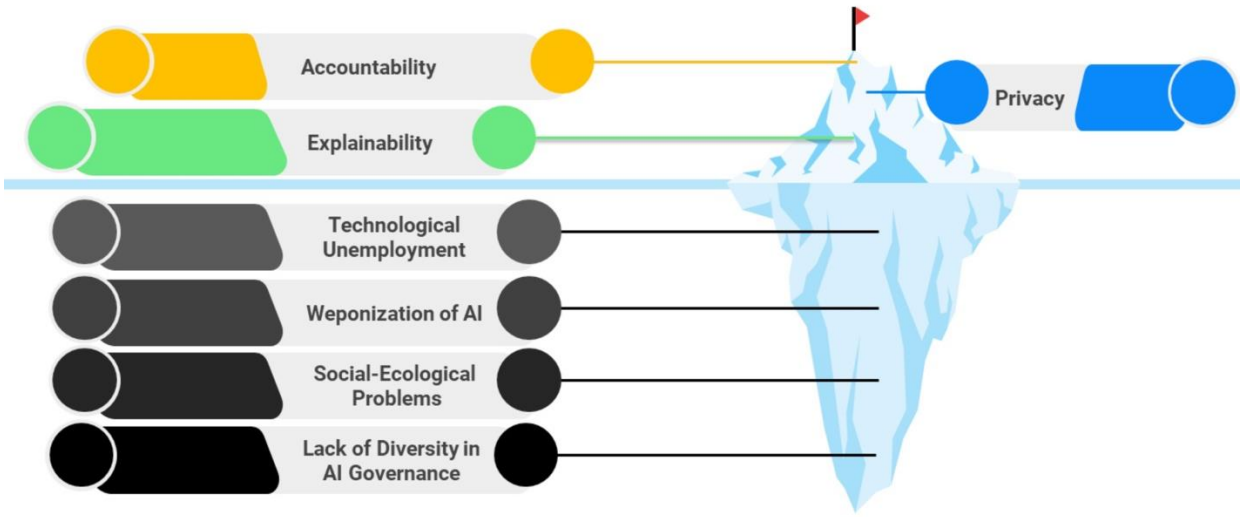

Good AI for the Present of Humanity

Nicholas Kluge Corrêa[1] and Mariana Dalblon[2]

[1]Master in Electrical Engineering and Ph.D. student in Philosophy (PUCRS), nicholas.correa@acad.pucrs.br, ORCID: 0000-0002-5633-6094

[2]Lawyer from the Laureate International Universities – Ritter dos Reis, mbdalblon@gmail.com

Graduate Program in Philosophy of the Pontifical Catholic University of Rio Grande do Sul – Av. Ipiranga, 6681 - Partenon, Porto Alegre - RS, 90619-900.

Abstract

There is a link between critical theory and some genres of literature that may be of interest to the current debate on AI ethics. While critical theory generally points out certain deficiencies in the present to criticize it, futurology, and literary genres such as Cyber-punk, extrapolate our current condition into possible dystopian futures to criticize the status quo. Given the advance of the AI industry in recent years, an increasing number of ethical matters have been pointed and debated, and we have converged into a couple of principles, like Accountability, Explainability, and Privacy as the main ones. But certainly, this can't be all. While most of the current debates around AI Ethics revolve around making AI "good", we see little effort made to make AI good for everyone. This raises questions as, what published ethical guidelines fail to cover? Like critical theory and literature warns us, what kind of future are we creating? At the expenditure of who? Does AI governance occur inclusively and diversely? In this study, We would like to present two aspects omitted or barely mentioned in the current debate on AI ethics. The current humanitarian costs of our new industrial automatize revolution and the lack of diversity in this whole modernization process.

Keywords: AI ethics, Technological unemployment, Humanitarian cost, Lack of diversity, Ethical guidelines.

I. Critical Theory and Cyberpunk

With the current progress in Artificial Intelligence (AI), increasingly intelligent systems are becoming a part of and helping us shape our society. The so-called 4th Industrial Revolution is the culmination of the digital revolution, where technologies such as robotics, nanotechnology, genetics, and artificial intelligence, promise to transform our world and the way we live (Mulhall, 2002). At the present moment, the most accessible and massively used technology in our society, of those mentioned above, is AI. One of the differences between this present moment of technological modernization, when compared to those of the past, is that machines are progressively surpassing our cognitive capabilities in several areas. Given the size and complexity that our society has grown, human beings alone are not able to cope with the demands of processes that are vital to our civilization, and we increasingly rely on the help of intelligent autonomous systems.

We may say, perhaps without many controversies, that our current society cannot exist in its present form without the help of such technologies. Samuel Butler (1863), in his work, "Darwin Among the Machines", questioned our "quiescent bondage" to technology. Butler argued that one day we would reach the point where society would no longer be able to separate itself from its technological creations because it would be equivalent to the suicide of the status quo. In Butler's words: "[…] this at once proves that the mischief is already done, that our servitude has commenced […]". In the end, whether all the technological modernization we experience will result in a future good for all humanity is still a question with no answer. And to us, this is an answer worth pursuing sooner rather than later. This uncertainty concerning the type of future we are creating has raised several critics and warnings from areas such as sociology, literature, and philosophy.

Contemporary critical theory, with its origins in sociology as well as literary criticism, proposes to conduct a reflexive and critical assessment of society and

culture to reveal and challenge deficits in their underlying power structures. We propose that there is a fruitful relationship between the criticism made by contemporary critical theorists, like Craig Calhoun (1995), Paul Virilio (1997), and Hartmut Rosa (2010), that can help ethics be more "what it was meant to be". We also would like to point that, while contemporary critical theory focuses on the present to point out deficits in our society, another possible form of criticism involves extrapolating the future to criticize the present. Cyberpunk, a subgenre of science fiction, seeks to show how our technological advances can lead our society to dystopian futures. Authors such as Philip K. Dick (Do Androids Dream of Electric Sheep?), John Brunner (Stand on Zanzibar), William Gibson (Neuromancer), surrounded by the technological innovations of the '80s and '90s, internet, AI, robotics, virtual reality, genetics, gave rise to a form of literature aimed to criticize the status quo. Fredric Jameson defines cyberpunk as: "[...] the supreme literal expression, if not of postmodernism, then of late capitalism itself" (Jameson, 1991, p. 417). Similar to Jameson, Jean Baudrillard (1994) proposed that given the rapid pace of social and cultural transformation we are experiencing, sociological studies are increasingly approaching what we call science fiction, where we progressively need to anticipate social change while it is happening.

What we see today, especially in the context of AI ethics, is a "kind of" response to the postmodern critique of cyberpunk, that is: how can we avoid the blind march into the dystopian future? In this context, the premise for security issues involving our technological advance is founded on an idea of a negative utopia, in the words of Robert Tally:

> First of all, the utopian impulse must be negative: identify the problem or problems that must be corrected. Far from presenting an idyllic, happy and fulfilled world, utopias should initially present the root causes of society's ills [...] to act as a criticism of the existing system (Tally, 2009, p. 115).

