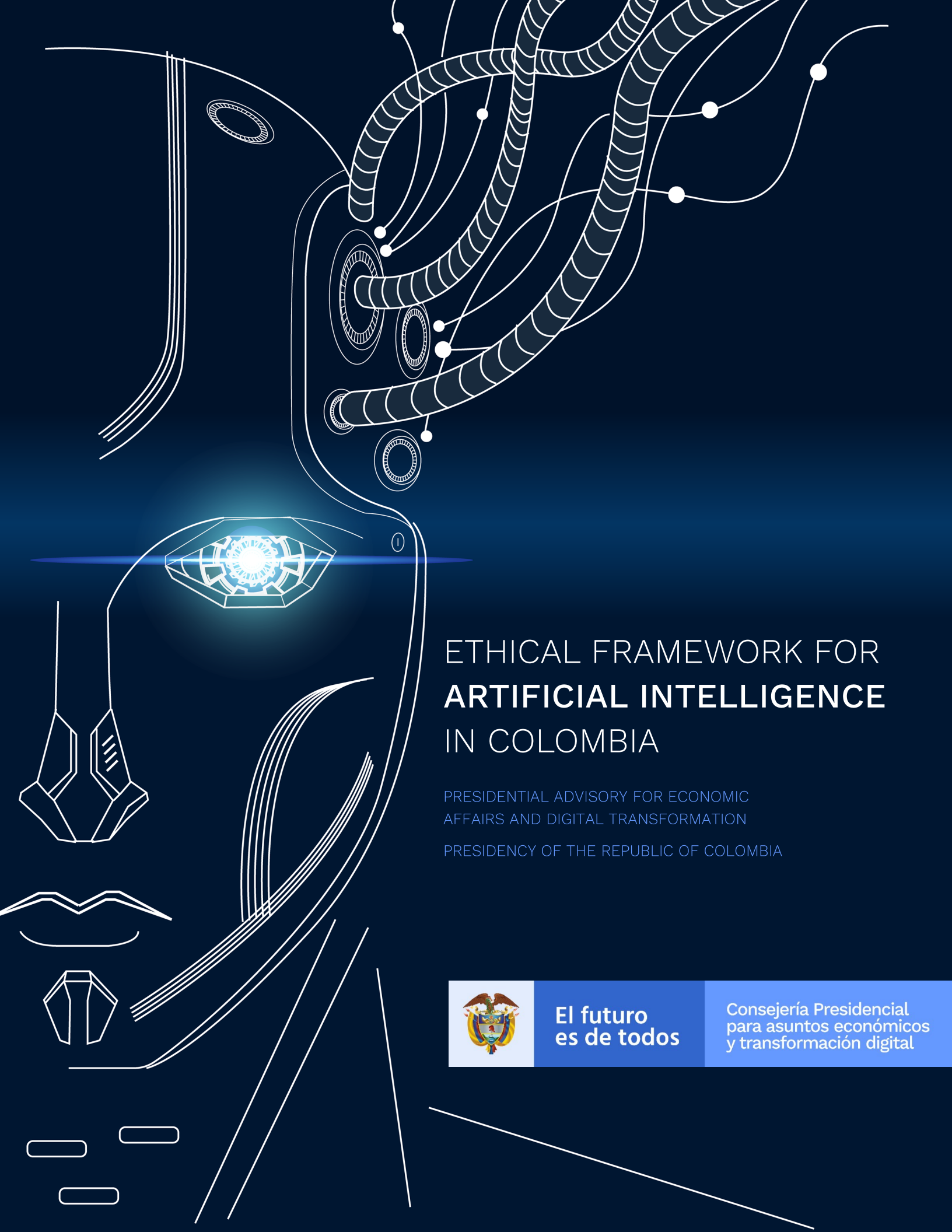


ETHICAL FRAMEWORK FOR **ARTIFICIAL INTELLIGENCE** IN COLOMBIA

PRESIDENTIAL ADVISORY FOR ECONOMIC AFFAIRS AND DIGITAL TRANSFORMATION

PRESIDENCY OF THE REPUBLIC OF COLOMBIA

El futuro es de todos

Consejería Presidencial para asuntos económicos y transformación digital

ETHICAL FRAMEWORK FOR **ARTIFICIAL INTELLIGENCE** IN COLOMBIA

PRESIDENTIAL ADVISORY FOR ECONOMIC AFFAIRS AND DIGITAL TRANSFORMATION

PRESIDENCY OF THE REPUBLIC OF COLOMBIA

Autor:

Armando Guío Español

Consultant of the CAF

Supervisors:

Víctor Manuel Muñoz Rodríguez

Presidential Advisor for Economic Affairs and Digital Transformation

Carlos Santiso

Director of State Digital Innovation at CAF

María Isabel Mejía Jaramillo

Senior Executive in Digital Government, Digital Innovation of the State in CAF

Date:

Agosto, 2020

Entities involved in the realization of the document

Acknowledgments

The vocation of this document is to serve as input to the urgent and necessary national conversation regarding the ethical framework for the development of artificial intelligence in Colombia. In November 2019, the Government of Colombia adopted a National Policy for Digital Transformation and Artificial Intelligence pertain in the document CONPES 3975. This document therefore seeks to nurture the national dialogue on the development of artificial intelligence within the framework of the national recovery. In this sense, it does not represent the official opinions of either the Government of Colombia or CAF Latin American Development Bank.

This document was prepared by Armando Guio Español, affiliate of the Berkman Klein Center for Internet & Society at Harvard University. The Directorate of Digital Innovation of the CAF Latin American Development Bank commissioned this document for the Presidential Council for Economic Affairs and Digital Transformation of the Presidency of the Republic of Colombia. It is part of CAF's agenda on the responsible use of artificial intelligence in the public sector in Latin America. It has been supervised by María Isabel Mejía with the support of Nathlie Gerbasie and reviewed by María Isabel Mejía, Victor Manuel Muñoz and Carlos Santiso.

CONTENIDO

The ethical discussion should start from an understanding of the principles and models that have been proposed at the global level, their impact on different stages of development of artificial intelligence, those that could be applied to the Colombian case and the existing implementation mechanisms.

I

INTRODUCTION

One of the fundamental tasks set by the Colombian Government is to create its own ethical framework for the development of Artificial Intelligence, applicable to both the public and private sectors, in the country. This implies the design of a transversal framework that can be appropriate by different actors, considering diversity of interests and opinions.

However, the discussion about ethics in Artificial Intelligence is not new and has been developing for several decades. Since the middle of the 20th century and even much earlier, philosophy has taken on the task of studying this subject in depth. Therefore, we should not ignore the history that surrounds this subject and the different efforts that have been generated to produce an understandable framework around the development of a technology that tests our understanding of reality.

This is why several elements arise for consideration in the construction of an ethical framework for artificial intelligence in Colombia:

1 Justification of this framework

One of the first questions that may arise when designing an ethical framework for artificial intelligence in Colombia is "why such a project should be considered in this country?". There are two fundamental elements that make this a necessary and prioritized exercise, beyond the fact that this is a global trend that several countries have been proposing.

Firstly, the effects that the implementation of artificial intelligence can have in the country are varied and there for have different ethical implications in the way this technology can be used. As we will see further on, international experience and collected worldwide evidence show that artificial intelligence systems used in various sectors such as justice, law enforcement and the financial market, among others, can lead to discriminatory and unfair practices with undesirable social implications. Furthermore, the use of this technology may conflict with several fundamental and human rights that Colombia has committed to respect and defend since the National Constitution of 91' and the ratified international treaties. This is why defined ethical limits are essential in the deployment of this technology in the country.

In addition to this, the current situation caused by COVID-19 means that automation and the use of artificial intelligence systems in various sectors may have accelerated the detriment of Colombian employments. This necessarily leads the country towards the discussion on how thousands of jobs can be replaced, the

position of human beings in the fourth industrial revolution and ultimately the role of the human beings in the implementation of these systems. This makes part of the role of human beings in the implementation of this technology and the degree of control they have over their decisions and actions.

Secondly, the ethics of artificial intelligence prove to be essential as a first step towards creating desirable social boundaries in the use of this technology. The ethical discussion provides a starting point and allows for a series of social consensuses on the use of this innovation, and which have less disruptive effects on innovation, as opposed to a regulatory framework. Therefore, it allows for a greater field of exploration of this technology, its future development and its effects, which will subsequently consolidate evidence to justify regulatory or other measures.

In this way, the ethical discussion permits the approach to this technology from a position of discussion and experimentation orientated towards the design of specific measures that materialize those ethical principles. However, the profound impact of the ethical discussion should not be underestimated, since it provides the fundamental elements on which we will surely proceed to design later on a regulation on the subject. That is why this ethical discussion is in the end is one withregulatory implications, since it seeks to analyze what will be the philosophical bases and the initial principles that guide the use of artificial intelligence in the country, consequently the regulations that will be established in this aspect.

2 State of technological development

The discussions on the ethical framework of artificial intelligence applicable to Colombia should focus on the current development of this technology, understanding its true functionalities and the different types of artificial intelligence that have been developed and are being applied. Presently, we have specific artificial intelligence systems limited to certain types of processing and tasks. Therefore, this is how there are Machine Learning and Deep learning systems, as two examples of the type of artificial intelligence that are available. It is on this state of development that a current discussion regarding the ethics of artificial intelligence must fall, this is how different authorities around the world have approached it.

Establish ethical discussions on whether artificial intelligence systems have the capacity to be conscious beings and our coexistence with this type of being is unaware that at currently we have not reached a state of development that responds more to fiction than to reality. What the literature has considered general artificial intelligence is not typical of the current technology that we find available. Discussing this type of artificial intelligence is not only tiresome, but can also divert us from the points that are currently more pressing, such as data processing or models that guide the current design of an algorithm and affect thousands of people today.

3 The different stages of Artificial Intelligence

The design of an ethical framework of these characteristics must consider the existence of different stages that comprehensively describe the development and deployment of an artificial intelligence system (*algorithmic chains*). Within this 'chain' we find four fundamental phases:

1. **Design.**
   This phase includes both the technological design of AI-based technologies (e.g. how the data is collected, the target audience for an AI tool) as well as the design of systems that govern AI and the implications to consider before moving to the next phase (AI and Inclusion Staff, s.f.).

2. **Development.**
   The development phase is the production phase of an autonomous system that follows the design process. Questions in the development category pertain to the incorporation of tools and methods into AI-based technologies, frameworks for AI development, and which AI tools are developed for whom (AI and Inclusion Staff, s.f.).

3. **Implementation.**
   The implementation stage includes the distribution, use, ubiquity and implementation of AI-based technologies within society at multiple levels, including local, national and global ecosystems (AI and Inclusion Staff, s.f.).

4. **Evaluation / Impact.**
   The evaluation / impact phase includes measuring and understanding the impact of AI technologies, including ways to assess the effects of autonomous systems on different actors within society (AI and Inclusion Staff, n.d.).

When addressing the ethical proposal of this framework and the proposed principles, reference can be made to any of these stages, which means that some implications circumscribe to some of these specific stages.

4 Inputs provided by the international scenario

Ethics in artificial intelligence have been developed around the design, development and implementation of this innovation, especially *machine learning* and *deep learning*. Therefore, it is desirable that the Colombian government also focuses its work on this type of technologies and on the principles applicable to these systems. However, the Colombian government does not have to start from scratch in this task. Not only would this prove to be inefficient and unnecessary, but alsounaware that for some years now a series of principles have been proposed worldwide to guide the development of this technology. Private entities, governments, NGOs and multilateral organizations have already been making proposals in this regard. Even within several of the national policies and strategies for artificial intelligence proposed since 2017, several countries have already dared to formulate a series of principles.

