From Looks to Essence: A Shift in Perspective with Physical Appearance Debiasing

Shivatmica Murgai[1]

[1]The International School of Bangalore (TISB), India

Abstract

As Natural Language Processing (NLP) gains popularity due to its powerful nature, it is crucial to consider its role in promoting stereotypes and society's implicit bias. While previous research addresses mitigating bias for gender or race, this study considers bias due to physical appearance in current text corpora, where algorithms mimic human prejudice from trained data. The methods to detect implicit bias in word embeddings include cosine similarity and analogies to evaluate a hard de-biasing approach. The objective of this research is to extend the existing bias detection and mitigation methods to physical appearance in text corpora to promote fairness, inclusivity, and reduce perpetuating stereotypes. To quantify the de-biasing approach, the Word Embedding Association Test (WEAT) score compares the original and de-biased word embeddings on text8 corpus. The proposed alternatives improve mitigating physical appearance bias on text8 corpora by 7.14%. The paper aims to make valuable contributions to the continuous efforts to promote ethical AI applications.

Keywords

Artificial Intelligence, Machine Learning, Natural Language Processing, Bias, Physical appearance, Gender, WEAT Score, De-bias

1. Introduction

Artificial Intelligence (AI) and Machine Learning (ML) have the potential to revolutionize the world and have proven to do so in the past. Its algorithms can serve a wide variety of purposes given enough data, but they are ingrained with a level of bias from training off of biased or skewed data which does not represent the population accurately.

Bias in natural language processing (NLP) models has raised significant ethical concerns and challenges the fairness of generated text. Biased language models not only reflect but also reinforce societal biases, leading to biased outcomes in various applications. They tend to discriminate certain groups, specifically minorities, leading to unfairness in key areas such as employment, healthcare and education.

Stereotypes associated with gender are deemed dangerous when they restrict people to certain professions, decisions, or affect their overall well-being. In children's literature, stereotypes with respect to gender [10], race [11], [12], [14], body image [9], and such are easily accepted and, extensively, internalized throughout society. The stereotypes or general messages conveyed through children's books are integral to a developing child's mindset.

NL4AI 2023: Seventh Workshop on Natural Language for Artificial Intelligence, November 6-7th, 2023, Rome, Italy [16]

shivatmica@gmail.com (S. Murgai)

CEUR Workshop Proceedings (CEUR-WS.org)

Discrimination against physical appearance can profoundly impact individuals' mental health and well-being, as well as their employment prospects [14], [15]. Derogatory or negative judgments lead to low self-esteem, body image issues, or even depression. The relentless pressure to conform to societal beauty standards often results in eating disorders or risky cosmetic procedures. This causes marginalization, perpetuating a spiral of inequality and reduced economic prospects for those targeted. To combat this bias, promoting an inclusive society to recognize the value of diversity, while prioritizing character over physical appearance, is of vital importance.

The Implicit Association Test (IAT) [1] measures the strength of associations between concepts and evaluations or stereotypes to reveal individuals' hidden or subconscious biases. The test involves categorizing stimuli, such as words or images, into different categories such as "good" or "bad," for example. The IAT test records the speed and accuracy of participants' responses to determine their implicit associations and subconscious bias.

Often, AI algorithms amplify prejudice and reinforce particular stereotypes, rather than correcting bias. This can be extremely harmful where AI is prominent in highly autonomous fields, such as applications in interview scanning, facial recognition, and the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) algorithm [2].

The main contributions of this research paper are:

1. We explore the use of cosine similarity to uncover implicit biases related to physical appearance and socioeconomic status.
2. We extend the existing work of de-biasing techniques with gender and race bias into the realm of physical appearance bias and derive their respective mathematical relations.
3. We use WEAT score [6] for bias quantification. WEAT score builds upon cosine similarity by providing a structured framework for analyzing and quantifying associations and biases in word embeddings, especially in the context of social attributes. The WEAT [6] score mimicks the structure of IAT [1] by comparing associations of opposing characteristics with sets of words like professions or adjectives with positive or negative connotations.

