filosofia unisinos | ISSN 1984-8234
Unisinos Journal of Philosophy

Filosofia Unisinos
Unisinos Journal of Philosophy
25(1): 1-17, 2024 | e25109

Unisinos – doi: 10.4013/fsu.2024.251.09

Article

Algorithmic injustice and human rights[1]

Injustiça algorítmica e direitos humanos

Denis Coitinho
https://orcid.org/0000-0002-2592-5590
Universidade do Vale do Rio dos Sinos - Unisinos, Programa de Pós-Graduação em Filosofia, São Leopoldo, RS, Brasil. Email: deniscs@unisinos.br

André Luiz Olivier da Silva
https://orcid.org/0000-0003-2828-0596
Universidade do Vale do Rio dos Sinos - Unisinos, Programa de Pós-Graduação em Direito, São Leopoldo, RS, Brasil. Email: andreluiz@unisinos.br

ABSTRACT

The central goal of this paper is to investigate the injustices that can occur with the use of new technologies, especially Artificial Intelligence (AI), focusing on the issues concerning respect to human rights and the protection of victims and the most vulnerable. We aim to study the impacts of AI in daily life and the possible threats to human dignity imposed by it, such as discrimination based on prejudices, identity-oriented stereotypes, and unequal access to health services. We characterize such cases as algorithmic injustices. In the final part of the text, we propose some strategies to confront this specific type of injustice.

Keywords: algorithmic injustices, artificial intelligence, victims, human rights.

RESUMO

O objetivo central deste artigo é investigar as injustiças que podem ser ocasionadas pelo uso de novas tecnologias, especialmente, a inteligência artificial, tendo por foco central o respeito aos direitos

[1] This article was prepared within the scope of the research project entitled "Human rights and artificial intelligence: from the violation of personality rights to the need for regulation of new technologies", which was included in the CNPq/MCTI Call No. 10/2023 – UNIVERSAL and received support of National Council for Scientific and Technological Development – CNPq.

humanos e a proteção dos mais vulneráveis ou vítimas. Quer-se estudar os impactos da IA em nossas vidas e a possível ameaça à dignidade humana, identificando as possíveis injustiças ocasionadas, tais como discriminações preconceituosas, estereótipos identitários e acesso desigual ao sistema de saúde, entre outros, caracterizando-se como injustiça algorítmica. Na parte final do texto, propomos algumas estratégias para enfrentar este tipo específico de injustiça.

Palavras-chave: injustiça algorítmica, inteligência artificial, vítimas, direitos humanos.

1 Initial considerations

Artificial Intelligence (AI)[2] oriented technology is already a constant presence in everyday routine and it appears to facilitate our lives. For instance, in using streaming services like Amazon Prime Video, Disney+, Netflix, and Globoplay, it is common to receive recommendations of movies or TV shows based on consumed content, as well as receiving recommendations of what wine to drink based on the history registered in apps like Evino. However, this technology is progressively reaching a field in which, most likely, the impact would be far more relevant. Another incursion is on decision-making in high-risk circumstances, establishing priorities between people, and making complex judgments that demand moral criteria. For this reason, reflection on the algorithms that feed those and similar products looks urgent. Algorithms are already in use by the police and the justice system in some countries. They determine who should be arrested based on facial recognition, and they also give advice to judges about parole and higher prison sentencing. In a substantial number of cases, the algorithms do not take into account the important racial questions involved, which can result in racism. Even innocuous apps like Spotify can reveal ethical problems in its algorithm, such as sexism. The research conducted by Andres Ferraro, Xavier Serra, and Christine Bauer shows that the platform unequally highlights male musicians, thus the female artists are less recommended. When they tested the algorithm, they discovered that the six first tracks were from male artists (Ferraro; Serra; Bauer, 2021, p. 249-252).

One notices that the basic idea of AI is to facilitate life, those algorithms can, theoretically, make decisions without cognitive biases, like tribalism and biases of gender, race, or class, logically and rationally, avoiding partial judgment and discrimination of any kind. But that might not occur as expected, since, inevitably, the decisions mentioned earlier involve ethical criteria and complex moral evaluation, which can lead to injustices. The problem we currently face is that some uses of AI, like facial recognition software, CV evaluation by employers, classification of pictures by topic, and software influencing penal-judicial decision-making, reveal worrisome disparities in performance on the basis of gender and race. Those disparities raise urgent questions about how AI usage can function to promote justice or consolidate some injustice, it can, for instance, reproduce institutional injustice (Shelby, 2007), and structural injustice, like racism, sexism, and class prejudice (Young, 2011).[3]

Recent developments show several cases of AI's potential to reflect human prejudice. Let's take the AI CV evaluation tool from Amazon as an example. The tool automatized sexism, it systematically favors

[2] AI refers to the use of machines and software to accomplish tasks that usually demand human intelligence. It is the capacity of a system, i.e, the software incorporated in a device, of executing tasks commonly associated with intelligent beings. Also, the capacity of a digital computer or robot controlled by a computer to execute tasks. See Frankenfield, 2022. Also see Copeland, 2001.

[3] Iris Young, in *Responsibility for Justice*, argues that injustice is more than simply a matter of people suffering an undeserving destiny, the way how institutional rules and social interactions conspire to limit the options of people is also part of it. For her, promoting justice and its consequences in the social structure implies the restructuring of institutions and social relations to prevent those threats to the basic well-being of people. See Young, 2011, p. 3-41.

resumes from men, apparently due to the fact that the system was trained with the data of resumes previously sent to Amazon, the majority being resumes from men (Dastin, 2018). In the same vein, one can point to the AI made by Google to classify pictures, the tool erroneously labeled black people as "Gorillas", which shows a racist stereotype (BBC, 2015). Or, we can consider the analyses made by Buolamwini and Gebru (2018) of three facial recognition software. They show the worse performance of the software in relation to women of darker skin, which can result in legal injustices towards black women (Buolamwini; Gebru, 2018, p. 1-8).

The relevance of the debate is of important note because it is also perceived in new police and political and social agendas. In fact, international documents, like the UN's "2030 Agenda"[4] and many others, in which Human Rights and social justice appear as fundamental pillars to guarantee human societies' future, including the thought of the impacts of AI in daily life, especially in the most vulnerable citizens. Having that in mind, the rest of the paper deals with the issue of algorithmic injustice and human rights. We first define algorithmic injustice and reflect on how it leads to disrespect to human rights, to later exemplify the phenomena. We analyze the cases of facial recognition software used by the police and software predicting the criminal recurrence of already convicted individuals employed by the justice system of some countries, how they (re)produce injustices, as well as the cases of algorithmic discrimination in decision-making processes. The final section of the paper proposes possible lines of action to face this difficult problem, one ethical and another political, with special emphasis on the demand for algorithmic transparency.

2 Algorithmic injustice

Prior to analyzing the cases of algorithmic injustice, it is productive to understand conceptually what is involved in this phenomenon. Let us begin, then, by independently understanding the terms "algorithm" and "injustice", to later offer an initial definition, clarifying the reasons why this type of injustice disrespects human rights.

