Journal of Economy Culture and Society
JECS 2024; 69 : 159–168
DOI: 10.26650/JECS2023-1415085

Review Article

Artificial Intelligence Bias and the Amplification of Inequalities in the Labor Market

Mahmut Özer[1], Matjaz Perc[2,3,4,5,6], H. Eren Suna[7]

[1]Prof. Dr., Turkish Grand National Assembly National Education, Culture, Youth and Sports Commission, Ankara, Türkiye
[2]Prof. Dr, Faculty of Natural Sciences and Mathematics, University of Maribor, 2000 Maribor, Slovenia
[3]Complexity Science Hub Vienna, 1080 Vienna, Austria
[4]Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 404, Taiwan
[5]Alma Mater Europaea, Slovenska ulica 17, 2000 Maribor, Slovenia
[6]Department of Physics, Kyung Hee University, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul, Republic of Korea
[7]Dr., Ministry of National Education, Paris Education Attaché, Paris, France

Corresponding author : H. Eren Suna
E-mail : herensuna@gmail.com

ABSTRACT

Artificial intelligence (AI) is now present in nearly every aspect of our daily lives. Furthermore, while this AI augmentation is generally beneficial, or at worst, nonproblematic, some instances warrant attention. In this study, we argue that AI bias resulting from training data sets in the labor market can significantly amplify minor inequalities, which later in life manifest as permanently lost opportunities and social status and wealth segregation. The Matthew effect is responsible for this phenomenon, except that the focus is not on the rich getting richer, but on the poor becoming even poorer. We demonstrate how frequently changing expectations for skills, competencies, and knowledge lead to AI failing to make impartial hiring decisions. Specifically, the bias in the training data sets used by AI affects the results, causing the disadvantaged to be overlooked while the privileged are frequently chosen. This simple AI bias contributes to growing social inequalities by reinforcing the Matthew effect, and it does so at much faster rates than previously. We assess these threats by studying data from various labor fields, including justice, security, healthcare, human resource management, and education.

Keywords: artificial intelligence; bias; Matthew effect; social inequality; misinformation

Submitted : 05.01.2024
Revision Requested : 03.02.2024
Last Revision Received : 08.02.2024
Accepted : 08.02.2024
Published Online : 28.03.2024

1. INTRODUCTION

Currently, there are serious attempts to digitize every aspect of our lives. In light of these efforts, measurement and assessment capacity has expanded in most fields, and processes are now managed using common and quantitative metrics. To put it another way, digitization studies enable the tracking of all processes using data-driven indicators and the analysis of intervention effects. This digitization process is being implemented in all areas, and substantial investments are being made to process and evaluate the continuously generated data. A variety of data mining, artificial intelligence (AI) algorithms, and machine learning (ML) techniques are commonly used to analyze data, generate specific information, and make accurate predictions (Perc, Özer & Hojnik, 2019).

As a result of advances in AI and ML, machines can now perform a variety of human behavioral and decision-making processes (Bozkurt & Gursoy, 2023; Castellucia & Le Metayer, 2019; Mahmut et al.,2022). Due to their superior calculation abilities, machines that learn from data recorded in various fields are becoming increasingly important in decision-making processes. It is practically impossible to discuss a field that does not involve some degree of AI or ML. It has become evident that AI has gone beyond being a support mechanism and has affected many aspects of society (Makridakis, 2017; Stefan, 2019). AI and automation have transformed the labor market, resulting in new jobs as skills and competencies are updated (Acemoğlu & Restrepo, 2018; Bozkurt & Gursoy, 2023; Harar, 2017).

Before delving into AI algorithm bias, we must first understand the algorithm's logic. The primary goal of these algorithms is to make predictions using numerical parameters determined in various subjects (Silberg & Manyika, 2019). In addition to the parameters in question, the algorithm requires a data set (*training set, training data*) for model development. In summary, AI makes predictions using a predetermined set of parameters based on current data. In this context, a key feature of AI is its goal of developing the most predictive models based on training data and parameters.

It is important to note that the quality of the training data is critical (Vicente & Matute, 2023). In fact, these data provide the raw material required for the algorithm to achieve its highest prediction accuracy. However, a significant problem area begins here. Because predictive power is the primary success factor for AI algorithms, many other important indicators remain in the shadows. In general, the nature of data sets and the groups they represent, among other factors, influence the outcome of AI algorithms. Despite this, the system can make highly accurate predictions without considering several important indicators, such as the output's potential results, group comparability, or ethical values and social norms (Silberg & Manyika, 2019). In light of this, even though AI algorithms have impressive computational and predictive power, there have been insufficient discussions about their logic and outputs.

The representativeness of the training data used by the AI algorithm is critical for producing high-quality results. Any deficiencies and negative aspects of the data, particularly in terms of quality, will harm the algorithms developed based on it (Aldoseri, Al-Khalifa, & Hamouda, 2023). It is important to note that if the data set used to train the AI algorithm contains an unequal number of races, ages, or genders, significant issues arise in the algorithm and the predictions made with this data set (Aquino, 2023). Furthermore, AI algorithms appear to perpetuate and replicate social biases due to their scale effect. This results in permanent and even deepening inequalities within a society.

