Uncovering Bias in Classic Children's Books: Are We Doing Better Today?

V. Bolzonella
Radboud University, Nijmegen, NL
veronica.bolzonella@ru.nl

ABSTRACT

We analyze gender, age, and socioeconomic bias in the perception of professions, animals and adjectives in classic children books. We demonstrate a strong bias against women in professions, in the form of under-representation, especially in professions associated with positions of power. We also show a bias in animals of large size being perceived more as male and rich. Moreover we demonstrate a strong aporophobia in adjectives, with the vast majority of positive adjective scoring high similarity with the "rich" direction, and negative adjectives scoring high similarity with the "poor" direction. Finally, we report that modern stories aimed at transmitting positive values show a reduction in aporophobia, and a reduction in sexism and reverse ageism in professions and animals perception. However, we report an increase in gender and age bias in the use of adjectives, seeing a stronger polarization of these towards one or the other direction of the bias axis.

1 INTRODUCTION

The bias in language that children are exposed contributes to the establishment of often discriminatory stereotypes. Exposing bias in children literature holds value as a form of understanding and evaluating the education of children, and consider bias accordingly in conversation and educational contexts. In this project we expose gender, age and socioeconomic bias present in classic chidren books. It is also interesting to explore how these biases changed in time, and evaluate if new generations are exposed to less, more or different biases.

To these ends, we aim to answer the following research questions:

What patterns can be found in the gender, age and socioeconomic bias in children books across words for professions, animals and adjectives?

Is there evidence of a change in these patterns in modern children stories?

We hypothesize that classic books present strong gender bias against women, reverse ageism (bias against young people), and aporophobia (bias against poverty). We also hypothesize a reduction in all biases in the extended dataset.

2 RELATED WORK

The method utilized in this project is inspired on that of Bolukbasi et al. [2], where the bias is quantified in relation to the location of words on the vector space of a custom trained Word2vec model. While [2] offers a framework to identify and remove bias from the model, this project in not concerned with the latter task. Other methods have been designed for the purpose of bias identification, such as NBIAS [18]. However, our method is advantaged over that of NBIAS in the way that it does not require annotation of text, which is a costly process. Caliskan [5] proposes another method, Word Embedding Association Test (WEAT), to quantify bias in word embeddings. However WEAT focuses on relative bias between two groups, and is not suitable for measuring global presence of bias in a corpus.

In successive work, Caliskan also extends the analysis of gender bias in embeddings with statistical methods, POS analysis, and cluster analysis [4]. For purpuse of simplicity, interpretability and generalizability of our results, we limit our method to that of embeddings projections on a bias axis, as done by Bolukbasi et al [2].

We contribute to the field of bias identification by analyzing the gender, age, and socioeconomic bias in professions, adjectives, and animals perceptions, in children's literature. These analysis will be performed in an analogous method to that of [2], aplpied to sets of target words and axis pairs in all fields above mentioned.

Moreover, we extend the method by comparing bias in different periods of time, more specifically that of classic books, written between the years 1765 and 1963, and that of modern stories written in the 21st century. This is done using the magnitude of vectors' shift towards a perpendicular or parallel direction to the bias axis after finetuning the model with modern stories.

The results will be compared with experts' findings on animal stereotypes in children books [13, 19] and movies [20], reports on wemen and youth misrepresentation [1, 12, 14].

3 METHOD

A Word2Vec model is trained with 3021 classic children books obtained by crawling the Children and Young Adults category in the Gutenberg Project [8]. The books are published between 1765 and 1963, and collectively written by over 1000 authors.

The embeddings are validated against manually labelled scores from the dataset in [3]. This dataset contains pairs (e.g. "sun" and "sunshine"), and a score on a 50 point scale assessing the similarity of the terms (where 50 indicates identical terms). We assess the validity of the Word2Vec model as the Pearson Correlation Coefficient between the similarity labels from the validation set and the cosine similarity calculated on the embeddings as $1 - cosine(\vec{w}_1, \vec{w}_2)$. Our trained models achieved a coefficient above 0.7 with high statistical significance. The bias analysis will be performed using the following method. The projection (cosine similarity) of words in the categories of interest (professions, animals, adjectives) is quantified against the bias axis.

TxMM'25-'26, Radboud University, Nijmegen, Netherlands

Each word achieves a score between -1 and +1, indicating association with one or the other direction of the bias axis. In particular, for each axis negative scores indicate respective association with "she", "young" and "poor".

The bias axis are made more stable by taking an average of a number of possible directions between word couples representing this bias. For example the gender axis is found by averaging the directions between pairs "he, she", "man, woman", "boy, girl", etc. The axis pairs are created manually, in collaboration with 3 annotators who proposed and selected the most representative pairs for each bias. The complete lists of axis pairs can be found in Appendix B.

A second Word2Vec model is then finetuned with 304 additional stories from the database Bedtime Stories [9], which features modern stories for children with the aim of enforcing positive values.

The same scores are obtained for the new embeddings, and the shift is quantified by measuring the difference in scores towards or against the direction of the axis that a word was previously associated with.

