

The background of the cover is a photograph of a street in Bogotá, Colombia, featuring colorful buildings and lush greenery. A white network overlay, consisting of interconnected nodes and lines, is superimposed on the image, extending from the top left towards the right side. The text is overlaid on a semi-transparent grey rectangular area on the left side of the image.

The Colombian case: adopting collaborative governance as a path for implementing ethical artificial intelligence

*Víctor Muñoz
Elena Tamayo
Armando Guio*



The Colombian case: adopting collaborative governance as a path for implementing ethical artificial intelligence is a research project that was developed during 2021 and was financed by the Center for Technology and Society Studies (CETyS) of the University of San Andrés, Argentina.

Researchers:

Armando Guío, Víctor Muñoz y Elena Tamayo

Design: Mónica Castellanos

Translations: Verónica Penelas

Document of the project The Colombian case: adopting collaborative governance as a path for implementing ethical artificial intelligence available at:

<http://hdl.handle.net/10908/18743>

Creative Commons Attribution/NonCommercial-NoDerivatives 4.0 International Public License.



The views expressed in the publications are solely those of the authors. They are not intended to reflect the views or perspectives of CETyS or any other organization involved in the project.

Abstract:

Artificial intelligence has permeated most industries from manufacturing, to healthcare, to food, to the creative industries. It has enormous potential to solve global issues we face today, but it also represents considerable risks in terms of discrimination, privacy, bias, inequality, safety, and security. The paper identifies the main risks of AI particularly for the Latin American region: discrimination, threats to civil liberties, and threats to security. This paper presents the challenges that Latin American countries face in the need to address ethical risks of AI while the concrete path for practical implementation of ethical AI remains unclear. Then, this paper analyzes the case of Colombia that has adopted a collaborative governance approach in the path of promoting ethical AI but that needs to deepen its practical implementation of AI. For this, the paper focuses on the 'Ethical Framework for Artificial Intelligence in Colombia', whose content and adoption process are both oriented towards the implementation of ethical AI, the first document in Latin America on this subject with a practical approach.

1.

Introduction

Artificial intelligence is every day more present and more influential in human lives. It permeated most industries from manufacturing, to healthcare, to food, to the creative industries. 40% of businesses are currently using artificial intelligence (Kumar, 2020). AI is also playing an essential role in resolving the most pressing issues that the world is currently facing: COVID – 19 (Council of Europe) and economic recovery (Inaa Group, 2020). AI also has the potential to contribute to resolve more structural issues such as inequalities, corruption and climate change (Castaño, 2020). However, artificial intelligence represents considerable risks in terms of discrimination, privacy, bias, inequality, safety and security (University of Cambridge, 2021). Even in cases in which AI is intended to have positive impacts it can have negative side effects as in the prioritization for the COVID – 19 vaccination process, in which people of color have been evidenced to be more vulnerable to the pandemic (Jercich, 2021). Therefore, adopting an ethical approach to the design, development and implementation of artificial intelligence appears indispensable for preventing risks and ensuring AI has positive impacts for users and society.

Artificial Intelligence refers to ‘a field of computer science dedicated to solving cognitive problems commonly associated with human intelligence or intelligent beings, understood as those that can adapt to changing situations. Its basis is the development of computer systems, data availability and algorithms’ (Colombia. National Policy for Digital Transformation and Artificial Intelligence, 2019, p.16). Artificial intelligence has the capacity to learn from experience and adapt itself to different circumstances, taking decisions, and producing autonomous output (Walz & Firth – Butterfield). AI ethics ‘comprise a set of values, principles, and techniques which employ widely accepted standards of right and wrong to guide moral conduct in the development and deployment of Artificial Intelligence technologies’ (Fourtané, 2020). AI ethics refers to the ethics of a family of technologies but not to specific applications of AI.

While there is a global consensus regarding the importance of ethics in AI, concretely implementing ethical AI remains challenging. Latin American countries have adhered to this global consensus as AI represents

pressing ethical issues in the sub-continent, however they face practical implementation challenges.

This paper first presents the challenges that Latin American countries face in the need to address ethical risks of AI while the concrete path for practical implementation of ethical AI remains unclear. Second, this paper analyzes the case of the 'Ethical Framework for Artificial Intelligence in Colombia' as a collaborative governance approach in the path of promoting ethical AI, and considers how collaborative governance can be better used in the Colombian case.

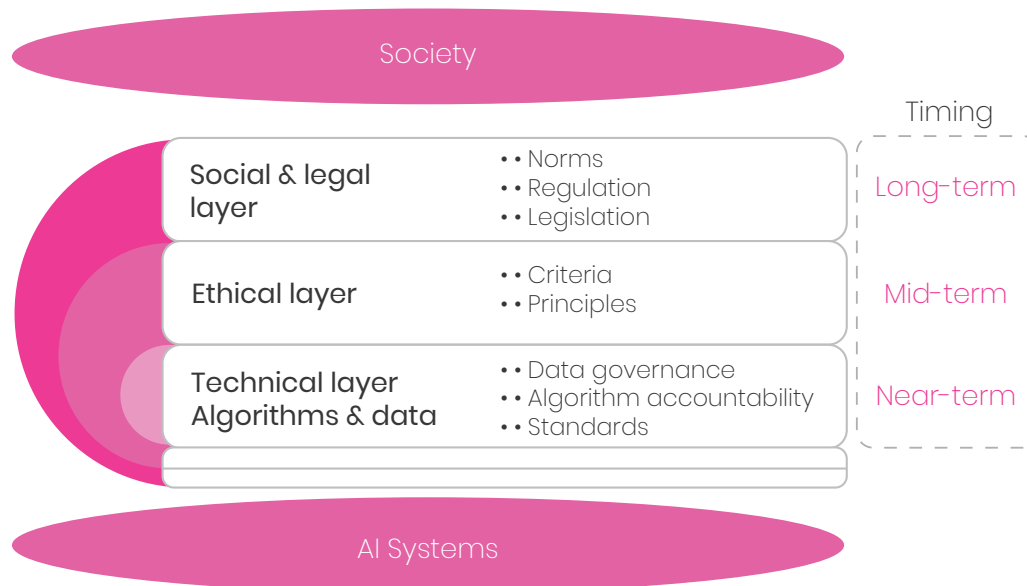
This paper will first establish a theoretical approach around AI ethics and governance. Then it will analyze the path towards implementing ethical AI in Latin America, asking: are ethics a tool to address AI challenges? And Why has it been difficult for nations worldwide to bridge the gap between principles and practical implementation? Then it will review a potential answer to at least part of the challenges of bridging said gap: Colombia's bet for collaborative governance, asking what further actions could be taken to effectively implement ethical AI in Colombia. Finally, the paper concludes by identifying specific recommendations based on the global experience for the Latin American region when implementing ethical AI.

