ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
Special Issue, https://doi.org/10.1590/1678-6971/eRAMD240056
Submitted: Apr. 2, 2024 | Approved: July 11, 2024

Gender bias in large language models: A job postings analysis

Viés de gênero em *large language models*: Uma análise em descrições de emprego

Guilherme Guimarães Nomelini[1] and Carla Bonato Marcolin[2]

[1] Juniper Creates, Toronto, Ontario, Canada
[2] Federal University of Uberlândia, Uberlândia, MG, Brazil

Authors' notes

Guilherme Guimarães Nomelini is now the data analytics and reporting manager of Juniper Creates; Carla Bonato Marcolin is now a full professor in the School of Business and Management at Federal University of Uberlândia (UFU).

Correspondence concerning this article should be addressed to Carla Bonato Marcolin, Avenida João Naves de Ávila, 2121, 1F, Santa Mônica, Uberlândia, Minas Gerais, Brazil, ZIP code 38400-902. E-mail: cbmarcolin@gmail.com

To cite this paper: Nomelini, G. C., & Marcolin, C. B. (2024). Gender bias in large language models: A job postings analysis. *Revista de Administração Mackenzie*, *25*(6). 1–27. https://doi.org/10.1590/1678-6971/eRAMD240056

RAM does not have information about open data regarding this manuscript.
RAM does not have permission from the authors or evaluators to publish this article's review.

Abstract

Purpose: This article aims to evaluate potential gender biases on job postings.

Originality/value: The management literature still seeks to understand the different social and economic impacts that the introduction of large language models (LLM) has produced. It is now recognized that such models are not neutral; they carry a large portion of the biases and discriminations found in human language. If used as support in writing job descriptions, it can contribute negatively to the preservation of gender inequality among occupations and the perpetuation of the sexual division of labor.

Design/methodology/approach: This research uses different LLMs embeddings to evaluate potential gender biases generated in job postings from two major platforms, LinkedIn and Vagas.com. More specifically, it evaluates gender biases and identifies the vectors' sensitivity within the context of gender inequality analysis.

Findings: The degree of consistency between architectures varies significantly as the words contained in job descriptions and the two gender vectors are altered, which means that even pre-trained models might not be reliable to understand gender bias. This lack of consistency indicates that the evaluation of gender bias might be different depending on the parameters. Also, a high sensitivity to pronouns was observed, and the difference between genders seems to be greater when we relate the unitary vectors "man" and "woman" with terms related to family.

Contribution/implication: The use of LLM for job postings must be carried out with caution, and efforts to mitigate gender biases must take place on the *corpus* to be modeled before data training occurs.

Keywords: gender bias, language analysis, word embeddings, LLMs, job postings

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

Resumo

Objetivo: Este artigo visa avaliar potenciais vieses de gênero em descrições de emprego.

Originalidade/valor: A literatura em administração ainda procura entender os diferentes impactos sociais e econômicos gerados a partir da introdução de grandes modelos de linguagem (*large language models* – LLM). Reconhece-se que tais modelos não são neutros; eles carregam um grande número de vieses e discriminações contidos na linguagem criada por humanos. Se usados como ferramenta para a redação de descrições de empregos, podem contribuir negativamente para a preservação da desigualdade de gênero entre as ocupações e para a perpetuação da divisão sexual do trabalho.

Design/metodologia/abordagem: Esta pesquisa faz uso de diferentes *word embeddings* (vetores de palavras) treinados em modelos LLM para avaliar potenciais vieses de gênero gerados em descrições de empregos de duas das principais plataformas de recrutamento e seleção, LinkedIn e Vagas.com. Mais especificamente, ela avalia vieses de gênero e identifica a sensibilidade dos vetores dentro do contexto de análise de desigualdade de gênero.

Resultados: O grau de consistência entre as arquiteturas varia significativamente quando os vetores contidos nas descrições de emprego e os dois vetores de gênero são modificados, o que evidencia que mesmo modelos pré-treinados podem não ser confiáveis para a verificação de vieses de gênero. Essa falta de consistência indica que a identificação de vieses de gênero pode ser impactada a depender dos parâmetros adotados. Ademais, observou-se uma alta sensibilidade a pronomes e que a diferença entre os gêneros parece ser maior quando os vetores unitários "homem" e "mulher" são comparados a termos relacionados à família.

Contribuição/Implicação: O uso de modelos LLM para a redação de descrições de emprego deve ser realizado com cautela e os esforços para mitigar os vieses de gênero precisam acontecer sobre o *corpus* a ser modelado, antes que o treinamento dos dados aconteça.

Palavras-chave: vieses de gênero, análise de linguagem, vetores de palavras, LLMs, descrições de emprego

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

INTRODUCTION

Despite various efforts towards a more gender-equitable society, in several parts of the world, women are underrepresented in different occupations, both in the public and private spheres, especially when related to leadership positions. In the sciences field, data from the United Nations Educational, Scientific and Cultural Organization (Unesco) indicates that the average proportion of women researchers in the world was 29.3% in 2016 (Unesco, 2019). As for the private sector, numbers from Fortune Magazine indicate that only 41 out of the top 500 largest American companies were led by women in 2020, which represents mere 8.1% of the analyzed companies (Hinchliffe, 2021).

In Brazil, this occupational discrepancy is quite critical, considering that women constitute the majority of the Brazilian population. Numbers from the Annual Relation of Social Information (RAIS), consolidated by the Ministry of Labor and Employment of Brazil, show that in 2021, women were more represented in occupations in the areas of health and education, such as beautician (96.20%), nail technician (95.86%), teacher in early childhood education (95.54%), and obstetric nurse (95.47%), and less represented in the fields of construction and industry, such as concrete pump operator (0.31%), construction carpenter (0.31%), and construction and earthmoving machinery maintenance mechanic (0.43%) (Brasil, 2021).

In this sense, the present study perceives that job postings can function as institutional tools capable of reinforcing the perpetuation of gender inequality, using language that implicitly contains gender biases (Doughman et al., 2021; Gaucher et al., 2011). If women are less attracted to job positions they perceive to be more valued and, consequently, dominated by men, male candidates end up considering themselves more suitable for those job vacancies (Gaucher et al., 2011). Thus, the choice of certain textual constructions can reflect an institution's preference for a person of a specific gender to fill a job vacancy within its organizational and hierarchical structure. Even if such a choice is made unconsciously, it can reinforce harmful stereotypes that hinder women's entry into certain job positions, especially those related to sciences, technology, engineering, and mathematics (STEM) fields (Giovannetti & Becker, 2023).