3.1 Target groups

Each target group includes terms and categories. The category labels were obtained by taking the majority label out of 3 annotators with a total Cohen's Kappa score of 0.891.

3.1.1 Professions. This category includes 95 professions with no definitional gender. The list does not include terms which are associated with a gender by their grammatical or syntactical form, such as "business woman" or "actress". Words with a stereotypical gender association, such "nurse" or "president" are included.

Professions are divided into the following categories:

- STEM, such as 'biologist', 'mathematician', 'programmer'.
- Healthcare, such as 'doctor', 'surgeon', 'nurse'.
- Business and Finance, such as 'banker', 'manager', 'salesman'.
- Government and Law, such as 'president', 'ambassador', 'lawyer'.
- Education and Academia, such as 'professor', 'philosopher', 'librarian'.
- Art, Media, and Entertainment, such as 'writer', 'dancer', 'singer'.
- Service and Support, such as 'driver', 'janitor', 'barber'
- Caregiving and Household, such as 'nanny', 'housekeeper', 'maid'.

3.1.2 Animals. Animals are an important presence in children stories. For example, think of the iconic characters in The Three Little Pigs [11] or The Tale of Peter Rabbit [16]. This is because animals, apart from stimulating children's imagination, also form a diverse and cross-cultural characters group, making it easier for writers to create inclusive characters that are easy to relate to by children across different social groups. Nonetheless the use of certain animals is often tied to anthropomorphism, the attribution of human traits and emotions to non-human entities. While animals are used for their diversity and inclusivity, these antromorphisms can still subtly result in children unconsciously learning wrongful associations. For example, the authors in [14] find that animals are the most unequal group among those studies in terms of gender representation.

The target words include 74 animals, classified into Pray, Predator, or Neutral, and in Big, Medium or Small. Notice that the animals do not need to be definitionally part of the category, but must be sentimentally. What we mean is that an animal such as "ant" is categorized as a neutral, even if it can be biologically defined as a predator. This is because ants are not stereotypically considered predators, hence the readers sentiment towards ants is not that of a predator. Finally, notice that some animals have a definitional gender (e.g., cow), but are still included in the list as these animals are common in children stories and can be insightful across other bias axes.

3.1.3 Adjectives. This list includes 136 adjective that can be used to describe a person, such as 'active', 'charming', 'beautiful'. These are classified as Positive, Neutral, and Negative.

4 RESULTS AND ANALYSIS

Note that the following report only includes significant findings, and excludes discussion of targets and bias for which no strong results were found.

4.1 Gender Bias

4.1.1 Gender bias in professions. Our findings suggest a strong gender bias in professions. The first striking property of these results is the disproportional representation, finding that less than 30% of professions carry a female connotation. This result is supported for example by reports from [15], who estimate a 20% to 32% share of the labor market being women between 1920 and 1960, mainly in the field of domestic and personal service. We also notice a strong prevalence of jobs in "Arts, Media and Entertainment" in the female direction of the axis, while most jobs that hold positions of power (particularly those in Law, Politics, and Finance), carry a strong association with the male direction of the axis (see Figure 2). These results are not surprising, and widely supported by historical evidence around the limited and polarized number of female workers during the time of writing of the books. Even if gender imbalance in the labor markets is well-known, showing that these patterns are embedded in material aimed at children uprising and education highlights how social expectations and constructs are transmitted in one's early life.

4.1.2 Gender bias in animals. Our findings suggest a strong correlation between gender and size of animals: those perceived as larger animals tend to be represented as male, whereas smaller animals are more often linked to female connotations. This is in line with our findings from the next session, which see the adjective 'giant' among the most masculine. This association demonstrates a perpetuation of the bias that associates men with power and strength and women with fragility and vulnerability. Related studies [6] also find high probability of female characters being portrayed as "submissive and dependent".

We also find an interesting outlier in the frog, as this is among the animals most similar to the male vector, despite being classified as a small animal. This finding is consistent to that of other work, such as in [19]. We associate this outlier with the strong popular

