# Thesis Project Portfolio

**Complexity Class Analysis with Machine Learning**

(Technical Report)

**Privacy, Prejudice, and Pixels: A Journey Through the Implications of Machine Learning**

(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

**Mitchell Hong**

Spring, 2024

Department of Computer Science

## Table of Contents

Sociotechnical Synthesis

Complexity Class Analysis with Machine Learning

Privacy, Prejudice, and Pixels: A Journey Through the Implications of Machine Learning

Prospectus

## Sociotechnical Synthesis

Machine Learning (ML) has emerged as one of the most revolutionary technologies in the 21st century. Consequently, the questions of when, where, and by whom it should be used arise. The collective societal response to these questions is 'whenever, wherever, and by whoever,' which is completely justifiable. ML technology has enabled coastal countries to predict flood patterns, law enforcement to detect faces with high precision and accuracy, and automobiles to drive without any human input. The prowess of ML is unquestionable, but efficacy cannot be simply evaluated on performance – social and ethical factors must be considered. There exists a delicate balance between harnessing the transformative potential of ML and mitigating its adverse social impacts. A holistic approach is necessary in advancing the application of ML technology throughout society. My research covers two topics. My technical paper highlights a powerful use case of ML. My research paper explores the implications of ML on data privacy and bias within society.

My technical paper investigates the cutting-edge application of ML in the analysis of complexity classes. The study of complexity classes or the “hardness” of problems has been a leading field in computer science research. Efficiency of software is bounded to the complexity class of the underlying problem the software solves. There are many solutions to the same problem; some more effective than others. This is determined by decomposing a solution into its constituent parts and analyzing them to determine the spatial and computational upper bounds. Based off of these analyses, various problems have been sorted into their respective complexity classes over the last few decades. My paper proposes the introduction of ML into complexity class analysis. Leveraging a description of a problem along with its associated complexity class, a machine learning model can be trained to predict an arbitrary problem’s complexity class. With this, computer scientists can automate the categorization of problem difficulty and potentially find discrepancies between their intuition and the output of the model. ML has the potential to assist in determining whether P = NP, one of the most debated fields in Computer Science research. If P = NP, then the class of problems solvable in polynomial time (P) would be equivalent to the class of problems whose solutions can be verified in polynomial time (NP), meaning that every problem whose solution can be quickly verified could also be quickly solved. This would alter the computational landscape in an enormous manner.

My research paper examines the implications of ML technology on society. ML, while innovative and powerful, has the potential to harm its stakeholders. Models are inherently limited by the quality of data they train on – they utilize a vast amount of data, some of which is gathered without sufficient ethical and moral considerations. When trained with sensitive data, models have the potential to leak information, exposing critical information that was once secure. If a model is trained on biased data, the model will output biased results. In critical environments like healthcare and law enforcement, biased results are detrimental and will perpetuate bias in a rapid manner. Due to the complexity of ML technology, matters of data privacy and bias are difficult to resolve. My research examines these complications and challenges associated with legal regulations.

My research is confined to discussions within academia and my own analytical insights, rather than to practical field applications. While it covers the broad, overarching problems of ML technology, much more attention needs to be focused on the minute details of ML in application-specific contexts. In more detail, ML will be applied differently across various sectors, making it crucial to distinguish its uses, implementation methods, and the individuals it will impact. Future research should address these topics, using a more technical and experimental approach – instead of an overview, build a model and demonstrate infringements on data privacy and bias.

I would like to extend sincere gratitude to the UVA faculty for their instrumental role in my growth and progression throughout my undergraduate career. A special thanks to my STS advisor, Professor Caitlyn Wylie, whose guidance and insights were valuable in shaping this work.