Identifying Words in Job Advertisements Responsible for Gender Bias in Candidate Ranking Systems via Counterfactual Learning

Deepak Kumar[1,2], Tessa Grosz[3], Elisabeth Greif[3], Navid Rekabsaz[1,2] and Markus Schedl[1,2]

[1]Johannes Kepler University Linz, Institute of Computational Perception, Multimedia Mining and Search Group, Linz, Austria
[2]Linz Institute of Technology, AI Lab, Human-centered AI Group, Linz, Austria
[3]Johannes Kepler University Linz, Institute for Legal Gender Studies, Linz, Austria

Abstract

Candidate ranking systems (CRSs) for vacancies can pose a significant risk in terms of ethical considerations if they are prone to gender bias or even have legal implications if discriminatory behavior is found. In the case of content-based CRSs, which identify suited candidates for a given job opening based on their resumes and the job advert, gender bias in these texts can also lead to discriminatory behavior of the CRS algorithm. We propose an algorithm to automatically identify gendered words in the job advertisement responsible for gender bias in the rankings. The algorithm determines the words with gendered connotations in the rank distribution for a given job advertisement using content-based job-candidate matching based on the actual biography of a candidate and a counterfactual version in which explicit gender-mentioning terms are swapped between male and female. To this end, we employ the neural network explainability method of integrated gradients to compute CRS's association of the job advertisement words with the gender of candidates, which we call the bias score of words. At the core of our CRS is a cross-encoder architecture. To showcase and validate our approach, we conduct a study investigating the gendered words identified by the proposed algorithm in job advertisements from a private dataset and biographies from the BIOS dataset. We analyze the gendered words along multiple job categories and different linguistic categories. Finally, we statistically and qualitatively compare them with standardized lists manually created by social psychologists to contrast the gender associations CRSs make with human associations.

Keywords

Candidate Ranking, Gender Bias, Explanation

1. Introduction and Background

Candidate selection for jobs has become a very difficult task for human recruiters to complete due to the vast amount of applicants for a job advertisement. This has led to the usage of candidate ranking systems (CRSs). As cases of gender discrimination have been observed in human recruiters [1], the introduction of CRS was believed to be a worthwhile antidote [2]. By now, however, bias in ranking systems is well documented and researched [3]. A real-world application concerns a CRS developed by Amazon, which was promptly discarded when its hiring decisions evidenced gender discrimination [4]. Further empirical evidence of bias in CRSs can be found all over the industry [5].

CRSs commonly leverage content like resumes and job advertisements [6], as in content-based recommendation, knowledge-based recommendation, or hybrid approaches. When dealing with text content, a large language model (LLM) such as BERT [7] is often used for processing. The use of LLMs for encoding textual information is very effective, but they are prone to gender bias [8]. The bias found in LLMs is essentially the association in the embedding space of LLMs of words in textual content with gender concepts represented by gender-identifying words. Hence, the gender-correlated words identified in the LLMs embedding space might not have any gender tones. In this work, we want to study these words for the gender bias occurring in CRSs. However, the words can also have gender targets, i.e., words in job advertisements written to attract certain gender candidates. The words in job advertisements with gender targets have been studied in social psychology [9, 10]. Furthermore, the lists by social psychologists, which capture the implicit bias of words towards the gender of candidates, have been evaluated in the literature for their correlation with candidate-perceived gender bias [11].

