

# Project Discussion

## Gender Bias in Word Embeddings

### Project Outline Report (Write-up)

- Information on Word Embeddings
- Information on Bias (Gender Bias)
- Technologies for De-biasing (General Idea)
- Timeline

### Points to consider:

- Approach:
  - Recognise
  - Algorithmically Identify
  - De-bias
- Factors that make Word Embeddings biased
- Do datasets reflect pre-concieved notions
- Use Pre-trained Models (Word2Vec, ELMO)
- Look for List of Biases (list of Gender Bias)
- Use data-sets like GoogleNews300

### Questions to consider:

- Is bias a function of corpus
- Other kinds of Bias
- Creative ways to de-bias

### Timeline:

#### Week 1:

- Explore Word Embeddings (Word2Vec, ELMO)
- Explore Pre-trained models
- Architecture

#### Week 2:

- Similarities and analogies
- Input stereotypes and non-stereotypes and check if analogies come up in WE

#### Week 3+:

- Qualitative Analysis
- Is bias a function of data or function of architecture or both
  - $Bias = f(data) \mid Bias = f(architecture)$

## Bonus:

- Types of other biases
- Other methods of removing bias