***24th Session of the* WORKING GROUP OF EXPERTS ON PEOPLE OF AFRICAN DESCENT**

**Data for Racial Justice**

*Geneva, 25-29 March 2019, Palais des Nations, Room XXI*

Questions Raised by Racial Statistics in the Age of Artificial Intelligence (AI), with Emphasis on the Justice System

**Pastor Murillo**

CERD Vice-Chair

This session has highlighted the crucial role played by statistics in providing *recognition,* ***justice*** *and development for* people of African descent. For our purposes, I would like to *explore* a subject which I believe to be a growing threat, not just to those historically discriminated against, such as people of African descent and women, but also to society at large.

I’m referring to *racial bias present in some of the algorithms used by artificial intelligence.* Algorithms are understood to be "a set of unambiguous, finite, deterministic (“not by chance”) rules to be followed in problem solving operations”. They are like the “heart” of robots, drones, self-driving vehicles and other intelligent machines[[1]](#footnote-1).

There is a growing consensus about the social impact of AI, in spite of its benefits. Within two decades almost 50% of all jobs around the world, including those considered “white collar”, will have been automated. This represents the loss of millions of jobs to automation.

The social and economic impact of the digital revolution is already reaching the political arena as we have witnessed in several recent elections, while some choose to train the spotlight on the current migration crisis, with ever-graver consequences as demonstrated, once again, in the painful examples of New Zealand and Italy.

In this context, that is, *questions raised by racial statistics in the age of artificial intelligence;* in the framework of the justice system, the example of the United States is significant for a number of convergingfactors: it is at the forefront of AI; after Brazil it is home to the second largest population of people of African descent in the Americas, currently estimated at around 45 million; in the USA – as in the rest of the Americas – there is widespread structural, systemic racism and racial discrimination towards this population, as illustrated by the *disproportionate number of people of African descent that are in prison.*

Even though the United States has only 4% of the world population, it generally makes up 25% of the world’s prison population - approximately 2 million people. Although only about 14% of Americans are of African descent they make up over 40% of its inmates. This disproportionate number of people of African descent in the United States prison system is similarly reflected on death row.

This dramatic situation speaks volumes about the racism and racial discrimination suffered by people of African descent in the United States, yet it is being further exacerbated by *racial bias* linked to *artificial intelligence* and specifically to *algorithms* used with increased frequency *around the world*.

In 2016, for example, several *ProPublica* journalists detected that one of the algorithms used in the American justice system was not impartial and was detrimental to people of African descent. The police hands out a questionnaire to all detainees and their answers are entered into a computer. An algorithm called *Compas* (US Correctional Offender Management Profiling for Alternative Sanctions) uses all that information to predict the likelihood of a person reoffending by assigning them a score.

This score is provided to the judge to assist in making “more objective” decisions based on data when handing down the verdict. The results are clear: those of African descent are sentenced to longer prison terms than whites[[2]](#footnote-2). In essence:

 “The information requested on the forms includes details such as **“your criminal record, whether any member of your family has committed a crime or been arrested**, whether you live in a dangerous neighbourhood, whether you have any friends who are members of a gang, your work and academic records.

(…) Additionally, there are questions that could be described as probing a criminal mind-set, such as: Do you agree or disagree with the following statement: A hungry person has a right to steal. Each response is scored from 1 to 10. This produces a risk assessment score which **determines whether somebody can be released on bail, should be sent to prison or be given a different sentence.** Once jailed, the algorithm also determines whether they should be released on parole. Compas and other such programs are used throughout the United States.”

The ProPublica study analysed the scores of 7.000 people arrested in the state of Florida over two years and the results were shocking. When comparing blacks and whites having the same background, age, gender, criminal record and criminal future (the likelihood of committing one, two or no crimes), **the black defendant has a 45% greater chance receiving a higher risk rating**. [[3]](#footnote-3)

Even though the algorithm does not request the race information of the suspect, the complete set of questions clearly allows us to conclude that, given the social circumstances associated to racism and racial segregation in the United Sates, people of African descent are at a disadvantage. Furthermore, the data that generates the scores is kept secret. The use of these algorithms in the United States justice system is normally provided by private security companies.

These machines are like a “…*black box* keeping secrets even from their own developers, who are incapable of understanding the route taken to a specific conclusion… when you are put on trial your verdict gives an explanation: the problem is that these algorithms are unclear. You are brought before an oracle that will hand down its decision” [[4]](#footnote-4) says an expert.

In addition to the controversy generated by the ProPublica results, experts agree that *algorithms “replicate the inequalities of the real world”.* There are many examples of this: *“experiments by Carnegie Mellon[[5]](#footnote-5) showed that significantly fewer women than men were shown online ads for better paying jobs in Google because the programs used by some hiring departments preferred names used by men”.*

Some experts believe that “…the problem stems from the data rather than from developer prejudice. For example, image recognition and classification machines *learn* from what they find on large Internet image banks. A Nature report discovered that 45 per cent of the photos in one of these banks are from the United States and depict mainly white people, despite the fact that the United States represents only 4 per cent of the world population. Whereas China and India, which represent one third of all people on the planet, constitute only 3 per cent of the images in that same bank”.

As we know, *facial recognition* systems are ever more popular and could make racial bias in algorithms even more pervasive. For example, *“Jacky Alcine, a young man from Brooklyn, discovered that Google Photos had labelled as “gorillas” a photo himself with a female friend.* The company corrected the situation due to the complaint”.

