IEOR 8100 – Foundations of Data Privacy, Spring 2021

Instructor: Prof. Rachel Cummings
Times: Thursday, 4:10-6:40pm
Location: 325 Pupin Laboratory (or Zoom)
Office Hours: by appointment
Email: rac2239@columbia.edu
Course Website: TBA
Prerequisites: Undergraduate-level courses on probability, algorithms, and proof-based mathematics. Instructor's permission required for undergraduate and Master's students.

Format: This course will be in the HyFlex format. The first two lectures will be entirely virtual. The remaining lectures will be taught in-person in the classroom, and will also be available on Zoom for students who are not able to join in person. Attendance at the in-person lectures is encouraged but not required. If it becomes unsafe to continue in-person sessions, this course may move fully remote at some point in the semester. There will be a survey the first week of classes to determine the learning needs of all students: workload, time zone, being quarantined, and other situations that impact their learning. Adjustments may be made to the syllabus based on the findings of this survey.

Mandatory Face Coverings:
Wearing a face covering to cover the mouth and nose is required at all times while on Columbia University property. Face coverings that cover nose and mouth are required to be worn during all in-person components of the course. Face covering use will be in addition to and is not a substitute for social distancing. Anyone not using a face covering when required will be asked to wear one or must leave the area. Refusal to comply with the requirement may result in discipline through the applicable conduct code for faculty, staff or students. Reasonable accommodations may also be made for those who are unable to wear a face covering for documented health reasons. For more information about face masks and coverings, review the COVID-19 Resource Guide for the Columbia Community.

Description:
How should we define privacy? How can we enable the analysis of data containing sensitive information about individuals while protecting the privacy of those individuals? What are the tradeoffs between useful analyses of large datasets, and the privacy of the individuals from whom the data are derived? This course will take a mathematically rigorous approach to addressing these and other questions at the foundations of research in data privacy.

Over the past decade, a new line of work on *differential privacy* has provided a framework for computing on sensitive datasets in which one can mathematically prove that individual-specific information does not leak. In addition to showing that many useful data analysis tasks can be accomplished while satisfying a strong privacy requirement, this line of work has also shown that differential privacy is quite rich theoretically and has potential for a significant impact on practice. This course will draw connections between differential privacy and a wide variety of

topics, including economics, statistics, information theory, game theory, optimization, probability, learning theory, geometry, and approximation algorithms.

By the end of the course, you should be able to:

- Understand the state of the art in differential privacy at a level sufficient to engage in research, apply the material in practice, and/or connect it to other areas
- Extract both the high-level ideas and technical details when reading a mathematical text and identify interesting questions that are not answered
- Explain and collaboratively work through an advanced subject with your peers
- Formulate and carry out an interesting, short-term independent research project, and present the work in both written and oral form

Topics covered:

Topics 1-5 below will form the core content of the class. Topics 6-9 are applications of differential privacy to other fields. Some or all of these topics will be covered, based on student interest. Topic 10 will be covered if time permits.

1. The definition of differential privacy
    - Motivation & interpretation
    - Properties of differential privacy
    - What differential privacy does and does not promise
2. Basic differentially private mechanisms
    - Randomized response
    - The Laplace mechanism
    - The exponential mechanism
    - Advanced composition theorems
3. Answering many queries with differential privacy
    - Sparse vector
    - SmallDB
    - Median mechanism
    - Private multiplicative weights
4. Average-case privacy
    - Propose-Test-Release
    - Subsample and Aggregate
    - Lipschitz extensions
5. Variant models of differential privacy
    - Local model vs centralized model
    - Privacy in federated learning
    - Reyni/concentrated differential privacy
6. Differential privacy & economics
    - Joint differential privacy
    - Private equilibrium computation
    - Markets for data
7. Differential privacy & optimization
    - Privacy-preserving linear/convex programming

    - Privacy-preserving ERM and regression
8. Differential privacy & statistics
    - Privacy as a tool for generalization
    - Reusable holdout
9. Differential privacy & learning theory
    - Privacy of learning tasks
    - Boosting and differential privacy
    - Query release vs. agnostic learning
10. Privacy problems outside the scope of differential privacy
    - behavioral tracking and targeted advertising
    - discrimination
    - surveillance
    - fairness

Grading and Format:
The main graded components of the course are as follows:

- 3 Homework assignments (10% each)
- Scribe notes (10%)
- Participation and engagement (20%)
- Final project (5% for the proposal, 5% for the presentation, and 30% for the write-up)

This breakdown may change during the term, particularly as enrollment levels settle.

Homework:
Homework assignments will be assigned roughly every two weeks in the first half of the semester, and will either be due at the beginning of class or will be submitted electronically. Each assignment will specify the due date and the submission method. These assignments will be a mix of problem sets and guided research problems, intended as a gentle introduction to research in differential privacy. You may not copy solutions directly from any source, but are encouraged to use academic references (e.g., textbooks, research papers) as needed, particularly for the research-oriented assignments.

Scribe notes:
You will be assigned one lecture to take scribe notes. Your job will be to type up notes (in LaTeX) from that lecture, to be shared with the rest of the class. You'll have one week from the date of the lecture to submit your .tex, .pdf, and .bib files. I'll provide a LaTeX template for your notes.

Final project:
You will be expected to do a final project on a topic of your choosing, either individually or in small groups. Your project should provide good opportunities to connect the course material to your other interests and get some exposure to doing original research in differential privacy. This will involve submitting a detailed project proposal for feedback, completing the proposed project, producing a written paper, and presenting your project in class at the end of the semester. Student presentations may take place on the last day of classes. More details will be posted early in the semester about project requirements and suggested topics.

Exams:
There will not be any exams for this course.

Textbook:
"The Algorithmic Foundations of Differential Privacy" by Cynthia Dwork and Aaron Roth. The pdf version is available for free; the paper version is on Amazon for $99. We will also be reading research papers, which will be posted on the course website throughout the semester.

Academic Honor Code:
Each assignment will specify the extent to which collaboration is encouraged or allowed. Please refer to Columbia University's Graduate Engineering Honor Code here: https://www.gradengineering.columbia.edu/academics/academic-integrity

Office of Disability Services:
Columbia has policies regarding disability accommodations, which are administered through Columbia Health and Disability Services (https://health.columbia.edu/content/disability-services). If you require special accommodations, please notify me ASAP.