stakeholders, and the absence of regulatory impact assessments, as pointed out by Querbach and Arndt (2017). The need for AI rules is perhaps an area that emerges as an opportunity for the application of new and effective forms of multi-sector involvement and participation. In turn, such involvement must be broad: deliberation, the formation of public policies and AI regulation must be informed by a multiplicity of knowledge disciplines (Marda, 2018).

Latonero (2019) provides a comprehensive guide to ethical and human rights considerations in the development and implementation of artificial intelligence, with special regard being given to the multiplicity of stakeholders and their roles. However, it is necessary to state that not all the same agents are an equally active part in Latin America, traditionally excluded from the design and programming chain of technological tools of massive application, and currently without significant AI technology production or exports. Therefore, it is also necessary to examine interdisciplinarity more deeply as a source of evidence for public policies in our contexts.

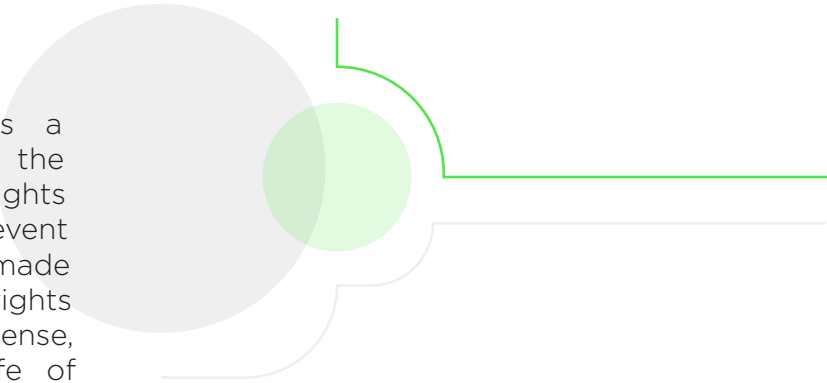
There are several efforts in the literature to put existing and mentioned ethical frameworks into practice. An example is provided by Mantelero (2018), who suggests a model for a human rights-friendly artificial intelligence deployment focused on controllers or data controllers. Mantelero proposes the prevention model of Human Rights, Ethics and Society Impact Assessment (HRESIA), so that controllers can assess values in the design of products and services, and the possibility of establishing a committee if required to judge the assessment. This contrasts, in turn, with the impact assessment on personal data required by the GDPR.



The impact analysis mechanism is a mechanism that best makes use of the framework of international human rights law. Given that states have a duty to prevent harmful impacts, the definition can be made on the basis of the body of human rights norms, in a broad and comprehensive sense, and in relation to all stages of life of algorithmic systems (McGregor et al., 2019). The incorporation of the characteristics listed by Wagner (2018b), which have already been reviewed, may focus on that range of interests.

Some of the aspects of prospective responsibility require additional detention by virtue of what they mean for the development of AI systems. By way of illustration rather than limitation, for being individually the object of academic discussions and the search for technical solutions, we can highlight three. Even though the literature tends to examine them from the perspective of bias and discrimination, the range of possible impacts is greater.

The first of the major aspects is the transparency of algorithmic systems in design, formulation, development, internal logic of operation, and implementation, in this way creating, according to Pasquale (2015), "black boxes" which are difficult to scrutinize and question. This may refer to the technical complexity of their operation, to the lack of disclosure of their code, which is protected by intellectual property rules (Noto La Diega, 2018), and to their being difficult to understand by people without training in technology, ethics or law. It serves here as a guide to induce effective regulation on this point by the GDPR and the duty to inform, demanding that the delivery of information to the owners of personal data who request it should be delivered in a "concise, transparent, intelligible and easily accessible manner, in a clear and simple language", allowing them to take action on decisions or exercise their rights of rectification or cancellation¹¹.