Thus, the first step in the review should be to determine the training data set's bias and representativeness. To increase the representativeness of training data sets, it is important to use data from various sources. Using data from multiple sources, such as training data, can help to reduce the risks associated with data protection. According to Rajpurkar et al. (2022), a federated model that focuses on training data and collects data from multiple points has a high potential for risk mitigation. Therefore, using data from different institutions in different geographical regions and hierarchical levels as much as possible will increase the quality of the training data and reduce biases by increasing the representativeness of the training data (Nazer et al., 2023; Zhang and Zhang, 2023). Robert Merton (1968), who developed the Matthew effect to explain social inequality based on the Biblical verse "*to everyone who has, more will be given*" in Matthew's Gospel, claims that the Matthew effect provides a powerful insight into social inequalities. The Matthew effect causes small differences to grow, so that an advantage leads to a greater advantage and a disadvantage leads to a greater disadvantage. When no interventions are made to balance the process, advantages and disadvantages accumulate on opposing sides, resulting in a further rift between the groups (Zuckerman, 1989).

The Matthew effect is increasing inequalities in various areas, including education, health, technology, and science (Rigney, 2010). For example, Zuckerman (1977) showed that more than half of Nobel laureates worked with previous Nobel laureates. A well-known scientist will most likely receive more citations than a lesser-known scientist (Perc, 2014). According to recent studies, citations for joint publications have increased significantly, as have citations for other publications by well-known scientists and those with whom they have previously collaborated (Li et al., 2019). Young scientists want to work with well-known scientists to increase their publications and citations to capitalize on the Matthew effect. The Matthew effect allows well-known scientists to easily collaborate with young scientists worldwide,

resulting in increased recognition, publications, and citations. The Matthew effect is characterized by a self-reinforcing positive feedback cycle in which more citations and publications lead to even more citations and publications, and more recognition leads to even more reputation.

By attracting well-known academicians, talented students, and scientific research funds, well-known scientific institutions position themselves at the center of the scientific community, contributing to a wider gap between scientific institutions. The Matthew effect, which is related to economic activity, allows advantageous centers to attract more earnings, talents, raw materials, and other resources to their centers while also widening the gap between the centers and the periphery (Rigney, 2010).

Hence, the Matthew effect strongly correlates with the results of the bias problem in AI algorithms. Because of the bias inherent in AI algorithms, the privileged are in a more advantageous position, whereas the disadvantaged are in a more detrimental position. Thus, this study addresses the bias problem caused by the data set used to train AI algorithms in various fields, including justice and security, health, human resource management, and education. The bias problem in AI algorithms is emphasized as a platform for the Matthew effect, perpetuating and exacerbating inequalities.

2. BIAS IN ARTIFICIAL INTELLIGENCE

Developing AI prediction systems typically involves three stages: measurement, training, and prediction (Kizilcec & Lee, 2022). The measurement phase generates numerical data on the area under consideration, whereas the training phase teaches the system using data. Finally, during the prediction phase, the system responds to new circumstances. Thus, the AI algorithm must train on high-quality data.

It directly impacts data quality in terms of representability whether algorithms produce biased results. During the training phase, a broad area must be represented and sampled using limited indicators (Kizilcec & Lee, 2022). Most data sets have low representation and comparability, increasing bias even more. Bias in AI algorithms is highly likely, given that it is directly related to the quality of the data set. Among the data sets commonly used today, only very few can adequately represent vulnerable and disadvantaged groups while allowing for a balanced distribution of comparisons between groups.

Because data sets represent real-world situations, their biases and deviations can differ depending on race, gender, and socioeconomic status. Accordingly, the data set from which an algorithm learns and operates determines its quality (Barocas & Selbst, 2016). Without control over AI algorithms, the data set's bias is transferred to the AI algorithms, resulting in biased decisions (Aquino, 2023; Notoutsi et al., 2020). A biased data set is used in AI algorithm training, replicating the bias learned from the data (Lum & Isaac, 2016). Biases in AI algorithms are thus more than technical errors; they reflect power structures, social, economic, and political systems within society (Ulnicane & Aden, 2023).

Bias caused by low data representation in AI algorithms can result in both observable and unobservable risks (Varsha, 2023). As previously stated, bias is a natural consequence of insufficient data-quality and representation. Furthermore, because AI strives to improve its predictive power by employing a diverse set of parameters, it may result in indirect biases due to the interaction of deficiencies in various parameters. It is possible to detect observable risks when inadequate aspects of the data and parameters are available; however, it is nearly impossible to detect bias when it is part of a big-data process with an interaction of a large number of parameters and no prior knowledge of inadequate aspects of the data.

