

ARTIFICIAL INTELLIGENCE: EXPLORING ETHICAL CHALLENGES AND ALGORITHMIC BIAS

Anvar Zokhidov¹, Turayev Baxodir Ergashevich², Ergashev Javohir Bahodirovich³

¹PhD Candidate in AI, the University of Texas at Dallas // Former Senior Manager in AI

²Doctor of Technical Sciences, Professor

³Management Development Institute of Singapore in Tashkent

Abstract. This article explores the ethical implications and potential biases of artificial intelligence (AI) in both complex decision-making processes and everyday life scenarios. AI systems, while increasingly capable and autonomous, can inadvertently reinforce existing social biases, exhibit discriminatory behavior towards certain groups, genders, or cultural backgrounds, and perpetuate harmful stereotypes. These ethical challenges are compounded by the rapid evolution of AI technologies, which often outpace the development of regulatory frameworks and societal understanding. By examining real-world examples and current research, this study highlights the importance of integrating ethical principles, fairness, and accountability into AI design and deployment. Addressing these issues is critical to ensuring that AI contributes positively to society while minimizing unintended harm.

Keywords: Artificial Intelligence, Ethics, Algorithmic Bias, Fairness, Discrimination, Social Impact, Responsible AI, Stereotypes

INTRODUCTION

Artificial intelligence (AI) has rapidly transformed from a simple vision of the future into an integral part of many people's lives. Examples of its usage can already be observed when applying for jobs in companies and firms, in the field of healthcare, education, banks and social networks. AI has completely changed these areas that used to function alone, and now they have become more intelligent. While this has certainly brought many benefits, it has also started to cause ethical issues and bias. AI doesn't just make decisions out of nothing. Initially, it needs to be trained to work with a huge amount of data, which serves as a source on which AI makes decisions. Unfortunately, any data on which AI learns is highly likely to contain information about various stereotypes, incorrect information, miscalculations, historical uncertainty and inequality. They can further enhance the AI's decision-making ability, leading to unpredictable consequences. Can you imagine that a machine is subjecting a person to extreme discrimination, and there's nothing you can do about it because you

can't argue with the machine, it's just doing what it was originally intended to do, but the distorted data misleads the machine during training. These are complex problems in the modern world, and they need to be addressed. The following paragraphs provide examples of the biases in AI that is already taking place around the world.

MATERIALS

AI Bias in Hiring and Employment.

Currently, there are more frequent cases when the resumes of candidates applying for jobs are initially reviewed by artificial intelligence, and not by humans. On the one hand, it has significantly improved the resume selection process. Traditional application tracking systems (ATS) are not sophisticated enough to take into account all the details, but artificial intelligence not only takes them into account, but also performs a certain level of logical reasoning, producing results about the candidate's resume. These resumes are then sent to the HR department, which conducts an analysis at another level. This step is considered extremely important, especially for large companies, as they receive thousands of job applications. It is impossible to view each of them, so the first level of filtering is necessary. The AI handles this pretty well. However, AI performs biased analysis. For example, it may give a male candidate more points than a female. Since AI makes predictions, it can be concluded that the female candidate will soon go on maternity leave, and hiring her here for a long time is risky. This is a common problem even without AI, but with AI, such cases can occur much more often. This should not be allowed. No one should have gender-based advantages that do not reflect their actual experience and skills. Moreover, artificial intelligence-based hiring systems may also be biased against candidates from certain educational institutions. If these systems had been trained on the company's historical data, they would probably have caught all the patterns and reinforced them. This can lead to candidates from prestigious universities having an advantage over those who graduated from lesser-known universities. Although prestigious universities produce exceptionally educated candidates, this does not mean that they should have advantages over others, since candidates from other universities may be as well educated as the first ones. Also, it should be thought about the potential benefits that will be missed just because of this priority. This can lead to companies missing out these candidates.

RESEARCH AND METHODS

AI in Healthcare: The Risk of Misdiagnosis.

