W I S D O M

Special Issue 1(2), 2022

PHILOSOPHICAL ISSUES OF ECONOMICS

YEREVAN – 2022

Anastasia LOBACHEVA,
Ekaterina KASHTANOVA

SOCIAL DISCRIMINATION IN THE EPOCH OF ARTIFICIAL INTELLIGENCE

Abstract

The article aims to study the genesis of understanding the causes of social discrimination from its traditional manifestations to the era of digitalization and artificial intelligence.

The methodological basis of this scientific article was formed by the approaches, methods and principles of scientific research. The authors independently check existing theories and previous results of practical research in the field of social discrimination, as well as discover new modern forms of its manifestation generated by the action of artificial intelligence and subject them to open discussion, offering their vision of solving the problem of neutralizing the risks spread by the use of artificial intelligence technologies concerning certain social groups of people. In this article, the authors continue their case studies in exploring the ethics of artificial intelligence and the benefits and risks of its ubiquity and use.

Keywords: artificial intelligence, social discrimination, digitalization, digital literacy, artificial intelligence bias.

Introduction

Recently, employers or scientists and each person are increasingly feeling the effect of artificial intelligence technologies in various areas of life.

The primary purpose of this article is to continue studying the issue of how much the use of artificial intelligence and algorithmic solutions aggravates the problem of social discrimination, which is inevitably associated with the operation of artificial intelligence. By artificial intelligence (AI), we mean the ability of a computer to learn, make decisions and perform actions inherent in human intelligence (Leonov, Kashtanova, & Lobacheva, 2021). Algorithmic decisions are programmes according to which AI operates.

AI is penetrating deeper and deeper into business and the world as a whole, influencing vital decisions, such as employment, obtaining loans or affordable healthcare. This increases the risk of social discrimination from AI. Managing and

mitigating this risk begins with understanding how such discrimination can occur and why it can be difficult to detect.

In the short term, the goal of preserving the beneficial effects of AI on society motivates research in many areas, from economics and law to technical topics, such as verification, validity, safety and control (Suen, Hung, & Lin, 2020). Insignificant fraud or implicit injustice in cyberspace is still a completely insignificant (to some extent even a side) effect in comparison with the global advantage that the AI system gives today - it learns to do what a human wants from it, adopting human's traditional, more often routine, functions.

In the long term, the main question is what will happen if the new powerful AI becomes much better and more efficient than people in solving all their problems? However, many experts are already expressing concern about this development of events and declare that if you do not learn how to coordinate the actions of AI,

then human power on earth will end. We believe that the existing research, including ours, will help form an understanding of the importance of this issue and draw the close attention of all interested parties to it.

Theoretical Basis

The issue of discrimination by AI is closely intertwined with the ongoing debate in the academic community about AI's ethics. For example, there is an opinion that AI algorithms undermine the social safety system, criminalize the poor, enhance discrimination and threaten our national values (Vinichenko, Narrainen, Melnichuk, & Chalid, 2020). According to other authors, if our entire society does not comply with ethics and requires fairness in information exchange and data transfer, discrimination caused by AI will continue to grow (Symitsi, Stamolampros, Daskalakis, & Korfiatis, 2020). In his scientific review (Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016), the author agrees that most of the relevant literature is devoted to explaining how discrimination is the result of today's discrimination biased evidence and decision making. We can agree with the view that the current causes of AI-driven discrimination come from the same conceptual problems that have characterized discrimination since its very formal interpretation in law and ethics (Sinhaa, Singhb, Guptaa, & Singha, 2020).

The concept of discrimination is widely used in everyday speech and many national laws and supranational codes. It is difficult enough to interpret the essence of discrimination in such a way as to consider all, or at least most of the meanings of this concept. In order to express n the best way the meaning of discrimination that can arise (and is already arising) with the advent and spread of AI, let us turn to its manifestations that existed in the epoch before the arrival of AI.

One of the brightest examples of social discrimination is the division of the population into castes in India. Furthermore, one of the reasons why the caste system was able to exist was func-

tional interdependence in Hindu society, which allows a wide range of differentiation without disturbing the social structure. The religious foundation embedded in the caste system made it easier for the higher castes to perpetuate differences and enjoy the growing privileges of suppressing the lower castes.

