Anthropological problem behind the discrimination generated from artificial intelligence algorithms

Problemática antropológica detrás de la discriminación generada a partir de los algoritmos de la inteligencia artificial

*Gabriela Morales Ramírez**
Universidad Panamericana, Mexico

Abstract

Artificial intelligence is currently at a point of development never seen before, promising great benefits that transcend into different social spheres. One problem in this regard is the apparent neutrality of the algorithms used in their programming and their impact on a large scale in relation to the discrimination generated from the biases immersed in them, coming from their designers. This is the result of a partial view of reality and the person himself. The solution to segregation can be found not only in the so-called parities, which are a response intended to compensate for errors in programming and result in inequalities in

---

* Graduated in Philosophy from the Universidad Panamericana. Master in Philosophy of Science in the area of Philosophical and Social Studies of Science and Technology. E-mail: gmoralesr@up.edu.mx https://orcid.org/0000-0003-1297-2977

---

opportunities and privileges for certain groups, but in a look at the whole person.

Keywords: biases, automatic thinking, algorithmic fairness, neutrality.

1. Introduction

The development and application of artificial intelligence (AI) can represent a considerable progress for science and technology, but it can also appear as a threat to people and their existence on the planet.

Until a few years ago, the use of AI was a topic of science fiction books in stories that seemed too distant and illusory for the time in which they were written, such as Isaac Asimov's I, Robot or Philip K. Dick's *Do Androids Dream of Electric Sheep?* or even cult movies such as the Wachowski sisters' *Matrix*. Today, technological advances and the ethical issues raised in these stories have caught up with us and even gone beyond us.

There is a great ignorance among the general population regarding what AI is, what its uses are, what consequences its development could bring, why it is morally correct, or not, to allocate so much money to its development, etc.

At the same time, there is a great disinterest on the part of governments and large companies dedicated to AI to make evaluations of the possible ethical, political, social, environmental and economic repercussions of their actions dedicated to it. As well as a lack of interest in making a transparent statement of its effects on the daily and future life of society.

In view of this situation, it is essential to address these issues in the light of different disciplines, but above all, to consider their anthropological implications. One of them concerns the so-called algorithmic biases that translate into statistical, structural, cognitive and social errors that bring with them disadvantages that are ethically objectionable because they give rise to discriminatory results or

systematically produce benefits for one group of individuals over others (11).

Therefore, this article will work on the following hypothesis: if AI algorithms contain cognitive biases coming from their designer, then discriminatory models that focus only on accidental aspects of the person are perpetuated.

In order to demonstrate this hypothesis, we will start with some considerations about the person and his or her dignity. We will then turn to cognitive biases and their link to discrimination against people. Later, it will be argued how these biases are transferred to the field of science and technology, especially to the field of AI from the training of algorithms. Some examples of their consequences will be shown and an anthropological basis will be proposed to outline a solution.

2. Some initial considerations on the person and his dignity

If one intends to make a study of the human being, one must start from his substantiality understood as "that, which possesses a totality in itself" (9, p. 11). That is to say, that it does not depend on something else to exist and that it has its own characteristics that distinguish it from anything else, it is endowed with "an existential density so strong that it remains itself through changes" (5, p. 29).

On the other hand, it is also convenient to refer to the so-called accidents that attend to characteristics that may or may not be present but do not modify the being. "That the human being is substance means then that all his qualities can be predicated of him: size, weight, color, age, sex, etc. and that, in turn, these will be accidental, that is to say, whether they are present or not, they will not affect the substance that already is" (9, p. 11).

The word persona, from the etymological point of view, refers to the term *prósopon* that alludes to the masks of the characters of the

ancient Greek theater. In Roman law, *personare* refers to the role of the individual in society. Later, Christianity picks up this term, but emphasizes on the social and human order affirming that person is predicated absolutely of all human beings, designates the uniqueness and unrepeatable character of each one in addition to the equality of all of them before God to reject any possible discrimination. St. Augustine of Hippo points to the idea of the person as a being who participates in the creator God, we all participate in the same way and it is the origin of equality. St. Thomas Aquinas, taking up the definition established by Boethius, says:

> In general, person indicates the individual substance of a rational nature. Individual is that which is indistinct in itself, but distinct from others. Therefore, in any nature, person means that which is distinct in that nature, as in human nature it indicates this flesh, these bones and this soul, which are the principles that individualize man. These principles, even though they do not mean person, nevertheless do enter into the meaning of human person (8).

To reduce the person to his rational dimension is a reductionism that forgets the volitional or affective dimension. In the same way, focusing on the intellectual part of the human being would leave aside his psyche and corporeality. The human being is an individual and unique being as well as a spiritual being who is capable of self-transcendence, of going out of himself. As Burgos (5, p. 29) points out, both men and women "are very special beings because of the intrinsic perfection they possess, which places them above and on a different plane from the rest of the beings of nature".

In modernity, Kant would allude to the notion of the dignity of the human being as the value he has in himself and that, therefore, eliminates any possibility of being bought, substituted or instrumentalised. Unlike objects that have a price, the person has an incalculable value for the mere fact of existing, it is she who gives value to things and the universe itself.