A decision may be made against groups that are less represented in the data set when data sets representing more specific groups than the entire population are used, that is, when the representativeness of the data set decreases (Nazer et al., 2023). Furthermore, differences in measurement sensitivity in the health field can lead to bias (Babic et al., 2021; Charpignon et al., 2023). In the fields of security and justice, groups with a higher representation in the data set are disadvantaged, particularly blacks, minorities, and those with low socioeconomic status (Lum & Isaac, 2016). Again, algorithms may be biased toward the more advantaged, whites, and those with higher socioeconomic status in education and human resource management. Thus, it is critical to continuously assess whether the algorithm has unintended consequences for a specific group in society based on the hypotheses or data set used in training (Nazer et al., 2023).

Because of their structural nature, AI-based algorithms can produce data-based biases. A couple of factors should be highlighted: the rapid change in competencies/skills required in the labor market because of technological advances (European Parliament, 2020; Napierala & Kvetan, 2023), and the emergence of AI-based decision-making approaches (Agrawal, Gans, & Goldfarb, 2019; OECD, 2023). As labor market demands shift, collecting reliable and representative data on these new expectations is critical. The current situation necessitates updating long-standing data sets that serve as the foundation for AI algorithms. The second issue is that algorithms may produce results that contradict the European

Pillar of Social Right principles and the United Nations' Sustainable Development Goals (European Commission, 2022). In this context, the examples we will see in various sectors in the next stage of this study will shed light on the biases caused by AI.

2.1. BIAS IN JUSTICE AND SECURITY

Risk assessment algorithms are widely used in criminal justice systems because they help determine whether punishments, bail, and parole are effective in deterring future reoffending (Bagaric, Hunter, & Stobbs, 2019). The issue of social inequality is also relevant in this context. It is well-known that recidivism prediction algorithms are unbalanced and discriminatory toward black people (Angwin et al., 2016). Even if a criminal does not reoffend, blacks are roughly twice as likely as white criminals to be incorrectly evaluated. Furthermore, this bias causes the software to classify white offenders as having a lower likelihood of reoffending, even if they do.

The fact that there are more African-American inmates than white Americans makes the risk assessment unfair to them (Bagaric, 2016). Blacks dominate the data set used to train the algorithm. Dressel and Farid (2018) found that the predictions made by software commonly used in American courts for risk assessment are no more accurate or fair than those made by people with no experience in criminal justice. Thus, it does not appear reasonable to expect AI algorithms to provide more accurate, objective, and equitable results in every situation and context than human evaluations.

AI algorithms in security software are common in the United States and Europe, particularly for identifying potential crime areas and criminals (Lum & Isaac, 2016). However, it has long been known that these software are primarily trained using criminal data from police stations, and therefore, discrimination in the data, particularly in terms of race, ethnicity, and socioeconomic status (SES), is directly reflected in the software's results. Lum and Isaac (2016) used a widely used software tool to identify areas where drug use was prevalent in a state. They concluded that, despite the high rate of drug use in that state, the software identified areas populated primarily by nonwhite residents and people with low SES backgrounds.

As a result, these regions account for the majority of drug-related arrests. Thus, by intensifying patrols in these areas, the likelihood of apprehending criminals increases, and the software can then direct police patrols back to the same area as a result of the new data. As a result, the data set is constantly growing, feeding the aforementioned bias. In this case, the subsequent training data used to update the algorithm will be biased, reinforcing the bias (O'Neil, 2016). The above-mentioned positive feedback loop works by finding a segment of society guilty in a biased and discriminatory manner (Lum & Isaac, 2016). As a result, while a segment of society that commits the same crime will not be adequately punished, research into the already disadvantaged and discriminated segments of society increases the likelihood of crime detection in this region, reinforcing social bias.

As we learn more about the biases of AI in this field, we are experimenting with different approaches to make more objective predictions. Researchers, for example, are developing algorithms in training data that focus on decision-making processes rather than group differences and historical records (Srinivas, 2023). By using additional algorithm systems calculating the extent of risk involved in each decision taken (Bagaric et al., 2022),we can take additional steps to determine how erroneous/biased the inferences made by artificial intelligence.

2.2. BIAS IN EDUCATION

The use of AI algorithms is also widespread in education. It is commonly used in education to identify potential failures and take preventative measures in advance (Holmes, Bialik, & Fadel, 2023). The COVID-19 pandemic has significantly increased the use of digital technologies in general, and AI algorithms in particular, in educational systems (Renz & Hilbig, 2023). In recent years, AI systems have been widely used to provide personalized solutions to students who have lost learning time due to the pandemic. Furthermore, approximately 40% of higher education institutions use predictive algorithms to identify students who are likely to drop out of classes or courses, and the use of such early warning systems increased during the Covid-19 pandemic (Bird, Castleman & Song, 2023). If such widely used algorithms make decisions favoring one group, educational inequalities could increase significantly. Extensive research has demonstrated the profound impact of educational inequalities in other fields (Bourdieu & Passeron, 2000; Coleman et al., 1966; Özer & Perc, 2022; Özer, 2023; Suna et al., 2020).

In education, AI applications can potentially mitigate the negative effects of the projected global short- and long-term teacher shortage. With its personalized learning capabilities, AI is regarded as one of the most important tools for meeting students' learning needs in areas where teachers are in short supply (Edwards & Cheok, 2018). Instead of replacing teachers, the goal here is for AI algorithms to assist students in learning in personalized digital environments.