Several studies in Linguistics, Social Psychology, and Anthropology have tried to understand the social or psychological factors that influence the language used by individuals in a society. However, in the last decade, with the advent of natural language processing (NLP) models, there has been a

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

growing number of textual productions generated by artificial intelligence, especially after the introduction of transformers in 2017. This new architecture provided a significant increase in parallelization in computational operations for NLP algorithms modeling (Vaswani et al., 2017) and drove the development of large language models (LLMs), such as Google's Gemini, Meta's LLaMA, and the most popular one of its kind, OpenAI's ChatGPT.

Nevertheless, the management literature still seeks to understand the different social and economic impacts that the introduction of LLM models has produced. It is now recognized that such models are not neutral; on the contrary, they carry a large portion of the biases and discriminations found in language produced by humans (Doughman et al., 2021; Bolukbasi et al., 2016; Caliskan et al., 2017). Thus, as their use becomes widespread and is multiplied in different contexts, LLMs may not only preserve the mentioned language distortions but also exacerbate them.

Although there is extensive literature on gender bias in AI (Doughman et al., 2021; Nadeem et al., 2020; Lambrecht & Tucker, 2018) and previous research considered gender bias in the use of AI in recruitment systems (Yang et al., 2022) we could not find research considering the study of job posting' texts after LLM models were launch. If LLMs are being used as support in writing job descriptions, such algorithms can contribute negatively to the preservation of gender inequality among occupations and the perpetuation of the sexual division of labor.

This research utilizes pre-trained word embeddings to assess gender biases based on job postings collected from two recruitment platforms, LinkedIn and Vagas.com. It uses LLMs to evaluate potential gender biases generated in these textual productions.

RELATED WORK

Gender inequality in the job market

The scientific production on gender inequalities in the job market has expanded significantly in the last few decades. Studies attempt to explain the reasons why men and women progress differently in their careers and the related social impacts. According to Meyer et al. (2015), women are confronted daily with gender stereotypes and the idea that they lack the aptitude to occupy certain spaces. This social construct prevents many women from preemptively directing their efforts toward STEM fields.

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

Jost et al. (2004) present arguments that support the theory of system justification, which assumes the existence of psychological mechanisms that legitimize social arrangements and the way society is organized, often at the expense of collective desires. The authors argue that there is a widespread ideological motive to justify the social order, which partially accounts for the internalization of the inferiorization of certain minority groups (including women) and is observed at an implicit and unconscious level in thought processes.

In industries where men make up most of the workforce, these mechanisms seem to have an even greater impact. When analyzing the job market in the Brazilian digital game development sector, Giovannetti and Becker (2023) identified that insecurity in applying for a job and differential treatment towards men and women during interviews are factors that make the selection process and the journey into the industry even more difficult for women. Furthermore, the overwhelming presence of men sends the message that women are overlooked in that environment, leading many female candidates to give up on applying for a position in the company or, extrapolating, in the industry.

In this regard, Gaucher et al. (2011) examined job advertisements on two recruitment platforms in Canada and concluded that the choice of terms used in the descriptions could function as a mechanism to maintain the status quo in corporate hierarchies and the definition of gender roles. The authors found that ads with more "masculine" words attracted fewer female candidates and diminished their sense of belonging in such occupations. The study also observed that job ads for fields typically dominated by men contained more "masculine" terms than those dominated by women.

In order to identify gender biases in the labor market, Rudinger et al. (2018) used the 60 occupations tested by Caliskan et al. (2017), as well as official data on the participation of men and women in the American labor market. The study sought to ascertain whether the evaluated models would complete sentences with terms that reflected gender biases regarding the selected occupations and compared their results with those obtained with humans. The behavior of the models replicates the labor statistics of the United States and the textual statistics of the data they use in their training. However, the authors reinforce that "these systems overgeneralize the attribute of gender, leading them to make errors that humans do not make on this evaluation" (Rudinger et al., 2018, p. 5).

Finally, Lambrecht and Tucker (2018) explored reasons why job advertising in STEM was more frequently shown to men than women, making it

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

more likely that men applied for the job. They stressed several explanations and concluded that advertising economics considered women click more expensive than men since marketing literature suggests that women control the household budget. This result indicates that the advertising market might contribute to bias due to economic reasons.

Language and bias

Language can have various definitions, depending on the assumptions adopted to analyze it. Even within the study of linguistics, its conceptualization can have different origins. Phonology, for example, represented by Sweet (1877), defines language as the expression of ideas through sounds combined in the form of words, which, in turn, when grouped into sentences, transform those ideas into thought. On the other hand, Chomsky (2002, p. 13) addresses language as "a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements", referring not only to its oral form but also to written language. In a broader sense, language can be interpreted as a communication system in which a set of symbols (written, spoken, gestures) is exchanged between a sending agent and a receiving agent.

Women and men use communication strategies that result in differences in speaking styles. According to Tannen (1987, 1994), women tend to speak more gently and empathetically to build a sense of collectivity. They use the first-person plural more when making a proposal, talk more about other people than about themselves, and, when in a group, they are more inclined to speak all at once as they see conversation as a cooperative act, a combined discourse without polarized positions. On the other hand, men are more direct when giving order, they speak more frequently in public situations of social interactions, and tend to use language to confirm their status within their own group (Kloch, 2000). This might be a result of the social role that men and women are expected to fulfill in the hierarchical structure.

From the social role theory aspect (Doughman et al., 2021), gender stereotypes are rooted in social roles designated to men and women. Men were expected to engage in activities with speed, strength, and more dynamics, and women should stay at home and develop activities linked with family and children. This was also transferred to language, using *he* as a generic term. Therefore, a recurrent process is posed, with a biased language reinforcing a biased world perception.