In medicine, only in the late 1950s social discrimination became an important research topic. Until that time, people with deviations from socially established norms were considered sick and having pathology, and society excluded opportunities for such people to receive education, normal life and recognition. In many parts of the world, people with disabilities were barred from participating in public affairs because of physical barriers to their mobility; and social discrimination against them at least made it difficult for them to study and work.

Symptoms of negative behaviour are common concerning people who are overweight or have some other physical peculiarities that deviate from the norms in a particular society.

The New Testament mentions that leprosy is also considered a determining factor in social discrimination; the sick person can only be cured with the help of a spiritual miracle, and even today, the isolation of sick people continues to be carried out as the most effective strategy to combat the spread of disease during a pandemic. In addition, the most famous manifestations of discrimination are racial discrimination, class discrimination and gender discrimination (Popkova & Gulzat, 2020).

Thus, social attitudes, both cultural and interpersonal, clearly affect the behaviour of the majority of people due to traditions, ideology, a system of values and beliefs, as well as the standard of a person cultivated in society in terms of appearance, manners, religious beliefs, activities and are the cause of the emergence of social discrimination.

Research Problem

Social discrimination is closely related to be-

longing to a group. However, as you know, no type of group membership has the right to it. The legal provision on discrimination is reflected in Article 26 of the International Covenant on Civil and Political Rights.

"All people are equal before the law and have the right for equal protection of the law without any form of discrimination. A fair law prohibits any discrimination and guarantees equality for all, and the law must provide effective protection against discrimination on any ground such as race, colour, sex, language, religion, political or other opinions, national or social origin, property, birth or another status" (International Covenant on Civil and Political Rights, 1966).

It would seem that all issues related to discrimination were studied and acquired legal status. However, there is a new manifestation of discrimination at the moment, which is very difficult to predict and foresee. This new wave of discrimination is associated with the spread and penetration of AI into all areas of life. Indeed, the truth is that AI is starting a technological revolution, and while it is just going to take over the world, there is a more pressing problem that we already face, and it is AI bias. What is it?

AI bias is a significant bias in the data used to create AI algorithms, ultimately leading to discrimination and other social consequences (Courtland, 2018). Let us take a simple example. Imagine that we want to create an algorithm that decides whether an applicant will be admitted to a university or not, and one of our input data will be the applicant's geographic location. Hypothetically if a person's location is strongly correlated with ethnicity, then our algorithm would indirectly give preference to specific ethnic groups over others. This is an example of bias in AI.

Main Results

Here are real examples of when AI algorithms showed bias and discrimination against others.

In October 2019, researchers found out that an algorithm used for more than 200 million people in US hospitals to predict which patients are likely to need additional care gave preference to white patients compared to black patients. Although the race itself was not a variable inherent in this algorithm, another variable strongly correlated with race was the history of health care costs. The rationale was that cost summarizes the amount of health care needs a particular person has. For various reasons, black patients, on average, had lower health care costs than white patients with the same illnesses.

Another example is Amazon, one of the most significant technological giants globally. Thus, it is no surprise that they actively use machine learning and AI. In 2015 Amazon realized that its algorithm for hiring employees was biased towards women. The reason for this was that the algorithm was based on the number of resumes submitted during the last ten years, and as most of the applicants were men, the AI was trained to prioritize men over women.

Digitalization, which is "served" to society under the slogan of "convenience", is civic digitalization. Furthermore, the spread of AI within society brings the complete destruction of privacy – an opportunity for social discrimination. The most striking example of social discrimination in the epoch of AI is the threat of introducing a system of social ratings. The social rating system is a system of assessment based on the socio-political behaviour of individuals, organizations and other institutions to determine their "social reputation", on the basis of which the policy of incentives and sanctions is implemented, according to the words of I. Ashmanov, the member of the Presidential Council for the Development of Civil Society and Human Rights and the entrepreneur in the field of IT and AI, in 2 hours in the Darknet, as an experiment, he acquired complete information about a person, including his bank accounts, assets, passport data, education, place of residence and work. The most exciting thing is that along with these data, "movement around the city during the day" was sold - the video path of a person from the cameras of the "Safe City" system with the transfer of images from the camera to camera. All this in-

formation cost less than 10 thousand rubles (Ashmanov, 2020).

