GuIA.ai

Responsible Use of Technology in Latin America and the Caribbean

Examples of models and tools from the global North and South. Opportunities and challenges from an ethical and human rights perspective

Elina Castillo Jiménez and
Cristina Martínez Pinto

November 2021

CETyS | LATAM DIGITAL CENTRO DE POLÍTICA DIGITAL PARA AMÉRICA LATINA

In collaboration with: fAIr LAC | BID Banco Interamericano de Desenvolvimento

Responsible Use of Technology in Latin America and the Caribbean

Authors: Elina Castillo Jiménez and Cristina Martínez Pinto

Contributor: Luz Elena González

Coordinators: Carolina Aguerre and Maia Levy Daniel

Revision: Gonzalo Bustos Frati and Matías Jackson

Design: Mónica Castellanos

Translation: Verónica Penelas

Responsible Use of Technology in Latin America and the Caribbean

Examples of models and tools from the global North and South. Opportunities and challenges from an ethical and human rights perspective

November 2021

Cristina Martínez Pinto

Specialist in civic technology, digital transformation and new technologies focusing on innovation and social impact. She is the founder and general director of PIT Policy Lab, the first spin-out of C Minds, where she explores the intersection between technology of public interest and technology public policies. She has experience as an advisor to international agencies and in the Mexican public service. She has a BA in International Affairs from the Tec of Monterrey and a Master's Degree in Public Policies from the University of Georgetown. She is part of the Board of Advisors of the Beeck Center for Social Impact and Innovation, she is Fellow of the One Day Technology Policy Accelerator of the American Scientific Federation, and she has been recently selected as an expert in the Digital Government and Access to Public Services area in the Advisor Group to make the First Open Government Plan of the CDMX.

Elina Castillo Jiménez

Feminist lawyer expert in human rights and public international law, committed to using her voice and skills to help build a fairer and more inclusive world. She combines her thematic expertise in human rights and digital activism to develop public policies, manage projects, design incidence and communication campaigns, and enable workshops and other spaces for dialogue, thanks to her seven years' experience advising governments, the civil society and international organizations. She is part of the Advisor Network of the PIT Policy Lab, as a specialist in human rights and digital activism.

Luz Elena González

International specialist in cooperation for development, with experience in the internationalization of metropolitan cities and legislative incidence. She is committed to the promotion of the ethical design of technology in its intersection with public policy to create more inclusive, sustainable and resilient cities in Latin America. She has worked as an advisor and project designer in the public sector. She has a BA in International Affairs from the Tec of Monterrey. She is currently a project assistant at the PIT Policy Lab, a member of the MCODER.ai community and a content curator at Women in AI Mexico.

Executive Summary

At a global level, the publication of documents addressing the importance and need to anchor ethical principles to the development and implementation of artificial intelligence systems has multiplied. However, few of them mention how those principles can be put into practice. This research work reviews the literature about the concept of r*esponsible use of technology*, presents a brief assessment of the models based on ethics and human rights for their governance, and outlines some advantages of its incorporation for organizations. In the second section, there is a documental revision of insights published by the World Economic Forum (WEF), and examples of tools that have been implemented by big technology companies of the global North are given. Regarding Latin American, some of the tools developed by the Inter-American Development Bank (IADB) within the framework of the initiative fAIr LAC are reviewed, and a combination of the models based on ethics and human rights in the region is proposed. Finally, regional best practices are explored, as well as examples that have been implemented by the private sector in Mexico, and we will end with some reflections about the challenges that lie ahead in the future for the region.

Contents

1. Introduction

In the last years, in line with the global boom of artificial intelligence (AI) as a transforming tool and a general purpose technology, the creation of concept frameworks based on a series of ethical principles to guide its development and adoption has become widespread. The Berkman Klein Center of Harvard made a comparative analysis of over 30 documents that explain many principles and their interpretations, identifying eight shared trends or thematic axes: privacy, responsibility, safety and protection, transparency and explainability, equality and non-discrimination, human control of technology, professional responsibility, and the promotion of human values (Fjeld & Nagy, 2020).

Even though there is consensus about the need to address these topics, it has been called into question, mainly by the private sector, that the publication of such principles does not have mechanisms to accompany its operationalization, that is to say, that they are only good intentions to aspire to, but do not translate into concrete actions in the life cycle of technology products or services and, therefore, there is no accountability for potential unwanted impacts.

In order to face this challenge, one of the most relevant initiatives from the global North has taken place in the World Economic Forum (WEF), where a pioneer working group was created in 2018 concerning the responsible use of technology (RUT), through which it carries out thematic research about innovative tools and methodologies available for companies to transform organizational cultures and implement strategies based on ethical and human rights (HR) approximations, and on lessons learned from behavioral sciences that seek to influence the behavior of people and working teams. Additionally, it organizes workshops and conversations with companies, ethicists, the civil society and the academic world to exchange knowledge and examples from the private sphere about how to put the ethical principles of AI into practice. With the participation of global companies such as Deloitte and IBM, and institutions such as the Markkula Center for Applied Ethics of the University of Santa Clara in the United States, this working group is co-led by Microsoft and Salesforce, who have teams focused on mainstreaming their respective AI ethical principles in coordination with the different areas in charge of the design and implementation of such technology through the catalog of the products and services they offer.

In Latin America and the Caribbean (LAC), the Inter-American Development Bank (IADB) launched the fAIr LAC initiative in 2019, which stated the need to address new conversations and actions to make the most of AI from a RUT and a context anchored in the regional reality. From this platform, work has been done to develop practical tools and the identification of use cases in LAC where the AI is at the service of the public interest, and governments, the academic world, the civil society, start-ups and companies are involved (Gómez Mont *et al*., 2020). Such cases show an encouraging regional picture; however, there are few examples where the development and implementation of AI and its intersection with ethics and HR are addressed.

It is worth mentioning that in LAC, and in the rest of the world, the context of the COVID-19 pandemic expedited the digital transformation of companies towards the digital economy, and of governments towards the provision of public services based on digital tools and remote work. In this process, the inequality gaps have been exacerbated, not only in terms of access to technology but also regarding education for digital competencies, underlining the need to boost investment in this field and move forward towards a flexible and agile regulation concerning new technologies, including AI (López Calva & Melguizo, 2021).

At the same time, the academic world has explored how, in spite of having a robust regional human rights system, there is still room to strengthen the bond between the development of AI for public interest and the incorporation of ethical and human rights frameworks (Lara Gálvez, 2020). The pandemic context has highlighted the need to have proper guidelines and tools to enable companies to incorporate elements of *due diligence* into the design and implementation of technologies from the human rights perspective.

In this sense, there is a possibility to work with global technology companies to co-design impact solutions at big scale, for example, with innovative small and medium-sized businesses experiencing an exponential growth in the region, by means of the development of new business models based on Artificial intelligence systems, or else by offering products or services that use this technology (Everis, 2021). Therefore, it is a priority to accompany the documentation with innovative examples that emerge from small and medium-sized businesses that are participating in exploratory exercises regarding the responsible and ethical use of AI as part of their business models in order to collect lessons, successes and challenges both in the phases of development and the implementation of this emerging technology.

The first section of this research focuses on a concept analysis where there is a revision of the RUT literature, some advantages of the use of RUT in organizations and companies are shared, and a brief assessment of the models based on ethics and HR for their governance, including benefits and criticism to both, is presented.

In the second section there is a documental analysis of the frameworks presented by the initiative of the working group of RUT of the WEF, including the phases suggested for its operationalization in the reports *Responsible Use of Technology* (2019) and *Ethics by Design* (2020). There are also examples of tools that companies like Microsoft, Salesforce and IBM, leaders in that working group, have documented and made available for the public to exemplify how they implement the landing of ethical principles in the life cycle of their technology products. Likewise, from a Latin American point of view, some of the programs and tools that the IADB has made available for start-ups in Latin America and the Caribbean in the framework of the fAIr LAC initiative are analyzed.

In the third and last section, challenges and opportunities are brought up from a public policy perspective in Latin America to obtain a combination of models based on ethics and HR, and a selection of experiences of the entrepreneurial sector in Mexico, who were part of the first generation of programs promoted by the IADB mentioned before, is shared. We end with some reflections to continue identifying use cases and enlarging their documentation in more Latin American countries.