This does not mean that the Colombian Government should simply make a selection process and apply principles it considers most relevant. The challenge lies in selecting those that respond to the country's needs, giving them meaning and content, establishing the implications they would have for the country's own context, and the way in which they would materialize in Colombia's public and private sectors.

Therefore, the proposal is that the Colombian Government should be aware of the international principles and direct all its efforts to establish which of these will be applicable to the country, how should they be understood in the different stages of development of artificial intelligence and design specific mechanisms for their implementation and materialization. Nonetheless, this does not prevent Colombia from generating its own principles, perhaps never before seen in the international scenario. However, before generating this type of proposal it is important to have extensive knowledge of the inputs already provided by the international system, in order to justify the need to propose new ones. Furthermore, this was pointed out in the CONPES 3975 of 2019, which, as a roadmap for the design of this ethical framework, established a series of principles that must be taken into account and the need to make use of international experience on the subject.

5 A cross-cutting ethic

Since the Colombian Government wants the proposed ethical framework to apply to both the public and private sectors, it is necessary to consider the realities and particularities of each sector. It is understandable that the Colombian Government seeks to develop a general framework, given a criterion of equality and the desire that there should be no asymmetries in how these sectors approaches this type of technology. Hoping that this framework will serve as a general input that will later allow each entity within one of these sectors to develop its own ethical framework that follows a series of particularities and its own implementation mechanisms.

However, the main consequence of having a transversal framework is that it makes it necessary to start from general principles that can be applied to society. Consequently, it will point out a series of mechanisms to implement these principles without entering into specifics regarding the entities and persons in charge of this process or the methodologies to followed in each sector, given that this will depend on the logic of each entity.

6 A national conversation

Given all the above, the implications of this ethical discussion are varied and profound for Colombian society. Through this effort, guidelines are being defined that will influence the way a generation of Colombians will approach the most important innovation known to humanity in the last centuries. These principles will have a high impact on different public policies, regulations and other official documents as well as innovation guide within the Fourth Industrial Revolution. This discussion will become the basis of a State policy and therefore requires a prior discussion with different sectors of society that will analyze the principles proposed by the Government and their implications.

That is why the final construction of this framework must be a democratic and inclusive exercise involving different actors at national and international levels. **The vocation of this document is to serve as an initial input in the starting point of the discussion that takes place in an organized manner and with clear criteria to guide this conversation.** A debate that does not have a roadmap with some concrete discussion elements will hardly generate a precise product. For this reason, this input is necessary, since it also gives further development to the

ethical issues already mentioned in the CONPES 3975 document of 2019, and previously highlighted by international entities such as the OECD or the Inter-American Development Bank. This last entity even pointed out in its 2020 report:

> "The Government AI Readiness Index 2019 , produced with the support of Oxford Insights and the International Development Research Centre (IDRC), shows that countries in the region face three challenges when it comes to harnessing the use of AI for the common good: adequate policies, capacity and resources. First, to date LAC lacks a coherent policy approach and defined ethical standards. Mexico, Colombia, Uruguay and Argentina are currently setting up AI policies and strategies. Colombia, for example, through the document CONPES 3975 defined its National Policy for Digital Transformation and Artificial Intelligence. There, concrete guidelines are identified that, through their implementation, will generate a coherent policy framework for the ethical and responsible development of AI".
>
> (Cabrol, González, Pombo, & Sanchez, 2020).

II

WHAT IS THE ETHICS OF ARTIFICIAL INTELLIGENCE?

The ethics of artificial intelligence has been seen as a branch of ethics that analyzes and evaluates the moral dilemmas that arise from the deployment of this technology in society. For centuries, the idea of conscious and thinking beings has been used by philosophy to analyze characteristics of human beings and whether these are transferable to other objects. Thus, concepts such as consciousness and its importance in this branch are developed, taking into account that through our individual consciousness we take knowledge of our most deeply rooted moral principles, we motivate ourselves to act accordingly and we evaluate our character, our behavior and ultimately ourselves, according to these principles (Giubilini, 2016).

However, in recent years a new approach to ethics of artificial intelligence has derived, based on the *Data Ethics*. This ethics focuses on the use and analysis of data and the various systems and innovations that interact with this information. Given the importance that data analysis and the rise of *big data has had in* recent years this is one of the most predominant ethics and many of the proposals and principles are derived from this study.

As Floridi and Taddeo assert, this ethic is based on the foundation provided by the ethics of computing and information technology, but at the same time, it perfects the perspective supported so far in this field of research, by changing the level of abstraction of ethical questions from being information-centered to data centered (Floridi & Taddeo, 2016).

As Floridi himself points out, in this case we have a transformation in perspective due to a change in the levels of abstraction (LoA). This means that the focus shifts from a more abstract and general definition of information and the various ethical dilemmas that can be derived from these concepts, to a data-centered approach, especially those used for the development and implementation of artificial intelligence systems (Floridi & Taddeo, 2016).

Consequently, a definition of data ethics is proposed which divides this discipline into three criteria for analysis: a new branch of ethics which studies and evaluates moral problems related to data (including their generation, recording, adaptation, processing, dissemination, and use), algorithms (including AI, artificial agents, *machine learning, and* robots), and corresponding practices (including responsible innovation, programming, *hacking,* and professional codes), in order to formulate and support morally good solutions (e.g., appropriate codes or correct values) (Floridi & Taddeo, 2016).

This also means that the focus of the ethical study must be on the computer systems and operations in which these data are used, rather than on the variety of digital technologies that facilitate them (Floridi & Taddeo, 2016). This ethic is valuable because it allows the development of good practices and behaviors considered morally good to address ethical dilemmas raised by the collection and analysis of large databases. This considers dilemmas ranging from the use of *big data* in biomedical research and in the social sciences, to profiling, publicity and philanthropy of data, as well as *open data* (Floridi & Taddeo, 2016).

Given its practical sense and the fact that it allows for the discussion of concrete ethical judgments and in the face of current technological developments, the proposal of a transversal framework by the Colombian Government should focus on this ethics. **The ethics of the data will be analyzed throughout this document as the ethics applicable to the artificial intelligence that is designed, implemented and developed in Colombia.**

The way this ethic is divided also will be taken into account when analyzing each of the proposed ethical principles, as well as determining how they relate to data ethics, algorithms and practices. This will allow the development of a comprehensive view of each of the proposed principles.

III

MAIN ETHICAL CHALLENGES AND UNDESIRED EFFECTS OF THE ETHICS OF ARTIFICIAL INTELLIGENCE

As noted earlier in this document, **the implementation of artificial intelligence has posed a number of ethical challenges, as well concerns about the impact it may have in some cases.**

As the French government pointed out, there are the following specific challenges to be addressed: (i) possible threats to freedom of will and responsibility; (ii) bias, discrimination and exclusion; (iii) algorithmic profiling: personalization versus collective benefits; (iv) seeking a new balance by preventing massive databases while increasing AI; (v) quality, quantity and relevance: the challenge of adapted data for AI, and (vi) human identity versus the challenge of artificial intelligence (CNIL, 2018).

According to the Web Foundation, the main negative effects that an artificial intelligence system can have are divided into *Harm and Algorithmic Discrimination.* Concerning algorithmic damage, the Web Foundation has explained that the values of each society underlie the definition of damage. By defining what an algorithm should not do (damage), solid limits emerge for what an algorithm's optimization function should be (broad objectives). Ensuring that algorithms are compatible with the diversity of values that exist around the world is certainly a challenge. Who should define and determine whether algorithms have produced damage? In what cases should we promote that those who may have been affected by an algorithm are integrated into the design process? (World Wide Web Foundation, 2017). The Web Foundation takes the definition of damage from the legal sphere, which defines damage as setbacks that are also considered wrong (World Wide Web Foundation, 2017). The above-mentioned risks are addressed from the concept of legitimacy that has been proposed by different international entities, such as the Web Foundation.

In turn, discrimination can occur in two ways (World Wide Web Foundation, 2017). Two people can be equal in relevant aspects, but be treated differently, or relevant differences between them are not recognized or taken into account and the two people are treated the same (World Wide Web Foundation, 2017). In the second scenario, by not taking into account these relevant details, the result is unfair and consequently wrong (World Wide Web Foundation, 2017). In this way, a person can reasonably expect an outcome that is unfairly impeded by an algorithm, constituting a harm (World Wide Web Foundation, 2017). High- and low-income countries face the same categories of damage and threats from algorithmic decision-making (World Wide Web Foundation, 2017).

However, the impact of these damages can be widely different, depending on existing legal protections and *accountability* mechanisms implemented, especially for marginalized groups (World Wide Web Foundation, 2017). In some countries, algorithmic discrimination and inaccurate predictions can result in unwanted advertising or other inconveniences to consumers' experiences, but for marginalized groups in fragile contexts, it is argued that algorithmic discrimination can lead to unchecked assaults and even fatal exclusions from public services and resources (World Wide Web Foundation, 2017).

As it can be seen, most of the concerns are about the discrimination that these systems can generate, such as deepening inequalities and the possibility of having systems that make decisions automatically, without control and guided by the prejudices and discrimination that have been established in their design or in the data they use.

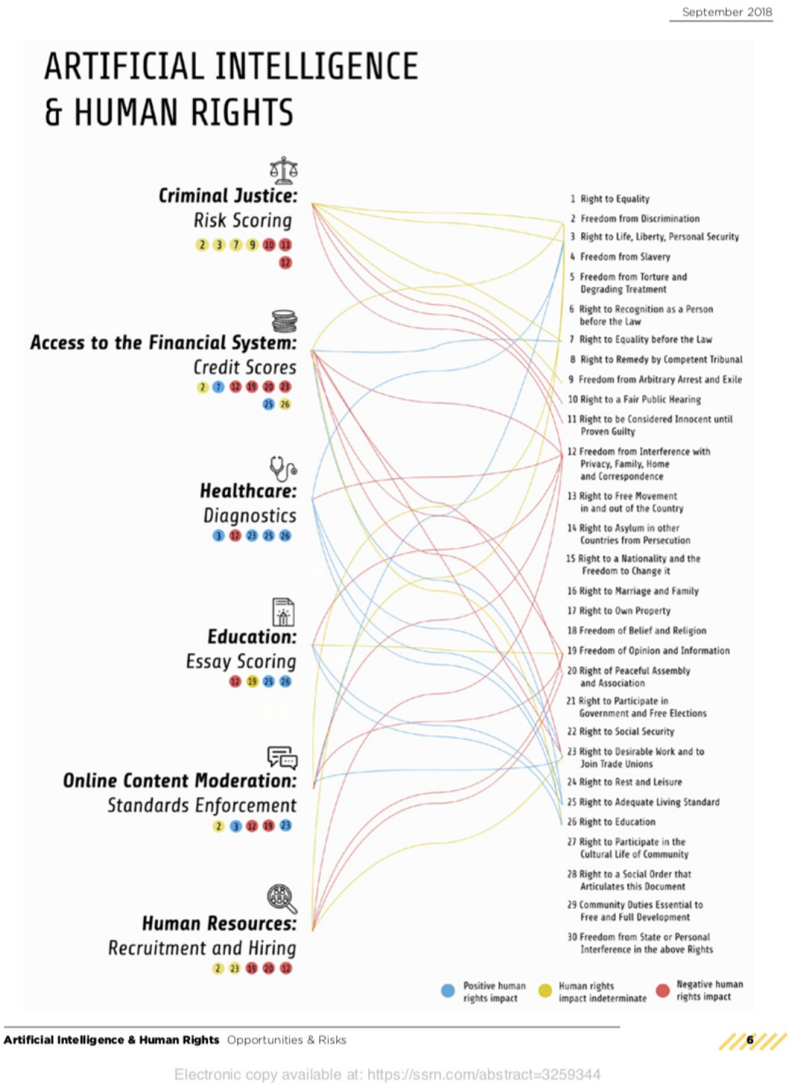
An example of this has been the **facial recognition systems** that use this technology. There is ample evidence that this technology causes serious harm, often to people of color and poor people (Crawford, et al., 2019). Consequently, there must be a suspension of all uses of facial recognition in the public and sensitive social domains, where this type of recognition has risks and consequences that cannot be remedied retroactively (Crawford, et al., 2019). Legislators should supplement this suspension with (1) transparency requirements that allow researchers, policy makers, and communities to understand and advise on how to best restrict and regulate facial recognition, and (2) protections that give the communities on which this technology is used the power to make their own assessments and reject implementation (Crawford, et al., 2019).