In contrast to these expert annotations, in this paper, we aim at identifying gender target words in job advertisements from the CRS's perspective instead of that of a human. More precisely, we want to identify words in job advertisements that cause CRSs to rank candidates dif-

RecSys in HR'23: The 3rd Workshop on Recommender Systems for Human Resources, in conjunction with the 17th ACM Conference on Recommender Systems, September 18–22, 2023, Singapore, Singapore.
deepak.kumar@jku.at (D. Kumar); elisabeth.greif@jku.at (T. Grosz); elisabeth.greif@jku.at (E. Greif); navid.rekabsaz@jku.at (N. Rekabsaz); markus.schedl@jku.at (M. Schedl)
0000-0002-4828-8328 (D. Kumar); 0000-0002-9072-9641 (T. Grosz); 0000-0003-4096-8510 (E. Greif); 0000-0003-1706-3406 (M. Schedl)

ferently only because of their gender. Our contribution is two-fold:

- We present an algorithm to identify the words in job advertisements causing gender bias in the candidate ranking.
- We analyze the words identified by our algorithm, and compare them with a list of words from previous studies curated by social psychologists.

This work can help us understand the distinction between the human and CRS's view of a bias-free job advertisement. Both perspectives are essential as the CRS's perspective will help us reduce algorithmic gender discrimination while the human perspective will help us create a job advertisement desirable to both male and female job seekers. Furthermore, the comparison between the two views may help us better understand the distinction or similarity in the potentially discriminatory nature of a job advertisement for CRSs and humans. As we will see, seemingly gender-neutral job advertisements can inadvertently lead to discriminatory practices by CRSs. Additionally, the use of biased language in job ads may not necessarily result in discriminatory behavior by CRSs. It is essential to remain mindful of these potential issues in order to cultivate a more inclusive and equitable hiring process.

The structure of the remaining paper is as follows: In Section 2, we introduce the CRS used in this work and explain our algorithm to identify words in job advertisements causing gender bias in the rankings of the CRS. Subsequently, in Section 3, we describe the dataset used in the experiments. Section 4 describes the setup of the experiments, and their results are presented and discussed in Section 5. Finally, Section 6 summarizes our work and gives directions for future research endeavors.

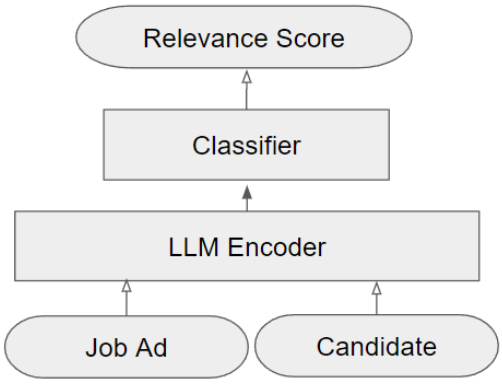

Figure 1: Cross-encoder as CRS model

2. Method

2.1. Candidate Ranking System

We use an LLM-based cross-encoder [12] as our CRS model to rank the candidates for given job ads. The architecture is shown in Figure 1. Our CRS takes a job-candidate pair as input and outputs the relevance score used for ranking the candidates ($RelevanceScore \in [0, 1]$). During inference, we rank the top 1000 candidates with bag-of-words based BM25 model [13] and then re-rank these 1000 using our CRS.

Algorithm 1 Red-word identification

Require: Gender is binary $g \in \{1, 0\}$
Given: Job ad j, and recommended candidate list C
Given: Trained ranking model M. $M(< j, c >)$ is $RelevanceScore$ used for ranking c, $c \in C$
Given: f to create gender counterfactual candidate. $f(c)$ is candidate c with opposite gender
Given: Bias threshold θ.
Job Ad $A_j : A_j = \{w_1, ..., w_n\}_j$, consist n many tokens
Integrated gradient $IG : IG(w_i, <j, c>, M)$
Bias score $S : S = \{s_1, ..., s_n\}$
Temporary bias score $T : T = \{t_1, ..., t_n\}$
$s_i \leftarrow 0 \forall s_i \in S$
$RedWords = \{\}$
for each $c \in \mathscr{C}$ **do**
  $t_i \leftarrow 0 \forall t_i \in T$
  for each $w_i \in j$ **do**
    $t_i \leftarrow |IG(w_i, <j, c>, M) - IG(w_i, <j, f(c)>, M)|$
  end for
  $T \leftarrow SoftMax(T)$
  for each $t_i \in T$ **do**
    if $t_i < \frac{1}{n}$ **then**
      $t_i \leftarrow 0$
    end if
  end for
  $S \leftarrow S \oplus T * \frac{1}{\log(rank(c) \in C+1)}$
end for
for each $s_i \in S$ **do**
  if $s_i > \theta$ **then**
    $RedWords.insert(w_i)$
  end if
end for
return $RedWords$

2.2. Job Ad Words for CRS Gender Bias

We approach the first contribution by identifying the words in a job advertisement responsible for gender bias in the ranking of candidates. For this purpose, we create for each candidate, an artificial gender-counterfactual

candidate. We do this by replacing gendered words, pronouns, and names in the candidate's textual materials (e.g., CV, biography, job-portal profile) with the corresponding word of the opposite gender.[1] Thus, for instance, "he" is changed to "she", and "hers" becomes "his".

Hereafter, we call the words in the job ad affecting the rank of candidates only based on the explicit mention of gender in the candidate's content, **red-words**. Red-words are identified by using Algorithm 1, which tries to find words in job advertisements salient for the difference between the relevance score for the original candidate and its gender-counterfactual candidate. To find the salience and assign bias score to words in the job advertisements, we use machine learning explainability methods. We use gradient-based CRS, hence, we choose the integrated gradient explainability method [14]. [2]

Integrated gradient of a word w_i in job ad j with candidate c and ranking model M is

$$IG(w_i, <j, c>, M) = \{(w_i - w_i') * \sum_{k=1}^{m} \frac{\delta M(<j', c'> + \frac{k}{m} * (<j, c> - <j', c'>))}{\delta w_i} * \frac{1}{m} \quad (1)$$
$$, w_i, w_i' \in j \text{ and } k \in \{1, ..., m\}\}$$

where m is the number integral approximation steps and $< j, c >$ is the original input to the model M, $< j', c' >$ is the same size masked input (i.e., for the model it is a blank job advertisement and candidate content of the same length as original input)

Algorithm 1 is used to identify words in job ad A_j that contribute to the difference in the relevance score of a candidate c and its gender counterfactual $f(C)$ by a trained CRS model M. Integrated gradient IG is used to identify the contribution T of words towards the difference, and we further use $SoftMax$ to normalize T. We scale the normalized T based on the rank of candidate c in the ranking by M. Finally, we use a bias threshold θ to cut off the less critical red-words. Counterfactual candidate creating transformation f replaces all the candidate's nouns and pronouns with that of the opposite gender. For simplicity, in this work, we restricted the transformation f to a binary behavior.

3. Dataset

We created the dataset for the experiment using biographies from BIOS dataset [16] and job advertisements from a private dataset from UK job portals. Firstly, to create the dataset, we employ an exact matching algorithm between the current job mentioned in the biography and the job for which the job advertisement is advertised to get the binary matching/relevancy ground truth labels. The size of the dataset created by matching is 2775 job posts and 322,337 biographies covering 24 different jobs. Of these 24 jobs, 10 have a job advertisement frequency of less than 5 and are removed. Further, a subset of the BIOS dataset is created such that the subset is balanced according to the job and gender of candidates. The size of the balanced subset of biographies is 1400, where each profession has 50 male and 50 female biographies. Thereafter, we split the job advertisement into train, test, and validation sets with stratification of job titles by 70:20:10 split, respectively. So, finally, we are left with 14 different job titles, 1400 biographies with 50 males and 50 females of each job title, and 2085 job advertisements with job distributions shown in Table 1

Table 1
Job title distribution in train, test, and validation set.