On the one hand, algorithms are computer programs' logical structures and currently make up AI systems with wide applications in society. They are made of programming codes, which feed on databases. Web structures, as they are known, are composed of codes. These codes are written in a language specific to the machine, functioning as a list of commands. When writing code, a sequence of tasks is established so that the machine understands and executes as expected. The act of organizing steps logically and stipulating a sequence of steps to perform a task is called programming logic. The narrative sequence of these events is called an algorithm. Thus, an algorithm behaves the way it was programmed to behave. The problem in question is that since the algorithm is crafted by the programmer, it may be the case that she works with restricted or even biased data, and thus, the algorithms will reproduce the restrictions or discriminations on a large scale, which may result in racism, sexism, ageism, or any number of arbitrary discriminations that must be fought if one has a prosperous and stable society in mind (Baer, 2019).[5] Injustice, on the other hand, is commonly understood as the absence of justice, being classified especially as the

[4] In 2012, the UN instated 17 sustainable development objectives for overcoming the biggest challenges of our time, like eradication of poverty and goals related to equality, care for the planet and the improvement of quality of life to all, it established 196 goals that aim at the balance of the three dimensions of sustainable development, i.e, the economic, social and environmental dimension. See *The 17 Goals*: https://sdgs.un.org/goals.

[5] More specifically, the problem at hand is due to the fact that with the spread of increasingly sophisticated AI tools, society will demand institutional responses to the problems caused by possible algorithmic discrimination, such as facial recognition systems accessing that non-whit people have a greater chance of greater criminal recurrence based on the analysis of their skin tone alone. Especially in Brazil, where a considerable part of the population is black or brown, in addition to indigenous populations, this is extremely worrying and must be carefully observed by society. It is not about being against this technology, but about being aware of possible injustices.

absence of impartiality and equity, or even as opposed to greed. According to tradition, justice is a normative-moral standard that is universal, being represented, above all, by the idea of equity and impartiality. Thus, injustice would occur in situations of partiality in a decision, such as a judicial decision that takes into account friendship, as well as in a circumstance in which equity is denied to citizens, as in the situation of denying equal rights to citizens for arbitrary reasons.[6]

In contrast to more orthodox views, Judith Shklar, in *The Faces of Injustice* (1990), characterizes injustice as a phenomenon both particular and individual, having many faces. For this reason, history, culture, and status play a fundamental role in its identification, accordingly, injustice is thought from the perspective of the victims. It is important to highlight that Shklar argues that the victims' perception is central to distinguishing between a fatality, bad luck, and injustice, insofar as it allows one to connect primarily with the victim's sense of injustice (Shklar, 1990, p. 1-14).[7]

In our opinion, Shklar correctly argues that the voice of the victims is privileged because it is the voice without which it is impossible to decide whether they have suffered an injustice or a fatality. From the point of view of the victim, whether the victimization results from a fatality or from social discrimination, injustice must include not only the immediate cause of the disaster but also the refusal of people and institutions to prevent and mitigate damage, which characterizes passive injustice. In his words: "My real object is personal and political injustice and the ways in which we respond to it as agents and especially as victims" (Shklar, 1990, p. 14).

The broader definition of injustice requires greater responsibility from both citizens and governmental agents in dealing with the suffering of others, which seems to be an adequate criterion for investigating the injustices befalling the most vulnerable citizens of any given society. We would deem it important to emphasize that, for her, no traditional model of justice offers an adequate conception of injustice because they are based on the unfounded belief that we know and make a stable and rigid distinction between injustice and bad luck. The problem is that this belief tends to ignore passive injustice, the victim's sense of injustice, and the aspect of injustice as a social phenomenon.[8] For her, traditional theories of justice:

> *(…) take it for granted that injustice is simply the absence of justice, and that once we know what is justice, we will know all we need. That belief may not, however, be true. One misses a great deal by looking only at justice. The sense of injustice, the difficulties of indetifying the victims of injustice, and the many ways in which we all learn to live with each other's injustices tend to be ignored, as is the relation of private injustice to the public order (Shklar, 1999, p. 15).*

The cases that can be classified as algorithmic injustice or algorithmic discrimination are circumstances in which the algorithm makes exclusionary decisions concerning a race, gender, or social class, discriminating against more vulnerable social groups. This phenomenon is characterized by ongoing discrimination against the most vulnerable through the functioning of software's algorithms, and it can

[6] Chaim Perelmann outlines the maxim "to be fair is to treat equally" as a universal conception of justice, and that the discussion intensifies when one asks whether everyone should be treated the same or whether we should establish distinctions. Thus, justice has an intrinsic connection with the way in which we perceive and treat equality/inequality. For him, to be fair is to treat in the same way beings who are equal from a certain point of view, who have the same characteristic in regards to the one that must be taken into account in the administration of justice (Perelman, 2005, p. 18-19).

[7] She begins by making an important distinction. She wonders how to know when a disaster is a fatality or an injustice? For her, if it is caused by an external force of nature, like a hurricane, it is a fatality and we must resign ourselves to suffering. On the other hand, if it is caused by an agent with bad intentions who makes a deliberate decision, then it is an injustice, so we should express strong indignation. For example, an earthquake is clearly a natural event, but the damage caused by it can have relevant personal or social causes, as in the case of builders not following the correct construction plan to save material, or even in the case of public authorities who have done nothing to prepare for that eventuality. See Shklar, 1990, p. 1-14.

[8] Shklar explains the sense of injustice as (i) a special kind of anger one feels when promised benefits are denied and when one does not get what one believes one is entitled to, and (ii) as a betrayal experienced when others thwart one's expectations. See Shklar, 1990, p. 83-126.

also be understood as an algorithmic bias.[9] This clearly violates human rights, as AI has no greater obstacles in invading the privacy and intimacy of users. This technology, therefore, encourages the discriminatory use of data, unequal treatment, and violation of the idea of equality that permeates every definition of human rights.

In this sense, the use of new technologies can harm human rights, since it violates the basic assumption that all human beings have equal freedom to make decisions about their own lives. We are not referring only to the human rights that have been institutionalized in the international community since the Universal Declaration of Human Rights of 1948 (and other international protocols and conventions), these normative texts expressly declare rights that must be protected, including from violations caused by new technologies. We are also talking about those rights that, from a moral point of view, concern the most basic and personal decisions a human being can make and correspond, in general terms, to decisions about what it means to live life. These are rights that concern a person's autonomy to make choices and decisions.

It should be remembered that the philosophical idea of human rights (Beitz, 2009; Campbell, 2001 and 2018; Cranston, 1973; Donelly, 2013; Nickel, 2007; Tasioulas, 2015) is closely linked to the notion of justice and is anchors itself specifically on the binomial of equality and freedom. The philosophical idea has evolved significantly throughout its history[10] and, with new technologies, perhaps these rights still need to acquire new meanings. Since the liberal revolutions and, in particular, after World War II, with the Declaration of 1948, the idea of human rights has been reaffirmed in a global community, giving rise to an institutional interpretation of human rights.

The normative devices about human rights were enunciated in an analogical world. That doesn't mean they fail to refer to rights today. Regardless of the impact of new technologies and AI, they make a lot of sense and should certainly be taken into account when thinking about human rights in a digital context. The Declaration of 1948 says, in its first article, that human beings are "free and equal in dignity and rights". They are therefore equal in rights and cannot be discriminated against. Art. 7, incidentally, highlights the "equal protection against any discrimination". Art. 12 repels interference or attacks against the human person "in his private life, in his family, in his home or his correspondence".

Likewise, the Brazilian Federal Constitution of 1988, also guarantees in Art. 5 the inviolability of freedom and equality, detailing the inviolability of "intimacy, private life, honor and image" (item X), and the "secrecy of correspondence" (item XII), also highlighting the right to claim compensation for damage caused by violations of these rights. Both the 1948 Declaration and the 1988 Brazilian Constitution are normative texts pointing to freedoms and equalities in the form of general and open principles, in addition to being documents designed to preserve and protect rights in an analog world, when the debate on AI was not passing mere projection of science fiction books. These devices are fundamental, but, because they were created in an analog world, they lack specific and adequate regulation so that the rights enshrined in the declarations can be protected at the present.

To get an idea of the size of the challenge that AI poses to humanity, it is enough to observe how privacy and intimacy were constituted, throughout history, in international human rights protocols. From

[9] Algorithmic bias is a type of distortion of judgment related to the construction of the algorithm, it can produce or magnify situations of racism, sexism and even violation of consumer rights. Algorithmic bias occurs when the algorithms used in decision-making processes reproduce prejudices and discrimination present in society. It can occur in different ways, such as with the use of historical data that contains prejudices, with the selection of variables that are correlated with discriminatory characteristics and with the lack of data on minority groups. In the health sector, for example, it can lead to incorrect diagnoses or lack of treatment for minority groups, and in the area of public safety, it can lead to discriminatory approaches and human rights violations. On the subject, see Baer, 2019.

[10] The idea of human rights has been expanded since the liberal declarations (Bobbio, 2004; Ignatieff, 2001; Raz, 2010 and 2011; Tasioulas, 2012). After the Second World War, when the paradigm for justifying these rights ceased to be morality (Gewirth, 1982; Griffin, 2008; Nino, 1989; Wellman, 2010) moving to a perspective of International Human Rights Law (Clayton, 2009; Donelly, 2012; Tasioulas, 2019), i.e., human rights came to be understood as institutionalized rights legally recognized by the international community.

1948 until very recently, privacy protection was limited to keeping paper mail delivered by the postal service secret. Today, paper correspondence is in disuse and nobody communicates how they used to, going back a few decades ago. Of course, correspondence, in the sense of exchanging and sharing data and information, never ceased to exist or was left aside. Communication – and correspondence – has only intensified, in the context of AI, privacy protection becomes even more relevant and complex, especially when dealing with personal data privacy. Privacy and intimacy (Posner, 1978 and 1979; Schoeman, 1984) remain fundamental and non-negotiable values in the protection of human rights (Roessler, 2004, 2015, and 2017), however, the context of AI demands new approaches and specific regulations precisely to ensure that privacy and intimacy are properly protected in an increasingly AI-driven world.

To avoid or mitigate the possible side effects (Tasioulas, 2019) that AI-driven technology can cause, technological advancement and its increasing application in various sectors of society reveal a range of practical problems and require an in-depth analysis of the possible impacts of AI on human rights. Artificial intelligence technologies, when applied on a large scale, can escape traditional mechanisms of human rights protection, whether with regard to privacy or with regard to equal treatment and the requirement for non-discrimination. In the sections that follow, we will see some examples of the phenomenon, such as facial recognition programs and judicial decision programs in the criminal sphere that are discriminatory, especially with the black population, as well as we will also address algorithmic discrimination in decision-making processes.

3 Facial recognition and Compas: two examples

Let's start with the facial recognition program that is being widely used around the world as a way of identifying people, whether in airports, buses, malls, residential buildings, and various public places, especially as a way of identifying criminals and arresting them. This type of technology is used as a way to ensure public safety, being used by police forces and various security agents. In Brazil, one can remember the famous case of Salvador's 2019 Carnival, in which a criminal was arrested by the Bahian police thanks to this technology.[11]

Facial recognition is a biometric system that uses machine learning techniques, in conjunction with artificial neural networks, to "match" an image with a person's profile. This system is a technique based on the fact that each person has a characteristic facial pattern in which, using some deep image analysis system, individuals can be identified. It is a technology that works through an AI algorithm, capable of analyzing in detail the image captured by a camera, defining an individual pattern for the verified faces. This way, it is possible to feed a database with the information extracted from this technology and consequently identify people. It is a technology made possible due to the formation of a significant database, bringing together large volumes of biometric information used in the process of identification. The need to use this new technology is probably associated with the process of accelerated urban and population growth in cities around the world, and the consequent increase in surveillance.[12]

Despite the benefits, the technology raises an important discussion about the accuracy of the analysis. This is a challenging task considering the complex and diverse societies in which we live. Due to several components, such as the technical angle at which the camera captures the image, the lighting

[11] A 19-year-old who was on the run from the police on a murder charge was arrested while enjoying the Salvador Carnival in 2019, dressed as a woman, after having an image of his face recorded by a facial recognition camera. The recognition was carried out by comparing the images of the people who had access to the circuits with the database of the Secretariat of Public Security (SSP). See report by Alves, 2019. And, in 2020, 42 fugitives from the Bahian police were arrested at Salvador's carnival using this technology, and in the 2023 carnival, there were 77.

[12] Costa and Kremer consider that, in this context, the demand for greater state control has increased, and it is common for public administrations to use these new forms of surveillance over people, which clearly brings forward the issue of invasion of privacy. See Costa; Kremer, 2022, p. 147.

and facial expressions, and the sociological framework from which the database is collected and filtered, it is necessary to reflect on the processes behind these programs coming from contexts normally ignored in society. There has been a significant increase in unfair arrests of non-white people resulting from the use of facial recognition programs, which may indicate that these algorithms are being trained from a biased database, with little diversity. It is important to note that although facial recognition programs have become more accurate, several studies have shown that they are susceptible to errors when accessing people with darker skin. This is probably because the technology uses data that do not represent the ethnic diversity of the population, resulting in identification errors.[13]

Analyzing specifically the Brazilian context, despite advances in public safety and criminal investigation, facial recognition programs have revealed a worrying side of racism, as cases of innocent black people being arrested for crimes they did not commit are not rare. The article by Hellen Guimarães gives us an interesting example of this (Piauí, 2021). She gathered several such cases, reporting the arrest of a data scientist, a motorcycle taxi driver, an app driver, a cellist, and a cultural producer, all black men. The problem is that the software used in facial recognition reproduces the perceptions of those who create them and the treatment given to the databases that feed them. Structural questions about how society and the state determine who are the individuals who pose danger and should be detained appear to be the reason. Her investigation reveals that socially vulnerable populations have been constantly subject to the automation of embarrassment and violence. Also, the technologies aid undue police approaches and untrue attribution of criminal records. This was the case of data scientist Raoni Lázaro Barbosa, unfairly arrested at the gate of his home in September 2021, accused of being part of the police wanted database for belonging to a militia in Duque de Caxias (where the data scientist never lived). He was confused with Raoni Ferreira dos Santos, codenamed *Gago*, accused of being part of a militia in Duque de Caxias. The mistake was based on the recognition through a photo that was not of him, but of the suspect.[14]

Another significant case was that of construction worker José Domingos Leitão, which occurred in December 2021, in Piauí. He was woken up by police officers at dawn with screams and kicks at the door of his house, after a facial recognition program confused him with the author of a crime that occurred in Brasília, approximately 1,200 kilometers away from where he lives. Both Raoni Barbosa and José Leitão had one thing in common: they were black men, revealing the racial biases contained in facial recognition algorithms, and how they gained other contours in the field of public safety. They have been a tool for reproducing and enhancing oppression that already exists in society because by delegating the task of identifying suspects to algorithms, criminal selectivity receives an appearance of neutrality. And the current conditions of production, storage, and updating of the databases of these systems by the powers that be are not at all transparent.[15]

The second significant example is the AI used in the judicial system with the role of predicting who will be a future criminal and influencing decisions, from bail to convictions. Software like this is built to make autonomous decisions, to predict future actions, stipulating the risk of types of behavior, such as committing a crime. It is clear that programs like COMPAS (Correctional Offender Management Profil-

[13] Studies carried out in the US by the NGO Innocence Project point out that in 75% of the 365 cases in which the NGO operates in New York, the innocence of an unjustly convicted person was proven based on mistaken photographic recognition through DNA tests. In 2019, the National Registry of Exonerations, a database on cases of miscarriage of justice in the US, pointed out that 29% of judicial errors are due to errors in the photographic recognition of people. See COLET NEA - CNJ, 2022.