2. Literature Review

Bolukbasi et al. [3] introduced cosine similarity to quantify gender bias, indicating biases in the embeddings, which are illustrated through different cosine similarity values from the word pairs. It compares the similarity between different word pairs to measure the association between nouns like occupations and a specific gender. Our research paper applies the cosine similarity method to physical appearance bias and will be discussed in Section III of this paper.

Kim et al [4] asserts that previous work on debiasing word embeddings has largely focused on individual and independent social categories, such as solely gender or race, and de-biases embeddings for this specific singular subspace. This is highly inefficient though since real-world corpora often presents multiple social categories, intersecting with each other. This paper proposes techniques to debias word embeddings for multiple social categories, where individual biases intersect non-trivially, known as intersectional bias. People part of several possibly marginalized social categories may experience bias unique to their intersection of social identities; the experiences and discrimination an African American woman would face vary vastly from that of an African American man or a white woman. An intersectional subspace is constructed to debias embeddings for multiple social categories using nonlinear geometry for individual biases, additionally supported by empirical evaluations. In our research, we extend the debias technique for physical appearance bias and discuss further details in Section III.

The authors in [5] provide a comprehensive survey of more than 300 papers on gender bias in NLP and describe the limitations of current approaches. Most of the studies have used English, Chinese and Spanish languages, which lacks a holistic view of bias in NLP. Lack of bias testing in various NLP models is another issue, which leads to ethical and societal detriment.

In this research paper, we start by discussing crucial aspects of using text corpus datasets and word embeddings. The paper then delves into physical appearance detection and proposes a de-bias method, extended from previous research regarding gender or race. The paper concludes with empirical evidence of these findings and outlines potential directions for future research.

3. Methodology

3.1. Word Embeddings & Dataset (Text Corpus)

Word embeddings, with common algorithms like GloVe [9] or Word2Vec [10], are vector representations of words where similar words will be mapped closer together in an embedding space. GloVe considers words' co-occurrences over an entire corpus and the embeddings relate to the probability that two words are used together in different contexts. However, Word2Vec uses the meaning of words in local context (and also has faster computation), which is much more useful for recognizing bias. Word2vec embeddings, for example, create these vectors based off of previous text's data.

The author experimented with Text8 [7] to train the word embeddings. Text8 is extensively used in NLP to train language models. One of the key advantages of the Text8 dataset is that it has gone through significant text preprocessing to remove unnecessary formatting, punctuation, and other noise, making it suitable for training word embedding models like Word2Vec or GloVe. These models learn to represent words as vectors based on the co-occurrence patterns of words in the dataset. It provides highly compact word vectors while maintaining a wide range of words and contexts.

3.2. Bias Detection Metrics

3.2.1. Cosine similarity: Metric to determine two words' relation

Cosine similarity is a metric to determine how closely related or similar two words are, whose range is $[-1, 1]$ according to the range of the cosine function. Cosine similarity, between the words a and b with x and y as their respective word embeddings, is defined as

$$\text{Cosine Similarity}(a, b) = \frac{x \bullet y}{\|x\|\|y\|} \tag{1}$$

where $x \bullet y$ is the dot product of the two vector word embeddings and $\|x\|$ is the norm or the length of $\|x\|$.

The similarity between two vectors is defined by the magnitude of the angle between them, θ. Additionally, the cosine of an angle is closest to 1 for the smallest angles, which implies greater similarity, and decreases as the magnitude of the angle increases, signifying smaller correlation. Conventionally, we'll consider that differences in cosine similarity scores are reflective of word associations in the trained text corpus, which therefore portrays society's bias. This metric can be employed to detect implicit bias relevant to gender, repulsion towards certain physical appearance, or even socioeconomic backgrounds.

This strengthens the findings of Harvard's Implicit Association Test (IAT), where people were more hesitant to match positive descriptives with negatively assumed characteristics.

Several words are plotted in Figure 1 using cosine similarity score between two words' respective vector embeddings. It depicts society's perceptions of positive and negative physical characteristics as well as unfair stereotypes or assumptions associated with these words. The lengths of the edges connecting different nodes are inversely proportional to the words' similarity. The distance between two nodes represents the correlation between two words, where a smaller distance connotes a greater cosine similarity score.