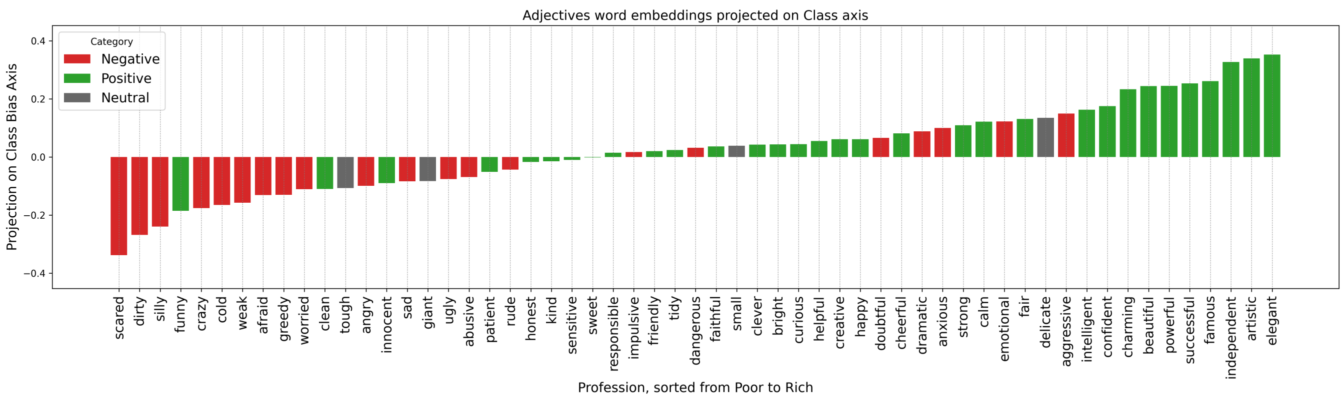

Figure 1: Adjectives projection on the Class axis
On the vertical axis the projection between -1 (identical direction to 'poor') and +1 (identical direction to 'rich), on the horizontal axis, some of the words in the adjectives target group, classified as positive (green) or negative (red).

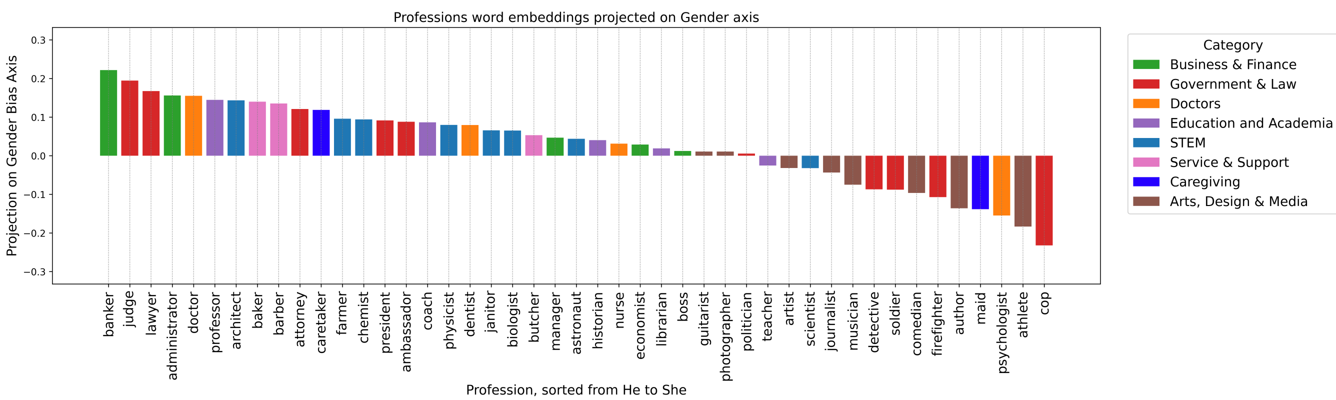

Figure 2: Professions' projection on the Gender axis
On the vertical axis the projection between -1 (identical direction to 'she') and +1 (identical direction to 'he), on the horizontal axis, some of the words in the professions target group, classified in different market groups

image of the frog as a male character coming from the stories in our dataset such as The Frog Prince (the original version of the Princess and the Frog) [10], or The Tale of Mr. Jeremy Fisher [17], both depicting the frog as a male character.

Finally, we also find a higher pray to predator ratio among the animals with female association (3:1), than in those with male association (1:1). Oter studies find similar results [20].

4.1.3 Gender bias in descriptions. Firstly we find that the majority of adjectives carry a female connotation. This could reflect a bias found in the work of Caliskan et al. [4], in which they find that "the top male-associated words are typically verbs while the top female-associated words are typically adjectives and adverbs". As a result adjectives appear in the corpus more often as descriptors of women characters than of men characters.

Moreover, we identify recurring gendered patterns across the adjectives, for example *innocent, dependent, sensitive, emotional* associated to women, and . As already mentioned above, this result reflects cultural expectations for women to remain submissive and passive, as well as a cultural perception of women as emotional, impulsive and dramatic.

There is no evidence indicating women are described positively more than men are, which is in contrast with the psychological phenomenon known as the women-are-wonderful effect [7].

4.2 Age Bias

4.2.1 Age bias in professions. We firstly notice that the majority of the professions are associated with adults. This result is not surprising, as we do not expect children to have a job. Nonetheless, an interesting observation is that the majority of the professions strongly associated with youth, are also found in the female direction of the gender axis. This indicates further sexism in the perception of professions.