In addition, today, we have given much of our decision making to complex machines. Automatic right of the system for decision-making, ranking algorithms and risk prediction models monitor and determine which families receive the necessary subsidies, who is shortlisted for employment, and who may be most inclined to cheating. There are cases when the system denied access, for example, to transport or shopping centres to people with a particular type of appearance because the security system based on AI determined this appears to be potentially dangerous (for example, similarity to the image of a terrorist).

The Report to the Council of Europe Anti-Discrimination Department from 2018 states that the anti-discrimination law has "several weaknesses in AI (Ross & Konyavsky, 2020).

Let us try to find out in what areas of human resources management the used decision-making algorithms and other types of AI create discriminatory effects or can create them in the foreseeable future.

Decision making based on artificial intelligence can lead to discrimination in several ways. One is to define a "target variable" and "class labels". In human resources management, the situation of choice most often arises, from the selection of personnel for work to the issues of promotion and dismissal. For example, a company needs an AI system to sort job applications in order to find good employees. But how should a "good" employee be defined? In other words, what should be the "good employee class labels"? Is a good employee the one who sells more products than all others? Or someone who is never late for work? Nevertheless, the candidate for the position, living further from the company's location, will be determined as potentially coming late.

Also, research on the applicability of AI in the field of personnel management leads to unexpected results (Chang, 2020). Interestingly, some managers are reluctant to agree with AI as they

are afraid that AI can discriminate against their job roles and importance as leaders, reducing their influence in the workplace. These managers tend to interpret AI as a threat to their careers and evaluate AI from a more subjective and negative point of view.

How does bias creep into a dispassionate set of algorithms that deal with complex, pure data? The answer to this question is quite simple.

AI is only as good as the data that powers it. Its quality depends on how well its creators have programmed it to think, make decisions, learn and act. As a result, AI may inherit or even reinforce the biases of its creators, who are often unaware of their own biases, or AI may use biased data.

In this regard, the logical question is who is the creator of AI, and, in fact, who makes the decisions? Perhaps few people notice it, but in our country and the world as a whole, a new digital class is emerging – concerning digital means of production. This class includes those who have free access to the personal data of citizens: for example, employees of the Multifunctional Centre (MFC) or the registry office, who have access to large databases of personal data of citizens, followed by programmers who write programmes to create databases, then system administrators who organize their work, the IT directors (CIO in the sphere of IT) and the officials who manage it all. However, the main difficulty arises because the official who exercises this management does not have enough digital competencies or lacks them.

Programmers and system administrators have an absolute sense of freedom, irresponsibility and impunity. They did not take the oath, perhaps; they signed some non-disclosure obligations, but the responsibility is only administrative, not criminal. An official makes his managerial decision based on the data provided to him by his representative. He cannot influence or change this data and cannot check them because of his digital illiteracy.

Making the average portrait of the modern creator of algorithms for AI, we saw the follow-

ing picture. According to I, these are people, mainly with a technical education, aged 20 to 30 years old, technocrats without notable convictions. Ashmanov, the new generation of programmers, belongs to the so-called "digital barbarians", who know only the digital sphere, and outside of it, they know almost nothing - neither about history, nor about culture, nor ethics. They are simply not interested. All types of ethics for them are concentrated in the algorithmization of life.

If they do not have ethical ideas, then from their point of view, distributing information about another person is not theft or a crime. AI systems make decisions of an ethical nature because decisions about people belong to the sphere of ethics, and this ethics is downloaded there by programmers who do not possess it. Even if the authors of the programmes claim that their algorithms are based on neutrality and inclusion, they develop them on behalf of someone else, and there is a great danger that this neutrality and inclusion are dictated by the persons who are the programme's customers. AI has the ability to form the decisions of individuals even without their knowledge, giving those who control algorithmic decisions full implicit power. In addition to issues of general cultural knowledge, the creators of algorithms and collectors of data used to test and launch them will also not be able to foresee all variants of the development of an event. A simple example is a car driven by a robot without a driver. What if the robot's creators forget to test its image recognition at night in heavy fog in the countryside?