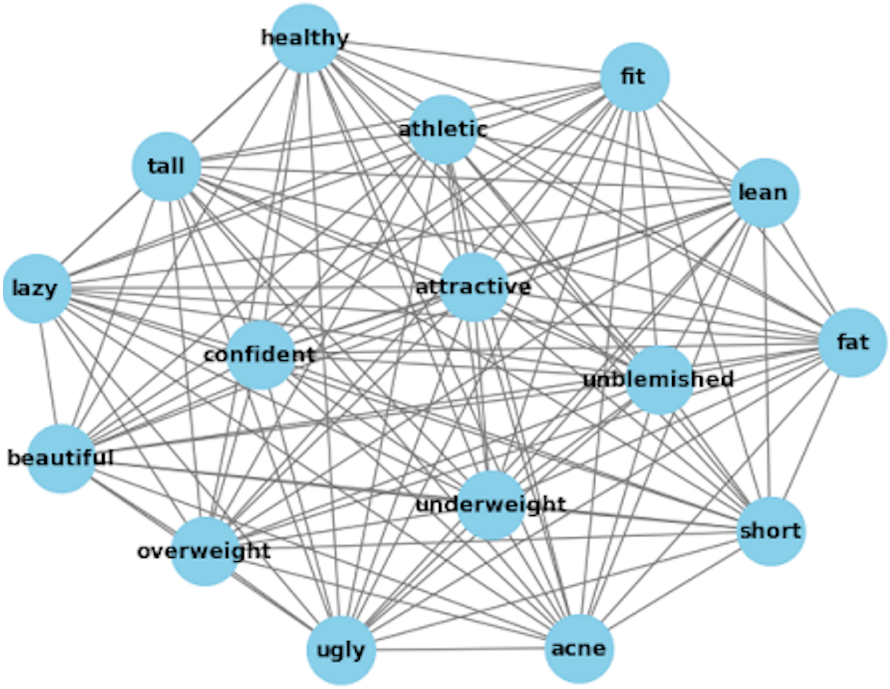


Figure 1: Word Association Network

For example, there are evident instances of bias with these correlations:

- "attractive", "athletic", "healthy", and "confident" are both closer to "fit" than "fat"
- "confident" is closer to "beautiful" than "ugly"
- "lazy" is closer to "overweight" than "underweight"
- "acne" is closer to "ugly" than "attractive", "confident", or "beautiful"

- “tall” is closer to “beautiful” than “short.”

For the above observations, if a word is relatively closer to a word than some other, it’s considered to have a higher cosine similarity score and the two words therefore are more associated according to society’s perceptions.

3.2.2. Analogies

In addition to the cosine similarity metric, analogies ensure that the relationship [3] between two pairs of words is nearly equivalent. A typical example of an analogy would be man : woman :: king : queen. For example, in the analogy $w_1 : w_2 :: v_1 : v_2$, the relationship between w_1 and w_2 that is the same as the relationship between v_1 and v_2. The relationship between two words is defined by subtracting the words’ respective word embeddings,

$$\text{word\_to\_vec\_map}[w_1] - \text{word\_to\_vec\_map}[w_2].$$

With this definition, the relationship between two words can be examined in contrast with the relationship of another two words. An “inverse” relationship would be where $w_1 - w_2$ is equivalent to $v_2 - v_1$ since the order is reversed of one side of the analogy.

Effectively, the relationship between w_1 and w_2 is examined and a word v_2 is found such that the relationship between v_2 and v_1 is identical to the former. The initial brute force method, by considering the word that provides the closest relationship with the other pair of words, did not hold to be as effective if the three provided words are not related.

For instance, in our experiments on Text8 corpus [7], analogies(“black”, “white”, “up”) outputs an entirely irrelevant word. Consequently, the “most_similar” function from word2vec more effectively comprehended the relationship between w_1 and w_2.

The output of analogies (“black”, “white”, “up”) was “down,” as desired. Although this seems to be more accurate than the brute force method, it still includes a large amount of bias in the context of gender. The model succumbs to the famous example of gender bias,

$$\text{woman : man :: homemaker : \_\_\_}$$

and outputs “machinist.”