Job Titles	#Train	#Test	#Validation
software engineer	365	104	53
senior software engineer	284	81	40
dentist	195	56	28
accountant	113	33	0
teacher	106	31	0
architect	102	29	15
nurse	99	28	0
paralegal	67	19	10
painter	49	14	7
psychologist	23	7	3
personal trainer	16	5	2
dietitian	16	4	2
interior designer	13	3	2
photographer	11	3	0

The biographies have been pre-processed by replacing real names with "bob" for males and "alice" for females. Additionally, counterfactual biographies have been generated by replacing gender-specific words with those of the opposite gender. The male-coded words used are "bob," "mr," "his," "he," "him," and "himself," while the female-coded words are "alice," "mrs," "hers," "she," "her," and "herself."

4. Experiment Setup

Experiments are conducted using a BERT-based cross encoder, i.e., CRS, over our collection of job advertisements and biographies of candidates (Section 3). CRS is trained for four epochs using the sigmoid variant of binary cross entropy loss[3] on our collection. We report ranking performance in terms of nDCG and bias in terms of true positive rate parity (TPRP) [17, 18]. TPRP in candidate recommendation for binary gender attribute

[1] We consider binary gender here. In the non-binary case, any gender other than the original can be considered the opposite gender.
[2] For non-gradient-based CRS explainability, SHAP [15] method can be used.
[3] https://pytorch.org/docs/stable/generated/torch.nn.BCEWithLogitsLoss.html, access: July 2023

Table 2
Most frequent words for each job title. Except for the last row, all words are red-words, i.e. obtained using BERT-based CRS and Algorithm 1. The last row contains examples from the expert-list.

Job	Top cleaned red-words
senior software engineer	software, senior, engineer, development, team, engineering, experience, design, code, java
software engineer	software, engineer, team, development, experience, technology, engineering, data, code, engineers
dentist	dental, dentist, practice, associate, nhs, care, patients, clinical, private, patient
paralegal	legal, para, team, firm, law, litigation, client, property, role, commercial
nurse	nurse, nursing, nurses, residents, home, training, registered, clinical, shifts, team
teacher	school, pupils, teaching, teachers, children, teacher, students, staff, schools, curriculum
architect	architect, projects, design, architectural, practice, residential, team, working, architects, experience
accountant	accountant, accounting, accounts, management, tax, finance, audit, reporting, business, experience
painter	painter, decor, painters, painting, looking, shift, working, refurbishment, email
Expert-list examples	lead, depend, support, logic, principle, depend, understand, active, child, superior

Table 3
Most frequent adjectives in red-words for each job title and examples from expert-list.

Job	Adjective red-words
senior software engineer	successful, flexible, strong, able, financial, new, creative, electronic, digital, continuous
software engineer	new, flexible, critical, strong, responsible, able, angular, scientific, successful, desirable
dentist	corporate, flexible, competitive, available, digital, clinical, highest, able, legal
paralegal	corporate, legal, financial, successful
nurse	available, clinical, residential, competitive
teacher	available, successful, able
architect	architectural, residential, talented
accountant	financial, statutory, relevant, reactive, desirable, hard
painter	internal
Expert-list examples	responsible, competitive, supportive, analytical, ambitious, confident, competent

$gender \in \{male, female\}$ and recommendation list Q_{A_j} for a job advertisement A_j is defined as

$$TPRP(A_j) = |P(c \in Q_{A_j}|gender = female, \rho(A_j, c) = 1) - P(c \in Q_{A_j}|gender = male, \rho(A_j, c) = 1)| \quad (2)$$

where $\rho(A_j, c) = 1$ implies that a candidate c sampled from the candidate set C is suitable for job advertisement A_j. Furthermore, $TPRP = 0$ implies that equal opportunity [19] fairness condition is achieved.

We used Algorithm 1 to create three lists of red words with different bias thresholds: 0.05, 0.02, and 0.002. It's worth noting that the likelihood of randomly selecting a token from CRS's input is about 1/512 or 0.002. For comparison, We use the word list by social psychologist [10] as our "expert-list". We also identified the parts of speech for the words in both the red-word and expert lists using NLTK wordnet [20].