Human intervention is another significant area, where compliance with integrating more human decisions is not enough to correct biases or prevent damage ("human in the loop"). Pasquale (2015) emphasizes the importance of a sense of human agency in decision-making processes and systems, highlighting multiple examples of harmful impacts usually attributed to algorithms. The arguments for a rejection of fully automated decisions are well summarized by Noto La Diega (2018). But the notion that integrating human decisions into processes operated primarily by machines constitutes sufficient protection of human rights has been questioned, and may even have a negative impact where it constitutes rather a formal gesture of humanization of decisions (Wagner 2018b). If we understand that there is a continuous flow among people's biases due to their historical material circumstances and the biases incorporated into decision systems, a greater interaction that detects the effects of those biases does not alter their existence. Thus, the level of effective agency that human intervention has in automated decisions is a relevant factor in fulfilling a damage prevention role and, ideally, in promoting a framework of fundamental rights in the implementation of AI systems (Wagner, 2019).

¹¹ For a more complete discussion, see also Kaminski (2019).

Linked to the above, another relevant aspect is related to participating in the formulation of technologies, both with regard to the representativeness of the information sources used by the automated learning algorithms (that is, the biases of their databases), as - more relevantly - with regard to the existence of inclusive and participatory processes in its development. In a sense analogous to that of public policy processes (highlighting the regulatory nature in AI application), the intervention of specialists and possibly affected individuals and communities is necessary.

In short, converting guides and guidelines into compliance lists is not a satisfactory approach. In addition to strict formal compliance with rules and principles, it is necessary to incorporate ethics sensitively to specific situations with operational guidelines for the decision-making process. Likewise, while recognizing the limitations of external regulation, it is necessary for technical implementation to fully accept values that seem abstract, which is an aspect of a formative and organizational nature that exceeds the purpose of this document¹².

Updating the relevant substantive regulatory framework and retrospective liability

The need for human rights-based regulation for the governance of artificial intelligence is primarily aimed at establishing an accountability framework by setting out the mechanisms that such regulation must address in relation to the processes that range from formulation to implementation and assessment. However, it is pointless to focus on the process without committing to the substance of fundamental rights, their limitations, their mechanisms of complaint, their forms of affectation (by means of defined acts or those responding to certain characteristics), the facilitation of collective search for responsibility, and the assurance of the possibility of acting against companies and state entities.

In a general sense, the damage classification, and consequently the object of protection subject to risk, must be considered in a comprehensive manner by system operators. We have mentioned in section 2.2. that efforts exist in the literature to list potentially affected human rights; however, their range is broader than that represented by usual concerns about individual involvement related to bias or discrimination. As we have mentioned before, considerations on democracy and self-determination, collective participation, and the environment are also included. A broader conception of damage is provided by international human rights law, as noted by McGregor et al. (2019), by allowing a concrete and universally applicable definition of damage accompanied by sophisticated and developed tests, including what constitutes direct, indirect and intersectional discrimination, as well as structural and unconscious biases, and extending the analysis to the impact on the rights of individuals and also of groups. This results in consequences such as the possible

¹² In this regard, it is worth mentioning the need to integrate analytical and technical tools for different potentially affected communities, not only in education but also for workers. This also includes the ability to serve as a counterweight, which requires organizational capacity and recognition of their interests as groups.



prohibition of the use of certain algorithms, where regardless of bias issues there may be significant impacts on human rights (McGregor et al., 2019).

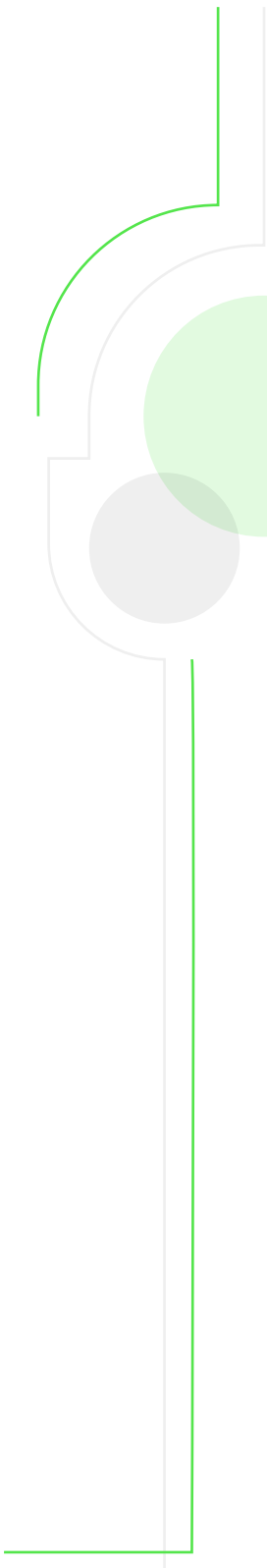
As we have already pointed out, the roles and responsibilities of states and companies, outlined by international human rights law, require substantive and operational implementation at the local level. Since states are primarily responsible for the effective protection of human rights, they also have an obligation to introduce legislative frameworks recognizing the individual and collective rights of people in their territories, as well as the legal obligations and duties of non-state entities (Yeung, 2019).

Another key area of regulation lies in the ability to exercise the rights related to personal data protection, especially where it stands as a complementary regulatory structure for the exercise of fundamental rights. This requires, in part, that such protection frameworks exist as is already the case in much of the world (Greenleaf, 2019), in addition to filling cooperation and coordination gaps (Aguerre, 2019). It also requires the provision of reforms that provide content and functionality to their rights in a context of algorithms and AI, with the GDPR as a common example of regulation with substantive and operational aspects.

On the same subject, with particular pragmatism, the Information Commissioner's Office of the United Kingdom provides practical recommendations about "Big Data" analysis that can be extended in general to information processing technologies in the public and private sectors (Information Commissioner's Office, 2017). The different

stages of an information privacy impact assessment arise in light of compliance with the regulations in force from the implementation of the GDPR in 2018, when the United Kingdom was part of the European Union. The lesson for Latin America on this is twofold: in addition to the need for a strong framework for the protection of personal data against the possibility of potentially harmful exploitation or use, the assessment of possible risks to information privacy should be an aspect necessarily embedded in decisions to implement and use advanced data processing tools. That is, both a finished protection framework (related to the GDPR), and an implementation guided by public control authorities that are useful for states and companies. It is necessary to recognize, in the Latin American context, the challenge posed by the plurality of different rules on the matter, the challenges posed by productive mechanisms of observance and supervision, the furtherance of compliance and responsibility programs, the cooperation between national authorities and the interoperability of legal frameworks, among other issues (Aguerre, 2019).

Furthermore, the need for regulation implies political decision-making on the consequences of criminal and civil liability, which are matters worthy of deeper exploration, separate from this research. In this regard, Yeung (2019) examines models of civil liability according to the classification between strict liability, conditional liability and immunity. Scherer (2016) proposes further liability reduction mechanisms by certifying AI development that assumes responsibility for prospective liability measures, especially regarding unanticipated



effects. At variance with this, McGregor et al. (2019) estimate that unpredictability, being known as such by the agent who decides to use an algorithm, does not allow for the making of statements of reduced responsibility in the abstract in accordance with international human rights law.

Finally, the establishment and refinement of judicial claim mechanisms is crucial. In addition to constituting a form of seeking remedy against private companies, it is a mechanism for claiming the state's own responsibility in preventing damage. International human rights law requires the state to take measures to prevent damage recurrence and to deliver reparations, in addition to asserting individual responsibilities. Therefore, it is critical that both states and companies provide complaint mechanisms which address the information asymmetry between the parties in order to facilitate the questioning of an algorithmic process or result (McGregor, 2019). Likewise, the existence of entities or agencies with associated complaint procedures can serve as an efficient and specialized mechanism (Scherer, 2016). In the same sense, litigation against state action based on algorithmic and AI decisions must be ensured¹³. In general, the liability aspect for AI and algorithm impacts, in the sense of the need for accountability, requires further exploration.

In summary, a normative update appropriate to the digital age, to the situation of global capitalism and to the context of developing and less advanced countries, could use international human rights law as a substantive and procedural basis. In addition to defining the state obligations that can be materialized internally, it provides guidelines for defining and examining the damages and risks that automated systems can create for individuals and communities. It also enables prevention through impact analysis based on these definitions for the entire cycle of AI formulation and use, and to build liability schemes for the materialization of these damages. And it crucially allows said analysis to be used in turn to formulate policies that address the materialization of individual and collective rights, providing technology with a role that reinforces fundamental rights.

¹³ As an example, the *AI Now Institute (2018)* focuses on litigation against public decisions delegated to algorithmic systems

5 Conclusions

The problem derived from the potential risks of algorithmic decision-making and the use of artificial intelligence, especially in areas of application in the public apparatus, has been discussed by governments, international institutions, academia, the technical community and civil society through the proposal of different governance frameworks and the formulation of ethical principles and guidelines. These proposals, which have proliferated in recent years, variably address the chain of development and operation of this family of technologies with different levels of detail and specificity.

These ethical frameworks have been widely criticized for being subservient to the current situation of concentrated economic and technological power, both due to the origins of the proposals and the exclusion of considerations belonging to the interests of people from less developed countries. Likewise, their weakness consists in not having frameworks for practical observance and accountability, as well as serving as a source to avoid discussion about effective state and international regulation. Even if they could be influential, the institutions and people dedicated to the study of ethics in technology are far from being effective agents of change.

However, the same frameworks, by virtue of their multiplicity and the novel attempts to systematize their content in order to make them operational, serve as an initial basis for action guided by society or its representatives in the State to incorporate fundamental rights considerations with normative force at different stages of AI development. This calls for a normative proposal to reinforce international human rights law, for coordination between states, and for the materialization of its content at the local level.

A proposal of regulatory basis is by no means a sufficient solution even if it is rooted in existing obligations under international human rights law. It is not enough to address risk prevention and remedy without also addressing the political economy, which is an important source of these threats and of the growing development gap among countries. However,

the international human rights law provides a framework to collect these other aspects, at least in part, from the perspective of the rights of collective exercise, including development interests and economic, social, cultural and environmental rights.

A proposal for regulatory basis could ideally focus on operational mechanisms of artificial intelligence governance backed by law and state action. That is, in the fundamental incorporation of aspects of human rights to regulate technology. However, in Latin American countries it is still necessary to ensure the existence of regulatory frameworks that are sufficiently protective of the substance of the wide range of fundamental rights, as is the case with civil liability rules and with the existence of rules for personal data protection in addition to technology policies sensitive to local contexts. Determining a technology governance framework that reflects a fundamental interest in self-determination and development is a political objective that should in turn inform regulatory action, taking into account public interest and community rights.

At the same time, an international effort is necessary to establish an effective counterweight to the large transnational companies that exploit data and technology. Regardless of the existence of international regulatory frameworks in the specific area of data processing, it is also necessary to strengthen mechanisms for forced observance at the international level, as is the case with the proposal for a new treaty on human and business rights. This would support a more robust margin of action than that currently provided solely by the norms of international human rights law and the United Nations Guiding Principles on Business and Human Rights.

Recognizing the diversity of tools that the regulatory state can use to reinforce the potentially positive aspects of AI and to prevent its risks, anchoring human rights can serve as a guide. Even if the tools may represent different levels of state regulation or self-regulation, national or international regulation, individualized or shared responsibilities, their value is maintained with the support of both the state apparatus and the international community. Incidentally, this requires conditions for the democratic exercise of power so as not to constitute an additional form of stripping people of control over their lives, individually and collectively. In the same sense, this requires the exercise of state power in a way that does not represent, as occurs in countless examples, a repressive attitude towards technology nor an instrumentalization of power to ensure and protect the political objectives of the authorities through the regulation of technology.

International human rights law, taking advantage of its own development for more than half a century and coupled with the latest technology-related considerations, provides a fundamental rights-based approach to address the formulation, design, development, deployment, application and assessment of decisions and AI systems. It serves as a guide to identify risk patterns and responsibilities among state and non-state agents, including transnational companies. And it does not exclude the existence of codes and guidelines based on ethics for the positive development of technological tools, especially where the regulation may be insufficiently detailed and by way of interpretation.

