IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.569

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 7, July 2021 ||

|DOI:10.15680/IJIRSET.2021.1007183|

Understanding Human Bias in Artificial Intelligence: Suggesting a Future for Being More Objective, And Less Subjective

Mira Kaul[1]

Student, DPS R.K. Puram, New Delhi, India[1]

ABSTRACT: Human life has become easier, dynamic, and versatile with the help of technological advancements of Artificial Intelligence and machine learning. Artificial intelligence not only makes the lives of humans easier by doing their tasks automatically and at a much faster rate but also allows for more productive work so that the human race can work towards progress at a faster rate. However, it must not be forgotten that the systems of artificial intelligence are also built by humans, and humans hold their own characteristics of being prejudiced and discriminatory towards certain groups of fellow humans. The machine indeed gives the output as per the input that is provided to it by the humans. Hence, if the input is biased, the output that is produced isn't completely objective but coloured by the stereotypical attitudes of the creator. This paper delineates some of the human biases that are evident in the systems of artificial intelligence and puts forward some suggestions to make the systems more objective, and less subjective.

KEYWORDS*:* artificial intelligence, human bias, algorithm bias, gender bias, racial bias.

I. INTRODUCTION

The power that artificial intelligence (hereinafter referred to as AI) holds in today's world is remarkable. It is an indispensable asset to every industry. Its influence is seen almost everywhere from the pop-up advertisements on our computer screens to self-driving cars, from economic modelling to healthcare systems, from the google search engine to digital smart assistants like Siri. It provides a basis for all computer learning and is also cost, time, and labour efficient. Its demand has been rising and shall continue to rise in the coming years.

However, it is important to understand that AI is a product of the data and algorithms humans feed into computers. The computers don't have a conscience of their own: their actions are based on the information they're fed. Therefore, any deviation in the information they're fed results in deviations in their actions. Any sort of bias in the data they're fed results in biased actions. This leads us to one of the biggest issues faced by the AI industry which is AI bias.

Bias in AI corrupts well-intentioned projects and has a negative impact on thousands, if not millions, of people across the globe. Therefore, it is necessary for people and companies to understand the deleterious effects of bias in AI and apply the best practices to make AI as fair as possible. A variety of studies have explored AI bias including racial, and gender bias across domains (Feldman & Peake, 2021; Makhortykh et al, 2021; Ntoutsi et al, 2020; Prates et al, 2020; Silberg & Manyika, 2019).

II. WHAT IS HUMAN BIAS IN AI?

AI bias, also known as algorithm bias, is errors that occur during building AI algorithms that result in unfair and biased outcomes. It is a situation in which a computer system due to human errors discriminates between different groups of people. This discrimination could be based on gender, race, age, occupation, etc. (Nelson, 2019).

Humans have a history of being biased towards certain communities. They have conscious and unconscious assumptions on race, gender, occupation, creed, etc. as a result of cultural conditioning and learning which was taught to them through a supremacy culture. They are quick to assume that the word "nurse" refers to a female whereas a word like "mathematician" refers to a male. Such biases are unconsciously instilled in their brains and

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 7, July 2021 ||

|DOI:10.15680/IJIRSET.2021.1007183|

are prominently visible in their actions. Bias in AI is a result of such actions. The data which humans feed the computers are filled with numerous biases which are then reflected in various AI models.

For instance, human fed data may make the AI model come to conclusions like the following:

1. Males are more qualified as business people than females.
2. The chances of a black criminal reoffending are higher than that of a white person.
3. Females are better off for jobs that don't require technical expertise.
4. A white man is more qualified for a skilled job than a black man.

These statements are extremely sexist and racist and are unfortunately instilled in many AI systems due to inaccuracy while framing, collecting, and preparing the data. There are numerous cases in which black men are given a higher recidivism score (a number that determines the likelihood that the criminal will re-offend) simply because of their race and women have been denied jobs simply because of their gender. Such biases pose a significant threat to gender, racial, and ideological equality, which are matters which most countries in the world are still trying to work on. In fact, AI bias may deteriorate the progress these countries have made on racial, gender, and ecological matters in the past years.

AI has a massive impact on the world and has the ability to influence large groups of people. Any ill action taken by AI has a negative impact and influence on society. Therefore, it is vital to address its setbacks.

III. WHY DOES BIAS EXIST IN AI?

Bias is instilled in AI due to various human shortcomings (Kantarci, 2021), which are explained below:

1. Incomplete data: This occurs when the data fed to the computers lack diversity, i.e., the sample is unrepresentative in nature. For instance, if a deep-learning algorithm is fed more photos of light-skinned faces than dark-skinned faces, the resulting face recognition system would inevitably be worse at recognizing darker-skinned faces.

2. Bias in Human Intelligence: This occurs when a computer system is trained alongside already existing human biases, for instance, preferring men over women. The computer system retains these biases.

IV. GENDER BIAS IN AI

AI bias in terms of gender refers to a situation in which a computer system discriminates between different genders (usually preferring men over women). It has the potential to widen the already existing gender gap, which can be threatening to the lives of women. It prefers males for activities that require some form of expertise and prefers females for activities that involve assisting and obeying others.

The following are some real-life examples that depict gender bias in AI:

1. Gender bias in virtual voice assistants: It is not a coincidence that all the major virtual personal assistants like Siri (for Apple), Alexa (for Amazon), or Cortana (for Microsoft) are females. These personal assistants are programmed to only obey the orders of the user. They don't have the right to refuse any of the commands given to them. In fact, they're apologetic to the insults given to them by humans. This clearly enforces the idea that females ought to do whatever they're told to without any questions and that they are secondary as compared to males. Moreover, using assistants for trivial work like making appointments or playing a song is derogatory towards the value of work done by real-life assistants (Adams, 2019).

2. Gender bias in Amazon's recruiting process: Amazon launched an AI project in 2014 which was solely based on reviewing the resumes of job applicants using AI algorithms so that recruiters don't spend time on manual resume screen tasks. However, in 2015 it was found that the AI system was not reviewing the candidates fairly. The AI system preferred males over females for job applicants. This bias existed due to a lack of diversity in the data used. To build this model, amazon used historical data from the previous 10 years. This data was not representative. A

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 7, July 2021 ||

|DOI:10.15680/IJIRSET.2021.1007183|

majority (about 60%) of the people selected from the IT industry were males, so the AI system learned that males were more preferable to work in the IT industry than women, and therefore while reviewing the resumes, the AI system dismissed female applicants.

3. Gender bias in Google translates: As per an Infosys report of 2019, gender bias in google translate is evident when translating between a gender-specific and a non-gender-specific language. It tends to give people like doctors, engineers, mathematicians, etc. a masculine form and people like assistants, nurses, etc. a feminine form. For example, the Turkish form of 'he/she is a doctor gets translated into its masculine form, and 'he/she is a nurse' into its feminine form.

V. RACIAL BIAS IN AI

Racial bias, like gender bias, is long prevalent in our society, which resulted in it being incorporated in our AI systems too. Racial discrimination in AI refers to a situation in which a computer system discriminates between different races, favouring lighter skin tones over darker skin tones.

The following are some real-life examples that depict racial bias in AI:

1. In 2020, a facial recognition system was used on a surveillance camera to recognise a black man stealing watches from a store in Detroit, Michigan. However, the wrong man was arrested. The face recognition system couldn't accurately tell the difference between black faces. This is because while making the model, a majority of the sample population were white faces. Hence, the system learned to accurately tell apart only white people (Pagliaccio, 2020).

2. In 2019, a healthcare prediction algorithm, which is used on more than 200 million US citizens, was designed to predict the level of healthcare required for the patients. However, it was soon found that the algorithm was favouring white patients over black patients. A study revealed that black patients unfairly tended to have a lower risk score than white patients, and had to pay more for the same level of treatment as white patients (Vartan, 2019).

3. In September 2020, Twitter had created an AI algorithm to recognize the most prominent figure in a picture. However, the algorithm had resulted in photo previews that heavily favored White faces over Black faces: the algorithm couldn't capture black faces in the pictures. Ultimately, this face-recognition feature was taken down by Twitter.

