

# Gender Bias Analysis

## Reproduce

### Prerequisites

To reproduce the findings, ensure you have the following prerequisites: - A corpus of text that is representative of natural language usage. `brown.sents(categories=['news'])` from the `nlTK` package was used for this specific model. - A word embedding model library, such as Gensim in Python. - Python environment with necessary libraries installed (numpy, sklearn, etc.).

### Steps

1. Load your corpus of text into the Python environment.
2. Train a word embedding model using the corpus. Example using Gensim:

```
from gensim.models import Word2Vec
sentences = // your text corpus as a list of tokenized sentences //
model = Word2Vec(sentences, vector_size=100, min_count=5, sg=0)
```

3. Perform a similarity search for the target words ('guy', 'woman', 'mrs', etc.):  

```
similar_words = model.wv.most_similar('target_word', topn=20)
```
4. Record the top 20 words and their similarity scores.

### Analysis

- Compare the list of similar words for male-associated and female-associated words.
- Observe the types of roles, attributes, and contexts associated with each gender.

## Bias Report

### Observations of Potential Gender Bias

1. 'guy':
  - Professions linked with 'guy': 'director', 'general', 'attorney' suggest traditional male roles.
  - To confirm bias, compare with professions linked to 'woman' or 'girl'.
2. 'woman':
  - 'issue' as a top similar word might imply negative connotations associated with women.
3. 'mrs' and 'mr':
  - A closer relationship of 'mrs' with domestic terms versus 'mr' may reflect traditional domestic gender roles.
4. 'woman' and 'junior':
  - If 'junior' is associated with 'woman' but not 'man', it may indicate a bias suggesting women are often in subordinate roles.
5. 'guy' and 'working':
  - A strong correlation might perpetuate the stereotype of men being primarily valued for their work.

### Conclusion

While the observations above highlight potential areas of bias, they should be contextualized within the broader dataset. It is crucial to consider societal norms and historical data representations that could contribute to these biases. To make definitive claims of bias, further statistical analysis and comparisons with control terms are necessary.