

The dangers of algorithmic bias in artificial intelligence

مخاطر التحيز الخوارزمي في الذكاء الاصطناعي

Timedjehdine Nadia¹

¹ Akli Mohand Oulhadj University Bouira (Algeria),

n.timedjehdine@univ-bouira.dz

Multidisciplinary Laboratory in Human, Environmental, and Social Sciences

Received:18/11/2024 Accepted: 13/01/2025 Published:01/02/2025

Abstract:

Algorithms serve as shorthand for describing decision-making processes that are the basis of technology, playing a crucial role in various aspects of our lives. They are based on formulas that people enter into a database in order to carry out particular activities. It has been noted, meanwhile, that algorithms do not always strive for inclusivity in all respects; they may restrict some user groups and exclude others based on racial, gender, and other characteristics. Which makes Algorithmic bias standing as a problem.

Therefore, this research paper will focus on the problem of algorithmic racism, which arose when data system programs used machine learning and artificial intelligence-based technologies to produce racial inequities in a variety of aspects of life. Therefore, it is essential for those responsible for software engineering to take necessary precautions and find solutions for users to achieve equality and inclusion, ensuring that all users worldwide are encompassed.

Keywords: Algorithms, Algorithmic bias, Artificial intelligence, the database, Machine learning.

ملخص:

الخوارزميات تعتبر طريقة مختصرة لوصف عمليات اتخاذ القرارات التي تمثل أساس التكنولوجيا وتشكل أهمية بالغة لحياتنا في مختلف المجالات، إذ نجد لها قائمة على تعليمات وصيغ يدخلها الفرد في قاعدة البيانات لكي تقوم بتنفيذ مهام محددة، لكن ما تمت ملاحظته هو أنها لم تسعى إلى أن تكون شاملة في جميع جوانبها فقد تقييد المنتجات البرمجية مجموعات من

Timedjehdine Nadia

المستخدمين فقط وتستثني الآخرين منهم انطلاقاً من عوامل قد تكون قائمة على العرق، الجنس وغيرها مما أثار مشكلة التحيز الخوارزمي.

لذلك فإنّ هذه الورقة البحثية ستسلط الضوء على الخوارزميات العنصرية التي ظهرت عندما قامت برامج أنظمة محركات البيانات بالتمييز بين الأفراد باستخدام التكنولوجيا القائمة على التعلم الآلي والذكاء الاصطناعي في العديد من جوانب الحياة اليومية، لذلك لا بد من المسؤولين عن هندسة البرمجيات اتخاذ الاحتياطات اللازمة وإيجاد حلول لكي نحقق المساواة والتنوع لنشمل جميع المستخدمين حول العالم.

كلمات مفتاحية: الخوارزمية، التحيز الخوارزمي، الذكاء الاصطناعي، قاعدة البيانات، التعلم الآلي.

Corresponding author: Timedjehdine Nadia

1. Introduction

Discussions about equality, diversity, and inclusion are progressively growing. However, this discussion is moving slowly in the field of software engineering. It may sound paradoxical, but the rise of algorithmic racism shows that the industry that develops innovative software solutions for billions of users globally does not always represent the diversity of the population it serves.

When data system programs generate racial inequities based on racial, gender, cultural, and religious affiliations, this is known as algorithmic racism. This causes problems in a variety of situations, which have an impact on people by making them feel excluded, denied their rights, and having their dignity compromised. Finding concrete ways to stop the phenomenon's spread is necessary to lessen such incidents.

What are the threats posed by algorithmic racism to individual security? What proposed solutions can help minimise the spread of algorithmic bias?

2. Algorithms in Artificial Intelligence

Artificial Intelligence (AI) is a topic that has been widely discussed, and the term "Artificial Intelligence" was coined by the American computer scientist John McCarthy when he began research on the subject in 1956. He convinced Marvin Minsky, along with Claude E. Shannon, the inventor of information theory, and Nathaniel Rochester, the designer of the first

The dangers of algorithmic bias in artificial intelligence

computer at IBM, to join a program with the goal of hypothesising that all aspects of learning and intelligence could, in theory, be accurately described in a way that would allow machines to mimic them. This brought together top experts to start a new field of science that involved the study of intelligent machines.