In the twentieth century, in *Populorum Progressio,* an encyclical dedicated to promoting cooperation among nations, Paul VI emphasized the social character of the human being:

> And it is not only this or that man, but all men are called to this full development (...) we are obliged to all and we cannot be disinterested in those who will come to increase still further the circle of the human family. Universal solidarity, which is a fact and a benefit for all, is also a duty (20).

It is a requirement for all of us to recognize ourselves as part of humans and to seek the development not only of some but also of all who belong to it. This will also lead us to speak of the term "human adulthood", which refers to the fact that all people should be able to access the possibility of building their being from a sufficient having. As conditions of possibility to decide to be and reach human adulthood, the need to have is presented, certain minimum conditions are required for their development and respond to their vocation (28). For this reason, Paul VI emphasizes:

> (People must) be freed from misery, find their own subsistence, health, a stable occupation with more security; participate even more in responsibilities, free from all oppression and sheltered from situations that offend their dignity as men; be more educated; In a word, to do, know and have more in order to be more: such is the aspiration of today's men, while a large number of them are condemned to live in conditions that make this legitimate desire illusory (20).

Paul VI's aspiration is that all people should be able to have access to a steady job, to be free from situations that threaten their dignity, to achieve stability and human adulthood. Instead, we find that in spite of the progress that humanity has made, for many human beings, reaching these goals appears to be a mere unattainable dream.

The treatment that a person should have, a being that does not depend on others and that is unique from the rest of existing things, that has a vocation and that is called to achieve it together with others,

is that of respect and recognition beyond its accidental qualities. The question before these initial reflections is why if all human beings are worthy, irreplaceable, we should not be instrumentalised, we have as an end to acquire human adulthood, etc. there are discriminatory practices that promote the fullness of some at the expense of others? The following is a brief reflection on the subject.

3. Discrimination and cognitive biases

Daniel Kahneman and Amos Tversky (14) in their text *Prospect Theory: An Analysis of Decision under Risk* were pioneers in pointing out that the decisions of human beings are neither objective nor informed. This put on the table that partial information added to beliefs, experiences; prejudices and previous knowledge intervene in the behavior and deliberation of individuals.

> People interpret reality and, based on this, they judge and act influenced by the information perceived by their senses and that which they receive and accumulate from their environment, in addition to mechanisms that are not always conscious, but that allow them to make immediate decisions and react to the challenges and questions they are presented with. This response, which is variable in each person and may or may not be attached to rational deliberation, is the product of mental mechanisms called cognitive biases that we use to simplify and facilitate our daily judgments and actions (12, p. 9).

Although human beings are endowed with intelligence, it is hardly admissible to think that all their decisions are accompanied only by reason and that they always reach conclusions endowed with objectivity. On the contrary, the evaluations made about reality are often partial and the decisions made based on them are loaded with previous ideas, opinions and convictions that have not necessarily been demonstrated or rationally justified.

Why can we be brilliant in some things and inoperative in others? Why do we perform some tasks with special skill and not others? Two

types of thinking have been proposed to answer this question: one is intuitive and automatic, which is uncontrolled, does not involve effort, is associative and fast, and the other is rather reflexive and rational, which in contrast is controlled, laborious, deductive, slow, follows rules and is self-conscious (27).

We use one system or another according to the situation we are facing. If a ball is coming at full speed towards us, we will try to dodge it without further reflection. If someone asks how much is 15,345 times 23 most people will use the reflexive system. The automatic system can be very useful, but relying on it completely can be a mistake because many of its conclusions are drawn immediately without any analysis or broad understanding of the problem behind them and are taken as if they are correct even though they are not necessarily so.

People usually have busy lives, which prevents them from reflecting at every moment. When they have to make judgments, because of the need to come up with immediate answers, they do so using basic and automatic rules. They are of course very practical, but they can also bring with them systematic biases known as cognitive biases.

A cognitive bias, then, refers to "the tendency to opt for a specific way of thinking, conditioned by intuition rather than discernment" (29, p. 59). These biases are understood as heuristic shortcuts that allow the human being to give a quick response to certain situations in the environment. This entails imposing on reality a selective and subjective filter of information that will lead the subject to make decisions or carry out wrong behaviors under certain contexts.

Cognitive biases have opened the discussion on how we think and decide, the autonomy with which we choose. The way in which our mind handles attitudes and reactions towards others that may be loaded with heuristics and affirmations in which there is no reflection and, therefore, obtain solutions that concentrate only on a part of reality, but do not contemplate relevant aspects to dictate a valid and true judgment.

A generalized idea as a society, for example, is that the recognition of the common dignity that makes us recognize the other as a person with equal rights and value is seen as a theoretical aspect without great relevance in day-to-day life. This leads to practices of violence, intolerance and marginalization and opens the door to having an incomplete view of people or only seeing some of their dimensions or accidental characteristics that are not involved in a person being a person or having a different or higher level.

Discrimination refers to the differentiation made between some things and others. In itself, it is not a problem, at least not in all cases, since it can serve to distinguish characteristics or determine the treatment to be given, for example, to a person and an object. However, there is a pejorative discrimination that deals with the different treatment given to some groups of human beings because of their gender, color, sexual orientation, among others, with the objective of "maintaining or establishing an oppressive relationship between groups or keeping them in a disadvantaged position" (24, p. 46).

Under pejorative discrimination, that is, when a difference is made between beings who share an ontological nature, the demand to create public policies and seek means to eradicate some distinctions that have been made towards oppressed or excluded groups in an active manner throughout history is indisputable. In this way, human rights will be guaranteed, aiming at the pretension of recognizing all people as free and equal in terms of dignity and rights, without any distinction due to contingent issues of the human being as mentioned above (24).

One could have the false belief that the exercise of automatic thinking is only used in the little transcendent activities of the day to day or in the immediate encounters, we have with other human beings. However, we find that biases and immediate responses are also present in areas, such as science and technology, which in principle integrate reflection, but which are infiltrated by automation precisely because of their ability to offer us resolutions that do not involve great effort and are dynamic according to the situation in which we are standing.

4. Cognitive biases in science and technology development

In the scientific field there is a very competitive struggle to obtain the monopoly of scientific authority, since this supposes having legitimacy. It should be noted that it is human beings who give meaning to scientific practices and their work. Derived from this, the influence of the psyche of those who carry out research with the axioms of science will be recognized.

Scientific knowledge and technological development are the result of the way in which scientists and technologists perform science but above all, of how they learn and conceive it in order to transmit it to others. Authors such as Popper (21) allude to this point when they point out that the choice of a purpose of this type must be the object of a decision that transcends rational argumentation, which concerns the individuality of the subject who works from previously internalized conventions and agreements, far removed from the rationality that later gives rise to science. In other words, not even scientists and technologists escape having a partial view of reality.

If these biases are placed in the field of scientific research, it is possible to speak of inferential illusions. This is because our reason works with premises that are nothing more than inferences. In view of this, we find that many of the scientific theses and classifications that have been accepted for a long time are now studied as a product of cognitive biases. Some of these are as follows:

a) Confirmation bias: involves accepting evidence that supports one's own ideas while adopting a skeptical attitude toward contrary theses by assuming them to be biased. In the scientific field, it is common for people to align the results obtained with their own certainties (29).

b) Halo effect: it occurs when a positive trait of the person is transferred to his research or to his whole person. For example, when it is assumed that an outstanding scientist is always

right and his observations and conclusions are always correct. In turn, this is linked to third parties who cite them as an indisputable source to support their arguments, giving rise to the so-called authority bias (29).

c) Framing effect: this bias occurs when the researcher already has a conclusion in mind and seeks to frame it with the results.

d) Illusion of control: refers to the tendency that it is possible, through control and manipulation, to govern or at least influence events on which it is not possible to fully act. It assumes that it could be observed without error or without any failure (29).

e) Adherence to ideas: scientists analyze the arguments that oppose them in an effort to discover flaws in such a way that they do not allow their results to be easily questioned (30).

Science deals with knowing and understanding the causes of phenomena while technology invents products that do not yet exist but are presented as a solution to current problems. The technological sector incorporates knowledge obtained thanks to scientific research coupled with market information, competitive prices, etc. If scientific progress works hand in hand with the technological field, the latter is not excluded from containing the biases mentioned above.

The situation is problematic because the results of various research projects are not confined to a laboratory or an academic paper, an example of which is AI, which is used to solve multiple practical events. They have repercussions in people's lives, markets and I would even dare to say, in the vision of the world that we have built hand in hand with the progress of science and technology.

5. Considerations around AI

Intelligence is defined in many ways. However, the definition of the philosopher Burgos (5) is taken up again because it emphasizes some of the aspects that show the difference between artificial and human

intelligences: "(it is) the capacity of the person to go beyond himself, transcending himself, accessing the world that surrounds him, understanding it and possessing it in an immaterial way" (p. 65). That is, this conception assumes that intelligence allows the human being to understand, know and access reality and in that sense, to possess it with special emphasis on the abstraction and immateriality of knowledge.

Meanwhile, AI itself is "a branch of computer science (that) deals with methods that enable a computer to solve tasks that, when solved by humans, require intelligence" (3, p. 5). Like other new technologies, AI is also characterized by the possibility of working with uncertainty, inaccuracy, fuzziness and probabilities (4).

In addition to this definition, it is possible to distinguish between types of AI: weak AI is "that in which machines simulate intelligent behavior using mathematics and computer science in a specific area of application and have the ability to learn" (3, p. 5). General AI is "a learning capability in general, including the ability to develop autonomously." (3, p. 5) Superintelligence or strong AI refers to a development superior to that of the human brain in many areas (3).

AI has reached a stage of development where it has the potential to significantly modify life on the planet through its application. Given the potential danger of AI advancement, in 2017 the *Asilomar Principles* were postulated to regulate its limits. Among other things, it is committed to the progress of "beneficial intelligence", a link between science and politics, transparency, accountability, security, service to the common good and specifically:

> 20. Precautionary capacity: in the absence of consensus, we should avoid making assumptions about the upper limits of future AI capabilities.
> 22. Risks: risks associated with AI systems, especially catastrophic or existential ones, should be subject to planning and mitigation efforts commensurate with their expected impact (22).