From the very beginning, the healthcare industry has evolved from year to year. Starting from treating people with herbs and ending with modern medical equipment that helps doctors diagnose

diseases and scientists creating new medicines. Over the past few years, even more advanced technologies have been created with the help of artificial intelligence. While AI-driven medical equipment can now detect diseases by analyzing patterns that hadn't even been considered before, one of the main advantages of AI- driven equipment is its ability to predict possible diseases in the future. Computer vision, which is an area of artificial intelligence, analyzes the results of CT scans, X-rays, and any other images reflecting the patient's current condition, and predicts possible disruptions that may occur in the near future. Moreover, these forecasts even specify the date range of these results. This is fantastic, because timely treatment can save many lives. However, it should be borne in mind that people of different races are susceptible to certain diseases in different ways. Dermatological treatment affects people with darker skin in different ways than people with lighter skin, and vice versa, and their diseases develop in different ways. If the AI has been trained mainly on one type, it may misdiagnose or incorrectly predict the disease and treatment. This can become dangerous, and it is also not easy to solve this issue. This bias can potentially be avoided if artificial intelligence-driven equipment is trained based on all possible data from all people around the world. We can hope that this will happen in the near future, but at the moment we do not have enough computational power to conduct such training, so this risk still exists.

Generative AI and Misinformation in Education

The boom of generative AI has revolutionized people's lives. This has had a significant impact on education, where the way people learn has changed dramatically through generative AI models such as ChatGPT and others. Most of these models allow students to get free access to almost any question they are interested in, even if these questions are academic in nature or require justification. This is fantastic because now students don't have to have a big budget to access expensive textbooks, online courses, individual tutors, and college. This provides advantages not only to students who cannot afford expensive education, but also to fairly wealthy people, since their time is money, and generative AI can significantly save their time. Back then, you had to spend a semester or more in college to learn a new skill, but now they can learn it much faster with generative AI at their own pace. However, this convenience comes with big risks that need to be considered. One of the main problems is how accurate the information generated by artificial intelligence is. Due to the bias or incorrectness inherent in any system, AI can generate misleading answers to questions. This may be not only an AI error, but also an incorrect prompt provided by the students. In this case, the answer may seem convincing, but it is inherently wrong. Of course, not everything is perfect in our world, and teachers make mistakes in traditional education, while at the same time providing knowledge. But teachers will probably figure out where they made a mistake before things go too far, which AI can't do because it provides the most likely outcome based on its training. Because misinformation is

spread too widely, a student who has already formed a solid but incorrect knowledge base will face serious difficulties in unlearning and re-learning the correct information from the very beginning. This takes even more time and leads to dangers that we cannot predict. An example would be a student who became a teacher with a misleading knowledge base from generative AI and started teaching other students. This would spread misleading information to students, their families and friends,

as well as to everyone they interact with, like a virus. Putting an end to this spread can be an extremely difficult task.

AI Bias in Banking and Financial Services

One of the greatest achievements of AI is in the banking sector. Most decisions are currently based on data and AI, rather than human inferences. One of the most significant achievements of AI is observed in the risk management departments of banks. Previously, a client would come to the bank to get a loan. They would fill out the application and wait for it to be approved or rejected. On the bank's side, the risk management department would analyze the client's application and make a decision on approving or rejecting the loan based on the client's initial data, which usually would include the credit history, transactions, age, demographic data, etc. This procedure would have taken a long time, since the bank employee would have scanned all this data and set the final credit rating manually, that is, they would have viewed each page, read and analyzed the customer's data. Currently, there is no need to do this manually. Instead, AI engineers create models that can read, analyze, and make a final decision on approving or rejecting a loan. The whole process takes several minutes or even seconds. But at the same time, AI tends to be more biased towards the clients with certain background. For example, if bank customers from a certain region went default on their loans more often than other customers, and this data was stored in the bank's database, then there is a high probability that the scoring model based on artificial intelligence caught this and became extremely conservative towards new customers from this region. As a result, these customers will mostly get their loans rejected, or even if they get approved, the interest they will have to pay will be higher compared to other customers. This is wrong even from a business point of view. The bank is losing good customers who can potentially bring more profit than losses.

