**Project Title:** Investigating Gender Bias in Language Generated by chatGPT-3

**Introduction:**

The emergence of language models like chatGPT-3 has led to new possibilities in natural language processing, including automated text generation such as scripts, blogs, stories etc.. However, there are concerns about the potential for gender bias in language generated by these models. Previous research has shown that language models can reproduce and amplify existing societal biases, including gender bias. Therefore, it is important to investigate the extent and nature of gender distribution in language generated by chatGPT-3, especially as these models become increasingly prevalent in various applications.

**Research Question:**

To what extent does the language generated by chatGPT-3 distribute gender across popular professions.

**Papers**:

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0245533#abstract0>

<https://www.cambridge.org/core/journals/british-journal-of-political-science/article/political-leadership-in-the-media-gender-bias-in-leader-stereotypes-during-campaign-and-routine-times/B197672D2B8A6BCB0A65920396151699>

**Methodology:**

*Data Acquisition and Processing:* We will throw some questions to ChatGPT-3 to generate some response. The corpus will then go through a prepossessing stage by stripping the entire corpus to individual sentences. We will then analyse the file using natural language processing techniques (such as sentiment analysis and Named entity recognition) to identify patterns of gender bias in the language generated.

**Approach**

The proposed approach consists of two main steps: name extraction and name classification.

Name extraction: This step involves identifying and extracting potential names from chatGPT-3 generated text using natural language processing (NLP) techniques such as tokenization, part-of-speech tagging, named entity recognition, and rule-based filtering. The output of this step is a list of name candidates with their corresponding text spans.

Name classification: This step involves assigning a gender label to each name candidate using machine learning (ML) techniques such as supervised learning, feature engineering, and model evaluation. The output of this step is a list of name candidates with their corresponding gender labels.

**Expected Outcomes:**

The study is expected to contribute to our understanding of the gender distribution in language generated by chatGPT-3, and how these biases manifest across different topics and contexts. The findings will provide insights into the ethical implications of using AI-generated language, especially in applications where gender distribution can have harmful consequences. Additionally, the study will offer recommendations on how to mitigate gender biases in language models like chatGPT-3.

**Conclusion:**

This study aims to investigate the extent and nature of gender distribution in language generated by chatGPT-3. The findings of this study will have implications for researchers, practitioners, and policymakers working in the field of natural language processing and AI. By identifying and addressing gender biases in language models like chatGPT-3, we can ensure that these models are more equitable, inclusive, and beneficial for all.