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

Studies conducted in the 1970s sought to understand how language can play an important role in maintaining a power imbalance between genders depending on how it is organized. As a channel for the transmission of political and ideological power, language is seen as an instrument of domination by oppressive groups (Rowbotham, 1974). The feminist theorist Dale Spender (1980) argues that, historically, men have been more present in public activities and discussions than women, which allowed them to control and adapt the language used, as their opinions and ideas prevailed as socially acceptable. This greater control over language, as a result of other social processes that granted more power to men, imposed a masculine interpretation of reality on society as a whole and forced women to internalize a worldview that was not their own (Corson, 1992). The male style of communication becomes the dominant pattern of speech or discourse and is then copied and associated with power, and it is more common for women to adopt the male style of speech than the other way around (Kloch, 2000; Tannen, 1987).

Communication between individuals is based on semantics: words are endowed with meanings, and the agents understand what is being communicated. This is how gender biases in language are formed, whose recurrent and unconscious use reinforces a cyclical system of inequalities. Leavy (2018) points out other ways in which gender biases can manifest. For example, when referring to women, it is common to use occupational terms associated with gender specification, such as "a female police officer," as a counter to the social expectation that police officers are male occupations. The word "man" as a synonym for "humanity" is also frequently used, such as "the arrival of man on the moon." In addition, it is more common for masculine gender words to come before feminine gender words when used together: "husband and wife," "sir and madam." Finally, the author also notes that terms related to men are much more frequent in language than those related to women.

Gender bias is presented in algorithms that deal with text as much as it is presented in language. From training data to the models themselves, they can lead to consequences in the real world when AI systems are used in decision-making, which can cause discriminatory decisions (Doughman et al., 2021).

This observation aligns with the findings of Webster et al. (2018), who show a significantly disproportionate presence of masculine and feminine pronouns in a *corpora* used for training language models. Considering the definite pronoun resolution dataset, 27% of the pronouns are feminine,

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

while in the Winograd schema challenge datasets, feminine examples represent 28% and 33% of the total.

These discrepancies generate biases that, in turn, find an efficient channel for their propagation in LLMs. The language distortions, captured and processed in large volumes, are reflected in the results of the modeling process of these algorithms, which, when applied massively in other contexts, assist in perpetuating these biases (Caliskan et al., 2017).

METHOD

Considering the gender differences in occupations, this study uses job postings published on online recruitment platforms to verify the application of terms that corroborate the existence of gender biases in the job descriptions. This section is divided into two parts. In the following subsection, the data collection, structure and variables, the model's training *corpus*, *corpus* the repository of word embeddings, and the architectures used to construct the vectors of job descriptions are explained. After that, the other subsection explains the definition of gender biases and discusses the construction of gender vectors, which will be used to calculate the biases. All the codes used here (Python) and the database are available by request to the first author.

Data collection, preprocessing, and vectorization

In order to obtain a list with a significant number of job descriptions, web scraping techniques were applied to two of the main Brazilian online recruitment platforms, LinkedIn and Vagas.com, considering the period from August 2022 to February 2023 and prioritizing the most recent postings at the time of the search, which resulted in 55,673 job postings, excluding duplicates.

After grouping these records into a single database, it was necessary to preprocess the data. The first step was to eliminate duplicate records and records containing null values. As a result, there were 34,881 unique records left, which were then subjected to the removal of punctuation marks, special characters, numbers, and emojis.

This research uses pre-built word embeddings by Hartmann et al. (2017), made available by the Interinstitutional Center for Computational Linguistics (NILC) at the University of São Paulo (USP) in a comprehensive

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

repository of multi-dimensional vectors. This *corpus* includes informative, scientific, didactic, and literary texts, among others. The authors used 1,395,926,282 tokens of 3,827,725 distinct types for constructing these vectors. NILC was the most complete Portuguese repository available at the time of this research.

For LLMs to be developed, these analytical units need to be converted into numbers. Word2Vec, a technique proposed by Mikolov et al. (2013), involves transforming words into unique vector representations that preserve semantic distinctions between each of them. Since Word2Vec requires a historical set of texts to construct word vectors, it is essential to highlight that this study does not disregard the important and ongoing discussion about the stereotypes that these algorithms may carry into the results they generate (Bolukbasi et al., 2016; Caliskan et al., 2017; Garg et al., 2018; Rudinger et al., 2018).

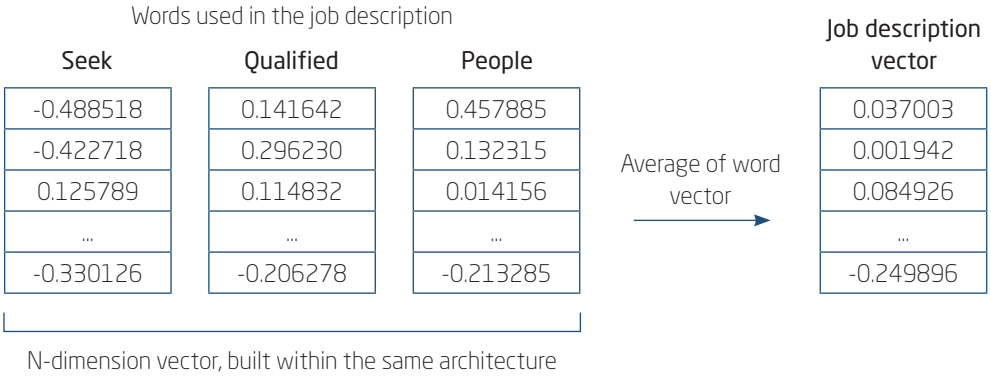
The word vectors constructed by Hartmann et al. (2017) consider two possible architectures for Word2Vec, and both are used within the scope of this research. Word vectors were built via CBOW and Skip-gram and extracted directly from the NILC repository. The model's performance grows with higher dimensionalities and approaches a near-optimal performance at approximately 300 dimensions. Increasing the dimension of the word vectors would convert in diminishing gains (Mikolov et al., 2013). As the choice of different dimensionalities affects the resulting word vector, we chose 300 as the maximum value and tested other lower values, choosing 50, 100, and 300. Six vectors were generated for each job description, combining the three adopted dimensionalities (50, 100, and 300) with the two Word2Vec architectures (CBOW and Skip-gram).