We can thus establish the critique proposed by critical theory, and cyberpunk itself, as a manifestation of the negative utopian impulse, using our possible future as a research subject. But do we see this spirit of critique in the current debate of AI ethics? In my opinion, very little. What we can say we see is a great deal of "Ethical Guidelines" that promise to ensure the utopian development of AI for all

(Russell, Dewey, & Tegmark, 2015; Amodei et al, 2016; Boddington, 2017; Goldsmith & Burton, 2017; Greene, Hofman, & Stark, 2019). But of course, if we choose to follow them. It's not like they are rules or laws, right? Would all these published ethical guidelines have any real normative power over the AI industry? Like Ryan Calo (2017), I also think that ethical guidelines end up serving more as a marketing strategy than a real effort to regulate the tech industry. As Calo says: "[...] Several efforts are underway, within the industry, academia, and several NGO's, to resolve the ethics of AI. But these efforts probably cannot replace policy making" (Calo, 2017, pp. 407-408).

To support this claim, We cite a controlled study conducted by McNamara et al (2018) at North Carolina State University. The sole purpose of this study was to investigate whether ethical guidelines have any effect on the decision-making of software engineers. In their research, the authors evaluated 63 software engineering students and 105 professional software developers, analyzing whether the ethical guidelines of the Association for Computing Machines[1] (ACM) would have any influence on moral dilemmas related to software production. The question the authors sought to answer was: Does the presence of a code of ethics influence ethical decisions related to software? And the answer was: "Despite our stated goal, we found no evidence that the code of ethics of the ACM influences ethical decision-making". Does this study show that maybe there is a hole in the academic education of software developers like applied ethics? Or perhaps we are just inefficiently using Ethics?

Several studies support the idea that ethical guidelines have little to no effect on decision-making in many different professional fields (Brief et al, 1996; Cleek & Leonard, 1998; Lere & Guamnitz, 2003; Osborn et al, 2009). And this idea resonates with several criticisms raised against the current state of AI ethics: Jobin et al (2020, p. 389);

i. "*Private sector involvement in the field of AI ethics has been questioned for potentially using soft policies as a way to turn a social problem into something technical or to completely avoid regulation*".

[1] ACM Code of Ethics and Professional Conduct. Disponível em: http://www.acm.org/binaries/content/assets/membership/images2/fac-stu-poster-code.pdf

Hagendorff (2020, p. 389). 99);

ii. "*AI ethics - or ethics in general - have no mechanisms to reinforce its normative claims*".

Rességuier and Rodrigues (2020, p. 1):

iii. "*Ethics have great powerful teeth. Unfortunately, we are barely using them in AI ethics - no wonder then that AI ethics is called toothless*".

And finally, Brent Mittelstadt:

> Statements reliant on vague normative concepts hide points of political and ethical conflict. 'Fairness', 'dignity', and other such abstract concepts are examples of "essentially contested concepts" which have many possible conflicting meanings requiring contextual interpretation through one's background political and philosophical beliefs. These different interpretations, genuinely held, lead to substantively different requirements in practice which will only be revealed once principles or concepts are translated and tested in practice. At best, this conceptual ambiguity allows for the context-sensitive specification of ethical requirements for AI. At worst, it masks fundamental, principled disagreement and drives AI Ethics towards moral relativism. At a minimum, any compromise reached thus far around core principles for AI Ethics does not reflect meaningful consensus on a common practical direction for 'good' AI development and governance (Mittelstadt, 2019, pp. 503).

The point we want to make is that the role of ethics is not to be a soft version of the law, even if the laws are based on ethical principles. That is not where ethics finds its power. Like critical theory and literature, the real application of ethics lies in challenging the status quo, seeking its deficits and blind spots. True ethicists concerned with the current state of the AI industry should not only reinforce the repetition and defense of the same concepts already cited by numerous published guidelines. Guidelines made by the very industry that it's (weirdly) "self-regulating" itself. But we should seek to re(visit) all the issues that are being forgotten. Issues like diversity, representativeness, anti-war policies, equality of income and wealth distribution, the preservation of our socio-ecological system, things that are rarely cited in this so-called Ethical Guidelines.

II. Safety Issues and AI Ethics

But what does this Ethical Guidelines talk about? Who makes this? Jobin et al (2019), in their meta-analysis, mapped all the countries responsible for producing the existing ethical guidelines, at that time, for AI regulation. Their research identified 84 documents containing ethical guidelines for intelligent autonomous systems, divided by eleven ethical principles: transparency, justice & equity, non-maleficence, responsibility, privacy, beneficence, freedom & autonomy, trust, dignity, sustainability, and solidarity, with convergence around five principles; transparency, justice, non-maleficence, responsibility, and privacy. Hagendorff's (2020) meta-analysis of the main ethical guidelines published in the last five years showed that the main ethical principles cited by them were similar to Jobin et al findings, accountability, explainability, and privacy, appearing in almost all guidelines. These principles can be described as follows:

1) *Accountability:* how to make the AI industry accountable for its technologies. For example, in the context of autonomous vehicles, what kind of guarantees and responsibilities should companies developing autonomous vehicles offer to society (Maxmen, 2018)?
2) *Explainability:* one of the greatest deficits in contemporary machine learning systems is that it is difficult to explain the internal process of these types of AI systems, especially when using architectures like deep neural nets (Mittelstadt et al, 2019);
3) *Privacy:* An interesting analogy with the second industrial revolution is that data is like coal for AI, and the big technology companies, like Google, Amazon, and Facebook, are the coal mines of today. The abundance of data that we produce daily ensures an inexhaustible source of information for the training of AI systems. However, the use of personal data without consent is one of the main preoccupations found in the literature involving AI ethics (Ekstrand et al, 2018).

Jurić et al (2020) conducted a similar study, a quantitative bibliographic survey on the recent expansion of AI safety research and its main topics of interest. The common motivation for short and long-term interests in AI safety and AI ethics is

the same: how to make the interaction between humans and AI safe and beneficial? And this is what a lot of the contemporary debate on AI ethics has delimited itself. Questions like how to make possible advanced AIs operating by reinforcement learning corrigible (Soares et al, 2015; Turner et al, 2020), or how to align the terminal goals of AI systems with our values (Soares, 2016; Russel, 2019), And even how to integrate human society in a post-Singularity era (Chalmers, 2010). As much as anyone versed in the field, the authors also don't want powerfully misaligned AI systems in the near/far future, turning its future light cone into paper clips or anything like that. Nor we desire any kind of hellish dystopian Singularity desolated future, and probably no one really does, and that's o literal no-brainer. Not wanting to reduce the importance of Alignment research and all the benefits it may bring to future humanity, but we ask what about present humanity? What are we doing to prevent the side-effects of AI and mass automation right now? How will survive to enjoy the pleasures of aligned AI in the future?

Furthermore, Hagendorff (2020) points out in his meta-analysis that the main principles, Accountability, Explainability, Privacy, mentioned in the most recent published ethical guidelines have a considerable technical effort made to ensure aspects such as transparency, legal accountability, and preservation of privacy. However, several other issues are still not mentioned by even half of the published ethical guides. According to Hagendorff, of the 22 most relevant published ethical guidelines into the last five years, only nine mention labor rights and technological unemployment, while only two mention the lack of diversity in the tech and AI industry. Since this is some of the most underrated problems, social problems, we think it's fair that we gave them a little bit more room in the current AI ethical debate, so we can see the current humanitarian cost and social risks we are facing, with "weak" AI run by a misalign world.