In view of these principles, some questions undoubtedly arise regarding, for example, the meaning of "beneficial intelligence" and who will be those who will receive that benefit, all humanity or just a few.

6. Biases in AI

AI is presented as a new technology introduced to the market only about sixty years ago. Although its development is quite early, it has been considered that it could be a viable option for decision making in matters, for example, social and economic.

In principle, it is seen as a tool to neutralize the subjectivity that has been associated with human decision making by eliminating discriminatory treatment and biases towards certain individuals or groups. However, systems using AI can have much broader effects and harm many more people without the mechanisms of social control and self-limitation that are present in human behavior (26, p. 2).

AI systems belong to the realm of weak AI, allowing them to perform tasks and provide solutions in particular areas of human knowledge. *Machine learning* or automated learning, also belonging to weak AI, refers to a set of techniques and methods that allow algorithms to extract correlations from data, which constitute the raw material from which learning processes can be automated and unsupervised predictions can be made (11).

We found that there are different forms of learning with respect to AI. One of them is the so-called supervised learning. In this case, the systems are subjected to a directed training process that aims to associate certain characteristics of the data with the labels that correspond to them. In other words, the data are analyzed in such a way as to find elements that allow one category or label to be distinguished from another. For example, if we want to train a model to identify faces in photographs, we would have to enter a database with photographs of people and labels that, at the same time, indicate in which part of the image the face of each of them appears.

Initially, the associations made by the AI will be incorrect, but they will be corrected until, even from the data, it will be able to arrive at new results with data never seen before and establish whether its conclusions are correct. A fundamental premise to consider is that the data with which the model will work in the future will be

somewhat like, but not the same as, those with which the model has been trained (11).

Instead of programming a computer to know how to recognize one image, it receives many, begins to make connections between them, and then weights their characteristics for later use in new images. For example, a photo of a dog is passed to the AI, followed by a photo of a Golden Retriever. The algorithm is informed that both are dogs and it will be able to identify any dog even if it does not have the same characteristics as the examples given initially (18).

Up to this point, it would be feasible to think that these are only issues related to the mere programming of a system without more. However, there are some variations in the data input, coming from what we will call from now on algorithmic biases, which can interfere in a devastating way in the quality of the predictions. As pointed out in the introduction, these biases refer to statistical, structural, cognitive and social errors that entail disadvantages that are ethically objectionable because they lead to results that discriminate against individuals or groups or systematically produce benefits for some over others (18). In other words, they refer to a probabilistic and statistical disparity that comes from an algorithm generated by a computer that follows very specific rules that allow it to make decisions established through different codes (16).

Statistics always have errors, so rather than stopping at this point, two questions arise. The first responds to the need to know if these errors are balanced among the different populations that make up the community and the second to understand where the inequity in the statistical rules has arisen.

The solution to the former refers to the fact that statistical rules are not learned by automated systems out of the blue, but have the possibility of containing biases present in their designer:

> Data are rarely neutral, they are linked to people's experiences and histories, so reducing them to mathematical models without taking into consideration the circumstances surrounding them in order to give them an apparent neutrality, leads inescapably to incomplete and wrong results (15, p. 279).

It is therefore essential to understand how they work, to make them evident and to control them in order to eradicate them and eliminate the discrimination they can bring with them (26, p. 5). The following are some examples:

a) Interaction bias: occurs when the programmer introduces a bias in the model, for example, when defining "success". When a selection of applicants to a university is made, if the programmer has defined a preference that applies only to those who come from certain educational institutions because they are considered academically superior, then there will be an interaction bias because those students who have not been part of these institutions will be rejected regardless of any aspect.

b) Latent bias: refers to when the AI makes inappropriate correlations between the data creating false links. For example, a manager has not hired a certain ethnic group and thinks that these people tend to live in certain areas of the city. When training for the AI is performed, based on previous decisions of that same manager, the system would learn not to select people living in those areas by automating the discarding of applications coming from that group of individuals.

c) Selection bias: when there is insufficient data representative of the diversity existing in a social environment, i.e., there is a disparity in the sample size (26, p. 5). If an AI were trained to predict the skills of the population of the university for the humanities, the algorithm used would be useless to make that prediction in any other university given the low representativeness of that population. Another case is that of:

> Joy Buolamwini, a computer scientist, (who) discovered that her face was not recognized by a facial recognition system while developing applications in a lab at her university's computer science department. Buolamwini discovered that the data (faces) they trained that type of system on were mainly white males. This explained why the system did not recognize his African American face (1).

It has been believed that the results offered by AI are more objective and neutral than those that a person would reach, since they exclude, for example, feelings and emotions, achieving better results that meet the needs of the group to which they are directed.