Figure 1 illustrates how the job description vector would be constructed for a publication that contained solely the phrase "Seeking qualified people," considering word vectors of n-dimensions and one of the two architectures mentioned:

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

Figure 1

Construction of job description vectors



Construction of gender vectors and gender bias

The vectors created from the job descriptions must be evaluated in comparison to another tool to highlight the existence of gender bias in the text. In this regard, proxy vectors were defined to represent sets of feminine and masculine words in their most direct conception, based on Garg et al. (2018) and supplemented with other relevant terms within the Brazilian socio-economic and cultural environment. To preserve comparability with the job description vectors, six gender vectors were also generated, considering the combination of the three dimensionalities (50, 100 and 300) and the Word2Vec architecture (CBOW and Skip-gram).

There are various measures that enable the comparison of similarity between vectors, such as Euclidean distance, dot product, or cosine similarity. While the first measure considers only the magnitude of the vectors, the dot product is a more comprehensive measure as it also considers the vectors' orientation (or direction). On the other hand, cosine similarity only considers the orientation between the vectors, meaning it is calculated from the cosine of the angle formed between them. Thus, this measure allows for identifying whether two vectors are similar or not, resulting in values ranging from -1 to 1. In this sense, if two vectors point in the same direction, the calculation result is 1; if they point in perpendicular directions, the result is 0; and finally, if they point in opposite directions, the result is -1.

Cosine similarity can be obtained according to the Equation (1), using the formula for Euclidean inner product, where $\vec{x}$ and $\vec{y}$ are two vectors with the same n-dimensions:

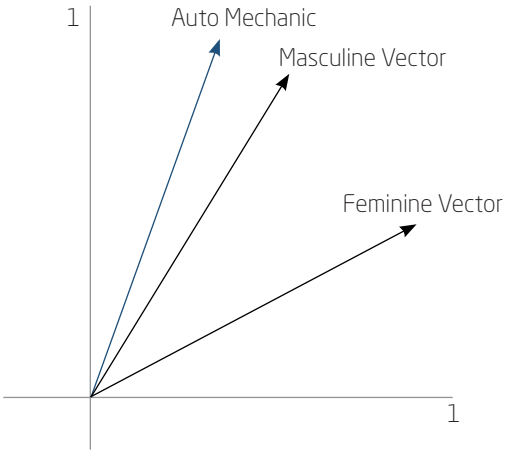
ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

$$sim_{cos} = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \, \|\vec{y}\|} \quad (1)$$

Within the scope of this study, the vectors representing a job description are compared to the vectors of feminine and masculine genders to determine which one shows higher similarity with the job description. The premise is that vector orientation is a sufficient metric for this comparison and, therefore, cosine similarity is suitable for the intended purpose. This premise is supported by word embeddings representing each word as a dimensional vector, resulting in words with similar semantic meanings tending to have vectors close together (Bolukbasi et al., 2016). The calculation was performed for all six vectors for each job description, maintaining the same dimensionality and architecture between them and their corresponding gender vectors. Figure 2 presents a simplified graphical representation to illustrate a possible result, assuming the vectors have only two dimensions.

Figure 2

Graphical representation of cosine similarity



Within the scope of this study, it was defined that if the vector of a job description and the vector of a specific gender show cosine similarity in the range of 0 to 1, the description aligns with that gender (with a higher magnitude of alignment as it approaches 1). In situations where this metric results in a similarity between -1 and 0, then the description does not align

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

with the gender to which it is being compared. In an ideal scenario, cosine similarity would be equal to 0, meaning the description neither aligns nor opposes a possible alignment with the compared gender.

The definition of gender bias adopted in this study refers to the difference in similarity between the job description vector and the two gender vectors. Descriptions with more significant differences in similarity are expected to reflect greater gender biases and directly align with candidates of a particular gender. On the other hand, it is understood that descriptions with smaller differences in similarity would be more neutral and, therefore, attractive to individuals of both genders.

RESULTS

Throughout this analysis, six tests were conducted to assess the impacts of term removals or inclusions in the job description vectors and/or gender vectors, as well as the sensitivity of these vectors. Also, to verify if a job description preserves a higher degree of similarity with the same gender in both architectures. The conducted tests were: (1) original job descriptions and original gender vectors; (2) original job descriptions and gender vectors with pronouns included; (3) job descriptions with pronouns removed and gender vectors with pronouns included; (4) job descriptions with pronouns removed and original gender vectors; (5) original job descriptions and reduced gender vectors; and finally, (6) original job descriptions and unitary gender vectors.

The first observation is that the cosine similarity between a job description vector and a gender vector varies considerably depending on the adopted dimensionality within the same architecture. In some situations, the similarity of certain descriptions had its sign inverted with the increase in dimensionality, which means that they were not only losing similarity, but gaining dissimilarity in relation to the gender with which they were being compared. In the most extreme case, a job description was observed to have a standard deviation of 0.119647 in the similarity between its vectors generated via CBOW and the feminine gender vectors created under the same architecture. This indicates that the models might not be stable and, thus, not reliable when generating less biased text. Table 1 below presents the analysis of the standard deviation of similarities, considering the architecture used in the vector creation and the target gender for comparison.

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

Table 1

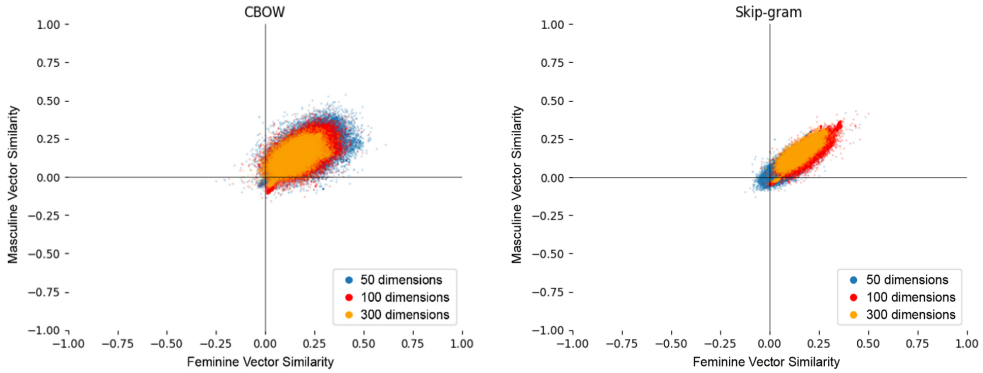
Analysis of cosine similarity standard deviation