III. Who are we leaving behind?

What would be the backbone of our technological industry? As farmers are the backbone of society, what would be the analog for the technological society? Not that farmers cease to be fundamental agents of the social structure, but I would like to look for an agent more connected to the support of our technological

civilization. We could say that engineers, computer scientists, mathematicians, technical experts, as well as other occupations involving specialized knowledge such as administrators, designers, entrepreneurs, are fundamental agents for the technological industry. But let's take a step back. It isn't as if microprocessors and GPUs were born in trees? The industry responsible for producing the entire technological machinery, the hardware, would be more fundamental than the agents previously mentioned because there would be no artificial intelligence without the artificial artifact that supports it. Now, one last step backward. All the components necessary to build the most sophisticated hardware depend on physical substrates that compose and support our entire technological infrastructure. After all, the smartest software ever created cannot act in the physical world without the hardware making the connection between the digital and the physical world.

The physical substrate we will focus on in this session is Tantalum. Tantalum is a metal usually mined in the form of a mineral called tantalite. Tantalite is a mixture of niobium (Nb) and tantalum (Ta) also known as "Coltan", short for columbite-tantalite. Coltan is mainly mined for tantalum extraction since this metal has several properties and useful applications for the tech industries. Nowadays, it is almost impossible not to have contact with this type of material. Practically all cell phones have capacitors containing tantalum powder or wire. Computers, cars, aircraft engines, surgical instruments, orthopedic implants, global positioning systems (GPS), missile parts, all use tantalum as a raw material in one way or another. As much as there are substitutes like aluminum, performance is not equal, so tantalum is still targeted as the primary choice in applications where performance and reliability are required (Usanov et al, 2013).

According to Burt (2010), the largest known tantalum reserves in the world are located in South America (41% of all tantalum in the world), specifically in the Amazon rainforest region, Brazil. However, even though Brazil has the largest natural reserves of tantalum in the world, it is Central Africa that produces the

largest amount of tantalum (37%) (Sutherland, 2011). According to Garside[2], the world's largest exporters of tantalum are the Democratic Republic of Congo and Rwanda, followed in 3rd place by Brazil. And Coltan mining in Africa is linked to one of the most atrocious humanitarian disasters in the history of the 20th and 21st centuries. A disaster whose damages can still be seen today in the Democratic Republic of Congo (DRC).

Ironically, one of the most fundamental elements for the maintenance and expansion of our technological society has an artisanal origin. Small-scale artisanal mining occurs all over the world. However, the world's largest producer of tantalum, the DRC, is also where artisanal small-scale mining occurs more often. This type of labor is carried out by small groups of individuals often formed by family units or workers' cooperatives, with little to no form of mechanization to aid them, acting informally and illegally. In short, artisanal mining is a method of mineral extraction that requires less investment and capital and can operate with agility, responding promptly to market needs, without the need to guarantee any form of workers' rights. Artisanal miners do not have a fixed wage, only receiving a fixed percentage of their production, which makes the workers hostage to the fluctuations of the tantalum price in the market (Dorner et al, 2012). In 2011 artisanal mining supported approximately 16% of the DRC population, 13.5 million people, making it one of the most profitable labor activities in the country, more than agriculture (Nest, 2011, p. 37).

How can a country as rich in natural resources as the DRC be called a geological scandal[3]? Instead of the country's riches being used for the development of the DRC, years of conflict, mismanagement, corruption, and international interests have imposed one of the lowest human development rates in the world[4]. Political instability in the DRC has led several mining industries to shut down their

[2] Garside, M. (18 de Fevereiro, 2020). Global tantalum production by country 2019. *Statista*. Disponível em: https://www.statista.com/statistics/1009165/global-tantalum-production-by-country/

[3] New Internationalist (2 de Maio, 2004). The Looting of the Congo. Disponível em: https://newint.org/features/2004/05/01/congo

[4] According to the United Nations Human Development Data (1990-2018), the DRC has a human development index of 0.459 (179th place in the human development ranking). Retrieved from http://hdr.undp.org/en/data

operations in eastern Congo. In response, artisanal mining has become the obvious solution to satisfy a market with constant demand, and the eastern Congo is still a region of conflict between the state and rebel groups. Between 2000 and 2001 there was an increase in the value of tantalum in the world market. This led to a massive expansion of artisanal mining in the DRC. This event was called the Dot-com boom, a stock market bubble caused by excessive speculation in Internet-related companies in the late 90s (Cassidy, 2009). During this time the DRC was facing its Second War, and between 1998 and 2003 more than 73,000 people were dying every month. It's estimated that between 1999 and 2000 the Rwandan army made at least $20 million a month due to the control of artisanal tantalum mining in eastern Congo (Montague, 2002). According to Sutherland (2011), it is estimated that at least 5,000,000 people were killed during the conflict at this time. 40% of these people were women and children, and a similar number were displaced due to the war, which reports the use of child soldiers and widespread sexual violence.

To provide a more "first-person" position on the exploitation of DRC's natural resources, we quote a small sample from a series of interviews conducted by Tegera et al (2002) with artisanal miners from the North Kivu region in eastern Congo:

> Question: What are the main problems linked to coltan mining?
>
> Answer: The young people who do this drop out of school, and even some teachers leave school. There is moral depravity in the mines: No morals, no difference between the sexes in this activity where you find men and women working naked. Prostitution is booming in these immoral mines, couples are formed just like that. There is also the rape of children under 16. There are drug and alcohol abuse even among the children. There are also child marriages. Agriculture is abandoned and left to the women. There are landslides, grazing land, and fields that are destroyed. Drugged fighters are dangerous.
>
> Question: What positive aspects do you see in coltan mining?
>
> Answer: I don’t see any positive aspects because it benefits foreigners who have a sales monopoly (Tegera, 2002, p. 14).

Currently, the Congolese tantalum trade is largely illegal, being conducted by criminal organizations and rebel militias. This is against the right of permanent sovereignty over the natural wealth of the state guaranteed by the UN Resolution of 1803[5]. This resolution guarantees the right of peoples and nations to permanent sovereignty over their natural resources, which must be exercised in the interest of their national development and the well-being of their people. Also, international interference must only occur in such a way as to benefit the development of the state in question. And this is not what is happening. Since accountability is such an important ethical principle for the tech industry ah AI Ethics, let us raise the question, who is responsible for this? Who is going to pay the bill for reckoning the DRC? Because we, as a global society, definitely should help support and rebuild the backbone of our technological civilization. Perhaps we should allow the International Criminal Court (ICC) to have jurisdiction over both individuals and companies, allowing companies to be prosecuted for crimes against humanity. Since ethical principles do not prevent entire nations from being harmed, perhaps the possibility of a company having to close its doors if it doesn't get its act straight is incentive enough.

IV. Who will lose their job?

The automation of processes that were previously carried out by human individuals has been one of the main sources of technological unemployment in the last two centuries (Peters, 2017). Many jobs and forms of occupation have not lasted more than a century in our society, such as telephone operators, typists, public pole lighters, night-time soil collectors, elevator operators, ice cutters, furnace burners, and many other labor occupations. Nowadays, with the use of IA companies can drastically reduce their need for human labor to lower their costs. However, the adoption of this management policy has two obvious consequences:

1) Wealth accumulation for companies oriented to the development of AI;
2) The unemployed population replaced by AI would find themselves without any source of income.

[5] United Nations, General Assembly, "Permanent Sovereignty over Natural Resources," General Assembly Resolution 1803, 14 December 1962. https://www.ohchr.org/Documents/ProfessionalInterest/resources.pdf

This reality is best summarized by Erik Brynjolfsson[6] in the following quote: "*It is one of the dirty secrets of the economy: technological progress makes the economy grow and creates wealth, but there is no economic law that says everyone will benefit*". Frey and Osborne (2013) estimated the probability of automation for 702 occupations in the US. The result showed an estimate that 47% of these occupations will be eliminated by technology over the next 20 years. This estimate can be used in other regions of the world, such as Latin America, which, according to a study published by the International Labour Organization (ILO), breaks a historical record of unemployment[7].

