



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 8, August 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.569

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Identifying Subconscious Gender Bias from Human Generated Data for Data Processing in Artificial Intelligence

Prem Ranade¹, Shivam Singh²

U.G. Student, Department of Information Technology, B. K. Birla College of Arts, Science and Commerce
(Autonomous), Kalyan, Maharashtra, India¹

U.G. Student, Department of Information Technology, B. K. Birla College of Arts, Science and Commerce
(Autonomous), Kalyan, Maharashtra, India²

ABSTRACT: In this modern era, everything is being automated and this has become possible because of the further development in the field of Artificial intelligence. The reduction in the price of computing resources has made it possible for AI models to train on very and very large amounts of data within less time. As every good thing has pros, there are cons as well and the same is with AI also. The data which is used for the training model is derived from many sources. One of the sources can be considered as people. Each person is different from one another and the same can be applied to data as well. The more the people, the more the variation in the data can be seen. One of the variations can be seen in the field of bias. There can be a lot of biases present in the data like relating to race, gender, demographic bias, etc. In this paper, we have shown that how much gender bias we have identified in our data and also the opinion of people regarding gender bias in society by using survey. And at the end of the paper, we have shown some measures to mitigate gender bias from one of the sources that are people.

KEYWORDS: Gender Bias, Survey, People, Artificial Intelligence, Data Processing.

I. INTRODUCTION

Nowadays, Artificial intelligence is considered a new and essential trend of this digital world. The importance of this field is growing day by day. Many programmers, developers, and researchers are contributing their knowledge, skills to the evolutionary AI technologies.

Artificial intelligence is used in plenty of different fields like Healthcare, Finance, Gaming, Education, and many more. For working in these fields, programmers have to process the big datasets which are written in Human readable languages. But because of some language faults and the presence of bias in people's minds, critical algorithmic biases can be introduced in AI.

There are many examples of bias in Human readable languages (like English). For example: "That person will die when he jumps in the river". In this English sentence; the person is implicitly considered as a male person without knowing the fact. That person could be a female also but normally we use 'He' which gives first preference to the Male entity.

This type of language fault can lead to gender bias in AI. Reducing gender bias in AI is as important as the Revolution in Artificial intelligence. Before removing or mitigating gender bias, there should be some effective techniques to identify gender bias in datasets.

Identification and mitigation of gender bias become essential tasks in AI. Researchers are finding new technologies and methodologies to contribute to the evolution of this field. And this process will further proceed till our whole AI era will not become completely biased-free.

The organization of the paper is as follows. In Section II, we have discussed some of the related works regarding gender bias in Artificial intelligence, its detection, and mitigation techniques in datasets and algorithms. Section III has a survey part with graphs. Section IV describes the results which we got from our survey. Section V contains the conclusion.



II. RELATED WORK

Artificial intelligence has plenty of sub-domains like NLP (Natural Language Processing), MT (Machine Translation), ML (Machine Learning), etc. When programmers perform Machine translation or any other processing on data which is in human-readable languages (like English), some algorithmic biases can be introduced in AI. For example, in [1] authors have explained how algorithmic bias in AI may discriminate against patients in health care services based on gender, ethnicity, income disparities, and race, etc.

The concept of gender bias is not something new. It has existed since the old era, the comparison done by Susan Leavy et al. [2] between the 19th-century British fiction corpora and the articles from The Guardian (2009-2018) newspaper showed how gender bias existed in the old era and the modern era. [3] showed the existence of gender bias was found in many areas like naming terms used for women, order of naming binomials, adjectives defining men and women, and news articles. In [4], authors have introduced the effect of gender bias on Word clustering, Machine translation, Multilingual applications & other (NLP) Natural language processing tasks in AI. Gender bias was also found in a great domain like Ted Talk ratings which was shown by Acharyya et al. [5] and the authors have also shown that racial bias and gender bias present in the Ted Talk rating dataset were removed by using pre-processing and in processing techniques respectively but post-processing technique was found to be incompatible. Junior et al. [6] by using the First Impressions dataset showed how a person's perception can introduce bias like face attractiveness bias and perceived age bias which can be problematic in personality recognition tasks, this shows the dataset bias. Sometimes pre-trained models also contain bias as shown by Nadeem et al. [7]. But not always it is a fault of an algorithm for bias as shown by Krishnan et al. [8] by finding that the cause of bias in the gender classification algorithms across gender-race groups was found because of the facial morphological structure not because of the architectural differences between the algorithms.