Since technological advance, its influence on the public apparatus and its consequences on the lives of individuals and communities are particularly fast-paced, the task of providing regulatory frameworks for new technologies that promote positive uses and prevent risks is an urgent task.

Bibliographic references

Agrawal, A., Gans, J. and Goldfarb, A. Prediction Machines: The simple economies of artificial intelligence, Harvard Business Review Press, Boston, 2018.

Aguerre, C. Digital trade in Latin America: mapping issues and approaches, on Digital Policy, Regulation and Governance, vol. 21 n. 1, 2019, pp. 2-18. <https://doi.org/10.1108/dprg-11-2018-0063>

Aguerre, C. Estrategias nacionales de IA y gobernanza de datos (en prensa) [National AI strategies and data governance (in print)]. Universidad de San Andrés, Buenos Aires, 2020.

AI Now Institute. Litigating algorithms: Challenging government use of algorithmic decision systems, 2018. Available at: <https://ainowinstitute.org/litigatingalgorithms-2019-us.pdf>

Anderson, K. and Waxman, M. C. Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can, Stanford University, The Hoover Institution (Jean Perkins Task Force on National Security and Law Essay Series); American University, WCL Research Paper 2013-11; Columbia Public Law Research Paper, 2013, pp. 13-351. <http://dx.doi.org/10.2139/ssrn.2250126>

Angwin, J., Larson J., Mattu, S. and Kirchner, L. Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks, ProPublica, 2016. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

Artificial Intelligence for Development. Government Artificial Intelligence Readiness Index, 2019. Available at: <https://ai4d.ai/index2019>.
Broussard, M. Artificial Unintelligence: How Computers Misunderstand the World. MIT Press, Cambridge, 2018.

Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlstrom, P., Henke, N. and Trench, M. Artificial Intelligence: the Next Digital Frontier? McKinsey Global Institute, 2017.

Calo, R. Artificial Intelligence Policy: A Primer and Roadmap, on Davis Law Review, University of California, vol. 51, 2017, pp. 399-435.
Campolo, A., Sanfilippo, M., Whittaker, M., and Crawford, K. AI Now 2017 Report. AI Now Institute, 2017. Available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf

Cath, C. Governing artificial intelligence: ethical, legal and technical opportunities and challenges, on Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, n. 376(2133), 2018. <http://doi.org/10.1098/rsta.2018.0080>

Cath, C., Wachter, S., Mittelstadt, B. D., Taddeo, M. and Floridi, L. Artificial Intelligence and the "Good Society": the US, EU, and UK approach, on Science and Engineering Ethics, 538(7625), 2017, pp. 1-24. <http://doi.org/10.1007/s11948-017-9901-7>

Citron, D. K. and Jurecic, Q. Platform Justice: content moderation at an inflection point, Hoover Institution, Aegis Series Paper No. 1811, 2018.

Couldry, N. and Mejías, U. Data colonialism: rethinking big data's relation to the contemporary subject, on Television and New Media, vol. 20 n. 4, 2019, pp. 336-349.

Crawford, K. Artificial Intelligence's White Guy Problem, on The New York Times, 26 June 2016. Available at: <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>.

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., Kak, A., Mathur, V., McElroy, E., Nill Sánchez, A., Raji, D., Rankin, J. L., Richardson, R., Schultz, J., Myers West, S., Whittaker, M. (2019). AI Now2019 Report. New York: AI Now Institute. Available at: https://ainowinstitute.org/AI_Now_2019_Report.html.

Daly, A., Hagendorff, T., Hui, L., Mann, M., Marda, V., Wagner, B., Wang, W. and Witteborn, S. Artificial Intelligence Governance and Ethics: Global Perspectives (Research Paper No. 2019-15), Chinese University of Hong Kong, Hong Kong, 2019.

Donahoe, E. and Metzger, M. M. Artificial Intelligence and Human rights, on *Journal of Democracy*, vol. 30 n. 2, 2019, pp. 115-126.