Thus, this work aims at analyzing in a descriptive and critical manner some of the reference frameworks developed to promote the RUT and its uses in LAC, from a governance perspective based on ethics and human rights. The authors expect to contribute insights to a regional conversation, to bring international best practices and examples of local best practices to the entrepreneurial and decision-making ecosystem, and to give visibility to methodological tools to diagnose and evaluate the challenges faced by their AI-based products and services, incorporating the communities affected and users into their design, in an inclusive manner.

Additionally, we seek to contribute with some ideas regarding a hybrid governance model based on ethics and human rights, building on some of the discussions that have taken place in the working groups of responsible use of technology of the WEF and in other international spaces. Lastly, we intend to make public in LAC the concept frameworks that this group has prepared and to give greater visibility and recognition to the tools developed by the IADB together with allied organizations in the region, in order to motivate its drive and harnessing from the highest leadership levels, both in private and public sectors.

In order to meet the objectives set, this research is based on a combination of methodologies, including literature analysis, conversations with key players of international institutions and agencies, such as the WEF and the IADB, to understand first-hand their progress and vision regarding the responsible use of technology, and with the founders of start-ups that took part of the selection of experiences from the entrepreneurial sector in Mexico to identify examples and best practices related to the ethical and practical approach of AI, as part of their participation in the first generation of programs developed by the IADB.

2. Responsible use of technology

When we talk about *responsible use of technology* (RUT), we refer to the working framework suggested by the World Economic Forum (2019) with the intention to introduce and own concepts of ethical thinking and human rights international standards into the value chain of the companies that design, develop, commercialize and use disruptive technologies. In the universe of reference frameworks, most are part of the working group of *responsible use of technology* of the WEF. The starting point of this effort has two big challenges: the formulation of new policies and regulations in the face of the constant technology development, and the ethical and human rights concerns entailed by the digitalization and adoption of disruptive technologies, combining the ethics and HR dimensions in artificial intelligence (AI), machine learning (ML), blockchain technologies, privacy policies, and management of data and the internet of things (IoT).

From the perspective of the WEF and those who participate in the project above, the concept of RUT is intrinsically related to the construction of a trust environment (both inside and outside the business), the relevance of assembling a vision of impact towards society, the need to make thorough risk assessments to identify and, if possible, reduce negative consequences, as well as the importance of building leaderships and having inclusive decision-making processes, and driving diversity in working teams, understanding the key role played by the context and the organizational culture of a company as enablers of systemic change (WEF, 2020). These concept frameworks are very valuable and useful for companies to be able to have a list of examples and implement the actions that best adapt to their own identities and needs. Specifically, the professor and researcher Virginia Dignum builds the concept of *responsible artificial intelligence* as the responsibility facing the power and potential of AI in the autonomous execution of tasks, ensuring that the purpose coded in the system is aligned with the purpose visualized for the product or service using AI (Dignum, 2012).

2.1. Advantages of RUT in organizations and companies

From an added-value perspective, it has been proven that having a space for ethical reflection and the formulation of concrete actions for their operationalization shows greater technology knowledge and corporate maturity (Sniderman, 2020). From a series of studies conducted by Deloitte about the acceleration of digital innovation and corporate leaderships in the context of digital transformation, as well as the international race to get ahead in the development of AI systems, the advising company stressed that attention to the ethical and human rights impact of technology reflects greater progress in the involvement of companies in digital transformation, compared to other companies in the same fields that are in the initial stages of digital assimilation. According to that study, 57% of the people interviewed in companies in the process of digital maturity reported that the internal leadership of their organizations allocated an adequate and useful time to reflect and communicate the social impacts of the digital initiatives that they launch (Kane, et al., 2019).

Similarly, the same study points out that 47% of the organizations have clear and defined processes of decision-making acknowledge the importance of the RUT. The figure goes down to 21% among those who consider that there is not such a clear process (2019). Therefore, the ethical proactiveness in the entrepreneurial ecosystem allows companies to have a broader picture regarding the technology development and use challenges, as these concerns are turned into an intrinsic part of their processes, so the following iterations of products and services will have previous learning and specific considerations to take into account in a cross-cutting manner.

Furthermore, paying attention to ethical considerations in all the development stages of technology, in the organizational culture and the business model is an opportunity to reinforce the trust of the customers and actors involved. The Global Report of Marketing Trends 2020 of Deloitte points out that trust in brands is an increasingly important aspect for the success of businesses and their relationships with customers, regulating agencies and the media, through transparency and openness about processes that guide decision-making (Deloitte, 2020).

On the other hand, the Ethisphere's 2021 World's Most Ethical Companies index, which brings together over 200 data points about culture and social practices, activities focused on ethics and diversity, and governance actions of 135 companies, compared the performance index of the companies classified in the Solactive Index of high capitalization of the U.S. Through this comparison, it was found out that there is an "ethics prime", giving a benefit to ethical companies with a growth of 7.1% over their counterparts in the period of five years between 2016 and 2021. In this respect, Ethisphere highlights that this benefit has been consistent since the measurement of companies started, and it argues that the result of ethical organizational and development practices is reflected in a better financial performance (Ethisphere, 2021).

Given the growing importance of corporate ethics and trust, the working group of RUT of the World Economic Forum presents operationalization models of the ethics and human rights perspectives, so that the advantages of both may be leveraged in the different stages of the cycle of technology products and services.

2.2. Brief assessment of the models based on ethics and human rights for RUT governance

There is consensus that technology use and design should incorporate minimal principles around the different thematic axes (Fjeld& Nagy, 2020). Now then, to put those values and principles into practice, it has been necessary to build models or directives that allow both state and non-state players to be able to identify gaps in RUT. This section puts forth a brief assessment of two of the main models used to that end: the governance models based on ethics and those based on human rights. Likewise, we will address some benchmarks regarding opportunities for their operationalization in LAC.

2.2.1. Ethics-based governance models: benefits and criticism

The ethics-based governance models have been widely used in the RUT (WEF, 2019), in part because of the possibility to adapt them to different geographical and cultural contexts. These models allow for the incorporation of considerations from different systems of values, traditions and cultures. Thus, ethics-based models make it possible to make decisions "in situations in which what is right and wrong, good and bad, is not clearly defined" (WEF, 2019). This way, ethics-based models are particularly attractive for multinational companies, as they enable dialogues that imply taking into account different perspectives that cannot be ruled out depending on more neutral elements (WEF, 2020).

Even though the ethics-based models prevail in the industry, they have not been free of criticism. One of the main criticisms is that, when cultural considerations are allowed, it may be possible to define different compliance standards, according to the value system of each context (Elsayed-Ali, 2018). For instance, questions such as who to sell, how to restrict the inadequate use of AI or how to reduce biases cannot always be addressed in a homogeneous way using ethics-based models (WEF, 2019).

Another criticism is the lack of concrete, measurable criteria to monitor the compliance with RUT standards. It has been highlighted how, in practice, ethics-based models are usually hard to operationalize for technical teams (Burt, 2020). Thus, these models run the risk of becoming aspirational statements rather than specific control criteria that have to be met. For example, out of 160 directives analyzed worldwide in a broad study, most of them are based on voluntary commitments or recommendations with little force to ensure their compliance (Algorithm Watch, 2021).

As a solution to this ambiguity, the creation of clear, measurable and monitored indicators has been offered for each ethics-based principle or criterion that technology-developing organizations adopt in their processes (Burt, 2020). However, this proposal would not help fully solve how the content of these ethics-based principles is defined. For example, what do we mean by "respect for equality and non-discrimination"? What concept frameworks are we going to use, if we depend on the local context? How do we measure compliance, if we do not know what it really refers to?

We believe that, even with the development of indicators and metrics, using ethics-based models to promote the RUT exclusively does not totally overcome the hurdles for its fulfillment. It does not get to solve the lack of follow-up mechanisms either: in the study mentioned before, for example, it is documented that *only* 10 of the 160 directives analyzed have practical implementation mechanisms (Algorithm Watch, 2021). Ethics-based models, in and of themselves, do not seem to be enough to incorporate the resources available in case of non-compliance and measures to protect and prevent the risk of damage into the cycles of design and use of technology.

As an alternative, different players, mainly from the civil society, have promoted the design of models based on more universal and firm criteria, which, as is the case of AI, transcend borders. The Toronto Declaration, for example, is a call to go beyond the ethics-based model and adopt one based on the international framework of human rights (Amnesty International & Access Now, 2018).

2.2.2. Human rights-based governance models: A common and unique framework?

By and large, the proposal of human rights-based governance models is justified by the argument that they are universal, binding and actionable, that is to say, they present a clear action framework, generally accepted by governments, companies and the civil society, with responsibilities both from the States and private actors. Unlike ethics, the starting point of a human rights-based RUT model are principles and regulations internationally recognized (WEF, 2019).

Indeed, the human rights-based models encompass criteria that are legally binding, which put forth a minimum number of conditions to be respected, named rights and obligations, as well as holders of those rights and obligations. In the context of RUT, the human rights-based models make it possible to identify roles and standards to be followed by governments and private actors, in the face of users of technology tools, as well as people who may be affected by them (WEF, 2019). Thus, elements like privacy and free access to information, for example, go from aspirational principles to rights of all the people, with a specific content defined by treaties and other legal instruments, and they should be respected by states and companies.

Even though the ethics-based models may be perceived as endogenous for those promoting the development of AI, in the case of HR models, they have been recently discussed and are more closely related to the responsibilities of States against the possible impact of this technology. From this last perspective, the regulating role of governments has been much debated. However, in the last decades, and from the international law of human rights, the role of companies in their safeguard has also been discussed.

This construction of parallelisms for non-state actors, such as companies involved in the development of AI, has been a process that starts off the recognition that, while at a normative level, States are the ultimate guarantors of human rights, not only governments have the power to affect people through AI and, therefore, not *only* States should be responsible under the human rights international law.

This way, the human rights-based RUT models include the UN Principles on Business and Human Rights as one of their cornerstones; they clearly state the duty of companies to respect them. Under this instrument, companies, regardless of the geographic scope of their operations, go from passive actors to agents who have an impact on their compliance and, therefore, should follow some benchmarks in their business models (UN, 2011).

In this sense, the Guiding Principles are an effort to make parallelisms between state responsibility, as defined in international law, and the responsibilities that all companies must assume to participate in the process to safeguard human rights at a domestic, and even transnational, level.

In this sense, it is worth mentioning that the Toronto Declaration constitutes a big effort to take human rights concepts to the field of AI, on the basis of antecedents such as the Guiding Principles. This Declaration outlines responsibilities of states and non-state actors regarding their use of ML systems, such as the mitigation of discriminatory effects and the provision of effective resources for individuals or communities negatively affected by technology (Amnesty International & Access Now, 2018).

Indeed, one of the concerns where human rights international law may be particularly useful, in the field of AI, is in topics about equality and non-discrimination. In fact, one of the main standards that the human rights model offers to the RUT is related to the principle of equality and non-discrimination that makes it possible to identify, with the inclusion of AI tools, potential biases and prejudices that may be replicated, and understand how to avoid some discriminations.

Human rights international law, especially in the Latin American and Caribbean tradition, has developed concepts and tests to identify the occurrence of structural, multiple or intersectional discrimination and to make room for the recognition of biases as structural causes (MacGregor, 2019). In this sense, when these criteria and tests are incorporated into the design and use of AI, we might decide to quit the development of some tool in particular, because the risk of replicating discrimination situations is high and unavoidable (MacGregor, 2019).

From this reasoning, we can identify that one of the responsibilities that companies have to assume, according to the human rights-based model, is conducting a human rights-based impact or risk analysis at the beginning of design, not only during the deployment and use (WEF, 2019). This is so because one of the benefits brought about by the human rights-based model, apart from minimum standards to meet, is the need to prevent their possible violations, among them the risk to generate or amplify situations of discrimination (Lara, 2020). For that, different actors in the AI field have proposed the use of tools, such as Human Rights Impact Assessments (HRIAs), during the full life cycle of AI systems (Latonero, 2020).

It is worth stressing that the HRIAs are provided by the Guiding Principles. Moreover, while the use of these tools is widely considered to be a best practice, they are not totally mandatory internationally yet for companies and even for state actors involved in the design and use of technology (Dawson, 2019). Thus, different stakeholders, in particular from the civil society and the academic world, have made the case for governments to gradually include the analysis of risks and impact on human rights as a mandatory condition that should be met to grant permits or licenses for the use of technology (Element AI, 2019). The Toronto Declaration is, precisely, an effort to that end.

The binding effect of human rights is the foundation of another benefit offered by the governance models based on them. What is often missing in the ethics-based models is the need to determine accountability mechanisms (WEF, 2019). According to the HR international law, this rights protection would be two-way: the prevention of human rights violations, mentioned before, and the assurance of those rights. In the cases when they are not fully realized, it should be possible to have access to an effective resource, and if it applies, to receive reparation for the damage caused.

Even though the human rights-based models might fill the gaps of the ethics-based models, they are not a perfect solution. In general, it is important for technology-developing companies to collaborate permanently with the civil society, the academic world and the communities affected, so as to operationalize the technicalities of human rights and make effective procedures that ensure them. For example, through this collaboration, technology-developing companies might be able to establish accountability resources or mechanisms that are effective and agile enough to be able to respond to the fast nature of the development and use of technology.

Likewise, even though one of the advantages of human rights is their binding nature, as they are legal norms, they should be reinforced by governments at domestic level. In general, the HR norms mostly referenced in the field of AI are Treaties, such as the International Covenant on Civil and Political Rights, generated at the core of multilateral agencies like the Organization of American States (OAS) or the United Nations (UN). For human rights-based models to be more effective, it is ideal to incorporate the obligation of companies to protect human rights in public policies related to AI systems at national or domestic level.

This incorporation might help face another criticism of the model, related to this last topic. As we said before, one of the vital standards of human rights are the Treaties, which, in turn, are interpreted by specialized bodies. Some voices, for example, researchers from the Berkman Klein Center of the Harvard University, have stressed that, even though the HR tenets are universal, they do not provide the necessary precision and certainty for companies to enforce them, in particular because the interpretation of these standards by courts and government authorities is not homogeneous across regional systems, like the European and the Inter-American ones (Dvoskin, 2019).

Although it is possible to find some differences in the interpretation of the content of certain rights between regional systems, there is an increasingly greater interchange between those systems as an exercise to strengthen their work. Likewise, it is important to underscore that companies are more and more frequently called to make a human rights-based risk analysis and make decisions according to those results by national judicial regulations and by the Inter-American system. As recent examples, the last report on AI and human rights issued in 2021by the High Commission of the United Nations for Rights and the standards of companies and human rights of the Inter-American Commission on Human Rights (CIDH) stand out. We are going to discuss both sources later on.

Now then, taking into account what we have explained before, instead of adopting one of the models, we propose that ethics-based and human rights-based governance models should complement one another. In a region like LAC, using both models in a complementary manner might help solve several challenges that both companies and governments face at the time to make sure that AI and other technologies are used responsibly. We are going to deal with this proposal later on.

3. Proposals to operationalize the responsible use of technology

From a Global North perspective, the RUT Working Group of the WEF recognizes that the ethical challenges of disruptive technologies are inherent to them, while the digitalization processes undertaken entail the need to question, rethink and outline our relationship with new technologies to take advantage of their benefits (BSR & WEF, 2019). In the document *Responsible Use of Technology,* the specific roles of the companies that develop disruptive technologies and those which obtain and implement them are highlighted. The former have the opportunity to outline the use policies that guide what can and cannot be done with their products, they can choose who to sell and facilitate information and risk-mitigation practices to their potential customers; the latter have the possibility to understand the impact of the use and implementation of technology through a human rights assessment, mitigate the risks of implementation by deciding how the technologies will be used, and take part of a permanent dialogue with the developing companies to prevent the uses of technology from affecting users. Moreover, specific actions are established in three stages or phases of the value chain: Design and Development, Deployment, and Implementation or Use. Next, the models of responsible use of technology, Ethics by Design, and the complementary model of Act-Enable-Influence are described, as well as examples of tools developed by Microsoft, Salesforce and IBM for their operationalization.

3.1. Model propounded in the report Responsible *Use of Technology*

The model propounded by the report Responsible Use of Technology (BSR & WEF, 2019) aims at the combined implementation of the ethical and human rights perspectives in the development of technology products and services. To that end, it suggests assuming the corporate responsibility regarding the observation of human rights and the need to act in the face of possible unfavorable impacts on them in corporate activities, as outlined by the UN Principles on Business and Human Rights, adopting the ideas of responsibility and capacity for action in each stage of the value chain of the products and services offered.

In the design and development phase proposed in this model, the approach is the recognition of ethical and human rights implications in the teams' prototypes and ideas by means of four distinctive practices. Firstly, if companies lack the necessary knowledge to identify situations of negative impact, they should hire specialized people and consulting agencies in charge of drafting policies for the companies based on the respect for human rights through the awareness of the different teams involved. Secondly, the participation in active discussions with actors relevant for developing technologies will make it possible to know the points of view of vulnerable populations that could be primarily affected, and make decisions in the early stages of the cycle of creation of products to reduce the associated risks. In this sense, there are already work frameworks that enable this interaction between developing companies and actors: the people-centered design, HR by design and responsible innovation are some of them (BSR & WEF,

2019). The third practice implies the use of technical tools to respond to possible scenarios of discrimination and bias, such as the Justice Toolbox of Accenture or AI Fairness 360 of IBM, or the use of databases representative of the populations affected by one product. The last practice is the "adversarial tests", in which a product or service is recalled from the contexts in which it will supposedly be used, and it is used under different conditions to assess the possible impact of its deployment (BSR & WEF, 2019).

The previous proposals require the multidisciplinary participation of those involved in the value chain, starting off the policy and sustainability teams, in the institutionalization of the people-centered design work frameworks, and the product design and engineering teams to consult the implementation of new processes and to give feedback on the first stages of the design of products, services and technologies (BSR & WEF, 2019).

In the deployment phase, the implementation of three measures paying attention to ethics and human rights are proposed for the sale and use of new products. Firstly, acceptable use policies (AUPs) are considered a suitable exercise to define ethical uses and their limits, but they build a legal and moral distance between people and the companies that create technology, so that the final responsibility lies on users, in addition to the fact that enforcing AUPs entails an associated challenge (BSR & WEF, 2019). Secondly, aimed at more commercial and industrial actors, creating "white or black lists" defines the allowed use of a technology in some sectors, through the restriction of possible buyers and industries of implementation; this is achieved by understanding the nature of the customer and the final use the product or service would have. White lists record potential consumers who have the green light to obtain certain technology, while black lists eliminate some companies or industries as possible customers. Both lists are made by means of an assessment of ethical, sustainability, equality and human rights criteria, among others. Finally, providing ethical and human rights training as guidelines for a better use of products during the sale process increases the awareness of those who purchase the product and opens the door for governments, universities and the civil society to own those guidelines (BSR & WEF 2019).

Finally, in the use phase, during the creation and production cycle, responsibility goes back to users and consumers, so it is important to take actions to distribute the responsibility between the sectors and industries involved. Technology companies create new products and services, but their ethical use has relevance in the wide variety of the adopting industries.

This methodological framework admits that there are limitations to the implementation of RUT practices. In the first place, it is necessary to include the entire value chain in these practices, instead of conducting them every now and then; therefore, public policy-makers, the civil society and other relevant actors should get involved to meet ethical and human rights challenges in the face of technology. This involvement requires a commitment from states to incorporate and implement policies for the RUT that companies, groups of the civil society and scholars cannot implement independently (BSR & WEF, 2019). At the same time, the complementary model of Act-Enable-Influence is presented as a path to bring about systemic changes in the value chain and in actors involved from the initiative of the companies committed to the RUT.

3.2. Model presented in the report Ethics by Design

Further to the ideas put forth in the document R*esponsible Use of Technology, the report Ethics by Design*, also by the WEF (2020), goes back to the model developed by Nicholas Epley and David Tannenbaum (2018), based on their investigation in the field of psychology and behavioral sciences, to design principles that encourage refection and ethical actions. That model is divided into three related phases: attention, construction and motivation (WEF, 2020).

The first pillar presumes that a design that is effective for ethical behavior will drive people to reflect routinely; for that, it recommends the use of awareness material, institutionalized organizational incentives as reminders within the selection and interview processes, or technical tools that assess possible ethical risks of algorithms, as well as the creation of shared languages that include slogans, mantras and co-defined missions.

The second pillar of *Ethics by Design* of the WEF acknowledges the influence of contexts in decision-making, and it suggests incorporating their diversity in the experiences lived by the people in the organization (2020). For that, the creation of socially diverse teams is proposed, which develop, together, ethical reasoning and RUT models based on ethical paradigms and existing policies, while such models are iterated with the people and communities that might be affected by the new technologies (WEF, 2020).

The last pillar argues that the change in people's behavior needs psycho-social motivations, for which it presents historical introspection activities of the people in the organization, as well as their decision-making from lived contexts, using the mission previously defined and the framework of ethical work as a signature (WEF, 2020). With that, this model also seeks to create empathetic relationships from individuals of the organizations towards customers and users, all this within a general narrative of the ethical impact of the products and services that they develop.

As can be seen, the leadership of the WEF in the development of operative models to institutionalize the models based on ethics and human rights offers results that can be applied in the whole process of creation of technology products and services. That is why, in line with the growth of the innovation and entrepreneurial ecosystem in disruptive technologies in Latin America and the Caribbean, it is relevant to present these models and examples of tools to companies and start-ups of the region. Taking elements from those models and tools in the entrepreneurial ecosystem in Latin America and the Caribbean will make it possible to enable a change of paradigm in which their use becomes a widespread practice outside companies and organizations based in the United States and Europe, leaving the construction of an international picture of the RUT sensitive to regional contexts and the specific needs of different populations.

3.3. Examples of tools for the RUT

In line with the principles and guidelines proposed by the models mentioned, Microsoft, Salesforce e IBM, leading companies in the RUT working group of the WEF, have documented examples that offer a variety of options for the teams involved in the life cycle of technology products

or services, adding a wide range of actors to include different perspectives about the impacts of AI systems. Next, we will share some tools that those companies have put into practice.Juego de juicio (Judgement Call)

Judgement Call

Going back to the five AI principles of Microsoft (justice, safety, privacy, inclusion, transparency and accountability), the game presents the teams in different phases of the life cycle of the technology product or service with the opportunity to assess the impact of such product or service from the perspective of another actor involved. The people participating receive a letter that assigns the role of some actor or group that the digital product has an impact on, a letter where they read one of the AI principles, and a letter with a number 1 to 5, representing the number of stars the hypothetical actor will give the product in the review. According to this information, each team member writes reviews explaining the impact of that technology (Markkula Center for Applied Ethics & WEF, 2021).

This initiative has proven to be very effective to motivate ethical thinking and collaboration in the teams inside the company (Ballard, et al., 2019). Some reasons why this tool has been successful are that the organizational leadership pushed in the direction of the development of solutions related to ethics, the game builds on the AI principles of Microsoft, and it considers existing practices in product development teams, so that their incorporation into the process does not entail a bigger work burden but a change of perspective about existing processes.

External Advising Council and Office for the Ethical and Human Use of AI

Salesforce opened the Office for the Ethical and Human Use of AI in 2018 as part of their Equality Office in order to ensure that technology services and products do not affect human rights but help protect them, and generate common values for the technology industry through the leadership of the company (Salesforce, 2020). Since the creation of that office, the commitment was made to create a space where the different actors involved in the decisions of the company, their needs and ethical perspectives are recognized. At the same time, an external advising council was created, with experts in ethics and RUT with varied perspectives that nurture the decision-making process (Salesforce, 2020) (WEF, 2020).