[14] In the article "Easy recognition errors, one isolated case after another", Hellen Guimarães shows that in all cases of unfair arrests the arrested citizens were black, which alerts us to investigate how algorithms and databases are being produced. In her words: "It was not the first time that recognition by photo resulted in injustice. In Rio, the feeling is that every day there is a different isolated case. Social class may even vary, but the victims are almost always black men". See Guimarães, 2021. See also Lemos, 2021.

[15] Research carried out by Ramon Costa and Bianca Kremer analyzed the ways in which facial recognition technologies affect fundamental rights, in particular, the rights of vulnerable groups in Brazil, such as black, transgender, transvestite and non-binary people. See Costa; Kremer, 2022, p. 150-151.

ing for Alternative Sanctions), which is being used in US forty-six states of the judicial system, were created with the objective of improving the criminal justice system and eliminating human prejudices. The US judicial system today uses the tool to predict who will re-offend in crime, having a role in decisions on parole, bail, and convictions. The COMPAS software uses an algorithm to assess the potential risk of recidivism. It has risk scales for general and violent recidivism and for pre-trial misconduct. Following the COMPAS Practitioner's Guide, the scales were designed using behavioral and psychological constructs of high relevance to criminal recidivism.[16]

But tests such as the one done by ProPublica have found the software is frequently wrong, as well as being biased against black people. The risk scores defined by the program for more than 7,000 people in Broward County, Florida, from 2013 to 2014 were analyzed. Then, journalists verified how many of these defendants were convicted of new crimes in the following two years, with the same reference used by the creators of the algorithm. The comparison showed that the program tends to misreport black defendants as future offenders, placing them in the category of possible repeat offenders nearly twice as often as white defendants. White defendants were also more often rated as less dangerous than black defendants. What this seems to reveal is a situation of discrimination and structural injustice by treating the black population as potentially criminal and the white population as potentially innocent (Larson et al., 2016).[17]

As shown by these two examples and recently demanded in calls for fairness, accountability, and transparency in AI, the use of these new technologies raises urgent issues of justice that we need to collectively face as a society. Above all, issues of racial discrimination and structural injustice (Rafanelli, 2022). And in a country like Brazil, with marked social inequalities, structural racism, as well as institutional prejudices against the most vulnerable, this urgency seems even greater. As well pointed out by Brito and Fernandes:

> *Even if it is not explicitly written in their code, the unconscious bias of the programmer will reflect discrimination of a social order, gender, or skin color and reverberate in the result found, making it apparently "scientific" (Brito; Fernandes, 2020, p. 90).*

The decisions resulting from algorithms can perpetuate already existing situations of discrimination, social injustice, and even gender injustice. It is important to emphasize that it is not the case that we are against the use of programs that make use of AI. They can potentially bring countless benefits to citizens, including helping the socially vulnerable. But we must be certain that this will not in fact generate more injustices or even perpetuate existing injustices.

4 AI and human supervision

Artificial intelligence is a field in computer sciences dedicated to the development of information systems taking the form of machines and robots, reproducing human-like behavior while carrying out certain

[16] COMPAS - Correctional Offender Management Profiling for Alternative Sanctions is a case management and decision support tool developed and owned by Northpointe (now Equivant) used by multiple U.S. courts to assess the likelihood of a defendant becoming a repeat offender. It is a program that uses AI algorithms to calculate the risk of criminals reoffending, suggesting parole or remaining in the prison system.

[17] ProPublica's research looked at more than 10,000 criminal defendants in Broward County, Florida, and compared their predicted recidivism rates to the rate that actually occurred over a two-year period. They then noted that when most defendants are booked into prison, they respond to a COMPAS questionnaire. Their responses are entered into the software to generate various scores, including predictions for "Risk of Reoffending" and "Risk of Violent Reoffending." COMPAS asks several questions that assess how likely you are to commit a crime again in the future, and this assessment is based on a point system, from one to ten, explains Julia Angwin, from ProPublica, an independent American organization dedicated to journalism. investigative. She says that the questions seek to know, for example, "if someone in the family has been arrested, if the person lives in an area with a high crime rate, if they have friends who are part of gangs, as well as their professional and educational history". See Larson et al., 2016.

tasks and making decisions. Among technology experts, the point that has attracted the most attention is the development of "machine learning" systems based on the training of algorithms. The training allows computers to learn skills, identify patterns, and make decisions without being explicitly programmed to do so (Belda, 2017; Russell, 2021 and 2022; Moravec, 1988; Searle, 2014). These systems constitute a powerful practical tool in everyday applications (Zekos, 2021), in natural language processing, in financial guidance, in pattern recognition, in industrial automation, among other uses that still leave much to be explored.

Some of these systems are designed to perform specific and limited tasks, while others would have the ability to understand, learn, and reason more broadly. Both are trained by extracting the relevant information from large data sets, but the degree of autonomy granted to them tends to vary. Only through the so-called "strong" AI could one design systems able to recognize patterns and, through statistical and predictive analysis, learn and develop new skills, making decisions even in unpredictable ways.

The reality of our time is the development of machines specialized in singular tasks, or specific tasks, which characterizes "weak" AI. Those systems have reduced autonomy in that they are limited to only carrying out specific tasks. Examples are the technology already present in people's daily lives, such as chatbots, product recommendation systems on internet pages, streaming services recommending movies and music, and security systems based on speech and vision recognition.

One can say that weak AI is made up of intelligent machines, but only intelligent enough to perform specific tasks. In any case, this type of technology already presents practical dilemmas and impacts on the legal sphere. In effect, the exponential advancement of technology makes it possible to increasingly focus on autonomous decision systems, pointing to the machines developing intelligence comparable to or greater than that of human beings. It has been said that when it occurs, in the near not-so-distant future, we will be faced with what several authors call "singularity" (Kurzweil, 2018), i.e., the singular moment from which the developed technology would reach a no-turning-back point for humanity, causing radical and unpredictable changes in the world. The technological singularity would mark the advent of a stage in which machines and robots would be capable of improving their own intelligence exponentially, surpassing the human capacity to reason and act.

The matter-of-fact reality is that the development of weak AI, however recent on the market, has already presented ethical challenges (Dubber, 2020; Hoven, 2015; Kumar, 2023; Liao, 2020; Müller, 2021; Verbeek, 2011). These challenges (which are also legal) will be enhanced if broad and general AI or even superintelligence (Tzimas, 2021) comes to be. The impact of new technologies brings with it possible implications in terms of human rights (Donahoe, 2019; Leslie, 2021; Livingston e Risse, 2019; Roumate, 2021; Vayena, 2016), and this is where the importance of identifying the exact circumstances where AI violate rights to, therefore, avoid or mitigate the damage that may be caused by its the misuse.