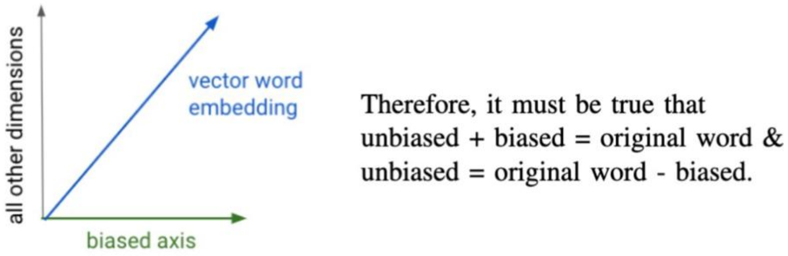
For bias related to physical appearance, these embeddings are dominated by stereotypes associating negative or positive adjectives with society’s assumptions of these characteristics. For instance, given the first three words of the following analogy, it provides the word “weak”,

$$\text{‘big : strong :: thin : weak’.}$$

It is evident from our empirical findings that these vectors contain an ingrained notion of bias. The paper now discusses ways to reduce bias with gender or physical appearance by hard de-biasing the word embeddings. We would leverage the approach suggested by Kim et al [4] to the physical appearance realm. This includes neutralizing or equalizing the vectors to fit an idea of gender equality or removing discrimination based on physical appearance for words that, in an ideal world, should not be correlated with specific stereotypical characteristics.

3.3. Generalizing de-biasing technique for an arbitrary subspace

A subspace can be defined as the relationship between two opposing adjectives, similar to how the gender subspace is defined as ”man” – ”woman” in [3]. The difference between the vectors that map to these antonyms defines the subspace, upon which the vectors are projected onto. The original word’s vector is the word embedding and the “biased” component is the projection of the original word’s embedding onto the biased axis of gender. The projection from the original word can be subtracted to obtain the unbiased vector.



Therefore, it must be true that unbiased + biased = original word & unbiased = original word - biased.

Fig. 2. Word embeddings vector component

This biased axis is defined as the difference between the vectors corresponding to antonyms representing the subspace. The subspace encapsulates the idea or direction of a specific category such as gender, race, or a physical characteristic.

The projection of the word’s vector **w** onto the biased axis can be calculated as

$$\mathbf{w}_{biased} = \frac{\mathbf{w} \bullet \mathbf{subspace}}{\|\mathbf{subspace}\|^2} \cdot \mathbf{subspace} \tag{2}$$

where w is the word’s embedding, **subspace** is the biased axis, and $\mathbf{w}_{biased}$ is the word’s biased component. Since hard de-biasing entirely removes this biased component, in a way, the definitions of the word embeddings can be updated to obtain

$$\mathbf{w}_{unbiased} = \mathbf{w} - \mathbf{w}_{biased} \tag{3}$$

Algorithm 1 Debias (*word*, *gender*, *word_to_vec_map*)

```
$w \leftarrow word\_to\_vec\_map[word]$
$w\_biased \leftarrow$ (Dot product of $w$ & $gender$) $\cdot$ $gender/$(norm of gender)$^2$
$w\_unbiased \leftarrow w - w\_biased$
return $w\_unbiased$
```

The amount of bias in a particular subspace can be compared by using the cosine similarity metric with the vector representing the subspace (i.e. defined as $a - b$). A negative score implies a word’s greater association with b, and if it is positive, then a.

Additionally, the magnitude of the similarity with this subspace is irrelevant to understanding which word is more closely related with a instead of b (very small positive values do not necessarily indicate they are skewed towards the word a, but rather comment upon correlation with the subspace vector, since the difference between a and b is considered). In other words, 1 does not entirely represent associated with a, and neither does -1 with b.

3.4. Equalization for an arbitrary subspace

Equalization applies to words equivalent in meaning but differing with respect to the corresponding subspace. Both these words should be equidistant to each category, to alter how a word is sometimes ingrained with bias relevant to physical appearance or gender.

Debiased vectors remove any notion of the subspace from them and the resulting vector word embedding from these words solely exists on the axis with the other $n - 1$ dimensions, excluding this biased axis. Equalization, whose techniques are discussed in [3], ensures that any word specific to a subspace's category, such as a gender-specific word, is equidistant from this axis of $n - 1$ dimensions.