Standard deviation	Architecture: CBOW Gender: Feminine	Architecture: CBOW Gender: Masculine	Architecture: Skip-gram Gender: Feminine	Architecture: Skip-gram Gender: Masculine
Average	0.031726	0.024470	0.041417	0.042176
Minimum	0.000065	0.000088	0.001182	0.000805
Maximum	0.119647	0.101894	0.072821	0.078932

This variation caused by the change in the dimensionality of cosine similarities within the same architecture is not entirely unexpected. Mathematically, it is known that in high-dimensional vector spaces, the probability that two normalized vectors are orthogonal is exceptionally high (Arora, 2013). When the vectors are orthogonal, the dot product between them is equal to 0; therefore, the equation (1) numerator is also 0. In other words, the higher the dimensionality of the vector space, the higher the probability that the cosine similarity between them will be lower. Figure 3 below shows that when the job description and gender vectors were constructed under 300 dimensions, the similarities decreased with both genders compared to other dimensionalities. This effect is observed in both CBOW and Skip-gram architectures.

Figure 3

Effects of dimensionality change on cosine similarities



Considering the variability found, it was decided to adopt the average of the similarities of vectors from the same architecture and a specific gender

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

so that the analysis could be developed based on an aggregated metric. This allowed us to have a singular point of comparison during the study, in order to evaluate other factors that may influence gender bias. As a result, the study's metrics were reduced to four: (1) average cosine similarities of CBOW-generated description vectors compared to the feminine gender vector generated by the same architecture; (2) average cosine similarities of CBOW-generated description vectors compared to the masculine gender vector generated by the same architecture; (3) average cosine similarities of Skip-gram-generated description vectors compared to the feminine gender vector generated by the same architecture; (4) average cosine similarities of Skip-gram-generated description vectors compared to the masculine gender vector generated by the same architecture.

A total of six tests were conducted to verify two key aspects:

1. The **consistency between CBOW and Skip-gram architectures** regarding the sign of the average cosine similarities resulting from the comparison to a specific gender vector. In other words, it aimed to test whether the average cosine similarities would have a positive or negative sign in both architectures or if the signs would be divergent depending on the architecture used in the vectorization process. In an ideal situation, the signs should remain the same regardless of the architecture used, meaning the vectorization process is reliable. The rule for consistency analysis is defined as in (2).

$$\left\{\left[(1)\geq(2)\right] \textit{ and } \left[(3)\geq(4)\right]\right\} \textit{ or } \left\{\left[(2)\geq(1)\right] \textit{ and } \left[(4)\geq(3)\right]\right\} \qquad (2)$$

2. The **balance of proportionality between the average similarities and the two genders**. Considering that the sample was collected randomly, it would be expected that the percentage of job descriptions that showed higher similarity to the masculine vector would be close to the percentage of descriptions that showed higher similarity to the feminine vector. In an ideal situation of unbiased text, the descriptions would be equally similar to masculine and feminine vectors. Although this study adopts such balance as a premise, the conducted experiments had their results compared among themselves, considering that it is not possible to guarantee that the two genders are balanced as expected. Therefore, considering the result of the consistency analysis, the verification of balance was conducted only on the set of job descriptions whose CBOW and Skip-gram architectures resulted in similar averages with the same sign.

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

Test 1: Original job descriptions and original gender vectors

The first test calculated the cosine similarities, considering the job descriptions in their original configuration as extracted and the original gender vectors described in section 3.2. Out of the total 34,881 analyzed descriptions, 27,556 (79%) showed consistency between the two architectures. Among the descriptions consistent between CBOW and Skip-gram, 18,849 (68.4%) had higher average similarities with the feminine gender, while only 8,707 (31.6%) resulted in averages closer to the masculine gender.

The consistency test results show that both architectures were able to generate word vectors that, on average, were grouped in descriptions that interacted with the same gender. The balance test, on the other hand, highlights a significant imbalance between the number of job positions that would interact more with the feminine gender and the number of positions that would interact more with the masculine gender.

Test 2: Original job descriptions and inclusion of pronouns in the original gender vectors

Language can comprise multiple categories of pronouns (demonstratives, relatives, indefinites) that are inflected according to the gender they refer to. In this second experiment, it was decided to include some demonstrative pronouns in the original gender vectors due to the inherent gender they carry. This inclusion aims to increase the number of words whose grammatical gender is equivalent to the vector in which they were included and, eventually, enhance the distinction between the two gender vectors.

With the new gender vectors, job descriptions showed a lower degree of consistency between the two architectures. In this second experiment, 22,470 (64.42%) of the descriptions maintained aligned cosine similarity averages regardless of the architecture used in the construction of their vectors. Out of this number, 11,340 (50.47%) had a higher average similarity to the masculine gender, and 11,130 (49.53%) resulted in a higher average similarity to the feminine gender. Despite observing fewer consistent descriptions between the architectures, the texts presented a substantially greater balance between the two genders. It is possible that the inclusion of words whose grammatical gender is inexorably associated with the gender of the vector they were added to result in a more significant distinction between the gender vectors. This can also mean that neutral language might have a significant impact on LLM models.

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

Test 3: Removal of pronouns from job descriptions and inclusion of pronouns in original gender vectors

The third experiment tested how the results could be impacted if the same feminine and masculine pronouns (added to the gender vectors in Test 2) were removed from the job descriptions. This new test adopts the assumption that such pronouns are used with a high frequency in the Portuguese language and may refer to words that do not necessarily have any semantic relation to the job vacancy being described. In this scenario, the gender vectors used in the previous experiment were replicated.

Removing pronouns from the job descriptions resulted in 22,406 (64.23%) job descriptions whose average similarity scores were consistent between the CBOW and Skip-gram architectures. Among the consistent descriptions, 11,352 (50.66%) communicated more directly with the masculine gender, and 11,054 (49.34%) were directed towards the feminine gender. There was a maintenance of consistency between the two architectures regarding the alignment of the average similarity score to a particular gender and a high degree of balance in dialogue with both genders.

Test 4: Removal of pronouns from job descriptions and original gender vectors

Due to the previous results, in a fourth experiment, the scenario tested was where the same feminine and masculine pronouns, included in the gender vectors of Test 2, were removed from the job descriptions and gender vectors. Test 4 seeks to ascertain if the inclusion of new terms in the gender vectors is the factor responsible for positively influencing the balance of the vectors since the results of Tests 2 and 3 were not strongly affected by the removal of pronouns from the job descriptions.