With continuous advances in algorithm development and data analysis, digital platforms have become increasingly adept at techniques that influence our decisions and actions in subtle and even imperceptible ways. AI systems can be employed to create detailed profiles of individuals and thus influence their decisions and behaviors. Social networks, for example, use sophisticated algorithms to personalize content, displaying information, news, and advertisements based on personal preferences and past behaviors. Manipulation can be used for commercial, political, or social purposes, aiming to direct marketing advertisements and digital content in a personalized way, often without people's knowledge or consent. In marketing and advertising strategies, highly persuasive and targeted advertisements are created based on predictions about needs and desires revealed in individual purchasing habits, influencing consumption choices. This is a type of manipulation that reduces the user's interest to access only ideas and opinions already similar to their own, closing them in an information bubble. The information bubble is limiting to their worldview and, within the scope of politics, it contributes to social polarization, perhaps even to the fragmentation of society. The spread of fake news and conspiracy theories can be amplified by algorithms, it can cause splits and ruptures in electoral decisions by way of distorting public perceptions on important issues to political and institutional dialogues.

Indeed, it is important to highlight that informational discrimination depends on other human rights violations, especially those that arise from the invasion of people's private lives, honor, image, and intimacy. Systems that handle large data sets can result in privacy violations (Beduschi, 2019; Goold, 2019; Nissenbaum, 2010), since sensitive personal information can be collected, stored, and shared without the informed consent of the subject generating the data. Adding insult to injury, by making information that should remain confidential public, an invasion of privacy can also occur. The implementation of surveillance systems, such as facial and voice recognition tools, can invade and violate people's privacy, collecting and monitoring personal data without people's knowledge. One can, for example, collect intimate moments between lovers, nude images, the sharing of secrets, and spiritual and religious communions, one can even capture the way a person smiles, frowns, or reveals a joke.

One cannot help but notice that invasion of privacy already involves a violation of human rights, regardless of the use made of the data invaded. Damage is already being caused to the human person and this damage ends up getting worse when the information collected is used in selective discrimination or to manipulate behavior by the guidance in making everyday decisions that the technology provides. The proliferation of false and realistic content, classified as deepfakes, reveals what is to come with the use of AI and how this technology can be used for disinformation with the sole purpose of damaging people's reputations. Deepfakes use real information and data from other people (such as a person's video image, the tone of their voice, and their facial features) to produce new false information (with the image and tone of voice of a person speaking about something she never said). The creation of fake videos invades the private lives and intimacy of real people and can cause serious consequences for human beings by manipulating behavior and directing decision-making.

The misuse of AI violates freedoms (Ferguson, 2017) and generates injustices and inequalities, it causes biases in automated and autonomous decision-making processes. Algorithmic bias refers to the biased steering of information. The steering is produced in AI systems from algorithms that are designed to make decisions based on historical data, past experiences, and identified patterns. As databases are fed by information that reproduces the reasoning and behavior of human beings, the algorithms of these systems can perpetuate and even amplify such biases, manipulating the autonomy of human beings in making a free decision, as we have already warned, highlighting the creation of profiles based on characteristics that stigmatize people with labels that only reflect prejudice and discrimination based on race, gender, ethnicity, etc. These profiles are being used to make important decisions in anyone's life, decisions that are automated in different sectors of society, such as in job selection processes or access to health, credit, or insurance services. Not without reason, AI has been associated with increasing cases of algorithmic discrimination caused by automated data systems perpetuating and even magnifying biases and prejudices against minority groups in society, resulting in unfair and unequal treatment.

To make matters worse, the opacity of AI systems and their lack of transparency makes the misuse of the technology more harmful, already displaying that it is not clear how the systems are programmed to make decisions or arrive at the conclusions they arrive at. Many artificial intelligence models are designed using complex algorithms and machine learning techniques, which can make it difficult for humans to understand exactly how the AI arrived at such an answer or result. When an AI system makes important decisions without explaining the reasoning behind them, it can raise questions about fairness (Binns, 2023; Rafanelli, 2022), liability, and potential discriminatory bias. Opacity can also make it difficult to detect potential flaws or errors in AI systems, which can have serious consequences in a myriad of contexts. This is an important challenge to be solved, so the search for approaches to make AI systems more transparent and explainable (Winikoff, 2021) is crucial to ensure that this technology is used in an ethical, fair, and reliable way, respecting human rights and promoting the benefit of society as a whole.

The main fear of implementing "strong" AI systems is the loss of human control in decision-making. Not only because the flow of information could generate distortions in the data collection and the discrimination of individuals, but because we are facing a unique moment for humanity. For the first time in

history, the making of important decisions in human life can be completely controlled by machines and algorithms. In this line of thought, Müller warns of the difficulty that citizens have in protecting digital property when accessing private data and personal identification, pointing to the loss of control and ownership of data. Says Müller:

> *(…) but it appears that present-day citizens have lost the degree of autonomy needed to escape while fully continuing with their life and work. We have lost ownership of our data, if "ownership" is the right relation here. Arguably, we have lost control of our data (Müller, 2021).*

This fear regarding the loss of control with new technologies puts the human person at the center of inquiry and concern, as has been highlighted by the European Union Parliament.[18] when referring to the centrality of the human person. Recently, Article 1 of the Amendment of the Proposal for the Artificial Intelligence Regulation, by the European Parliament, determined the "adoption of human-centered and reliable artificial intelligence and ensuring a high level of protection of health, safety, fundamental rights, democracy, the rule of law and the environment against the harmful effects of artificial intelligence systems in the Union"[19]. Brazil, as everything so far indicates, pretends to follow the same path.[20]

The European Parliament is also studying regulating this technology based on the diagnosis of the risks that its use poses to users. The initial discussion distinguishes between unacceptable risk, high risk, and limited risk. To reduce risks and guarantee the protection of human rights, the technology must always be subject to human control. Systems that put humans and the environment at very high risk must be prohibited, such as, for example, the production of voice-activated robots that are used to interact and play with children. These robot toys cannot encourage aggressive behavior or behavior that could put them at risk of death, such as in cases of suicidal ideation.

In this way, we can verify that decision-making must always be linked to the control of human will. Hence the expression "human-in-the-loop" (Gordon, 2023, p. 52). This approach seeks to balance AI automation with human supervision and intervention. It means that in certain critical decisions or situations, a qualified human must be part of the decision-making process, reviewing, monitoring, or correcting the AI's actions. The presence of the "human-in-the-loop" is crucial for decision-making, which cannot be completely subject to the command of machines. Only human control at various stages of decision-making, as occurs in legal processes, can identify and correct possible biases, ensure that AI is operating within ethical and legal limits, and ensure transparency in decisions. This approach also provides greater confidence to users and society in general, knowing that there is human supervision in the use of technology. As a result, automated judicial decisions can only be made by AI when there

[18] EU law on AI: first regulation of artificial intelligence. Current affairs: European Parliament, 2023. Available at: <https://www.europarl.europa.eu/news/pt/headlines/society/20230601STO93804/lei-da-ue-sobre-ia-primeira-regulamentacao-de-inteligencia-artificial>

[19] EUROPEAN PARLIAMENT. Proposal for Artificial Intelligence Regulation. Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_PT.html. Accessed on: June 25, 2023.

