

How Can We Utilize Natural Language Processing to Identify Bias in Job Descriptions?

Nisanth Nimashakavi

July 2023

1 Abstract

In the pursuit of creating fairer hiring practices and promoting workforce diversity, this research project explores the potential of Natural Language Processing (NLP) techniques to identify and rectify biases in job descriptions. The language used in job postings can inadvertently perpetuate biases and deter applicants from underrepresented backgrounds. Leveraging cutting-edge NLP methods, this study aims to automatically detect and address biases, fostering a more inclusive recruitment process. By examining the biases within job descriptions, organizations can attract a more diverse range of applicants and cultivate an inclusive workplace culture. Through the application of NLP, this research seeks to drive positive change in recruitment practices, ultimately contributing to a more equitable job market.

2 Introduction

In today's competitive job market, the language employed in job descriptions plays a pivotal role in shaping the perceptions of potential applicants. While these descriptions serve as the initial point of contact between organizations and prospective employees, they can inadvertently carry biases that reflect deep-seated societal prejudices.[14][1]Such biases, often subtle yet insidious, can deter a diverse range of candidates, particularly those from underrepresented backgrounds, from applying to certain roles. [4]This not only limits the diversity of the applicant pool but also hampers the chances of selection for many deserving candidates.

The implications of biased job descriptions extend beyond individual candidates. They can perpetuate stereotypes, reinforce traditional norms, and hinder the broader goal of achieving workplace diversity. [3]Recognizing the profound impact of language on recruitment, this research endeavors to harness the capabilities of Natural Language Processing (NLP) to detect and rectify biases in job descriptions. The overarching aim is to foster a more inclusive recruit-

ment process, one where every candidate, irrespective of their background, feels welcomed and valued.

The significance of this research lies in its potential to revolutionize hiring practices. By ensuring fairness and inclusivity right from the initial stages of recruitment, organizations can attract a wider spectrum of talent.[7] This not only enhances workforce diversity but also contributes to a richer, more inclusive workplace culture, known to drive innovation and improve overall performance.

However, the challenge of bias detection, especially in textual data, is multifaceted. Within the realm of machine learning, bias detection pertains to the meticulous identification of prejudices, unfairness, or any skewed patterns evident in a model's predictions. These biases are often a reflection of historical imbalances or disparities in training data. [8][10]Delving deeper, the process of bias detection amalgamates both qualitative and quantitative methodologies. It seeks to discern specific terms and language patterns that might inadvertently lean towards or against certain demographic groups. For example, while descriptors like "aggressive" or "dominant" may inadvertently resonate more with male applicants, terms such as "nurturing" or "supportive" could be perceived as favoring female candidates. [6][13][21]

With the advent of advanced NLP algorithms, particularly those trained on expansive datasets, there is newfound optimism in the fight against bias. These algorithms possess the ability to autonomously detect potentially biased terms, either by juxtaposing them against a pre-established lexicon of biased words or by recognizing patterns synonymous with biased language. [11][17]Through this research, we aim to harness these technological advancements, paving the way for a more equitable and inclusive recruitment landscape.

3 Methods

3.1 Data Collection:

The cornerstone of a successful machine learning model is the data upon which it is built. A model's accuracy, reliability, and generalizability are directly influenced by the quality, diversity, and volume of its training data. In the realm of this research, our primary emphasis has been on curating an exhaustive dataset of job descriptions, which serves as a rich tapestry of information, encapsulating various roles, responsibilities, and qualifications across different sectors.To ensure a broad spectrum of data, we opted for the publicly available EMSCAD dataset. [20]This dataset stands out for several reasons:

- **Diversity of Roles:** The EMSCAD dataset encompasses a wide range of job roles, from entry-level positions to senior management roles, ensuring a holistic representation of the job market.[20]
- **Variety of Industries:** It covers multiple industries, including but not limited to, technology, healthcare, finance, and education. This diversity

ensures that our model is not biased towards any particular sector and can generalize across different fields.

- **Comprehensiveness:** Beyond just job titles and basic responsibilities, the EMSCAD dataset delves into detailed job descriptions, providing insights into specific qualifications, skills, and experiences required for each role.[20] This depth of information is crucial for creating a nuanced and detailed machine learning model.
- **Authenticity:** Being publicly available, the dataset has been used and vetted by researchers and professionals worldwide, adding to its credibility and reliability.

3.2 Data Preprocessing:

The process of preparing raw data for machine learning models is a critical step, ensuring that the data is in the right format, is clean, and devoid of any inconsistencies or redundancies. For our research, we employed the following preprocessing steps on the EMSCAD dataset[20]:

- **Removal of Extraneous Data:** The initial step involved pruning the dataset to retain only the job descriptions. This was done by eliminating columns and data points that were not pertinent to our research objectives. By focusing solely on the job descriptions, we ensured that our model would be trained on the most relevant and context-rich information.
- **Token Limitation:** Machine learning models, especially those dealing with natural language processing, often have token limitations due to computational constraints. In our case, to ensure compatibility with existing bias detection models, we filtered out rows containing more than 2024 tokens. This step ensures that the data fed into the model is within its processing capacity, thereby optimizing performance and accuracy.
- **Bias Column Initialization:** Post the cleaning and filtering processes, we introduced a new column named 'bias'. This column is initialized with empty values and is designed to hold binary data - 'true' or 'false'. A 'true' value indicates the presence of bias in the job description, while 'false' denotes its absence. This column serves as a placeholder, set to be populated in the subsequent stages of our research, based on the outputs of our bias detection algorithms.

3.3 Bias Detection Model:

Detecting biases in textual data, especially in sensitive areas like job descriptions, is a task of paramount importance. The subtleties and nuances with which biases can be embedded make it a challenging endeavor. Leveraging the preprocessed EMSCAD dataset[20], we adopted a multi-pronged approach to ensure robust and accurate bias detection:

- **Implementation of Prior Models:** The field of bias detection has seen the development of numerous models, each with its strengths and methodologies. We began by selecting a variety of these prior models, ranging from deep learning architectures to traditional NLP-based classifiers. The idea was to harness the collective power of these diverse models to ensure comprehensive bias detection.
- **Cross-referencing Results:** Instead of relying on the output of a single model, we adopted an ensemble-like approach. After obtaining results from the multiple models, we cross-referenced their outputs. Only the job descriptions flagged as biased by all models were considered to have a high likelihood of containing bias. This method reduces the chances of false positives, ensuring that only the most consistent biases across all models are highlighted.
- **Manual Verification:** While automated models are powerful, human judgment remains an invaluable asset, especially in tasks as nuanced as bias detection. To ensure the validity of our results, the flagged job descriptions were manually reviewed. This step served as a final filter, catching any potential oversights by the automated models and validating the true positives.

4 Literature Review

4.1 WEFE: The Word Embeddings Fairness Evaluation Framework

The WEFE repository introduces a groundbreaking framework dedicated to the evaluation and quantification of fairness within word embeddings. Word embeddings, which have become a cornerstone in modern natural language processing tasks, can inadvertently capture and perpetuate societal biases present in the data they are trained on. The Word Embeddings Fairness Evaluation Framework, or WEFE, is a Python library meticulously designed to address this challenge. It offers a comprehensive suite of metrics and visual tools that empower users to measure and understand potential biases lurking within their embeddings. With its well-structured documentation, user-friendly tutorials, and illustrative examples, WEFE stands as a beacon for researchers and practitioners aiming to ensure the ethical use of word embeddings in their applications. [2]

4.2 HonestyMeter: A Beacon of Truth in the Age of Misinformation

In an era where misinformation can spread like wildfire, the HonestyMeter project emerges as a much-needed solution to evaluate the honesty and credibility of news articles. With the proliferation of news sources and the rise of fake

news, discerning the truth has become increasingly challenging for the average reader. HonestyMeter, through its sophisticated machine learning models, provides an objective platform to assess the trustworthiness of news content. By offering an unbiased measure of news article reliability, it empowers readers to make informed decisions and fosters a more informed and discerning public.[15]

4.3 Bias Detection: Illuminating the Shadows of Bias in Textual Data

Bias, often subtle and insidious, can permeate textual data in ways that are not immediately apparent. The bias-detection project stands at the forefront of efforts to detect, analyze, and understand these biases. Leveraging state-of-the-art machine learning models, this project specializes in identifying potential biases across diverse datasets, with a particular emphasis on news articles. From preprocessing scripts to model training and evaluation tools, the repository provides a holistic toolkit for researchers and practitioners. The overarching mission is clear: to illuminate the biases present in textual data and equip users with the tools to address and mitigate them.[9]

4.4 Can Contextual Biasing Remain Effective with Whisper and GPT-2?

This research delves into the realm of end-to-end automatic speech recognition (ASR) and large language models, specifically Whisper and GPT-2. Despite the vast amounts of training data used for these models, there can still be poor ASR performance for infrequent content words. The study introduces an adapted tree-constrained pointer generator (TCPGen) component for Whisper, combined with a unique training scheme. This approach aims to adjust the final output dynamically without changing any Whisper model parameters. Experiments conducted across three datasets showed a significant reduction in errors on biasing words.[22]