Envision AI Workshop

This Microsoft exercise helps teams reflect on the development of an impact analysis. In it, participants review real scenarios of the AI system development process of Project Tokyo, established in 2016, whose goal is to boost human capacities through technology. The objective of this scenario review, from innovation centered on people and the understanding of technology needs of different groups, is to provide tools and resources to mitigate possible risks of technology use and identify the differentiated impact on users (WEF, 2021). In the context of the event Forum Europe 2021, a short version of the workshop was organized, where attendees had the opportunity to put into practice the artificial intelligence principles of Microsoft through interactive discussion and case-based reflection (ForumEurope, 2021).

Notifications in AI Platforms

Einstein is the AI platform of Salesforce. One of their functions to alert people who use this service with commercial purposes and who might be using sensitive variables, such as age, race or gender, is through guides called Einstein Discovery. The tool indicates data problems that might skew results, for example discrepancies between classes, high correlation between sensitive variables, too many outliers or the presence of a variable correlated with the results (Salesforce, 2020).

Community Jury

The tool of the Community Jury, proposed by Microsoft, allows teams in different stages of the life cycle of a technology product or service to interact with the populations it may impact. The first step to carry out a community jury is to establish a relevant, representative and diverse group of actors, who will have the role of the jury. The teams that develop a new service or product present the purpose and possible cases for the use of the technology, as well as potential dangers and benefits. Participants review this information and share perspectives about the impact of the product on them, with the help of a neutral moderator who ensures their equal participation. The key topics that come up in the discussion are incorporated as opportunities and challenges for the co-creation of solutions (Markkula Center for Applied Ethics & WEF, 2021).

InterpretML

This Microsoft tool aims at creating more transparent, understandable and interpretable machine learning models. Through InterpretML, we can generate *global* explanations about the behavior of a model and local explanations for individual predictions of that model. Moreover, it adds decision trees and mechanisms to generate explainable and transparent models, as well as tools to decode black box models (Markkula Center for Applied Ethics & WEF, 2021).

Error Analysis

Similar to InterpretML, error analysis is a tool used at Microsoft that looks for areas of opportunity for the explainability of ML models, but this tool is focused on the identification of biases and errors in data. Through visualizations, data science teams can identify classifications within a model that may have errors, and the data groups whose error margin stands out from the parameters set, as well as their error percentage distribution (Markkula Center for Applied Ethics & WEF, 2021).

Tools Open Source of IBM

In 2020, IBM donated a series of open code tools to the Linux Foundation. Among them are *AI Fairness 360*, which has over 70 metrics of AI justice and 10 bias mitigation algorithms; *AI Explainability 360*, which uses eight algorithms to render the ML operation transparent and

explain it, considering the variety of players involved in the life cycle of AI; *the Adversarial Robustness 360 Toolbox*, which allows AI developers to defend themselves from attacks through learning based on the simulation of possible attacks and the evaluation of response models, and the *Uncertainty Quantification Toolbox*, which allows developers to know the limits of trust you can have in AI by means of the quantification of uncertainty in the decision-making of a given system (Markkula Center for Applied Ethics & WEF, 2021).

Closed Tools of IBM

Cloud Pak for Data is an IBM tool package, used in the full data analysis cycle inside the company, which allows users to participate in the monitoring of reliable AI models. One of the most important tools is Watson Studio, used together with the AI solution Watson, to ensure the quality of data used for learning, scan models to detect biases and identify populations that algorithms may have a negative impact on, and generate recommendations to modify machine learning models; all this is explained using words that people in the business field and the scientific data community can understand (Markkula Center for Applied Ethics & WEF, 2021).

The tools presented by Salesforce, IBM and Microsoft within the framework of the work group may be useful as a starting point to understand how to operationalize the RUT, but as the document *Microsoft Case Study* (2021) points out, the culture change in the organization is crucial for these principles, work frameworks and tools to become an inherent part of the process to develop new technologies. That is why, when we think of the possible implementation of those tools en Latin America and the Caribbean, it is necessary to adapt the expectations, capacities and needs of start-ups and multinational and national technology companies to the context of the region, in keeping with their priorities.

In the case of LAC, in accordance with the need to have reference frameworks adapted to the political, economic and social context of the region, the fAIr LAC initiative of the IADB Group was launched as a meeting point for the actors of the public and private sectors, the academic world and entrepreneurs to generate networks that contribute to the ethical and responsible implementation of AI in the region to solve public problems (IADB, s.f). The initiative appeared at the end of 2019, stemming from the increase of projects with a component of private or sensitive data use, as well as an alignment with the Vision 2025 of the IADB, where tools that simplify the digitalization of companies, especially the smallest ones, and the responsible adoption of new technologies by the public and private sectors are democratized.

In May 2021, the AI Ethics Self-assessment for Actors of the Entrepreneurial Ecosystem was published, as the first product of the fAIr LAC *Entrepreneurial Journey* project. It is a guide in Spanish for companies, investors and accelerators regarding ethical aspects to consider in the development and implementation of technology solutions based on AI data management according to their maturity level. The result of this exercise is translated to the identification of areas of improvement that will make it possible to enhance the quality of products, services, processes, models and developing systems, and identify the main areas of risk in terms of ethics.

What is innovative about this document lies in two main aspects: the first one is that it does not set the weight of self-regulation only on entrepreneurs, but it also involves two vital players for the ecosystem: project funders and accelerators; the second one is that the guiding questions belong to three levels of development in companies, from the early stages of devising, to small and medium-sized businesses (SMBs) developing innovating products. Likewise, even though the document is thought of mainly to guide the development and implementation of AI-based solutions, it is also useful for solutions based on data management (Rosales, Buenadicha & Narita, 2021). An important aspect of the tool is that it was co-created with entrepreneurs, and it is being tested within the entrepreneurial ecosystem of Ibero-America.

To sum up, both the WEF, from a mainly Eurocentrist vision, and the IADB Group, with a focus on LAC, are international agencies that have been pioneers in the generation of knowledge, tools, initiatives and projects that seek to operationalize the RUT in close collaboration with global con technology companies and with the technology entrepreneurial ecosystem of the region.

4. Opportunities and challenges for public policy in Latin America

The generation of ethical and human rights frameworks for the RUT in Latin America is a pending task. The frameworks proposed by the WEF are the result of a general context of the global companies and the organizations that develop them together, so the specificity of contexts and needs of the LAC region are not taken into account. In this sense, even though we have the opportunity to use the existing frameworks and tools, and document their uses in global companies with a presence in the region, it is necessary to open up the conversation to the entrepreneurial ecosystem of Latin America, so that emerging companies can participate in the construction of ethical governance and human rights models in their initiatives and, in turn, report a hybrid model for the region, which clearly distinguishes between the models, providing clarity and discretion for their implementation of the RUT.

4.1. Governance models for Latin America and the Caribbean

One of the main challenges for the applicability of ethics-based governance models for the RUT in the region is the high inequality levels, since LAC is the most unequal region in the world (UN, 2020). Therefore, it is not enough to look at the gaps between the global North and South in terms of technology development, but it is also necessary to look inside. The development of AI in LAC must be carried out from the recognition of inequalities in our region and from the possibility to have regulatory standards that may adapt, in the shape of handbooks and other tools, to the design and development of AI. In that sense, as we mentioned before, human rights-based models may pave the way to meet these goals.

Moreover, even though there has been significant progress, making sure that the protection of human rights is binding for companies, as it is for States, is still a challenge (De Asís, 2020). As we said before, at the level of the universal system of human rights, although we have the UN Principles on Business and Human Rights, this is a softlaw instrument, that is to say, it is not binding, like treaties. Similarly, in the Latin American and Caribbean context, although the need for companies to take risks and the impact of their operations on the human rights of people is acknowledged, the Inter-American System of Human Rights still argues that states are the ultimate guarantors and they have to play their regulatory and overseeing role to make sure companies comply with those obligations (CIDH, 2019).

However, we can identify some encouraging indications both at regional and universal levels. In the Latin American tradition of human rights, there is a concept of “due diligence” to refer to any action that the States can and must adopt to prevent the violation of human rights (CIDH, 2019). In the most recent analysis of the Inter-American Commission of Human Rights on business and human rights (2019), this concept has been used to establish what role human rights must play in corporate practices and, in particular, assess and prevent or reduce possible damages to rights as a result of their activities. This obligation is particularly relevant in the field

of AI, because it goes beyond avoiding the wrongful or inappropriate use of AI systems in the public sphere, since it also implies the creation of mechanisms that make it possible to oversee the actions of non-state players vis-à-vis right holders (CIDH, 2019).