In this era of Artificial intelligence, reducing barriers like data and algorithmic gender bias becomes an essential task. Before removing or mitigating gender bias there should be some effective and possible techniques to detect gender bias from human-readable data. In [9], the authors achieved Automatic detection of Gender issues in Machine translation using BERT-based perturbation, Challenge Dataset & some essential steps to identify sentence pairs. To introduce a form of biases in data, authors from Ji Ho Park et al. [10] have introduced some methodologies like DE, GS, FT on datasets that contain unbiased test sets for each gender. In Rachel Rudinger et al. [11], the authors have used a dataset Wino-gender schema of Wino-grad schema style set to uncover gender bias in coreference resolution. Also using the Wino-MT challenge set which is a concatenation of Wino-gender & Wino-bias co-reference test sets, authors from Stanovsky et al. [12] have introduced the First large-scale multilingual quantitative evidence for gender bias in AI.

It is possible to reduce or remove gender bias in DL & other fields of AI using benchmark dataset, bias mitigation algorithm & pre-processing as shown by [13]. When we perform bias mitigation on data in Machine translation, the overall translation quality of the data can be degraded. To improve the quality of the data as well as to reduce gender bias in MT, the authors from Basta et al. [14] used PreSent & SpeakerId techniques which are implemented in a decoder-based neural MT system. There are other bias mitigation techniques in NLP & MT like EWC & Lattice rescoring approach without degradation in overall translation quality introduced by [15]. By training sets of word embeddings as well as gender-neutral versions with the standard Glove algorithm & Post-process method, the authors from [16] have reduced gender bias in MT and other NLP tasks. Various sources of gender bias and various mitigation techniques in the Natural Processing Technique were shown by Sun et al. [17] to find and mitigate bias from dataset and algorithm. The authors from [18] showed various problems like exclusion, overgeneralization, bias confirmation, topic overexposure, and dual-use along with their solutions. To solve problems related to bias and discrimination we require a cross-disciplinary perspective from fields like ethics, legal and social as shown by Ferrer et al. [19].

III. METHODOLOGY

To show gender bias in society, we created an online survey form using Google Forms. Then we distributed it on social media platform to get responses. Overall, we got 60 responses.

Types of questions we used:

1. Type 1 question was a riddle to check the subconscious gender bias of participants which we took from the Quora platform [20].



A man and his son get into a terrible car crash. The father dies, and the boy is badly injured. In the hospital, the surgeon looks at the patient and exclaims, "I can't operate on this boy, he's my son!" Can you guess who is that Surgeon ?

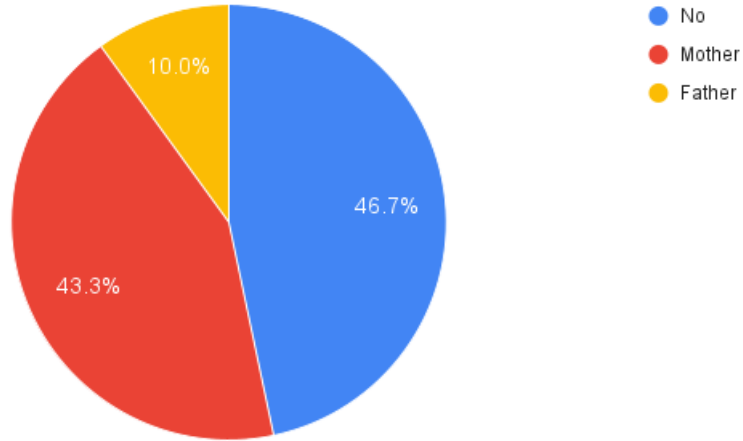


Figure 1

2. In the Type 2 question, we gave some words to the participants and asked them to associate each word with their right options. We took seven words which are cook, rude, arguing, helping, hard worker, teacher, and emotional.

