Thesis Project Portfolio

Methods to Reduce Bias in Machine Learning Applications

(Technical Report)

Unfairness In Machine Learning: Reaching an Ethical Closure

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Henry T. Todd

Spring, 2023

Department of Computer Science

Table of Contents

Executive Summary

The development of artificial intelligence has gone largely without regulation and is highly dependent on the data sets from which it is trained. The technical paper considers potential methods to reduce the propagation of bias in artificial intelligences from big data through use of traditional statistical methods, while the STS paper considers the ethical use of AI systems, how bias in AI affects society, and the need for and feasibility of regulation. The technical report provides a review of current research on the limitations of AI due to its reliance on big data and analyzes the potential use of statistical methods to mitigate them. Methods for mitigating bias include collecting better data sets, reducing bias in preexisting data sets, and providing context on training data alongside outputted predictions. The product of the paper will be an analysis on the effectiveness of such methods on reducing error propagation from big data. The STS paper discusses the meaning of ethical AI, current issues relating to bias, and considers the current regulations in place as well as the potential for future regulations. The paper analyzes how the reduction of biased artificial intelligence might be enforced at a policy level.

Artificial intelligence algorithms are inherently tied to the data from which they are trained, however developers of AI methods make little use of traditional statistical methods to provide context to output decisions. While both the technical report and the STS research paper consider the issue of bias in artificial intelligence, the former looks at the data sets provided to algorithms as a place to minimize the bias propagated by AI. Traditional statistical methods can be used to analyze the bias of these data sets, which can both provide context to the algorithms which use them as well as be used to "clean" the data. These cleaned data sets can then be used to develop more ethical algorithms; however, issues arise from tampering with the data of preserving properties of the initial data set. The technical report is a review of these methods

currently under review as a means of reducing bias and identifies areas which need additional research.

Artificial intelligence poses novel complications in regulation as compared with previous data technologies. The development and use of artificial intelligence algorithms requires significant amounts of data in training and testing, as well as continued data input and output during use. These data sets may be biased or otherwise imperfect in ways that artificial intelligence algorithms do not adequately account for as opposed to traditional statistical applications. Whereas the technical report considers the details of these statistical methods and their applicability to reduce bias in AI, the STS research paper addresses the regulatory ability to mandate methods for reducing bias. The paper does so by analyzing whether previous legislation regarding data protection rights is sufficient to ensure ethical AI development and implementation, and considering what additional regulations have been proposed and might be effective. The social construction of artificial intelligence will be used to determine the various interpretations of artificial intelligence among social groups. The paper analyzes the social groups, their interpretation of AI, and how AI algorithms have developed in response. These findings help guide the future legislation governing the ethical use of data sets in the development of artificial intelligence algorithms, and what data these trained models can store and interact with.

The technical report and STS research paper provide a tightly coupled review of the issue of bias in artificial intelligence and consider means for reducing such bias and its impacts on society from two angles. The technical report focuses on the potential for statistical methods to provide context to biased outputs, adjust biased data sets, or even guide the collection of data to develop unbiased data sets. The STS paper then considers the legal side of enforcing artificial

intelligence. Writing the two papers in conjunction provided a more complete study considering all aspects of bias in artificial intelligence and the feasibility of reducing its impacts on society. The technical report details the difficulty in creating data sets which are not biased, while the STS paper indicates the difficulty in governing artificial intelligence as a specific technology as opposed to an application of data. The two papers consider the issue of bias at various levels. For example, the technical report is aimed at improving artificial intelligence, but does not specifically discuss AI methods. Instead, it focuses on the data used by such methods. The STS paper, on the other hand, considers the issues of bias inherent in artificial intelligence algorithms in addition to the data. The paper considers difficulties, as a result, of governing the development of AI outside of data protection rights.