However, up to now and all along the region, there are still prohibitive regulatory prescriptions that do not give any answer to doubts regarding what companies can do when they want to start a process of responsible development, of positive actions, to incorporate protection principles in the design and use of AI systems. Now then, if we go back to the concept of "due diligence" in the face of the AI design and use cycle, we argue that, apart from these prohibitive tenets, the companies involved can also consider, in advance, the implications that technology development may have in terms of specific rights and take actions to reduce that risk.

Everything says that the universal system of human rights protection points in that direction. From the United Nations Educational, Scientific and Cultural Organization (UNESCO), in July 2021, progress has been made in the development of a recommendation for the ethical use of AI with the promotion of human rights, an important step for its member states (UNESCO, 2021). In September 2021, the United Nations Office of the High Commissioner for Human Rights, in a revolutionary report, underscored the argument that both States and companies should guarantee that a "due diligence" be conducted not only since the development, deployment and operation, and it even — and this is new — acknowledges the role of States not only as regulatory agencies but also as "customers", suggesting that they should conduct the due diligence also when they consider the acquisition of technology (UN, 2021).

Additionally, it proposes "moratorium" in the development and use of AI whose risks for human rights pose a serious challenge, at least until there are effective safeguards to mitigate or prevent possible damages (UN, 2021). Even though it does not talk about a binding obligation for companies at an international level yet, it definitively reinforces the role of states at a domestic level to ensure that companies that develop technology adopt human rights models for the RUT (UN, 2021). Although it is true that not all the entities participating in the process of design and use of AI have, today, the capacity to incorporate criteria of responsible use based on one model or the other from the beginning, it is not less true that, as seen from both models, it is imperative to make tools aiming at that objective.

Prioritizing a human rights-based model would allow for a responsible use of AI from the beginning, reducing the possibilities to incur in greater investment to correct biases in the long term or, in the worst-case scenario, losses due to lack of use. A concrete example comes from the intersection between the use of technology and the rights to privacy and free speech, in particular, in the field of social media. Some authors have suggested getting closer to a governance model based on human rights and sustained by the responsibility of companies (Kaye, 2019). In particular, Kaye proposes the figure of "multi-actor councils" as decentralized agencies, including local activists and users as members, and that they should use the human rights standards as foundation, so that companies have a "language to coordinate their international position, in ways that respect democratic norms and counter authoritarian attitudes" (Kaye, 2019). Nowadays, it seems that companies like Facebook, with its Content Advisor Council, made up by experts in Law and other areas, have incorporated proposals like this (Botero, 2020).

In conclusion, it is ideal to be able to combine both models, those based on ethics and on human rights. However, in a widely unequal context like LAC, prioritizing human rights-based models would make it possible to follow criteria that apply not only to the whole region, but to other parts of the world; it would enable a greater interchange and construction of knowledge through borders, with the considerations of place for the context of LAC beyond the principles of well-being or maleficiency. In this sense, the authors state that human rights-based models offer more homogeneity and damage prevention, while ethics-based models enable decision-making where several options might be feasible, especially in contexts where the protection of democratic or equality values should be a priority.

Here is where the value of documenting pilots and use cases that may become best practices for the responsible use of AI comes into play. The models based on ethics and human rights propose criteria that help distinguish the actions that should not be taken. What to do, how to put those principles into practice (positive obligations, in the language of human rights; or moral mandates, in the language of ethics) is a field that continues under development and, to be able to make a contribution to it, it is necessary to look at what has been done. In the next subsection, and from this perspective, some examples of best practice in the region are presented.

4.2. Regional best practices for the RUT

As mentioned before, even though there are differences between the countries in the global North and South in terms of technology development, LAC has experienced an exponential growth in the artificial intelligence ecosystem in the last years. It is here that we ask ourselves, how have the RUT principles been incorporated into the framework of such technology development? Pursuing the knowledge development at a regional level, next are some examples that have been shown in reports about the impact of AI in entrepreneurship, and initiatives implemented from a multi-sectoral collaboration approach (Open Loop, led by Facebook) and by start-ups in Mexico that have been part of the programs and tools made available by the IADB in the framework of fAIr LAC.

4.2.1. Reports: the impact of artificial intelligence on entrepreneurship

The 2018 and 2020 editions of the reports made by the collaboration between Endeavor and Everis, company of NTT Data, delve deeper into the evolution of the characteristics of the emerging companies that implement AI systems in their productive processes or offer the technology as a commercial proposal in Latin America. In 2020, the analysis indicated that the AI entrepreneurial hub employed over 38,000 people in 490 companies of the region; that number accounts for an increase of more than twice the companies analyzed in the 2018 edition of the same report (Everis, 2018 & Everis, 2020).

Similarly, the most recent version includes a section that describes the main concerns of companies using AI. The most commonly mentioned dilemma was “finding a solution that is a compromise between automation and human intervention” (p.60), explained by means of AI applications that use chatbots, voicebots and other automation tasks to interact with people. In

the area of technology ethics and RUT, this dilemma stems from the right of people, as users, to know that they are interacting with an automated system although it may seem to be human; and the difficulty to eliminate unwanted or abusive human interactions in the communication with AI systems.

Other concerns were mentioned to a lesser extent, like algorithmic bias, data portability and protection, privacy, and the need to build trust in users (Everis, 2020); it is highlighted that "there is not a deeper vision of the ethical implications of AI" (p.66). All the issues mentioned by entrepreneurs as dilemmas of the use of AI are related to the *responsible use of technology*, and the models previously described pick them up again in an interactive way in different stages of the development of AI systems, to raise awareness about their impact on the product or service deployed. Given the current state of the AI ecosystem in Latin America, growing and in expansion, it should be expected that the models of RUT based on ethics and human rights have not been explored in depth. However, this is an opportunity for the young entrepreneurial ecosystem to generate its own RUT tools, and they should become ubiquitous in the development of AI technologies in the region.

4.2.2. Open Loop Mexico

The Open Loop initiative, led by Facebook and implemented globally, has a Public Policy Prototype in Mexico (the first one of its type in the region), focusing on transparency and explainability (T&E) of artificial intelligence systems. This effort is led by the EON Resilience Laboratory of C Minds and the IADB by means of fAIr LAC, and it has the support of the National Institute of Transparency, Access to Information and Data Protection (INAI, for its acronym in Spanish). In this initiative, eleven start-ups implemented transparency and explainability prototypes, which will lay the foundations to develop public policy recommendations in the Mexican context (Facebook, C Minds & IADB, 2021).

The core theme of the program is the strengthening of T&E of AI systems of the participating companies. T&E are part of the principles of ethical use and implementation of AI which, in the first place, allow users to know that they are interacting with an AI system and, also, helps them understand its process, purpose, capacities and limitations. Likewise, the regulatory framework tested in the public policy prototype has at the core of its creation the safeguard of privacy and personal data protection rights, as well as the importance of the selection of training data, algorithms and the model itself to guarantee its reliable and transparent implementation to reduce the negative impacts on people's rights and freedoms.

For that purpose, C Minds created "missions" that lasted between one and two weeks for the participating companies to conduct an internal study, analysis or implementation, and then share their learning in forms with questions defined by C Minds and Facebook. This made it possible to analyze the effectiveness of the policy suggested and generate recommendations for the INAI. It is important to mention that the companies that tested the framework were not familiar with the T&E topic; an important learning was that, in general, they did not have much knowledge about the topic of AI ethics, so it was necessary to redesign the program to cover that issue and then focus on T&E. The answer of the companies was positive, as they shared a great deal of learning that they are implementing in their systems.

It is necessary to replicate the multi-actor initiatives like this one to create frameworks that steer the use of AI systems in a responsible way, especially for the people in charge of the development of technology. This is so because a lack of communication between developers and the companies using the AI systems was identified. Moreover, among the lessons of the program, its implementers found that once a company participates in this type of program, there are not more specialized programs to continue supporting them.