The results of using AI technologies already provide an opportunity for all interested parties (corporations-monopolists, government, etc.) to collect, store and analyze a vast amount of data. This information can be used with complete impunity to increase efficiency and profit. At the same time, the possible consequences of technological breakthroughs and government innovations for certain groups of the population will remain unaccounted for - intentionally or unintentionally – we will never know. At the same

time, an individual person, a living person, having will, emotions, desires and needs, will generally remain on the sidelines from what is happening. Thus, a step is taken towards a society where there is no place for the individual, where the AI itself writes the algorithms, and robots make the decisions. They, of course, will strive to make decisions that correspond to the preferences of the majority, but the flip side of these algorithmic decisions is the inability to go beyond the framework determined by this decision. This is especially dangerous for the younger generation, for whom the acquisition of experience to act independently will be practically inaccessible according to their own opinion.

The creators of algorithms for AI simply cannot consider every piece of data that represents the amplitude of the personality and the needs, desires, and hopes of this person. Who is collecting data today? Do the people who reflect the data points even know what the data is used for, or did they just agree with the terms of service provided because they had no real choice? Who makes money from this data? How can anyone know how his / her data is being processed and for what purposes to justify those purposes? There is no transparency here, and data use monitoring is a farce. All this is hidden from outsiders. "Who owns the information, he owns the world" - this phrase of Rothschild after the famous scam with the purchase of securities is quite relevant today. The economic system we live today has such a nature that data will be used to enrich and/or protect a group of certain people rather than an individual person.

In the future, based on algorithms created by AI, there may be a gap between people who understand digital technologies (mainly the most prosperous, who are mostly in demand in the created digital ecosystem) and those who do not have digital competence or, due to own various reasons, do not want to master it. The algorithmic solutions themselves will be able to instantly provoke disagreements of any kind between different groups of the population using the media, as AI knows almost everything about the prefer-

ences not only of the groups classified according to some criterion but also according to the preferred method of obtaining information (television, Internet, social networks, etc.).

Furthermore, traditionally discrimination caused by AI is associated with the threat of mass unemployment and its consequences. Indeed, if an algorithm can efficiently represent a task, a machine can easily perform it.

Discussion

So, let us formulate the main scientific results that we obtained when determining the possibility of discrimination in the new digital age.

We identified explicit and latent problems of the consequences of the distribution of biased AI.

We call social discrimination caused by the limitations of the creators of one or another AI technology an explicit or main problem. Among the so-called latent or related problems, we highlight the following: algorithmic lack of transparency, cybersecurity vulnerabilities caused by the lack of protection against threats from new fraudsters in the network, unfairness and bias, lack of competition, adverse consequences for employees, breaches of privacy and data protection and, as a result, possible harm to a person's reputation, irresponsibility of developers and users for damage and lack of reporting on data use.

We identified possible threats of discrimination due to the distribution of AI and presented their essential characteristics (Table 1).

Table 1.

Potential AI Discrimination Threats and the Manifestation of These Threats

Potential discrimination threat as a result of AI distribution	Manifestation of threat
Data, algorithms and predictive modelling domination over human judgment and emotion	• the impossibility of taking into account the broadest characteristics and peculiarities of each personality; • AI algorithms developed for the company seek to maximize profits rather than maximize the public good; • persons who have access to the management of AI and databases have the opportunity to manipulate people; • disappearance of personal confidentiality; • lack of control and transparency of actions; • criticism of AI algorithms will be belittled and suppressed, and rejected due to the prevalence of digital logic over the process; • people lose their free will due to the need to follow the algorithm
Algorithmically organized AI systems contain bias	• AI algorithms are developed using data selected by certain privileged participants – in the interests of consumers like themselves; • programmers who create algorithms for AI are an unrepresentative subgroup of the population; • AI values efficiency more than fairness; • Producers of AI algorithms (corporations and governments) tune the algorithms in such a way as to make choices that are favourable for themselves
AI deepens differences	• users who are "quarantined" in various ideological areas may lose the human ability to empathize; • Non-active users of AI will be in an unfavourable position; • anything that the algorithms consider risky or less profitable will have negative consequences;