4.5 Balancing Gender Bias in Job Advertisements with Text-Level Bias Mitigation

The article addresses the issue of gender bias in job advertisements. Despite advancements towards gender equality in the labor market, gender segregation and disparities in labor market outcomes persist. The study highlights that merely removing explicit gendered words from job ads isn't sufficient, as certain implicit traits are often associated with specific genders. The authors propose an algorithm that evaluates gender bias in text and offers guidance on debiasing the text by suggesting alternative wording.[17]

4.6 Analysing Gender Bias in IT Job Postings: A Pre-Study Based on Samples from the German Job Market

This research paper focuses on the gender bias present in IT job postings, particularly in the German job market. The study acknowledges the current challenge faced by many industrial nations, known as the "war for talents," where there's a significant competition for skilled workers. The paper emphasizes that while job postings have evolved with the inclusion of logos, images, and videos, text remains the primary source of information for applicants. The study presents an approach to measure and mitigate gender bias in job postings. [5]

4.7 A Machine Learning Approach to Recognize Bias and Discrimination in Job Advertisements

The article delves into the increasing digitization of recruitment processes, particularly the use of job application tracking systems (ATS) for online job advertisements. The authors introduce machine learning models designed to identify and classify biased and discriminatory language in job descriptions. The developed system can pinpoint five major categories of biased language: masculine-coded, feminine-coded, exclusive, LGBTQ-coded, and demographic and racial language. The study underscores the importance of addressing bias during the attraction phase of hiring, promoting the creation of more inclusive job advertisements. [12]

5 Limitations of this paper

- **Limitations of the Methods:** While the methodologies employed in this paper were designed with rigor and thoroughness, it's essential to acknowledge the inherent limitations that come with them. Recognizing these limitations not only provides transparency but also paves the way for future improvements. Here are the key limitations associated with our approach:
- **Human Review Biases:** While manual verification serves as a valuable checkpoint, it's important to note that human reviewers come with their own set of biases, both conscious and unconscious. These biases can influence their judgment, potentially leading to inconsistent or subjective evaluations. Even with training, it's challenging to ensure complete objectivity in human assessments.
- **Potential Biases in Original Models:** The prior models we employed for bias detection were trained on different datasets, potentially under different conditions. If these original datasets contained biases, or if the models were designed with certain assumptions, these biases could be propagated into our current analysis. It's a classic case of "garbage in,

garbage out” – if the input or the tool is biased, the output is likely to reflect that bias.

- **Token Length Limitation:** By filtering out job descriptions with more than 2024 tokens, we inadvertently excluded lengthier, more detailed descriptions from our analysis. These longer descriptions might contain nuanced biases that shorter descriptions don’t, and our decision to exclude them might lead to a lack of comprehensive bias detection.
- **Data Volume Constraints:** The quantity of data plays a crucial role in training robust machine learning models. While the EMSCAD dataset is comprehensive, it’s still a finite set of job descriptions.[20] There’s always a possibility that certain biases, especially those that are rare or industry-specific, might not be adequately represented in our dataset. This limitation can affect the generalizability of our models to real-world scenarios.

6 Conclusion

This research has embarked on an ambitious journey to explore the potential of Natural Language Processing (NLP) in identifying and mitigating biases in job descriptions. Through the utilization of the EMSCAD dataset and the implementation of a multi-pronged approach for bias detection, we have taken significant strides towards fostering a more equitable and inclusive recruitment landscape.[20] Our methodology, which amalgamates both machine learning models and human expertise, aims to offer a robust and comprehensive solution to the pervasive issue of biased language in job postings.

However, it is crucial to acknowledge the limitations of this study. From the inherent biases in human reviews to the potential shortcomings in the original models and data constraints, these limitations offer avenues for future research and refinement. Despite these challenges, the study serves as a foundational step in leveraging technology to combat biases in recruitment processes.

The implications of this research are far-reaching. By addressing biases at the very first point of contact between employers and potential employees, we can significantly impact the diversity and inclusivity of workplaces. This not only enriches organizational culture but also drives innovation and overall performance.

In conclusion, while the journey towards completely unbiased and equitable recruitment is long and fraught with challenges, the advancements in NLP and machine learning offer a beacon of hope. This research contributes to that journey by providing a framework that can be adapted and expanded upon, ultimately aiming for a job market where opportunities are genuinely equal for all, regardless of their background.