4.2.2 Age bias in animals. While we notice a slight tendency for larger animals being associated with adults, a more evident result is the association of all predators as adults. This can be supported by the general understanding that predators only develop their pray

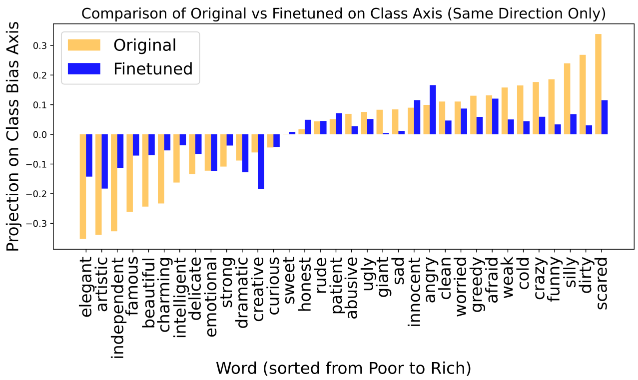

Figure 3: Comparison of scores before and after finetuning with modern stories for some adjectives.

abilities as adults, rather than this representing a bias coming from antrophormism.

5 SOCIOECONOMIC BIAS

5.0.1 Class bias in animals. We see a strong correlation between large animals and the "rich" direction of the bias axis. Similarly to the discussion previously mentioned, large animals are associated with strength and power, two features that are stereotypically associated with rich people.

5.0.2 Class bias in descriptions. As can be seen in Figure 1, adjectives show the strongest display of aporophobia, with a ranking that features the great majority of positive adjectives associated to 'rich' (e.g., *elegant, artistic, independent, successful*), and almost all negative adjectives associated to 'poor' (e.g. *scared, dirty, silly, weak*).

This finding can be explained by the social and economic hierarchies determined by wealth as a measure of material, moral and cultural superiority. This demonstrates that classic children literature illustrates poverty as fear, ignorance, and weakness, reflecting and reinforcing societal prejudices.

5.1 Bias Shift

In Table 1 are the total shifts for each bias and target group. A positive shift refers to a cumulative shift of target vectors towards a perpendicular angle to the axis, while a negative shift indicates vectors got cumulatively more aligned with the axis, enforcing the bias. We notice an overall improvement, with a consistent positive shift across all biases for Profession and Animals, as well as across all target groups for the socioeconomic bias (see Figure 3). A general increase in women representation, which would explain a reduction in gender bias, is consistent with findings of [13]. We do, however, report a negative shift in gender and age bias for adjectives.

6 DISCUSSION AND OUTLOOK

This research demonstrates strong biases, particularly sexism in professions, animal figures, and descriptions, reverse ageism in animal figures, and aporophobia in professions, animal figures, and descriptions. The focus on children literature shows that children

Bias	Professions	Adjectives	Animals
Gender	0.6141	**-2.7171**	1.5164
Age	4.3681	**-1.1669**	2.8369
Class	4.8430	6.6441	3.1435

Table 1: Total shifts of biases by target group

are exposed to these prejudice and social constructs since early development. It is therefore fundamental that language in books aimed at the young public are carefully examined for conscious and unconscious bias.

Our findings indicate a general decrease in gender, age and socioeconomic bias in modern children stories. Nonetheless some bias is still present and requires further improving. This and similar work is of value to address present prejudice that is being transmitted to children since a young age, particularly in the education sector. Future research direction should focus on developing similar frameworks capable of identifying multi directional bias, such as that of race and ethnicity, where a single axis is not sufficient for defining all bias directions. We also propose the analysis of non-binary gender to future work. This was not considered in this study due to the lack of representation of these groups in classic books. Finally, we identify a gap in the research of aporaphobia and socioecnomic bias in word embeddings compared to other biases. We therefore a more in depth analysis of aporaphobia in literature is a potential direction for future work.

Below we report some notable limitation of our work. Firstly, the target words, axis pairs, and categories are derived from three annotators. More annotators would result in more stable axis, less biased target groups, and more reliable categories. Secondly, we finetuned the model with stories freely available online from a single source. There are not representative of all stories and children books today. A larger database of modern children books, particularly edited books, is a more representative set of books read by children nowadays. Finally, most of our findings are validated through comparison with a small set of expert findings. We therefore want to note that findings require more thorough validation by careful examination of the dataset and statistical methods, which were not explored in this project due to time constrains.

REFERENCES

[1] Anjali Adukia, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and Teodora Szasz. 2023. What We Teach About Race and Gender: Representation in Images and Text of Children's Books. *The Quarterly Journal of Economics* 138, 4 (2023), 2225–2285. https://doi.org/10.1093/qje/qjad028

[2] Tolga Bolukbasi, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. 2016. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In *Advances in Neural Information Processing Systems 29 (NeurIPS 2016)*.

[3] Elia Bruni, Nam-Khanh Tran, and Marco Baroni. 2014. Multimodal Distributional Semantics. *Journal of Artificial Intelligence Research* 49 (2014), 1–47. https://doi.org/10.1613/jair.4135

[4] Aylin Caliskan, Pimparkar Parth Ajay, Tessa Charlesworth, Robert Wolfe, and Mahzarin R. Banaji. 2022. Gender Bias in Word Embeddings: A Comprehensive Analysis of Frequency, Syntax, and Semantics. In *Proceedings of the 2022 AAAI/ACM Conference on AI,*

Ethics, and Society (Oxford, United Kingdom) *(AIES '22)*. Association for Computing Machinery, New York, NY, USA, 156–170. https://doi.org/10.1145/3514094.3534162

[5] Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356, 6334 (2017), 183–186.

[6] I. L. Child, E. H. Potter, and E. M. Levine. 1946. Children's Textbooks and Personality Development: An Exploration in the Social Psychology of Education. *Psychological Monographs* 60 (1946), 1–144.

[7] Alice H. Eagly, Antonio Mladinic, and Stacey Otto. 1991. Are women evaluated more favorably than men? An analysis of attitudes, beliefs, and emotions. *Psychology of Women Quarterly* 15 (1991), 203–216. https://doi.org/10.1111/j.1471-6402.1991.tb00792.x

[8] Project Gutenberg Literary Archive Foundation. 1971. Project Gutenberg. https://www.gutenberg.org/

[9] FreeStoriesForKids.com. 2008. FreeStoriesForKids. https://freestoriesforkids.com/

[10] Jacob Grimm and Wilhelm Grimm. 1812. *The Frog Prince (Der Froschkönig / Iron Henry)*. Realschulbuchhandlung.

[11] James Orchard Halliwell-Phillipps. 1886. *The Three Little Pigs*. London: John C. Nimmo.

[12] Rebecca Harlin and Hani Morgan. 2009. Review of Research: Gender, Racial and Ethnic Misrepresentation in Children's Books: A Comparative Look. *Childhood Education* 85, 3 (2009). https://doi.org/10.1080/00094056.2009.10521389

[13] J.L. Massman. 1979. Animal Stereotypes in Children's Picture Books. (1979). https://scholarworks.uni.edu/cgi/viewcontent.cgi?article=4820&context=grp Article from University of Northern Iowa ScholarWorks.

[14] Janice McCabe, Emily Fairchild, Liz Grauerholz, Bernice A. Pescosolido, and Daniel Tope. 2011. Gender in Twentieth-Century Children's Books: Patterns of Disparity in Titles and Central Characters. *Gender & Society* 25, 2 (2011), 197–226. https://doi.org/10.1177/0891243211398358

[15] U.S. Department of Labor. 2020. Occupations of Women in the Labor Force Since 1920. https://www.dol.gov/agencies/wb/data/occupations-decades-100

[16] Beatrix Potter. 1902. *The Tale of Peter Rabbit*. Frederick Warne & Co.

[17] Beatrix Potter. 1906. *The Tale of Mr. Jeremy Fisher*. Frederick Warne & Co.

[18] Shaina Raza, Muskan Garg, Deepak John Reji, Syed Raza Bashir, and Chen Ding. 2023. {NBIAS}: A Natural Language Processing Framework for Bias Identification in Text. *arXiv preprint arXiv:2308.01681* (2023). https://arxiv.org/abs/2308.01681

[19] Melanie Walsh. 2025. The Sneaky Gender Bias in Picture Books: Animal Characters. *Publishers Weekly* (aug 2025). https://www.publishersweekly.com/pw/by-topic/children/childrens-industry-news/article/98304-the-sneaky-gender-bias-in-picture-books-animal-characters.html Web exclusive article, August 05 2025.

[20] Lara A. Wood. 2025. Mr Predator and Mrs Prey: gender stereotypes in children's films correlate with explicit and implicit gender stereotyping. *Social Development* 34, 3 (2025).

A WORK REPORT

B AXIS PAIRS

Youth Term	Elder Term
toddler	grandparent
child	elder
boy	senior
girl	grandmother
schoolboy	retiree
schoolgirl	octogenarian
teenager	elderly-man
adolescent	elderly-woman
youth	aged-person
baby	grandfather
infant	grandmother
kids	seniors
minor	pensioner
newborn	old-man
youngster	old-woman

Table 2: Pairs of words used for the age bias axis.

Lower-Income Term	Higher-Income Term
poor	wealthy
impoverished	millionaire
destitute	billionaire
low-income	affluent
homeless	opulent
needy	high-income
underprivileged	aristocrat
working-class	tycoon
struggling	elite
day-laborer	magnate
pauper	heiress
beggar	heir
underclass	prosperous
broke	luxurious
disadvantaged	upper-class

Table 3: Pairs of words used for the socioeconomic bias axis.

He	She
monastery	convent
spokesman	spokeswoman
Catholic_priest	nun
Dad	Mom
Men	Women
councilman	councilwoman
grandpa	grandma
grandsons	granddaughters
prostate_cancer	ovarian_cancer
testosterone	estrogen
uncle	aunt
wives	husbands
Father	Mother
Grandpa	Grandma
He	She
boy	girl
boys	girls
brother	sister
brothers	sisters
businessman	businesswoman
chairman	chairwoman
colt	filly
congressman	congresswoman
dad	mom
dads	moms
dudes	gals
ex_girlfriend	ex_boyfriend
father	mother
fatherhood	motherhood
fathers	mothers
fella	granny
fraternity	sorority
gelding	mare
gentleman	lady
gentlemen	ladies
grandfather	grandmother
grandson	granddaughter
he	she
himself	herself
his	her
king	queen
kings	queens
male	female
males	females
man	woman
men	women
nephew	niece
prince	princess
schoolboy	schoolgirl
son	daughter
sons	daughters
twin_brother	twin_sister