In the second quarter of 2020, Brazil registered 12.8 million unemployed people[8], 1.8 million more than at the end of 2019[9]. The pandemic of the new coronavirus has helped accelerate job losses in other sectors as well. Call centers, a sector that is already being rapidly automated, due to the current pandemic has had to close several workplaces due to unsafe working conditions. This caused several companies to increase the use of chatbots and virtual assistants (Hao, 2020). Recently, in May 2020 more than 90 university professors were fired from the Laureate group, responsible for universities such as Anhembi Morumbi, the FMU University Center, and other universities in Brazil. The fired professionals were all responsible for teaching disciplines in a distance education format (EAD). The Laureate group replaced these professionals with "monitors" and autonomous tools for proofreading (Domenici, 2020).

Although autonomous vehicles are not yet a publicly available technology, their test versions are circulating in several US cities. It is estimated that by 2021 at least five major automotive companies will have autonomous cars and trucks available for the general public (Maxmen, 2018). But what will be the effect of automation

[6] Interview by Rotman, D. (2013). How technology is destroying jobs. [online] MIT Technology Review, June 12, 2013. Retrivied from https://www.technologyreview.com/2013/06/12/178008/how-technology-is-destroying-jobs/

[7] Retrivied from https://www.ilo.org/caribbean/newsroom/WCMS_749692/lang--en/index.htm

[8] Retrivied from https://www.ibge.gov.br/explica/desemprego.php

[9] Retrivied from https://www.ibge.gov.br/estatisticas/sociais/trabalho/9173-pesquisa-nacional-por-amostra-de-domicilios-continua-trimestral.html?=&t=series-historicas&utm_source=landing&utm_medium=explica&utm_campaign=desemprego

on the transportation industry and its workers? How much of the working population will be affected? According to the company Uber[10], more than one million drivers work for the company in Brazil. According to the Brazilian National Agency for Land Transport[11], the country's truck fleet has 1.941 million units. Of this total, 703,000 vehicles are owned by independent truck drivers. Meanwhile, data regarding the number of delivery workers in Brazil is difficult to obtain. According to Eufrásio and Goulart (2020), the company iFood has registered 170 thousand deliverers in Brazil, and the Rappi platform has 200 thousand deliverers throughout Latin America. Brazil also has the largest number of motorcycle delivery workers in the world, according to Sindimoto - SP[12] (São Paulo State Union of Motorcycle, Cyclist, and Mototaxi Messengers), in 2017 Brazil had more than 1.85 million motorcycle delivery workers (representing 30% of the workforce in Brazil). Would our social support network for unemployed individuals be ready to deal with this demand?

Another preoccupying aspect is the perceptible growth of informal work alternatives (Antunes, 2009), characterized as another humanitarian problem involving labor exploitation. An example easy to cite is the emergence of click working. Click work is a type of essential task for training AI systems. Usually, machine learning requires large amounts of labeled data to become proficient in certain tasks. Thus, human individuals are hired to perform extremely easy and boring digital tasks such as image classification (Irani, 2016; Silberman et al, 2018). Companies that hire such a workforce usually do not offer a minimum wage, paying less than 2 dollars per hour, and sometimes even charging commissions for each transaction made by the workers (yes, these individuals must pay to receive for their service). If all this is not absurd enough, workers have difficulty receiving payment, obtaining technical assistance, or any other kind of support from the companies they work for (Harris, 2014). In the end, behind

[10] Retrivied from https://www.uber.com/pt-BR/newsroom/fatos-e-dados-sobre-uber/

[11]Retrivied from https://agenciabrasil.ebc.com.br/economia/noticia/2019-12/governo-lanca-programa-de-incentivo-caminhoneiros-autonomos#:~:text=Segundo%20dados%20da%20Ag%C3%AAncia%20Nacional,apenas%2026%20mil%20s%C3%A3o%20cooperados.

[12] Retrivied from http://www.sindimotosp.com.br/noticias/noticia146.html

almost every intelligent classification algorithm is a bunch of developing country workers spending their days saying what pictures are cats and what are dogs.

The fact that large AI companies pay pennies for the kind of essential work that makes machines learning efficient and valuable demonstrates a certain indifference on the part of these companies to notions of economic egalitarianism, income equality, and the value of human capital. These new labor relations without analyzing it under the light of the ILO guidelines put the lives of these individuals at risk. According to the ILO Declaration on Social Justice for a Fair Globalization made in 2008[13], it is necessary to establish four minimum objectives to be followed by the entire global society:

i. promote employment by creating a sustainable institutional and economic environment;
ii. adopting and expanding social protection measures - social security and worker protection - that are sustainable and adapted to national circumstances;
iii. promote social dialogue and tripartism as the most appropriate methods;
iv. respect, promote, and apply fundamental principles and rights at work, which are of particular importance, both as to rights and as conditions necessary for the full realization of strategic objectives.

The working modality mentioned above, click working, is in direct violation of ILO guidelines. In addition to the ILO, the United Nations also provides in Article 23[14] of The Universal Declaration of Human Rights that every individual must receive a dignified, fair, and satisfactory remuneration that assures him/her and his/her family an existence compatible with human dignity. Thus, it must be a reason to warn everyone that, although there are new labor relations, they cannot be allowed to be precarious in the name of technological advancement.

[13] Retrivied from http://ilo.org/global/about-the-ilo/mission-and-objectives/WCMS_371208/lang--en/index.htm#:~:text=Adopted%20in%202008%20by%20the,in%20the%20era%20of%20globalization.

[14] Retrivied from http://www.capital.sp.gov.br/cidadao/familia-e-assistencia-social/conheca-seus-direitos/declaracao-universal-dos-direitos-humanos#:~:text=Todo%20ser%20humano%20que%20trabalha,outros%20meios%20de%20prote%C3%A7%C3%A3o%20social.

And how have the AI ethics responded to this possibly inevitable wave of technological unemployment on our way? How can we distribute the new goods and services generated by this economy sustained by intelligent automation? Solutions to this problem range from Universal Basic Income (Russell et al, 2015), to legal clauses that ensure that companies involved in the AI industry are committed to sharing their profits with society, "Windfall's Clause" (O'Keefe et al, 2020). However, these solutions may generate other types of problems. Justin Gest (2016), in his book "The new minority: White working-class politics in an age of immigration and inequality" describes how unemployment is correlated with suicide, divorce, homicide, theft, and other social problems related to marginalization and exclusion.

With this in mind, a critique of solutions such as "UBI's for all" and the "Windfall clause" would be that they are just advertising solutions. So that it would appear that there is a plan to combat technological unemployment, and while the wave keeps coming, no real strategy is being implemented. We believe that strategies to mitigate the side effects of mass unemployment should be implemented in all levels of society, and not only regarding the utopian hypothesis that there will be money for everyone. What we need are strategies aiming at the implementation of educational and inclusive processes that enable the relocation of unemployed individuals within the new labor market.

Finally, the question of whether "individuals would be happy and fulfilled if they did not have to contribute/work for society" is a complex existential problem. But for writers like Voltaire[15], "Work saves us from three great evils: boredom, addiction and need". Fyodor Dostoevsky[16] shows a similar reflection in his work

[15] Voltaire. Candide. Published by Cramer, Marc-Michel Rey, Jean Nourse, Lambert, and others. 1759.