AI algorithms can provide personalized suggestions to students throughout the learning process while closely monitoring their progress. While measures are being discussed, digital technologies have become even more important in assisting students in mitigating the impact of massive teacher shortages predicted in the United States, the European Union, and African countries (UNESCO, 2023).

Recent studies have examined how AI algorithms can increase educational inequalities (Kizilcec & Lee, 2022; UNESCO, 2023). Biases produced by algorithms are typically due to the inadequacy of the features used to represent the field and other quality issues in the training data set. For example, suppose academic potential is an indicator of admission to an educational institution, and SAT or ACT scores are used to represent academic potential. In that case, the algorithm's predictions will inevitably prioritize the applicants' SES, because these scores are highly correlated with SES (Sackett et al., 2009). As SES increases, students are more likely to succeed on high-stakes tests like the SAT or ACT. In this way, educational institutions are more likely to accept students with higher SES. Accordingly, the Matthew effect will be at work, as advantage will lead to more advantage and disadvantage will lead to more disadvantage, sharpening existing inequalities (Merton, 1968; Özer & Perc, 2022; Özer, 2023).

During the COVID-19 pandemic, the UK discovered that the grading algorithm used produced grades that favored private school students and, therefore, students with privileged SES levels, increasing the disadvantages of disadvantaged students and rearranging the grades in response to complaints (Smith, 2020). Conversely, under- or over-representing a group in data sets used in the learning phase can result in estimation errors in education and other fields (Chawla et al., 2002). As Baker and Hawn (2021) point out, if black students are punished more than white students in the same incident of in-school violence, how will an algorithm trained on this training set treat both groups fairly?

Research shows that an algorithm that uses training data to predict student dropout potential favors men while making fewer accurate predictions for women (Ocumpaugh et al., 2014). Evidence shows that school dropout risk estimates are less accurate for underrepresented minority groups (Bird, Castleman, & Song, 2023). These types of estimation errors cause misallocation of resources to reduce underachievement and the risk of school dropout, resulting in a decrease in the effectiveness of these critical interventions. As a result, those who are most vulnerable receive less support (Bird, Castleman, & Song, 2023).

Four sections discuss the four steps required to prevent biased predictions produced by AI in education and achieve more objective results (Baker & Hawn, 2021). Some of the most notable are as follows (Baker & Hawn, 2021): the collection and enrichment of comprehensive data regarding gender, age, ethnicity, and national origin. Continuously monitoring the balance of demographic characteristics within the entire data set by collecting data on demographic characteristics, using bias metrics to review AI predictions, maintaining access to the data sets on which the algorithm is based, and incorporating disadvantaged and vulnerable groups into the algorithm development process.

2.3. BIAS IN HEALTHCARE

The possibility of biased AI algorithms in the healthcare sector increases and deepens existing inequalities (King, 2022; Mittermaier, Raza, Kvedar, 2023). It is common for the problem to arise when an algorithm trained on data from one hospital in one region is applied to another with significantly different characteristics (Aquino, 2023). The same phenomenon occurs across races (Obermayer et al., 2019). In particular, health inequalities tend to perpetuate inequalities by reducing the number of representations in the data set during the AI algorithm's learning process (Seyyed-Kalantari et al., 2021).

Recent studies show that those who benefit the most from healthcare and spend the most receive better care (Bates et al., 2014; Obermayer et al., 2019). In this case, the algorithm ignores the needs of those who receive less service to a greater extent, reducing follow-up screening and thus increasing the number of cancer patients who go undiagnosed or untreated (Mittermaier, Raza, & Kvedar, 2023). In their study, Obermeyer et al. (2019) demonstrated that when health expenditures are used as an indicator, whites benefit significantly more from high-risk care programs. Correcting this issue would increase the healthcare service's additional assistance to black patients from 17.7% to 46.5%.

Based on a training set of images of dermatological lesions primarily taken from white patients, an AI algorithm used to identify dermatological lesions during the learning phase performs 50% less accurately in patients with darker skin colors than the claimed accuracy level (Kamulegeya et al., 2019). As previously noted in healthcare, if the learning set is primarily targeted at one gender, performance has been shown to decrease significantly when the other gender is used in the testing phase (Larrazabal et al., 2020). Furthermore, the precision of algorithms that predict skin cancer based on patients' skin images decreases due to sun exposure, particularly during the summer (Babic et al., 2021).

Celi et al. (2022) reviewed medical and surgical studies published on AI algorithms in the PubMed database in 2019. A key finding of this study is that data sets from China and the United States dominate AI algorithm training,

particularly in fields such as radiology, pathology, and ophthalmology, where imaging is commonly used in diagnosis. Therefore, AI has a large amount of data from these countries in relevant medical fields. As a result, AI algorithms can predict more accurately for patients from these countries while being less accurate for patients from other countries.