After equalizing, the only difference between the two words is their direction on the biased axis. The stem of these words is the same since excluding the projection onto the biased axis (or the biased component) of each vector should result in the same vector component. Since the words' biased components are in opposite directions, the vectors' sum consists of no biased components because the addition causes it to cancel out. By taking the average of the two opposite vectors, the unbiased component is obtained.

Algorithm 2 Equalize(*pair_of_words*, *biased_axis*, *word_to_vec_map*)

```
$w1, w2 \leftarrow word\_to\_vec\_map[pair\_of\_words]$
$x \leftarrow$ Average of word vectors $w1$ and $w2$
$x\_biased \leftarrow$ Biased component of $x$ (using projection)
$x\_unbiased \leftarrow x - x\_biased$
$w1\_biased, w2\_biased \leftarrow$ Biased component of $x$
    (using projection of the vectors onto the biased axis)
$eq\_w1\_biased, eq\_w2\_biased \leftarrow$ Biased components
    used to equalize both vectors according to equations for
    eq_w1_biased and eq_w2_biased
return $eq\_w1\_biased + x\_unbiased, eq\_w2\_biased + x\_unbiased$
    // combines equalized components with the unbiased component to get
    equalized word vectors for $w1$ and $w2$ below
```

Let w_1 and w_2 be two words opposite in the biased axis. The biased component of their average, $x = \frac{w_1+w_2}{2}$, can be obtained, and then subtracted from x to obtain the unbiased vector. The unbiased and biased components of x are calculated by replacing the subspace projected onto earlier (while de-biasing) with the biased axis.

Both words, w_1 and w_2 are now projected onto the biased axis, once again subspace with the biased axis. Finally, the biased components of the two words' vector embeddings are obtained to be the following:

$$\textbf{eq\_w1\_biased} \quad = \sqrt{|1 - ||\mathbf{x}_{unbiased}||^2|} \cdot \frac{\mathbf{eq}_{w1_{biased}} - \mathbf{x}_{biased}}{||(\mathbf{eq}\_w1 - \mathbf{x}_{unbiased}) - \mathbf{x}_{biased}||}, \tag{4}$$

$$\textbf{eq\_w2\_biased} \quad = \sqrt{|1 - ||\mathbf{x}_{unbiased}||^2|} \cdot \frac{\mathbf{eq}_{w2biased} - \mathbf{x}_{biased}}{||(\mathbf{eq}\_w2 - \mathbf{x}_{unbiased}) - \mathbf{x}_{biased}||}. \tag{5}$$

We have extended the concept provided for gender bias in Bolukbasi at el. [3] to physical appearance bias in the above equations. To find the equalized word embeddings, the above biased components are added to $\mathbf{x}_{unbiased}$. Therefore, the following is true

$$\textbf{eq\_w1} = \textbf{eq\_w1\_biased} + \mathbf{x}_{unbiased}, \tag{6}$$

$$\textbf{eq\_w2} = \textbf{eq\_w2\_biased} + \mathbf{x}_{unbiased}. \tag{7}$$

By equalizing, both words, *a* and *b* are become equidistant to the vector representing the subspace, which means that 1 and −1 would represent the two words that make up the subspace. Neutralization prevents words skewing to one category of the subspace specifically, while ensuring that the embeddings retain their original meaning.

4. Experiments & Results

4.1. Equalization example for gender bias

This is necessary since gender-specific words, such as "hero" and "heroine," are not equidistant from the biased axis of gender.

The subspace is defined as

$$\text{gender} = \text{word\_to\_vec\_map["she"]} - \text{word\_to\_vec\_map["he"]}.$$

(The word "man" was not used since it is more commonly used to connote a person such as in "mankind" instead of a male.)

	Gender	Cosine Similarity score with gender
Before equalizing	"he"	$-0.23826271 \approx -0.24$
Before equalizing	"she"	$0.62127906 \approx 0.62$
After equalizing	"he"	$-0.62998897 \approx -0.63$
After equalizing	"she"	$0.62998885 \approx 0.63$

Table 1
Cosine similarity score with gender & nouns or professions

The subspace is defined as the gender vector to equalize upon. As mentioned earlier, −1 does not entirely correlate to men or 1 to women, as shown in Table I.

