Thesis Project Portfolio

Generating Custom Real-World Activity Data to Train an Artificial Intelligence Cloud Cybersecurity Model

(Technical Report)

Equity in Artificial Intelligence: Identifying Bias and its Causes in Intelligent Systems

(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 in Computer Science

Claire Williams

Spring 2024

Department of Computer Science

Table of Contents

Sociotechnical Synthesis

As Artificial Intelligence (AI) becomes more prevalent in our society, we must be increasingly aware of AI's inherent limitations, as AI can have serious consequences on marginalized populations when it introduces and perpetuates bias. My Science Technology and Society (STS) and technical papers explore the nature and origin of these biases and their effects. My technical report details a project automating the creation of model training data for a cloud cybersecurity application. My STS project addresses the questions of how we can identify bias in AI to prevent systematizing it and where these biases come from. Both projects highlight the importance of careful consideration of bias in AI.

In my technical component, I detail a software tool designed to automate the creation of training data for an intelligent model and address some of the major considerations during the early stages of the design process. Amazon Web Services (AWS) Detective is an AWS service that facilitates investigations of cybersecurity threats to a user's cloud resources. Detective does this by maintaining a detailed graph model of a user's activity such that users can more easily visualize and analyze their data. Detective runs algorithms on these activity graphs to help identify atypical patterns that may be related to security findings and, for Detective's security engineers to hone these algorithms, they need training data of dummy activity data. This generated training data needs to range from small and simple to very large and complicated to represent a range of customer needs. It must also be as similar to real data as possible to properly inform the model. Additionally, security engineers should be able to inject specific attacks into the activity data to see if the model identifies the attacks well. Through the design and implementation of this tool, I found that developers must be hyperaware of the training data inputted into a model as it greatly affects the model's abilities.

My STS research explores bias in AI more generally. The National Institute of Standards and Technology defines three types of bias in AI: statistical/computational biases, human biases, and systemic biases. These are the biases that can arise from training data, the humans building the AI, and existing societal biases that are "learned" by AI, respectively. In my STS research, I conduct a literature review and cite various real-world examples of these types of bias, including decreased accuracy of facial recognition technology for people of color, disproportionately punitive decision-making by predictive policing models, and language models learning stereotypical meanings behind words. When analyzing this evidence using the Social Construction of Technology (SCOT) framework, I concluded that it is paramount that bias in AI be seriously considered starting in the early stages of development and continuously throughout development. This is because bias in AI, when present, is fundamentally baked into the system starting from the onset. Therefore, the training data, human developers, and existing societal biases must be considered during the design phase, or as early as possible, to mitigate the possible consequences. If not, these biases will be codified by the system, making them significantly harder to identify and eliminate.

Both my STS and technical projects gave me an acute understanding of the effects of the inputs to an AI on its performance. My technical portion demonstrated some considerations when gathering training data and how foundational it is to the performance of a model. My STS research demonstrated the potential hazards that can come from training data, developers, and existing societal biases. Combining these perspectives illustrates the importance of critical design decisions in the early stages of development to create models that are both equitable and high-performing.