Table 4: Pairs of words used for the gender bias axis. These are derived from the list used in [2]

C SCORES

Word	Gender (O)	Gender (F)	Age (O)	Age (F)	Class (O)	Class (F)
administrator	0.1445	-0.0127	0.1559	0.0169	0.3586	0.0352
ambassador	0.0044	-0.0330	0.0879	-0.0510	0.2760	0.0171
analyst	0.1587	-0.0335	-0.0162	-0.1307	-0.0218	0.0011
architect	0.1398	0.0786	0.1435	-0.0188	0.2931	0.0535
artist	-0.0571	-0.0170	-0.0317	-0.1565	0.2442	0.0422
astronaut	-0.0760	-0.1838	0.0440	-0.1372	-0.0720	-0.0066
athlete	0.1809	-0.0739	-0.1834	-0.1903	0.1138	0.0012
attorney	0.1308	-0.0119	0.1211	0.1064	0.1693	0.0952
author	-0.0299	-0.0771	-0.1361	-0.1363	0.1839	0.2478
baker	0.0901	0.1046	0.1398	-0.0095	-0.1202	-0.0085
banker	0.0899	-0.0093	0.2217	0.0582	0.2110	0.0551
barber	0.0877	0.0685	0.1354	-0.0584	-0.0902	-0.0401
biologist	0.0986	0.0569	0.0653	-0.0425	0.2249	0.1140
boss	0.1752	0.0818	0.0124	0.0720	-0.1136	-0.1255
butcher	0.1875	0.0213	0.0532	-0.0693	-0.1625	-0.0140
chef	0.0504	0.1394	-0.0241	-0.0059	0.1152	-0.0633
chemist	0.0277	0.1147	0.0942	-0.0725	0.1062	-0.0211
coach	0.0545	-0.0666	0.0865	-0.0919	0.1228	-0.0386
comedian	0.0826	0.0121	-0.0964	-0.2839	0.1602	-0.0121
cop	-0.0721	-0.1206	-0.2322	-0.0517	-0.1826	-0.1593
dentist	-0.0592	-0.0469	0.0796	-0.0551	-0.1723	-0.0003
detective	0.0895	-0.0026	-0.0870	-0.0103	-0.0059	0.0181
doctor	0.0259	-0.0129	0.1551	-0.0369	-0.0768	-0.1045
economist	0.0911	-0.0944	0.0289	-0.0537	0.2250	0.1637
farmer	0.1482	-0.0029	0.0957	0.0813	-0.0657	-0.0412
firefighter	0.1222	-0.0614	-0.1072	-0.1176	0.0444	0.0082
guitarist	-0.0219	-0.0434	0.0110	-0.0923	0.0483	0.0216
historian	0.1989	0.0799	0.0403	0.0020	0.2754	0.0665
janitor	0.0463	-0.0695	0.0659	-0.0496	-0.1020	-0.1250
journalist	0.0381	-0.0720	-0.0433	-0.1560	0.1844	0.0598
judge	0.0870	-0.0428	0.1947	0.0056	0.1439	0.0481
lawyer	0.1245	-0.0206	0.1675	0.0472	0.1889	0.0451
librarian	-0.0152	-0.0691	0.0191	-0.1442	0.1479	0.0446
manager	0.1446	0.0738	0.0469	-0.0036	0.1193	-0.0288
mathematician	0.1411	0.0164	0.0015	-0.0151	0.1854	0.1122
musician	0.0176	-0.0173	-0.0751	-0.0696	0.1256	0.0049
nurse	-0.4289	-0.2050	0.0314	-0.0354	-0.1195	-0.1532
photographer	-0.0166	0.0912	0.0107	-0.0241	0.0989	-0.0883
physicist	0.1874	0.0442	0.0797	-0.1418	0.1591	0.1370
politician	0.2230	-0.0099	0.0055	-0.1469	0.2699	0.0271
president	0.1253	-0.0077	0.0914	0.0080	0.2254	0.0509
professor	0.2129	0.0441	0.1447	-0.1105	-0.0347	-0.0285
psychologist	-0.0963	-0.1454	-0.1548	-0.2012	0.0817	0.0633
researcher	0.0338	-0.0962	-0.0034	-0.0556	0.1414	0.0768
scientist	0.2400	0.0778	-0.0320	-0.0467	-0.0250	-0.0541
soldier	0.2154	0.0097	-0.0879	-0.0357	-0.1169	-0.1304
teacher	-0.2001	-0.2363	-0.0253	-0.1782	-0.0464	-0.0694

Table 5: Scores for the projection of all professions against each bias axis before and after tuning

Word	Gender (O)	Gender (F)	Age (O)	Age (F)	Class (O)	Class (F)
abusive	0.1089	-0.0401	-0.0514	-0.2411	-0.0690	-0.0274
afraid	-0.0327	-0.1081	0.0723	-0.1107	-0.1315	-0.1206
aggressive	0.0263	-0.0535	-0.1395	-0.2397	0.1492	-0.0488
angry	-0.0128	-0.0876	0.0616	-0.2154	-0.0992	-0.1658
anxious	-0.0718	-0.0553	-0.0322	-0.1124	0.0997	-0.1001
artistic	-0.2588	-0.1209	-0.0868	-0.1478	0.3393	0.1830
beautiful	-0.2414	-0.1872	-0.0210	-0.1372	0.2437	0.0706
bright	-0.1237	-0.1160	-0.0939	-0.0388	0.0431	-0.0135
calm	-0.1106	-0.0884	-0.0752	-0.0155	0.1219	-0.0069
charming	-0.3109	-0.2411	-0.0821	-0.1865	0.2334	0.0541
cheerful	-0.1222	-0.1353	-0.0358	-0.1020	0.0813	-0.0429
clean	-0.1060	-0.1097	-0.0676	-0.0427	-0.1106	-0.0467
clever	-0.0634	-0.1477	-0.0521	-0.1211	0.0427	-0.0174
cold	-0.0822	-0.0583	0.0123	-0.1539	-0.1650	-0.0440
confident	0.0179	-0.0402	-0.1140	-0.1243	0.1751	-0.0490
crazy	-0.0723	0.0298	-0.0413	-0.0807	-0.1764	-0.0595
creative	-0.0650	-0.0126	-0.2396	-0.0275	0.0612	0.1836
curious	-0.0756	-0.0036	-0.0573	-0.1152	0.0443	0.0424
dangerous	0.0534	-0.0550	-0.0683	-0.2013	0.0319	-0.0395
delicate	-0.3206	-0.1382	-0.2448	-0.1286	0.1347	0.0660
dirty	-0.0650	-0.1099	-0.1737	-0.1305	-0.2682	-0.0301
doubtful	0.0070	-0.0627	-0.1257	-0.1566	0.0655	-0.0439
dramatic	-0.1600	-0.0812	-0.2065	-0.2121	0.0883	0.1279
elegant	-0.1332	-0.1751	0.0295	-0.1230	0.3529	0.1426
fair	-0.1718	-0.2145	-0.0891	-0.1583	0.1305	-0.0469
faithful	0.0588	-0.0304	-0.0011	0.0028	0.0360	-0.0805
famous	0.1815	0.1280	0.1015	-0.1765	0.2612	0.0716
friendly	-0.0180	-0.1595	0.0121	-0.2672	0.0197	-0.2166
funny	-0.0841	-0.0770	-0.0379	-0.1369	-0.1856	-0.0335
giant	0.2523	0.1575	0.0693	0.0493	-0.0828	-0.0047
greedy	-0.0216	0.0439	-0.1380	-0.0340	-0.1302	-0.0590
happy	-0.1195	-0.1534	-0.0864	-0.1069	0.0612	-0.0052
helpful	-0.1436	-0.2222	-0.1573	-0.1469	0.0548	-0.0250
honest	0.1120	-0.0661	-0.0522	-0.0343	-0.0170	-0.0491
impulsive	-0.1881	-0.1055	-0.2072	-0.1050	0.0171	-0.1276
independent	-0.0056	-0.0566	-0.1980	-0.1023	0.3270	0.1130
innocent	-0.1164	-0.1072	-0.3093	-0.1862	-0.0900	-0.1154
intelligent	-0.0007	-0.0857	-0.1776	-0.1482	0.1627	0.0369
kind	-0.0573	-0.0995	-0.0260	-0.1129	-0.0147	0.0021
patient	-0.1499	-0.0190	-0.1013	-0.0077	-0.0512	-0.0714
powerful	0.1302	0.0588	-0.0009	-0.1456	0.2450	-0.0055
responsible	-0.0166	-0.1861	-0.0680	-0.0451	0.0149	-0.0071
rude	-0.0529	-0.1615	-0.0326	-0.1482	-0.0436	-0.0451
sad	-0.1645	-0.1914	-0.0892	-0.1199	-0.0841	-0.0116
scared	-0.0651	-0.0973	-0.0876	-0.1636	-0.3383	-0.1150
silly	-0.1657	-0.1868	-0.1773	-0.1460	-0.2394	-0.0681
small	0.0056	-0.1195	-0.1165	-0.1270	0.0390	-0.0111
strong	0.0674	-0.0859	0.0282	-0.0137	0.1089	0.0378
successful	0.0556	-0.0223	-0.0621	-0.1820	0.2534	-0.0059
sweet	-0.2672	-0.1571	-0.1042	-0.0728	-0.0015	-0.0082
tidy	-0.3244	-0.2671	-0.0242	0.0513	0.0243	-0.0461
tough	0.1481	-0.0541	0.0001	0.0082	-0.1075	0.0153
ugly	-0.0794	-0.1429	-0.0504	-0.1713	-0.0758	-0.0516
weak	-0.0237	-0.1055	-0.2143	-0.1583	-0.1578	-0.0501
worried	-0.1290	-0.1452	0.0068	-0.1202	-0.1107	-0.0868
emotional	-0.1728	-0.1325	-0.2621	-0.1017	0.1225	0.1228
sensitive	-0.1866	-0.0130	-0.2903	-0.2552	-0.0101	0.0136