[16] [...]I ask you: what can be expected of man since he is a being endowed with strange qualities? Shower upon him every earthly blessing, drown him in a sea of happiness, so that nothing but bubbles of bliss can be seen on the surface; give him economic prosperity, such that he should have nothing else to do but sleep, eat cakes and busy himself with the continuation of his species, and even then out of sheer ingratitude, sheer spite, man would play you some nasty trick. He would even risk his cakes and would deliberately desire the most fatal rubbish, the most uneconomical absurdity, simply to introduce into all this positive good sense his fatal fantastic element. It is just his fantastic dreams, his vulgar folly that he will desire to retain, simply in order to prove to himself, as though that were so necessary, that men still are men [...] (Dostoevsky, 1992 (1864), Chapter 8, p. 21).

"Notes from the Underground". We cannot speak for these authors, but for ourselves, we think that the idea that the abolition of human work will bring about the flourishing of society is, at best, a mistaken assumption.

V. "Bring The Rest In"

It is no surprise that issues such as labor exploitation, violation of humanitarian rights, and mass unemployment, which are often social problems related to developing countries, are not raised much in the current debate on AI Ethics. And perhaps this is because all published ethical guidelines are produced by a minority of our global society. Would this minority be qualified to defend the morality and ethics of the entire global society?

Interesting research on AI ethics to be cited, and certainly one of the largest research on experimental ethics ever done in history, was the Moral Machine[17] experiment conducted by Awad et al (2018). The experiment was an online survey designed to explore moral dilemmas faced by autonomous vehicles, using the formal structure of the well-known trolley problem. A dilemma where the decision-maker must choose between sacrificing the lives of X individuals or letting an out-of-control trolley cart kill Y individuals. The platform has achieved great reach, gathering 40 million decisions in ten languages from 10 million people in several different countries. As said, this was probably the greatest experiment in ethics ever made in history. In the experiment global moral preferences were summarized in nine groups, which characterize decision patterns, such as the preference for inaction, the preference for saving females rather than males, the preference for sparing more people, etc.

Measuring the individual variations of preferences based on the demographic and geographic data of the participants significant cross-cultural ethical differences were observed. This allowed the grouping of three large social clusters: Eastern (mainly formed by Islamic and Confucian countries and cultures), Western (formed by Protestant, Catholic and Orthodox countries in Europe) and Southern (formed by Latin American countries in Central and South America, and several African countries). The distributions among the three clusters revealed significant

[17] Retrivied from https://www.moralmachine.net/

differences in preferences. For example, the Eastern cluster was characteristic in preferring to save more lives and spare pedestrians to the detriment of drivers. The Western cluster had a strong preference for inaction and sparing the younger ones. Meanwhile, the countries belonging to the Southern cluster showed a much stronger preference concerning saving females.

The findings of Awad et al serve as evidence in favor of the argument that ethical principles cannot be based on a single vision (Western, Eastern, Eurocentric, etc.) of AI ethics. Therefore, ethical guidelines should be created inside an intercultural context. When we think about AI governance and regulation for the technological industry and the use of AI, an optimistic view would be that we have a global consensus on how we should direct this progress. But a more realistic view shows us that we are in an era where the term "technological colonialism" is becoming increasingly meaningful. Mohamed et al (2020) use tools from post-colonial and decolonial theories to critically analyze how the inequality of technological power between different states continues to perpetuate a new form of exploitation and control of developing countries.

In the technological context, this new form of colonialism can be analyzed by the values that guide technological development, and more specifically, the development of AI ethical guidelines. Besides the epistemic values that form our notion of scientific objectivity, e.g., generalization and falsification, there is a strong consensus in the literature that non-epistemic values guide and shape scientific reasoning, and in our case of interest, the interpretation and application of technological developments (Laudan, 1968; Nissenbaum 2001; Douglas, 2007; DiSalvo 2012; Friedman et al. 2013; Elliott and McKaughan 2014, Bueter, 2015).

When non-epistemic values guide the technological progress of AI, an obvious question may come to mind: what values are being taken into consideration? We need to remember that when ethical guidelines are written, they are made to reflect the core values of the culture and society responsible for writing them, not in the defense of states, cultures, or segregated populations. And it is exactly the segregated, the periphery of society, that is the first to feel the side effects of the rapid changes caused by our technological advance. The post-colonial ills that persist in many states can be represented by the structural biases in the political

and social fabric of their society. And with recent advances in artificial intelligence, such biases have gained a new characteristic: the automation of prejudice.

Tools created to optimize processes in several areas end up becoming oppressive paraphernalia, favoring certain social groups over others. Examples of how classification algorithms can act in a biased way are not difficult to find in the literature, and a simple internet search can yield the reader countless cases of how such systems threaten the very notion of human dignity. The Brazilian government has recently adopted the use of video-monitoring and facial recognition technologies (AI). Through Decree Nº 793[18] of October 2019, the former Minister of Justice Sérgio Moro presented the decree as a way to modernize Brazil's police forces. However, what has been happening is a step backward concerning issues such as transparency, accountability, and protection of personal data. According to Nunes (2019), the type of policy being adopted has only increased the mass incarceration of peripheral and segregated populations. First, the facial recognition techniques used by Brazilian police forces are not accurate, something that can generate arbitrary arrests and human rights violations. According to a report made available by the Criminal Defense Coordination and the Board of Studies and Research on Access to Justice of the Public Defender's Office of Rio de Janeiro[19], between June 1, 2019, and March 10, 2020, there were at least 58 cases of erroneous photographic recognition, resulting in unjust accusations, and even the imprisonment of innocent individuals. 70% of the unjustly accused were black. Second, Nunes points out that since the implementation of such systems the black population has been disproportionately affected. In 2019, 90.5% of those arrested by facial recognition and video-monitoring systems are black.

Other cases of algorithmic bias toward gender and sexual orientation are cited by Costanza-Chock (2018), which showed that intelligent airport screening systems systematically signal transsexual individuals for security searches. A controversial study by Wang and Kosinski (2017), where the authors stated that "classification algorithms can infer sexual orientation from facial images," caused a series of

[18] Diário Oficial da União - Portaria Nº 793. Retrivied from https://www.in.gov.br/en/web/dou/-/portaria-n-793-de-24-de-outubro-de-2019-223853575

[19] Public Defender's Office of Rio de Janeiro. Retrivied from http://www.defensoria.rj.def.br/uploads/imagens/d12a8206c9044a3e92716341a99b2f6f.pdf

criticism from the LGBTQ + community (Agüera y Arcas et al, 2018). Kosinski made even more controversial statements in an interview for The Guardian (Levin, 2017), stating that soon intelligent algorithms will be able to measure IQ, political orientation, and criminal inclinations from facial images only.

Another unethical aspect that we can cite is that developing countries are being used as a "test area" for new technologies. For example, before the involvement with Donald Trump's political campaigns, Ted Cruz, and the separation of the UK from the EU, Cambridge Analytica tested its tools in the 2015 elections in Nigeria, and 2017 in Kenya. These countries were chosen for the company's beta phase due to milder data protection laws, which facilitated the unscrupulous use of prediction and classification systems to influence these countries' elections, i.e., social engineering architected by foreign agents equipped with intelligent autonomous systems (Nyabola, 2018). It is worth noting that the series of lawsuits that led to the company's bankruptcy in 2018 only occurred after its acts against the democracy of countries such as the U.S. and the United Kingdom came to light, and not for its interference in Niheria and Kenya.