A. AI Ethics and Governance: a theoretical approach

The approach to AI ethics and governance in the region, and the specific case of Colombia, will be analyzed within the framework of two AI ethics and governance approaches proposed by academics. The first model is the layered model, which consists of using 'layers' to develop AI governance.

In this model, three layers are taken into account to develop AI governance in the near, mid, and long term. The first layer is the technical one, which refers to algorithms and data, and it constitutes the foundation for the AI governance ecosystem. The second layer is the ethical layer, which is meant to aid regulators in articulating the ethical concerns that AI technologies of all kinds bring along. The main purpose of this layer is to have a system in place to assess actions driven by algorithms according to ethical criteria and principles. The top layer is the social and legal layer, which should aid in creating institutions and distribu-

ting responsibilities when developing AI related regulations (Gasser et al, 2017). These three layers provide a way to approach AI governance that is structured and is designed for regulators to take into account the different complexities that these technologies bring along with them:



(2017)

The second model that serves as part of our theoretical framework is the use of professional norms for the governance of AI. Professional norms are developed by private institutions when incorporating AI into their processes and governments can include them in the existing regulations or build AI governance in accordance with industry-set standards (Gasser et al, 2019). Using norms established by companies looking to set ethical standards can be solid ground on which a sturdy governance model can be built. However, the norms selected need to be carefully analyzed, as there is no guarantee that they will in fact advance the ethical principles that a governance model seeks to fulfill.

Certain strategies for implementing these models can be forms of soft law. Three main advantages of adopting a soft law guide are first that procedures are less complex than in binding rules, second a guide is flexible enough for adapting to technical developments (Walz et al., 2019), and third as it is not binding, companies and entities are less apprehensive towards using AI due to eventual sanctions. As this soft law guide is implemented by the public sector it gains public credibility for the private sector to adopt ethical guides for AI in a second instance (2019).

Nevertheless, it is key to keep in mind that the meaning of what an ethical AI system exactly is remains unclear. According to Peter-Paul Verbeek Chairperson of the UNESCO Commission on the Ethics of Scientific Knowledge and Technology (COMEST) there are three aspects of ethical AI: ethical conception, understandable technology for final consumers and ethical frameworks provided by governments for the use of AI (Maubant, 2019). Ethical AI is also described as a system aligned with values and the anticipation of potential bias and unethical results (Hian Teng, 2020). However, there is a lack of clarity on what the exact materialization of an ethical AI system, product or service is.

2.

The path towards implementing ethical AI in Latin America, beyond theoretical principles

A. Can ethics be a tool to address AI challenges?

Artificial intelligence represents opportunities but also particular risks for Latin American countries, which is why most countries from the region have adhered to international AI ethical principles. In terms of specific opportunities, AI has the potential to add one percentage point to economic growth in Latin America (Ovanessoff & Plastino, 2017). For instance, according to an Argentinean think tank artificial intelligence could contribute to an increase of 4,4% of economic growth in the next decade (Plaw, Carvalho & Ramírez, 2021).

However, artificial intelligence also represents significant risks in Latin America related to discrimination, civil liberties, and security. Regarding discrimination, the training of AI models with data which is not representative of the population can generate discrimination risks. According to the results of a study from 2018, the error rate for white male was 8% while the error rate for women of color was 34,7% for a system that

was trained mainly with individuals of white skin (Buolamwini y Gebru, 2018). This risk is particularly high in Latin America as most systems are trained in high-income countries, and the region is ethnically diverse (Red MTCI Américas). These unethical uses of AI causing discrimination can cause marginalization and also deepen inequalities, in the continent with the highest inequality rates in the world (Lissardy, 2020). This is the main reason why countries need to implement practical methods towards ethical AI and go beyond theoretical principles.

Artificial intelligence also represents risks in terms of civil liberties, regarding the right to privacy but also the right to freedom. As summarized in an article for Agência Brasil, journalist Jonas Valente gathers cases from different countries in the regions in which individuals' privacy was bridged. For instance, in Argentina the local administration of the Salta's Province launched a Microsoft platform seeking to identify women with early pregnancy based on the non-consensual use of personal data such as their name and address (Valente, 2020). In Brazil the right to freedom was compromised as a woman was imprisoned when a system of security cameras in Rio de Janeiro confused her with a fugitive (2020).

Concerning security, although cyberattacks are not exclusively due to AI, the expansion of this technology increases the exposure of personal data (Cabrol, González, Pombo & Sánchez, 2020). In Latin America there has been an increase in personal data leakage (Cabrol, González, Pombo & Sánchez, 2020). For instance, in Mexico in 2016 the information from 94 million people from the National Electoral Institute became exposed (Baraniuk, 2016).

As ethical risks derived from artificial intelligence are every day more present in Latin America it is essential for the region to insure an ethical implementation of AI. In this optic, countries need to promote mechanisms for managing and mitigating risks (Cabrol, González, Pombo & Sánchez, 2020). To this is added that the lack of mechanisms to ensure an ethical implementation of artificial intelligence can dissuade organizations from implementing the technology because of concerns related to bias, privacy and accountability (Walz & Firth – Butterfield).

As a result of the increasing risks of artificial intelligence, there is a global trend highlighting the relevance of ethics in the field of AI governance. The adoption of ethical principles and frameworks across the world evidences this global consensus. According to an informal survey, 10 sets of ethical principles were proposed by the end of 2017 (Winfield & Jirotko, 2018). Forty-two countries signed up to the OECD Principles on Artificial Intelligence. Likewise, UNESCO is seeking to establish along with

193 countries the ethical foundations of artificial intelligence. In the last ten years there has been a considerable increase in the ethical frameworks and principles (Lara, 2020), and there have been more than 84 public – private initiatives also presenting principles and guides for an ethical AI implementation (Mittelstadt, 2019). Latin American countries are increasingly adopting international principles. For instance, Argentina, Brazil, Colombia, Costa Rica and Peru, among others, adopted the OECD Principles on Artificial Intelligence. However, beyond adopting the AI principles, most countries in the region do not have a clear approach on how to promote and implement ethical AI (Cabrol, González, Pombo & Sánchez, 2020).

Hence, ethics do appear to be a tool towards addressing AI challenges and risks. However, bridging the gap between principles and practical implementation remains an unsolved challenge for Latin America, as well as globally.