Thus, artificial intelligence simulates human capabilities on electronic chips and circuits to enable machines to perform tasks and solve problems similar to those carried out by humans. John McCarthy notes that a large portion of AI research focuses on figuring out how to encode and input the difficulties that the outside world presents to machines. AI researchers have the freedom to use methods that may not be observable in humans or involve a greater level of computation than humans are capable of performing. All these inputs are encoded using mathematical equations and computational symbols known as algorithms.

At first, the objectives of artificial intelligence were to perform extremely specialised tasks like speech recognition, computer vision, pattern recognition in data, and prediction. Narrow AI is the term used to describe this concentration on particular intelligent tasks. Later on, though, its goal shifted to developing Strong AI, which aims to produce computers with cognitive capacities that are identical to those of humans. In order to create intelligent machines that are more cognitively capable than humans, researchers are now pursuing Strong AI as their ultimate goal. (D.Luxton, 2016, p. 3)

The concept of the algorithm has evolved over the centuries. It was Abu-Abdullah Muhammed ibn-Musa Al-Khawarizmi who first introduced the term algorithm, based on step-by-step reliable procedures for solving equations. Later, the concept was further developed by Alonzo Church and Alan Turing, who expanded it to include the notion of computability, formalizing the idea of an algorithm. It became understood as a finite sequence of precise instructions that can be executed on computing systems. These instructions do not only process numbers through fixed mathematical models. (Osoba, Osonde ; Welser, Wiliam ;, 2017, p. 5)

Algorithms represent a set of instructions created by an individual to solve a problem or a class of problems. In order to generate outputs that

Timedjehdine Nadia

address a particular issue, they also rely on data and analysis based on what is input and programmed into them. Therefore, algorithmic bias refers to the process of encoding and interpreting input data, followed by the generation of outputs—often without conscious awareness—that may overlook equity and justice. (Ronnie de Souza , Santos; Fernando de Lima, Luiz ; Magalhaes, CESAR;, 2023, p. 1).

An algorithm is a fundamental and precise computational formula developed to initiate programmed computational steps in order to perform a specific task. These algorithms, embedded within software, produce outputs based on the input data they receive. It is important to note that algorithms represent the backbone of all software systems that govern various institutions and companies, meaning they exist wherever technology is present. Humans, especially software developers who are trained in various programming languages, program them. (ADIB-MOGHADDAM, 2023, p. 23)

We can thus say that an algorithm is a computational rule upon which all technological devices rely in their computational framework. It works by receiving a set of encoded inputs, then translating and representing them in order to send us outputs that solve a specific and defined problem.

In this context, we can also refer to the concept of machine learning, which is based on algorithms. Machine learning derives its efficiencies from various scientific fields, including data science, statistics, and optimization. It uses probability-driven mathematical models and statistics to make predictions, leveraging available data to infer insights about unavailable data.

Algorithms have become ubiquitous and serve as shorthand for describing statistical decision-making processes. Whether in the economic sector (such as in insurance companies or marketing), in healthcare and learning devices, in weather prediction, or in equipment used by athletes, algorithms are now embedded in nearly everything.

The dangers of algorithmic bias in artificial intelligence

3. Algorithmic Racism

The concept of algorithmic racism has gained widespread attention, especially after numerous events revealed its presence. People from various backgrounds have suffered from behaviours and outcomes resulting from differences and discrimination based on race, skin colour, gender, and other factors. So, we ask: What do we mean by algorithmic racism? And how has it infiltrated artificial intelligence systems and come to dominate us?

Algorithmic Racism is the term used to describe the behaviour of technological solutions that restrict users based on their race, gender, or other individual characteristics. Algorithmic racism occurs when data-driven systems generate racial disparities, leading to flaws in various contexts, especially since algorithms are based on the concept of artificial intelligence. (M.Blank, Rebecca; Dabady, Marilyn; F.Citro, Constance;, 2004, p. 8)

Algorithmic Racism is a concept related to the fairness of software, referring to outcomes that arise from racial bias in technology. In other words, it occurs when technology produces unequal and unfair results for individuals from a particular racial group.