4. AI is commonly used in the criminal justice system to flag prisoners which have a high chance of reoffending. All prisoners are given a recidivism score, which indicates the chances of the prisoner reoffending. Soon researchers found that the recidivism scores given to dark-skinned people were generally higher than those given to light-skinned people, even if there wasn't much difference between the profiles of the people except their race (Hao, 2019).

VI. SUGGESTIONS

1. Only the fairness of the people who create it can predict the fairness of AI. It is well established that biases in AI are simply a reflection of the existing human prejudices. Since humans can't ever be expected to be completely unbiased, AI also can't be expected to be completely unbiased (Cremer, 2020; Satell & Abdel-Magied, 2020).

2. While bias in AI can't be completely eliminated, it can most definitely be reduced. In this context, the following are steps are suggested for people building AI models to reduce bias in AI:

3. Diverse people should be building this technology: People of different races, gender, background, etc. should be working on building each model as people from diverse backgrounds will have diverse opinions on how to go about building this technology. Moreover, different types of people will be able to detect different types of biases.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.569|

|| Volume 10, Issue 7, July 2021 ||

|DOI:10.15680/IJIRSET.2021.1007183|

4. Sample data should be more diverse: It is very important to carefully choose the data that is being fed to the computer system. If the sample is diverse, i.e., the people belonging to different races, genders, sexualities, etc., the chances of the computer system getting biased information is less, which leads to fairer outcomes.

5. Evaluate the fairness of the model before the operation: All models undergo an evaluation process before operation. This evaluation is usually based on the precision of the model. Evaluation in terms of bias should also be added to the evaluation process to detect any form of bias in it.

6. Continuously review the models in operation: There should be a method present in each company to keep reviewing the performance of the model in terms of bias to make sure the system is as fair as possible in the long run.

VII. CONCLUSION

The good news about AI is that it is completely within the control of humans. Humans have the power to make right and wise decisions using this technology. They can bring about more equality (racial, gender, and occupational) and opportunities for people all over the world if it is used the right way.

AI can increase the accuracy of work done by humans while being time and labour efficient, create various jobs for people in the IT sector, make life easier for people, and help humans achieve great things which they can't on their own. If different people come together and make the right decisions while building this technology, the limits it can reach are unbounded.

REFERENCES

[1] Adams, R. (September 22, 2019). Artificial Intelligence has a gender bias problem – just ask Siri. The Conversation.
[2] Cremer, D. D. (September 03, 2020). What Does Building a Fair AI Really Entail? Harvard Business Review.
[3] Feldman, T., & Peake, A. (2021). On the Basis of Sex: A Review of Gender Bias in Machine Learning Applications. arXiv preprint arXiv:2104.02532.
[4] Hao, K. (January 21, 2019). AI is sending people to jail – and getting it wrong. MIT Technology Review.
[5] Human Bias in AI. Infosys Limited 2019.
[6] Kartan, A. (June 06, 2021). Bias in AI: What it is, Types & Examples of Bias & Tools to fix it. AI Multiple.
[7] Makhortykh, M., Urman, A., & Ulloa, R. (2021). Detecting race and gender bias in visual representation of AI on web search engines.
[8] Nelson, G. S. (2019). Bias in artificial intelligence. North Carolina medical journal, 80(4), 220-222.
[9] Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M. E., ... & Staab, S. (2020). Bias in data‐driven artificial intelligence systems—An introductory survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1356.
[10] Pagliaccio, S. (November 9, 2020). Understanding Gender and Racial Bias in AI, Part 1. UX Matters.
[11] Prates, M. O., Avelar, P. H., & Lamb, L. C. (2019). Assessing gender bias in machine translation: a case study with google translate. Neural Computing and Applications, 1-19.
[12] Satell, G., & Abdel-Magied, Y. (October 20, 2020). AI Fairness Isn't Just an Ethical Issue. Harvard Business Review.
[13] Silberg, J., & Manyika, J. (2019). Notes from the AI frontier: Tackling bias in AI (and in humans). McKinsey Global Institute (June 2019).
[14] Vartan, S. (October 24, 2019). Racial Bias Found in a Major Health Care Risk Algorithm. Scientific American.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.569

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details