Leveraging Natural Language Processing Techniques and Bidirectional Encoder Representations From Transformers (BERT) for Detecting Bias and Toxicity in Journalism and Monkeypox Related News Media

Singh, Simar (School: Stockdale High School)

In an era where news media significantly shapes public perception, frames societal belief systems, and influences decision-making during responses to public health and global crises, there is an urgent need for equitable, unbiased journalism. Despite the positive impact of news media platforms, they have also contributed to widespread misinformation which amplifies bias and toxicity. This research aims to address the urgent need for bias and toxicity detection tools in identifying types of biases, levels of bias, and toxicity within news articles. This is the first study to leverage three machine learning models using NLP techniques such as Logistic Regression, XGBoost, and BERT (Bidirectional Encoder Representations from Transformers) which is a cutting edge, state of the art, large language model. I hypothesized that BERT would outperform other models in detecting nuanced bias and toxicity in journalistic text. My models accurately identified bias type, bias level, and toxicity in news articles, with performance evaluated through accuracy, F1 score, confusion matrices, and SHAP values. The fine-tuned BERT model surpassed Logistic Regression, XGBoost, and even outperformed the prior SimCSE model, setting a new benchmark for bias detection in media. Lastly, I applied my BERT model on a set of Mpox related articles, illustrating and reproducing its ability to flag bias and toxicity. This showcases its potential as a groundbreaking editorial tool that can equip journalists to edit and refine their content for fairness and objectivity before publication. Future work will integrate this cutting edge BERT model within news media outlets to promote factual, equitable, and scientifically substantiated information which is critically necessary in the current media landscape.

Awards Won: