DOI: 10.5433/2176-6665.2024v29n3e50076en
Received on 15/03/2024;
approved on 10/08/2024;
published on 30/10/2024.

Territories of Knowledge: the Bioethical Debate on Responsible and Decolonial AI

Territórios do Conhecimento: o Debate Bioético da IA Responsável e Decolonial

*Elen Nas[1]

Abstract

The article explores the alignment of bioethics and decolonial perspectives with the premises 'responsible artificial intelligence' (RAI). It proposes a careful examination of the implicit conflicts in terms such as 'decoloniality' and 'territories of knowledge'. The article presents the similarities between the debates on biopolitics, necropolitics, and bioethics, associating them with the possible impacts of artificial intelligence (AI). Given the inevitable reach of AI in all spheres of society, the decolonial perspective explains how monoculture reinforces epistemic ideas with oppressive potential on minorities and groups that have been targeted from colonial practices to the present day. When presenting the principles of RAI, the article highlights the risk of embracing, without the necessary critical thinking, the formal rules imported from the Global North as "THE" solution to mitigate the possible impacts of AI, when educational and communication actions within the culture become necessary, and they will only be possible if RAI becomes Slow AI.

Keywords: artificial intelligence, bioethics, decoloniality, responsible AI.

Resumo

O artigo explora os alinhamentos das perspectivas bioéticas e decoloniais com as premissas da inteligência artificial responsável (IAR). Propõe, desse modo, exames atentos sobre os conflitos implícitos nos termos 'decolonialidade' e 'territórios do conhecimento'. Apresenta as aproximações dos debates acerca da biopolítica e necropolítica com a bioética, associando-os aos possíveis impactos da inteligência artificial (IA). Dado o inevitável alcance da IA em todas as esferas da sociedade, o olhar decolonial explicita o quanto a monocultura reforça ideias epistêmicas com potencial opressor sobre minorias e grupos que têm sido alvos desde as práticas coloniais até os dias atuais. Ao apresentar os princípios da IAR, o artigo ressalta o risco de absorver, sem o necessário pensamento crítico, as regras formais importadas do Norte Global como "A" solução para mitigar os possíveis impactos da IA, quando se fazem necessárias ações educativas e de comunicação dentro da cultura, que só serão possíveis partindo-se do entendimento de que a IAR é uma IA Lenta.

Palavras-Chave: inteligência artificial, bioética, decolonialidade, IA responsável.

[1] University of São Paulo, Institute of Advanced Studies, Oscar Sala Chair (IEA/USP, São Paulo, SP, Brasil). ORCID: https://orcid.org/0000-0002-6275-2799.

1 - Introduction

When we think of territoriality, geographic space comes to mind. However, as philosopher Lewis Gordon (2021, p. 8) points out in "Freedom, Justice and Decolonization" euromodernity extends this idea to identities. In the book he argues that a special form of alienation has reduces entire groups to categories such as Indigenous or native, black, colonized, and enslaved. As a result, people who have had their culture devastated begin to permanently suffer from a kind of melancholy and mourning for the separation of territories where 'home' and belonging are not restricted to geographic space, but also to their identities. The loss represents a composition of elements ranging from the relationship with space and language, to the dignity of existing within relations. Therefore, not only is geographic space of ethnic groups been violated, but also territories of knowledge. This knowledge has forced some of these ethnic groups into obscurity, and, in their place, there is only a non-place, not being accepted, not being part of a world where victimized groups cease to BE and become 'problems'.

As the entire world turns to the potential and limits of artificial intelligence (AI), we have an opportunity to reignite debates of decoloniality. Debates that might otherwise have been restricted to small niches. Because AI represents the accumulated scientific and philosophical knowledge of the Western tradition, its applications revolutionize various fields of society, from the most visible to the invisible.

Bioethics discusses the ethics of technical-scientific advancements, and their impact on humanity, society, and the environment, from health care to new ways of communicating and educating. Discussions in bioethics uncovers the humanist values that emerge in the modern world. Bioethicsconfronts not only in the inconsistencies and contradictions, but also where there could be consensus on what is - or is not - acceptable.

Thus, this text aims to bring bioethical reflections on the challenges of life on earth due to the unbridled race of artificial intelligence (AI) developments applied throughout the world. These AI developments have been applied in a ubiquitous, pervasive manner, with both visible and invisible impacts. There are techno-scientific projects that have a direct threat to life. Bioethics emerges as a bridge between the sciences and humanities facing techno-scientific projects and therefore understands 'humanity' as a non-negotiable good.

Seeking comprehension of such threats from a power relations perspective has led bioethics on an exploratory journey that begins with attention to ethics applied to the life sciences, from classical Western philosophy to the intersectional themes of contemporary human sciences. Therefore, a first step is to understand power relations and their expressions of control – biopowers and biopolitics – that are embedded in Western epistemologies (and how this shows up in knowledge and language). As argued in the work of Michel Foucault (1988, 2001, 2008), new policies of the modern era have generated habits that are inserted into and interfere in life in a more incisive way, based on how knowledge is organized and applied. In other words, Foucault drew our attention to the fact that 'the truth' evoked by institutions, whether of scientific knowledge or political-administrative structures, does not exist outside of power or without power. It is a truth of this world, its power structures that are produced thanks to multiple coercions (Foucault, 1993, p. 12). Its influence on behavior demonstrates a kind of 'training' capable of influencing agency over one's own body. Foucault argues that in the 18th century, power was exercised in a way that was interconnected within the social body, rather than imposed upon it from the outside. (Foucault, 1993, p. 131).

Structural changes in lifestyles have increased the impact of technologies and industrial production methods on our bodies. This permeates everything from the relationship with work and time control, to policing and influence over the imagination. This context has taken oppression to another level, that of *re*pression, in which there is a shift from regimes fundamentally based on punishment to another, in which surveillance stands out as an instrument of authority (Foucault, 1993, p. 130).

Recognizing that control mechanisms are expressions of authority, is crucial to grasping how these control mechanisms are used to obstruct the dissemination of decolonial thought across institutions. Thinking about decoloniality in AI goes beyond denouncing the improper and non-consensual use of data. It goes beyond how algorithmic classifications and orientations of this data perpetuate historical inequalities and prejudices. At the root of these problems there are 'territories of untouched knowledge', in their very forms of validation and understanding. For decolonization to have a real-world impact, it needs to be an inclusive, broad movement rather than something controlled by a few scholars or institutions. The voices of a few theorists related to African philosophy and indigenous knowledge will not be able, alone, to transform the *status quo*. It is necessary to build bridges between the knowledge that governs our entire educational structure at all levels and other knowledge that is part of the cultural heritage of our territory. And the act of building is the act of innovating, and transforming, and requires creative thinking in non-linear methodologies: it requires processes in spiral movements and an acknowledgment of the knowledge that is constructed beyond the boundaries of school and university. And where and how does this knowledge appear or is it invisible? Today, this knowledge appears in computational infrastructures. For this reason, thinking about decoloniality for AI requires attention to technological infrastructures, which are composed of knowledge and social practices:

the data they produce reflect access or exclusion, violence or institutional protection, over-representation or under-representation of people's perspectives and imaginaries. In the same way, the definitions of priorities regarding technological developments are determined based on the interests of those who dominate them.

In addition to the concepts of *biopolitics* presented by Michel Foucault (2008), Achille Mbembe (2016) argues that conflicts are inherent to the power dynamics that dictate which lives matter most and which matter least. The result of such a scale of values for life is what he called *necropolitics*.

Despite universal principles of human rights (UN, 1948), the tradition of Western thought continues to be influenced by an epistemology that is organized in binary distinctions such as the Aristotelian hylomorphic principle and the understanding of *zoe* and *bios*. *Zoe* is biological life in general (bare live), while ***bios*** refers to the particular, qualitative way a life is lived within a particular society or culture. For Western politics, the fundamental dual category is one is either a friend or an enemy. However, a more nuanced understanding is expressed in the tensions between *bios* and *zoe*, which is that *zoe* - the 'bare life' - is a life without political existence, a life excluded from citizenship rights. And *bios*, a life that belongs to the protection of law. (Agamben, 1998). Bios is the life included in the equations with a greater wealth of detail; a life which goes beyond just categories of information that directly feed statistics into computational algorithms.

Therefore, while emerging technologies have a broad reach, being present in the daily lives of more than half of the global population, we can ask ourselves what knowledge and territories they most correspond to, what policies and world ideals will be embedded in them, and who will they tend to harm. 'Bare life' is the material for filling these technologies for better serving 'citizen life'; and this citizenship, is not, even in the 21st century, accessible to all people as a universal right.

Taking some uses of AI in the health field as an example, to think about decoloniality, is an invitation to observe the forms of knowledge that generate data, organize, classify, and produce results. This is what I have called the 'bioethics of non-presence' as a continuous exercise in analyzing facts beyond 'evidence' (Nas, 2021).

Ruha Benjamin (2019), for example, argues that algorithmic bias is also expressed in invisible racism arising from assumptions about health risks and assessment of access to treatment based on the characteristics of the black population, whose socioeconomic situation is generally precarious. This fact that creates contrasts in the Americas is one of the consequences of colonialism. The author points out that, in the health environment, in many cases, it is assumed that black people are stronger and do not need anesthesia; for the same reason, these social subjects receive more rushed care than people with white characteristics. Thus, we can see implicit prejudices about which lives have greater or lesser value within the social imaginary; such practices are interpreted as data that generate statistics and, when encapsulated within an AI, they become oppressive so-called 'truths'. Since oppression recurrently affects the groups that suffered most from colonialism in the Modern Era, we face the risk of AI inaugurating a new colonialism that imposes itself on different layers, from the most objective and superficial level to the deepest and most subjective.

Colonialism impacted the body's relationship with the territory this body lives in. This impacts how one can move in a territorial space, that is, whether or not one has the right to come and go, or whether or not one has sovereignty over one's body. Added to this is the extent to which the ways of seeing and knowing the world are inscribed in the body as ways of behaving, dressing, organizing the community, and educating. Patriarchy and Euromodern colonialism imposed what they understood as superior ways of acting, even claiming authority over the knowledge of what ethics should be, and what should be considered more effective, advanced, and appropriate.

Thus, even though communication through this text will require words that are widely adopted in our territory as part of the legacy of the Western European model in our universities, and that we have benefits in many areas of science and humanities, the decolonial reading aims to draw attention to how patterns of violence, exploitation, subordination and uses of people, bodies, groups, and territories are also present in this knowledge in invisible ways. With AI, these patterns may become even more pulverized and difficult to identify. Therefore, when the world turns to efforts to regulate AI from the perspective of implementing Responsible AI (RAI) as a framework capable of condensing the principles of justice, transparency, explainability, and accountability, we must be careful that the proposed legislation does not become a new social contract enlightened by experts, which in practice only maintains inequalities and imbalances expressed in biased understandings of the law, granting the privileges of access and application of justice to the same hegemonic actors today, under the same perspectives

of colonialist culture[2]. And since information systems operate in closed boxes through AI, we must ask ourselves to what extent they are reinforcing the culture of colonial violence through means that make it even harder to identify. It also reflects how hegemonic knowledge models exercise authority, often disqualifying narratives in 'territories of knowledge' that do not follow linear and border-bound models.

Even in the tradition of Western knowledge, which has men and white people at the top of its hierarchies, there are a large number of authors and works produced in the last 50 years that understand the need to rescue interdisciplinary dialogue. Interdisciplinary dialogue facilitates a better understanding of a culture that is moving towards a 'hybridization' resulting from the human-machine relationship. (Nas, 2020). Overall, challenging the ideals of *formality* in knowledge organization and research faces ongoing obstacles.

According to Gilbert Simondon (1995, p. 49), the concept of 'form' is an example of the influence of political life on theory. Thus, the separation between 'form' and 'matter' reflects the structure of thought formed in classical antiquity in which 'form' comes from ideas of qualified life, of citizens, and 'matter' is what fills the form, and serves purposes. Matter is understood as passivity or the absence of freedom of choice (Nascimento, 2017, p. 135).

When exploring the impact of RAI, we should therefore return to fundamental epistemological questions: what is knowledge, what is it's purpose, and/or how is its relationship with life understood? In doing so, we can situate where 'form' contributes to discoveries, and where it can only reproduce *knowledge as a 'product' of authority*. Fundamentally, whether 'shapeless' knowledge (a knowledge that deviates from the known, current, or hegemonic patterns) is still knowledge, or whether, like a 'defective' product (a brick with a crack or in a proportion different from the rectilinear format), it should be discarded[3].