References

- [1] E. O. Arceo-Gómez, R. M. Campos-Vázquez, R. Y. B. Salas, and S. López-Araiza. Gender stereotypes in job advertisements: What do they imply for the gender salary gap? 2020. Available online at: <http://conference.iza.org/conferencefiles>.
- [2] Pablo Badilla, Felipe Bravo-Marquez, and Jorge Pérez. Wefe: The word embeddings fairness evaluation framework. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 430–436. International Joint Conferences on Artificial Intelligence Organization, 7 2020.
- [3] S. L. Bem and D. J. Bem. Does sex-biased job advertising “aid and abet” sex discrimination? *J. Appl. Soc. Psychol.*, 3:6–18, 1973.
- [4] M. Bertrand. Gender in the twenty-first century. *AEA Papers Proc.*, 110:1–24, 2020.
- [5] Stephan Bohm, Olena Linnyk, Jens Kohl, Tim Weber, Ingolf Teetz, Katarzyna Bandurka, and Martin Kersting. Analysing gender bias in it job postings: A pre-study based on samples from the german job market. *ResearchGate*, 2020.
- [6] T. Bolukbasi, K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Adv. Neural Inf. Process. Syst.*, volume 29, pages 4349–4357, 2016.
- [7] M. P. Born and T. W. Taris. The impact of the wording of employment advertisements on students’ inclination to apply for a job. *J. Soc. Psychol.*, 150:485–502, 2010.
- [8] A. Caliskan, J. J. Bryson, and A. Narayanan. Semantics derived automatically from language corpora contain human-like biases. *Science*, 356:183–186, 2017.
- [9] Aylin Caliskan. Detecting and mitigating bias in natural language processing, May 2021.
- [10] L. L. Carli. Gender, language, and influence. *J. Pers. Soc. Psychol.*, 59:941, 1990.
- [11] L. Ding, D. Yu, J. Xie, W. Guo, S. Hu, M. Liu, and et al. Word embeddings via causal inference: gender bias reducing and semantic information preserving. 2021. arXiv preprint arXiv:2112.05194.
- [12] Richard Frissen, Kolawole John Adebayo, and Rohan Nanda. A machine learning approach to recognize bias and discrimination in job advertisements. *AI Society*, 38:1025–1038, 2022.

- [13] N. Garg, L. Schiebinger, D. Jurafsky, and J. Zou. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proc. Natl. Acad. Sci. U.S.A.*, 115:E3635–E3644, 2018.
 - [14] D. Gaucher, J. Friesen, and A. C. Kay. Evidence that gendered wording in job advertisements exists and sustains gender inequality. *J. Pers. Soc. Psychol.*, 101:109, 2011.
 - [15] HonestyMeter. A beacon of truth in the age of misinformation, 2023. Retrieved August 24, 2023.
 - [16] L. K. Horvath and S. Sczesny. Reducing women’s lack of fit with leadership positions? effects of the wording of job advertisements. *Eur. J. Work Organ. Psychol.*, 25:316–328, 2016.
 - [17] Shenggang Hu, Jabir Alshehabi Alani, Karen D. Hughes, Nicole Denier, Alla Konnikov, Lei Ding, Jinhan Xie, Yang Hu, Monideepa Tarafdar, Bei Jiang, Linglong Kong, and Hongsheng Dai. Balancing gender bias in job advertisements with text-level bias mitigation. *Big Data*, 5, 2022.
 - [18] Nicholas Meade, Elinor Poole-Dayana, and Siva Reddy. An empirical survey of the effectiveness of debiasing techniques for pre-trained language models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1878–1898, Dublin, Ireland, May 2022. Association for Computational Linguistics.
 - [19] Moin Nadeem, Anna Bethke, and Siva Reddy. Stereoset: Measuring stereotypical bias in pretrained language models, 2020.
 - [20] Amruth Jith Raj V R. Recruitment scam. <https://www.kaggle.com/datasets/amruthjithrajvr/recruitment-scam>, 2021.
 - [21] S. Sczesny, M. Formanowicz, and F. Moser. Can gender-fair language reduce gender stereotyping and discrimination? *Front. Psychol.*, 7:25, 2016.
 - [22] Guangzhi Sun, Xianrui Zheng, Chao Zhang, and Philip C. Woodland. Can contextual biasing remain effective with whisper and gpt-2? *arXiv preprint arXiv:2306.01942*, 2023.
- [20] [9] [1] [3] [4] [6] [7] [5] [8] [10] [11]
 [12] [13] [14] [16] [17]
 [18]
 [19]
 [22]
 [2]
 [15]