Despite this, algorithmic systems are sometimes nothing more than "opinions written in code", according to Cathy O'Neil, mathematician and data expert. It is therefore important to consider that they are not just algorithms or mathematical models, but that they have an impact on people's lives. The author states: "I worried about the separation between technical models and real people and about the moral repercussions of that separation" (17, p. 42).

We forget that it is humans who develop and design this technology, which implies that the biases they possess could be transferred to the AI consciously or unconsciously. In this regard, Coeckelbergh (7) says: "often the bias is not intentional: it is common for developers, users and others involved, such as the management of a company, not to anticipate the discriminatory effects against certain groups or individuals" (p. 117).

This leads us to the fact that, if the initial variables and data with which the AI has been trained are biased with prejudices, its results, no matter how good the algorithm used by the AI, will be flawed. If these algorithms are used in a social program, in analyzing whether a megaproject is viable in a territory where a certain group resides, whether or not a person deserves to be subject to credit or hired by a company, etc. then the approval does not depend on mere data, but has behind it a whole contextual framework that will have to be identified, analyzed and that constitutes an irreplaceable part of the development of AI algorithms.

Biases learned by AI are not isolated cases but have been identified in different environments. For example, Clearview AI promised to predict where a crime was going to be committed and identify the perpetrator. It stopped being used in many countries such as Canada when they realized "the tendency to identify people with non-Caucasian features as criminals" (6). In other words, the fact of having Latino or African-American features, minorities in many

of the territories where this system was used, presupposed a greater willingness to commit criminal acts.

Another case is that of *Amazon* when it tried to employ an AI-based recruitment system. However, the system was biased against women because if the words "female" or "women" appeared on the resume when applying for technical roles they automatically received low ratings (25). *Amazon's* approach was to train its recruiting tool based on the identification of the most used keywords in the resumes of the best employees, but without the ability to understand the social context.

According to these examples, we find that algorithmic biases bring with them repercussions that are becoming more and more accentuated. That is, they do not only affect the ten or five hundred people who were not accepted for a job, but generate a general dismissal of certain groups who are denied opportunities because of something as irrelevant as their ethnicity or gender beyond their capabilities.

In the end, we will always have in companies, top managers who comply with stereotypes or in prisons people with a certain skin color under the common belief that this is the right and normal thing to do. We cannot forget that neither "...politics, science, art, religious forms..., are ethically neutral or inhuman or antisocial by nature. It is a human activity and, precisely because it is human, it must respond to humanizing criteria" (19, p. 77). We should not blindly trust an algorithm to make decisions without first ensuring that it has been analyzed and that it has admissible criteria when it comes to evaluating people whose lives have the possibility of being greatly disrupted by partial or unreflective views.

In view of this, we proceed to establish some guidelines that lead us to at least a momentary solution.

7. How to integrate equity in the design of algorithms?

Eliminating discrimination, inequalities and promoting respect for dignity are not new issues, but have been addressed from multiple

perspectives. Today, the challenge does not change when AI plays a role, but it involves certain nuances.

So far, we have said that the person is a unique being, that he or she does not depend on others to exist, irreplaceable, etc. It has also been pointed out that he is a rational being but that his decision-making and worldview is not necessarily guided by it alone. Rather, other factors come into play, such as beliefs and biases that open the door to cognitive biases which, transferred to technology development, can filter into the design of AI algorithms based on data and statistical rules.

Unfortunately, it is impossible to achieve zero error both in humans and in what they produce or project. It would be desirable to achieve excellence in algorithms and statistics avoiding any failure, but in view of their inaccessibility, the application of the so-called "parities" has been proposed to mitigate them:

a) Demographic parity: "refers to a demographic distribution in which it is sought that people who are part of a group of interest are equally represented in a demographic population" (16, p. 141). In other words, there should be a quota that allows a balance to be introduced in the data entered into the algorithm, depending on the case, similar numbers between men and women, Caucasians and African-Americans, among others.

b) Parity of thresholds: establishes whether a decision is admissible as fair by measuring people according to the same criteria without considering their ethnic origin (16). Beyond the differences implied by belonging to a country, having skin of a certain color, gender, the same evaluation should be used with some people as with others. If tests are going to be administered using AI for obtaining a job or access to higher education, the score and difficulty requirements should be the same if they are applied to Americans and Salvadorans.

c) Error parity: this refers to the possibility that when a decision is made based on a statistical rule, there may be an error that

can only be verified a posteriori. If an algorithm were equitable, it would be used in different population groups and would tend to err with the same frequency so that both false positives and false negatives are generated (16). That is, any category of people divided by any criterion is guaranteed to integrate failure as a possible outcome.

It is presented as viable to apply these parities because the responsibility and impact changes when a new technology, in this case AI, transcends the direct person-to-person relationship and has the capacity to normalize and institutionalize biases in a society, not only in the present but also in the long term. It is not the intention here to delve into Jonas' Principle of Responsibility, but it does add to the discussion a central point of his proposal:

> The good and evil for which the action was to be concerned resided in the vicinity of the act, either in the praxis itself or in its immediate scope; they were not a matter of distant planning. This proximity of ends applies to both time and space. The effective scope of the action was scarce. The time for foresight, determination of purpose and possible attribution of responsibility was short. In addition, control over circumstances was limited. Righteous conduct had immediate criteria and almost immediate fulfillment (13, p. 29-30).