B. The challenges when bridging the gap between principles and practical implementation

While there is a global awareness of the need for ethical AI, there is no general clarity around concrete ways to implement it. In other words, while international awareness of the relevance of ethical principles for AI has increased, evidence is insufficient as to whether this awareness has translated into practice through effective ethical governance (Winfield & Jirotko, 2018). AI has been implemented in several countries of the region, however there is a lack of ethical frameworks as well as of governance and regulatory approaches guiding the ethical use of AI (Universidad de San Andrés, 2020).

In addition, firms in high – income and in developing countries have taken different approaches in seeking to incorporate ethical principles, such as establishing their own high-level ethical principles, conducting ethical impact assessments (CBI, 2019) and setting ethical subcommittees (Cio from IDG, 2019). However, these initiatives are scattered and there is a lack of governance approaches orienting public and private organizations to effectively address ethical challenges and to insure responsible artificial intelligence. Furthermore, according to a study conducted by the Association for Computing Machinery measuring the influence of codes of ethics in the development of ethical softwa-

re, there is no evidence showing that the existence of a code of ethics influences the decisions made by software developers (Gálvez, 2020). This situation of increasing awareness of the importance of ethical AI and development of ethical principles could lead to 'ethics washing' in which AI principles and governance initiatives continue, while in practice unethical uses of AI continue (Daly et al., 2019).

There are five main factors explaining the gap between principles and concrete implementation. First, the lack of a global consensus of what ethical AI systems actually are, as discussed in the theoretical framework section of this paper.

The second factor is that the characteristics of AI and mainly the "black box" problem represent specific ethical challenges compared to other technologies. The "black box" issue consists of the possibility of understanding inputs and outputs but not the process of how inputs bring to outputs (Web Foundation, 2018). This challenge can become especially pressing in Latin America as most research, development, and innovation in the AI field has taken place in other regions, which has resulted in knowledge gaps (Cabrol et al., 2020). The missing knowledge can hinder opportunities for understanding the black box. It can also turn governments reliant on private black box systems that threaten transparency, which is key for democratic regimes and can affect local governance, due to the growing need of implementing this technology. It is for instance the case in Uruguay where according to the Web Foundation the government relies on a private black box system in the process of allocating police resources (Web Foundation, 2018).

The third factor is the disconnection between policy makers, who are generally aware of the importance of ethics, and AI developers and implementers, whose priorities are oriented towards innovation and economic development. In the same optic there is a risk of AI projects being designed and developed based on priorities regarding technology progress instead of seeking to address social issues. This disconnection can be exacerbated in Latin America given that the design and development of AI systems generally takes place in high – income countries (Cabrol et al., 2020).

The fourth factor potentially explaining this gap is insufficient education in AI ethics for designers and implementers of this technology. Academia has found that this factor can also explain why ethics has received little attention in engineering literature (Vakkuri et al., 2020). While ensuring that technical dimensions of an AI system follow ethical parameters is essential, they remain insufficient for ethical AI (Walz, et al., 2019). As AI systems are designed, constructed, programmed and implemen-

ted by human beings', education in ethics is a structural need to insure ethical AI systems (2019).

The fifth factor explaining why there is a gap between ethical principles and ethical AI is that principles remain at a theoretical level without practical implementation tools and accountability mechanisms (Lara, 2020). In Latin America, the 'Ethical Framework for Artificial Intelligence in Colombia' is the only policy document describing concrete tools to implement ethical principles (Guio, 2020). This document not only establishes nine principles to guarantee the ethical design, implementation, and use in Colombia, but also provides eight specific tools to ensure the AI systems being used by the government comply with the established principles (2020). For instance, one of the tools is the national algorithm registry, a system inspired by initiatives previously launched in Amsterdam and Helsinki (Johnson, 2020). The national algorithm registry for ethics is a platform where all public entities using AI systems must report their algorithms as well as the steps they are taking to ensure the system is aligned with the ethical principles outlined in the Framework (Guio, 2020). This initiative will be further explored in the section that reviews Colombia's efforts towards ethical AI. While this represents the most practical approach to implementing ethical principles in the region today, it is not a perfect solution, blind spots as to the risks of AI remain.

Consequently, while there is a pressing need for Latin American countries to implement an ethical and responsible AI, countries encounter challenges affecting their implementation opportunities. To overcome this difficulty Colombia has implemented a strategy which includes a long term collaborative governance approach.

3.

The Colombian example: Collaborative Governance as a path to bridging the gap

A. The collaborative governance approach

Collaborative governance between the public and private sector can be an alternative to implement ethical AI. Collaborative governance consists of “each party (public and private) has a hand in defining not only the means by which a goal is achieved but the details of the goal itself” (Donahue et al., 2006). The private sector includes not only private companies but NGO and civil society as well.

Collaborative governance has five main advantages. The first advantage of the collaborative approach lies in overcoming information asymmetries. As private companies have increasing knowledge, expertise and capabilities in artificial intelligence (Quélin et al., 2017) their involvement is key for the public sector to learn about market failures and identify the areas where their intervention is the most needed (Fernández-Arias et al., 2016). Collaborative governance is an asset for overcoming the challenge of fragmentation between public and private entities, and technical knowledge, previously presented. Overcoming information asymmetries is especially relevant in the field of artificial intelligence where the pace of technical evolution compared to regulatory evolution is faster than in other sectors (Malyshev et al., 2019).

The second advantage of collaboration between the public and the private sector is legitimacy. Collaboration between the public and private sector can increase legitimacy, in tasks where the exclusive action from the government can be perceived as inappropriate (Donahue et al., 2006). In artificial intelligence, undertaking this approach is especially relevant as private companies are the ones that develop and implement the technology in most cases.

The third advantage is agility. Artificial intelligence is developing at an unattainable pace for traditional regulation led exclusively by the government (Walz et al., 2019). Traditional regulation has proven to not achieve the rapid pace of change for disruptive technologies, and so working together with the private sector becomes a suitable alternative to keeping up with the latest developments of AI. Collaborative governance in which governments cooperate with the private sector developing AI is essential for governments to protect citizens up-front instead of reactively (2019).

The fourth advantage is that collaborative governance also represents an asset for companies to implement ethical principles. By implementing ethical principles and collaborating with governments companies,

R&D efforts are not wasted on systems that will be later constrained by regulations as they go against citizen's protection (Walz et al., 2019).