A standardised method has been identified to explore issues of racial disparity and discrimination in the real world through:

1. Developing a regression model that includes a variable for race and variables for other relevant observed characteristics.
2. The impact of the previous variable on the outcome variation is defined as discrimination.
3. Bias in the excluded variables is observed. Whenever the dataset contains only a limited number of characteristics that may be reasonable factors in the process under study, sample selection bias will occur. This is because, by excluding individuals whose characteristics differ from those of the individuals represented in the data, the research will be biased towards the represented group at the expense of others. (M.Blank, Rebecca; Dabady, Marilyn; F.Citro, Constance;, 2004, p. 8)

It is true that the algorithm works properly in freeing the cognitive capacity of the decision-maker to engage in other important deliberations. However, the opacity of algorithms makes it difficult to assess their

Timedjehdine Nadia

accuracy, evaluate risks, and assess fairness in social applications, as it can obscure the causal understanding behind decisions. These issues may be less harmful if the algorithms were almost infallible. (Osoba, Osonde ; Welser, Wiliam ;, 2017, p. 23)

Algorithms are present in all aspects of our daily lives because technology has overwhelmed us. We find that government agencies rely on algorithms for tax audits and social welfare checks, as well as the police through DNA analysis and other forensic examinations. Therefore, tampering with their inputs and racial bias is dangerous.

Therefore, we can assert that algorithms, due to their programming and the data (inputs) they receive through computational processes, only produce outputs when they rely on their inputs. They cannot independently identify or distinguish one group from another. What has been observed is a set of classifications based on race, such as black people versus white people, religious beliefs, cultures, and so on. Therefore, caution is necessary because these issues are not easily overcome and will create outcomes and dimensions that are difficult to resolve.

4. Risks of Algorithmic Bias

The concept of algorithmic racism has spread globally, especially in countries that heavily rely on technology. This raises the question: What are the potential risks of biased algorithms that affect individuals worldwide and threaten their safety?

Artificial intelligence raises questions about racism and gender bias. The development and control of AI are currently carried out by a small group of people, most of whom are white, heterosexual men. This leads to an incomplete representation of other populations, such as Black people. Data is directly documented and programmed into computational systems, which is why there has been a call for the implementation of software justice.

Examples of ethical dilemmas related to artificial intelligence include AI tools that perpetuate discrimination in housing, such as in tenant selection and mortgage qualification. Another example is the AI hiring tool from Amazon, which was shut down after one year of use due to bias. There are also AI algorithms that create risk scores to predict which patients are likely to develop certain diseases, such as skin cancer, and these algorithms

The dangers of algorithmic bias in artificial intelligence
discriminate against racial minorities. (J.Salazar, 2023)

Software justice is an emerging concept in software engineering that focuses on exploring strategies and techniques to develop fairer and more inclusive data-driven systems. (Ronnie de Souza , Santos; Fernando de Lima, Luiz ; Magalhaes, CESAR;, 2023, p. 1)

For example, for Black users, many data-driven programs result in the following: higher car insurance costs, lower credit scores, more expensive or inaccessible mortgage loans, assignment to less prestigious positions, rejection of life-saving healthcare, reliance on a more punitive criminal justice system, and Black communities being programmed as more prone to committing crimes, which leads to excessive police targeting. (Ronnie de Souza , Santos; Fernando de Lima, Luiz ; Magalhaes, CESAR;, 2023, p. 2)

Therefore, all these disadvantages must be taken into account. The software industry needs to take more responsibility for the impacts of algorithmic racism, which misguides people's lives, and begin initiating discussions about the risks of biased algorithms on our society.

In 1942, Gordon Allport, a psychologist, stated:

"A fatal non sequitur occurs in the reasoning that if 80% of the delinquents who come have an 80% chance of becoming recidivists. The truth of the matter is that this delinquent has either 100% certainty of becoming a repeater or 100% certainty of going straight." (SLOBOGIN, 2021, p. 43)

Some lawyers also adopt this view. In 1944, Judge "Quinn" from the Minnesota Supreme Court wrote that it is wrong to limit someone based on their previous behaviour, but rather on statistical evidence related to the behaviour of other individuals. In line with these views, courts have excluded statistical guessing. (SLOBOGIN, 2021, p. 43 44)

Critics of risk algorithms in the criminal justice system have often complained about the unauthorised uses of RAIs (Risk Assessment Instruments), which are relied upon to measure criminal tendencies and determine sentences for offenders. They have criticised what they perceive as injustices in using these tools to assess risks, even if RAIs meet the most stringent requirements for relevance and validity. However, their use should be avoided or minimised, as they exacerbate racial biases and other forms of