That is to say, previously, ethical concerns referred to a closeness in the actions of individuals, responsibility and consequences did not exceed the short term. In the case of AI and the algorithms used, as mentioned in the previous section, they have consequences not only in the present life of individuals but even in future generations that will be affected by the inequality and exclusion resulting from their results.

Added to the parities, a path that would help reduce these difficulties would also be algorithmic audits that, among other things, request information from those responsible for both the design, development and implementation of the algorithm. The methodology used to create it; data on the learning process and operation of the

system; the databases used during the training and a clear definition of the possible vulnerable groups affected by the implementation of the algorithm (10).

By itself, the audit would allow access to an external analysis that verifies that the algorithm is free of biases and that, if it has them; there is a way to mitigate them. On the other hand, knowing who is behind them provides a guideline to establish responsibilities and understand interests and even contexts behind them. Having clarity with the databases implies transparency and openness to admit that it is impossible to include all the variables and that is why the other's point of view is important.

Before going to market, all algorithms should have been audited and passed evaluations aimed at proving that segregation towards certain people is not the common response.

8. Back to square one

The central problem of algorithmic biases, one would think, lies in the data that are fed into the systems, the low representativeness of some groups, etc., so that a first solution would be to introduce parities. Paradoxically, these bring with them the option that the error appears with the same constancy in some groups as in others. In other words, it is not only a few groups that are affected, but there is the possibility that anyone could be affected. This begs the question, is this desirable?

The answer does not lie in perfecting facial recognition or determining a parity quota to ensure that the number of data entered into the system is admissible. In the end, this does not guarantee that discrimination will be eliminated (2,14). The real solution is to look at the whole person. We must recognize that we have before us a valuable being, worthy, deserving of reaching human adulthood, of achieving the best version of himself and responding to his calling.

Cognitive biases are still present in the way we observe the world and make decisions. Even so, it is possible to lessen their impact or

even eradicate them if, before deciding who is worth more or better, we reflect on everything we have learned, the beliefs we have acquired, and see the person instead.

Algorithmic biases are nothing more than the reflection of a society that is divided by unjustified prejudices, that throughout history has been concerned with differentiating rather than building bridges, and that has put other interests before the person.

> To explain how discrimination occurs through AI systems exclusively from a technical point of view would be a limitation (...), AI is a socio-technical concept that can only be explained by taking into consideration purposes, motives and social relations that influence its development and implementation (15, p. 281).

Despite this, it is not a reason to stop using AI at present, much less to stop its development because, as Idoia Salazar, president of Odise IA, the Observatory of the Social and Ethical Impact of Artificial Intelligence, points out:

> AI is software with the ability to analyze data, draw conclusions, make decisions autonomously and learn. It is a technology with enormous possibilities to help us have a better life if it is used for good (23).

It is an opportunity to rethink the way we treat our fellow human beings and seek measures to mitigate possible damage to society, whether intentional or a mere accident of thoughtlessness, as there is responsibility in this.

9. Final comments

Discrimination comes from differentiating people by focusing on accidental characteristics such as skin color, sexual preference, age, height, as if these defined their being and it depended on these particularities to be more or less human and therefore to have more or less value.

Discrimination against others is due to a lack of attention to what it truly means to be a person and to respond to the respect that each one of us deserves for being ontologically worthy.

According to the hypothesis put forward, as mentioned above, cognitive biases are constantly present in our way of reasoning and acting towards the world and people, which leads us to have a partial vision of reality that would allow us, for example, to see in others only their accidents and not their substantiality as persons. This applies to all people, even to scientists, technologists, algorithm designers, etc. As mentioned above, biases are often introduced unintentionally. To blame them would be, once again, to pretend to work with automatons exempt from human faculties such as reason, affectivity, will, freedom and their own biography.

On the other hand, we cannot forget that if these latter actors also use automatic and fast thinking to obtain answers and from there programmed, opinions, prejudices and unjustified beliefs can easily be filtered out. Faced with this, then, we as humanity are presented with at least three options: reflect on the biases, determine their negative impact and seek means to eradicate them, or simply ignore them or deny their existence.

Taking these biases into the realm of algorithms and AI machine learning represents major problems because their impact transcends the practical realm in the lives of people and their environment. The intention is not to build a wall to impede scientific and technological progress, but rather to establish minimum guidelines to ensure that AI is used for the benefit of human beings, that it achieves inclusive and equitable progress for all, eradicating the illusion of neutrality and that it is, above all, capable of responding to the demands of society.

Behind these biases and segregation itself lies a rejection of what is different or of what is shown as separate from an "I" or an "us" that is perpetuated from some human beings towards others. When these sectioned views of reality are transferred to algorithms, which will have an impact on groups of people, discrimination by one aspect or another increase exponentially and even becomes normalized.