The fifth advantage is that actively engaging in collaborative governance by contributing to social benefit and citizens' well-being represents a competitive advantage for companies in the market as their products or services are more human-centric and therefore more accepted by customers and society (Walz et al. 2019)

B. The role of collaborative governance in Colombia's efforts towards ethical AI

Colombia is listed as the first Latin American country to adopt an Ethical Framework for the implementation of artificial intelligence (OECD, 2021). An analysis of the Ethical Framework can show a collaborative governance approach for implementing ethical AI, considering that this document uses collaboration between the public and the private sector as one of the vectors of Colombia's digital transformation strategy. This innovative approach could be furthered and oriented towards ethical artificial intelligence. Colombia has used three main tools for collaborative governance when it comes to ethical AI: the discussions around the Ethical Framework for Artificial Intelligence, the Ethical Algorithm Register, and the hiring of an ethicist for the AI Task Force within the President's office.

The Ethical Framework for Artificial Intelligence in Colombia

The 'Ethical Framework for Artificial Intelligence in Colombia' (Guio, 2021) can be considered a milestone in Colombia's trajectory in promoting the implementation of ethical AI. The Ethical Framework responds to the need of a guide for implementing ethical AI in the public sector (Flórez et al., 2020). This Framework constitutes a form of soft law, as it leaves room to adapt to new technological developments, and as it remains non-binding, entities are more likely to use and develop AI systems with a problem-solving approach. It is also built upon recommendations on how to deal with AI challenges as they may come up: the Framework recommends that the government develop strategies of education and research on AI ethics; that sustainable and trustworthy AI systems are

designed through prioritizing technical objectives such as trustworthiness, security, and solidity, in order to ensure that AI systems are safe; and that ethics are strengthened in human rights programs (Guio, 2021).

The Framework bases its tools for implementing ethical AI upon nine principles. The nine principles included are: transparency and explanation, privacy, human control of AI systems' decisions, security, responsibility, non discrimination, inclusion, prevalence of the rights of children, and social benefit (Guio, 2020). Each principle includes three dimensions: data ethics, ethics of algorithms and ethics of practices. These dimensions envisage how principles can be implemented not only by AI designers and developers but also by AI implementers (Guio, 2021). The transparency and explanation principles are key for citizens to understand AI systems and as a result participate in their conception and implementation.

Colombia used a collaborative governance approach for the process of creating and implementing the Framework. For the development of the document, the government created a first draft, which was the product of the analysis and assessment of other efforts in this field all over the world. The government then proceeded to arrange discussion tables around this first draft, where different sectors of society weighed in on their needs and concerns. Each of these discussion tables reached a set of conclusions that were implemented into what is now the Ethical Framework.

International organizations, NGOs, representatives of the private sector, students and university directives, participated in these discussion tables in which they were able to share their experiences, make comments and provide feedback on the framework. As artificial intelligence impacts multiple sectors of society, considering different visions and actors with different interests is essential for the framework in order to take into account different priorities (Castaño, 2020).

The private sector also had the opportunity of participating in a technical table which opened space for collaborative governance because it allowed participating regulators to understand where certain tools could hinder innovation if they were implemented. In the same optic, participation from the private sector is also relevant for implementation since it puts the challenges companies face when implementing the ethical framework in the forefront and builds governance systems with these challenges in mind.

The education sector also provided comments and feedback about the framework in another technical table. This point of view was essential

as representatives from universities highlighted the importance of incorporating ethics into artificial intelligence curriculums, which can become an asset to overcome the disconnection between policy makers and developers, and also the need of ensuring that the human beings behind AI act based on ethical parameters (Walz et al., 2019).

In the process of adopting the Framework, Colombia has sought to promote multi-stakeholder participation. To do this, the government has implemented tools outlined in the Framework, such as the Ethical Algorithm Register. Another tool is the implementation of an AI Task Force within the presidency's office which includes an Ethicist, as well as orienting specific AI projects from the public sector towards social benefit. We will now take a look at each of these tools in more detail.

The Ethical Algorithm Register

As formerly mentioned, the Framework also establishes concrete implementation tools as algorithm assessment, smart explanation and definition and risk management (Guio, 2021). For instance, it encourages entities that use the framework for their work in AI to use the Ethical Algorithm Register, a platform in which entities report their use of AI in projects, which allows the government to monitor advances of the ethical implementation of AI and reinforces citizen participation, as users have access to a platform on which they can submit questions, concerns, or good practices. Entities are also expected to perform tasks such as algorithm assessment, to map the state of ethical principle implementation for a particular AI system. Data cleansing is another of the measures included in the Framework, which aims to limit prejudice and errors in the data used for the development of AI systems. Other tools listed in the framework include intelligent explanation, legitimacy assessment, ethical risks definition and management, internal codes of conduct and ethics, and privacy impact assessments (Guio, 2021).

The Ethical Algorithm Register is also a key asset for ensuring that AI ethical principles are implemented in practice and for monitoring ethical AI implementation. The Colombian algorithm register is based on the models of Amsterdam (City of Amsterdam Algorithm Register Beta, 2020) and Helsinki (City of Helsinki AI Register, 2020) seeking to ensure transparency in the public implementation of AI (Johnson, 2021), but at a national level. In the registry all AI projects provide a short description of the project, how AI is being implemented, what are the ethical risks of

the use of AI in the project and how they are being mitigated, how non-discrimination is insured and the contact details of the entity managing the project.

The first advantage of the Ethical Algorithm Register is that as the information from all AI projects in the public sector is published, entities have an incentive to implement effective tools to guarantee the ethical implementation of AI. The noncompliance with ethical principles and tools does not lead to sanctions (Lara,202), but instead there is a positive “competition” between entities motivating them to implement ethical principles without needing to regulate. The ethical reputation of projects in the eyes of the public becomes essential.

The second advantage is citizen’ engagement. As all projects have to provide contact information there is a possibility of direct engagement between citizens and AI implementers. Citizen engagement also generates an accountability mechanism from projects towards citizens, as citizens can provide feedback concerning the ethical implementation of AI and projects need to justify their ethical implementation of the technology, without the need of introducing sanctions. Citizen’ monitoring and watch over the implementation of the AI projects in the Ethical Algorithm Register is also a way of implementing human oversight. Accountability also contributes to building citizens’ trust in AI (Floridi, 2020). Finally, the algorithm register increases citizen’s understanding of what AI is and the impact it is having on their daily lives (Floridi, 2020).

The AI Task Force

Colombia has also taken relevant steps in the path of implementing ethical AI through the inclusion of an ethicist as a key member of the AI Task Force (AI Project). Colombia is launching an AI Task Force among the Presidency of the Republic in charge of the implementation of the AI policy and facilitating the use of AI in the public sector (OECD, 2021). One of the functions of the ethicist is ensuring the implementation of ethical AI in the different projects in which the Office participates. This individual could also be in charge of evaluating and managing the ethical risks of AI deployment, decide over the appropriate level of human intervention in decisions making influenced by AI, managing the training and selection process of AI models, monitor the AI models that have been implemented in order to take the remediation measures if necessary, and revising communication channels to promote feedback con-

cerning ethical components of AI projects (Guio, 2020). This is in line with other global initiatives, such as Singapore's AI governance model, which includes the Advisory Council on the Ethical Use of AI and Data (OECD, 2021).