Timedjehdine Nadia

discrimination, justify the deprivation of liberty using irrelevant factors unrelated to culpability, and allow for detention based on inhumane and non-individualised determinations. (SLOBOGIN, 2021, p. 86)

Another city decided to use algorithmic methods to guide law enforcement activities, with the justification being that predictive police algorithms relied on multivariable equations rather than human decisions. However, other reports showed systematic biases in the criminal risk assessment algorithms used during sentencing hearings across the United States. (Osoba, Osonde ; Welser, Wiliam ;, 2017, p. 4)

A lot of behaviours have been observed that demonstrate the bias of algorithms in artificial intelligence. For example, the Beauty AI contest was held under the title:

"WELCOME TO THE FIRST INTERNATIONAL BEAUTY CONTEST JUDGED BY ARTIFICIAL INTELLIGENCE" (BENJAMIN, 2019, p. 50).

This contest was organised by Youth Laboratories, based in Australia and Hong Kong, in collaboration with several companies that worked together to organise the first beauty contest judged by a robot. The process involved the following steps:

1. Contestants download the Beauty AI app.
2. Contestants take selfies.
3. The robotic judging panel examines all the photos submitted by the contestants.
4. The panel then selects a king and queen winner.
5. The winner is announced, and the news is spread worldwide.

The symmetry of facial features, skin tone, gender, race, and many other factors would be used to judge the participants since conditions were imposed, such as the participants not using makeup, growing facial hair, or wearing spectacles. To guarantee that the competition would be held under precise and unambiguous guidelines, more than 6,000 applications were submitted from roughly 100 nations. (BENJAMIN, 2019, p. 50)

Participants expressed their discomfort because the robots seemed to favour people with lighter skin tones. All 44 winners were white, and only one of the finalists had dark skin. Upon investigation, it was found that the

The dangers of algorithmic bias in artificial intelligence

reason behind this outcome was the algorithms on which the robots had been programmed, leading to biased and racist results.

Bernard Harcourt remarked that “the idea that you could come up with a culturally neutral, racially neutral conception of beauty is simply mindboggling.” (BENJAMIN, 2019, p. 51)

Many Black victims have had their bodies turned into properties controlled by police officers in their swift efforts to maintain order. The unjust power of technology now raises fear, as it operates on gender and racial segregation. What is observed is that Black individuals, particularly those living in the South compared to other continents, are controlled by white men and masters, following a hierarchical structure in the world of Silicon Valley.

There are many real-life examples of the harm artificial intelligence can cause. In 2018, Joy Buolamwini, a student at the Massachusetts Institute of Technology (MIT), wrote about her experience discovering that facial recognition algorithms in her lab, which are used worldwide, failed to detect black faces. She even had to wear a white mask for the computer to recognise her as a person. (Raikes, 2023)

Thus, it is essential to think about the people behind racism, as these machines and robots are only subject to what they have been programmed to do. For example, John McWhorter (linguist in Time magazine) says:

"Machines cannot, themselves, be racists. Even equipped with artificial intelligence, they have neither brains nor intention." (BENJAMIN, 2019, p. 60)

Concerns of bias in the employment of computer systems for a variety of tasks, including scheduling, job matching, aircraft routing, and computerised legal help for immigration, were examined by Batya Friedman and Helen Nissenbaum. Although their discussion appeared to centre on the usage of computer systems, their criticism was focused on the processes—algorithms—that these systems employ to get outcomes. For instance, it was discovered that certain hospitals had a bias in favouring residents and single people over married inhabitants. (Osoba, Osonde ; Welser, Wiliam ;, 2017, p. 8)

There are algorithms developed to evaluate visas to some countries that contain the classification tool, it has been noticed that they have been

Timedjehdine Nadia

programmed on the lack of transparency in the analysis method, which led them to making unfair decisions where they judged people as very dangerous to their country and not based on carefully studied personal criteria.

It happened between 2015 and 2020 that people who applied for a visa to enter the UK by filling out papers in the usual way and then handing over that data to an algorithm to evaluate it and give it a rating: red, amber or green. Those who were evaluated in Green were allowed to enter, but those who were evaluated in red were considered the most dangerous category and underwent further tests to verify the data by returning to race, gender, etc. and make the final decision. All this suggests algorithmic bias.