Elish, M. C. and Boyd, D. Situating methods in the magic of Big Data and AI, on *Communication Monographs* vol. 85 n. 1, 2017, pp. 57-80. <http://doi.org/10.1080/03637751.2017.1375130>

Eubanks, V. *Automating inequality: how high-tech tools profile, police, and punish the poor*, St. Martin's Press, New York, 2018.

Evans, M.C. and Cvitanovic, C. An introduction to achieving policy impact for early career researchers, on *Palgrave Common* vol. 4, 88, 2018. <http://doi.org/10.1057/s41599-018-0144-2>

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. C. and Srikumar, M. *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, 2020 [online]. Available at: <https://cyber.harvard.edu/publication/2020/principled-ai>

Gasser, U. and Almeida, V. A Layered Model for AI Governance, on *IEEE Internet Computing* vol. 21 n. 6, 2017, pp. 58-62. <http://doi.org/10.1109/mic.2017.4180835>

Greenleaf, G. *Global Data Privacy Laws 2019: 132 National Laws & Many Bills*, on *Privacy Laws & Business International Report* n. 157, 2019, pp. 14-18. Available at: <https://ssrn.com/abstract=3381593>

Gurumurthy, A. and Chami, N. *The Wicked Problem of AI Governance*, on *Artificial Intelligence in India*, vol. 2, Friedrich Ebert Stiftung, 2019. Available at: <http://library.fes.de/pdf-files/bueros/indien/15763.pdf>.

Hasselbalch, G. Making sense of data ethics. The powers behind the data ethics debate in European policymaking, on *Internet Policy Review*, vol. 8 n. 2, 2019. <http://doi.org/10.14763/2019.2H401>

Information Commissioner's Office. *Big data, artificial intelligence, machine learning and data protection*, 2017.

Jobin, A., Ienca, M. and Vayena, E. The global landscape of AI ethics guidelines, on *Nature Machine Intelligence*, vol. 1 n. 9, 2019, pp. 389-399. <http://doi.org/10.1038/s42256-019-0088-2>

Kaminski, M. E. The Right to Explanation, Explained, on *Berkeley Technology Law Journal*, vol. 34, 2019, pp. 189-218. <http://doi.org/10.15779/Z38TD9N83H>

Koene, A., Clifton, C., Hatada, Y., Webb, H., Patel, M., Machado, C., LaViolette, J. and Richardson, R. A governance framework for algorithmic accountability and transparency, European Parliamentary Research Service, 2019.

Latonero, M. Governing Artificial Intelligence: Upholding Human Rights and Human Dignity. Data & Society, 2019. Available at: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf.

Liu, H.-Y and Zawieska, K. From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence, on Ethics and Information Technology, vol. 19 n. 3, 2017. <http://doi.org/10.1007/s10676-017-9443-3>

Mantegna, M. OK Computer: ¿de qué hablamos cuando hablamos de IA?, [OK Computer: what do we talk about when we talk about AI?] on: I. Petrella, P. Marzocca and A. Beun (compilers), Ideas para la Argentina del 2030 [Ideas for Argentina in 2030], Área Cuatro, Buenos Aires, 2019.

Mantelero, A. AI and Big Data: A blueprint for a human rights, social and ethical impact assessment, on Computer Law & Security Review, vol. 34 n. 4, 2018, pp. 754-772. <http://doi.org/10.1016/j.clsr.2018.05.017>

Marda, V. Artificial intelligence policy in India: a framework for engaging the limits of data-driven decision-making, on Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133), 2018. <http://doi.org/10.1098/rsta.2018.0087>

McGregor, L., Murray, D. and Ng, V. International Human Rights Law as a framework for algorithmic accountability, on International and Comparative Law Quarterly, vol. 68 n. 2, 2019, pp. 309-343. Available at: <http://doi.org/10.1017/S0020589319000046>

McNamara, A., Smith, J. and Murphy-Hill, E. Does ACM's code of ethics change ethical decision making in software development?, Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE '18), 2018, pp. 729-733. <http://doi.org/10.1145/3236024.3264833>

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. The ethics of algorithms: Mapping the debate, on Big Data & Society, vol. 3 n. 2, 2016, pp. 1-21. <http://doi.org/10.1177/2053951716679679>

Morozov, E. Digital Socialism? The calculation debate in the age of big data, on New Left Review, n. 116/117, 2019, pp. 33-67.

