



CALIFORNIA SCIENCE & ENGINEERING FAIR 2018 PROJECT SUMMARY

Name(s) Nitya Parthasarathy	Project Number 38094
Project Title BiasCheck: An Artificial Intelligence Based Tool to Evaluate Bias in Social Media	
Abstract Objectives/Goals The objective of this study is to establish a scientific basis using Artificial Intelligence (AI) techniques to study various forms of bias and stereotypes in social media. Methods/Materials Computer for coding and running AI programs. Public domain datasets were obtained from the Internet. Results New statistical metrics, some of which were adapted from diverse areas such as Information Theory and Language Modelling were introduced to evaluate gender bias in social media. These models were then further substantiated with novel algorithms using AI for bias prediction. Using Bayesian models as well as numerous sophisticated Neural Networks, the effectiveness of AI algorithms in studying biased text is then demonstrated on large social datasets. Incorporating these ideas, a web-based BiasCheck software is developed to automatically assess a BiasScore for any blog, webpage or document. Though particular emphasis is placed on gender bias evaluation, results are shown to readily extend to other types of bias evaluation. Conclusions/Discussion Comprehensive results were provided to demonstrate the presence of male and female gender stereotypes in social media. Furthermore, statistical techniques identified positive social sentiment for gender associated with specific behavior. Female gender was generally identified with softer roles while male gender was identified with leadership roles. AI algorithms (using numerous classifiers) developed were able to pick up this bias and aptly identify the gender in a sentence from surrounding words. In particular, female stereotypes was picked up more accurately indicating the presence of more overt bias for the female gender. Further, AI models also yielded interesting insights into social behavioral perceptions whereby a providing man was identified as successful whereas a providing woman was tagged as delicate! Existence of bias in movie review datasets as a function of movie genre was also evaluated and shown to be more prevalent in specific categories. A valuable social commentary is also provided by studying the evolution of bias over time. In summary, a new direction of applying statistical techniques and AI for social good has been established in this work uncovering a rich set of topics for future study.	
Summary Statement A comprehensive scientific basis is developed for evaluating bias in social media using novel statistical techniques and Artificial Intelligence algorithms.	
Help Received I developed and coded the AI algorithms myself. I discussed results with Prof. Sameer Singh in the department of computer science at UCI.	