Examining the steps used to reduce the bias of AI applications in the health field reveals that the emphasis is primarily on enriching the data on which AI predictions are based. Overall, these steps include involving representatives from various social groups (particularly vulnerable groups) in algorithm development and evaluation, refraining from making predictions when the data is insufficiently diverse to make objective predictions, or using statistical methods (such as synthetic data) to provide unbiased predictions, and assessing the algorithm's results. Furthermore, it is critical to continuously evaluate algorithm outcomes and their coherence with the Translational Evaluation of Healthcare AI (TEHAI), DECIDE-AI, the Consolidated Standards for Reporting Trials-Artificial Intelligence (CONSORT-AI), and the predictive model risk of bias assessment tool developed to reduce the possibility of bias in the algorithm (Nazer et al., 2023). Furthermore, the World Health Organization's (2021) guidance on the ethics and governance of AI in healthcare has become a highly relevant source of information.

2.4. BIAS IN HUMAN RESOURCES MANAGEMENT

One of the most common applications of AI algorithms today is human resource management (HRM). AI algorithms determine which candidates are best suited for each job by analyzing the match between the skills required by the job and the skills possessed by the candidates (ILO, 2023; Köchling & Wehner, 2020). Furthermore, AI algorithms are used in this field to select the best candidates from a talent pool for advancement opportunities (Franca, 2023; Köchling & Wehner, 2020). AI algorithms are widely used to identify the best candidates for promotions within a company.

With the widespread use of AI algorithms, their applications have expanded beyond assisting HRM managers to evaluating a candidate's fit for a position and tracking their performance. According to Oracle's 2019 "Future Workplace AI at Work Global Study," approximately 50% of employers use at least one AI-based application, a significant increase from 32% the previous year (Oracle, 2019). AI algorithms are regarded as the most substantial advantage in HRM because they are free of many biases that people encounter during the evaluation process, such as the halo effect, bias, gender bias, and others in most categories of unconscious bias (Storm et al., 2023). Contrary to popular belief, training AI algorithms with large amounts of recruitment and promotion data can bias them.

According to the findings in the literature, AI and ML algorithms can produce less biased results if the data is sufficiently represented, but they are not completely free of bias. In 1988, the UK Commission for Racial Equality penalized a medical school in the UK for using software that discriminated against women and candidates with non-European names (Silberg & Manyika, 2019). Several studies have shown that AI algorithms are susceptible to the aforementioned biases if the groups represented in the data set are unbalanced and the data do not adequately reflect the population (European Network Against Racism, 2020; Köchling & Wehner, 2020; Tuffaha, 2023).

The proposed steps to prevent biases in HRM aim to train experts in the field and improve the quality of the data that forms the basis of the algorithm (Franca et al., 2023; IFOW, 2020). As a result, all HRM professionals should receive training on bias in prediction. A group of experts should assess AI's predictions in recruitment and talent management across all fields. The patterns observed in AI predictions based on demographic characteristics should be investigated, and the results should be evaluated. It is advised to assess compliance with the Age Discrimination in Employment Act and the Equal Employment Opportunity Commission frameworks.

3. DISCUSSION AND CONCLUSIONS

Using AI, ML, and deep learning (DL) is now prevalent in every aspect of life. In particular, the generation of digital data due to measurements in every field has resulted in a massive data collection, and the ability to process and analyze this massive amount of data has become a significant opportunity. At this point, AI and related algorithms are emerging, potentially significantly benefiting society. As a result, AI algorithms provide powerful support mechanisms for almost every aspect of life, from education to health, economics, and security, by making predictions based on big data as constantly learning systems.

Many companies are investing heavily in AI, ML, and DL systems, which have enormously increasing predictive power. Despite the widespread assumption that technology will be neutral toward humans, evidence suggests that smart algorithms reproduce societal inequalities, perpetuating or even increasing them. The bias problem in AI algorithms essentially benefits one society group while increasing the disadvantage for others (Mehrabi et al., 2021). As a result, the increasing number of examples has piqued the interest of both experts in the field and the general public.

AI systems can generate inequalities in various ways, including the training data set, modeling approaches, biases imposed by variables used, and the system's reaction to data that differs from the training data set when used in practice. To reduce the inequalities caused by these systems, continuous monitoring must begin at the design stage and continue through implementation and beyond. In this study, we discuss and demonstrate the risks of bias in AI algorithms using examples from various fields such as health, education, justice, and security. Prediction systems typically consist of three stages: measurement, training, and prediction. Although bias can occur at any of these stages through various mechanisms, the data sets used in the training phase remain the primary source of bias. Because the training data are based on real-world measurements, it directly reflects inequalities in all social areas, including education, health, justice, human resource management, and security. Because AI algorithms learn from this data, which includes current inequalities, their predictions about future situations will reflect these inequalities. These decisions exacerbate the disadvantages of previously disadvantaged groups. As a result of AI algorithms' bias, advantages in social spheres lead to more advantages, while disadvantages lead to more disadvantages. This results in a continuation and deepening of inequalities.