Muñoz León, F. Epistemología de la techne: a propósito del fraude informático, on Revista Chilena de Derecho y Tecnología, [Epistemology of techne: on the subject of computer fraud, in Chilean Journal of Law and Technology] vol. 2 n. 2, 2013, pp. 247-260. <http://doi.org/10.5354/0719-2584.2013.30314>. Available at: <https://rchdt.uchile.cl/index.php/RCHDT/article/view/30314>

Nemitz, P. Constitutional democracy and technology in the age of artificial intelligence, on *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376, 2018. <http://doi.org/10.1098/rsta.2018.0089>

Noble, S.U. *Algorithms of oppression: how search engines reinforce racism*. New York University Press, New York, 2018.

Noto La Diega, G. Against the Dehumanisation of Decision-Making - Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information, on *JIPITEC*, vol. 9 n. 1, 2018, pp. 3-34.

O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Crown Publishers, New York, 2016.

Pasquale, F. *The Black Box Society: The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge, MA, 2015.

Querbach, T. and Arndt, C. Regulatory policy in Latin America: An analysis of the state of play OECD Regulatory Policy Working Papers, No. 7, OECD Publishing, Paris, 2017. <http://dx.doi.org/10.1787/2cb29d8c-en>

Raso, F., Hilligoss, H., Krishnamurthy, V., Bavitz, C. and Kim, L. *Artificial Intelligence & Human Rights: Opportunities and risks*, Berkman-Klein Center for Internet & Society, Harvard University, 2018.

Scherer, M. Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies, on *Harvard Journal of Law & Technology*, vol. 29 n. 2, 2016, pp. 353-400.

Selbst, A. D. and Barocas, S. The intuitive appeal of explainable machines, on *Fordham Law Review*, vol. 87, 2018, pp. 1085-1139.

Taddeo, M. and Floridi, L. How AI Can Be a Force for Good, on *Science*, n. 361 (6404), 2018, pp. 751-52.

UNCTAD. Informe sobre el comercio y el desarrollo, [Trade and Development Report] 2018 [online]. Available at: https://unctad.org/es/PublicationsLibrary/tdr2018_es.pdf.

Veale, M., Binns, R. and Edwards, L. Algorithms that remember: model inversion attacks and data protection law, on *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 376 (2133), 2018. <http://doi.org/10.1098/rsta.2018.0083>

Wagner, B. *Algorithms and Human Rights: Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications*, Council of Europe, Committee of Experts on Internet Intermediaries (MSI-NET), 2018a.

Wagner, B. Ethics as an Escape From Regulation: From 'Ethics-Washing' to Ethics-Shopping? on Bayamlioglu, E., Baraliuc, I., Janssens, L.A.W. and Hildebrandt, M. (editors), *Being Profiled: Cogitas Ergo Sum*, Amsterdam University Press, Amsterdam, 2018b, pp. 84-89).

<https://doi.org/10.2307/j.ctvhrd092.18>

Wagner, B. Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems, on *Policy & Internet*, vol. 11 n. 1, 2019, pp. 104-122. <http://doi.org/10.1002/poi3.198>

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S.M., Richardson, R., Schultz, J. and Schwartz, O. *AI Now Report 2018*. AI Now Institute, 2018. Available at:

https://ainowinstitute.org/AI_Now_2018_Report.pdf.

World Wide Web Foundation. *Artificial Intelligence: The Road Ahead in Low and Middle-Income Countries*, 2017.

Yeung, K. A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework, MSI-AUT (2018)05 rev. Council of Europe, Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence (MSI-AUT), 2019 [online]. Available at:

<https://rm.coe.int/a-study-of-the-implications-of-advanced-digital-technologies-including/168094ad40>.

Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, PublicAffairs, New York, 2019.

Disclaimer. The opinions expressed in the publication are the sole responsibility of the authors. Said opinions do not intend to reflect the opinions or perspectives of CETyS or any other organization involved in the project.