2 - Responsible AI

The advent of the internet embodies the supraterritorial mode of new industrial and financial policies that ultimately influence government policies in all areas. Globalization inaugurates the shift from physical space to the cloud, governed

2 We can cite many everyday examples from Brazilian political life, such as the case of Rafael Braga, who was arrested as an activist in 2013, in Rio de Janeiro. While white middle-class protestors were released, he was held in custody, even though the lawyer had presented himself for the release of all those involved in the demonstrations. Another example is the statements by far-right politicians minimizing the seriousness of the murder of Rio de Janeiro City Councilwoman Marielle Franco in 2018 because as a black person, and a woman, it was 'natural' that her life mattered less. There are many examples, that are borne out in the Atlas of Violence (IPEA, 2023), which identifies that almost 80% of homicide victims in Brazil are black and brown men, also 70% among women (also black and brown).

3 This debate is significant to me since I paused my academic life (which began in the Social Sciences with award-winning work in Political Science) for over 20 years when I won two music festivals and went abroad. But in times of analog technology and concentration of resources in industry agents, knowledge and talent incapable of becoming a product is insignificant. A technocracy dictates that the 'bad' can be 'good' as long as it is saleable, It means to have the 'right' forms which can be easily accepted. So, the 'good' is 'useless' if it is not adequately formatted by the technologies in vogue and, mainly, if it cannot be a 'product'. This comment does not deviate from what is presented here: in an industrial society, everything must 'work' like a factory, including – and unfortunately – the production of knowledge in the academic sphere.

by the supremacy of calculations and algorithms. The neoliberal policy begins to disregard as much responsibility as possible for the impacts of business on territories and lives (Bauman, 1999)[4].

Thus, although our new technological developments and information arrive through immaterial means and become "models" encompassed in software, applications, and devices, there are still gaps in how to measure their impacts on territories and lives, individuals and collectives, human and non-human.

While the debate on digital colonialism points to the widespread extraction of data held by large technology companies, this debate becomes more intense when we identify where these companies are based. These are new forms of imperialism and colonialism that expand the initial model of exploiting territory and extracting minerals to new exploitations of lives through data extraction and mining. When this data becomes part of an AI that presents itself as good for all humanity, it is necessary to observe how inherent flaws of the system can be amplified through AI and cause harm.

Considering that Brazil does not develop fundamental technologies and the main global artificial intelligence products and tools, we should ask ourselves how society can obtain information to acquire and develop AI technologies capable of promoting radical innovations for the common good. Instead, we tend to make superficial adjustments to existing Western products, services and systems. While this issue is currently on the agenda of several countries, the fact that AI-intensive economies will increasingly distance themselves from other perspectives in a race for future productivity, it will also perpetuate knowledge territories represented mainly by white males of the Global North. The only way to change this scenario would be through greater popular participation, indicating the ways to appropriate the technologies, also deliberating on whether to implement - or not - them according to the sectors, defining priorities and limits.

Institutions like ACM, UNESCO, and OECD have formulated different instrumental principles for responsible artificial intelligence. They propose to minimize specific risks of biases in these technologies, taking values such as equity, autonomy, privacy, proportionality, responsibility, and accountability into consideration (OECD, 2019; UNESCO, 2023)[5]. Seeking to minimize risks through standards and principles is not the same as ensuring full compliance with ethical and justice principles for AI developments from infrastructure to human-algorithmic interactions (HAI), human-computer interactions (HCI), and human-robot interactions (HRI). Or, if we think about the 'principlism' bioethical approach, of respect for autonomy, the principle of beneficence, non-maleficence, and justice, we realize that AI does not provide guarantees regarding the observance of such principles.

4 "To be free from the responsibility of the consequences is the most coveted and cherished gain the new mobility brings to free-floating, locally unbound capital. The costs of coping with the consequences need not be now counted in the calculation of the 'effectiveness' of the investment." (Bauman, 1999, p. 16-17).

5 I thank prof. Virgílio Almeida for his ongoing support and insights on Responsible AI in the article's initial version.

When we discuss responsibility, we implicitly recognize a call to an ethical awareness of not causing harm. This raises the question of how we can ensure that companies and institutions adhere to the principles of transparency, interpretability, explainability, auditability, and accountability in AI.

In the Brazilian context, we can consider to what extent such principles are aligned with institutional policies when the issue is transparency and providing explanations for decisions. While reports and recommendations are produced in the North in alignment with the interests of the richest countries, it may make sense for them to seek to apply the rules to protect their citizens while their companies profit in Brazil, where the observance of justice is flawed and selective, within a policy with colonial influences that are still based on coronelism and clientelism.

Thus, if these principles guide the development of AI applications in various areas, such as health, education, finance, or social media, Responsible AI should – in theory – combine AI governance (data and models) with training the right people to implement it. The definition of 'right people' should be a composition of a diversity of expertise, knowledge fields, sectoral representation of third parties in civil society, and make sure a plurality of views are included. We understand that, for pragmatic reasons, there is a tendency to avoid the inclusion of perspectives that are 'dissonant' from disciplinary fields, however, current challenges require a greater understanding of the ethics of hospitality[6] as a challenge of humanity that is a priority. That means the implementation of AI applications and its regulations cannot be a merely technical debate. As always mentioned, the noblest goals of AI developments are to improve the quality of life for all and amplify the power of human intelligence. In that case, we need a *'Slow AI'*[7]. Because the methods of AI development need reformulation.

The documentation of each stage of the process, its sources of information, and the decisions made when creating the algorithm tend to slow down the race of AI. Fundamentally validating the applications by listening to society, the parties impacted, and the parties that suffer most from discrimination and misinterpretations during the process. This type of review of the methods also requires curricular training for scientists and programmers in the Humanities, as getting to understand ethics in depth, and its new critical contours through bioethics. It is not just about understanding ethics from a theoretical point of view, but rather what perspectives on life these theories bring and what would be the best approach to ensure compliance with the principles of justice with attention to inclusion and diversity.

Over the past two decades, advances in AI research have been attributed to the combination of increased computing power, the availability of large volumes of data, and advances in machine learning algorithms (Mohamed; Png; Isaac, 2020). The decolonial approach to artificial intelligence identifies in contemporary data collection practices a new way of governing and distributing power in societies and economies (Couldry; Mejias, 2023). Whereas historical colonialism was an extractive model that

6 The ethics of hospitality can be found in various authors, such as the post-structuralist thought of Jacques Derrida, and we can also draw a parallel with the African concept of *ubuntu*, in which humans complement each other, communicate, and exercise their power relationally. This can also be paralleled with classical virtue ethics, in which 'doing good' is a way of practicing and refining virtues.