However, these risks may end up not only affecting society in general, but also the fundamental and specific human rights of each individual. Through a research of case studies, members of the Berkman Klein Center of Harvard University defined that the artificial intelligence systems that had been implemented in the United States in different sectors had had an impact on different human rights, such as freedom of expression, due process, right of association and equality, among others.

From their research, they were able to determine the following specific impacts of this technology on several of these rights:

- **Artificial Intelligence systems in the criminal justice system to determine an individual's level of risk and degree of recidivism:** according to the study these have a negative impact on the right to a fair public hearing, the right to be considered innocent until proven guilty and freedom from interference with privacy, family, housing and correspondence.
- **Artificial Intelligence systems in the financial system to determine an individual's credit risk:** according to the study these have a negative impact on freedom of interference with privacy, family, housing and correspondence, on freedom of opinion and information, on the right to freedom of association and on the right to desired work and unionization.

- **Artificial Intelligence systems in the health system for medical diagnosis:** they have a negative impact on the freedom of interference with privacy, family, housing and correspondence.
- **Artificial Intelligence systems in the education system for essay grading:** they have a negative impact on freedom from interference with privacy, family, housing and correspondence.
- **Artificial Intelligence systems to moderate online content and implement the proposed participation standards:** they have a negative impact on the freedom of interference with privacy, family, housing and correspondence and on the freedom of opinion and information.
- **Artificial Intelligence systems in human resources departments for recruitment and selection of candidates:** according to the study these have a negative impact on freedom of interference with privacy, family, housing and correspondence, on freedom of opinion and information and on the right to freedom of association.



(Raso, Hilligoss, Krishnamurthy, Bavitz, & Kim, 2018)

However, this does not mean that the technology *per se* has a negative impact on the human rights described. Many of the results obtained are related to the way the systems were implemented in each of these sectors and the way the information provided by this technology was interpreted. This study then provides relevant evidence about the potential risks that exist, but this should not lead to a generalization or consideration of absolute judgments, such as that the use of this technological innovation will always have negative effects on the health sector or the criminal system. It is clear that there may be sectors and practices in which the technology may carry greater risks, but these must be evaluated on a case-by-case basis and under particular contexts. The current state of evidence does not allow for generalizations in this regard.

As well, evidence has been presented in the Latin American region on how artificial intelligence can have undesirable effects. In this case, it is also worth highlighting the work of the Web Foundation that analyzed the use of artificial intelligence systems by the authorities of Argentina and Uruguay (World Wide Web Foundation, 2018).

The government of Salta Province, Argentina, implemented a system to predict teenage pregnancy and school dropout (World Wide Web Foundation, 2018). The case illustrates how a government with limited resources seeks to use technology to solve urgent social problems (World Wide Web Foundation, 2018). The government implemented a mechanism to coordinate the collection of data from 200,000 people living in vulnerable populations through NGOs and government officials, along with a *machine learning* model to generate predictions about school dropouts and teenage pregnancy among members of this population (World Wide Web Foundation, 2018). The implementation triggered a lot of interest from other governments and criticism from activists who considered that this violated privacy and did not solve the causes of the problem (World Wide Web Foundation, 2018). The implementation had transparent phases and others less transparent or opaque (World Wide Web Foundation, 2018).

Until the government collects and consolidates information on the impact of these tools, it is not possible to determine whether their implementation tends to yield fair or unfair results (World Wide Web Foundation, 2018). Women's rights activists have questioned the decision to implement such tools without a framework that incorporates them into a policy that addresses the structural inequity suffered by the populations they claim to support (World Wide Web Foundation, 2018).

In Uruguay, the government acquired Predpol, a police software to predict crimes (World Wide Web Foundation, 2018). It is a problematic case because of its low degree of transparency and the discrimination and exclusion dynamics that can be reinforced (World Wide Web Foundation, 2018). In less than three years the Ministry of the Interior discontinued the program and replaced it with retrospective statistical tools, developed by the ministry's own team, which were considered more useful (World Wide Web Foundation, 2018).

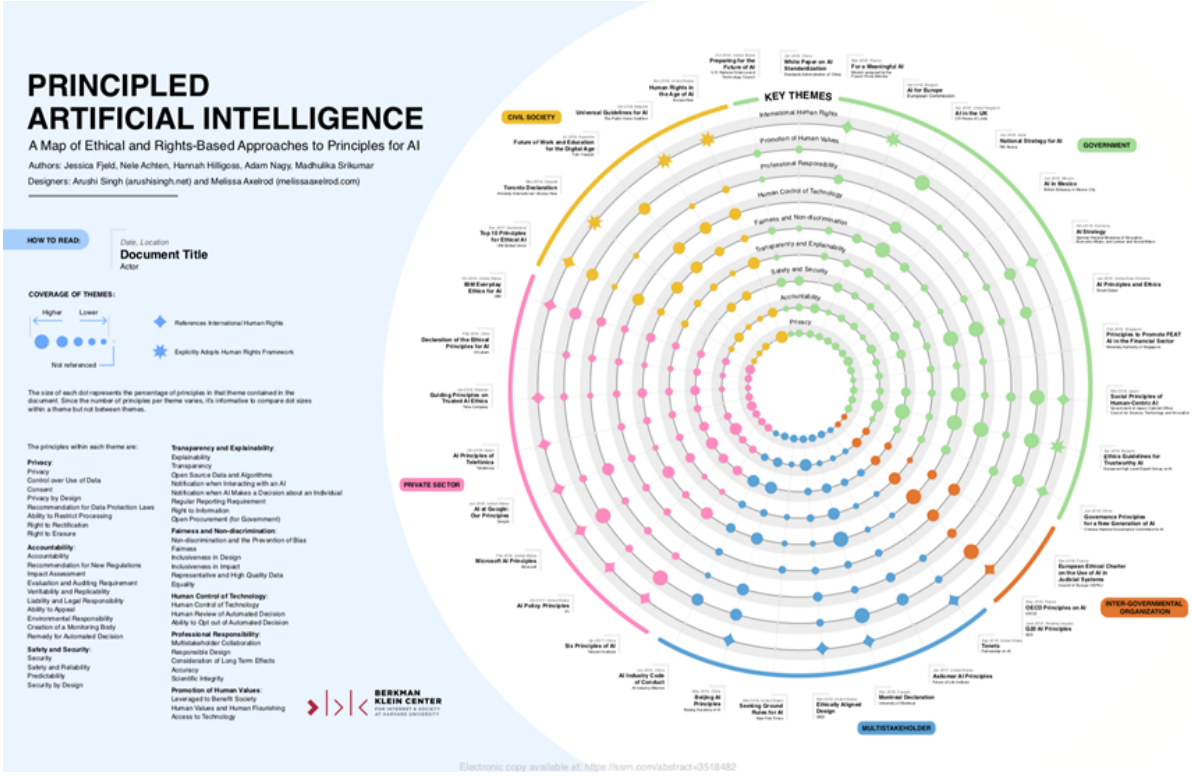
In this case, there is a risk of discrimination, and local and international organizations have argued that tools such as PredPol tend to replicate data training biases and historical power dynamics between law enforcement and minority or disadvantaged populations and are used to justify police presence in marginal areas (World Wide Web Foundation, 2018).

Colombia must clearly avoid the problems previously described even in other countries of the region and the ethical questions that have been generated in this regard. All of the above demonstrates the need for ethical principles to guide the design, implementation and use of this technology. These should be principles that address several of the specific problems already described and that have the possibility of changing and evolving according to the changes that the technology experiences. The implementation of these principles is also a priority task in order to mitigate several of the risks of the gradual use of this innovation in different sectors of society, while we learn more about how it works and obtain more evidence that allows us to consider and justify greater state intervention in this regard.

IV

THE PROPOSED ETHICAL PRINCIPLES FOR ARTIFICIAL INTELLIGENCE

As previously pointed out throughout this document, the ethical principles of artificial intelligence have been developing in recent years and therefore it is essential that the Colombian government, instead of presenting new principles, consider the existing ones and their impact on the country and its innovation ecosystem. Currently, even mapping these principles is no longer necessary since there are available inputs, such as the research conducted by the Berkman Klein Center published in January 2020, which presents a complete mapping of the principles of artificial intelligence proposed worldwide by various entities such as governments, international organizations, NGOs and private actors, among others. These principles and the implications of how they have been addressed at the global level are presented below:



(Fjeld, Achten, Hilligoss, Nagy, & Srikumar, 2020)

Several elements of analysis emerge from this study:

1 The prevailing principles at the international level

As it can be seen, the predominant principles on the world stage are those of privacy, responsibility, security, transparency and explanation, justice and non-discrimination, human control and supervision of technology, promotion of

human values and those outlined in the international human rights framework.

Privacy is a predominant principle that draws the attention of all the sectors analyzed in this study. This is explained by the large amount of data and information that is used today for the design, development and implementation of this technology. The data are not only fundamental for the operation of these systems, but also form part of practices known as artificial intelligence training. Training databases are currently common for the development of these systems, especially *deep learning systems*. The way in which these data and their processing can affect the privacy of individuals is a recurrent concern at the international level, a point that will be further explored.

In addition, security has become a cross-cutting concern considering the various attacks that have occurred to this technology, such as the *adversarial machine learning*, and the impact that an attack of this nature can have. This is not only a concern for governments, but also for private actors and social groups representing users and consumers.

One of the predominant principles that generates more discussion is the one catalogued as human control, given the implications that this principle can have in the development of this technology and the discussions that it can generate. This is of particular concern to governments and civil society that fear the escalation and deployment of artificial intelligence systems that have no control whatsoever and have greater autonomy in their decisions. On the contrary, the private sector may be more inclined to facilitate the autonomy of these systems and have clear limits to human intervention and control. As it will be seen below, this principle has different implications and it does not only mean that humans control the decisions and outcomes that the systems reflect, but that this technology should privilege human values and its design should reflect the characteristics of humanity itself, which can be somewhat complex and should be viewed with caution.

A principle closely linked to the ethics of practice is that of professional responsibility, which is particularly important and prevalent in the private sector and by *multi-stakeholder* bodies. For these entities, the responsibility of individuals behind the development of this technology is essential and consensus is sought.

2 Many of the concerns are around the use of the data

As noted earlier in this paper, the protection of privacy can be seen as the major concern on which there is almost a consensus among different sectors of society. The ability to profile through the use of data or make decisions based on poor quality personal information makes privacy a horizontal concern.