5. Results and Discussion

On the test set for candidate ranking, the CRS achieves a score of 0.82 nDCG@10. While the ideal nDCG score is 1, this performance is still considered decent [21]. However, the result is significantly biased, as the TPRP score shows. The average of TPRP over all job advertisements is 0.326, and according to Equation. 2, the ideal value of TPRP is 0. The top red-words according to the bias score of tokens (S in Algorithm 1) for each job are shown in Table 2 after removing punctuation, stopwords, numbers, and words with less than three letters. The majority of words here are relevant for identifying the job, unlike the terms in the expert-list examples which are not related to any specific job but are rather generic. This behavior is expected as CRS's training objective needs to focus on job-related words and, as a result, will get affected more by bias due to these terms. Contrary to this, the expert-list focuses on words describing the properties of the candidate, and hence, humans are more likely to associate them with the candidate's gender. The red-words for jobs with less than

Table 4
Red-word distribution for each job using BERT-based CRS. J is set of all job ads of a particular job title. $RedWords_j$ and $ExpertWords_j$ are words found in job ads by Algorithm 1 and by expert-list words respectively.

Job Title	$\frac{\sum_{j\in J} RedWords_j}{\lvert J\rvert}$			$\frac{\sum_{j\in J} RedWords_j \bigcap ExpertWords_j}{\lvert J\rvert}$			$\frac{\sum_{j\in J} ExpertWords_j}{\lvert J\rvert}$	$\lvert J\rvert$
	$\theta = 0.05$	$\theta = 0.02$	$\theta = 0.002$	$\theta = 0.05$	$\theta = 0.02$	$\theta = 0.002$		
software engineer	6.0	12.7	32.5	0.2	0.3	0.6	4.5	104
senior software engineer	5.8	13.2	35.3	0.1	0.3	0.7	4.8	81
dentist	5.3	9.3	20.8	0.2	0.3	0.4	3.7	56
accountant	6.3	14.7	41.2	0.1	0.3	1.0	3.9	33
teacher	5.6	14.7	44.7	0.2	0.5	1.3	5.7	31
architect	5.9	14.2	42.8	0.1	0.4	1.3	3.3	29
nurse	6.7	15.3	43.2	0.1	0.3	1.1	5.3	28
paralegal	6.5	16.8	53.2	0.1	0.3	1.4	4.0	19
painter	7.5	19.4	51.1	0.1	0.4	1.1	1.9	14
psychologist	6.0	22.3	71.6	0.4	0.7	2.6	5.6	7
personal trainer	4.4	16.6	85.6	0.0	0.0	1.8	2.2	5
dietitian	4.8	20.5	80.5	0.0	0.8	2.5	5.8	4
interior designer	3.7	17.0	110.7	0.0	0.0	1.0	1.3	3
photographer	6.0	19.7	69.3	0.0	0.7	0.7	1.7	3