7 The idea of a '*slow AI*' intuitively appeared to me as a new and original element, however, when searching the web, I found a short article on a blog that, in addition to aligning with the references in this text, reinforces the call that information systems, to be ethical, must prioritize people's voices in order to promote a more just and inclusive society (Conroy, 2023).

reorganized societies at multiple levels, the decolonial approach to AI focuses not only on large digital platforms, such as social networks and search engines, but also on broader data collection habits in all aspects of social, economic, and political life. The term "algorithmic coloniality" has been used to expand the notion of data colonialism in the context of algorithmic interactions in societies (Mohamed; Png; Isaac, 2020). Algorithms make decisions, define resources, and shape individual and collective sociocultural and political behavior (Mendonça; Almeida; Filgueiras, 2024). The language of decoloniality brought to the context of AI offers a new reading for fundamental concepts to be sought for algorithmic systems, which are justice, accountability, and transparency in the decisions made by these systems. In the context of Decolonial AI, algorithms fit into a taxonomy of a colonial vision, such as institutionalized algorithmic oppression, algorithmic exploitation, and algorithmic dispossession (Zuboff, 2019).

The part that remains invisible in algorithmic oppression is the knowledge systems understood as neutral, scientific, effective, and accurate. Furthermore, there is a fantasy that elevates AI to a superhuman and supernatural level, as an inevitable force beyond our ability to understand or control. It carries an imagination about a future that can liberate humans from all limitations through science. Thus, Timnit Gebru argues that if the plan is to make AI contemplate the values of a future with equity and humanity, it needs to be brought back to Earth (DAIR, 2021). The fantasies and romanticization of technologies are also part of a system of privileges sustained by extreme inequalities and exploitation of peoples and territories, since the beginning of the Industrial Revolution.

While it is not possible to undo the history of genocide and enslavement of peoples in colonized territories, we can still change the course of future events by understanding that the risks already highlighted in the use of AI show weaknesses that cannot be 'fixed' through some 'patching' in the algorithm. The future that is emerging is that, in the political and social sphere, nothing will change and may even get worse if actions are not taken to slow down the implementation of AI in favor of responsibility.

The Declaration of Principles for Responsible Algorithmic Systems (ACM, 2022) considers that AIs are increasingly being used in all spheres of society with the potential for great impact. From a bioethical perspective, we will talk about impacts on lives and not only the explicit harms but also the invisible ones that transform the views one has about oneself, about relationships, and the ways of seeing and understanding the world. Again, what we have as indicators for regulation are generic proposals such as the need for responsible AI to comply with the principles of legitimacy and competence; harm minimization; security and privacy; transparency; interpretability and explainability; maintainability; contestability and auditability, with recommendations for AI system builders and operators to compare what human decisions would be like in the contexts to which the AI intends to respond; for developers to conduct tests with comprehensive impact assessments; for policymakers to invest in audits to evaluate the applications of IAR in the AI development and implementation processes; and that AI system operators develop awareness of their decisions in the process of developing algorithms (Almeida; Nas, 2024).

In practical terms, these recommendations reach a level of sophistication to the extent that until now all training involving systems architecture has focused fundamentally on technical aspects, following methodologies and logic that resolve doubts and uncertainties with approximate answers, often granting them the *status* of truth. To change these ways of knowing, developing, and carrying out projects, it is necessary to invest in a significant paradigmatic shift. It means questioning how we understand technology and its relationships with humans, nature, and society. Furthermore, it is necessary to present ethical and philosophical knowledge to developers in a way that is not only prescriptive but also critical, that is, through the exercise of reflection, something that large language models in AI fail to provide with precision.

Thus, Responsible AI faces different obstacles, such as the difficulty in carrying out audits on AI systems capable of enforcing the need to be accountable to society (Raji; Chock; Buolamwini, 2023). The first step for companies to protect themselves from external audits is to promote the practice internally by hiring professionals dedicated to ethical debates around Responsible AI. However, such implementations fall short of what is needed and reveal cases of internal conflicts. Such as that of computer scientist Timnit Gebru, who was fired from Google (Hao, 2020) due to co-authoring the research "On the dangers of stochastic parrots: Can language models be too big?" (Bender *et al*., 2021). The work pointed out problems regarding the environmental cost of training language models, and racist and sexist biases, among other dangers such as the generation of misinformation.

In recent years, ethical debates surrounding the development of emerging technologies have often been cited as obstacles to the race for AI and robotics innovation. Therefore, there has been a slow process for companies to adopt new visions that reflect hiring personnel capable of proposing methodologies and reflections to meet the demands for ethical and responsible AI. In the absence of structural work that is consolidated within companies through internal teams qualified to understand the social impacts of AI, a second alternative is to hire consultancies to audit the systems under development. These consultancies only indicate their view on possible ways to mitigate the problems identified, without having authority over any modifications that should be made by the companies themselves. In addition, there is no guarantee that companies will provide all the information due to the protection of confidentiality and privacy, and the information they choose to disclose. In addition, there is a contractual limit on confidentiality for consultancies.

Researchers Raji, Costanza Chock, and Buolamwini (2023) emphasize that a "third party" is needed to audit systems so that a sector independent of companies can act more assertively, highlighting potential problems, and doubts, reporting them, if necessary. Thus, the authors suggest that such auditors should have legal support and protection to share their conclusions transparently. More than that, for representatives of society to participate in audits, a qualification process must be made viable. Therefore, these are educational proposals that require debate, investment, and time. The authors state that the current context does not allow for the effective participation of *third-party auditors*. They provide regulatory proposals to guarantee accreditation, protection, and support for external and independent auditors in favor of equitable policies for fair and responsible AI. The freedom to publicly disclose the

results of an audit must be defended and not restricted, even if the transparency of information is confused with the nature of denunciation when the models investigated are already in use and are potentially causing harm, as in the cases presented by *Pro Publica* - an independent investigative journalism agency -, which revealed biases in AI predictions that, in addition to being erroneous, were biased, as they considered black-skinned individuals to be at greater risk of reoffending in illegal acts than those with light skin (Angwin *et al*., 2022).

In short, for RAI applications to be effective and have a real impact, they depend on structural changes in research (R&D) and its methods, a fact that confronts the territories and frontiers of knowledge. Positively, RAI, if implemented as proposed in this approach, is an opportunity to promote interdisciplinary debates and stimulate changes.

3 – AI Decoloniality

The debate on the decoloniality of AI is recent and is related to the recurring ethical problems with the advancement of AI technologies. For example, when AI reinforces stereotypes of beauty and black women are often identified as 'ugly' (Araújo; Meira; Almeida, 2016) and with pejorative adjectives (Noble, 2018) – as are *Latinas* in the US –, there is a reinforcement of old prejudices that in the colonial context represented ideas used to justify the subordination and exploitation of the groups that suffer the most from the consequences of modern industrial colonialism.

Automated search engines often reproduce prejudices against women (Noble, 2018) and tend to emphasize stereotypes in which the image of beauty is white (Magno *et al*., 2016). This was the case even though the internet search was conducted in Brazil, a country where the majority of the population is black and brown. The decolonial perspective, therefore, seeks to reach the less evident elements that impact the subjectivities between ethical and aesthetic perspectives that are not easily visible within technological developments. Such invisible elements in products and artifacts are often suppressed, such as environmental disasters and damage to life in general (Nas, 2021).

Thus, decoloniality applies to artificial intelligence to the extent that ethical deviations in AI that impact and cause harm to life are related to structural problems of coloniality. The fundamental question in which the concept of decoloniality emerges is: when there are failures in the interpretation, recognition, and recommendation of algorithms, what are the consequences, and for whom? (Nascimento, 2019).

Failures commonly occur because models are trained and designed within the perspectives of developed countries, their values, and ways of seeing and understanding humans and the world. Technocolonialism arises when everything moves towards dependence on technologies that are created and have all the infrastructure management in the 'Global North'. These are technologies that carry the knowledge of an ethnocentric, anthropocentric, and speciesist techno-utopia when they announce that AI can improve the lives of women and children in vulnerable situations around the world (Global Grand Challenges, 2023). In practice, this discourse corroborates the imposition of a Western ideology that grants a subordinate

status to the knowledge and perspectives of other cultures, and, in the case of AI, the coloniality of knowledge is reflected in the *datasets* that are centered on Western culture (Mboa Nkoudou, 2023).

A monoculture appears when AI frequently presents light-skinned individuals as biotypes of "human" or "person", often males. This generates in terms of image production, a monoculture or cultural supremacy that makes diversity invisible, potentially influencing blockages on otherness (Nas, 2023b).If we want to list the fundamental characteristics of AI decoloniality, it refers to the underrepresentation of data related to cultural and epistemic diversity. Thus, the injustices experienced by black and indigenous people in the colonial era and monopolistic resource extraction have become prototypes of future exploitations applied globally through AI algorithms (Miller, 2022).

Currently, when AI tools are used to hire or fire someone, criteria are applied that in practice tend to subjugate people and groups that have historically been part of the oppressed portion of power structures. Thus, the decolonial perspective proposes a view that goes beyond simply seeing AI failures as cases to be analyzed separately, or as representing types of prejudices that are already criticized by the humanist ethics of the modern industrial world. The fissure between theory and practice created by this ethics has exceptions judged by systems of authority in which other ways of knowing and explaining the world are deliberately ignored and seen as obstructions.

The decolonial perspective of AI goes beyond the limits of its applications in the most diverse contexts, such as health, work, education, art, or public safety, among others. In all of these contexts, some biases demonstrate that technology serves some better and many worse, thus reflecting the major problems of concentration of wealth *versus* situations of extreme poverty in configurations of forces that separate those who are closest to and those who are furthest from certain privileges. Decolonial criticism is political and also concerns culture; for this reason, it is not technophobic, since there is no way to eliminate from culture what technologies represent in life today. Thus, it does not oppose techno-scientific developments but proposes a reflection on what is wanted, who it serves, what the impacts are, and who they affect. And, based on such awareness, discover possible paths to coexistence, based on conceptual restructuring and creative ways of thinking about emerging technologies.

Research challenges for Decolonial AI include the need for decentralized infrastructures that collect data through ethnographic and qualitative research as the basis for new local developments of AI technologies, aiming to organize data responsibly so that networked environments represent diversity and meet ethical, social, and environmental aspirations, that value affirmative action and seek alignment with sustainable development objectives, combating racism, social injustices, and environmental problems (DecolonizAI, [2022]). Equally important is establishing collaborations with researchers from the Global South and the African and Latin American diaspora in the North.

Mboa Nkoudou (2023), from the Montreal Centre of International Expertise for Artificial Intelligence (CEIMIA), proposes: prioritizing local needs by encouraging the participation of communities in identifying the problems they wish to solve; investing in the collection and development of databases that represent the diverse cultures of the territory; that cultural and academic institutions and AI developments should consider the knowledge of indigenous peoples; building databases with the

proactive participation of marginalized groups; that educators and media should resist dominant narratives, engaging in critical perspectives and offering technological literacy programs that question technological narratives and align them with local experiences and cultures; that educational and government institutions should invest in local talent for AI developments; that Ethics Committees and AI organizations should adopt ethical practices that include communities to ensure that such developments reflect local social values and principles; encourage international collaborations with equity in decision-making; that the Legislative and Judiciary branches must guarantee digital sovereignty with policies and laws appropriate to the context in order to ensure that data remains under the control of the community and for its benefit; that the global AI community must celebrate local achievements through the media in order to highlight advances in representation in the face of dominant narratives.

The recommendations cover and concisely consider the interrelationship between scientific developments and attention to the culture of peoples and territories, concerning knowledge distinct from hegemonic practices. Above all, they are an invitation to understand that not only are the knowledge and ways of life more distant from the West, but that a large part of the population of colonized lands also has in their DNA the marks of ethnic groups that were extinguished by genocide, kidnapping, and enslavement, resulting in melancholy as mourning for separation and the loss of a world replaced by another to which they do not belong with dignity (Nas, 2023a). The inability of DNA tests to reveal ancestries that were erased due to a lack of data also illustrates how the scientific method embedded in AI creates false expectations, over-representing groups in which there is a greater quantity of data and under-representing others for which there is insufficient information, thus contributing to the invisibility of ethnic groups that disappeared due to genocide and miscegenation policies. Lewis Gordon argues that the concept of race is fabricated and recalls that Antenor Firmin questioned the rigor of the claims of Euromodern human sciences, because "rigorous science adapts to the demands of its object. It does not try to force reality to fit its presuppositions" (Gordon, 2023, p. 107). Here we have the basic epistemological problem of how an algorithm is defined: everything starts from a formula (form) that will determine a sequence of actions and results.

4 – Discussion

AI systems can have transformative and long-term effects on individuals and society. To manage these impacts responsibly and direct the development of AI systems towards the public benefit, the principles of Responsible AI establish guidelines on how to design, develop, implement, and audit services based on artificial intelligence technologies. The proposal of a 'Decolonial AI' (outlined above) is to pay attention to the risks involved in an oppressive AI that reinforces monoculture and epistemicides. When Cathy O'Neal reveals how the "algorithms of mass destruction" operate (O'Neil, 2017, 2021), we can imagine the metaphor of the Trojan Horse as an AI that presents itself with the best intentions – in the imposing and 'magical' appearance of the technology that makes it a 'gift from the gods' –while there are invisible elements capable of causing irreparable damage.

The possibility of harm has brought the ethical debate to technologies, so that, when searching for the terms "AI Ethics" AND "ethics in computing" on the Dimensions platform ([2024]), the results on 01/03/2024 were modest, with 27 articles published in the last four years. Considering that this is a search directory that accesses the databases of journals and universities (mostly) in English and from the Global North, we understand that the ethical debate around emerging technologies has only been intensifying in recent years and, while new AI developments have been drawing attention to the possible risks and how to mitigate them, the production on this subject appears in publications that are not accessed by research directories, such as extra-academic, institutional and non-governmental reports.

Given this lack of a reflective history regarding ethics for technologies, the argument is that, unlike understanding ethics as a uniquely human capacity, it is necessary to prepare it as part of the systems, so that AI has an ethical agency (Bertoncini; Serafim, 2023). The question is based on that, understanding technology as solely instrumental is an outdated view corresponding to the first moment of the technological-industrial revolution, in which the characteristics are distinct from contemporary information technologies when software and devices condense a greater volume of information that feeds continuous learning systems through human-computer interaction and other means.

It is a consequence of large gaps in the 'territories of knowledge', from intersections between the philosophy of technology in its existential and political dimensions, to the social and Hard sciences'. From an ethics capable of recognizing the 'right of the things' as moral agents (Nas, 2023c) to metaethical perspectives aligned with worldviews distinct from the Eurocentric, such as Amerindian perspectivism (Maciel, 2019), which advocates the defense of rights for rivers, mountains, and everything that exists.

Because of the gaps created by fenced-off disciplinary territories, these are still marginal epistemologies that cause discomfort in the territories of knowledge that govern technological developments.

The debate regarding the challenges of making AI 'ethical *by design*' resembles 'ethical laundering' (Bietti, 2021)[8], since ethical propositions are difficult to apply from a pragmatic perspective it cannot be reduced to merely technical issues. Likewise, demanding the explainability of AI decisions is as complex as explaining human decisions and providing transparency.

In the ethical debate, there is no consensus among experts on the approach to be applied if the intention is to translate ethics through algorithms in computer systems.

Initially, traditional ethical concepts concern individual conduct, while information systems bring together the responsibilities of a group of 'actors', including designers, engineers, developers, companies, and users (Taddeo; Floridi, 2018). Therefore, the context presents challenges for outlining what 'distributed

8 The discussion on *ethical washing* (Bietti, 2021) emphasizes that the view on ethics should be understood as a philosophical exercise of asking questions as a way of achieving knowledge which is only possible by constructing it through reflection. Above all, hiring philosophers or ethical consultants to adapt companies to demands reinforces the idea of exclusive knowledge of a few technicians and specialists who are far from the world in the 'ivory towers'.

ethics' would be (Floridi, 2014; Floridi; Sanders, 2004). In other words, an open debate on new ethical perspectives is necessary when AI advances so rapidly. Such debates are still modest and require another time that involves reflection, listening, and multisectoral, interdisciplinary, and inclusive meetings.

To overcome the inertia caused by the gaps between theory and practice, what is needed and feasible, the solution points to a profound reform in the educational system. Thus, we ask ourselves: how long will Philosophy departments continue to ignore the challenges posed by a world mediated by information technologies, encouraging students to follow the traditional and conservative path of delving into authors seen and reviewed thousands of times? Because they are training philosophers who will have as much difficulty in talking to computer scientists as the other way around. Bioethics continues to be a new and speculative field that discusses disputes between ethical perspectives to evaluate specific contexts. The so-called 'exact' sciences, among other areas of knowledge, could benefit from approaching bioethical debates if they incorporate them into their curricula, opening up fields of reflection on ethics in research from the beginning of a project. Likewise, the arts and design tend to contribute to the search for possible paths of innovation in the sphere of knowledge.

Due to epistemological flaws, intangible technologies are often compared to tangible ones; however, unlike them, they are a set of measured and disseminated information, models of human learning applied to computing, and use the concept of intelligence as a way of training, recognizing and replicating existing, sedimented knowledge. Furthermore, they represent statistics about repetitions, identified as patterns, and therefore, 'truth models' that in practice are falsifiable.

Thus, AI becomes a prosthesis of human intelligence (and its gaps) where the users of the systems cease to be 'clients' and become 'products' and guinea pigs for algorithmic experiments (Garcia-Vigil, 2021). That characteristic threatens human values within ethical and justice principles, considering the meanings of biopolitics and necropolitics discussed here. In this way, the condensed and consolidated knowledge of techno-politics remains in progress, even if criticized, setting precedents for technofascisms to impact lives, bodies, hearts, and minds.

5 – Final Remarks

The problem of bias in knowledge, which defines and separates territories, can be examined when we consider that ethical issues are mediated by assessments of the consequences of acts, by value judgments about what is "good", "acceptable", "bad", and "unacceptable". Since these are issues dependent on narratives, points of view, and contexts, they also replicate power dynamics and disputes based on worldviews. Thus, AI governance efforts towards regulation may be indicative, but not definitive, given that they are incapable of mitigating structural problems of social injustices embedded in algorithms, biased data, and the possible misuse of AI. Since those who suffer the consequences most are the subjects excluded from decisions and opportunities, computer systems only replicate the problems that predate the existence of artificial intelligence, such as universal access to rights considered essential to human dignity (UN, 1948).

Thus, this article proposes that understanding what constitutes 'Responsible AI' should be broader than technical efforts in the legal, legislative, and governmental spheres, since simply establishing standards does not guarantee compliance with them nor their wide application. The existence of legislation does not prevent injustices, just as the possibility of punishment does not prevent many companies from disrespecting essential citizenship rights.

Merely technical solutions can represent *ethical washing* by companies, making their ethical approach performative and inefficient.

Thus, to make the principles of Responsible AI effective, governance mechanisms are needed to bring together representatives from different sectors of knowledge and professional activity, with the application of creative, alternative, and innovative methodologies, remembering that there is no innovation without risk, and part of the risk is to welcome what is unknown or 'discordant'. It is important to treat 'otherness' with hospitality. Here we are not talking strictly about individuals in their diversity of colors, orientations, and preferences, but above all about what they bring incorporated into their existence as expression in the world, with ideas and ways of living that are not defined by labels. Finally, it is necessary to bring together those who dare to disagree productively, by not conforming to how all sectors of society become machines of continuous production, where each element must fulfill an alienating purpose. It is urgent to welcome creative, dissident voices and those fighting against oppression in the territories of knowledge, understanding that bioethics, RAI, and decoloniality are concepts and fields of knowledge under construction.[9]

References

ACM – ASSOCIATION FOR COMPUTING MACHINERY. Statement on principles for responsible algorithmic systems. *ACM Bulletins*, New York, 1 nov. 2022. Available at: https://www.acm.org/articles/bulletins/2022/november/tpc-statement-responsible-algorithmic-systems. Access on 15 mar. 2024.