Colombia is also seeking to implement ethical AI in concrete projects, where the use of AI is mainly oriented towards social benefit (one of the principles of the Framework). For instance, concerning agriculture development, Agrosavia the Colombian Corporation of Agricultural and Livestock Research has a project using AI for improving soil quality and implementing fertilization plans (Muñoz et al., 2021). Farmers send the characteristics of the soil to Agrosavia that digitizes the information, analyses it through AI algorithms and sends recommendations back to the farmer (2021). Another example is the use of AI in the children's adoption field. The institute in charge of adoption at a national level uses an AI system in the interview process in interviews conducted by the psychosocial teams in evaluating the suitability of the applicant family or individual (ICBF ADA, Govcodashboard). Several projects of this sort have been implemented in which AI is used for social benefit. There is an opportunity to deepen the adoption and evaluation of ethical principles and tools in these projects through the algorithm registry, which will be presented in the next part.

C. More tools for collaborative governance and ethical AI

While Colombia seems to have successfully embarked on the path towards collaborative AI governance, there are opportunities for the country to deepen its efforts in order to ensure a practical implementation of ethical AI. The main mechanisms for concretely implementing ethical artificial intelligence include regulatory sandboxes, policy prototyping, implementation of actionable principles for AI, use of co-creation spaces, ethics curriculum development, and activation of stakeholder expertise. .

Regulatory sandboxes

Regulatory sandboxes, which Colombia has started to implement, are a way of deepening the collaborative governance approach and of concretely implementing ethical AI. As AI is in permanent evolution, ethical AI must evolve at an agile pace which is why experimentation is essential. Regulatory sandboxes apply collaborative governance as the risks of unethical use of AI rise in a controlled environment allowing the government and companies to take the appropriate measures.

Regulatory sandboxes are also an opportunity to deepen the implementation of ethical artificial intelligence as they allow to identify potential ethical risks at an early stage, following the example of Norway (Iapp, 2020). The Norwegian Data Protection Agency is establishing a regulatory sandbox seeking to promote innovation from an ethical approach to data. This sandbox provides companies with a perspective on personal data protection in order to adjust their products and ensures that the artificial intelligence programs that are being developed respect fundamental rights.

Colombia developed the 'Conceptual model for the design of regulatory sandboxes & beaches in AI'. In the framework of this model for instance, the Regulatory FinTech Sandbox allows financial institutions to test cash-in (deposit of funds) and cash-out (withdrawal of funds) transactions through exchange platforms (cryptocurrencies). The sandbox has helped in reducing information asymmetries as the Fintech ecosystem has the chance to analyze the advantages and the risks of these transactions (Colombia Fintech, 2020). Through developing this sandbox and creating systems for regulatory sandboxes to exist within the AI ecosystem in Colombia, the country has an opportunity to reinforce the monitoring of AI ethics using this experimental regulatory model.

Along with regulatory sandboxes, it is possible to implement similar structures for determining the accurate ethical measures required for the development and deployment of AI systems. Regulatory experimentation spaces are very attractive for fast-paced technologies like AI because they allow for innovative and interdisciplinary developments that traditional regulatory paths can overlook (Ranchordas, 2021). An example of this has been implemented in Norway, in which private companies and public entities can access free guidance on personal data protection. The main goal behind this sandbox is to develop responsible and trustworthy AI, focusing mainly on the use of data, which, as we have seen, is the foundation of any AI system. (Olsen, 2020).

Policy prototyping:

Policy prototyping is also an instrument for ensuring that artificial intelligence developments are ethical and responsible. Policy prototyping takes place in a controlled environment in which policies are tested as well as their impact on citizens and on business models (Castaño, 2020). Policy prototyping allows to identify the costs and challenges for companies of implementing ethical AI (Mittelstadt, 2019). This identification process is key for implementation since companies are already aware of the challenges they will address. Based on the results of the applying policy prototypes in a controlled environment it is possible to adjust policies given the risk they represent for society and business (Gomes de Andrade, 2021). These adjustments ensure that AI developments follow ethical principles.

A concrete example of prototyping for AI ethics is a prototyping experiment developed by the Open Loop program in Finland from which they drew the following recommendations for lawmakers (Gomes de Andrade, 2021):

- 1.** Determining high-risk AI applications is better through procedure rather than prescription: a codified risk assessment procedure can help entities assess risk based on their specific context (beyond blindly prescribing assumed solutions to mitigate risks) which will allow for a more accurate result (2021).
- 2.** Using this procedural risk assessment will then help determine the necessary regulatory requirements (2021).
- 3.** Developing concrete guidelines when implementing the regulation helps to overcome ambiguities in the regulation and provides clarity for entities on how to act (2021).
- 4.** Specificity in the definition of risks aids in preventing uncertainties when understanding how AI systems can impact individuals and society. Through collaboration with academia, civil society, and industry, policymakers can develop clear guidelines to identify the specific types of risks and harms that entities will need to look out for (2021).
- 5.** Documentation of risk assessment and decision-making processes needs to include justification for mitigation processes in order to determine how effective the implemented risk management measures were and what can work (or not) for other entities in the future (2021).
- 6.** Developing a sound taxonomy of the different actors involved in risk assessment can help to appropriately assign who is in charge of identifying, assessing, and mitigating risks, and to better understand the stakeholders affected by AI (2021).

7. Identifying the values that may be impacted by AI implementation as well as the values that may be in tension with one another can aid in better decision-making.
8. Combining new processes with established ones will improve the overall approach, instead of having multiple disconnected initiatives that can later lead to inefficiency and inaccuracy when assessing risks (2021).

Actionable Principles for AI:

These principles are meant to aid in the practical implementation of ethics principles, as they often fail to provide concrete ways to implement them. AI ethics principles are essential, but they need to be grounded into actionable methodologies that allow concrete action (Stix, 2021). It is important to highlight that actionable principles are not meant to replace ethics principles, but rather to contribute to working towards established goals (2021).

An example of an actionable principle is human control and oversight of AI systems, which makes part of Colombia's AI Ethics Framework (Guio, 2021). This principle brings along concrete actions and initiatives, such as the National Ethical Algorithm Registry discussed earlier in this paper, instead of simply naming a characteristic that would make an AI system more trustworthy. Most of the ethical principles used around Latin America (including some in Colombia's Framework) cannot be considered actionable, as they are not grounded on methodologies that permit concrete action.