It is critical to demonstrate similar sensitivity during the modeling stage and assumptions and weighting factors that do not increase inequalities (Erdi, 2020; Nazer et al., 2023). For example, in the context of modeling the system for determining patients who require advanced health care based solely on their health expenditures, inequalities naturally arise in access to health services. Therefore, in health expenditures, and in this case, those who already have disadvantages in accessing health services cannot benefit from advanced care services due to modeling bias. As a result, disadvantages increase and inequalities persist (Obermeyer et al., 2019). The most effective way to reduce inequalities caused and exacerbated by AI systems is to ensure that models are not biased and to use higher quality training data sets (Baker and Hawn, 2021).AI algorithms appear to have found a powerful channel for influencing the Matthew effect, which exacerbates inequalities. As a result, because these algorithms provide extremely fast mechanisms, they may contribute to further increasing social inequality. An algorithm prediction, for example, will increase monitoring of a previously identified criminal group. As monitoring intensifies in these groups, the likelihood of detecting crime increases, providing the algorithm with current knowledge against this group and even increasing this group's disadvantage, resulting in a vicious cycle.

The same holds true for students whose socioeconomic status negatively affects their academic performance. If the algorithm's features are sensitive to SES, those with a socioeconomic advantage will continue to increase their advantage. Because of the bias in these algorithms, the increase in inequalities affects not only education, but all other fields as well. The inequalities created by AI algorithms greatly accelerate the accumulation of advantage or disadvantage, which Zuckerman (1977) identifies as the Matthew effect's main characteristic. As a result, society risks experiencing greater and deeper inequalities than before these smart technologies were implemented.

Therefore, the results produced by AI algorithms require far more attention. Transparency and openness in information sharing are critical components of AI algorithm development and data use (Erdi, 2020). However, resolving the bias issue is difficult because most companies are hesitant in this regard, particularly for security reasons (Lum & Isaac, 2016). Using non-biased key variables and collecting data sets of higher quality are the most effective ways to mitigate the effects of bias problems caused by representation and measurement (Baker & Hawn, 2021).

The recommendations offered in various fields to reduce bias in AI results are very similar (Bagaric et al., 2022; Baker & Hawn, 2021; IFOW, 2020). For AI to provide more objective inferences, the outputs should be integrated into human decisions as part of the current situation. Statistical methods should be used to monitor the algorithm output to detect possible biases. All social groups, particularly vulnerable groups, should be involved in algorithm development and revision processes, feedback should be solicited continuously, and inferences should not be drawn from incomplete data. The careful implementation of these precautions is critical for reducing existing biases and raising awareness when biased results are discovered.

Conversely, given the irresistible opportunities that AI systems will provide in the future, as well as the fact that they are impossible to control, slow down, or stop, there has been an increasing discussion of the risks and threats that AI systems may pose to humanity, as well as their achievements (Suleyman, 2023). Meanwhile, because smart systems are likely to cause more complex problems than bias, especially as their cognitive abilities increase in the coming years, it is critical to understand their impact on our lives and societies.

Peer-review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- M.Ö., M.P.; Data Acquisition- M.Ö., M.P., H.E.S.; Data Analysis/Interpretation- M.Ö., M.P., H.E.S.; Drafting Manuscript- M.Ö., M.P., H.E.S.; Critical Revision of Manuscript- M.Ö., M.P., H.E.S.; Final Approval and Accountability- M.Ö., M.P., H.E.S.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

ORCID :

Mahmut Özer 0000-0001-8722-8670
Matjaz Perc 0000-0002-3087-541X
H. Eren Suna 0000-0002-6874-7472

REFERENCES

Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation and work. *NBER Working Paper 24196*. Cambridge: National Bureau of Economic Research.

Aldoseri, A., Al-Khalifa, K.N., & Hamouda, A.M. (2023). Re-thinking data strategy and integration for artificial intelligence: Concepts, opportunities, and challenges. *Applied Sciences, 13*(12), 7082.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*.

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial intelligence: The ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives, 33*(2), 31–50.

Aquino, Y. S. J. (2023). Making decisions: Bias in artificial intelligence and data-driven diagnostic tools. *Australian Journal of General Practice, 52*(7), 439–442.

Babic, B., Gerke, S., Evgeniou, T., & Cohen, I. G. (2021). Beware explanations from AI in health care. *Science, 373*(6552), 284–286.

Bagaric, M. (2016). Three things that a baseline study shows don't cause indigenous over-imprisonment: Three things that might but shouldn't and three reforms that will reduce indigenous over-imprisonment. *Harvard J Racial & Ethnic* Just, 103.

Bagaric, M., Hunter, D., & Stobbs, N. (2019). Erasing the bias against using artificial intelligence to predict future criminality: Algorithms are color blind and never dash. *University of Cincinnati Law Review, 88*(4), 1037–1081.

Bagaric, M., Svilar, J., Bull, M., Hunter, D., & Stobbs, N. (2022). The solution to the pervasive bias and discrimination in the criminal justice system: Transparent and fair artificial intelligence. *American Criminal Law Review, 59*(95), 95–148.

Baker, R. S., & Hawn, A. (2021). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education, 32*, 1052–1092.

Barocas, S., & Selbst, A.D. (2016). Big data's disparate impact. *California Law Review, 104*, 671–732.

Bates, D.W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. *Health Affairs, 33*(7), 1123–1131.