RESULT AND DISCUSSION

Social Media and Echo Chamber Effect

Facebook and Instagram are the places where we all, as a rule, interact with artificial intelligence, without even suspecting it. It is almost impossible to find a person who does not check their accounts on FaceBook, Instagram or any other platform on a daily basis. Every time they search

a content they don't see random results. These platforms use "Recommendation Systems", which are advanced algorithms that allow you to recommend content based on a user's personal information, such as age, country of residence, gender, and even their name. Artificial intelligence has significantly improved these "Recommendation Systems." Neural networks are now used to analyze your data and, therefore, recommend content that is most likely to interest you. This is the reason why we mostly get similar content from our social media list. It's interesting because we consume content that we really enjoy and that makes us feel satisfied. However, this has a huge disadvantage. These "Recommendation Systems" become extremely biased towards ourselves. Instead of getting some new information from content that is not in our area of interest, which is very important because we must constantly learn something new in order to be able to withstand any future challenges, we receive similar information day after day. Now imagine a situation where a user initially started viewing similar content from multiple accounts. As a result, the other accounts with similar content will be recommended to this user. What if these accounts usually only show the bad side of any news? Every day, the user will receive only bad news, which can eventually lead to depression and even illness. These biases are extremely difficult to get rid of, because we humans like to see what we want to see, and artificial intelligence is great at providing us with what we want. However, this is very harmful to us.

Ethical Concerns in AI development

As artificial intelligence is increasingly integrated into the most important aspects of society, there are more and more ethical issues related to its development and proper use. The main idea of AI is to mimic a human and make decisions the way a human would, but they often have biases derived from training data. Governments around the world face serious ethical dilemmas.

One of these dilemmas is the ethical responsibility underlying AI. Who should be responsible for the decision made by the AI in case it made a wrong inference? Should it be the developer who trained the AI model, the author of the initial training data published, the scientists and researchers who created the architecture of the AI model, or the AI model itself? In the last case, how should the AI model be asked to take the responsibility for its actions?

Another dilemma is that AI is a "black box," especially when it comes to advanced models that involve neural networks. Even developers cannot always explain the decisions made by these models. Thus, the lack of transparency makes AI unpredictable.

Addressing Bias and Ensuring Ethical AI in the Future.

To reduce bias and meet ethical standards, AI must be trained to work with proper data. The data should be selected appropriately, making it more diverse and representative. This way, the AI

will no longer be biased towards a particular person or aspect. In addition, AI should help people, not replace them. People should consider it as a tool that helps them in their work or daily routine, but does not do their job instead, as human judgment always has the highest priority. Finally, special tools should be used to test AI models and systems for fairness, ethical standards, and bias. All this, therefore, will help solve the problem of ethics and bias in AI.

CONCLUSION

Despite the fact that artificial intelligence is certainly useful for our society, it is important not to forget about its biases and important ethical issues. AI is growing, hence there is a growing risk of encountering stereotypes in various aspects of people's lives. This can lead not only to minor problems, but also eventually turn into a problem that we cannot solve.

Solving these problems is the responsibility not only of AI developers, but also of society, institutions, and governments. To combat bias, it is necessary to train models on diverse data, ensure transparency in the decision-making process and set up the ethical implications.

In the future, we need to develop ethical guidelines for AI as it becomes more sophisticated. The goal is to help people, not replace them. These recommendations will help reduce risks, but at the same time they will not become an obstacle to growth.