Co-creation spaces:

Collaboration among the public and private sectors is an essential part of reaching ethical and trustworthy AI systems. The business trend of co-creation has now been implemented in government spaces: "In a co-creation effort, multiple stakeholders come together to develop new practices that traditionally would have emerged only from a bureaucratic, top-down process (if, indeed, those practices would have emerged at all). Change, moreover, occurs not just at the level of an organization, but also across an entire value chain" (Gouillart, 2015).

However it is key to keep in mind that when new spaces for opportunity are created, some groups benefit while others are left behind (Leino, 2020): “(...) citizens are not all the same: they fall in several categories and some of them have more resources to participate in co-creative processes than others. Thus, it is relevant to know who participates in and whose voices get heard through these processes” (2020). This can happen for a multitude of reasons, including citizens or institutions not being aware of initiatives in time, or deciding not to participate for several reasons.

Ethics curriculum development

As it was previously stated, decisions regarding AI are often taken by engineers with limited education in ethics (Walz et al., 2019). Colombia has begun to build awareness of the need for ethical courses in education programs, as it was highlighted in the technical table led by the education sector regarding the Ethical Framework for Artificial Intelligence in Colombia. However, there is a need for universities to implement courses in AI ethics structurally related with the core of AI programs. The repository of material to include social inquiry in AI programs that the World Economic Forum Global Future Councils on Artificial Intelligence and Robotics is creating can be an asset (2019). Implementing courses in AI ethics in different curriculums is also an asset for the long term in guaranteeing sustainable AI from a social, environmental and economic perspective.

Activating stakeholder expertise.

Expertise is not only developed within academia: it can also exist in civil society organizations and in the industry. For instance, one way in which public entities can access this knowledge to later use it to develop better regulations is launching consultation processes: for instance, the European Parliament launched one of these processes in which it asked individuals, organizations, and corporations about the future of robotics and AI, with questionnaires around ethics, liabilities, safety and security, and institutional oversight (ITU, 2018). Another example of similar initiatives was the White House AI Workshop Series, in which the US government invited experts to contemplate safety and control, legal issues, and social good in events all over the country (2018).

4.

Conclusions

As Latin American countries face increasing ethical risks related to AI, they adhere to the global trend of adopting an ethical approach for the design, development, and implementation of artificial intelligence. However, bridging the gap between theoretical principles and implementing ethical AI in practice has continued to be a global challenge which Latin American countries face as well.

In the case of Colombia, the country has developed a practical tool in the Ethical Framework through a collaborative governance approach, which can potentially be an asset in overcoming this challenge. There are opportunities for the country to deepen collaborative governance, based on the steps it has already taken, to ensure effective implementation of ethical AI in concrete projects through regulatory sandboxes, policy prototyping, education on AI ethics, and gathering stakeholder expertise.

Taking into account the actions Colombia has already taken towards implementing ethical AI, as well as the overview of other tools that can help countries reach ethical AI through collaborative governance, we have put together the following recommendations that can help Colombia move forward in its goals of reaching ethical AI and working towards digital transformation. This can be achieved through a combination of regulatory sandboxes specifically for AI ethics, policy prototyping, actionable principles for AI, co-creation spaces, ethics curriculum development, and activating stakeholder expertise:

- ◆ The Colombian government should consider implementing an initiative similar to the Norwegian Data Protection Agency's ethics regulatory sandbox, in order to advance its purpose of reaching ethical AI standards.
- ◆ Colombia has the potential to deepen its ethical approach towards artificial intelligence by deepening its collaborative governance strategy, and more concretely by implementing regulatory sandboxes and policy prototyping, aiming to ensure ethical and responsible artificial intelligence. Implementing policy prototyping would be an asset for deepening collaborative governance and ensuring the implementation of ethical AI from the private sector. Policy prototyping would allow private companies to test the implementation of the 'Ethical Framework for Artificial Intelligence in Colombia' in a controlled environment and based on the conclusions of the experimental exercise adapt the Framework in order to draft a version for the private sector.

- ◆ The Colombian government should develop a set of actionable principles to strengthen its approach to AI governance, taking into account technical and non-technical measures: technical measures address ethics principles like explainability and privacy, while non-technical measures address principles like citizen participation.
- ◆ By implementing co-creation spaces to regulate AI, the Colombian government will gain access to the expertise in academia and in the industry and be able to develop accurate regulations in a more efficient manner.
- ◆ The Colombian government needs to be weary of creating further gaps when developing co-creation spaces, as well as to look for ways to include those who are excluded under inevitable circumstances. Co-creation spaces are not inherently inclusive, and this needs to be at the forefront of their design and risk mitigation.
- ◆ It is essential for Colombia to provide ethics courses as a structural dimension of AI curriculums.
- ◆ We believe the Colombian government would benefit greatly from implementing a strategy to gather stakeholder expertise from several entities.

These recommendations are extremely relevant to the current discussion taking place among the international community. For instance, UNESCO has developed a Recommendation on the ethics of artificial intelligence, which is meant to serve as a guide for member states when regulating AI (UNESCO, 2021). Aligned with this Recommendation, we believe the Colombian government should take sturdier action when regulating AI, as it pertains to ethical risks.





Works Cited

Baraniuk, C. (2016) 'Millions of Mexican voter records" were accessible online". BBC [Online]. Retrieved from: <https://www.bbc.com/news/technology-36128745> [Accessed 5 April 2021]

Cabrol, M. González, N. Pombo, C. Sánchez R. (2020) 'Adopción ética y responsable de la Inteligencia Artificial en América Latina y el Caribe'. BID fAlr LAC [online]. Retrieved from: https://publications.iadb.org/publications/spanish/document/fAlr_LAC_Adopci%C3%B3n_%C3%A9tica_y_responsable_de_la_inteligencia_artificial_en_Am%C3%A9rica_Latina_y_el_Caribe_es.pdf [Accessed 6 April 2021]

Castaño, D. (2020) 'La gobernanza de la Inteligencia Artificial en América Latina: entre la regulación estatal, la privacidad y la ética digital'. CETyS and Universidad de San Andrés. Retrieved from: https://guia.ai/wp-content/uploads/2020/07/Casta%C3%B1o-La-gobernanza-de-la-Inteligencia-Artificial-en-América-Latina_compressed.pdf [Accessed 7 April 2021]