Bird, K. A., Castleman, B. L., & Song, Y. (2023). Are algorithms biased in education? Exploring racial bias in predicting community college student success. Ed Working Paper No. 23–717. Annenberg Institute: Brown University.

Bourdieu, P., & Passeron, J. C. (2000). *Reproduction in education, society and culture.* Sage Publishing.

Bozkurt, V., & Gursoy, D. (2023). The artificial intelligence paradox: Opportunity or threat for humanity?, *International Journal of Human–Computer Interaction*, doi: 10.1080/10447318.2023.2297114

Castelluccia, C., & Le Métayer, D. (2019). *Understanding algorithmic decision-making: opportunities and challenges*. European Parliament.

Celi L. A., Cellini, J., Charpignon, M. L., Dee, E. C., Dernoncourt, F., Eber, R. et al (2022). Sources of bias in artificial intelligence that perpetuate healthcare disparities: A global review. *PLOS Digit Health*, 31; 1(3), e0000022.

Charpignon, M. L., Byers, J., Cabral, S., Celi, L. A., Fernandes, F., Gallifant, J. et al. (2023). Critical bias in critical care devices. *Critical Care Clinics, 39*(4), 795–813.

Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *The Journal of Artificial Intelligence Research, 16*, 321–357.

Coleman, J. S., Campbell , E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity.* US Office of Education.

Colon-Rodriguez, C. J. (2023, 12 July). Shedding light on healthcare algorithmic and artificial intelligence bias. *US Department of Health and Human Services Office of Minority Health.* Retrieved from https://minorityhealth.hhs.gov/news/shedding-light-healthcare-algorithmic-and-artificial-intelligence-bias

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances, 4*, eaao5580.

Edwards, B., & Cheok, A. D. (2018). Why not robot teachers: Artificial intelligence for addressing teacher shortage. *Applied Artificial Intelligence, 32*(4), 345–360.

Erdi, P. (2020). *Ranking: The unwritten rules of the social game we all play.* Oxford University Press.

European Commission (2022). *Mitigating diversity biases of AI in the labor market.* Horizon Europe Project Report.

European Network Against Racism (2020). *Artificial intelligence in HR: How to address racial biases and algorithmic discrimination in HR?*. ENAR Publishing.

European Parliament (2020). *Skills and jobs for future labour markets: European policies and Skills Agendas 2010–2020*. EMPL in Focus.

Franca, T. J. F., Mamede, H. S., Barroso, J. M. P., & dos Santos, V. M. P. D. (2023). Artificial intelligence applied to potential assessment and talent identification in an organizational context. *Heliyon, 9*(4), e14694.

Hale, K. E. (2020). *Using artificial intelligence to circumvent the teacher shortage in special education: A phenomenological investigation.* Doctoral Dissertation, Liberty University, USA.

Harari, Y. N. (2017). Reboot for the AI revolution. *Nature, 550*(19), 324–327.

Holmes, W., Bialik, M., & Fadel, C. (2023). Artificial intelligence in education. In *Data ethics: Building trust: How digital technologies can serve humanity* (pp. 621–653). Globethics Publications.

IFOW (2020). *Artificial intelligence in hiring: Assessing impacts on equality.* Institute for the Future of Work. Retrieved from https://assets-global.website-files.com/64d5f73a7fc5e8a240310c4d/64d5f73b7fc5e8a240310ea0_5f71d338891671faa84de443_IFOW%2B-%2BAssessing%2Bimpacts%2Bon%2Bequality.pdf

ILO (2023). *Artificial intelligence in human resource management: A challenge for the human-centred agenda?*. ILO Working Paper 95.

Kamulegeya, L. H., Okello, M., Bwanika, J. M., et al. (2019). Using artificial intelligence on dermatology conditions in Uganda: A case for diversity in training data sets for machine learning. *African Health Sciences, 23*(2), 753–763.

King, M. (2022). Harmful biases in artificial intelligence. *The Lancet Psychiatry, 9*(11), E48.

Kizilcec, R.F., & Lee, H. (2022). Algorithmic fairness in education. In Holmes W., Porayska-Pomsta, K. (Eds), *The ethics of artificial intelligence in education* (pp. 174-202). Taylor & Francis.

Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Bus Res, 13*, 795–848.

Larrazabal, A. J., Nieto, N., Peterson, V., Milone, D. H., & Ferrante, E. (2020). Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proc Natl Acad Sci USA, 117*(23), 12592–12594.

Li, W., Aste, T., Caccioli, F., & Livan, G. (2019). Early co-authorship with top scientists predicts success in academic careers, *Nature Communications, 10*, 51–70.

Lum, K., & Isaac, W. (2016). To predict and serve?. *Significance, 13*(5), 14–19.

Mahmud, H. A. K. M., Islam, N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change, 175*, 121390.

Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures, 90*, 46–60.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35.

Merton, R. K. (1968). The Matthew effect in science. *Science, 159*, 53–63.

Mittermaier, M., Raza, M. M., Kvedar, J. C. (2023). Bias in AI-based models for medical applications: Challenges and mitigation strategies. *NPJ Digital Medicine, 6*, 113.