It is not as if developing countries don't have access to technology or algorithmic tools. A report entitled "The Global AI Agenda" written by MIT Technology Review Insights[20] collected data showing how the use of technologies involving AI is growing in Latin American countries. According to the report, almost 80% of large Latin American companies use AI in some form, and more than half of the companies interviewed (55%) cite customer service (with chatbots) and recommendation systems used in e-commerce as their main applications. However, one of the main limitations that the interviewed companies reported was the limited participation of Latin America in the development of global governance structures involving the use and development of AI. The European, North American, and Chinese dominance in the development of such guidelines make their integration in the Latin American context difficult and sometimes impractical.

[20] MIT Technology Review Insigh. (2020). The global AI agenda: Latin America. The global AI agenda series. Retrivied from https://mittrinsights.s3.amazonaws.com/AIagenda2020/LatAmAIagenda.pdf

Carman and Rosman (2020) raised the same issue, however, they focused on the African continent. For these authors, the use of technologies involving AI and the establishment of foreign governance structures is a delicate issue in the African context, given the continent's long history of imposing external values. In 2019 such concerns culminated in several G20 participating countries, such as India, Indonesia, and South Africa, refusing to sign the Osaka Track, an international declaration regulating aspects of e-commerce and data flow from the WTO (World Trade Organization) (Kanth, 2019). The refusal happened because the interests of these countries, as of several others, were not being represented in this document, denying political autonomy for the states themselves to go through their digital industrialization.

In the meta-analyses made by Jobin et al (2019) of the 84 documents exposed none had any connection with any organization in South America, Africa, or the Middle East, showing that countries corresponding to more than half of the globe are being excluded from the debate on the ethical principles that should guide the future transformation of our society. Also, in Hagendorff's meta-analysis (2020), none of the recently published ethical guidelines originated in the regions mentioned. Another trend that Hagendorff points out is the lack of representation of women in the ethical debate on AI (in fact, in the entire AI industry). Excluding the ethical guidelines proposed by the AI Now[21] research institute, an organization deliberately led by women, the proportion of female and male authors is 31.3%.

The under-representation of women within the technology industry can have a real impact on the way such systems operate in society. According to Corbyn (2018), for every four jobs held in the silicon valley, only one is held by a woman. In the UK[22], women account for only 16% of the technology industry. As the use of AI becomes more and more widespread, in domains increasingly close to our privacy, such as personal assistants and chatbots, as most of the developers of these systems are men, prejudices and sexist biases are often imbued in our

[21] Retrivied from https://ainowinstitute.org/

[22] Wise: 2019 Workforce Statistics – One million women in STEM in the UK. Retrivied from https://www.wisecampaign.org.uk/statistics/2019-workforce-statistics-one-million-women-in-stem-in-the-uk/

technology. And cases, where IAs developed with the best intentions become producers of insanely sexist content, are not uncommon in today's literature.

In an interview with The Guardian (Balch, 2020), Eugenia Kuyda, developer of the personal chatbot Replika says: "For IAs to be friendly to us [...] the main qualities that should attract the public are inherently feminine, so it is really important to have women creating these products". And for this to happen, the current data sets we have available for training agents developed by machine learning are not at all appropriate. Sets of text, films, and images, carry with them various social biases. For example, Brown et al (2020) reported that their language model (GPT-3), trained with over 570 GB of text collected from the Internet, often associates adjectives related to appearance, such as "beautiful," "beautiful," and other words of more misogynistic content, with the pronoun "she," while male pronouns are associated with a much broader spectrum of adjectives.

In the same Guardian report, Lauren Kunze, chief executive of AI developer Pandorabots[23], says: "You simply cannot use unsupervised machine learning to train IAs for adult conversation, because systems trained in data sets like Twitter and Reddit become Hitler-loving sex robots". Gender issues end up becoming just another inequity in our current scenario of technological development. If we want to create truly "beneficial and friendly" AI, we need ethnic and cultural diversity to be guiding principles in this process. And this is not what is happening. The importance of this ethical principle is rarely cited in the literature.

Garcia (2019) also points out that virtually the entire Southern Hemisphere is under-represented in the debate on AI governance. As most developing countries do not yet have an AI industry capable of competing with more developed countries, the Global South is depending on the goodwill of other governments in a new colonial technological regime. According to Lee (2017):

> Unless they [developing countries] want to plunge their people into poverty, they will be forced to negotiate with the country that provides AI software - China or the United States - becoming essentially economic dependent on that country.

Garcia warns Global South leaders in a similar way:

[23] Retrivied from https://www.pandorabots.com/mitsuku/

> […]Leaders and scholars in the Global South cannot afford to remain idle while others make decisions and allow the militarization of AI to go deeper, unhindered, and increasingly deadly. Developing countries should not be relegated to the role of bystanders, technology users, or (even worse) victims. (Garcia, 2019, pp. 19-20).

For Green (2019), "good is not good enough", i.e., the limited definitions of what is "correct" or "morally justifiable" within areas responsible for technological development, e. g., computer science, software engineering, computer engineering, need to achieve an understanding of what the "social good" means. Although still very modest, there is a concern to increase diversity. Either by making developing countries and minority groups more active or seeking to impose new notions of equity, urgency, necessity, and historical restoration in the drafting of norms and guidelines for the tech industry. ÓhÉigeartaigh et al (2020) point out that the Academy has a key role to play in promoting greater intercultural cooperation on issues related to AI governance and ethics. The authors make the following recommendations:

1) Demand that research agendas involving ethics and governance of AI be the fruit of international and intercultural cooperation;
2) To translate existing literature, especially the main documents related to ethics and governance of AI, so that the language barrier does not interfere with intercultural participation;
3) Ensure that major research conferences on AI and AI ethics and governance are held in alternative continents and countries;
4) Establish exchange programs for students and young researchers, such as doctoral and post-doctoral researchers, to encourage intercultural collaboration among researchers at the beginning of their careers.

On the other hand, Mohamed et al (2020) defend measures of technological decolonization, i. e., the reappropriation of political, economic, and social structures by developing countries, to seek true emancipation and technological autonomy. Technological decolonization would occur in two ways: (1) by breaking the bond of dependency between states with more advanced technological tools, and states that can benefit from technological innovation; and (2) by dissolving the

ills left by the colonial relationship, a revitalization of underlying power structures, and all their structural biases. This form of decolonization allows for a rescue of the identity of the state and the marginalized population, placing on a critical view all the dominant forms of knowledge, values, norms, and beliefs characteristic of the current power structure. Thus we can characterize three fundamental aspects of a decolonization process, which should guide both steps just mentioned:

1) An affirmation of the identity of every culture, seeking that this historic rescue be a fundamental part of the formation of any Global governance structure;
2) Instead of a universalist view of ethics and global governance, the adoption of a pluralistic positioning. An additive and inclusive approach to knowledge generation will allow new forms of governance to emerge genuinely;
3) Peripheral and segregated populations have to be the orienting points of a critical analysis of global normative structures. If non-epistemic values are guiding our technological advance, we need to ensure that they are the values of all, and not only of those who are ahead in the technological "race" for the development of more proficient AI.