4.2. De-biasing Word Embeddings for Physical Appearance

Physical appearance is a broad category which may include various definitions of the subspace to debias upon. For instance, bias due to weight, appearance, and attractiveness, with their associated societal stereotypes can be explored. Word 1 and Word 2 are antonyms to describe a single subspace, defining a category of physical appearance such as height, weight, or complexion, for example.

In essence, these distinct subspaces could de-bias word embeddings individually, and cosine similarity is calculated again using Word 2's de-biased word embedding. This updated cosine similarity, between the Word 1 and the de-biased word embedding of Word 2, is calculated and these results are shown in the rightmost column of Table 2, where de-biasing projects the vector onto each subspace and subtracts the resulting biased component. The table depicts a subset of the words considered for this research paper.

The subspace in each scenario is defined as the vector corresponding to the difference between the two antonyms like Word 2 − Word 1. For instance, the cosine similarity score of the word "actor" with the subspace weight (defined as "tall" − "short") is 0.227, which is reduced to $2.71 \cdot 10^{-8}$. This decrease elucidates the effectiveness of de-biasing.

However, this would be an extremely targeted, and consequently limited, approach, which is not scalable. This includes several pairs of positive and negative characteristics based on society's inherently flawed standards and expectations, as well as common stereotypes, jobs, and assumptions for these characteristics. A single subspace is instead constructed from individual subspaces (defined as the difference of antonyms) by averaging the vectors representing subspaces for different categories of physical characteristics.

4.3. Identification of biased stereotypes and characteristics

Cosine similarity scores and analogies were especially beneficial to highlight words that presented profound associations with certain attributes of physical appearance, such as in Table 2. If a word such as "beautiful" has a positive correlation with the vector representing height, or more generally, physical appearance, it is evident this conveys some idea of bias. The examples shown in Table 2 exemplify humans' pyschological biases towards people's inborne physical characteristics; this prejudice inevitably leaks into algorithms trained off of this biased data. Some words may exhibit strong positive correlations with favorable or less favorable physical characteristics, allowing for their categorization.

4.4. WEAT Score

The Word Embedding Association Test (WEAT) score [6] is a measure of the association between sets of words based on their word embeddings. It is commonly used to evaluate the presence of biased associations in word embeddings. A positive WEAT score between two sets of words indicates greater association, while a negative score indicates weaker association.

The original WEAT score on the text8 corpus was 0.35, indicating the existence of physical appearance bias in word embeddings. After employing debiasing techniques, WEAT score decreased to 0.325, improving by 7.14%, quantifying average efficacy of this approach in mitigating physical appearance bias in word embeddings.

Word 1	Word 2	Orig CSS	Debiased CSS
Fat	Lazy	0.357	0.222
Fit	Lazy	0.142	0.259
Overweight	Healthy	0.344	0.321
Underweight	Healthy	0.247	0.292
Tall	Beautiful	0.340	0.177
Short	Beautiful	0.110	0.228
Pimples	Attractive	0.032	0.095
Unblemished	Attractive	0.161	0.106
Ugly	Kind	0.228	0.352
Pretty	Kind	0.471	0.369
Villager	Uneducated	0.252	0.143
Metropolitan	Uneducated	0.076	0.172
Bald	Attractive	0.132	—

Table 2
Cosine similarity Score (CSS) illustrating bias towards physical appearance, depicting a subset of words considered for the research

5. Conclusion

The research successfully identified instances of physical appearance bias in word embeddings using cosine similarity and analogy methods. The de-biasing approach showed promising results in mitigating physical appearance bias. Maintaining a fine balance between retaining contextual information while mitigating biases is a challenging but crucial aspect in achieving effective debiasing results. Leveraging WEAT score, the magnitude of such biases was quantified, providing a meaningful metric to assess the level of bias reduction achieved through our debiasing efforts. After debiasing certain biased words from the corpus, WEAT score improved by 7.14%. The proposed de-biasing approach showed promising results in reducing bias but there are challenges in achieving complete de-biased dataset. As AI continues to evolve, future research has great potential to effectively transform the world into a socially responsible AI-powered society.