Napierala, J., Kvetan, V. (2023). Changing job skills in a changing world. In: Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S., Vespe, M. (eds) *Handbook of computational social science for policy* (pp. 243–259). Springer, Cham.

Nazer, L. H., Zatarah, R., Waldrip, S., Ke, J. X. C., Moukheiber, M., Khanna, A. K., et al. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digital Health, 2*(6), e0000278.

Ntoutsi, E., Fafalios, P., Gadiraju, U., et al. (2020). Bias in data-driven artificial intelligence systems-an introductory survey. *WIREs Data Mining Knowl Discov, 10*(3), e1356.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366, 447–453.

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology, 45*(3), 487–501.

OECD (2023). *OECD employment outlook 2023: Artificial intelligence and the labour market.* OECD Publishing.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy.* Crown Books.

Oracle (2019). *AI in human resources: The time is now*. Retrieved from https://www.oracle.com/a/ocom/docs/applications/hcm/oracle-ai-in-hr-wp.pdf

Özer, M., & Perc, M. (2022). Improving equality in the education system of Türkiye. *İstanbul University Journal of Sociology, 2022, 42*(2), 325–334.

Özer, M. (2023). The Matthew effect Turkish education system. *Bartın University Journal of Faculty of Education, 12*(4), 704–712.

Perc, M. (2014). The Matthew effect in empirical data. *Journal of Royal Society Interface, 11*(98), 20140378.

Perc, M., Özer, M., & Hojnik, J. (2019). Social and juristic challenges of artificial intelligence. *Palgrave Communications, 5*, 61.

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine, 28*, 31–38.

Renz, A., & Hilbig, R. (Eds.). (2023). *Digital transformation of educational institutions accelerated by covid-19: A digital dynamic capabilities approach.* Emerald Publishing.

Rigney, D. (2010). *The Matthew effect: How advantage begets further advantage.* Columbia University Press. New York.

Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance. *Psychological Bulletin, 135*(1), 1–22.

Seyyed-Kalantari, L., Zhang, H., McDermott, M. B. A., Chen, I. Y., & Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Med, 27*, 2176–2182.

Smith, H. (2020). Algorithmic bias: should students pay the price? *AI & Society, 35*, 1077–1078.

Silberg, J., & Manyika, J. (2019). Notes from the AI frontier: Tackling bias in AI (and in humans). McKinsey Global Institute.

Srinivas, N. (2023, February 24). The ethical debate of AI in criminal justice: Balancing efficiency and human rights. *Manage Engine* Insights. Retrieved from https://insights.manageengine.com/artificial-intelligence/the-ethical-debate-of-ai-in-criminal-justice-balancing-efficiency-and-human-rights//

Stefan, V. (2019). *Artificial intelligence and its impact on young people*. Seminar Report from the Council of Europe.

Storm, K. I. L., Reiss, L. K., Guenther, E. A., Clar-Novak, M., & Muhr, S.L. (2023). Unconscious bias in the HRM literature: Towards a critical-reflexive approach. *Human Resource Management Review, 33*(3), 100969.

Suleyman, M. (2023). *The coming wave: Technology, power, and the twenty-first century's greatest dilemma.* Crown: New York.

Suna, H. E., Tanberkan, H., Gür, B. S., Perc, M., & Özer, M. (2020). Socioeconomic status and school type as predictors of academic achievement. *Journal of Economics culture and Society, 61*, 41–64.

Tuffaha, M. (2023). The impact of artificial intelligence bias on human resource management functions: Systematic literature review and future research directions. *European Journal of Business and Innovation Research, 11*(4), 35–58.

Ulnicane, I., & Aden, A. (2023). Power and politics in framing bias in artificial intelligence policy. *Review of Policy Research, 40*, 665–687.

UNESCO (2023). *The teachers we need for the education we want: The global imperative to reverse the teacher shortage.* UNESCO & Education 2030 Factsheet.

Varsha, P. S. (2023). How can we manage biases in artificial intelligence systems-A systematic literature review. *International Journal of Information Management Data Insights, 3*, 100165.

Vicente, L., & Matute, H. (2023). Humans inherit artificial intelligence biases. *Scientific Reports, 13*, 15737.

Zhang, J., Zhang, Z. M. (2023). Ethics and governance of trustwothy medical artificial intelligence. *BMC Medical Informatics and Decision Making, 23*, 7.

Zuckerman, H. A. (1977). *Scientific elite: Nobel laureates in the United States.* New York: Free Press.

Zuckerman, H. (1989). Accumulation of advantage and disadvantage: The theory and its intellectual biography. In Mongardini, C., Tabboni, S. (Eds), *Robert K. Merton and contemporary sociology* (pp. 153–176). New Brunswick, NJ: Transaction.

How cite this article

Ozer, M., Perc, M., & Suna H.E. (2024). Artificial intelligence bias and the amplification of inequalities in the labor market. *Journal of Economy Culture and Society, 69*, 159–168. https://doi.org/10.26650/JECS2023-1415085