It is worth asking what the relationship is between this principle and data protection regulations and even many others that directly protect privacy. If regulations already exist, it would not be understandable that this is seen only as an ethical principle. However, privacy as a principle seeks to protect the privacy of individuals, but understands that this protection can go beyond the data and the way in which its protection has traditionally been conceptualized and regulated. Considering that existing legal frameworks must be adapted to this technology ignores that it has a disruptive effect that makes it necessary to rethink even these rules. Therefore, privacy as a principle seeks to ensure that such protection of privacy continues to be privileged, regardless of the legal model that ultimately results.

Thus, this principle questions the way in which artificial intelligence and its implementation can affect the privacy of a subject, not only from the point of view of the use of its data, but from the point of view of its fundamental freedoms, its capacity of decision and free self-determination.

However, this principle is predominant given the current characteristics of this technology. Nonetheless, the development of new algorithms may mean a reduced use of data that will completely transform our understanding of this technology. Therefore, the conceptualization of this ethical principle should not be intrinsically linked to data, and especially to personal data.

3 Governments have privileged transparency

Transparency has become a predominant principle within the proposals of governments, which is due to a more extensive phenomenon. In recent years, good

governance has been linked to transparent practices and transparent access to information. The transparency of artificial intelligence joins these efforts by different governments and allows access to more information about how public entities are making decisions and how they are using technology. Therefore, it is not strange to see that this is a predominant principle in governments and in the different strategies and official documents that Harvard University points out in its report. It is clear that the use of artificial intelligence for the design of public policy or in the provision of public services will be mediated by this principle.

4 The private sector seeks to define responsibilities in the use of these systems

For its part, the private sector seems to have a special interest in determining the *accountability* driven from the design and implementation of this technology. Given the legal, economic and social implications this may have, private actors have an interest in establishing a clear understanding of this principle and its limitations. Distributing the responsibilities that can be derived from the deployment of this technology can have profound economic effects on the design of your product, on the development of the business model and even on the contracting models to be used with your users. Therefore, it is clear that private actors want to influence the way this principle and its implications will be understood.

5 Civil society is concerned about the protection of fundamental rights and discrimination

The mapping carried out by Harvard University shows that several representative groups of civil society are interested in the application of the human rights framework and how this will be reflected in the ethics of artificial intelligence. This does not mean that it is the only sector that has this concern, but it does mean that it is one of the sectors that has given most relevance to the framework. Specifically, these concerns revolve around the justice and discrimination that can occur through this technology.

In addition, several civil society groups are emphasizing the need for greater human control of this technology and are concerned about how technology can displace humans in decision-making and in various productive activities. Concerns about discrimination are echoed by case studies carried out showing how different socially and historically marginalized groups have been affected by artificial intelligence systems. Reports from civil society representatives have placed special emphasis on this type of evidence.

Additional to this is the concern that governments may use technology to generate public policy and provide public services, which may be discriminatory against a specific group, even without this purpose. An example of this is the use of artificial intelligence in surveillance and crime-fighting systems that may have a laudable primary purpose, but may be discriminatory to a certain social group. The principle of non-discrimination seeks to limit these consequences in the deployment of this technology.

Thus, the analysis of the world mapping of principles applicable to artificial intelligence carried out by the Berkman Klein Center of Harvard University allows us to **identify interests and concerns that main sectors of society have in relation to the selection of these principles and their implications.** Reports, official documents and research generated by different actors on the subject allow us to affirm this. It would not be surprising if these concerns and interests also materialize in opinions of different actors that participate in the discussions on the ethical framework of artificial intelligence applicable in Colombia. Therefore, the principles proposed by the Colombian Government to energize this discussion must consider this context and the expectations that different social sectors have in this regard.

V

ADOPTION OF ETHICAL PRINCIPLES FOR ARTIFICIAL INTELLIGENCE IN COLOMBIA

Based on the mapping of principles described above and the way in which different sectors of society worldwide have approached the subject, the following principles are suggested as those that should guide the design, development, implementation and evaluation of artificial intelligence systems in Colombia. The principles are presented in a short and simple manner, considering that they are intended to be transversal and applicable to Colombian society in general. However, the effect and meaning of each of these principles must be understood under the type of ethics being defined. A principle such as privacy will have different implications if analyzed from the perspective of data ethics or algorithm ethics. Likewise, these principles are described through a general postulate. Therefore, the meaning of each of the principles are analyzed under the three ethics defined above.

1 Transparency

Transparency should be understood as openness to provide full information on the design, operation and impact of artificial intelligence systems. Such information should not compromise the confidentiality of the business model and innovation, making it susceptible to unauthorized use.

In data ethics: transparency has the effect of providing information about the sources of the information used for the design of this technology, its characteristics and the purposes for which the data, especially personal data, will be used. This has an effect on the data used for training and implementation of these systems, since the criteria used for their classification and processing must be transparent in their collection.

In the ethics of the algorithms: it means transparency in the model behind the algorithms, informing the criteria that leads these systems to have certain types of results. At this point, it is relevant to provide information to citizens about the inputs used in the design of systems and the results that may be presented (*input and output transparency*). In addition, complete information must be provided on the false positives or negatives that an algorithm has developed and the percentages of accuracy (*algorithm accuracy*).

In ethics of practices: it consists of providing complete information about the teams and people involved in the design and development of these systems, the objectives they pursue and the conduct manuals they have generated to carry out their work. These manuals must be publicly accessible. As a practice for those in charge of implementing this technology, the use of open source systems should be privileged, especially within State entities.

2 Explanation

This principle is seen at times as complementary to transparency, but differs in that the information that is shared and made known to the public in a transparent manner must be understandable both to the developers and users of the system and to those individuals who may be affected by its decisions and results. Therefore, it goes beyond transparency and seeks to make the content of the information and the way it is presented easily accessible, considering the complexities that surround this technology.

In data ethics: the explanation makes it possible for people to understand the importance of data in the design and development of these systems, how they collect and process information, and the purposes for which they do so, especially when processing personal data.

In the ethics of the algorithms: it means that the algorithms can be comprehended in the sense that it is allowed to understand the initial objectives that it looks for and that they are proper of the model, and the expected results and also those obtained. It is clear that the systems known as "*Black Box*" this principle can enter into tension, since the capacity of explanation can be limited. Even the developers of this type of technology and experts in the field are unable to understand fully the processing that takes place and the way in which the systems reach certain results. However, the aim is also to explain in an understandable way how these systems work, because they are called *Black Box* and the implications this may have. Similarly, and in any case, understandable and clear information must be presented about the objectives pursued by this system in its development and implementation.

***In the ethics of practice*:** it means providing clear and accurate information about the roles of the people involved in the design, development and implementation of this technology. In addition, clear, accurate and understandable information must be generated throughout the process about how these systems are being evaluated and specific mechanisms must be generated to share information about the results obtained, especially with the communities that are being impacted by these systems.

3 Privacy

Artificial intelligence must be preceded by a respect for people's privacy and their private sphere that prevents the use of information that they have not authorized and the profiling of individuals through this technology.

In data ethics: in this field the principle leads to the need to have authorization for the use of personal information, when this is not public data or under the exceptions specifically indicated by the law, describing the specific purposes and objectives pursued with the treatment (training, operation, etc.). It also requires the development of mechanisms to improve the quality of the data used and to achieve constant updating of the information. This leads to the possibility for the impacted population to correct erroneous or mistaken personal information being used for the development or operation of these systems, without the functionality of the technology being able to limit this type of request.

In the algorithms: the design of the algorithms must be respectful of the privacy of individuals and therefore the decision criteria should not be based on personal characteristics and their own private sphere. The use of personal information should be limited only to that which is necessary for the proper functioning of the system to avoid false positives or false negatives. Designers should avoid the development of technologies that facilitate the profiling of people, under criteria which are not previously known and authorized by them. The use of information to improve the operation or performance of these systems must be informed to the people who are in possession of that information.

In the ethics of practices: there must be internal procedures that develop good practices in the use of information and in the responses and explanations given to users impacted by these technologies. Priority should be given to measures of proven responsibility that allow the implementation of risk management tools for privacy, as well as mechanisms for analyzing the impact on privacy (*data protection impact assestments*). The design and development teams must generate criteria to identify those cases in which profiling may occur, its impact and how negative results can be avoided from this process. In this way, not only is privacy protected in an individual sense, but also in as a collective, avoiding the generation of undesired social classifications or profiles.

4 Human control of the decisions of an artificial intelligence system (*Human-in-the-loop* and *Human-over-the-loop*)

This principle is applicable to artificial intelligence systems that have a certain autonomy in decision making, making the human being have total control over decision making, especially in an implementation stage (*Human-in-the-loop*). Once a greater level of maturity of the technology has been reached in the country, there will be a greater level of autonomy in decision making, with mechanisms for human intervention, especially when undesired results are presented. This transition will take into account the social impact that may exist, especially in the future of work, given the displacement of human beings from certain activities.

***In data ethics*:** the collection and processing must be done according to the parameters and criteria established by human beings.

***In the ethics of algorithms*:** the algorithms must allow and facilitate decision making, but they must initially guide the decision making and cannot act in an automated way and according to suggested models.

***In the ethics of practice*:** artificial intelligence systems should not be used to interact with citizens without the control of a human being. Automated response and conversation systems must have mechanisms for humans to intervene and participate at any time. Practices that promote relationships with these systems without verifying that human beings are behind the contents or responses that are generated should be avoided.

5 Security

Artificial intelligence systems must not affect the physical and mental health and integrity of the human beings with whom they interact.

In data ethics: mechanisms must be implemented to ensure that this information will maintain its confidentiality, integrity and cannot be altered at any time. Mechanisms must be generated that can avoid this type of alteration to the information used by these systems and the way it is processed (*adversarial machine learning*).

In the ethics of algorithms: the implementation of algorithms and their design must follow a system of risks that allows establishing the possible affectations that certain results can generate and the possibility of avoiding it. In no case, an algorithm must lead to a result that puts at risk the integrity of a human being. Such decisions can only be made by human beings, and in any case, the algorithms will serve as models to guide the decisions that humans make regarding the life and integrity of others (e.g. health sector or national security sector).

In the ethics of practices: practices that put artificial intelligence systems at risk should be avoided and codes of conduct should generate parameters to avoid those activities that endanger the integrity and physical security of people.

6 Responsibility

There is a duty to answer for the results produced by an artificial intelligence system and the effects it may have. Solidarity in the responsibility of the designers, developers and people who implement this technology will be assumed, for the damages that the use of this technology has on an individual, unless it is sufficiently demonstrated that the responsibility falls on only one of these actors.

In data ethics: entities that collect and process data for the design, development and implementation of artificial intelligence systems must all be held accountable for this information and must be responsible for its integrity and processing purposes. In no case should responsibility fall to only one of these actors.

In the ethics of algorithms: there is responsibility of the people who design an algorithm for those results that come to generate and the criteria used to reach certain answers. However, their responsibility is not derived until implementation, since at this stage the results and their impact will be the responsibility of the person or entity that is in charge of using these systems and making decisions based on them.

In the ethics of practice: those involved in the development of this technology must establish clear responsibilities in the chain of design, production and implementation. Within the work teams there must be a clear distribution of functions and responsibility in their development and fulfillment. Practices and agreements between actors that limit responsibility should be restricted according to the way this principle has been established.

7 Non-discrimination

Artificial intelligence systems cannot have results or responses that threaten the well-being of a specific group or limit the rights of historically marginalized populations. Such decisions can only be made by human beings, under the criteria that the human rights framework allows in each case. The functionality of an artificial intelligence system should not be limited to a specific group on the basis of gender, race, religion, disability, age or sexual orientation.

In data ethics: the data used should be analyzed in such a way as to mitigate as much as possible the likelihood of using information that contains *biases*, either in its content, classification or the use that has been made of it. Preference should be given to mechanisms that allow a prior analysis of a set of data and the possible problems it may have.

