Algorithms for Responsible AI

Contact info:
Instructor: Prof. Juba Ziani
Email: jziani3@gatech.edu
Office: George Tower 1206

Class time and location:
TBD

Goals:
In this class, we will study the ethical and societal concerns that arise in big data, data-driven decision-making, and machine learning.

The course will be a project/research-based course. The first part of the class will introduce students to i) the basics of data privacy, focusing in particular on the current state-of-the-art notion of privacy---*differential privacy* then ii) bias and unfairness in machine learning and decision-making. Part 1 will cover formal algorithm and technical tools to get started in these areas of research. Part 2 of the class will be a mix of a class/research project (with possibility of eventually submitting the results of each student's project to a conference or a journal) as well as a reading group on fairness and privacy. Students will be expected to do in-class presentations on recent papers/research throughout Part 2 of the semester.

Learning objectives:
- Understanding the fundamental issues that stem from the use of sensitive data
- Being proficient with formal notions of privacy and fairness
- Understanding and applying formal tools to guarantee differential privacy and statistical/individual fairness.
- Being able to build a research agenda around privacy and/or fairness.

Pre-requisites:
Basic probabilities, basic machine learning (regression and classification), undergraduate algorithms, and mathematical proofs.
Understanding of randomized algorithms is a plus but not required.

Workload:
- Class participation (10%)
- Regular class presentations and write-ups on research papers (40%)
- Semester-long project on a topic of your choice related to the class (50%)
- No homework/exams

Recommended reading:

Differential Privacy.
"The Algorithmic Foundations of Differential Privacy" by Cynthia Dwork and Aaron Roth.

"*Differential Privacy in Artificial Intelligence: From Theory to Practice*" *by Ferdinando Fioretto and Pascal Van Hentenryck*

Fairness in machine learning.
"Fairness and Machine Learning: Limitations and Opportunities" by Solon Barocas, Moritz Hardt, and Arvind Narayanan.

Tentative (incomplete) List of Topics:

Introduction to Privacy and to Fairness
Definition and main idea behind Differential Privacy (DP)
Basic DP mechanisms Part I (randomized response, laplace)
Basic DP mechanisms part II (exponential mechanism)
Basic DP properties (group privacy, post-processing, composition).
Approximate DP and Gaussian Mechanism. Advanced Composition.
Formalizing fairness
Statistical Fairness (2 lectures: definitions, and incompatibilities)
Statistical Fairness case study: the COMPAS dataset