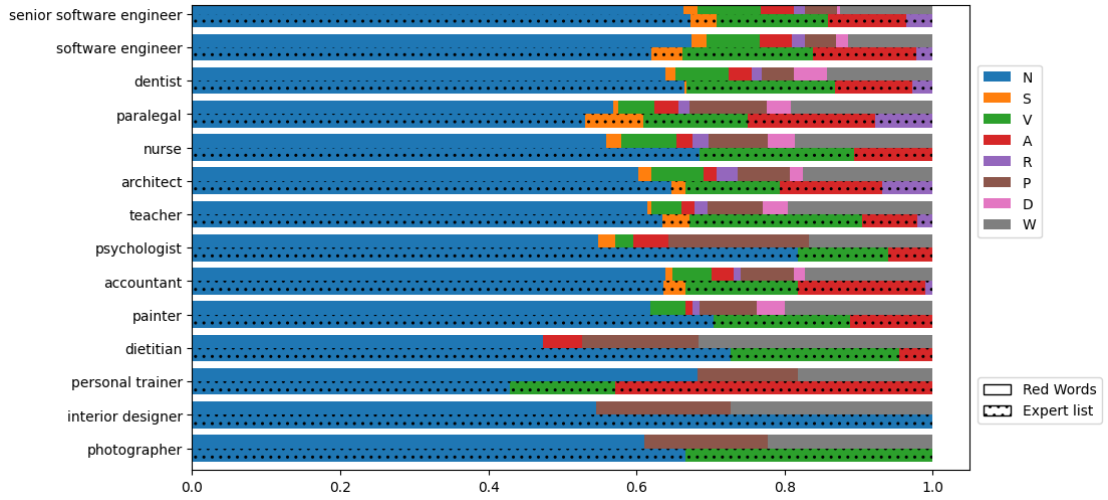

Figure 2: For $\theta = 0.05$, distribution of parts of speech (POS): noun (N), verb (V), adjective (A), adjective satellite (S), adverb (R), digits (D), stopwords (W), and unknown parts (P).

15 job advertisements in the test dataset are not reported. We observe later that with the increase in the number of job advertisements, the curated red-words exhibit less variance.

On average, job advertisements contain about 250-400 words. Table 4 displays the average number of red-words per job for each profession. With the relaxation of bias threshold value θ, the number of red-words increases, and also, the number of common words with the expert-list increases. Here, with the increase in the number of job advertisements, the red-words for more relaxed θ values show less variability.

The common words between the expert-list and red-words are few, which aligns with the observation from Table 2, i.e., both lists of words are different from each other. The number of words given by red-words for $\theta = 0.05$ and the expert-list are close, although, the common words between them are very low. We compare their parts of speech distribution in Figure 2 to better understand their composition. The expert-list does not contain any digits, stopwords, and unknown parts, removing these from red-words will make its distribution

Title: Senior Software Engineer Description: Requirements: Strong knowledge of the .NET framework Strong knowledge of TypeScript and React 6+ years of proven relevant work experience in a similar role Experience with Azure and ARM Templates Strong experience in SOLID OO, enterprise integration skills, and microservice architecture skills Strong understanding of BDD, TDD, and SOLID Strong experience with web application development and deployment Strong knowledge of continuous integration processes and pipelines Excellent verbal and written English skills If you have .NET experience and have strong experience with the skill set above, and the role looks like a great fit, then please send your updated CV to [redacted] and give me a call on [redacted] to discuss your application in further detail. **** INTERVIEW IMMEDIATELY - FAST OFFER **** Role: Senior Software Engineer Salary: Up to £62,550 Location: Reading - Mostly Remote

(a) Senior Software Engineer

Title: Clinical sister Description: Do you want to be part of an award winning and dynamic social enterprise that: is renowned for providing high quality care and is ranked 'Good' by the CQC; is a for-better-profit organisation, reinvesting any surplus back into our health and care services and our local community; is friendly, ambitious, welcomes innovation and rewards excellence; offers superior benefits; everything you get in the NHS and more; and whose achievements reflect the passion, dedication and commitment demonstrated by our staff across all services? Our vision is to be a successful, vibrant, community interest company that benefits the communities we serve. So if you want be a part of this, we would love to hear from you. Job overview Are you an experience nurse looking for a new challenge? Are you passionate about supporting patients holistically to leave hospital earlier? This vacancy is available for part time and full time hours. If you answered yes, we've got the perfect role for you! Medway Community Healthcare are developing our inpatient rehabilitation services and have a new project and we want you involved! We have recently opened up Harmony House and as a part of our Inpatient services offer you the opportunity to work across our sites including Amherst Court Endeavour suite (Stroke rehabilitation) and Britannia Suite (Intermediate care rehabilitation). You will work as part of a well-established multidisciplinary team supporting patients on their road to recovery following hospital admission or step up from the community. You will develop your skills around rehabilitation, discharge planning and specialist inpatient care. The successful candidates will be required to work on a rota basis including nights, weekends and bank holidays. Applicants can indicate a preference of stroke or intermediate care on their applications in the supporting statement but this will also be discussed at interview. Main duties of the job To work as a key member of the Intermediate care and stroke multidisciplinary team to deliver patient-centred nursing and rehabilitation services. The post holder will work across our sites delivering intermediate care and rehabilitation including Endeavour ward (Amherst Court stroke rehabilitation), Britannia ward (Amherst Court Intermediate care rehabilitation) and Harmony House (Intermediate care rehabilitation & step