In the ethics of algorithms: algorithms must be able to respond to the needs and interests of different population groups. The adequate performance of an algorithm cannot be limited to a specific population group. There must be a constant monitoring of false positives and negatives cases highlighted by a system which establishes the criteria of sex, race, religion, disability, age or sexual orientation in order for these not to affect these results.

In the ethics of the practices: a diverse group of the population must participate in the design and impact matrixes which allow a prompt response towards any possible discrimination. There must be a constant analysis of these impacts and even consider mechanisms to immediately withdraw systems that have discriminatory effects.

8 Inclusion

It is the active participation of historically marginalized populations in the design, development and implementation and evaluation of artificial intelligence systems used in Colombia. The State must use artificial intelligence systems that have met inclusion criteria and respond to the specific needs of these groups.

In data ethics: this implies using data that are representative and part of different social groups, whether for the design, training or operation of these systems. To this end, the availability of data sets from historically less represented groups should be increased.

***In the ethics of algorithms*:** the variables that have been included within the algorithm recognize the effects that it may have in particular contexts and the possibility that a specific group is privileged, avoiding such a design.

***In the ethics of practice*:** groups responsible for design, development and implementation should take into account different sectors of society and evaluation committees should be established to avoid discriminatory practices against groups such as women, people of African descent, indigenous people or members of the LGBTI+ community amongst others. Colombia must lead efforts to avoid the design of artificial intelligence systems that foresee women as personal assistants and beings at the service of consumers. The development of artificial intelligence systems that do not have a gender is desirable.

9 Prevalence of the rights of children and adolescents

Artificial intelligence systems must recognize, respect and privilege the rights of children and adolescents. In no case is the implementation of an initial intelligent system justified to the detriment of their best interests. We must advocate for strengthening education programs and strategies that facilitate this population's understanding of this technology and facilitate their interaction with this innovation.

In data ethics: data from this population cannot be used, except in those activities that relate to their best interests.

In the ethics of algorithms: the design and development of algorithms must be understandable to children and adolescents, especially when they have an impact on their development and well-being. The design of any algorithm that has an harmful impact on children and especially on practices such as *bullying and* discrimination should be avoided.

In the ethics of the practices: children and adolescents must be considered in the development of these systems when they are appropriate to their activities, establishing specific participation mechanisms that also allow them to evaluate the impact that these systems have on this population. Training and education programs must be generated to enable children to know and understand the characteristics of this technology and its implications, with emphasis on ethical training.

10 Social Benefit

The artificial intelligence systems implemented in Colombia must allow or be directly related to an activity that generates a clear and determinable social benefit. Such benefit can be materialized in the reduction of costs, the increase of productivity and the facilitation in the provision of public services, amongst others. Artificial intelligence systems that pursue other types of purposes should not be implemented in the public sector and their use in other sectors should be discouraged.

In data ethics: easy access to data and public data infrastructure must be prioritized for the development of artificial intelligence systems which show a clear social benefit, in the design of public policy and the provision of public services.

In the ethics of algorithms: the models and algorithms used must have as their ultimate goal a result linked to a socially recognized end, so it must be shown how the expected results relate to that social purpose.

In the ethics of the practices: the people who work in the design, development and implementation of this technology in Colombia must know the main social difficulties that the country faces and establish how this innovation and the desired implementation can help to solve it. The State must promote the use of this technology within a process of digital transformation that seeks to reduce gaps and decrease existing inequalities. For this same reason, programs should be established that promote public challenges in the use of artificial intelligence (*AIckaton*) and aim to solve a specific social problems.

VI

TOOLS FOR THE IMPLEMENTATION OF THE PROPOSED PRINCIPLES

One of the main doubts that may arise when analyzing these principles is to establish specific mechanisms for their implementation and for them to be materialized in the country and in the innovation and technology ecosystem they seek to impact. Below, a series of concrete mechanisms and tools that facilitate this impact will be presented, several of which have already been explored in other countries. Subsequently, the way in which each of these tools relates to the principles described in this ethical framework will be specifically shown.

However, none of these measures ensures total success in the implementation of these principles or that they are fully realized. This is a complex task and one that is being tested worldwide. It may also be necessary to consider new tools and strategies that have not been described in this document. This is entirely valid and desirable, as the tools to be described are not an exhaustive and exhaustive list. Similarly, an entity may consider that not all, but some, of these measures need to be implemented, according to the context and needs of each entity. The most important thing in the implementation of these types of measures is that they succeed in materializing and embodying the objectives sought with the principles already described.

1 Algorithm Assessment

This tool has been explored in recent years by various world authorities. The New Zealand data authority produced one of the main reports on this subject. From the report and analysis of cases of implementation of this technology in public entities of this country, the authority seeks to ensure that citizens of New Zealand are informed about the use of government algorithms and the weights and counterweights that exist to manage their use (Stats New Zealand, 2018). The New Zealand government's report builds on the Principles for the Safe and Effective Use of Data and Analysis developed by the Privacy Commissioner and the Government's Chief Data Officer, and makes recommendations to improve transparency and *accountability in the government's* use of algorithms (Stats New Zealand, 2018). The results of the report provide an opportunity for agencies to review and refresh the process they use to manage algorithms and will help shape the work of the Government's Chief Data Keeper and the Government's Chief Digital Officer to promote innovation and good practice across the data system. (Stats New Zealand, 2018).

Among the results of the report and the recommendations made is that, although some of the government agencies describe the use of algorithms with a standard of good practice, there is no consistency across the government (Stats New Zealand, 2018). There are significant opportunities for agencies to improve descriptions on how algorithms inform or impact decision making, particularly in those cases where there is a degree of automatic decision making or where the algorithms support decisions that have a significant impact on individuals or groups (Stats New Zealand, 2018).

Additionally, while some government agencies have formal processes for reviewing algorithms during development and operation, most do not (Stats New Zealand, 2018). There is no consistency across the government in including these processes within the organizational police, rather than relying on individual accountability (Stats New Zealand, 2018). This suggests that there is ample room for improvement, both to support decision makers and to ensure the continued improvement of the algorithms (Stats New Zealand, 2018).

Most government agencies said they hope to develop operational algorithms that rely on artificial intelligence in the future (Stats New Zealand, 2018). It will be challenging to explain clearly how these types of algorithms work and support decision making and how a given result is arrived at (Stats New Zealand, 2018). As technology evolves, this will continue to be an area where government agencies must balance the importance of human oversight with possible efficiencies in service delivery.

The Colombian government should consider preparing this type of analysis and report, such as the one carried out by the New Zealand government, the results of which we mentioned above. This would not only allow it to have a constant mapping of those transformative projects within the public sector in which it is using this technology, but also the way in which principles such as those indicated in this document are being implemented within the implementation and deployment of that technology.

2 Algorithm auditing

This is a proposal that has been led by several civil society entities. The French government has been one of the main promoters of its implementation in the public sector (Kayser-Bril, 2019). Even within its national strategy, the French government

considered the creation of a national audit platform of algorithms, especially those used by the government (Kayser-Bril, 2019). However, this proposal is still under discussion within that country's parliament (Kayser-Bril, 2019).

Many algorithmic behaviors that we might consider antisocial can be detected through appropriate audits, for example, explicitly exploring consumer service behavior, search results, or advertising, and quantitatively measuring results such as gender discrimination in a controlled environment (Kearns & Roth, 2020). However, to date, these audits have been carried out mainly on an ad-hoc basis and in isolation, usually by academics or journalists, and often in violation of the terms of service of the companies being audited (Kearns & Roth, 2020). Consequently, it is necessary to seek more systematic, continuous and legal ways to audit algorithms (Kearns & Roth, 2020). Regulating algorithms is different and more complicated than regulating human decision making (Kearns & Roth, 2020). That regulation must be based on what we have called ethical algorithm design, which is now being developed by a community of hundreds of researchers (Kearns & Roth, 2020). Ethical algorithm design begins with a precise understanding of the types of behavior we want the algorithms to avoid (so that we know what to look for in an audit) and then continues with the design and implementation of algorithms that avoid those behaviors (Kearns & Roth, 2020).

3 Data cleansing

This type of measure seeks to limit the biases and errors in the data used in the development of this technology. For this purpose, a series of steps have been generated that allow a process of debugging, correction and updating of this information, within which it is worth highlighting the following:

1. ***Monitor errors*:** keep track and observe trends on where most errors come from, as this will make it easier to identify and fix incorrect or corrupt data (Gimenez, 2018). This is especially important if other solutions are integrated with the main administrative software, so that errors do not obstruct the work of other departments (Gimenez, 2018).

2. ***Standardize processes***: it is important to standardize the point of entry and review its importance (Gimenez, 2018). By standardizing the data process, a good entry point is ensured and the risk of duplication is reduced (Gimenez, 2018).
3. ***Validate accuracy***: validating the accuracy of data once the existing database has been cleaned (Gimenez, 2018). Research and investment in data tools that help clean up data in real time is recommended (Gimenez, 2018). Some tools even use AI or *machine learning to* improve accuracy (Gimenez, 2018).
4. ***Searching for Duplicate Data***: Identifying duplicates can help save time when analyzing data (Gimenez, 2018). This can be avoided by finding and using the data cleansing tools mentioned above, which can analyze mass data and automate the process (Gimenez, 2018).
5. ***Analyze***: Once the data has been standardized, validated and reviewed for duplicates, third parties should be used to aggregate the data (Gimenez, 2018). Reliable external sources can collect information first hand, then clean up and compile the data to provide more complete information for business intelligence and analysis (Gimenez, 2018).
6. ***Communicate with the team***: Communicate the new standardized cleaning process to the team. Now that the data has been cleaned, it is important to keep it that way (Gimenez, 2018). This will help develop and strengthen consumer segmentation and send better targeted information to consumers and prospects, so it is important that the whole team is on the same page (Gimenez, 2018).

This is a particularly relevant measure in sectors that may be susceptible to the use of data that may be more biased or that may be "contaminated". It is also a highly recommended measure when entities are faced with databases of which there are several doubts as to their quality.

4 Smart explanation

As pointed out previously, the principle of explanation is one of the greatest challenges in terms of its materialization given the complexity of various artificial intelligence systems and that, in some cases, their operation is not entirely comprehensible, even to experts in the field. For this reason, a model has been proposed that we can consider as an "intelligent explanation" that seeks to provide citizens with understandable information about this innovation as long as there is a cost-benefit analysis that justifies this measure. In this case, it is recognized that the

explanation may be costly and time-consuming, and should therefore only proceed in those specific cases where access to this type of information presents more benefits than costs. This is why it is considered an intelligent explanation.

Thus, one must think about why and when explanations are useful enough to overcome the costs (Doshi-Velez & Kortz, 2017). Requiring all IA systems to explain all decisions can result in less efficient systems, forced design decisions, and a bias toward explainable but insufficient results (Doshi-Velez & Kortz, 2017). For example, the surcharges of forcing a toaster to explain why it thinks the bread is ready may prevent a company from implementing a smart toaster feature, due to engineering challenges or concerns about legal implications (Doshi-Velez & Kortz, 2017). On the other hand, we may be willing to accept the economic costs of a more explicable but less precise credit approval system for the social benefit of being able to verify that it is not discriminatory (Doshi-Velez & Kortz, 2017). There are societal rules about when we need explanations and these rules should apply to AI systems as well (Doshi-Velez & Kortz, 2017).