REFERENCES

- [1]. “Standards at a glance,” AI Standards Hub, <https://aistandardshub.org/resource/main-training-page-example/4-the-main-stages-of-standards-development/>.
- [2]. Boddington, P. AI Ethics: A Textbook / P. Boddington. — Singapore: Springer, 2023. — (Artificial Intelligence: Foundations, Theory, and Algorithms). — [Electronic resource]. Access mode: <https://doi.org/10.1007/978-981-19-9382-4>.
- [3]. Christian, B. The Alignment Problem: Machine Learning and Human Values / B. Christian. — New York: W. W. Norton & Company, 2020. — 496 p.
- [4]. Belenguer, L. AI Bias: Exploring Discriminatory Algorithmic Decision-Making Models and the Application of Possible Machine-Centric Solutions Adapted from the Pharmaceutical Industry / L. Belenguer. — AI Ethics, Vol. 2, No. 4, 2022, pp. 771-787. — [Electronic resource]. Access mode: <https://doi.org/10.1007/s43681-022-00138-8>.
- [5]. There’s More to AI Bias Than Biased Data, NIST Report Highlights. — National Institute of Standards and Technology (NIST), March 16, 2022. — [Electronic resource]. Access mode: <https://www.nist.gov/news-events/news/2022/03/theres-more-ai-bias-biased-data-nist-report-highlights>.

- [6]. Manyika, J., Silberg, J., Presten, B. What Do We Do About the Biases in AI? / J. Manyika, J. Silberg,
- [7]. B. Presten. — Harvard Business Review, October 25, 2019. — [Electronic resource]. Access mode: <https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai>.
- [8]. Hanna, M. G., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J., Deebajah, M., Rashidi, H. H. Ethical and Bias Considerations in Artificial Intelligence/Machine Learning / M.
- [9]. G. Hanna, L. Pantanowitz, B. Jackson, et al. — Modern Pathology. — [Electronic resource]. Access mode: <https://modernpathology.org/>.
- [10]. Labrague, L. J., AL Sabei, S., AL Yahyaei, A. Artificial Intelligence in Nursing Education: A Review of AI-Based Teaching Pedagogies / L. J. Labrague, S. AL Sabei, A. AL Yahyaei. — Teaching and Learning in Nursing, 2025. — [Electronic resource]. Access mode: <https://doi.org/10.1016/j.teln.2025.01.019>.
- [11]. Gieselmann, M., Erdsiek, D., Rost, V., Sassenberg, K. Do Managers Accept Artificial Intelligence? Insights into the Role of Business Area and AI Functionality / M. Gieselmann, D. Erdsiek, V. Rost, K. Sassenberg. — Journal of Economic Psychology, Vol. 108, 2025, p. 102804. — [Electronic resource]. Access mode: <https://doi.org/10.1016/j.joep.2025.102804>.
- [12]. Lee, J., Su, H., Ji, D.-Y., Minami, T. Engineering Artificial Intelligence: Framework, Challenges, and Future Direction / J. Lee, H. Su, D.-Y. Ji, T. Minami. — arXiv preprint, 2025. — [Electronic resource]. Access mode: <https://arxiv.org/abs/2504.02269>.
- [13]. Shen, Y., Ma, H.-F., Yang, L. Analytical Discovery of Manifold with Machine Learning / Y. Shen, H.-F. Ma, L. Yang. — arXiv preprint, 2025. — [Electronic resource]. Access mode: <https://arxiv.org/abs/2504.02511>.
- [14]. Grezes, F. Reservoir Computing: A New Paradigm for Neural Networks / F. Grezes. — arXiv preprint, 2025. — [Electronic resource]. Access mode: <https://arxiv.org/abs/2504.02639>.
- [15]. Zhuohan Ge, Nicole Hu A Survey of Large Language Models in Mental Health Disorder Detection on Social Media — arXiv preprint, 2025. — [Electronic resource]. Access mode: <https://arxiv.org/abs/2504.02800>.
- [16]. Barbero, F., Arroyo, Á., Gu, X., Perivolaropoulos, C., Bronstein, M., Veličković, P., Pascanu, R. Why do LLMs attend to the first token? / F. Barbero, Á. Arroyo, X. Gu, C. Perivolaropoulos, M. Bronstein, <https://arxiv.org/abs/2504.02572>