Table 6: Scores for the projection of all adjectives against each bias axis before and after tuning

Word	Gender (O)	Gender (F)	Age (O)	Age (F)	Class (O)	Class (F)
ant	-0.0223	-0.0145	-0.1190	0.0553	-0.0584	-0.0241
bear	0.0338	0.0286	-0.0190	0.0223	-0.1620	-0.0031
bee	-0.1574	-0.1231	-0.1962	0.0198	-0.1404	-0.0478
butterfly	-0.2380	-0.1236	-0.2921	-0.0714	-0.0202	-0.0342
cat	-0.1687	-0.0281	-0.0327	0.0037	-0.1074	-0.0214
chicken	-0.2144	-0.0013	-0.1120	0.0078	-0.0793	0.0115
cow	-0.0958	-0.0103	-0.0624	0.0490	-0.1126	-0.0297
dog	0.0656	0.0455	-0.0730	-0.0326	-0.1957	-0.0992
donkey	0.0604	0.0440	-0.0374	0.0589	-0.0842	0.0028
duck	-0.1129	-0.1010	-0.0693	-0.0425	-0.0953	-0.0591
elephant	0.1436	0.0689	-0.0132	0.0099	-0.0115	-0.0092
fish	0.1127	0.1284	-0.1079	0.0024	-0.0711	-0.0303
fox	0.1071	-0.0617	0.1116	0.0651	-0.1318	0.0134
frog	0.1336	-0.0551	0.0680	0.0391	-0.0251	0.0072
giraffe	0.1646	0.1403	0.0117	0.1019	-0.1451	-0.0626
goat	0.0093	0.0509	-0.0235	0.0992	-0.1826	-0.0329
goose	-0.2010	-0.1041	-0.1140	-0.0227	-0.0972	0.0491
horse	0.1602	0.1032	0.0375	0.0749	0.0221	-0.0350
lion	0.2026	0.1166	-0.0134	0.0560	-0.0701	0.0411
monkey	0.0819	0.0631	-0.1046	-0.0244	-0.2330	-0.0945
mouse	-0.1189	-0.1552	-0.1235	0.0358	-0.1520	0.0716
owl	-0.0040	-0.1583	0.1193	0.0194	-0.0511	0.0064
panda	0.1857	-0.1324	0.0637	-0.0904	0.1275	-0.0232
parrot	-0.1555	-0.0916	0.0393	-0.0359	0.0037	0.0078
pig	-0.0732	0.0469	-0.1240	0.0321	-0.1543	0.0371
rabbit	0.0735	-0.1252	0.0453	0.0433	-0.1931	0.0029
sheep	0.0200	-0.0562	-0.0363	0.0330	-0.0562	-0.0139
snake	0.0101	0.0201	-0.0814	-0.0091	-0.1515	-0.0932
spider	-0.1648	-0.0702	-0.1213	0.0829	-0.0787	-0.0068
tiger	0.1065	0.1049	0.0049	-0.0189	-0.0789	-0.0394
turtle	0.0165	-0.0309	-0.0052	-0.0145	-0.1209	-0.0746
wolf	0.1265	-0.0494	0.0174	0.0494	-0.2454	0.0367
zebra	0.0769	0.0300	-0.0490	0.0069	-0.0038	0.0653
bird	-0.1075	-0.1380	-0.2218	-0.0755	-0.0977	-0.0236
deer	0.1058	-0.0534	-0.0463	-0.0505	-0.0425	-0.0146
dolphin	0.1254	0.1157	-0.1104	-0.0372	0.0126	-0.0897
crocodile	0.0725	0.0571	0.0169	0.0523	-0.0862	-0.0109
camel	0.0934	0.1273	0.0341	0.1303	0.0308	-0.0108
bat	0.0807	-0.0485	-0.1001	-0.2424	-0.1739	-0.0734
beaver	0.1263	-0.0612	0.1276	0.0494	0.0261	0.0457
caterpillar	-0.0911	-0.0599	-0.1811	0.1042	-0.1582	0.0520
crow	0.0911	-0.0853	0.1150	-0.0155	-0.1506	-0.0239
eagle	0.0989	0.0181	-0.0431	-0.0736	0.0000	-0.0183
lizard	0.1021	0.0002	-0.0352	0.0363	-0.0575	-0.0528
cheetah	0.1629	0.0353	-0.0003	-0.0825	-0.0218	-0.0209
chimpanzee	0.1225	0.0408	0.0423	-0.1012	0.0241	0.0513
gorilla	0.0793	0.0203	-0.0081	-0.0347	-0.0252	0.0530
grasshopper	0.0778	-0.0696	-0.0923	-0.0165	-0.1735	-0.0879
hamster	0.1522	0.0598	0.0042	0.0090	-0.0259	0.0285

Table 7: Scores for the projection of all animals against each bias axis before and after tuning