CIO from IDG. '10 tendências em Dados e Analytics para CIOs ficarem de olho'. Available from: <https://cio.com.br/tendencias/10-tendencias-em-dados-e-analytics-para-cios-ficarem-de-olho/>

City of Amsterdam Algorithm Register Beta (2021) [online]. Retrieved from: <https://algoritmeregister.amsterdam.nl/en/ai-register/> [Accessed 7 April 2021]

City of Helsinki AI Register (2021) [online]. Retrieved from: <https://ai.hel.fi/en/ai-register/> [Accessed 7 April 2021]

Daly, A. Hagendorff, T. Li, H. Mann, M. Marda, V. Wagner, B. Wang, W. Witteborn, S. (2019) 'Artificial Intelligence, Governance and Ethics: Global Perspectives'. Faculty of Law The University of Hong Kong [online]. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3414805 [Accessed 5 April 2021]

Donauhe, J., Zeckhauser, R. (2006) 'Public – Private Colaboration'. Oxford University Press. Pp. 496 – 525. Retrieved from: https://scholar.harvard.edu/files/rzeckhauser/files/public-private_collaboration.pdf [Accessed 20 March 2021]

Fernández-Arias, E. Sabel, C., Stein, E., Trejos, A., 'Two to Tango. Public-Private Collaboration for Productive Development Policies'. Interamerican Development Bank. Retrieved from: https://books.google.com.co/books?hl=fr&lr=&id=0OZ3DwAAQBAJ&oi=fnd&pg=PR6&dq=research+about+advantages+of+public+private+collaboration&ots=sA6hCoR-_1&sig=krghSA5APxHWGtKG_8enHAXonKw#v=onepage&q=research%20about%20advantages%20of%20public%20private%20collaboration&f=false

Flórez, M.L. & Vargas, J. (2020) 'El impacto de herramientas de inteligencia artificial: un análisis en el sector público en Colombia'. CETyS and Universidad de San Andrés [online]. Retrieved from: <https://guia.ai/wp-content/uploads/2020/05/GECTI-El-impacto-de-herramientas-de-inteligencia-artificial.pdf> [Accessed 5 April 2021]

Floridi, L. (2020) 'Artificial Intelligence as a Public Service: Learning from Amsterdam and Helsinki'. Philosophy & Technology. DOI: 10.1007. Retrieved from: <https://doi.org/10.1007/s13347-020-00434-3>

Fourtané S. 'Ethics of AI: Benefits and Risks of Artificial Intelligence Systems'. Interesting Engineering [online]. Retrieved from: <https://interestingengineering.com/ethics-of-ai-benefits-and-risks-of-artificial-intelligence-systems> [Accessed 18 April 2021]

Gasser, U., & Almeida, V. (November 2017). "A Layered Model for AI Governance." IEEE Internet Computing 21 (6) : 58–62. doi:10.1109/mic.2017.4180835.

Gasser, U., & Schmitt, C. "The Role of Professional Norms in the Governance of Artificial Intelligence" (April 25, 2019). Forthcoming in: Markus D. Dubber, Frank Pasquale, and Sunit Das (eds.), *The Oxford Handbook of Ethics of AI*, Oxford University Press, Available at SSRN: <https://ssrn.com/abstract=3378267> or <http://dx.doi.org/10.2139/ssrn.3378267>

Gómez, C., May, C., Martínez, C. & Martín, A.M. (2020) 'Artificial Intelligence for Social Good in Latin America and the Caribbean: the Regional Landscape and 12 Country Snapshots'. IDB fAlr LAC [online]. Retrieved from: <https://publications.iadb.org/publications/english/document/Artificial-Intelligence-for-Social-Good-in-Latin-America-and-the-Caribbean-The-Regional-Landscape-and-12-Country-Snapshots.pdf> [Accessed 5 April 2021]

Gomes de Andrade N. & Kontschieder V. (2021) 'AI Impact Assessment: A Policy Prototyping Experiment'. Open Loop. PP 7 – 57

Gouillart F. & Hallett T. (2015). 'Co-Creation in Government'. Stanford Social Innovation Review. [Online]. Retrieved from https://ssir.org/articles/entry/co_creation_in_government [Accessed June 30 2021]

Guio, A. (2021) 'Consejo Internacional de Inteligencia Artificial para Colombia'. Colombian Presidency. Retrieved from: <https://dapre.presidencia.gov.co/AtencionCiudadana/convocatorias-consultas/consejo-internacional-de-inteligencia-artificial-para-colombia> [Accessed 5 April 2021]

Guio, A. (2020) 'Task Force for the Development and Implementation of Artificial Intelligence in Colombia' Retrieved from: <https://dapre.presidencia.gov.co/TD/TASK-FORCE-DEVELOPMENT-IMPLEMENTATION-ARTIFICIAL-INTELLIGENCE-COLOMBIA.pdf> [Accessed 10 December 2020]

Guio, A. (2020) 'Marco Etico para la Implementación de la Inteligencia Artificial en Colombia'. Colombian Presidency [online]. Retrieved from: <https://dapre.presidencia.gov.co/TD/MARCO-ETICO-PARA-LA-INTELLIGENCIA-ARTIFICIAL-EN-COLOMBIA.pdf> [Accessed 10 December 2020]

Iapp (2020) 'Norwegian DPA creating regulatory sandbox for AI' [online]. Retrieved from: <https://iapp.org/news/a/norwegian-dpa-creating-regulatory-sandbox-for-ai/> [Accessed 6 April 2021]

ICBF (2021). Govco Dashboard. Presidency of the Republic of Colombia [online] Retrieved from: <https://govcodashboardcovid19.shinyapps.io/ProyectosTD/> [Accessed 18 April 2021]

Inaa Group. International Association of Independent Accountants. (2020). 'How AI is the Key to Economic Growth Beyond COVID -19' [online]. Retrieved from: <https://www.inaa.org/how-ai-is-the-key-to-economic-growth-beyond-covid-19/> [Accessed 18 April 2021]

ITU. (2018). 'Module on Setting the Stage for AI Governance: Interfaces, Infrastructures, and Institutions for Policymakers and Regulators'. Artificial Intelligence (AI) for Development Series. [online] Retrieved from https://www.itu.int/en/ITU-D/Conferences/GSR/Documents/GSR2018/documents/AISeries_GovernanceModule_GSR18.pdf [Accessed 30 June 2021]

Jercich K. (2021) 'AI has advantages for COVID – 19 vaccine, rollout, but potential dangers too'. Healthcare IT News [online]. Retrieved from: <https://www.healthcareitnews.com/news/ai-has-advantages-covid-19-vaccine-rollout-potential-dangers-too> [Accessed 18 April 2021]