In this new test, 27,895 (79.97%) job description vectors showed consistent average similarity scores with respect to the two architectures used in their construction. Out of this total, 19,232 (68.94%) publications were more aligned with the feminine gender, while 8,663 (31.06%) of them engaged more directly with the masculine gender. The result shows that choosing specific terms in constructing gender vectors can significantly affect the consistency between architectures and the balance between genders.

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

Test 5: Original Job Descriptions and Reduction of Terms in Original Gender Vectors

Considering the results obtained in Test 4, it questions whether the larger number of terms in the gender vectors would be responsible for increasing consistency between architectures. To test this hypothesis, a fifth experiment was conducted using a reduced version of the original gender vectors. In this test, job descriptions were used without removing the previously tested pronouns.

Adopting reduced gender vectors resulted in 22,278 (63.87%) description vectors consistent between both architectures. Among them, 15,320 (68.77%) were closer to the feminine gender vector, while 6,958 (31.23%) were closer to the masculine gender vector. The result obtained here demonstrates that the increase in the number of terms considered in the gender vectors (Test 2) is insufficient to prove that the variation in the number of words is the predominant factor for impact analysis on the degree of consistency between architectures. In other words, the effect on consistency would not be of a mathematical nature. One can question whether the semantic relations carried by the chosen set of terms would have a greater effect on this metric, regardless of the number of words that constitute the gender vectors.

Test 6: Original job descriptions and reduction of terms in original gender vectors to a unitary gender vector

Expanding the test carried out in Test 5, the effects investigated were the ones if the gender vectors were reduced to unitary vectors, meaning gender vectors that contained only the word "woman" or the word "man." The choice of such words is justified by the semantics inherent in the definition of the term "gender." The idea behind this experiment is to verify if a more significant reduction in the number of terms in the gender vectors causes any effect on the consistency between architectures and what that effect would be.

With the reduction of gender vectors to a single-term vector, the degree of consistency between CBOW and Skip-gram architectures experienced a significant decline, and the number of inconsistent vectors became predominant. Only 14,658 (42.02%) job description vectors remained consistent between both architectures. Regarding the balance between genders, Test 6 resulted in a slight increase in the difference in the number of job descriptions closer

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

to a certain gender. Among the consistent vectors, 10,135 (69.14%) job descriptions were closer to the feminine gender vector, and 4,523 (30.86%) were closer to the masculine gender vector.

DISCUSSION

As observed, it is evident that any alteration to job descriptions and gender vectors, whether aggregative or reductive, impacts the results when using word embedding aggregation. Table 2 presents a summary of the experiments, highlighting the comparison of the effects on the degree of consistency between architectures and gender balance.

Table 2

Summary of experiments

Experiment	Consistency	Balance (Fem % / Masc %)
(1) Original job descriptions and original gender vectors	79.00%	68.40% / 31.60%
(2) Original job descriptions and gender vectors with added pronouns	64.42%	50.47% / 49.53%
(3) Job descriptions with pronouns removed and gender vectors with added pronouns	64.23%	50.66% / 49.34%
(4) Job descriptions with pronouns removed and original gender vectors	79.97%	68.94% / 31.06%
(5) Original job descriptions and reduced gender vectors	63.87%	68.77% / 31.23%
(6) Original job descriptions and unitary gender vectors	42.02%	69.14% / 30.86%

It is also observed that this sensitivity to term alterations appears to be lower on job description vectors than on gender vectors. Comparatively, when evaluating the results of Tests (1) and (4), which differ only in the presence/absence of pronouns in the composition of job description vectors, it can be seen that the degree of consistency remains similar, as well as the gender balance. The same does not happen when job description vectors are kept unchanged, and only gender vectors are modified. When comparing Tests (1) and (2), (3) and (4), (1) and (5), and (1) and (6), a greater impact on the consistency between architectures is observed. However, except for the comparison between (3) and (4), the degree of gender balance is preserved

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

in this situation. Finally, the analysis of the results indicates that the inclusion of pronouns in job descriptions may lead to a greater balance in similarities between descriptions and masculine and feminine gender vectors, although further studies are needed to corroborate this finding.

Therefore, the degree of consistency between architectures varies significantly as the words in job descriptions and the two gender vectors are altered. The tests conducted confirm the existence of such variation when descriptions are compared to gender vectors, that is, at the level of aggregations (averages) of word embeddings based on the words contained in these descriptions and the words that compose the gender vectors. However, this oscillation may have been generated when the word embeddings were created, and, therefore, the result captured by the experiments only presents the aggregation of variations occurring at the lowest level of this study, i.e., during the model training phase.

In order to verify this hypothesis, the unitary vector containing only the word "man" was compared to other unitary vectors that, in turn, contained some of the words present in the original masculine gender vector. The words "male," "boy," "father," and "son" were chosen. Furthermore, the similarity between the unitary vector "man" and the reduced masculine gender vector, tested in (5), was also examined. The same verification was performed for their equivalent vectors of the feminine gender: unitary vector "woman" compared to "female," "girl," "mother," and "daughter," as well as to the reduced feminine gender vector "woman." Table 3 displays the results of this analysis.

The study of the resulting standard deviations demonstrates that cosine similarity between the aforementioned vectors indeed varies substantially depending on the adopted dimensionality within the same architecture at the word embedding level. In the most extreme situation, the cosine similarity between the unitary vectors "woman" and "female" in the CBOW architecture decreases only slightly when we increase the dimensionality from 50 (0.771802) to 100 (0.721664). However, it decreases significantly when we increase it to 300 (0.428330), resulting in a standard deviation of 0.151486.

These numbers show that words that share a common characteristic (in this case, gender) can have widely disparate cosine similarities simply because their respective vectors were constructed with different dimensionalities. It can be expected that words that do not have a common characteristic may also exhibit discrepant similarities with changes in the number of dimensions, although further testing is necessary to confirm this.

