CSC 2547HF – Fall 2019
AI and Ethics: Mathematical Foundations and Algorithms

Lectures: Tuesday 3-5, HS 106
Instructors: Toniann Pitassi (toni@cs.toronto.edu) & Richard Zemel (zemel@cs.toronto.edu)
Office hours: By appointment
Tutors: Elliot Creager (creager@cs.toronto.edu) & Alex Edmonds (edmonds@cs.toronto.edu)
Web Page: http://www.cs.toronto.edu/ toni/Courses/Fairness/fair.html

This is an introductory-level graduate course on social and ethical aspects of machine learning. Machine learning systems are becoming increasingly important in many domains where they are used to make predictions and decisions that often have life-altering consequences. Examples include machine learning algorithms for criminal sentencing, health insurance decisions, car insurance rates, and targeted advertising to name just a few. As these systems are becoming ubiquitious it is of extreme importance to address issues of privacy, fairness and accountability.

The focus of the course will be mathematical formalisms and algorithms for studying ethics in AI/ML systems. We will assume that the students have some background in machine learning, including both its algorithmic as well as theoretical aspects.

The course will delve into two important topics: privacy and fairness:

Fairness in ML. How can we develop classification algorithms that incorporate both human decision makers as well as ML systems, that are fair with respect to subgroups of a population? This is a rapidly growing field. We will discuss the main sources of unfairness, some definitions that have been proposed, and problems and tradeoffs with these definitions. Then we will study several approaches to fairness in classification, including: fair representational learning, achieving fairness in the multiclass setting, game theoretic approaches to fairness, dynamics of fairness, and causal approaches.

Privacy in ML. What kind of privacy should be guaranteed when using data to learn? We will discuss blatant abuses of privacy, and survey some common approaches to achieving privacy in ML, for example cryptographic methods, and differential privacy. We will review the basics of differential privacy, with a focus on differentially private learning (from both an applied as well as theoretical perspective), and how privacy is strongly connected to generalization.

Course Material: The topics and formalisms in this course are very new and evolving. still very new and that has fluid boundaries and evolving formalisms. We will provide a reading list, and hope that the presentations will provoke discussion, arguments, and new ideas. So please read the highlighted materials ahead of lecture and come prepared with your questions, comments and critiques. You will benefit the most from the material if you have time to engage with it.

Evaluation:

You are expected to come to class, and come prepared by reading the papers in advance. We will typically give a lecture on the paper, followed by a discussion.

Class attendance/participation (10% of grade)
2 assignments (each worth 25% of grade)
Poster presentation (40% of grade)

Schedule:

1) Sep 10: Introduction & Definitions of Fairness in CS
2) Sep 17: Paradoxes; Tradeoffs; Impossibility results
3) Sep 24: Fairness in Classification
4) Oct 1: Recent Approaches to Fair Classification
5) Oct 8: Privacy in AI Systems
6) Oct 15: Making ML System Private
7) Oct 22: Causality and Fairness
8) Oct 29: Fairness in Dynamic Settings

   Nov 5: Reading Week

9) Nov 12: Extending the Formulation: Multi-group Fairness
10) Nov 19: Fairness in Economics & Game Theory
11) Nov 26: Beyond Classification: Other Domains & Viewpoints
12) Dec 3: Poster Session