(b) Nurse

Figure 3: Sample job advertisement: red-words marked by red-color, expert-list words marked by green color for $\theta = 0.05$.

similar to the expert-list for jobs with more than 15 job advertisements. The prevailing part of speech in both the red-words and the expert-list is observed to be the noun, followed by the verb, adjective, adjective satellite, and adverb. Although the two lists have different words (common words are few), the parts of speech distributions are very similar after removing digits, stopwords, and unknown parts. The existence of digits, stopwords, and unknown parts can only be justified due to their relationship with more meaningful words and needs further investigation. The small overlap between the expert-list and red-words observed in the Table 2 can also be seen in the Table 3. Table 3 presents the most frequent adjective red-words for each job title and overall most frequent expert-list adjectives. Here, the top 2 expert list adjectives, "responsible" and "competitive", also appear in the red-words of "software engineer", "dentist", and "nurse". Some other words common between both lists are "support", "commit", "child", and "principle". These words do not show any specific difference from other words in expert-list and might not appear in a more refined red-list. But anything conclusive cannot be deduced from these common words and require further investigation.

Sample job advertisements are shown in Figure 3. Here, red-words are highlighted with red color, and expert-list words are highlighted with green color. As can be seen, the two methods highlight different types of words based on their association with either the job or the candidate. In the example of the "Senior Software Engineer" shown in Figure 3a, the words in red are mainly related to the job of a software engineer, while the word "understanding" from the expert-list is a description of candidate and is not associated with the job of a software engineer. Similarly, in the example of "Nurse" (Figure 3b), the red-words "care" and "rehabilitation" aligns with the job of the nurse. The expert-list in this example describes the employer and the candidate because the same terms can be used to describe both candidate and the employer. Here, the term "ambitious" can be used for both the employer (ambitious care house) and the candidate (ambitious nurse). Both

examples (see Figure 3) and Table 4 confirm that gender-biased wording of a job advertisement is quite distinct from the words causing bias in CRSs. Hence, to come a bit closer to our goal of a bias-free recruitment process, we have to give attention to both red-words and expert-list.

6. Conclusion and Future Work

We present an algorithm to create a list of words from job advertisements, which CRS associate with the gender of candidates. In contrast to a well-established list generated by social psychologists, this list addresses CRSs' gender bias instead of the perceived gender bias of experts. So, a gender bias-free wording of a job advertisement is different for an expert and a CRS leveraging LLMs, and this distinction should be kept in mind while debiasing the candidate selection process. Although expert-list and red-words contain different words, their composition is similar in terms of parts of speech distribution.

As for future work, we plan to investigate more thoroughly possible similarities and differences between the two lists of words. Further, we want to understand how both lists affect the debiasing of the candidate selection process. Also, we plan to improve our algorithm after understanding the reasons for the existence of digits, stopwords, and unknown parts of speech in red-words.

Acknowledgments

This research is funded by the Austrian Science Fund (FWF): DFH-23 and P33526; and by the State of Upper Austria and the Federal Ministry of Education, Science, and Research, through grants LIT-2020-9-SEE-113 and LIT-2021-YOU-215.

References

[1] I. P. Levin, R. M. Rouwenhorst, H. M. Trisko, Separating gender biases in screening and selecting candidates for hiring and firing, Social Behavior and Personality: an international journal 33 (2005) 793–804.