3.2.3. Start-ups in Mexico

Two experiences driven by the entrepreneurial sector in Mexico are shared below, from conversations with their leadership and documentation of examples, concerning their participation in programs developed by the IADB Group in the framework of fAIr LAC.

helKi

helKi is a Mexican company that was born with the dream of democratizing information and best practices of early child development for all the families and caretakers. To accomplish that vision, they have developed a series of principles inspired in different organizations that are leaders in social impact and technology forefront worldwide. In order to operationalize the principles they have committed to, they have joined PASIA, a start-up acceleration program on artificial intelligence in the framework of fAIr LAC Jalisco; moreover, they are one of eleven companies participating in the Open Loop Mexico initiative, and are part of the program *fAIr LAC Entrepreneurial Journey* of the IADB Group.

For Irene Velasco, CEO and founder of the company, the self-assessment tool of the *fAIr LAC Entrepreneurial Journey* allows companies in an early stage to focus algorithms on impact, mitigating risks and informing decisions. By means of the self-assessment, they were able to increase the awareness of the organization regarding the negative consequences of impact minimization actions without an ethics perspective, devise an action plan to define critical topics to be solved in line with the growth of helKI and create open conversations about the ethics of organizations with allies, investors and customers, taking on the role of ethics influencers.

Among the main milestones reached after their participation in the different programs mentioned, the following can be mentioned:

- Defining processes to delimit access and ensure information safety.
- Implementing changes in the drafting of information to avoid unfair biases.
- Creating a change theory where the technology risk is described, so as not to lose sight on it in the impact measurement.
- Communicating the team and living the values guiding the company as part of the organization culture, and having the beneficiary at the core of decision-making.

OS City

OS City is a start-up focused on the creation of digital services centered on citizens. It is part of the portfolio of UNICEF, CAF and Ethereum Foundation. They have formal alliances with the IADB through their blockchain network, called LAC Chain, and with many international communities, such as RadicalxChange and Singularity University.

Jesús Cepeda, co-founder of the company, has highlighted the usefulness of the self-assessment tool of the *fAIr LAC Entrepreneurial Journey* of the IADB Group to deal with the lack of knowledge about biases and its mitigation in the models they develop, while political risks are brought to the table and the participation and democracy perspective is addressed. Even though they do not have a reference framework formally defined inside the company, there is an observance of the principles and values of international communities like the Young Global Leaders of the WEF and the Council of Ethical Artificial Intelligence of the IADB Group.

Likewise, he considers that one of the main challenges to advance the RUT in the region is related to educating customers on the importance and the value of explainability and the responsible use of the tools they develop.

5. Conclusions/Challenges

The trends in the creation of concept frameworks based on ethical principles for the development of technology point towards privacy, responsibility, safety and protection, transparency and explainability, equality and non-discrimination, human control of technology, professional responsibility and the promotion of human values as thematic axes guiding the conversation (Fjeld & Nagy, 2020). However, even though there are good intentions by the AI system development community, the scarcity of insights and mechanisms of operationalization and accompaniment of the ethical frameworks has halted its adoption. For that reason, the World Economic Forum created the RUT working group in order to do thematic research to make innovating methodologies and tools available for organizations and companies, so they can transform their strategies through the combination of ethical and human rights perspectives in each phase of the life cycle of technology products and services. As the AI ecosystem in Latin America is under construction, a space for reflection on the regional priorities of the RUT, the context of its implementation and the need to create work frameworks exploring how to combine the ethics and human rights perspectives to maximize the benefit for users is generated.

From the research done, the authors identify some indications of the next steps or aspects to delve into through further investigation. One of those points is the detection of opportunities and challenges for the technology ecosystem in LAC to incorporate the ethical and human rights perspectives in the organizational culture and the creation of new technology products and services. With this work, we hope to lay some foundations to promote a greater interchange and research on how to design and implement a hybrid model of ethics and human rights from LAC, acknowledging the benefits this exercise could bring, as we have pointed out.

Likewise, from the global outlook to the LAC region, there is an opportunity for technology companies that participate in the working group of the WEF to test the tools that have been developed and implemented in the Global North in the Latin American context, by means of their operation in the countries of the region. Meanwhile, growing technology entrepreneurships in the LAC region will benefit from the incorporation of the ethics and protection of human rights perspectives as a core element into the value chain, using the tools explored in this article, generating their own strategies, and documenting the results to strengthen the ecosystem. There is an opportunity to promote greater interchange between these actors to move forward to the incorporation of such tools in LAC.

In order to meet those goals, it will be necessary to take action on the lack of knowledge by companies and start-ups in the region on the difference between models based on ethics and human rights. The approximation from the programs offered by international agencies like the IADB makes no distinction between them, so only an ethics perspective approach is taken. As a response to this challenge, the systematization and documentation of experiences in the region from the private sector (start-ups), as was the case in the examples from Mexico,

offers an opportunity to detect existing lines of work, as well as priorities of the entrepreneurial ecosystem, trends and areas of opportunity. A first step in the direction of this purpose is the quiz online that is part of the study to learn about the adoption of ethical frameworks in the development of algorithms in companies and start-ups in Latin America; that study is being carried out from the GuIA initiative itself by the Latam Digital Center and the CETyS of the University of San Andrés, with the support of fAIr LAC.

Drawing on in-depth knowledge of the guiding principles of the ethical and human rights perspectives in the existing RUT initiatives in the LAC region, it will be more feasible to generate a hybrid model, customized for the region, which contributes to give shape to the responsibilities of the states, the private initiative, the civil society and the world of academics. This model would make it possible to relate both perspectives to the specific areas of incidence of the actors involved, so that the regional context and the needs detected guide its use. The Observatory of Artificial Intelligence set in motion by the IADB, together with the Regional Office of Sciences of the UNESCO for Latin America and the Caribbean, has the potential to become a space for mainstreaming public policies committed with the sustainable development and the responsible use of technology (UN Mexico, 2021).

Finally, the study of both perspectives in the private initiative gives us the opportunity to explore how to take these tools to the public sector as best practices that could be implemented to face the challenges entailed by the lack of operationalization of both models (ethical and human rights), and generate a broader discussion about the responsibilities of national governments, as well as the need to develop regional alliances and agreements in the ethical decision-making and the protection of human rights, with respect to the development of products and services based on disruptive technologies like artificial intelligence.

6. Acknowledgements

We are grateful to the people who contributed with insights, ideas and examples of international and regional best practices to prepare this report.

Name	Institution
César Rosales	IADB Lab (International agency)
Tetsuro Narita	IADB Lab (International agency)
Irene Velasco	Helki (Start-up)
Claudia Del Pozo	C Minds (Action Tank)
Jesús Cepeda	OS City (Start-up)
Daniel Lim	Salesforce (Fellow in charge of the Group Responsible for Technology at the WEF in the period 2018-2021)
Emily Ratte	WEF (Working group Responsible Use of Technology)

7. References

Ad Hoc Expert Group (2020) for the Preparation of a Draft text of a Recommendation the Ethics of 1. Algorithm Watch. (2020). In the realm of paper tigers – exploring the failings of AI ethics guidelines. Available on: https://algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/

Amnesty International & Access Now (2018). The Toronto Declaration. Available on: https://www.toronto-declaration.org/

Ballard, et al. (2019). Judgment Call the Game: Using Value Sensitive Design and Design Fiction to Surface Ethical Concerns Related to Technology. Association for Computing Machinery. Available on: https://dl.acm.org/doi/10.1145/3322276.3323697

Inter-American Development Bank (s.f). fAIr LAC: Quiénes somos. Available on: https://fairlac.iadb.org/es/quienes-somos

5Botero, C. (2013). Relatoría Especial para la Libertad de Expresión. Inter-American Commission on Human Rights. Available on: http://www.oas.org/es/cidh/expresion/docs/informes/2014_04_08_Internet_WEB.pdf

6Botero, C., et al. (2020). ¿Quién decide cuáles publicaciones se eliminan de Facebook? A partir de ahora, nosotros. New York Times. Available on: https://www.nytimes.com/es/2020/05/06/espanol/opinion/facebook-junta-supervision.html