For example, we find that the Chinese regime and the issue of Uyghurs, the large Muslim minority in the country, are being monitored, and we find that during 2019, the facial recognition feature was used to check whether people passing in front of the cameras are Uyghurs or not, the newspaper wrote, quoting one of the companies that if one Uyghur lives in a neighborhood, and six Uyghurs appear within twenty days, alerts are immediately sent to the local police. All this shows the algorithmic bias based on race and religion, people should not be monitored and arbitrary decisions should be made against them, everyone is equal.

Scientists argue that the most powerful machine learning mechanisms are deep neural networks and are more accurate than humans in detecting the sexual orientation of individuals from facial images. Research published in social psychology for a long time confirms that prediction models aimed at gender alone allowed the detection of gay males and lesbian females, all this is due to facial features, but not necessarily the results obtained are accurate, biased and erroneous information may be entered, the results will not be objective and accurate, people are accused without benefit.

“Angwin” and others pointed out the reformative risk assessment system for managing criminal offenders, COMPAS, used in “Northpointe”. This program is used in sentencing and parole hearings across the country. Cases have shown that the system misrepresents the risk of recidivism among different offenders and contains racial discrimination. For example, black offenders were classified as higher risk than white offenders, even

The dangers of algorithmic bias in artificial intelligence

when white offenders committed more serious crimes. They were accused of being more likely to return to violent crime. (Osoba, Osonde ; Welser, Wiliam ;, 2017, p. 13)

Racist algorithms will continue to destroy lives all over the world and especially in light of the rampant digitization that came with the coronavirus pandemic. Consider the murder of George Florida, which sparked the global "Black Lives Matter" movement and is likely to re-energize every time an innocent black man is killed in the United States. since then, companies such as Clearview AI have been accused of racism, fueling forms of surveillance, and the justice system for their lack of inclusivity.

There is something inevitable about the overwhelming power of Technology, where self-conscious digital life forms will come to rule the world. But none of these studies clearly linked the racism of the Enlightenment with the future. The monsters of today have been reinforced in the past as indicated by Mary Shelley who imagined in her mind a time when we will cross the threshold of human and post-human through what she wrote in her author Frankenstein.

These racist algorithms can lead to unfair criminalisation because machine learning based on biased data inherently leads to biased outcomes. For example, offenders with different demographic characteristics systematically have different chances of being arrested and varying levels of punishment. Therefore, necessary actions must be taken. We need learning algorithms that are optimised based on a certain level of social acceptance, in addition to any performance metrics that work to improve them internally for task performance.

5. What Are the Proposed Solutions to Avoid Algorithmic Bias?

As a result of algorithmic racism and the violations it has caused across various fields—such as economics, education, healthcare, security, and social justice—it has become necessary to monitor these algorithms and investigate the causes of racism. Ultimately, these algorithms are a human product, and their growing impact has led to the denial of rights for many individuals due to differences in personal characteristics.

Timedjehdine Nadia

We conclude that several strategies, including diversifying datasets, can address this algorithmic racism. This can be accomplished by incorporating broad and thorough computational models that are inclusive and equitable without isolating any group from others. Using various machine-learning models and emphasising social elements—like encouraging structural training on the idea of racism in research settings and among algorithm developers—are crucial. It is also critical to increase diversity in software teams. To guarantee inclusivity in justice and prevent prejudice, the teams should be varied and not only made up of white people.

If we are to rely on algorithms for making independent decisions, as is the case in many fields, they must be equipped with tools to review the causal factors behind key decisions and provide clearer justifications or explanations for their outcomes. This is particularly important to justify statistically disproportionate and biased results that favour one group over another. (Osoba, Osonde ; Welser, Wiliam ;, 2017, p. 23)

Silicon Valley entrepreneurs support business independence from governmental restriction as well as individual liberty. However, according to a survey of a large number of people who are internet entrepreneurs, the majority of them supported immigration protection, single-payer healthcare, social benefits for the underprivileged, more taxes on the wealthy, and more. Despite this, a large number of foreign government regulation workers were against the policies. (BENJAMIN, 2019, p. 13)