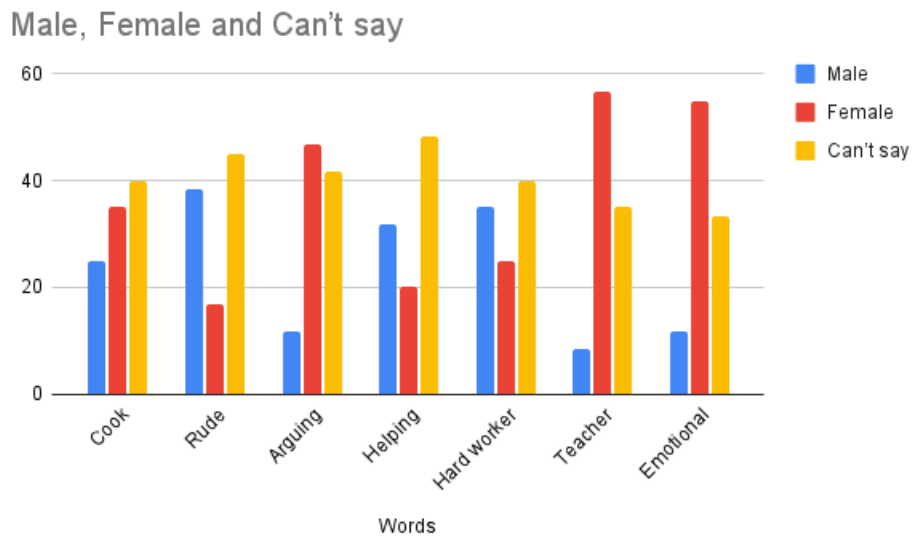


Figure 2

3. In the type 3 question, we asked some questions to get the personal opinion of participants. The questions are given in the below pie charts with their responses.