It is expected that *facial recognition cameras* will soon be part of everyday life and will determine the way we’re treated when entering shops, hotels or other places that serve the public. This service will not necessarily be at the hand of a human, given the rate at which we are being replaced by robots, as seen in a Japanese hotel chain where guests can spend the night without ever interacting with a human from check-in to check-out.

Similarly, in the United States, where the faces of 50% of the population can be found in the police databases, an algorithm was recently implemented to try to identify gays.

There are countless cases that expose the risks of automation. “In a trial run in Tempe, Arizona, *a self-driving Uber* ran over a woman who was carelessly crossing the street. *The reason given was that the vehicle was unable to recognize her as a human being.”*

Self-driving cars are set to be one of the most widespread uses for artificial intelligence and one where its impact will be felt the most. By 2022 it is expected that they will already be in use in the state of Florida, USA.

Concerns around the *ethics and governance of artificial intelligence* is already on the agenda of the world’s most prestigious universities and of the self-same companies leading the digital revolution, as they start to understand that programmers need to consider the impact and the improper use of their creations. [[6]](#footnote-6)

According to professor Meredith Broussard of NYU, “…computers are no more objective than people nor more impartial simply because they operate on the basis of questions and answers managed by mathematical processes. No technological innovation will ever distance us from the essential problems inherent in human nature for the simple reason that the design is human”, she says.

While recognising that artificial intelligence is prejudiced, *techno-optimists* believe that this can be prevented and even corrected. Dr. Gemma Galdon believes that *“an algorithm learns from what it sees, but not fixing it will make it worse”*, even though she admits that the bias is present in real life.

# Professor Carlos Gómez Abajo also admits that some algorithms have proven to be sexist or racist, as can be seen in many search engines that are not designed to compensate for human error, which is why he believes that “rectifying data after the fact or introducing limitations helps combat discrimination”.

At Pompeu Fabra University, Barcelona, Professor Carlos Castillo is working on an algorithm that retroactively corrects for discrimination based on gender, origin or physical appearance in *online* searches.

So-called *disparate impac*t must also be taken into account. This is a concept drawn from US Labour Legislation and is the different effect that one same algorithm has on different groups of people. This is different from “disparate treatment” in that this is intentional and direct, where *disparate impact* is indirect… The method consists of “correcting” the scores of the protected group (being discriminated against) retroactively in order to redistribute them in the same way as the unprotected group. [[7]](#footnote-7)

Meanwhile, in Germany, a system is being used that “…includes mathematical limitations during the algorithm’s training or learning period (when it is supplied with previous searches as learning examples), so that when it is used it won’t take into account sensitive data such as gender, race, physical appearance or origin”. However, it is worth restating that algorithms are ever more self-governing, blurring the lines between reality and fiction.

Nobody disputes that machines do learn. The question lies in how ready we are to bear the ethical burdens that this implies. Alongside scientific goals such as cancer diagnosis and 5G-tech remote control surgery, lurk less worthy and even frivolous goals.

The struggle against racial and gender bias in artificial intelligence must mobilize us all. *Microsoft* was recently obliged to retire a robot that defended the holocaust and glorified Hitler. An algorithm was identified in Facebook that was inclined to ignore black skinned users as a target for advertising campaigns and algorithms for certain financial institutions for credit ratings prefer white males aged 30 to 50 because of their better payment history. The self-same profile, in fact, of the algorithm’s creators. Gender inequality is also blatant “only 22% of the world’s AI professionals (LinkedIn) are women, compared to 78% men. The future is not bright in this field given the huge gender gap in artificial intelligence.” [[8]](#footnote-8)

**Distinguished Delegates:**

**Ladies and Gentlemen:**

During its next session, CERD will debate a draft general recommendation on **racial profiling**, which will examine ethical aspects of cyber security and will seek to provide guidelines to prevent and sanction racial bias in artificial intelligence. <https://www.ohchr.org/EN/HRBodies/CERD/Pages/GC36.aspx>

With this presentation I invite you to reflect upon these issues and submit contributions on whatever aspects you feel are appropriate such as: the criteria that companies should apply in preventing racial bias in the programming and use of algorithms; notable examples of artificial intelligence related racial discrimination; research and best practice in prevention or response to racial bias stemming from artificial intelligence; ethical aspects of machine learning and relevant criteria to ensure understanding and transparency in the programming process, as pertains to issues of race.

**Thank you**

1. <https://dcodingames.com/que-es-un-algoritmo/> [↑](#footnote-ref-1)
2. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [↑](#footnote-ref-2)
3. <https://www.bbc.com/mundo/noticias-37679463> [↑](#footnote-ref-3)
4. [https://elpais.com](https://elpais.com/) 23 November 2018. [↑](#footnote-ref-4)
5. <https://www.cmu.edu/news/stories/archives/2015/july/online-ads-research.html> [↑](#footnote-ref-5)
6. The reknowned Professor *Joichi Ito of MIT* gives a class on this subject [↑](#footnote-ref-6)
7. Is the technique designed by the University of Utah Informatics Professor Suresh’s team. [↑](#footnote-ref-7)
8. World Economic Forum Report 2018 - <https://www.weforum.org/reports/the-global-gender-gap-report-2018> [↑](#footnote-ref-8)