[20] In terms of domestic legislation, Brazil has the *Marco Civil da Internet* (Law No. 12,965, of April 23, 2014) and the General Data Protection Law (Law No. 13,709, of August 14, 2018). The Civil Rights Framework for the Internet establishes guidelines to protect human rights and guarantee the full development of the personality when using digital media. Among the topics it aims to protect are network neutrality; privacy and intimacy; the responsibility of service providers (they cannot be held responsible for content generated by third parties, unless they fail to comply with a specific court order to remove the content); freedom of expression, as long as it does not violate the rights of third parties or incite illegal practices; and the in secrecy storage of data for a limited period, unless there is a court order. In turn, the General Personal Data Protection Law – LGPD regulates the processing of personal data, granting its holders a series of rights, including access to personal data, the correction of incorrect information, the deletion of unnecessary data, the portability of data to other services and obtaining clear and transparent information about data processing. As an important point, the LGPD always requires the consent of the data subject, so that data should only be used for a specific purpose. Otherwise, the user's authorization and consent will always be required. In addition, the Data Protection Officer (DPO) was created, which involves the obligation of companies to designate a professional responsible for acting as a point of contact between the controller, data subjects and the National Data Protection Authority. Data (ANPD).

is the visible handprint, so to speak, of a person who is responsible for the decision, including being responsible for the production of damages and violations of rights.

5 Final considerations

Therefore, it is urgent to reflect on solutions to face this problem. We can think of two important dimensions, one ethical and the other political. In the ethical dimension, we can strive as individuals and as a society to develop the virtue of justice, understood as a moral-political sensitivity to recognize discrimination based on race, sex, and class and seek to combat them, identifying, above all, the structural racism (one of the causes of algorithmic injustice) and seek efficient ways to eradicate it. But what exactly is this virtue of justice?

Justice is acknowledged as one of the most important moral virtues, with the specificity of being a public virtue. Like every virtue, it is a permanent character trait that is constitutive of happiness or flourishing, that is, of a successful life. Being a moral virtue, it can be taken as a multiple character trait, connecting the agents' diverse emotions, choices, values, desires, perceptions, attitudes, interests, expectations, also, their sensitivity. For example, kindness implies appropriate attention to other people's feelings, requiring a sensitivity to facts about others' feelings as reasons for acting a certain way and a sensitivity to facts about rightness as reasons for acting a certain way. With this in mind, we can understand justice as a character trait, that is, as a propensity to act in a certain way, namely, to act fairly, for certain reasons, that is, for the pursuit of justice.[21]

As already identified by the philosophical tradition, the virtue of justice has an intrinsic relationship with others. The virtue of justice is a character trait, established by habit, to give people what is due to them, whether in terms of goods to be distributed or in the form of punishment for an illegal act. In this sense, justice is the perfect moral virtue in relation to others, which clearly reveals the public character of this virtue, apparently encompassing all other moral virtues, such as generosity, benevolence, clemency, and equity, among others. Justice, therefore, is both a moral quality of the individual and a virtue of citizenship, since it is a central and unifying virtue of individual and political existence, enabling both personal and collective happiness. It is the agent's ability to recognize the relevant contours of the case to give what is due to others from a willingness to achieve justice and it is also a public virtue to ensure social stability correctly.

Now, how does one acquire this important virtue? We propose that an efficient strategy would be to, first, acknowledge the problem, which could be possible by engaging in educational practices, such as classes, events, and courses that have this concern for the well-being of other people in mind. The educational practices could reflect on how new technologies can negatively affect the most vulnerable citizens. From this, incentives could be created to establish practices to combat this injustice. We think that these educational practices are not utopic, as both in schools and universities, events concerned with expanding students' reflection on topics that deal with equal rights and combating racism and sexism are already taking place. The difference, then, boils down to making these institutional educational practices more constant and incorporated into educational training, to the point of helping one's development of the virtue of justice, considering that these practices would aim to develop and foster a more reflective and sensitive personal character in agents on the topic of discrimination suffered by the most vulnerable citizens (with special attention to the discriminatory impacts of the use of AI).

[21] McDowell, for example, correctly defines virtue as a propensity to act in certain ways for certain reasons, consisting of a perceptual capacity to identify the relevant circumstances of the case, having an appetitive component in a presumed sensitivity. Thus, virtue would be an ability to recognize the demands in which the situation imposes a certain type of behavior, requiring complex sensitivity. See McDowell, 1997, p. 141-147.

Furthermore, we believe that in the political and legal dimension, we must create measures to monitor the uses of AI, establishing legislation that has as its central focus the respect for human rights and the protection of the most vulnerable, as well as requiring algorithmic transparency, so that the society can know how the algorithm is produced.

Among the main impacts of possible violations is the need to protect privacy and personal data. Within a massive collection of information, there is a risk of violations of people's privacy and confidentiality, as well as the misuse of data for discriminatory or harmful purposes. Furthermore, AI can perpetuate and amplify biases and discrimination that already exist in society, resulting in automated decisions that can be unfair and harmful to certain social groups. Not to mention the impact of this technology on the job market and workers' rights, driven by increasingly automated systems, could lead to mass unemployment and economic inequality. Therefore, the need for regulation seems indisputable, leaving a more in-depth discussion on ethical and legal responsibility in decision-making made by machines and robots. These questions lead to the debate on the justification of the rights and duties constituting the legal framework for the regulation of new technologies and artificial intelligence. Therefore, it is essential to discuss and debate the items of a specific regulation updating the right to privacy and intimacy, as well as the right not to be discriminated against, in the context of new technologies, especially those referring to systems programmed by machine learning and the use of algorithms.

But on what basis should this regulation be made? Answering the question about the best path to follow in possible regulation is what several countries are beginning to do with the aim of determining the regulatory frameworks to regulate AI. In any case, we can highlight, as has already been said, the centrality of the human person in decision-making. We can leave a series of important decisions in human life in the hands of AI, but only if these decisions are monitored and validated by human beings.

Artificial Intelligence systems must respect the self-determination of the data subject through the expression of informed consent. These systems must be designed and trained in a way that eliminates or mitigates biases and discrimination, to ensure a fair and equitable application of the technology. Protecting human rights requires AI systems to be transparent in their operations and decisions. It is important that users and society, in general, can understand how AI makes its decisions (Müller, 2021), enabling accountability for any errors or inappropriate behaviors (O´Neil, 2020). Transparency and accountability (understood as the systems' ability to account for their own functioning) will allow developers and operators of AI systems to be held accountable for possible damages caused by their technologies (Završnik, 2023; Navas Navarro, 2022). It will be possible to inquire and address the legal personality of intelligent machines and robots, with the aim of assigning them rights and duties, just as happens with a legal entity, such as companies. From this notion of responsibility, it becomes clear that there is a right on the part of users to have a reasonable explanation about the way algorithms make decisions.

The impact of this technology on possible human rights violations reveals the need to develop and implement appropriate regulations and public policies to ensure that AI is used ethically, transparently, and in compliance with human rights. Regulation must safeguard digital property, protecting the privacy and intimacy of individuals' data and information. Regulation is important for identifying possible biases and, consequently, developing regulatory policies that encourage technological advancement while promoting equality, transparency, and responsibility. This entails creating assessment mechanisms and regulations to mitigate algorithmic discrimination, ensuring that AI is developed and implemented fairly and per the principles of equality and non-discrimination enshrined in human rights charters. Regulating AI involves defining quality and security standards, ensuring the privacy of the data used, preventing unfair discrimination, and establishing clear responsibilities in the event of damage or abuse. Regulation is essential to mitigate risks and ensure that AI is a beneficial tool for society as a whole.

References

ALVES, A. T. 2019. Flagrado por câmera vestido de mulher no carnaval da BA matou homem após vítima passar perto dele de moto em alta velocidade. *G1*, 07/03/2023. Disponível em: https://g1.globo.com/ba/bahia/carnaval/2019/noticia/2019/03/07/flagrado-por-camera-vestido-de-mulher-no-carnaval-na-ba-matou-homem-apos-vitima-passar-perto-dele-de-moto-em-alta-velocidade.ghtml. Acesso em 07/15/2023.