VI. Conclusion

Currently, we see several issues disregarded from the debate related to AI ethics. While technical problems related to machine learning already have a strong literary background, social issues like income inequality, technological unemployment, humanitarian violations, and the total lack of diversity in the AI and tech industry, should be given an equal amount of attention. We need to remember that are prissily these issues that are drastically affecting the lives of millions of individuals today, and not our lack of understanding of what these systems are doing or how to better protect our personal information. Humanitarian rights, the sovereignty of a country, the right to decent labor conditions and payment, and the respect for all diverse expressions of humanity, are far more important problems to be addressed. Like critical theory, AI ethics must focus on highlighting the neglected aspects of our society and its relation to the technological industry, challenging its power structures, so that the promise of

— beneficial AI for all — can be fulfilled. Not just as an ideal for the future of humanity, but for the present people to.

Acknowledgments

We would like to thank the Academic Excellence Program (PROEX) of CAPES Foundation (Coordination for the Improvement of Higher Education Personnel), and the Graduate Program in Philosophy of the Pontifical Catholic University of Rio Grande do Sul, Brazil.

Funding

Research supported by the Academic Excellence Program (PROEX) of CAPES Foundation (Coordination for the Improvement of Higher Education Personnel), Brazil.

References

Agüera y Arcas, B., Todorov, A., & Mitchell, M. (2018). Do algorithms reveal sexual orientation or just expose our stereotypes? *Medium*. https://link.medium.com/GO7FJgFgM1.

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. ArXiv. https://arxiv.org/abs/1606.06565

Antunes, R . (2009). Século XXI: nova era da precarização estrutural do trabalho? Infoproletários: degradação real do trabalho virtual. https://projetoaletheia.files.wordpress.com/2014/08/seculo-xxi-era-da-precarizacao.pdf

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.F., & Rahwan, I. (2018). The Moral Machine Experiment. *Nature*. 563. doi: 10.1038/s41586-018-0637-6

Balch, O. (2020). AI and me: friendship chatbots are on the rise, but is there a gendered design flaw? *The Guardian*. https://www.theguardian.com/careers/2020/may/07/ai-and-me-friendship-chatbots-are-on-the-rise-but-is-there-a-gendered-design-flaw

Baudrillard, J. (1994). *Simulacra and Simulation*. Michigan, EUA, University of Michigan Press.

Boddington, P. (2017). *Towards a code of ethics for artifcial intelligence*. Springer International Publishing. DOI: 10.1007/978-3-319-60648-4

Brief, A.P., Dukerich, J.M., Brown, P.R., Brett, J.F. (1996). What's wrong with the treadway commission report? Experimental analyses of the effects of personal values and codes of conduct on fraudulent financial reporting. *Journal of Business Ethics*, 15(2), 183–198. https://doi.org/10.1007/BF00705586

Brown, T., Mann, B., Ryder, N. el al . (2020). Language Models are Few-Shot Learners. OpenAI. https://arxiv.org/pdf/2005.14165.pdf

Bueter, A. (2015). The irreducibility of value-freedom to theory assessment. *Studies in History and Philosophy of Science Part A*, 49, 18–26.

Burt, R. (2010). Tantalum - a Rare Metal in Abundance? T.I.C. Bulletin.

Butler, S. (1863, June 13). Darwin Among the Machines. The Press, Christchurch, New Zealand.
https://web.archive.org/web/20060524131242/http://www.nzetc.org/tm/scholarly/tei-ButFir-t1-g1-t1-g1-t4-body.html

Calhoun, C. (1995). Critical Social Theory: Culture, History, and the Challenge of Difference. Wiley-Blackwell.

Calo, R. (2017). Artifcial intelligence policy: a primer and roadmap. *SSRN Journal*, 399–435. http://dx.doi.org/10.2139/ssrn.3015350

Carman, M., & Rosman, B. (2020). Applying a principle of explicability to AI research in Africa: should we do it? *Ethics Inf Technol*. https://doi.org/10.1007/s10676-020-09534-2

Cassidy, J. (2009). *Dot.con: How America Lost Its Mind and Its Money in the Internet Era*. Harper Collins.

Chalmers, D. (2010). The singularity: A philosophical analysis. *Journal of Consciousness Studies*, 17 (9-10), 9–10.

Cleek, M.A., & Leonard, S.L. (1998). Can corporate codes of ethics influence behavior? *Journal of Business Ethics*, 17(6), 619–630. https://doi.org/10.1023/A:1017969921581

Corbyn, Z. (2018). Why sexism is rife in Silicon Valley. *The Guardian.* https://www.theguardian.com/world/2018/mar/17/sexual-harassment-silicon-valley-emily-chang-brotopia-interview

Costanza-Chock, S. (2018). Design justice, AI, and escape from the matrix of domination. *Journal of Design and Science*. doi:10.21428/96c8d426

DiSalvo, C. (2012). *Adversarial design (design thinking, design theory)*. Cambridge: MIT Press.

Domenici, T. (2020). Após uso de robôs, Laureate agora demite professores de EAD. Publica, Agência de Jornalismo Investigativo. https://apublica.org/2020/05/apos-uso-de-robos-laureate-agora-demite-professores-de-ead/

Dorner, U., Franken, G., Liedtke, M., & Sievers, H. (2012). Artisanal and Small-Scale Mining (ASM). *POLINARES*. Disponível em: https://www.polinares.eu/artisanal-and-small-scale-mining-asm

Douglas, H. (2007). *Rejecting the ideal of value-free science*. In Kincaid, H., Dupre, J., Wylie, A. (Eds.) ´ *Value-free science: ideals and illusions*? chap 6 (pp. 120–141). Oxford: Oxford university press.

Ekstrand, M.D., Joshaghani, R., & Mehrpouyan, H. (2018). Privacy for all: Ensuring fair and equitable privacy protections. Proceedings of the 1st Conference on Fairness, Accountability and Transparency (pp. 1–13).

Elliott, K.C., & McKaughan, D.J. (2014). Nonepistemic values and the multiple goals of science. *Philosophy of Science*, 81(1), 1–21.

Eufrásio, J., & Goulart, G. (22 de Setembro, 2020). Entregadores por app têm entre 19 e 40 anos e ganham, em média, R$ 2,1 mil. Correio Braziliense. https://www.correiobraziliense.com.br/app/noticia/cidades/2020/07/20/interna_cidadesdf,873540/entregadores-por-app-tem-entre-19-e-40-anos-e-ganham-r-2-1-mil.shtml#:~:text=O%20que%20dizem%20as%20empresas,dar%20o%20recorte%20na%20capital

Frey, C., & Osborne, M. (2013). The Future of Employment: How Susceptible Are Jobs to Computerisation? Technical Report, Oxford Martin School, University of Oxford, Oxford, UK. https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf

Friedman, B., Kahn, P.H., Borning, A., Huldtgren, A. (2013). *Value sensitive design and information systems*. In *Early engagement and new technologies: opening up the laboratory* (pp. 55–95). Berlin: Springer.

Garcia, E.V. (2019). The Militarization of Artificial Intelligence: A Wake-Up Call for the Global South, *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.3452323

Gest, J. (2016). *The new minority: White working class politics in an age of immigration and inequality*. Oxford, UK, Oxford University Press.

Goldsmith, J., & Burton, E. (2017). Why teaching ethics to AI practitioners is important. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (pp. 4863–4840). https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14271/13992

Green, B. (2019). "Good" isn't good enough. In *NeurIPS workshop on AI for social good*. https://www.benzevgreen.com/wp-content/uploads/2019/11/19-ai4sg.pdf

Greene, D., Hofman, A.L., & Stark, L. (2019). Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artifcial intelligence and machine learning. In Hawaii international conference on system sciences (pp. 1–10). doi: 10.24251/HICSS.2019.258

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 30, 99–120. https://doi.org/10.1007/s11023-020-09526-7.