It is impossible to enclose people in categories because that would imply falling into a poor and cut observance of them. Paradoxically, human intelligence and the products developed by it, such as science and technology, operate in this way. We are forced to divide reality, generate models, include and exclude variables because we cannot know everything at the same instant. If this happens with the world, it is an illusion to pretend to know a multidimensional and ineffable being such as the person in an absolute way and from that to build the rest.

Knowing the limitations, in turn, is what opens the possibility of avoiding thinking that our gaze, that of scientists or technologists, is unique and all encompassing. It is practically impossible to keep ourselves in reflective thinking and to analyze each of the steps we take. However, if we manage to contrast the algorithms not with other algorithms or reduce their study to their effectiveness according to criteria established by some, but with the examination of more people and all that is behind them, such as their contexts, histories, ways of understanding the world, we will gradually create algorithms that respond better to what it entails to see the person.

Today there is no rule that allows us to build algorithms free of discriminatory biases, but there are people who are able to apprehend a little of who is in front of them. Like a jigsaw puzzle, one places the piece that someone else may had not seen or was trying to put in the wrong place. For example, when constructing an algorithm someone has considered that it will include equally men and women of productive age to generate health care insurance and considers only those with a paid job. Someone else realizes that he has forgotten the domestic workers who also perform indispensable work to keep society moving and who could be eligible for insurance, even if their work is not paid for monetarily. In other words, it is the other person who helps us to see this and many other nuances and circumstances of the person off the radar, to contrast with ourselves, until we gradually build the whole.

A further step will be to recognize that algorithms cannot be considered universal or permanent. They must be under constant

revision according to firm foundations such as dignity, respect for difference that allow us to navigate the new AI horizon. Once a flaw is identified in them, we will have to, in the first instance, rely on parities and algorithmic audits but in the face of multiple inconsistencies we will have to discard them and build new ones. It will be the disparity of the other that breaks the fragility of the biases and prejudices that up to now we have accepted it as immovable.

The AI is a computational system subjected to the decision between different options programmed according to an algorithm, at least until now, unable to leave itself. The person is able to choose, create, account for failures and invent new scenarios beyond his individual situation. To expand the analysis in the design of algorithms is to recognize the complexity of the person and to realize that it is in our hands the important decision to face our preconceived ideas that we have adopted without justification and to get involved in dazzling ourselves before the immensity of the other.

It is a reality that zero error does not exist both in statistics and in our thinking, but it generates an important change to put on the table that, given the situation in which we are standing today technologically speaking, we cannot stop studying and denouncing the algorithms that perpetuate models of injustice and discrimination towards people.

The anthropological perspective on the subject here discussed is indispensable because any progress in the field of knowledge or technology must be based on a correct view of the person. If we do not know exactly what or who the person is, his or her faculties, why he or she is different from other creatures, why he or she is worthy, then algorithms would simply obtain answers.

It is true that one ambition is to make processes more efficient, but the starting point and the end is the person. The person gives meaning to our work as humanity, as well as the demand that we make so that everyone is recognized as valuable and has the same opportunities. This under the aspiration of building together a society made up of people who can reach their fullness and develop with excellence.

References

1. Adetunji J. Los sesgos en inteligencia artificial, el reflejo de una sociedad injusta: The Conversation [Internet]. 2021 May 17 [Accessed 2022 Apr 20]. Available at: https://theconversation.com/los-sesgos-en-inteligencia-artificial-el-reflejo-de-una-sociedad-injusta-160820

2. Baeza R, Muñoz C. Académicos viendo Netflix: sesgos codificados. CIPER Académico [Internet]. 2021 May 8 [Accessed 2022 Nov 2]. Available at: https://www.ciperchile.cl/2021/05/08/academicos-viendo-netflix-sesgos-codificados/

3. Schmiedchen F, Bartosch U, Bauberger S, Stefan S, von Damm T, Engels R, Rehbein M, Stapf-Finé H, Sülzen A. Informe sobre los principios Asilomar en Inteligencia Artificial. Berlín: Grupo de Estudio Evaluación de la tecnología de la digitalización de la Federación de Científicos Alemanes; 2018. https://vdw-ev.de/wp-content/uploads/2019/05/Informe-sobre-los-principios-Asilomar-en-Inteligencia-Artificial_final.pdf

4. BIKTOM. Künstliche Intelligenz verstehen als Automation des Entscheidens. Berlín: Leitfaden; 2018. https://www.bitkom.org/sites/default/files/file/import/Bitkom-Leitfaden-KI-verstehen-als-Automation-des-Entscheidens-2-Mai-2017.pdf

5. Burgos J. Antropología Breve. España: Palabra; 2010.

6. Charte F. Qué peligro implican los sesgos en los modelos de inteligencia artificial: Campus MVP [Internet]. 2021 May 17 [Accessed 2022 Apr 25]. Available at: https://www.campusmvp.es/recursos/post/que-peligro-implican-los-sesgos-en-los-modelos-de-inteligencia-artificial.aspx

7. Coeckelbergh M. Ética de la inteligencia artificial. España: Cátedra; 2021.

8. De Aquino T. S Th.: HJG [Internet]. 2012 Sep [Accessed 2022 Nov 2]. Available at: https://hjg.com.ar/sumat/. I, q. 29, a. 4.