BAER, T. 2019. *Understand, Manage, and Prevent Algorithmic Bias: A Guide for Business Users and Data Scientists*. New York: Apress.

BBC. 2015. *Google apologises for Photo app's racist blunder*. 1 July 2015. https://www.bbc.com/ news/technology-33347866. Accessed on 12/072023.

BEDUSCHI, A. 2019. Digital identity: Contemporary challenges for data protection, privacy and non-discrimination rights. *Big Data & Society*, **6**(2): 1-6. Available at: <https://doi.org/10.1177/2053951719855091>. Accessed on 16 set. 2022.

BEITZ, C. 2009. *The idea of human rights*. Oxford/New York: Oxford University Press. 236p.

BELDA, I. 2017. *Inteligencia artificial*: de los circuitos a las máquinas pensantes. Barcelona: RBA Libros.

BINNS, R. 2018. Fairness in machine learning: lessons from political philosophy. *Proceedings of Machine Learning Research*, **8**(1): 1–11. Available at: <http://proceedings.mlr.press/v81/binns18a/binns18a.pdf>. Accessed on 12 jul. 2023.

BOBBIO, N. 2004. *A era dos direitos*. Rio de Janeiro: Elsevier.

BORGESIUS, F. Z. 2018. *Discrimination, artificial intelligence, and algorithmic decision-making*. Estrasburgo: Conselho da Europa.

BRITO, T. S.; FERNANDES, R. S. 2020. Inteligência Artificial e a Crise do Poder Judiciário Brasileiro: Linhas Introdutórias sobre a Experiência Norte-americana, Brasileira e sua aplicação no Direito Brasileiro. *Revista Acadêmica da Faculdade de Direito do Recife*, **91**(2): 84-107.

BUOLAMWINI, J.; GEBRU, T. 2018. Gender shades: intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, **81**: 1–15.

CAMPBELL, T.; BOURNE, K. 2018. *Political and Legal Approaches to Human Rights*. New York: Routledge.

CAMPBELL, T.; EWING, K.; TOMKINS, A. 2001. *Sceptical Essays on Human Rights*. New York: Oxford University Press. 423p.

CLAYTON, R.; TOMLINSON, H. (eds). 2009. *The law of human rights*. 2. ed. New York/Oxford: Oxford University Press, v. 1. 2193p.

CONSELHO NACIONAL DE JUSTIÇA – CNJ. 2022. Coletânea Reflexões sobre o Reconhecimento de Pessoas: caminhos para o aprimoramento do sistema de justiça criminal. Brasília: CNJ.

COPELAND, J. 2001. *Artificial Intelligence: A Philosophical Introduction*. Oxford: Blackwell.

COSTA, R.; KREMER, B. 2022. *Direitos Fundamentais & Justiça*, **16**(1): 145-167.

CRANSTON, M. 1973. *What are human rights?* London: Bodley Head. 170 p.

DASTIN, J. 2018. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 9 October. Available at: https:// www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against- women-idUSKCN1MK08G. Accessed on 14/072023.

DONAHOE, E.; METZGER, M. M.D. 2019. Artificial intelligence and human rights. *Journal of Democracy*, **30**(2): 115–126. Available at: <https://doi.org/10.1353/jod.2019.0029>. Accessed on 12 jul. 2023.

DONNELLY, J. 2012. *International Human Rights.* Philadelphia: Westview Press.

DONNELLY, J. 2013. *Universal Human Rights in Theory and Practice*. Ithaca, NY and London: Cornell University Press.

DUBBER, M. D.; PASQUALE, F.; DAS, S. 2020. *The Oxford handbook of ethics of AI.* Oxford: Oxford University Press.

EUROPEAN COMMISSION. 2021. *Ethics Guideline for Trustworthy AI.* Available at: <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1.html>. Accessed on 29 set. 2022.

FERGUSON, A. G. 2017. *The rise of big data policing:* surveillance, race, and the future of law enforcement. New York: New York University Press.

FERRRO, A.; SERRA, X.; BAUER, C. 2021. Break the Loop: Gender Imbalance in Music Recommenders, *CHIIR '21*, March, 14-19: 249-254.

FRANKENFIELD, J. 2022. Artificial Intelligence: What It Is and How It is Used. *Ivestopedia*. Available at: https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp. Accessed on 12/07/2023.

GEWIRTH, A. 1982. *Human Rights:* Essays on Justification and Applications. Chicago: University of Chicago Press.

GOOLD, B. 2019. More than Privacy: Thinking Ethically about Public Area Surveillance. In: LEVER, A., POAMA, A. *The Routledge Handbook of Ethics and Public Policy.* London: Routledge, p. 102-114.

GORDON, J-S. 2023. *The impact of artificial intelligence on human rights legislation*: a plea for an AI Convention. Cham: Palgrave Macmillan.

GRIFFIN, J. 2008. *On human rights*. Oxford/New York: Oxford University Press. 340p.

GUIMARÃES, H. 2021. Nos erros de reconhecimento facial, um "caso isolado" atrás do outro. Revista *Piauí*, 24 set 2021.

HOVEN, J.; VERMAAS, P.; POEL, I. (eds.). 2015. *Handbook of Ethics, Values, and Technological Design:* sources, theory, values and application domains, Dordrecht: Springer.

IGNATIEFF, M. 2001. *Human rights as politics and idolatry.* Princeton/Oxford: Princeton University Press. 187p.

KUMAR, S.; CHOUDHURY, S. 2023. Normative ethics, human rights, and artificial intelligence. *AI and Ethics,* (3): 441–450. Available at: <https://doi.org/10.1007/s43681-022-00170-8>. Accessed on 30 jun. 2023.

KURZWEIL, R. 2018. *A singularidade está próxima:* quando os humanos transcendem a biologia. São Paulo: Itaú Cultural - Iluminuras.

LARSON, J.; MATTU, S.; ANGWIN, J. 2016. How We Analyzed the COMPAS Recidivism Algorithm. *ProPublica*.

LEI DA UE SOBRE IA: primeira regulamentação de inteligência artificial. 2023. *Atualidade: Parlamento Europeu*. Disponível em: <https://www.europarl.europa.eu/news/pt/headlines/society/20230601STO93804/lei-da-ue-sobre-ia-primeira-regulamentacao-de-inteligencia-artificial>. Acesso em: 16 jul. 2023.

LEMOS, Marcela. Polícia admite erro e cientista de dados da IBM preso por 22 dias é solto. 2021. *UOL Notícias*, Rio de Janeiro, 9 set 2021.

LESLIE, D.; BURR, C.; AITKEN, M.; COWLS, J.; KATELL, M.; BRIGGS, M. 2021. *Artificial intelligence, human rights, democracy, and the rule of law:* a primer. Council of Europe, The Alan Turing Institute. Available at: <https://edoc.coe.int/en/artificial-intelligence/10206-artificial-intelligence-human-rights-democracy-and-the-rule-of-law-a-primer.html>. Accessed on 12 jul. 2023.

LIAO, S. M. (Ed.). 2020. *Ethics of Artificial Intelligence.* New York: Oxford University Press.

LIVINGSTON, S.; RISSE, M. 2019. The future impact of artificial intelligence on humans and human rights.