In fact, the author of *The Art of Computer Programming*, the field's bible, or "the Yoda of Silicon Valley," recently commented that he feels "algorithms are getting too prominent in the world. It started out that computer scientists were worried nobody was listening to us. Now I'm worried that too many people are listening." (BENJAMIN, 2019, p. 16)

In June 2020, over 1,400 mathematicians signed an open letter demanding that audits be added to all systems previously used. The researchers wrote, "It is very easy to create a scientific cover for racism." Kory Crieder, who led the legal challenges against flawed statistical systems in the UK, added that many systems are designed to manage groups of people with far less power, social capital, and money. There is a troubling aspect to algorithmic governance being a way to contain and monitor poor people of

The dangers of algorithmic bias in artificial intelligence

color. (BURGESS, 2021, p. 60 69)

In order to combat algorithmic racism, Polamini highlights that it will be challenging to identify any face that deviates from the norm if training datasets are not diverse. She therefore supports diversity in research settings. In order to promote greater diversity among AI programmers and to increase the quantity and calibre of training datasets utilised to train systems, she established the Algorithmic Justice League. (Raikes, 2023)

We can thus say that algorithms have shown even more concerning forms of arbitrary classification, upon which AI systems are built. These algorithms have been created to detect our criminal tendencies, genetic medical history, sexual preferences, racial origin, personal traits, religious inclinations, and even our political stances. Therefore, caution must be taken with everything resulting from this. Many scholars and decision-makers have called for the diversification of people, races, and genders in scientific research labs and among software engineers to ensure greater neutrality and objectivity. Additionally, they argue that all decisions made by these systems should be reviewed and not relied upon without assessment in order to reduce these incidents and ensure that the rights and freedoms of others are not violated or marginalised.

6. Conclusion

In conclusion, we can deduce that algorithms are the backbone of all software systems that govern electronic programs and computing machines. Since they are programmed by a group of researchers and software developers, they only execute the inputs and data they are provided with in order to offer solutions to problems. Even when relying on machine learning techniques, algorithms are not conscious and do not make biased decisions on their own. The spread of algorithmic racism is, therefore, a result of the software developers and engineers who programmed them to isolate and exclude certain groups based on a set of criteria.

It is crucial to find solutions to address this issue, which has worsened and threatened individuals. Researchers are expected to address this problem by educating software engineers and ensuring their diversity—both racially and religiously—in research environments to avoid bias. Additionally, the

Timedjehdine Nadia

results produced by algorithms should be reviewed and scrutinised before being applied, especially in the field of criminal justice, as they are prone to bias. We also hope that future studies will explore more perspectives from software professionals regarding algorithmic racism through qualitative research.

7. Bibliography List:

1. ADIB-MOGHADDAM, A. (2023). IS ARTIFICIAL INTELLIGENCE RACIST? The Ethics of AI and the Future of Humanity. London: BLOOMSBURY ACADEMIC.
2. J.Salazar, L. (2023, 2 27). OUR THOUGHTS From Bias to Inclusion: Training Data and AI Ethics. Retrieved from AI: www.AI4SP.org
3. Raikes, R. (2023, 4 21). AI Can Be Racist: Let's Make Sure IT Works For Everyone. Retrieved from forbes: www.forbes.com
4. BENJAMIN, R. (2019). RACE AFTER TECHNOLOGY. Cambridge: Polity Press.
5. BURGESS, M. (2021). ARTIFICIAL INTELLIGENCE how machine learning will shape the next decade. London: copyright.
6. D.Luxton, D. (2016). Artificial Intelligence in Behavioral and Mental Health Care. San Diego, USA: Naval Health Research Center.
7. M.Blank, Rebecca; Dabady, Marilyn; F.Citro, Constance;. (2004). MEASURING RACIAL DISCRIMINATION. Washington: THE NATIONAL ACADEMIES PRESS.
8. Osoba, Osonde ; Welser, Wiliam ;. (2017). An Intelligence in Our Image. Santa Monica: RAND.
9. Ronnie de Souza , Santos; Fernando de Lima, Luiz ; Magalhaes, CESAR;. (2023, jun 27). The Perspective of Software Professionals on Algorithmic Racism. 1.
10. SLOBOGIN, C. (2021). Just Algorithms using science to reduce incarceration and inform a jurisrudence of risk. UNITED KINGDOM: CAMBRIDGE.