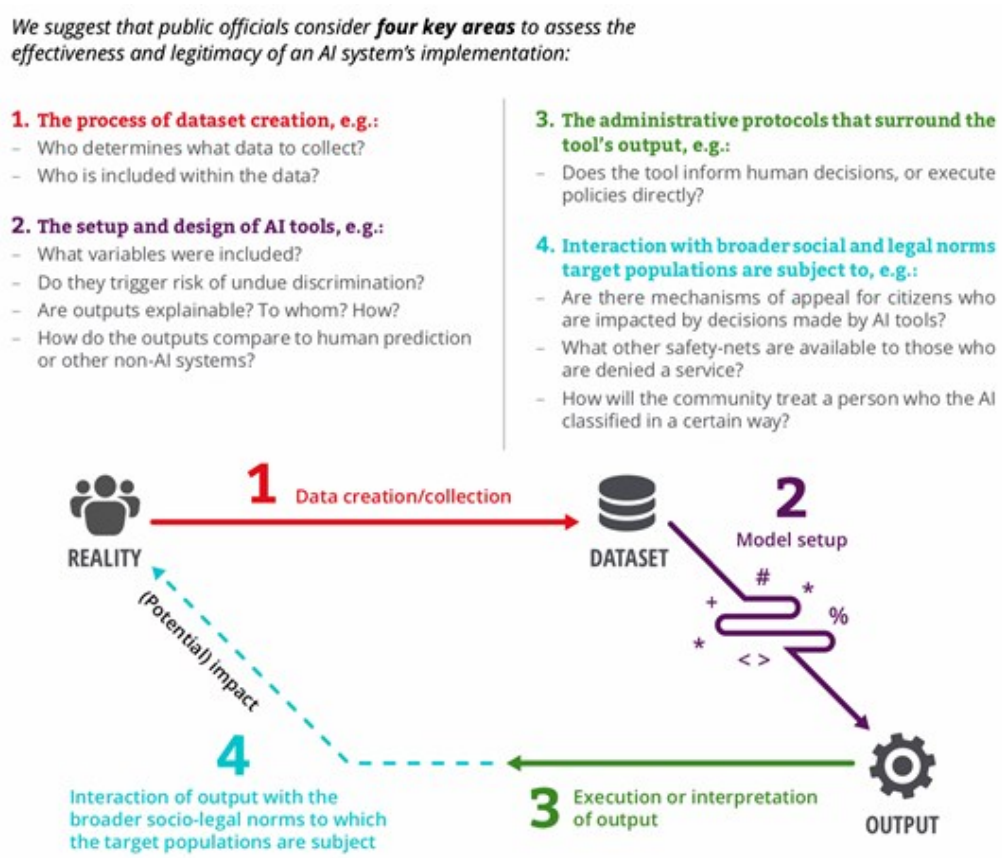
By doing this, we prevent AI systems from having free passes which avoid having the same level of scrutiny that humans can have, consequently avoiding asking too much of AI systems to the point of obstructing innovation and progress (Doshi-Velez & Kortz, 2017). Even this modest step will have its challenges, and as we resolve them we will get a better sense of whether and where the explanation requirements should be different for AI systems and for humans (Doshi-Velez & Kortz, 2017). Since we have little data to determine the actual costs of requiring AI systems to provide explanations, the role of explanation in ensuring accountability must also be re-evaluated from time to time, to adapt to the ever-changing technology landscape (Doshi-Velez & Kortz, 2017).

5 Legitimacy Evaluation

The Web Foundation has developed a model to **evaluate the legitimacy of the implementation of artificial intelligence systems, especially by public entities.** Legitimacy in implementation occurs when the procedure is explainable and has traceable responsibilities, allowing the precise definition of who is involved in the

different operations of the design and development of an artificial intelligence system, together with results that are non-discriminatory, fair and where the impact of false positives and negatives can be determined and minimized. (World Wide Web Foundation, 2018).

However, in order to determine whether this legitimacy is being presented, four specific steps are proposed to be followed by entities that are implementing this technology in advance:



(World Wide Web Foundation, 2018)

6 Sustainable and reliable system design

To ensure that an artificial intelligence system functions safely, the technical objectives of accuracy, reliability, safety and robustness must be prioritized (Leslie, 2019). This requires the technical team to examine in depth how to build a system that will operate accurately and reliably, in accordance with the expectations of its

designers, even when confronted with unexpected changes, anomalies and disturbances (Leslie, 2019). Building an AI system that meets these safety goals also requires rigorous testing, validation and reassessment, as well as the integration of appropriate monitoring and control mechanisms into its real-world operation (Leslie, 2019).

In *machine learning*, the accuracy of a model is the proportion of examples by which it generates a correct result (Leslie, 2019). The measure of this performance is also sometimes characterized as an error rate or the fraction of cases where the model produces an incorrect result (Leslie, 2019). It should be noted that sometimes the choice of an acceptable error rate or level of accuracy can be accepted according to the use of the specific needs of the application (Leslie, 2019). In other cases, it can be determined by a previously established standard (Leslie, 2019).

As a measurement of performance, accuracy must be a central component in establishing and qualifying a team's approach to safe AI (Leslie, 2019). Specifying a reasonable level of performance for a system often requires refinement or change in the measurement of accuracy (Leslie, 2019). For example, if certain errors are more significant or costly than others, a total cost metric can be included in the model so that the cost of some errors can be higher than others (Leslie, 2019). In addition, if the precision and sensitivity of the system to detect rare events is a priority, the precision and memory technique can be used (Leslie, 2019). This method of addressing unbalanced classifications allows the proportion of correct system detections to be weighed against the proportion of actual rare event detections (Leslie, 2019).

In general, measuring accuracy in uncertainty is a challenge that should be given significant attention (Leslie, 2019). The level of confidence in the artificial intelligence system will depend largely on the problems inherent in attempts to model the changing and chaotic reality (Leslie, 2019). Accuracy concerns must address issues of unavoidable noise that will be present in the sample data, architectural uncertainties generated by the possibility that a given model may lack relevant characteristics of the underlying distribution, and unavoidable changes in the data over time (Leslie, 2019).

On the other hand, the goal of **reliability** is for an AI system to behave exactly as its designers intended and anticipated (Leslie, 2019). A reliable system adheres to the specifications for which it was programmed (Leslie, 2019)-. Reliability is then a

measure of consistency and can determine the confidence in the security of a system, based on the credibility with which its operation conforms to the intended functionality (Leslie, 2019).

The goal of **security,** on the other hand, encompasses the protection of several operational dimensions of an AI system when confronted with a potential adverse attack (Leslie, 2019). A secure system is capable of maintaining the integrity of the information that constitutes it (Leslie, 2019). This includes protecting your architecture from unauthorized modifications or damages to any of its parts (Leslie, 2019). A secure system must also be continuously functional and accessible to its authorized users, keeping information private and confidentially secure even under hostile and adverse conditions (Leslie, 2019).

Finally, the goal of **robustness** can be thought of as the purpose of a reliable and accurate AI system functioning under harsh conditions, which may include adverse intervention, implementer errors, or distorted executions by an automated learner (Leslie, 2019). The measure of robustness is therefore the strength of a system's integrity and the consistency of its operation in response to difficult conditions, adverse attacks, disruptions, data poisoning and its undesirable enhanced learning behavior (Leslie, 2019).

Among the measures to be implemented are :

1. Run extensive simulations during the test phase, so that appropriate restriction measures can be programmed into the system (Leslie, 2019).
2. Continually inspect and monitor the system, so that its behavior can be better predicted and understood (Leslie, 2019).
3. Find ways to make the system easier to interpret, so it can better evaluate its decisions (Leslie, 2019).
4. Wiring mechanisms in the system that allow humans to override and shut down the system (Leslie, 2019).

This should involve rigorous protocols for testing, validating, verifying and monitoring system safety, as well as self-assessments of the safety performance of AI systems by relevant team members at each stage of the workflow (Leslie, 2019). These self-assessments should evaluate how the design and implementation practices of the equipment are consistent with the safety objectives of accuracy, reliability, safety and robustness (Leslie, 2019). The self-assessment should be contained in a single document in a form that allows for review and re-evaluation (Leslie, 2019).

Two measures that must be implemented on a mandatory basis and prior to the deployment of this technology in the country are *F-1 scores* (Shmueli, 2019) and *confusion matrix* (Narkhede, 2018). Both are methodologies that have allowed the precise definition of the results obtained by one of these systems, giving the possibility to make early improvements in these systems.

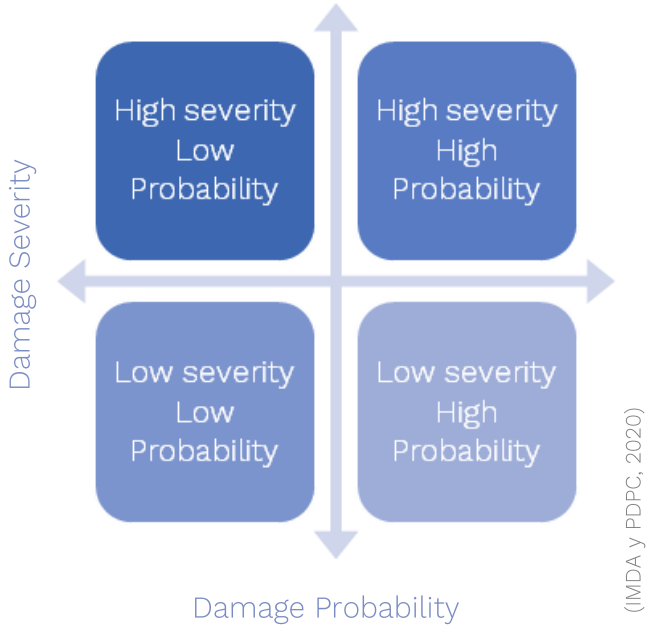
7 Definition and risk management

Organizations may consider implementing a consistent system of risk management and internal controls that address the risks involved in deploying a particular AI model (IMDA and PDPC, 2020). These measures include:

1. Use reasonable efforts to ensure that the databases used to train the AI model are adequate for the intended purpose, and assess and manage the risks of inaccuracy or bias, as well as review the exceptions identified during model training (IMDA and PDPC, 2020). Virtually no database is completely unbiased (IMDA and PDPC, 2020). Organizations should strive to understand the ways in which databases may be biased and address this in their security measures and deployment strategies (IMDA and PDPC, 2020).
2. Establish monitoring and reporting systems and processes to ensure that competent management is informed regarding performance and other issues related to the system deployed (IMDA and PDPC, 2020). Where appropriate, monitoring may include autonomous monitoring to effectively scale up human supervision (IMDA and PDPC, 2020). AI systems can be designed to report the confidence level of their predictions and explanatory features can focus on why the model achieved a certain level of confidence (IMDA and PDPC, 2020).
3. Ensure appropriate knowledge transfer when there are changes in key personnel related to AI activities, which will reduce the risk of staff turnover by creating a gap in internal governance (IMDA and PDPC, 2020).
4. Review internal governance structures and measures when there are significant changes to the organizational structure or key personnel involved (IMDA and PDPC, 2020).

5. Periodically review the internal governance structure and measures to ensure its continued relevance and effectiveness (IMDA and PDPC, 2020).

The model framework proposed by IMDA and PDPC also proposes a matrix to classify the probability and severity of harm to an individual as a result of an organization's decision about that individual (IMDA and PDPC, 2020). The definition of damage and the computation of probability and severity depend on the context and vary from sector to sector (IMDA and PDPC, 2020). For example, the harm associated with an incorrect diagnosis of a patient's medical condition will depend on the harm associated with an incorrect product recommendation (IMDA and PDPC, 2020).