6. Future Work

This study only focused on debiasing certain identified biased words. This could be extended to identify and mitigate bias in the entire text8 or Google News corpora. Instead of considering individual subspaces, the different subspaces defined within the category of physical appearance are correlated in potentially a non-linear relation. Futher research could examine the relationships between these words to further improve debiasing methodology and create a singular subspace encapsulating society's standards of physical appearance.

7. Acknowledgements

The author would like to sincerely thank Mr. Catalin Voss, CTO and Co-founder of Ello (helloello.com), and Prof. Nick Haber, Assistant Professor at Stanford University, for their review feedback, continuous support, and guidance throughout the research study.

References

[1] Loyola Marymount University. “Test Your Implicit Bias - Implicit Association Test (IAT) - Loyola Marymount University.” Lmu.edu, 2023, resources.lmu.edu/dei/initiativesprograms/implicitbiasinitiative, Accessed 17 June 2023.

[2] Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks; Angwin, Julia and Larson, Jeff and Mattu, Surya and Kirchner, Lauren, 2016

[3] Man is to Computer Programmer as Woman is to Homemaker, Debiasing Word Embeddings, Tolga Bolukbasi and Kai-Wei Chang and James Zou and Venkatesh Saligrama and Adam Kalai, 2016, https://doi.org/10.48550/arXiv.1607.06520

[4] Nayoung Kim, Huan L, Proceedings of the 29th International Conference on Computational Linguistics, October 12–17, 2022 https://aclanthology.org/2022.coling-1.110.pdf

[5] A Survey on Gender Bias in Natural Language Processing KAROLINA STAŃCZAK, University of Copenhagen ISABELLE AUGENSTEIN, University of Copenhagen https://arxiv.org/pdf/2112.14168.pdf

[6] Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183-186.

[7] Mahoney, M. *Text8 Corpus*, 2006. https://mattmahoney.net/dc/text8.zip

[8] Mikolov, T., Chen, K., Corrado, G., & Dean, J. *Distributed Representations of Words and Phrases and their Compositionality*, 2013. *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS 2013)*, pp. 3111-3119. https://code.google.com/archive/p/word2vec/

[9] GloVe: Global Vectors for Word Representation, Pennington, Jeffrey and Socher, Richard and Manning, Christopher D, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014

[10] Distributed Representations of Words and Phrases and their Compositionality; Mikolov, Tomas and Sutskever, Ilya and Chen, Kai and Corrado, Greg and Dean, Jeff, 2013

[11] Media’s Influence on Body Image Disturbance and Eating Disorders: We’ve Reviled Them, Now Can We Rehabilitate Them? J. Kevin Thompson, Leslie J. Heinberg, Martina M. Altabe, Madeline M. Tantleff-Dunn, Journal of Social Issues, 1999

[12] Gender Stereotypes, Naomi Ellemers Annual Review of Psychology 2018 69:1, 275-298

[13] The Consequences of Race for Police Officers’ Responses to Criminal Suspects, Plant, Ashby, Peruche, Michelle, 10.1111/j.0956-7976.2005.00800.x, Psychological science, 2005/04/01

[14] A Race-Based Size Bias for Black Adolescent Boys: Size, Innocence, and Threat; Freiburger, Erin, Sim, Mattea, Halberstadt, Amy, Hugenberg, Kurt; 10.1177/01461672231167978; Personality & social psychology bulletin; 2005/04/01

[15] Perceived Discrimination and Physical, Cognitive, and Emotional Health in Older Adulthood; Sutin, Angelina, Stephan, Yannick Carretta, Henry, Terracciano, Antonio 2014/03/01; American Journal of Geriatric Psychiatry

[16] Elisa Bassignana, Dominique Brunato, Marco Polignano, Alan Ramponi, Preface to the Seventh Workshop on Natural Language for Artificial Intelligence (NL4AI), in: Proceedings of the Seventh Workshop on Natural Language for Artificial Intelligence (NL4AI 2023) co-located with 22th International Conference of the Italian Association for Artificial Intelligence (AI* IA 2023), 2023.