Hao, K. (2020, May 14). The pandemic is emptying call centers. AI chatbots are swooping in. MIT Technology Review. https://www.technologyreview.com/2020/05/14/1001716/ai-chatbots-take-call-center-jobs-during-coronavirus-pandemic/

Harris, M. (2014). Amazon's Mechanical Turk workers protest: “I am a human being, not an algorithm”. The Guardian. http://bit.ly/2EcZvMS

Irani, L. (2016). The hidden faces of automation. *XRDS: Crossroads, The ACM Magazine for Students*, 23(2), 34–37. https://doi.org/10.1145/3014390

Jameson, F. (1991). *Postmodernism, Or, The Cultural Logic of Late Capitalism*. North Carolina, USA, Duke University Press.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nat Mach Intell*, 1, 389–399. https://doi.org/10.1038/s42256-019-0088-2

Jurić, M., Šandić, A., & Brcic, M. (2020). AI safety: state of the field through quantitative lens. 43rd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). https://arxiv.org/ftp/arxiv/papers/2002/2002.05671.pdf

Kanth, D.R. (2019). India boycotts ‘Osaka Track’ at G20 summit. *Live Mint*. https://www.livemint.com/news/world/india-boycotts-osaka-track-at-g20-summit-1561897592466.html

Laudan, L. (1968). Theories of scientific method from Plato to mach: a bibliographical review. *History of science*, 7(1), 1–63.

Lee, K. F. (2017). The real threat of artificial intelligence. *The New York Times*. https://www.nytimes.com/2017/06/24/opinion/sunday/artificial-intelligence-economic-inequality.html

Lere, J.C., & Gaumnitz, B.R. (2003). The Impact of Codes of Ethics on Decision Making: Some Insights from Information Economics. *Journal of Business Ethics*, 48, 365–379. https://doi.org/10.1023/B:BUSI.0000005747.37500.c8

Levin, T. S. (2017). Face-reading AI will be able to detect your politics and IQ, professor says. *The Guardian*. https://www.theguardian.com/technology/2017/sep/12/artificial-intelligence-face-recognition-michal-kosinski

Maxmen, A. (2018). Self-driving car dilemmas reveal that moral choices are not universal. *Nature*, 562 (7728), 469–470. doi:10.1038/d41586-018-07135-0

McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM’s code of ethics change ethical decision making in software development? ESEC/FSE 2018: Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (pp. 729–733). https://doi.org/10.1145/3236024.3264833

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nat Mach Intell*, 1, 501–507. https://doi.org/10.1038/s42256-019-0114-4

Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. In FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 279–288). https://doi.org/10.1145/3287560.3287574

Mohamed, S., Png, MT., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*. https://doi.org/10.1007/s13347-020-00405-8

Montague, D. (2002). Stolen Goods: Coltan and Conflict in the Democratic Republic of Congo. *SAIS Review*, 22(1), 103–118. doi:10.1353/sais.2002.0016.

Mulhall, D. (2002). *Our molecular future: how nanotechnology, robotics, genetics, and artificial intelligence will transform our world*. Amherst, NY, Prometheus Books.

Nest, M. (2011). *Coltan*. Cambridge: Polity Press.

Nissenbaum, H. (2001). How computer systems embody values. *Computer*, 34(3), 120–119.

Nunes, P. (2019). EXCLUSIVO: levantamento revela que 90,5% dos presos por monitoramento facial no Brasil são negros. *The Intercept Brasil*. https://theintercept.com/2019/11/21/presos-monitoramento-facial-brasil-negros/

Nyabola, N. (2018). *Digital democracy, analogue politics: how the Internet era is transforming politics in Kenya*. Zed Books Ltd.

O'Keefe, C., Cihon, P., Flynn, C., Garfinkel, B., Leung, J., & Dafoe, A. (2020). The Windfall Clause: Distributing the Benefits of AI. Centre for the Governance of AI Research Report. Future of Humanity Institute, University of Oxford. https://www.fhi.ox.ac.uk/windfallclause/

ÓhÉigeartaigh, S.S., Whittlestone, J., Liu, Y., Zeng, Y., & Liu, Z. (2020).Overcoming Barriers to Cross-cultural Cooperation in AI Ethics and Governance. *Philos. Technol*. https://doi.org/10.1007/s13347-020-00402-x

Osborn, M., Day, R., Komesaroff, P., & Mant, A. (2009). Do ethical Guidelines make a difference to decision-making?. *Internal medicine journal*, 39(12), 800–805. https://doi.org/10.1111/j.1445-5994.2009.01954.x

Rességuier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society*, 1-5. https://doi.org/10.1177/2053951720942541

Rosa, H. (2010). *High-Speed Society: Social Acceleration, Power, and Modernity*. USA, Pennsylvania State University Press.

Russel, S. (2019). Human Compatible: Artificial Intelligence and the Problem of Control. London, UK. Penguin.

Russell, S., Dewey, D., & Tegmark, M. (2015). An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence. Open Letter. Signed by 8,600 people. https://futureoflife.org/data/documents/research_priorities.pdf

Silberman, M.S., Tomlinson, B., LaPlante, R., Ross, J., Irani, L., & Zaldivar, A. (2018). Responsible research with crowds. *Communications of the ACM*, 61(3), 39–41. doi: 10.1145/3180492

Soares, N. (2016). Value Learning Problem. In *Ethics for Artificial Intelligence Workshop, 25th International Joint Conference on Artificial Intelligence* (IJCAI-2016) New York, NY, 9–15. https://intelligence.org/files/ValueLearningProblem.pdf.

Soares, N., Fallenstein, B., Yudkowsky, E., & Armstrong, S. (2015). Corrigibility. In Artificial Intelligence and Ethics, ed. T. Walsh, AAAI Technical Report WS-15-02. Palo Alto, CA: AAAI Press.

Sutherland, E. (2011). Coltan, the Congo and your cell phone. *SSRN*. Disponível em: http://dx.doi.org/10.2139/ssrn.1752822

Tally, R. (2009). *Radical Alternatives: The Persistence of Utopia in the Postmodern*. In *New Essays on the Frankfurt School of Critical Theory*, A. J. Drake (Ed.), Newcastle: Cambridge Scholars Publishing.

Tegera, A., MIkolo, S., & Johnson, D. (2002). The Coltan Phenomenon: How a rare mineral has changed the life of the population of war-torn North Kivu province in the East of the Democratic Republic of Congo', Pole Institute. http://archive.niza.nl/docs/200212171307317066.pdf

Turner, A., Smith, L., Shah, R., & Tadepalli, P. (2020). Optimal Farsighted Agents Tend to Seek Power. ArXiv. https://arxiv.org/pdf/1912.01683.pdf

Usanov, A., De Ridder, M., Auping, W., Lingemann, S., Espinoza, L., Ericsson, M., & Liedtke, M. (2013). *Coltan, Congo & Conflict: POLINARES CASE STUDY* (pp. 15-28, Rep.). Hague Centre for Strategic Studies. https://hcss.nl/sites/default/files/files/reports/HCSS_21_05_13_Coltan_Congo_Conflict_web.pdf

Virilio, P. (1997). *Open Sky*. London, UK, Verso.

Wang, Y., & Kosinski, M. (2017). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. https://doi.org/10.1037/pspa0000098