(IMDA y PDPC, 2020)

However, it is not only private actors that are considering a risk management approach, but also governments that are doing so. For example, European Union members have noted the current lack of a common European framework (European Commission, 2020). The German Data Ethics Commission has called for a five-level risk-based system of regulation, which would go from no regulation for the most innocuous AI systems to a complete ban for the most dangerous (European Commission, 2020). Denmark recently deployed the prototype of a Data Ethics Seal (European Commission, 2020). Malta introduced a voluntary certification scheme for AI (European Commission, 2020). If the European Union fails to provide approximation for all its member countries, there is a real risk of fragmentation in the internal market, which would undermine the objectives of confidence, legal certainty and market penetration (European Commission, 2020).

8 Differential Privacy

The general objective of differential privacy is to ensure that different types of statistical analysis do not compromise privacy and that privacy is preserved if, after the analysis, the analyzer knows nothing about the characteristics of a database, so that the information made public is not harmful to an individual (Garg, 2019). As proposed by Cynthia Dwork, differential privacy describes a promise made by the data owner to the data subject that the data of the subject will not be affected by allowing the use of his or her data in a study or analysis, no matter what other studies, databases or sources of information are available (Garg, 2019). To define privacy in the context of a simple database, by removing a person from the database and the query does not change, then that person's privacy would be completely protected (Garg, 2019). This means that when removing the person from the query it is checked that there was no data leakage in the query result (Garg, 2019).

9 Internal codes of conduct and/or ethics

The objective of codes of conduct and/or ethics is to establish the expected behaviors of those who develop, deploy, and use data-based technologies, to ensure that all people in this chain comply with the ethical principles for data initiatives: respect for individuals, respect for human rights, participation, and accountability for decisions (UK Department of Health and Social Care, 2019).

The minimum elements that each code must contain are: (i) the principles followed by each institution or entity; (ii) the scope; (iii) whether or not there is exclusivity, taking into account that exclusivity may limit the benefits in different systems; (iv) value, taking into account that technology providers may generate significant value from any product outside the scope of the initial product, which should be recognized; (v) intellectual property; (vi) liability; (vii) audit; (viii) bias and how to prevent it, taking into account that the greatest threat to data-based technology is the actual or potential presence of bias, for which any commercial agreement should be identified, as well as how it is handled, by whom and at whose expense; and (x) the different roles (UK Department of Health and Social Care, 2019).

In addition, the UK Cabinet Office published a Data Science Ethics Framework that aims to help researchers as *big data* methods begin to be used in the public sector (ICO, 2017). This Ethical Framework includes six principles: (i) start with clear user needs and public benefit; (ii) use data and tools that are minimally intrusive; (iii) create robust data science methods; (iv) be alert to public perceptions; (v) be as open and accountable as possible; and (vi) keep data secure (ICO, 2017).

There is a role in all this for ethics councils or boards, both within entities and institutions and at the national level (ICO, 2017). A large organization can have its own ethics board, which can ensure that its ethical principles are applied and can assess difficult situations such as balancing legitimate interests and privacy rights (ICO, 2017). An important element is the relationship between the ethics board and employees with responsibilities for data and analysis, such as the data protection officer (ICO, 2017).

10 Education and research strategies in the ethics of artificial intelligence

The French government has made this one of its priorities, therefore proposes the following measures:

1. **Encourage the education of all actors involved in algorithmic chain** (designers, professionals, citizens) **on ethics** (CNIL, 2018).
2. **Increase initiatives for ethical AI research and launch a national cause in a general interest research project** (CNIL, 2018).

On the other hand, it is interesting to take into consideration other projects developed by other entities with experience in the field, such as MIT, which seeks to bring the ethics of artificial intelligence into the school curriculum in a constructive and innovative manner. The ultimate goal is to enable students to see the adaptability of artificial intelligence, from a technical and societal point of view, to empower students with tools to design AI with ethical mindsets (MIT Media Lab Staff, n.d.). The MIT project seeks to develop an open source curriculum for high school students in the area of artificial intelligence (MIT Media Lab Staff, n.d.).

Through a series of lessons and activities, students learn technical concepts and the ethical implications of those concepts, such as algorithmic bias (MIT Media Lab Staff, n.d.). During the curriculum, students learn to think about algorithms as opinions, they are taught to consider direct and indirect *stakeholders,* and are involved in designing activities to re-imagine familiar artificial intelligence systems (MIT Media Lab Staff, s.f.).

11 Privacy Impact Assessments

A privacy impact analysis is an important tool that can help identify and mitigate privacy risks before processing personal data (ICO, 2017). Under the European Union's General Data Protection Regulation (GDPR) it is likely that conducting a privacy impact analysis will be a requirement for the analysis of mass data involving the processing of personal data (ICO, 2017). The unique features of mass data analysis can make some steps in a privacy impact analysis more difficult, but these challenges can be overcome (ICO, 2017).

12 Ethical approach to data (*litmus test*)

An ethical approach to the processing of personal data in the context of *big data* is an important compliance tool (ICO, 2017). Ethics boards at the organizational and national levels can help evaluate elements and ensure the application of ethical principles (ICO, 2017). Ethical approaches to the use of personal data can help build trust with individuals (ICO, 2017). There is a role in setting the standards for *big data* and promoting best practices across industries (ICO, 2017). Sometimes these principles are condensed into a simple *litmus test to* remind employees to think about them when planning new uses for the data, for example, if they want a family member's data to be used in a certain way (ICO, 2017).

13 Personal data stores

The use of *data stores* can help to address issues of fairness and lack of transparency by giving individuals greater control over their personal data (ICO, 2017). *Data stores* can support the concept of data portability, which is part of the GDPR, in relation to the re-use of an individual's personal data under his control (ICO, 2017). It has been suggested that one way to increase an individual's control over the use of his or her data is through what are usually known as personal *data stores,* or sometimes personal information management services (ICO, 2017). These are services provided by third parties that contain the persons data on their behalf and make them available to organizations when individuals wish to do so (ICO, 2017). This is also a measure that can help Colombia with its compliance with the OECD's general privacy principles.

14 Strengthen business ethics and corporate programs and human rights

While companies generally indicate they operate responsibly or ethically, different indicators, standards and certifications provide some assurance to *stakeholders* as to whether the company is indeed doing what they say they are doing, i.e. whether it is meeting its commitments (Institute of Business Ethics, 2012). These mechanisms can also be an incentive for companies (Institute of Business Ethics, 2012).

Respecting human rights is a criterion in several indicators and standards, such as the UN Global Compact, the ISO 26000 standard for social responsibility and the Dow Jones Sustainability Index (Institute of Business Ethics, 2012). FTSE4Good Index requirements vary depending on whether companies operate in sectors and countries with high or low human rights impact (Institute of Business Ethics, 2012). To qualify for inclusion, high-impact businesses must implement, among others: (i) a human rights policy; (ii) human rights training; (iii) human rights responsibility and *accountability* at the board level; (iv) *stakeholder* engagement with local communities; and (v) regular monitoring and reporting mechanisms on activities and progress (Institute of Business Ethics, 2012).

Given that this is an issue that is constantly developing and in which implementation guidelines are currently being designed to make companies accountable for human rights violations, as governments may be, it is essential that the Colombian Government permanently evaluate its commitments in the face of developments by international organizations, such as the UN, on this issue. Like so, the Government can implement different mechanisms for companies to comply with human rights and mitigate the risks and threats that different technologies imply for human rights.

15 Governance models to ensure the ethics of artificial intelligence

Broadly speaking, a model AI Governance Framework has been proposed that contains guidance on measures to promote the responsible use of artificial intelligence that organizations should adopt in the following key areas (IMDA and PDPC, 2020):

1. Adapt internal governance structures and measures to incorporate values, risks and responsibilities related to algorithmic decision-making (IMDA y PDPC, 2020).
2. Determine the level of human involvement in decision making increased by AI (IMDA y PDPC, 2020).
3. Operations management: considering elements in the development, selection and maintenance of AI models (IMDA and PDPC, 2020).
4. Strategies for communicating and interacting with an organization's *stakeholders* and managing relationships with them (IMDA and PDPC, 2020).

Organizations that adopt this model may not consider all elements to be relevant, as the model is designed to be flexible and organizations can adapt it to fit their needs by adopting the relevant elements (IMDA and PDPC, 2020).

Within the elements that a particular entity should have, the following clear roles and responsibilities for ethical AI deployment should be implemented (IMDA and PDPC, 2020):

1. Responsibility for and oversight of the various stages and activities involved in AI deployment should be assigned to the appropriate personnel and/or departments (IMDA y PDPC, 2020). If necessary and possible, consideration should be given to establishing a coordinating body, which has relevant expertise and appropriate representation from across the organization (IMDA y PDPC, 2020).
2. Staff and/or departments with internal IA governance functions should be fully aware of their roles and responsibilities, be adequately trained and have the necessary resources and guidance to carry out their duties (IMDA and PDPC, 2020).

Key roles and responsibilities that may be assigned include:

1. Use any risk management framework and apply risk control measures to : (i) assess and manage the risks of deploying AI, including any potential adverse impact to individuals; (ii) decide on the appropriate level of involvement in AI-assisted decision making; and (iii) manage the AI training model and selection process (IMDA y PDPC, 2020).
2. Maintain, monitor, document and review the AI models that have been deployed, with a view to taking remedial action if necessary (IMDA and PDPC, 2020).
3. Review communication channels and interactions with *stakeholders* to provide effective outreach and feedback channels (IMDA y PDPC, 2020).
4. Ensure that relevant personnel dealing with AI systems are adequately trained (IMDA and PDPC, 2020). When applicable and necessary, staff working and interacting directly with AI models may need to be trained to interpret AI model output and decisions, and to detect and manage biases in the data (IMDA and PDPC, 2020). Other staff whose work requires interaction with the AI system must be trained to at least be alert to and sensitive to the benefits, risks and limitations of using AI, so that they know when to alert the experts in the field within their organizations (IMDA and PDPC, 2020).

VII

RELATIONSHIP BETWEEN THE PROPOSED PRINCIPLES AND THE IMPLEMENTATION TOOLS

Below is how the implementation tools described above interact with each of the principles. As can be seen, some tools can facilitate the implementation of all the principles, while other measures have a more targeted impact on specific principles:

	Transparency	Explenation	Privacy	Human Control	Security	Responsability	Non-discrimination	Inclusion	Prevalence of the rights of children and asolescent	Social Benefit
Algorithm assessment	x	x	x	x	x	x	x	x	x	x
Algorithm auditing	x	x	x		x					
Data cleansing			x		x	x	x			
Smart explanation	x	x					x			x
Legitimacy Evaluation	x			x			x	x	x	x
Sustainable and reliable system design					x	x				
Definition and risk management:			x	x	x	x				
Differential privacy:			x		x					
Internal codes of conduct and/or ethics:	x	x	x	x	x	x	x	x	x	x
Education and research strategies in the ethics of artificial intelligence	x	x	x	x	x	x	x	x	x	x
Privacy Impact Assessments			x		x		x			
Ethical approach to data (litmus test):			x	x	x	x	x			
Personal data stores			x							
Strengthen business ethics and corporate programs and	x	x	x	x	x	x	x	x	x	x

VIII

RECOMMENDATIONS FOR THE COLOMBIAN GOVERNMENT

As noted above, this framework should serve as an initial input to develop a broader discussion of the issue. Therefore, we point out a series of recommendations that the Colombian Government, at the head of the Presidency of the Republic, should consider in order to have this discussion and the subsequent adoption of an ethical framework for artificial intelligence in both the public and private sectors:

1. Prioritize the publication of an ethical framework for artificial intelligence given the risks identified and prior to further deployment of this technology. This means that the Colombian Government should highlight the importance of this ethical framework and why it should be prioritize, especially in those entities that have already implemented this technology or are about to do so.