Johnson, K. (2020) 'Amsterdam and Helsinki launch algorithm registries to bring transparency to public deployments of AI' [online]. Retrieved from: <https://venturebeat.com/2020/09/28/amsterdam-and-helsinki-launch-algorithm-registries-to-bring-transparency-to-public-deployments-of-ai/> [Accessed 6 April 2021]

Kumar, V. (2020) 'What is the current state of omnipresent AI across business' Analytics Insight [online]. Retrieved from: <https://www.analyticsinsight.net/ai-everywhere-prepare/> [Accessed 6 April 2021]

Lara, J.C. (2020) 'Perspectivas de derechos humanos en la implementación de marcos éticos para la inteligencia artificial'. CETyS Universidad de San Andrés [online]. Retrieved from: <https://guia.ai/wp-content/uploads/2020/05/Lara-Perspectivas-de-derechos-humanos-en-la-implementacion-de-marcos-eticos-para-la-inteligencia-artificial.pdf> [Accessed 20 March 2021]

Lissardy, G. (2020) 'Por qué América Latina es 'la región mas desigual del planeta'. BBC [online]. Retrieved from: <https://www.bbc.com/mundo/noticias-america-latina-51390621> [Accessed 6 April 2021]

Malyshev, N & Kauffman, C. (2019) 'Regulatory effectiveness in the era of digitalization'. OECD [online]. Retrieved from: <https://www.oecd.org/gov/regulatory-policy/Regulatory-effectiveness-in-the-era-of-digitalization.pdf>

Maubant, T. (2019) 'L'éthique, les principes de l'IA et la pratique au coeur de la Conférence de l'UNESCO "Youth voices and the future of Artificial Intelligence"'. ActuaA [online]. Retrieved from : <https://www.actuaia.com/actualite/lethique-les-principes-de-lia-et-la-pratique-au-coeur-de-la-conference-de-lunesco-youth-voices-and-the-future-of-artificial-intelligence/> [Accessed 6 April 2021]

Mittelstadt B. (2019) 'Principles alone cannot guarantee ethical AI'. Oxford Internet Institute. SSRN. DOI: 10.2139

Muñoz, V. Hormechea, C. (2021) '¿Cómo vamos avanzando en la Transformación Digital?' Retrieved from: <https://dapre.presidencia.gov.co/TD/Como-vamos-avanzando-en-la-TD-070421.pdf> [Accessed 12 April 2021]

OECD (2021) 'National AI Policies and Strategies'. [online] Retrieved from: <https://www.oecd.ai/dashboards> [Accessed 9 April 2021]

Olsen, B.K. (December 14, 2020). 'Sandbox For Responsible Artificial Intelligence'. Available on <https://data-ethics.eu/sandbox-for-responsible-artificial-intelligence/> [Accessed 30 June 2021]

Oxford Insights (2019) 'Government Artificial Intelligence Readiness Index 2019'. [online]. Retrieved from: <https://www.oxfordinsights.com/ai-readiness2019> [Accessed 9 April 2021]

Ovanessof, A. & Plastino E. (2017) 'How Artificial Intelligence can drive South America's growth'. Accenture [online]. Retrieved from: https://www.accenture.com/_acnmedia/pdf-49/accenture-how-artificial-intelligence-can-drive-south-americas-growth.pdf [Accessed 6 April 2021]

Quélin, B., Kiveleniece, I., Lazzarini, S. (2017) 'Public – private Collaboration, Hybridity and Social Value: Towards New Theoreticall Perspectives'. Journal of Management Studies. doi.org/10.1111/joms.12274

Ranchordas, S. (May 4, 2021) 'Experimental Regulations for AI: Sandboxes for Morals and Mores'. University of Groningen Faculty of Law Research Paper 1, forthcoming in *Morals and Machines* (vol.1, 2021), Available at SSRN: <https://ssrn.com/abstract=3839744> or <http://dx.doi.org/10.2139/ssrn.3839744>

Red MTCI Américas 'La Diversidad Etnica en la Región de las Américas' [online]. Retrieved from: <https://mtci.bvsalud.org/la-diversidad-etnica-en-la-region-de-las-americas/> [Accessed 5 April 2021]

Stix, C. (2021) Actionable Principles for Artificial Intelligence Policy: Three Pathways. *Sci Eng Ethics* 27, 1. <https://doi.org/10.1007/s11948-020-00277-3>

UNESCO. (2021). "Elaboration of a Recommendation on the ethics of artificial intelligence". [online] Retrieved from: <https://en.unesco.org/artificial-intelligence/ethics> [Accessed 30 June 2021].

Universidad de San Andrés 'Proyecto GulA. Implementaciones éticas de Inteligencia Artificial en América Latina y el Caribe' [online]. Retrieved from: <https://udesa.edu.ar/noticias/proyecto-guia-implementaciones-eticas-de-inteligencia-artificial-en-america-latina-y-el> [Accessed 5 April 2021]

University of Cambridge 'Risks from Artificial Intelligence'. Center for the study of existential risks [online]. Retrieved from: <https://www.cser.ac.uk/research/risks-from-artificial-intelligence/> [Accessed 6 April 2021]

Valente, J. (2020) 'Riscos da inteligência artificial levantam alerta e suscitam respostas'. Agência Brasil [online]. Retrieved from: <https://agenciabrasil.ebc.com.br/geral/noticia/2020-08/riscos-da-inteligencia-artificial-levantam-alerta-e-suscitam-respostas> [Accessed 10 December 2020]

Vakkuri, V., Kemell, K., Jantunen, M., Abrahamsson, P. (2020). 'This is Just a Prototype': How AI Ethics Are Ignored in Software Startup Like Environments'. Springer. Retrieved from: https://link.springer.com/chapter/10.1007/978-3-030-49392-9_13

Walz A., Firth-Butterfield, M. (2019) 'Implementing Ethics Into Artificial Intelligence: A Contribution, From A Legal Perspective, To The Development Of An AI Governance Regime' 18 *Duke Law & Technology Review* 176-231

Winfield, A., Jirotko, M. (2018) 'Ethical governance is essential to building trust in robotics and artificial intelligence systems'. *Phil.Trans.R.Soc.A376*: 20180085. <http://dx.doi.org/10.1098/rsta.2018.0085>. Retrieved from: <https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2018.0085>

World Wide Web Foundation (2018) 'Algorithms and Artificial Intelligence in Latin America A Study of Implementation by Government in Argentina and Uruguay' [online]. Retrieved from: http://webfoundation.org/docs/2018/09/WF_AI-in-LA_Report_Screen_AW.pdf [Accessed 7 April 2021]