AGAMBEN, Giorgio. *Homo Sacer*: Sovereign Power and Bare Life. Translated by Daniel Heller-Roazen. California: Stanford University Press. 1998.

ALMEIDA, Virgílio; NAS, Elen. Desafios da IA responsável na pesquisa científica. *Revista da USP*, São Paulo, n. 141, p. 17-28, jun. 2024. DOI: https://doi.org/10.11606/issn.2316-9036.i141p17-28.

ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren. Machine bias. *In*: MARTIN, Kristen. *Ethics of data and analytics*. Sebastopol: Auerbach Publications, 2022. p. 254-264.

ARAÚJO, Camila Souza; MEIRA, Wagner; ALMEIDA, Virgílio. Identifying stereotypes in the online perception of physical attractiveness. *In*: SOCINFO – SOCIAL INFORMATICS INTERNATIONAL CONFERENCE, 8., 2016, Bellevue. *Anais* […]. Bellevue: Springer International Publishing, 2016. p. 419-437. Available at: https://repositorio.ufmg.br/bitstream/1843/ESBF-ALLHYX/1/camilasouzaaraujo.pdf. Accessed: 15 mar. 2024.

BAUMAN, Zygmunt. *Globalização*: as consequências humanas. São Paulo: Zahar, 1999.

BENDER, Emily M.; GEBRU, Timnit; MCMILLAN-MAJOR, Angelina; SHMITCHELL, Shmargaret. On the dangers of stochastic parrots: can language models be too big?. *In*: ACM CONFERENCE

[9] I would like to express my gratitude to Alice Gibson, Ph.D. in Philosophy, for reviewing the text and providing valuable suggestions.

ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 4., 2021, Barcelona. *Anais* […]. Barcelona: ACM, 2021. p. 610-623. DOI: https://doi.org/10.1145/3442188.3445922.

BENJAMIN, Ruha. Assessing risk, automating racism. *Science*, Washington, D.C, v. 366, n. 6464, p. 421-422, 2019. DOI 10.1126/science.aaz3873.

BERTONCINI, Ana Luize Corrêa; SERAFIM, Mauricio C. Ethical content in artificial intelligence systems: a demand explained in three critical points. *Frontiers in Psychology*, Lausanne, v. 14, p. 1074787, mar. 2023. DOI: https://doi.org/10.3389/fpsyg.2023.1074787.

BIETTI, Elettra. From ethics washing to ethics bashing: a moral philosophy view on tech ethics. *Journal of Social Computing*, Piscataway, v. 2+, n. 3, p. 266-283, 2021.

CONROY, Maggie. The ethics of slow AI: why taking time to develop technology matters. *Data Lab Notes*, Online, Appleton, 2023. Available at: https://datalabnotes.com/slow-ai/. Accessed: 14 mar. 2024.

COULDRY, Nick; MEJIAS, Ulises Ali. The decolonial turn in data and technology research: what is at stake and where is it heading? *Information, Communication & Society*, UK, v. 26, n. 4, p. 786-802, 2023. DOI: 10.1080/1369118X.2021.1986102.

DAIR - DISTRIBUTED AI RESEARCH INSTITUTE. *Timnit Gebru launches independent AI research institute on anniversary of ouster from Google*. Oakland: DAIR, 2021. Available at: https://www.dair-institute.org/press-release/. Accessed: 14 mar. 2024.

DECOLONIZAI. *Sobre o projeto*. [2022]. Available at: https://www.decolonizai.com/sobre-o-projeto/. Accessed: 30 jan. 2024.

DIMENSIONS. Linked research data from idea to impact. [2024]. Available at: https://www.dimensions.ai/. Accessed: 14 jan. 2024.

FLORIDI, Luciano. *'Distributed morality'*: the ethics of information. Oxford: Oxford Academic, 2014.

FLORIDI, Luciano; SANDERS, Jeff W. On the morality of artificial agents. *Minds and Machines*, Berlin, v. 14, p. 349-379, 2004. Available at: https://link.springer.com/article/10.1023/B:MIND.0000035461.63578.9d. Accessed: 14 mar. 2024.

FOUCAULT, Martins. *Nascimento da biopolítica*: curso dado no Collège de France (1978-1979). São Paulo: Martins Fontes, 2008.

FOUCAULT, Michel. *História da sexualidade I*: a vontade de saber. Tradução de Maria Thereza da Costa Albuquerque e J. A. Guilhon Albuquerque. Rio de Janeiro: Edições Graal, 1988.

FOUCAULT, Michel. *Microfísica do poder*. Rio de Janeiro: Edições Graal, 1993.

FOUCAULT, Michel. Outros espaços. *In*: FOUCAULT, Michel. *Ditos e escritos*. Barueri: Forense Universitária, 2001. v. 3, p. 411-422.

GARCIA-VIGIL, José L. Reflexiones en torno a la ética, la inteligencia humana y la inteligencia artificial. *Gaceta Médica de México*, Ciudad de México, v. 157, n. 3, p. 311-314, maio/jun. 2021. DOI: https://doi.org/10.24875/gmm.20000818.

GLOBAL GRAND CHALLENGES. *Catalyzing equitable artificial intelligence (AI) use.* Seattle: Bill & Melinda Gates Foundation, 2023. Available at: https://gcgh.grandchallenges.org/challenge/catalyzing-equitable-artificial-intelligence-ai-use. Accessed: 30 jan. 2024.

GORDON, Lewis Ricardo. *Medo da consciência negra*. Tradução de José Geraldo Couto. São Paulo: Todavia, 2023.

GORDON, Lewis Ricardo. *Freedom, Justice, and Decolonization*. New York: Taylor & Francis. 2021.

HAO, Karen. We read the paper that forced Timnit Gebru out of Google: here's what it says. *MIT Technology Review*, Massachusetts, 4 dez. 2020. Available at: https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/. Accessed: 4 jan. 2024.

IPEA - INSTITUTO DE PESQUISA ECONÔMICA APLICADA. *Atlas da violência*. Brasília, DF: IPEA, [2023]. Available at: https://www.ipea.gov.br/atlasviolencia/. Accessed: 14 mar. 2023.