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

Table 3

Variation in cosine similarities

Architecture	Metric	Dimensionality	Man					Woman				
			Male	Boy	Father	Son	[Unitary]	Female	Girl	Mother	Daughter	[Unitary]
CBOW	Similarity	50	0.743532	0.878105	0.659888	0.619285	0.879140	0.771802	0.911119	0.876158	0.848581	0.927940
		100	0.612263	0.812364	0.537618	0.495975	0.855912	0.721664	0.876010	0.812091	0.795420	0.898871
		300	0.473384	0.640942	0.332757	0.330186	0.761768	0.428330	0.730436	0.724832	0.689429	0.798306
	Average		0.609726	0.777137	0.510088	0.481815	0.832273	0.640599	0.839188	0.804358	0.777810	0.875039
	Standard deviation		0.110302	0.099974	0.134962	0.118448	0.050748	0.151486	0.078224	0.062017	0.066156	0.055541
Skip-gram	Similarity	50	0.632489	0.842850	0.591982	0.500285	0.850894	0.597705	0.868054	0.886398	0.774394	0.872228
		100	0.513901	0.685327	0.491224	0.394867	0.781938	0.522034	0.836663	0.816939	0.722011	0.859626
		300	0.419860	0.539157	0.314790	0.282449	0.721733	0.357365	0.724109	0.707481	0.637049	0.778925
	Average		0.522083	0.689111	0.465998	0.392534	0.784855	0.492368	0.809609	0.803606	0.711151	0.836926
	Standard deviation		0.086998	0.124011	0.114560	0.088947	0.052770	0.100336	0.061801	0.073649	0.056594	0.041334
CBOW + Skip-gram	Combined average		0.565905	0.733124	0.488043	0.437174	0.808564	0.566483	0.824399	0.803982	0.744481	0.855983
	Combined standard deviation		0.108572	0.120929	0.127104	0.113858	0.056940	0.148326	0.072027	0.068083	0.070004	0.052534

From the results displayed in Table 3, it is also noted that the CBOW architecture can generate higher cosine similarities compared to the Skip-gram architecture. In all evaluated cases, the cosine similarity averages between the vectors of 50, 100 and 300 dimensions in the CBOW architecture were greater than the averages obtained in the alternative architecture. This finding may partly explain why the degree of consistency between the architectures varies considerably with removing or including words in job descriptions or gender vectors, as indicated by Tests 1-6. This result aligns with previous literature about gender bias (Doughman et al., 2021; Bolukbasi et al., 2016), which indicates that LLMs are sensitive to pronouns.

Another important observation is that, on average, the unitary vectors of the feminine gender present higher similarities with the unitary vector "woman" compared to the similarities between their corresponding vectors of the masculine gender. Among the obtained results, only the comparison between "man" and "male" generated higher similarity than its feminine counterpart, i.e., the pair "woman" and "female." In general, it is noticed that the unitary vector "woman" is closer to the words it was compared to and, on a larger scale, may be closer to other words in the dictionary considered in the construction of the word embeddings used in this study.

Finally, it is observed that the difference between genders seems to be greater when we relate the unitary vectors "man" and "woman" with terms related to family. The vector "woman" presents higher similarities with the words "mother" and "daughter" than its feminine counterpart, i.e., "man" compared to "father" and "son." The data modeling may indicate that the *corpus* reflects a social reality, considering that women are the primary caregivers. This result aligns with the findings of Caliskan et al. (2017), which attest that women are more associated with terms related to family ("home," "parents," "children," "family," "cousins," "marriage," "wedding," "relatives") in word embeddings.

CONCLUSIONS

There are several factors contributing to the maintenance and perpetuation of gender stereotypes and roles in society. This study sought to demonstrate that, beyond individual preferences, ideologies, or values, gender biases may be presented in LLMs, functioning as an "invisible force" to preserve a status quo that marginalizes women from certain occupations.

Language can be identified as a latent mechanism in maintaining gender distortions in the job market since its manifestation reflects the conse-

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

quences of a social process that disproportionately favors the power and influence of a hegemonic group. According to the findings of this study, the closer proximity between terms related to family and the word "woman" is not surprising.

Large language models, when trained on texts where such associations are normalized, bring to light the biases and prejudices of a culture that are recurrently reproduced by the language constructed by that same culture. At the same time, considering the expansive power and reach of LLMs linked to their increasing popularity today, this study draws attention to their enormous potential to propagate the distortions already present in the language and create new distortions through their unconscious use.

Gender bias, like any other kind of bias, is inherent in language; therefore, its elimination would be impossible (Caliskan et al., 2017). Thus, the presence of gender biases is not a problem specific to Word2Vec or any other LLM technique but rather an issue that pervades language and, consequently, any means that humans use to communicate.

There are current proposals for mechanisms to neutralize or mitigate gender biases in LLMs (Bolukbasi et al., 2016; Zhao et al., 2019). However, this study shares the understanding of Caliskan et al. (2017) who see such models as important tools to elucidate and highlight the presence of biases. Removing these distortions would alter the artificial intelligence's perception of the reality it is modeling and, consequently, the quality of the information resulting from this process. Such information is fundamental to creating a new consciousness and perception of the world that is less biased, which, through language, will be captured by new modeling processes and generate more equitable word embeddings.

Thus, this analysis supports the proposal that efforts to mitigate gender biases must take place on the *corpus* to be modeled before data training occurs (Leavy, 2018). The first step to generating unbiased algorithms is to promote a greater awareness that gender distortions are incorporated into everyday language (Caliskan et al., 2017; Leavy, 2018) without ignoring the extensive feminist literature that has shed light on this problem for centuries. It reinforces the advocacy for a more neutral and less sexist language and the need for a more significant presence of women in knowledge areas related to artificial intelligence, who, as the main victims of this cyclical system of oppression, would be the best agents in identifying gender biases in training corpora and defining and implementing what they define as "justice" (Leavy, 2018). This change is social, above all.

Based on the evidence found here, it can be concluded that gender bias analysis through word embeddings must be carried out with caution, espe-

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

cially when based on gender vectors. It has been shown that Word2Vec architectures are highly sensitive to any alteration made to gender vectors. Although further studies are needed to determine the most appropriate gender vectors for this type of analysis in Portuguese, it is necessary to evaluate to what extent word embeddings may be unbalanced in terms of gender because of the modeling process or their origin. Another limitation is that our evidence supports only Portuguese and only considers job posting texts since word embeddings in another language or another context might be differently distributed among a vectorial space, which can bring different results.

