Bias is a prejudice that can be held against a thing, person, or group which is usually based on previously held beliefs or societal conventions. It is a concept that can greatly impact many different aspects of our society, leading to harmful actions such as race, gender, and economic discrimination. One example of this was shown in the study performed by Marriane Bertrand and Sendhil Mullainathan which looked at the response rate for job applicants with traditionally white names against those with traditionally African American names. This study found significant differences in the call back rates for applicants based on the race associated with their names. The significance of the results from this study show how impactful and harmful bias can be whether intentional or not. Furthermore, the study shows that it is so important to find ways in which to identify the biases that exist in order to help reduce their impact on society. Brian Nosek and Mahzarin Banaji also performed a study looking bias by studying how the group association of their subjects and stereotypical associations of certain subjects with certain groups impacted the attitudes of each group towards the subject. This study found that those who associated themselves with the female group felt more negative attitudes towards math when presented with the stereotype between males and math. This shows that another impact of stereotypes and biases is that they can greatly influence the attitudes of a certain group towards a certain subject. These biases and many others that exist in our society are also likely represented in many of the web pages that exist on the internet. Due to the widespread reach of the internet, this allows for the biases on these web pages to be further perpetuated throughout society, reaching an even wider audience. This project is a web plug-in which is able to detect and highlight such biases in web pages, as well as provide statistics regarding the bias content on web pages.

This project will identify biases using a machine learning algorithm trained on data sets which contain biases. Finding a sufficient data set is one area that presents a technical difficulty because it can be difficult to find data sets containing the necessary information. It is even more difficult to find a large enough data set to sufficiently test and train the algorithm. Using other research papers which relate to the technology we are creating, such as Ailyn Caliskan and others' paper, will allow us to discover where we may be able to find the data sets needed in order to test and train their algorithm. Their article describes using pre trained GloVe embeddings to train their algorithm. The GloVe algorithm contains biases in the form of vectors that are created by finding the associations between words based on how commonly they are found together. The algorithm can then be implemented into our web plug-in and used to parse web pages and determine the biases that exist in the web pages.

The technology that this project is creating to detect bias can be used in many beneficial ways. The harmful impacts that biases can have is one reason why it is so important to create technology that can detect them. By being able to detect the biases that exist on web pages it will allow for more awareness of their existence in the information presented online. In a time where people have begun to question the legitimacy of the information provided in news reports, this technology can be used by those who use the internet as the source of their news in order to detect the biases that exist within the content they are consuming. Being able to detect these biases will make people more aware of their presence and influence in the information that they are consuming. Companies can also use this technology in order to evaluate the potential biases displayed in the information on the internet that relates to them. As discussed above, viewing biased information can impact the attitudes of certain groups towards the subjects associated with the bias. Giving companies the ability to identify the biases associated with them allows them to evaluate how they are portrayed online. They can then use this information to further evaluate how the biases portrayed may impact how they are viewed by possible clients and consumers.

Works Cited

1. A. Caliskan, J. J. Bryson, and A. Narayanan. Semantics de- rived automatically from language corpora contain human- like biases. *Science*, 356(6334):183–186, 2017.
2. B. A. Nosek, M. Banaji, A. G. Greenwald, Group Dyn. 6, 101–115 (2002).

3. M. Bertrand, S. Mullainathan, Am. Econ. Rev. 94, 991–1013 (2004).