[2] K. A. Houser, Can ai solve the diversity problem in the tech industry: Mitigating noise and bias in employment decision-making, Stan. Tech. L. Rev. 22 (2019) 290.

[3] G. K. Patro, L. Porcaro, L. Mitchell, Q. Zhang, M. Zehlike, N. Garg, Fair ranking: a critical review, challenges, and future directions, in: 2022 ACM Conference on Fairness, Accountability, and Transparency, 2022, pp. 1929–1942.

[4] J. Dastin, Amazon scraps secret ai recruiting tool that showed bias against women, in: Ethics of data and analytics, Auerbach Publications, 2018, pp. 296–299.

[5] J. Sánchez-Monedero, L. Dencik, L. Edwards, What does it mean to 'solve' the problem of discrimination in hiring? social, technical and legal perspectives from the uk on automated hiring systems, in: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 458–468. URL: https://doi.org/10.1145/3351095.3372849. doi:10.1145/3351095.3372849.

[6] M. N. Freire, L. N. de Castro, e-recruitment recommender systems: a systematic review, Knowledge and Information Systems 63 (2021) 1–20.

[7] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423. doi:10.18653/v1/N19-1423.

[8] N. Rekabsaz, S. Kopeinik, M. Schedl, Societal biases in retrieved contents: Measurement framework and adversarial mitigation of bert rankers, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 306–316.

[9] S. L. Bem, D. J. Bem, Does sex-biased job advertising "aid and abet" sex discrimination? 1, Journal of Applied Social Psychology 3 (1973) 6–18.

[10] D. Gaucher, J. Friesen, A. C. Kay, Evidence that gendered wording in job advertisements exists and sustains gender inequality., Journal of personality and social psychology 101 (2011) 109.

[11] S. Tang, X. Zhang, J. Cryan, M. J. Metzger, H. Zheng, B. Y. Zhao, Gender bias in the job market: A longitudinal analysis, Proceedings of the ACM on Human-Computer Interaction 1 (2017) 1–19.

[12] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 3982–3992.

[13] J. Lin, X. Ma, S.-C. Lin, J.-H. Yang, R. Pradeep, R. Nogueira, Pyserini: A python toolkit for reproducible information retrieval research with sparse and dense representations, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 2356–2362.

[14] M. Sundararajan, A. Taly, Q. Yan, Axiomatic attribu-

tion for deep networks, in: International conference on machine learning, PMLR, 2017, pp. 3319–3328.
[15] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Advances in neural information processing systems 30 (2017).
[16] M. De-Arteaga, A. Romanov, H. Wallach, J. Chayes, C. Borgs, A. Chouldechova, S. Geyik, K. Kenthapadi, A. T. Kalai, Bias in bios: A case study of semantic representation bias in a high-stakes setting, in: proceedings of the Conference on Fairness, Accountability, and Transparency, 2019, pp. 120–128.
[17] A. N. Carey, X. Wu, The causal fairness field guide: Perspectives from social and formal sciences, Frontiers in Big Data 5 (2022) 892837.
[18] C. Rus, J. Luppes, H. Oosterhuis, G. H. Schoenmacker, Closing the gender wage gap: Adversarial fairness in job recommendation, in: 2nd Workshop on Recommender Systems for Human Resources, RecSys-in-HR 2022, CEUR-WS, 2022.
[19] M. Hardt, E. Price, E. Price, N. Srebro, Equality of opportunity in supervised learning, in: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 29, Curran Associates, Inc., 2016. URL: https://proceedings.neurips.cc/paper_files/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf.
[20] C. Fellbaum, WordNet: An electronic lexical database, MIT press, 1998.
[21] Y. Wang, L. Wang, Y. Li, D. He, T.-Y. Liu, A theoretical analysis of ndcg type ranking measures, in: Conference on learning theory, PMLR, 2013, pp. 25–54.