9. De los Ríos M. ¿Quién es el ser humano? Bioética. Aporte para un debate necesario. Mexico: Fundación Rafael Preciado Hernández; 2018. p. 11-27.

10. Eticas Research and Consulting SL. Guía de Auditoría Algorítmica [Internet]. 2021 [Accessed 2023 Jan 3]. Available at: https://www.eticasconsulting.com/wp-content/uploads/2021/01/Eticas-consulting.pdf

11. Ferrante E. Inteligencia artificial y sesgos algorítmicos ¿Por qué deberían importarnos? Nueva Sociedad: Fundación Friedrich Ebert, 2021; (294):27-36.

12. González L. Discriminación, discriminación peyorativa y la Declaración Universal de los Derechos Humanos. Aguilar A. Discriminación, sesgos cognitivos y derechos humanos: perspectivas y debates transdisciplinarios. Mexico: UNAM; 2022. p. 9-12.

13. Jonas H. El principio de responsabilidad. Ensayo de una ética para la civilización tecnológica. Barcelona: Herder; 1995.

14. Kanheman D, Tversky A. Prospect Theory: An Analysis of Decision under Risk. Econometrica. 1979; 47(2):263-291. https://doi.org/10.2307/1914185

15. Muñoz C. La discriminación en una sociedad automatizada: Contribuciones desde América Latina. Rev. chil. derecho tecnol. (online) [Internet]. 2021 Jun 30 [Accessed 2023 feb 4]; 10(1):271-307. Available at: https://rchdt.uchile.cl/index.php/RCHDT/article/view/58793

16. Noriega A. Discriminación algorítmica y costo de equidad. Aguilar A. Discriminación, sesgos cognitivos y derechos humanos: perspectivas y debates transdisciplinarios. Mexico: UNAM; 2022. p. 139-144. https://biblio.juridicas.unam.mx/bjv/detalle-libro/7065-discriminacion-sesgos-cognitivos-y-derechos-humanos-perspectivas-y-debates-transdisciplinarios-coleccion-pudh

17. O´Neil K. Weapons ot Math Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown; 2016.

18. Ortega A. La imparable marcha de los robots. España: Alianza; 2016.

19. Osorio B. Antropología de la donación: el don como principio de la acción humana. Escritos. 2015; 23(50):67-82.

20. Pablo VI. Carta Encíclica Populorum Progressio del Papa Pablo VI a los obispos, sacerdotes, religiosos y fieles de todo el mundo y a todos los hombres de buena voluntad sobre la necesidad de promover el desarrollo de los pueblos [Internet]. 1967 Mar 26 [Accessed 2022 Nov 3]. Available at: http://w2. vatican. va/content/paul-vi/es/encyclicals/documents/hf_p-vi_enc_26031967_populorum. Html

21. Popper K. La lógica de la investigación científica. Madrid: Tecnos; 1977.

22. ROBOTechnics. Principios de Asilomar de la Inteligencia Artificial [Internet]. 2017 Nov 11 [Accessed 2022 Apr 22]. Available at: https://www.robotechnics.es/asilomar/

23. De los sesgos a la manipulación, la cuestión ética es ineludible en el desarrollo de la inteligencia artificial: Nektiu [Internet]. 2021 Jun 24 [Accessed 2022 Apr 24]. Available at: https://nektiu.com/de-los-sesgos-a-la-manipulacion-la-cuestion-etica-es-ineludible-en-el-desarrollo-de-la-inteligencia-artificial/

24. Risse M. Sobre los sesgos cognitivos y los derechos humanos. Aguilar A. Discriminación, sesgos cognitivos y derechos humanos: perspectivas y debates transdisciplinarios. Mexico: UNAM; 2022. p. 46-57.

25. Sabán A. Amazon desecha una IA de reclutamiento por su sesgo contra las mujeres: Genbeta [Internet]. 2018 Oct 10 [Accessed 2022 Apr 23]. Available at: https://www.genbeta.com/actualidad/amazon-desecha-ia-reclutamiento-su-sesgo-mujeres.

26. Sánchez M. Prevenir y controlar la discriminación algorítmica. RC D [Internet]. 2021 [Accessed 2022 Apr 18]; (427). Available at https://www.researchgate.net/publication/358207305_Prevenir_y_controlar_la_discriminacion_algoritmica

27. Sunstein C, Thaler R. Un pequeño empujón. El impulso que necesitas para tomar mejores decisiones sobre salud, dinero y felicidad. US: Taurus; 2008.

28. Verdoy A. El concepto de progreso en la doctrina de Montini. Sols J. La humanidad en camino. Medio siglo de la Encíclica Populorum Progessio. Barcelona: Herder; 2019. p. 12-83.

29. Villarruel-Fuentes M. El quehacer del científico: una perspectiva crítica desde referentes psicológicos. Revista Ensayos Pedagógicos, 2019; 14(1):55-68.

30. Vinck D. Ciencias y sociedad: Sociología del trabajo científico. Barcelona: Gedisa; 2014.