Do you think gender bias still exists in our society?
60 responses

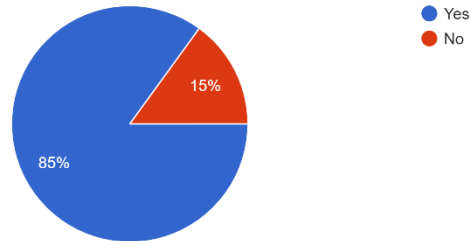


Figure 3a

Have you ever experienced or seen gender bias around you?
60 responses

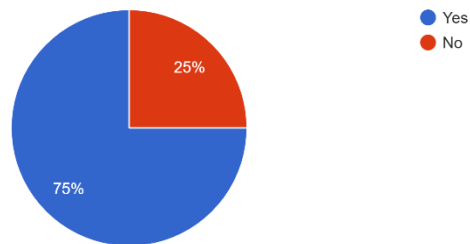


Figure 3b

Have you ever felt biased?
60 responses

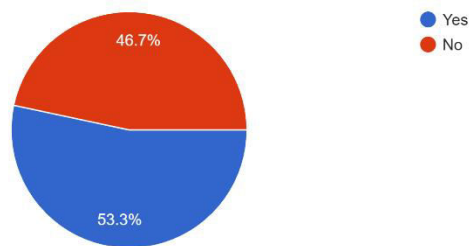


Figure 3c



4. In the type 4 question, we took responses on reasons and solutions of gender bias.

Why does gender bias still exist in society?
60 responses

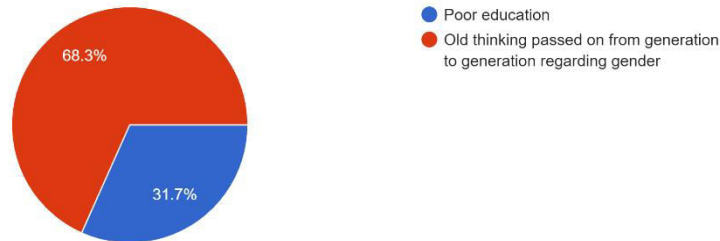


Figure 4a

Which is the best method to reduce gender bias in society?
60 responses

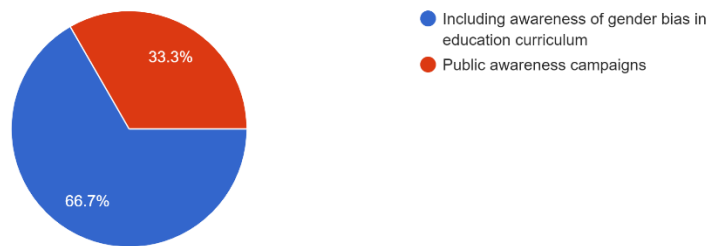


Figure 4b

IV. EXPERIMENTAL RESULTS

From Figure 1, it is observed that nearly 43.3% of participants guessed it correctly that is the answer Mother, 10% guessed Father which shows the presence of gender bias and 46.7% chose No which indicates that they did not even guess, the Surgeon may also be a woman which also shows the presence of gender bias in society.

From Figure 2, a strong bias against females can be observed especially in the fields of Teacher and Emotional. Keeping Can't say option aside, the male is more related with Rude and Hard worker fields. A surprising one is the Teacher field because the word Teacher is generally used for both males and females but in our survey, we found that it is more inclined towards females which shows the presence of gender bias in data.

From the above results, we have identified that the data is biased.

From Figure 3a, the results show that gender bias still exists in our society, and from Figure 3b, we came to know that many participants have seen or experienced gender bias around them. But in Figure 3c, when it was asked personally about gender bias many participants chose no option but it was still less than yes.

From Figure 4a, when asked for the reason, 68.3% of participants selected old thinking passed on from generation-to-generation option, and from Figure 4b, when asked for a solution, including awareness of gender bias in education curriculum got the greater number of responses that is 66.7%.

V. CONCLUSION

Various methods and techniques have been developed for mitigating gender bias in datasets and algorithms like preprocessing of data, in-processing of algorithm, and post-processing of output. These all methods are best and they have paved a path for further development in the field of Artificial intelligence. But they require a lot of time to implement. The processing time of cleaning data and modifying algorithms can be reduced if the data which has been generated from the source like people can be reduced at the first stage only. This means if the people are already aware



of gender bias and have proper knowledge regarding gender bias, the data which is going to get generated from them will always be less biased. In our survey from Figure 4b, we found that most participants chose the option of including gender bias in the education curriculum which can be considered as the best solution because public awareness campaigns might not help that much as most people might not accept the teachings of campaigns as their bias is already strong enough to not get affected but including knowledge of gender bias in education curriculum is going to help us more in mitigating gender bias as when the children are in school from that level, we are trying to make them aware regarding gender bias from their roots which they are going to understand more when they grow up.

Less gender bias in data means the time require for processing and cleaning of data will get reduced which in turn is going to help the researchers to utilize their time for fine-tuning and increasing the accuracy of the algorithms.

VI. ACKNOWLEDGMENT

We would like to thank Prof. Swapna Augustine Nikale, Department of Information Technology, B. K. Birla College Kalyan for providing us with guidance throughout our research work and also to all the participants who gave their responses for our survey.