2. The Government, under the leadership of the Presidential Council for Economic Affairs and Digital Transformation and the Ministry of Information and Communication Technologies, must sponsor a dialogue with different sectors of society given the transversal effect that a framework of these characteristics has. However, this dialogue must take place on a concrete basis that allows for the orientation of the discussion and its objectives. This proposal is expected to be an initial input for this purpose. The dialogue should involve national and international actors.

3. Following the discussions, adjustments to the framework should be made. These adjustments should incorporate the main conclusions drawn from these discussions.

4. Government entities should lead a campaign to disseminate the framework in order to enable its knowledge by a wide sector of society and the reception of comments from different sectors.

5. An ethical framework of these characteristics must be based on general principles, capable of adapting to the technological changes that will occur. The principles proposed should not ignore the international proposals that have already been made and their implications in different sectors.

6 An ethical framework for artificial intelligence must not ignore the impact that the proposed principles will have on each of the moments in the algorithmic chain and on what has been called the ethics of data, algorithms and practices.

7 The tools for implementing the principles are essential and constitute one of the most relevant elements of the framework. Without these, the framework lacks tools to be truly realized within Colombian society. The government must deepen its design and form of implementation, developing an organized plan for that purpose.

8 In addition to the above, the implementation process must be accompanied by a strategy of sensitization and training in the public sector so that the framework is understood, especially the tools of materialization that are finally adopted.

9 Metrics must be generated to evaluate the results and impact of the proposed principles. This impact should measure the effect that the adopted framework has on the adoption of this emerging technology and the cases of affectation to citizens and their fundamental rights

10 Although initially it is expected that these principles should not be adopted through binding regulations, the Colombian Government must constantly analyze the need to make them known through some legal instrument and the characteristics of the same.

11 Given the novelty of the subject, it is possible that a constant iteration of this process will emerge that will allow a refinement of this framework, considering the changes and new challenges that an emerging technology in constant transformation such as artificial intelligence can bring.

References:

AI and Inclusion Staff. (n.d.). *AI and Inclusion*. Retrieved from https://aiandinclusion.org/

Cabrol, M., González, N., Pombo, C., & Sanchez, R. (2020, January). *Ethical and responsible adoption of Artificial Intelligence in Latin America and the Caribbean.* Retrieved from Interamerican Development Bank: https://publications.iadb.org/publications/spanish/document/fAIr_LAC_Adopci%C3%B3n_%C3%A9tica_y_responsable_de_la_inteligencia_artificial_en_Am%C3%A9rica_Latina_y_el_Caribe_es.pdf

CNIL. (2018, Mayo 25). *Algorithms and artificial intelligence: CNIL's report on the ethical issues.* Retrieved from CNIL: https://www.cnil.fr/en/algorithms-and-artificial-intelligence-cnils-report-ethical-issues

Comisión Europea. (2020, Febrero 19). *White Paper On Artificial Intelligence - A European approach to excellence and trust.* Retrieved from European Commission: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

Crawford, K., Dobbe, R., Dryer, T., Fried, G., Green, B., Kaziunas, E., . . . Whitt. (2019, Diciembre). *AI Now Institute.* Retrieved from AI Now 2019 Report: https://ainowinstitute.org/AI_Now_2019_Report.pdf

Doshi-Velez, F., & Kortz, M. (2017). *Accountability of AI Under the Law: The Role of Explanation.* Retrieved from Berkman Klein Center Working Group on Explanation and the Law: http://nrs.harvard.edu/urn-3:HUL.InstRepos:34372584

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020, Enero). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI.* Retrieved from Berkman Klein Center for Internet & Society: https://dash.harvard.edu/handle/1/42160420

Floridi, L., & Taddeo, M. (2016, December). What is Data Ethics? *Phil. Trans. R. Soc. A, 374*(2083).

Garg, A. (2019, Junio 11). *Differential Privacy and Deep Learning.* Retrieved from Medium: https://webcache.googleusercontent.com/search?q=cache:osZUXYGNnusJ:https://mc.ai/differential-privacy-and-deep-learning-2/+&cd=20&hl=en&ct=clnk&gl=co

Gimenez, L. (2018, Mayo 24). *6 steps for data cleaning and why it matters.* Retrieved from Geotab: https://www.geotab.com/blog/data-cleaning/

Giubilini, A. (2016, Marzo 14). *Conscience.* Retrieved from Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/entries/conscience/

ICO. (2017). *Big data, artificial intelligence, machine learning and data protection.* Retrieved from Information Commissioner's Office: https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

IMDA y PDPC. (2020). *Model Artificial Intelligence Governance Framework.* Retrieved from PDPC Singapore: https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf

Institute of Business Ethics. (2012, Julio). *Business Ethics Briefing Issue 26.* Retrieved from Business and Human Rights: https://www.business-humanrights.org/sites/default/files/media/business-ethics-human-rights-briefing_.pdf

Kayser-Bril, N. (2019, Enero 29). *Report Automating Society France.* Retrieved from Algorithm Watch: https://algorithmwatch.org/en/automating-society-france/

Kearns, M., & Roth, A. (2020, Enero 13). *Ethical algorithm design should guide technology regularion.* Retrieved from Brookings: https://www.brookings.edu/research/ethical-algorithm-design-should-guide-technology-regulation/

Leslie, D. (2019). *Understanding artificial intelligence ethics and safety A guide for the responsible design and implementation of AI systems in the public sector.* Retrieved from The Alan Turing Institute: https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf

MIT Media Lab Staff. (n.d.). *AI + Ethics Curriculum for Middle School.* Retrieved from MIT Media Lab: https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/

Narkhede, S. (2018, Mayo 9). *Understanding Confusion Matrix.* Retrieved from Towards Data Science: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

Raso, F. A., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Kim, L. (2018, Septiembre 25). *Artificial Intelligence & Human Rights: Opportunities & Risks.* Retrieved from Berkman Klein Center For Internet & Society at Harvard University: https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights

Shmueli, B. (2019, Julio 3). *Multi-Class Metrics Made Simple, Part II: the F1-score* . Retrieved from Towards Data Science: https://towardsdatascience.com/multi-class-metrics-made-simple-part-ii-the-f1-score-ebe8b2c2ca1

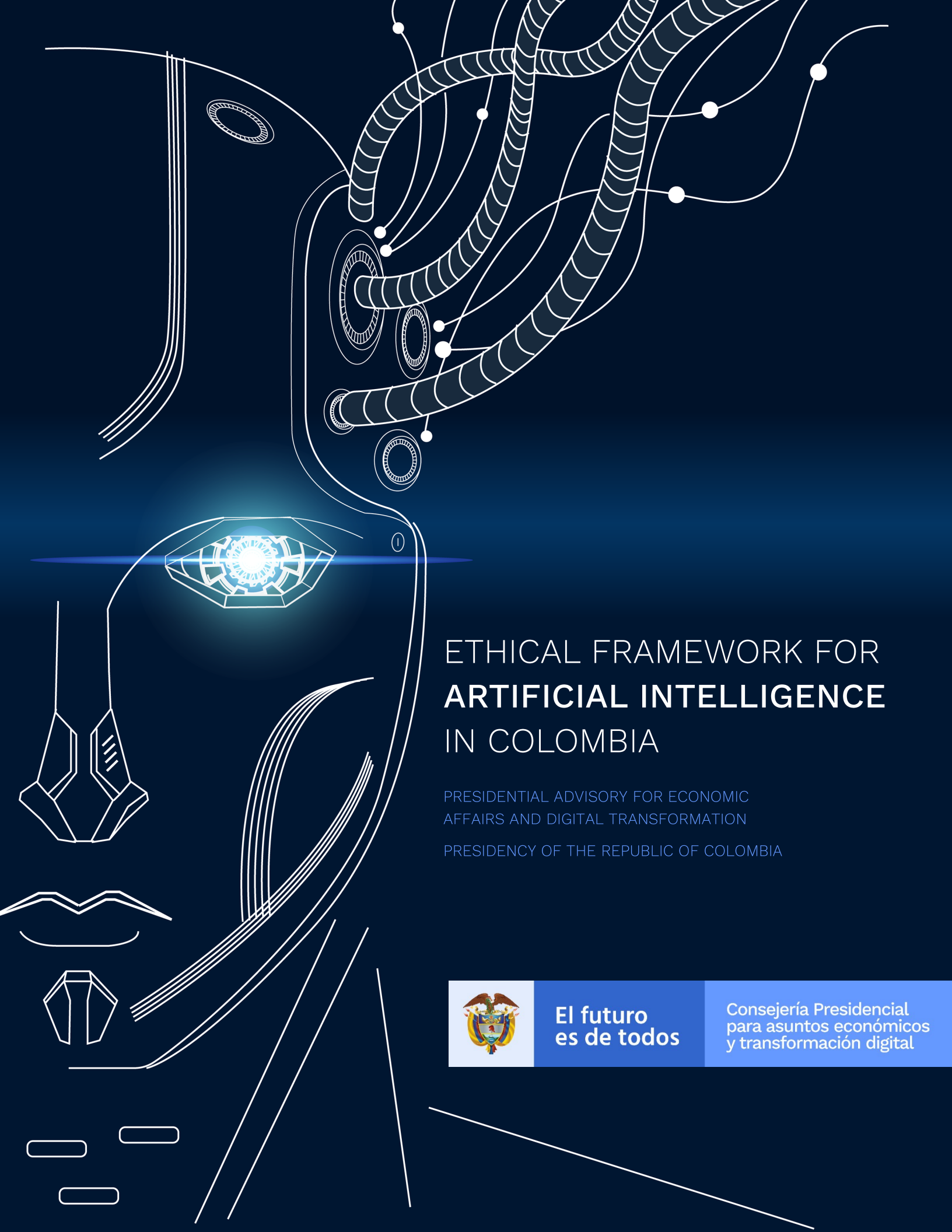
Stats New Zealand. (2018, Octubre). *Algorithm assessment report.* Retrieved from New Zealand Government: https://www.data.govt.nz/assets/Uploads/Algorithm-Assessment-Report-Oct-2018.pdf

UK Department of Health and Social Care. (2019, Julio 18). *Code of conduct for data-driven health and care technology.* Retrieved from UK Department of Health and Social Care: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology

World Wide Web Foundation. (2017, Julio). *ALGORITHMIC ACCOUNTABILITY Applying the concept to different country contexts.* Retrieved from World Wide Web Foundation: https://webfoundation.org/docs/2017/07/Algorithms_Report_WF.pdf

World Wide Web Foundation. (2018). *HOW ARE GOVERNMENTS IN LATIN AMERICA USING ARTIFICIAL INTELLIGENCE? A proposal for effective and legitimate implementations of AI systems in the public sector.* Retrieved from World Wide Web Foundation: http://webfoundation.org/docs/2018/07/AI-in-Latin-America_Overview.pdf

World Wide Web Foundation. (2018, Septiembre). *World Wide Web Foundation.* Retrieved from ALGORITHMS AND ARTIFICIAL INTELLIGENCE IN LATIN AMERICA A Study of Implementation by Governments in Argentina and Uruguay: http://webfoundation.org/docs/2018/09/WF_AI-in-LA_Report_Screen_AW.pdf



ETHICAL FRAMEWORK FOR **ARTIFICIAL INTELLIGENCE** IN COLOMBIA

PRESIDENTIAL ADVISORY FOR ECONOMIC AFFAIRS AND DIGITAL TRANSFORMATION

PRESIDENCY OF THE REPUBLIC OF COLOMBIA

El futuro es de todos

Consejería Presidencial para asuntos económicos y transformación digital