MACIEL, Lucas da Costa. Perspectivismo ameríndio. *In*: ENCICLOPÉDIA de Antropologia. São Paulo: USP, 2019. Available at: https://ea.fflch.usp.br/conceito/perspectivismo-amerindio. Accessed: 4 jun. 2024.

MAGNO, Gabriel; ARAÚJO, Camila Souza; MEIRA JUNIOR, Wagner; ALMEIDA, Virgilio. Stereotypes in search engine results: understanding the role of local and global factors. arXiv preprint, [*s.l.*], nov. 2016. Available at: https://arxiv.org/search/?query=Stereotypes+in+search+engine+results%3A+understanding+the+role+of+local+and+global+factors&searchtype=all&abstracts=show&order=-announced_date_first&size=50. Accessed: 4 jun. 2024.

MBEMBE, Achille. Necropolítica. *Arte & Ensaios*, Rio de Janeiro, n. 32, p. 122-151, 2016. DOI: https://doi.org/10.60001/ae.n32.p122%20-%20151

MBOA NKOUDOU, Thomas Hervé. We need a decolonized appropriation of AI in Africa. *Nature Human Behaviour*, London, v. 7, n. 11, p. 1810-1811, 2023. DOI: 10.1038/s41562-023-01741-3.

MENDONÇA, Ricardo Fabrino; ALMEIDA, Emeritus Virgilio; FILGUEIRAS, Fernando. *Algorithmic Institutionalism:* the changing rules of social and political life. Oxford: University Press, 2024.

MILLER, Katharine. The movement to decolonize AI: centering dignity over dependency. *HAI - Institute for Human-Centered AI*, Stanford, 21 mar. 2022. Available at: https://hai.stanford.edu/news/movement-decolonize-ai-centering-dignity-over-dependency. Accessed: 30 jan. 2024.

MOHAMED, Shakir; PNG, Marie-Therese; ISAAC, William. Decolonial AI: decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy and Technology*, Stanford, v. 33, n. 4, p. 659-684, 2020. Available at: https://doi.org/10.1007/s13347-020-00405-8. Accessed: 4 jun. 2024.

NAS, Elen. *Arte eletrônica*: elo perdido. São Paulo: Amazon Kindle, 2020. *E-book*.

NAS, Elen. *Bioethics of nonpresence:* body, philosophy and machines. São Paulo: Amazon Kindle, 2021. *E-book*.

NAS, Elen. Descolonizar o conhecimento: a perspectiva de Lewis Gordon. *Desenvolvimento Social*, Montes Claros, v. 29, n. 2, p. 189-199, 2023a. DOI 10.46551/issn2179-6807v29n2p189-199.

NAS, Elen. O Manifesto das Coisas: apontamentos para liberalização das vozes suprimidas. *Aurora*. Revista de Arte, Mídia e Política, v. 16, n. 48, p. 5-20, 2023c. DOI: https://doi.org/10.23925/1982-6672.2023v16i48p5-20.

NAS, Elen; AZEVEDO, Telma; LONGHI, Fernando; TERCEIRO, Luciana; VALENTE, Tânia. Future visions for a decolonized future of HCI: thick descriptions of a survey interaction to discuss the colonization of imagination. INTERNATIONAL CONFERENCE ON HUMAN-COMPUTER INTERACTION, Cham, p. 109-116, 2023b. DOI: https://doi.org/10.17613/30bc-6j76.

NASCIMENTO, Elen Cristina Carvalho. Reflexões bioéticas na era da inteligência artificial. *In*: CASTRO, João Cardoso; NIEMEYER-GUIMARÃES, Márcio; SIQUEIRA-BATISTA, Rodrigo (ed.). *Caminhos da bioética*. Teresópolis: Editora Unifeso, 2019. p. 345-362.

NASCIMENTO, Elen Cristina Carvalho. *Arte eletrônica*: elo perdido entre ciência, design e tecnologia. 2017. Dissertação (Mestrado em Design) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2017. Available at: https://www.maxwell.vrac.puc-rio.br/30067/30067.PDF. Accessed: 14 mar. 2024.

NOBLE, Safiya Umoja. *Algorithms of oppression*. New York: New York University Press, 2018.

OECD – ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. *AI principles overview.* Paris: OCDE.AI, 2019. Available at: https://oecd.ai/en/ai-principles. Accessed: 14 mar. 2024.

O'NEIL, Cathy. *Algoritmos de destruição em massa*. São Paulo: Editora Rua do Sabão, 2021.

O'NEIL, Cathy. *Weapons of math destruction*: how big data increases inequality and threatens democracy. New York: Crown Publishers, 2016.

ONU – ORGANIZAÇÃO DAS NAÇÕES UNIDAS. *Declaração universal dos direitos humanos*. Paris: ONU, 1948. Available at: https://declaracao1948.com.br/declaracao-universal/declaracao-direitos-humanos/. Accessed: 14 mar. 2016.

RAJI, Inioluwa Deborah; CHOCK, Sasha Costanza; BUOLAMWINI, Drjoy. Change from the outside: towards credible third-party audits of AI systems. *MacArthur Foundation*, Chicago, 7 jun. 2023. Available at: https://www.macfound.org/press/grantee-publications/outside-scrutiny-to-change-ai-systems. Accessed: 14 mar. 2024.

SIMONDON, Gilbert. *L'individu et sa genèse physico-biologique*. Grenoble: Éditions Jérome Millon, 1995.

TADDEO, Mariarosaria; FLORIDI, Luciano. How AI can be a force for good. *Science*, Washington, D.C, v. 361, n. 6404, p. 751-752, 2018. DOI 10.1126/science.aat5991.

UNESCO – THE UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. *Recommendations on the ethics of artificial intelligence*. Paris: UNESCO, 2023. Available at: https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence. Accessed: 14 mar. 2024.

ZUBOFF, Shoshana. *The age of surveillance capitalism*: the fight for a human future at the new frontier of power. New York: Public Affairs, 2019.

Author's Minibio:

Elen Nas. PhD in Bioethics, Applied Ethics, and Public Health from the Federal University of Rio de Janeiro (2021). Postdoctoral researcher at the Oscar Sala Chair of the Institute of Advanced Studies of the University of São Paulo. Research funded by the Internet Governance Committee in Brazil/IEA-USP (Project 642). E-mail: elennas@usp.br.

Reviewer 2: Icaro Ferraz Vidal Junior, Orcid;
Section Editor: Raquel Kritsch, Orcid.