	• massive increase of productivity gains on account of automation will increase inequality between workers and capital owners
The rise in unemployment as a result of AI distribution	• AI is cleverer, more efficient, more productive and cheaper than an employee for whom it is necessary to create working conditions and to ensure that his/her rights are respected; • violation of the economic model of the market, according to which capital is exchanged for labour to ensure economic growth (if the labour force is no longer part of this model)

We will likely need additional regulation to protect justice and human rights in the field of AI. However, regulating AI as a whole is not an unequivocally correct approach as the use of AI systems is too varied for a single set of rules. You should also take into account national, sectoral, geographical and other peculiarities when drawing up such rules. More research is needed, and more discussion and debate are needed.

We believe that another result of our research is the development of recommendations for minimizing and avoiding bias and discrimination as a result of large-scale civil digitalization and the distribution of AI.

One of the main reasons that can create AI bias and exacerbate differences is the lack of digital literacy and digital competencies among a large number of the population, not to mention the lack of knowledge about the operation of the AI decision-making mechanism and the development of algorithms based on big data. Therefore, it is necessary to develop digital competencies massively and from a very early age, introducing them into the compulsory public education curriculum to make the general public understand how AI algorithms function.

The next step is to ensure transparency of information on how data is collected and used and to develop public understanding of who is responsible for their use and non-proliferation. After all, it is no secret that despite the massive growth in cybercrime, the facts of criminal prosecution and punishment for them are practically unknown. According to the Central Bank, in 2020, the fraud "in the digital volume of transactions without the consent of the client" increased by 38%, and the amount of money stolen increased by 52% over the year and amounted to 10 billion rubles (Central Bank of Russia).

People today are very interested, for example, in the information about where and under what conditions food or clothes are produced. In the same way, we should ask ourselves how our personal data is collected, our opinions in any polls and how, and most importantly, who makes decisions subsequently. What is the sequence of transmission of this information? Are assumptions allowed, what criteria were used to select the information and data, and how relevant they are. Which parties are interested in making decisions, and how influential these parties are. In other words, at the moment, only very few people understand and, most importantly, are aware of the effect of those AI technologies capable of creating and changing the existing reality. However, as we have already noted, those who create and develop algorithms are not responsible for society. It is necessary to overcome this circumstance in the near future and develop an approach that will aim to oblige AI developers to consider human rights at every stage of development categorically. In turn, this step will act as a guarantee that the algorithms implemented in society will eliminate, not exacerbate, social inequality.

Control mechanisms should include stricter data access protocols. They should also include a mandatory list of responsible persons indicating their level of responsibility and the conclusion of nondisclosure agreements. It is necessary to provide the possibility of remote monitoring of repeated access to the information by this or that responsible person, system failure functions, setting access times, and the impossibility of selling information to third parties without the consent

of the regulatory authorities. By the way, many legislators and regulators are now claiming that the vast server farms of Google and Facebook need to become more transparent and understandable. These monopolists have the size, scale and, in some ways, the importance of nuclear power plants and refineries, but with little or no regulatory oversight. This situation must change.

Conclusion

Thus, we can summarize all the above-mentioned and formulate the following requirement to avoid digital bias - algorithmic transparency should be established as a fundamental requirement for all AI-based decision-making.

One more nuance. Algorithmic accountability, in our opinion, is a large-scale project that requires the involvement of various specialists and public representatives. Acknowledging bias is often a matter of perspective, and people with different racial or other identity and the economic background will notice different biases. Building diverse teams will help to reduce the potential risk of AI bias. The algorithm team should consist of data scientists and business leaders, government officials and professionals with various backgrounds and experience, such as lawyers, accountants, sociologists and ethicists, journalists and religious leaders. Everyone will have his/her perspective on the threat of bias and how to help mitigate it.

The assessment of predictive models based on AI decisions must necessarily include an assessment from social groups. As we learn from the examples above, we should try our best to ensure that such indicators as factual accuracy and false-positive results are consistent when comparing different social groups, be it gender, ethnicity or age.