REFERENCES

- [1] Gregory S. Nelson, "Bias in artificial intelligence," *N. C. Med. J.*, vol. 80, no. 4, pp. 220–222, 2019.
- [2] Susan Leavy, Gerardine Meaney, Karen Wade, and Derek Greene, "Mitigating Gender Bias in Machine Learning Data Sets," in *International Workshop on Algorithmic Bias in Search and Recommendation*, 2020, pp. 12–26.
- [3] Susan Leavy, "Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning," in *Proceedings of the 1st international workshop on gender equality in software engineering*, 2018, pp. 14–16.
- [4] M. R. Costa-jussà, "An analysis of gender bias studies in natural language processing," *Nat. Mach. Intell.*, vol. 1, no. 11, pp. 495–496, Nov. 2019, doi: 10.1038/s42256-019-0105-5.
- [5] R. Acharyya, S. Das, A. Chattoraj, O. Sengupta, and M. I. Tanveer, "Detection and Mitigation of Bias in Ted Talk Ratings," *ArXiv200300683 Cs*, Mar. 2020, Accessed: Jul. 11, 2021. [Online]. Available: <http://arxiv.org/abs/2003.00683>
- [6] J. C. S. J. Junior, A. Lapedriza, C. Palmero, X. Baro, and S. Escalera, "Person Perception Biases Exposed: Revisiting the First Impressions Dataset," in *2021 IEEE Winter Conference on Applications of Computer Vision Workshops (WACVW)*, Waikola, HI, USA, Jan. 2021, pp. 13–21. doi: 10.1109/WACVW52041.2021.00006.
- [7] M. Nadeem, A. Bethke, and S. Reddy, "StereoSet: Measuring stereotypical bias in pretrained language models," *ArXiv200409456 Cs*, Apr. 2020, Accessed: Jul. 21, 2021. [Online]. Available: <http://arxiv.org/abs/2004.09456>
- [8] A. Krishnan, A. Almadan, and A. Rattani, "Understanding Fairness of Gender Classification Algorithms Across Gender-Race Groups," *ArXiv200911491 Cs*, Sep. 2020, Accessed: Jul. 06, 2021. [Online]. Available: <http://arxiv.org/abs/2009.11491>
- [9] H. Gonen and K. Webster, "Automatically Identifying Gender Issues in Machine Translation using Perturbations," *ArXiv200414065 Cs*, Oct. 2020, Accessed: Aug. 19, 2021. [Online]. Available: <http://arxiv.org/abs/2004.14065>
- [10] Ji Ho Park, Jamin Shin, and Pascale Fung, "Reducing gender bias in abusive language detection," *ArXiv Prepr. ArXiv180807231*, 2018.
- [11] Rachel Rudinger, Jason Naradowsky, Brian Leonard, and Benjamin Van Durme, "Gender Bias in Coreference Resolution," *ArXiv180409301 Cs*, Apr. 2018, Accessed: Jul. 08, 2021. [Online]. Available: <http://arxiv.org/abs/1804.09301>
- [12] G. Stanovsky, N. A. Smith, and L. Zettlemoyer, "Evaluating Gender Bias in Machine Translation," *ArXiv190600591 Cs*, Jun. 2019, Accessed: Jul. 12, 2021. [Online]. Available: <http://arxiv.org/abs/1906.00591>
- [13] T. Feldman and A. Peake, "End-To-End Bias Mitigation: Removing Gender Bias in Deep Learning," p. 10.
- [14] C. Basta, M. R. Costa-jussà, and J. A. R. Fonollosa, "Towards Mitigating Gender Bias in a decoder-based Neural Machine Translation model by Adding Contextual Information," in *Proceedings of the The Fourth Widening Natural Language Processing Workshop*, Seattle, USA, 2020, pp. 99–102. doi: 10.18653/v1/2020.winlp-1.25.
- [15] D. Saunders and B. Byrne, "Reducing Gender Bias in Neural Machine Translation as a Domain Adaptation Problem," *ArXiv200404498 Cs*, Jul. 2020, Accessed: Jul. 21, 2021. [Online]. Available: <http://arxiv.org/abs/2004.04498>



- [16] J. E. Font and M. R. Costa-jussà, "Equalizing Gender Biases in Neural Machine Translation with Word Embeddings Techniques," *ArXiv190103116 Cs*, Jun. 2019, Accessed: Jul. 21, 2021. [Online]. Available: <http://arxiv.org/abs/1901.03116>
- [17] T. Sun *et al.*, "Mitigating Gender Bias in Natural Language Processing: Literature Review," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 2019, pp. 1630–1640. doi: 10.18653/v1/P19-1159.
- [18] D. Hovy and S. L. Spruit, "The Social Impact of Natural Language Processing," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Berlin, Germany, 2016, pp. 591–598. doi: 10.18653/v1/P16-2096.
- [19] X. Ferrer, T. van Nuenen, J. M. Such, M. Côté, and N. Criado, "Bias and Discrimination in AI: a cross-disciplinary perspective," *IEEE Technol. Soc. Mag.*, vol. 40, no. 2, pp. 72–80, Jun. 2021, doi: 10.1109/MTS.2021.3056293.
- [20] <https://www.quora.com/A-father-and-his-son-are-in-a-car-accident-The-father-dies-at-the-scene-and-the-son-is-rushed-to-the-hospital-At-the-hospital-the-surgeon-looks-at-the-boy-and-says-I-cant-operate-on-this-boy-he-is-my-son-How-can-this-be>



**Impact Factor:
7.569**



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details