Ethics & International Affairs, **33**(2): 141–158, 2019. Available at: <https://www.cambridge.org/core/journals/ethics-and-international-affairs/article/abs/future-impact-of-artificial-intelligence-on-humans-and-human rights/2016EDC9A61F68615EBF9AFA8DE91BF8>. Accessed on 17 set. 2022.

McDOWELL, J. 1979. Virtue and Reason. *The Monist*, **62**: 331-350. Rep. In: CRISP, R.; SLOTE, M. (Eds.). 1997. *Virtue Ethics*. New York: Oxford University Press, p. 141-162.

MORAVEC, H. P. 1988. *Mind children: the future of robot and human intelligence*. Cambridge, Massachussetts: Harvard University Press.

MÜLLER, V. C. 2021. Ethics of Artificial Intelligence and Robotics, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), summer Edition. Available at: <https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/>. Accessed on 30 may 2023.

NAVAS NAVARRO, S. 2022. *Daños ocasionados por sistemas de inteligencia artificial*: especial atención a sua futura regulacións. Granada: Editorial Comares.

NICKEL, J. W. 2007. *Making sense of human rights.* 2. ed. Malden: Blackwell.

NINO, C. S. 1989. Ética y *derechos humanos:* un ensayo de fundamentación. Barcelona: Ariel. 494p.

NISSENBAUM, H. 2010. *Privacy in Context*: Technology, Policy, and the Integrity of Social Life. Stanford: Stanford University Press.

O'NEIL, C. 2020. *Algoritmo de destruição em massa***.** Tradução de Rafael Abraham. Santo André: Editora Rua do Sabão.

PARLAMENTO EUROPEU. 2023. Proposta de Regulamento de Inteligência Artificial. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_PT.html. Acesso em: 25 jun 2023.

PERELMANN, C. 2005. Ética e Direito. São Paulo: Martins Fontes.

POSNER, R. 1979. Privacy, Secrecy, and Reputation. *Buffalo Law Review,* **28**(1): 1-55.

POSNER, R. 1978. The Right to Privacy. *Georgia Law Review*, **12**(3): 393-422.

RAAB, C. D. 2020. Information privacy, impact assessment, and the place of ethics. *Computer Law & Security Review*, **1**(37): 1–16. Available at: <https://ssrn.com/abstract=3549927>. Accessed on 30 jun. 2023.

RAFANELLI, L. 2022. Justice, injustice, and artificial intelligence: lessons from political theory and philosophy. *Big Data & Society*, **9**(1): 1-5.

RAZ, J. 2010. Human Rights in the Emerging World Order. *Transnational Legal Theory*, **1**: 31–47. Available at: <https://scholarship.law.columbia.edu/faculty_scholarship/1607/>. Accessed on 15 mar. 2022.

RAZ, J. 2011. Human Rights whitout Foundations. In.: BESSON, S.; TASIOULAS, J. (eds). *The Philosophy of International Law.* Oxford: Oxford University Press.

RISSE, M. 2019. Human rights and artificial intelligence: an urgently needed agenda. *Human Rights Quarterly*, **41**(1): 1-16. Available at: <https://doi.org/10.1353/hrq.2019.0000>. Accessed on 30 jun. 2023.

ROESSLER, B; MOKROSINSKA, D. (eds.). 2015. *Social Dimensions of Privacy*: Interdisciplinary Perspectives. Cambridge: Cambridge University Press.

ROESSLER, B. 2017. Privacy as a Human Right. *Proceedings of the Aristotelian Society*, **117**(2): 187–206.

ROESSLER, B. 2004. *The Value of Privacy.* Cambridge: Polity.

ROUMATE, F. 2021. Artificial intelligence, ethics and international human rights. *Law International Review of Information Ethics*, **29**(3): 1-10. Available at: <https://doi.org/10.29173/irie422>. Accessed on 30 jun. 2023.

RUSSELL, S. J. 2022. *Artificial intelligence*: a modern approach. Hoboken: Pearson.

RUSSELL, S. J. 2021. *Inteligência artificial a nosso favor*: como manter o controle sobre a tecnologia. Tradução Berilo Vargas. São Paulo: Companhia das Letras.

SCHOEMAN, F. 1984. Privacy and Intimate Information. In: SCHOEMAN, F. (ed.). *Philosophical Dimensions of Privacy: An Anthology.* Cambridge: Cambridge University Press, p. 403–408.

SEARLE, J. R. 2014. What Your Computer Can't Know. *The New York Review of Books*. October 9, 2014.

SHELBY, T. 2007. Justice, deviance, and the dark ghetto. *Philosophy & Public Affairs*, **35**(2): 126–160.

SHKLAR, J. 1992. *The Faces of Injustice*. New Haven: Yale University Press.

TASIOULAS, J. 2019. First Steps Towards an Ethics of Robots and Artificial Intelligence. *Journal of Practical Ethics*, **7**(1): 49-83. Available at: <http://www.jpe.ox.ac.uk/wp-content/uploads/2019/06/Tasioulas.pdf>. Accessed on 30 ago. 2022.

TASIOULAS, J. 2015. On the Foundation of Human Rights. In: CRUFT, S. R.; LIAO, M. R. M. *Philosophical Foundations of Human Rights*. Oxford: Oxford University Press, p. 45-71.

TASIOULAS, J. 2019. Saving Human Rights from Human Rights Law. *Vanderbilt Law Review,* **52**(5): 1167–1207. Available at: <https://scholarship.law.vanderbilt.edu/vjtl/vol52/iss5/2>. Accessed on 30 ago. 2022.

TASIOULAS, J. 2012. Towards a Philosophy of Human Rights. *Current Legal Problems*, **65**(1): 1–30, jan. Available at: <https://doi.org/10.1093/clp/cus013>. Accessed on 30 aug. 2022.

TZIMAS, T. 2021. *Legal and ethical challenges of artificial intelligence from an international law perspective*. Cham: Springe.

UNITED NATIONS. *The 17 Goals.* Department of Economic and Social Affairs. Sustainable Development. Available at: https://sdgs.un.org/goals. Acessed on 30 aug. 2023.

VAYENA, E.; TASIOULAS, J. 2016. The dynamics of big data and human rights: the case of scientific research. *Philosophical Transactions of the Royal Society:* Mathematical, Physical and Engineering Sciences, **374**(2083): 1–14. Available at: <http://dx.doi.org/10.1098/rsta.2016.0129>. Accessed on 30 jul. 2023.

VERBEEK, P.-P. 2011. *Moralizing technology:* understanding and designing the morality of things. Chicago: The University of Chicago Press.

WELLMAN, C. 2010. *The Moral Dimensions of Human Rights*. Oxford: Oxford University Press.

WINIKOFF, M.; SARDELIČ, J. 2021. Artificial intelligence and the right to explanation as a human right. *IEEE Internet Computing*, **25**(2): 116-120. Available at: <https://ieeexplore.ieee.org/document/9420081>. Accessed on 30 jun. 2023.

YOUNG, I. M. 2011. *Responsibility for Justice*. New York: Oxford University Press.

ZAVRŠNIK, A.; SIMONČIČ, K. (ed.). 2023. *Artificial intelligence, social harms and human rights*. Cham: Palgrave Macmillan. Available at: <https://doi.org/10.1007/978-3-031-19149-7>. Accessed on 30 jun. 2023.

ZEKOS, G. 2021. *Economics and law of artificial intelligence*: finance, economic impacts, risk management and governance. Switzerland: Springer International Publishing AG.

Submetido em 20 de setembro de 2023.

Aceito em 11 de janeiro de 2024.