**Identifying and Manipulating Personality Characteristics in PretrainedLanguage Models**

*Graham Caron*

Language modeling is widely used in the field of natural language processing (NLP) for tasks that require text generation. While previous research has looked at various forms of bias in pretrained language models from popular machine learning libraries, there is an absence of research into personality bias in these models. Identifying and manipulating personality characteristics in pretrained language models has the potential to improve their implementation. This study develops testing procedures to identify the Big Five Dimensions of personality in transformer-based language models, BERT and GPT2. By providing additional context to the models though a series of experiments, the work not only demonstrates a tendency for measured personality characteristics in BERT and GPT2 to change but also, the ability for them to be manipulated in a in a predictable way.