Furthermore, it is necessary to consistently regulate the issues of using AI at the legislative level. Moreover, here again, all the above-mentioned requirements should be provided – persons should make such decisions with a high level of digital literacy, developers' teams should

include representatives of various professions, and the decision-making process should be based on openness and accessibility principles and transparency. Leaders at the highest level must understand the need for responsible AI – that is, AI which is ethical, reliable, safe, well-managed, compatible and explainable. Social discrimination caused by the action of AI is not inevitable; everything depends on us, on how we, the nation and civilized society, can put an end to it.

References

Ashmanov, I. (2020). Monstrous Anti-utopia in Reality. Retrieved March 20, 2022, from https://pandoraopen.ru/2020-11-16/igor-ashmanov-nachinaetsya-kakaya-to-chudovishhnaya-antiutopiya-v-realnosti/

Central Bank of Russia (2020). *Obzor otchetnosti ob incidentakh informacionnoi bezopasnosti pri perevode denezhnykh sredstv* (Review of the reporting of information security incidents in money transfers, in Russian). Retrieved April 5, 2022, from https://cbr.ru/analytics/-ib/review_1q_2q_2020/

Chang, K. (2020). Artificial intelligence in personnel management: The development of APM model. *The Bottom Line, 33*(4), 377-388. https://doi.org/10.1108-/BL-08-2020-0055

Courtland, R (2018, June 20). Bias detectives: The researchers striving to make algorithms fair. *Nature*. Retrieved March 15, 2022, from https://www.nature.-com/articles/d41586-018-05469-3

Leonov, V. A., Kashtanova, E. V., & Lobacheva, A. S. (2021). Ethical Aspects of the Use of Artificial Intelligence in the Social Sphere and Management Environment. *Social and Behavioural Sciences, 118,* 989-998. doi:10.15405/epsbs.2021.04.-02.118

Mezhdunarodnyi pakt o grazhdanskikh i politicheskikh pravakh (International Cove-

nant on Civil and Political Rights, in Russian). (1966, December 16). Retrieved April 1, 2022, from https://-www.un.org/ru/documents/decl_conv/conventions/pactpol.shtml

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, Joly-December,* 1-21. doi:10.1177/2053951716679679

Popkova, E. G., & Gulzat, K. (2020). Contradiction of the digital economy: Public well-being vs. cyber threats. *Lecture Notes in Networks and Systems, 87*, 112-124. doi:10.1007/978-3-030-295-86-8_13

Ross, G., & Konyavsky, V. (2020) New method for digital economy user's protection. *Lecture notes in networks and systems, 78*, 221-230. doi:10.1007/978-3-030-22493-6_20.

Sinhaa, N., Singhb, P., Guptaa, M., & Singha, P. (2020). Robotics at workplace: An integrated twitter analytics – SEM based approach to behavioural intention to accept. *International Journal of Information Management, 55,* 102210.

Suen, H. Y., Hung, K. E., & Lin, C. L. (2020). Intelligent video interview agent used to predict communication skill and perceived personality traits. *Human-Centric Computing and Information Sciences, 10,* 3, 1-12. https://doi.org/-10.1186/s13673-020-0208-3

Symitsi, E., Stamolampros, P., Daskalakis, G., & Korfiatis, N. (2021). The Informational Value of Employee Online Reviews. *European Journal of Operational Research*, 288, 605-619.

Vinichenko, M. V., Narrainen, G. S., Melnichuk, A. V., & Chalid P. (2020). The influence of artificial intelligence on human activities. *Frontier Information Technology and Systems Research in Cooperative Economics, Studies in Systems, Decision and Control, 316*, 561-570. doi:10.1007/978-3-030-57831-2_60

ABOUT AUTHORS

Anastasia LOBACHEVA - PhD in Economics, Associate Professor of the Department of Personnel Management of the State University of Management, Moscow, Russian Federation.
ORCID iD: https://orcid.org/0000-0002-4210-9018
E-mail: aslobacheva@mail.ru

Ekaterina KASHTANOVA - PhD in Economics, Associate Professor of the Department of Personnel Management at the State University of Management, Moscow, Russian Federation.
ORCID iD: https://orcid.org/0000-0002-5949-9198
E-mail: kashtanovae@mail.ru