Burt, A. (2020). Ethical Frameworks for AI Aren't Enough. Harvard Business Review. Available on: https://hbr.org/2020/11/ethical-frameworks-for-ai-arent-enough

Business for Social Responsibility & World Economic Forum (2019). White Paper. Responsible Use of Technology. Available on: http://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology.pdf

9Inter-American Commission on Human Rights, (2019). Empresas y Derechos Humanos: Estándares Interamericanos. Available on: http://www.oas.org/es/cidh/informes/pdfs/EmpresasDDHH.pdf

Dawson, P. (2019). Closing the Human Rights Gap in AI Governance. Element AI. https://s3.amazonaws.com/element-ai-website-bucket/whitepaper-closing-the-human-rights-gap-in-ai-governance.pdf

De Asís, R. (2020). Inteligencia Artificial y Derechos Humanos. Materiales de Filosofía del Derecho. Universidad Carlos III de Madrid. Available on: https://e-archivo.uc3m.es/bitstream/handle/10016/30453/WF-20-04.pdf?sequence=1&isAllowed=y

Dvoskin, B. (2019). Why International Human Rights Law Cannot Replace Content Moderation. Berkman Klein Center for Internet & Society at Harvard University. Available on: https://medium.com/berkman-klein-center/why-international-human-rights-law-cannot-replace-content-moderation-d3fc8dd4344c

Deloitte (2021). Tech Trends Report. Available on: https://www2.deloitte.com/content/dam/insights/articles/6730_TT-Landing-page/DI_2021-Tech-Trends.pdf

Dignum, V. (2019). Responsible Artificial Intelligence. In Artificial Intelligence: Foundations, Theory, and Algorithms. Springer International Publishing. https://doi.org/10.1007/978-3-030-30371-6

Elsayed-Ali, S. (2018). New Human Rights principles on artificial intelligence. Open Global Rights. Available on: https://www.openglobalrights.org/new-human-rights-principles-on-artificial-intelligence/?lang=English

Epley & Tannenbaum (2018). Treating ethics as a design problem. Behavioral Science & Policy. 3(2). Available on: https://behavioralpolicy.org/articles/treating-ethics-as-a-design-problem/

Ethisphere (2021). Ethisphere Ethics Index. Available on: https://ethisphere.com/2021-wme-announcement/?__hstc=146412722.b875d3aa3b462032d211c1e2adf7ffc9.1624468978657.1624468978657.1624468978657.1&__hssc=146412722.3.1624468978658&__hsfp=1162251756

Everis (2021). El impacto de la IA en el emprendimiento en América Latina. Available on: https://www.diariosustentable.com/wp-content/uploads/2021/02/ESTUDIO.pdf

Facebook, C Minds, BID (2021). Transparencia y explicabilidad de los sistemas de IA: Prototipado de Políticas Públicas en México. Available on: https://es.cminds.co/prototipo-politica-ia

Fjeld & Nagy (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. Berkman Klein Center for Internet and Society at Harvard University. Available on: https://cyber.harvard.edu/publication/2020/principled-ai

Forum Europe (2021). Envision AI Workshop (short version). Available on: https://datasciencelawforum.eu/netherlands/

Gómez Mont, et al. (2020). La Inteligencia Artificial al Servicio del Bien Social en América Latina y el Caribe: Panorámica Regional e Instantáneas de Doce Países. fAIr LAC – Inter-American Development Bank. Available on: https://publications.iadb.org/es/la-inteligencia-artificial-al-servicio-del-bien-social-en-america-latina-y-el-caribe-panor%C3%A1mica-regional-e-instant%C3%A1neas-de-doce-paises

Kane, et al. (2019). Accelerating Digital Innovation Inside and Out: Agile Teams, Ecosystems and Ethics. MIT Sloan Management Review. Available on: https://sloanreview.mit.edu/projects/accelerating-digital-innovation-inside-and-out/?gclid=CjwKCAjwz_WGBhA1EiwAUAxIcdBVmk5CsScKkeVr6QbP6KmZtlDJx-1Vo2i-Ju7wVtIxTmXdGMPCMWxoCW0UQAvD_BwE

Keller, D. (2019). Dolphins in the Net: Internet Content Filters and the Advocate General's Glawischnig-Piesczek v. Facebook Ireland Opinion. Available on: https://cyberlaw.stanford.edu/files/Dolphins-in-the-Net-AG-Analysis.pdf

Kaye, D. (2019). A New Constitution for Content Moderation. One Zero. Available on: https://onezero.medium.com/a-new-constitution-for-content-moderation-6249af611bd

Latonero, M. (2018). Governing Artificial Intelligence: Upholding Human Rights & Dignity. Data Society. Available on: https://datasociety.net/library/governing-artificial-intelligence/

Lara Gálvez, J. C. (2020). Perspectivas de Derechos Humanos en la implementación de Marcos Éticos para la Inteligencia Artificial. En C. Aguerre, (Ed.). Inteligencia Artificial en América Latina y el Caribe. Ética, Gobernanza y Políticas. Buenos Aires: CETyS University of San Andrés. Available on:

https://guia.ai/wp-content/uploads/2020/05/Lara-Perspectivas-de-derechos-humanos-en-la-implementacion-de-marcos-eticos-para-la-inteligencia-artificial.pdf

López Calva & Melguizo (2021). El impulso para digitalizar bien a Latinoamérica. El País. Available on: https://elpais.com/planeta-futuro/2021-02-18/el-impulso-para-digitalizar-bien-a-america-latina.html

Mac-Gregor, E. (2017). Acerca de la discriminación cultural histórica en razón de la posición económica de los trabajadores sometidos a trabajo esclavo. Instituto de Estudios Constitucionales del Estado de Querétaro. https://archivos.juridicas.unam.mx/www/bjv/libros/10/4745/10.pdf

Markkula Center for Applied Ethics & WEF (2021). Responsible Use of Technology: The Microsoft Case Study. Available on: http://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology_2021.pdf

Markkula Center for Applied Ethics, Deloitte & WEF (2020). White Paper. Ethics by Design: An organizational approach to responsible use of technology. Available on: http://www3.weforum.org/docs/WEF_Ethics_by_Design_2020.pdf

Markkula Center for Applied Ethics & WEF (2021). Responsible Use of Technology: The IBM Case Study. White Paper. https://www3.weforum.org/docs/WEF_Responsible_Use_of_Technology_The_IBM_Case_Study_2021.pdf

UN Mexico (2021). BID y UNESCO lanzan Observatorio de Inteligencia Artificial. https://www.onu.org.mx/bid-y-unesco-lanzan-observatorio-de-inteligencia-artificial/

Rosales, Buenadicha & Narita. (2021). Autoevaluación ética de IA para actores del Ecosistema Emprendedor. Inter-American Development Bank. Available on: https://publications.iadb.org/publications/spanish/document/Autoevaluacion-etica-de-IA-para-actores-del-ecosistema-emprendedor-Guia-de-aplicacion.pdf

Salesforce (2020). How Salesforce Infuses Ethics into its AI. Available on: https://www.salesforce.com/news/stories/how-salesforce-infuses-ethics-into-its-ai/

Sniderman, Bannister & Buckley (2020). Ethical tech: Making ethics a priority in today's digital organization. Available on: https://www2.deloitte.com/us/en/insights/topics/digital-transformation/make-ethical-technology-a-priority.html?id=us:2sm:3tw:4di6289:5awa:6di:MMDDYY:HANDLE:dr26:author&pkid=1006571#endnote-

United Nations (2011). Guiding Principles on Business and Human Rights. Available on: https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf

UNESCO (2021). AI Ethics: Another step closer to the adoption of UNESCO's Recommendation. Available on: https://en.unesco.org/news/ai-ethics-another-step-closer-adoption-unescos-recommendation-0

United Nations Office of the High Commissioner for Human Rights (2021). Artificial intelligence risks to privacy demand urgent action. Available on: https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=27469&LangID=E

***Disclaimer** The opinions expressed in this publication are those of the authors. They do not purport to reflect the opinions or views of CETyS, CLD or of any other organization involved in the project.*