It was also observed that the choice between the two architectures proposed by Word2Vec considerably alters the results since they may reflect a greater or lesser gender imbalance depending on the terms used in the analyzed texts. In this study's sample of job descriptions, the Skip-gram architecture showed a lower degree of discrepancy between the unitary vectors "woman" and "man," even though it resulted in a greater imbalance in the overall repository of terms compared to CBOW.

The development of gender bias analyses through word embeddings requires prior studies on imbalances resulting from data modeling, i.e., the language used, which varies among cultures and social contexts within the same culture. The analysis of gender biases in the context of the job market might produce different results if the analysis were conducted in a context where the language used is more equitable and where one group of people is not dominant over another. Knowing that biases are inherent in language, a possible solution to correct such distortions would be to train the model on a dataset that contains as few gender biases as possible.

In this sense, when relying on LLM to write job postings, it is important to consider that language is a social construction and can play distinct roles in different cultures and industries. The use of pronouns and words historically associated with a specific gender can affect how the text is written by the models and distributed by the algorithms. Considering these aspects can help to convey the effect of gender bias to a lesser degree in the labor market and, more specifically, in recruitment processes.

REFERENCES

Arora, S. (2013). High dimensional geometry, curse of dimensionality, dimension reduction [Lecture notes]. In *Princeton University*. https://www.cs.princeton.edu/courses/archive/fall13/cos521/lecnotes/lec11.pdf

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to Homemaker? Debiasing word embeddings. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 4356–4364. https://doi.org/10.48550/arXiv.1607.06520

Brasil. (2021). Relação Anual de Informações Sociais (RAIS). In *Ministério do Trabalho e Emprego. Programa de Disseminação das Estatísticas do Trabalho.*

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. https://doi.org/10.48550/arXiv.1608.07187

Chomsky, N. (2002). *Syntactic Structures* (2nd ed.). Mouton de Gruyter.

Corson, D. J. (1992). Language, gender and education: A critical review linking social justice and power. *Gender and Education*, *4*(3), 229–254. https://doi.org/10.1080/0954025920040304

Doughman, J., Khreich, W., El Gharib, M., Wiss, M., & Berjawi, Z. (2021). Gender bias in text: Origin, taxonomy, and implications. *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing*, 34–44. https://doi.org/10.18653/v1/2021.gebnlp-1.5

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Gaucher, D., Friesen, J., & Kay, A. C. (2011). Evidence that gendered wording in job advertisements exists and sustains gender inequality. *Journal of Personality and Social Psychology*, *101*(1), 109–128. https://doi.org/10.1037/a0022530

Giovannetti, V., & Becker, J. L. (2023). Elas são a maioria do volume de jogadores, mas não programam. *Cadernos De Pesquisa, 53,* e10233. https://doi.org/ 10.1590/1980531410233

Hartmann, N. S., Fonseca, E., Shulby, C. D., Treviso, M. V., Rodrigues, J. S., & Aluísio, S. M. (2017). *Portuguese word embeddings: evaluating on word analogies and natural language tasks*. https://doi.org/10.48550/arXiv.1708.06025

Hinchliffe, E. (2021, June 2). *The female CEOs on the Fortune 500 just broke three all-time records*. Fortune. https://fortune.com/2021/06/02/female-ceos-fortune-500-2021-women-ceo-list-roz-brewer-walgreens-karen-lynch-cvs-thasunda-brown-duckett-tiaa/

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*(6), 881–919. https://doi.org/10.1111/j.1467-9221.2004.00402.x

Kloch, Z. (2000). Language and gender: Social and psychological determinants in communication. *Psychology of Language and Communication*, *4*(2), 45–58.

Lambrecht, A., & Tucker, C. (2018). Algorithmic bias? An empirical study into apparent gender-based discrimination in the display of STEM career ads. SSRN. https://dx.doi.org/10.2139/ssrn.2852260

Leavy, S. (2018). Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning. *Proceedings - International Conference on Software Engineering*, 14–16. https://doi.org/10.1145/3195570.3195580

Meyer, M., Cimpian, A., & Leslie, S. J. (2015). Women are underrepresented in fields where success is believed to require brilliance. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00235

Mikolov, T., Chen, K., Corrado, G. S., & Dean, J. (2013). Efficient estimation of word representations in vector space. *Proceedings of Workshop at ICLR, 2013*. https://doi.org/10.48550/arXiv.1301.3781

Nadeem, A., Abedin, B., & Marjanovic, O. (2020). Gender bias in AI: A review of contributing factors and mitigating strategies. *ACIS 2020 Proceedings, 27*.

Rowbotham, S. (1974). *Woman's consciousness, man's world*. Penguin Books.

Rudinger, R., Naradowsky, J., Leonard, B., & Van Durme, B. (2018). Gender bias in coreference resolution. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 8–14. https://doi.org/10.48550/arXiv.1804.09301

Spender, D. (1980). *Man made language*. Routledge & Kegan Paul.

Sweet, H. (1877). *A handbook of phonetics*. Clarendon Press.

Tannen, D. (1987). *That's not what I meant!* (2nd ed.). Ballantine Books.

Tannen, D. (1994). *Talking from 9 to 5: Women and men at work*. William Morrow & Company.

UNESCO. (2019). *Women in science*. http://uis.unesco.org

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, abs/1706.03762*, 5998–6008. https://doi.org/10.48550/arXiv.1706.03762

ISSN 1678-6971 • RAM. Rev. Adm. Mackenzie, São Paulo, *25*(6), eRAMD240056, 2024
https://doi.org/10.1590/1678-6971/eRAMD240056

Webster, K., Recasens, M., Axelrod, V., & Baldridge, J. (2018). Mind the GAP: A balanced *corpus* of gendered ambiguous pronouns. *Transactions of the Association for Computational Linguistics*, *6*, 605–617. https://doi.org/https://doi.org/10.48550/arXiv.1810.05201

Yang, J., Njoto, S., Cheong, M., Ruppanner, L., & Frermann, L. (2022). *Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS)*, 140–150. https://doi.org/10.18653/v1/2022.nlpcss-1.15

Zhao, J., Wang, T., Yatskar, M., Cotterell, R., Ordonez, V., & Chang, K.-W. (2019). Gender bias in contextualized word embeddings. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *1*, 629–634. https://doi.org/10.18653/v1/N19-1064