IEOR 4575 – Policy for Privacy Technologies, Fall 2021

Instructor: Prof. Rachel Cummings
Times: Tues/Thurs, 2:40-3:55pm
Format: In-person
Location: 415 Schapiro CEPSR
Office Hours: Tues/Thurs, 4-4:30pm or by appointment
Office: Mudd 535E
Email: rac2239@columbia.edu
Course Website: TBA
Prerequisites: None. Students should have background in either algorithms or policy, but are not expected to have background in both. No prior knowledge of privacy will be assumed.
Students from any relevant department or degree program are encouraged to enroll.

Description:
Privacy concerns are becoming a major obstacle to using data. It's often unclear how current regulations should translate into technology, and the changing legal landscape surrounding privacy can cause valuable data to go unused. How can data scientists make use of potentially sensitive data, while providing rigorous privacy guarantees to the individuals who provided data? The computer science and privacy engineering communities have developed a rich library of technical tools to address this problem. However, each privacy technology satisfies its own formalization of "privacy" or "security", and many come with privacy parameters to quantify precisely how much privacy is being provided.

The widespread deployment of these tools to new application domains brings about new policy challenges, such as deciding which privacy technology is appropriate for a particular use-case, determining appropriate values of relevant privacy parameters, and explaining technical – and oftentimes nuanced – privacy guarantees to non-experts. This course will survey the technical and policy-relevant details of existing privacy technologies, including anonymization, differential privacy, encryption, secure multiparty computation, regulation, data minimization, contextual integrity, and others. The course format will combine technical lectures with real-world and case-based discussion of the policy decisions that arise when implementing each privacy technology.

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

- Understand the key concepts, strengths, weaknesses, and primary use cases for several commonly-used privacy technologies
- Discuss policy considerations for the deployment of several commonly-used privacy technologies
- Analyze real-world business scenarios to identify appropriate privacy technologies for the given use case
- Collaboratively create policy recommendations for privacy technologies with your peers, backed by technical justifications, and present the work in both written and oral form

Topics covered:
Topics 1-4 below will form the core content of the class. Topics 5-7 will be covered briefly if time permits.

1. Anonymization
    - Personally Identifiable Information (PII)
    - k-anonymity and l-diversity
    - Famous "failings" of anonymization
    - Attacks against anonymized databases
2. Differential privacy
    - Privacy definition and properties
    - Algorithms for achieving differential privacy
    - Local vs central models of DP
    - Challenges with choosing privacy parameters
    - Challenges and opportunities for deploying DP in practice
3. Cryptography
    - Security definition and properties
    - Multi-party computation
    - Homomorphic encryption
    - Challenges and opportunities for regulating crypto
    - Relationship between DP and crypto
4. Regulation
    - GDPR/CCPA
    - "Singling out"
    - Individual vs aggregate data
    - Data memorization and deletion
    - Relationship with DP and crypto
5. Federated learning and data minimization
    - Data flow model
    - Relationship with DP and crypto
6. Contextual integrity
    - Formal definition
    - Relationship with DP and crypto
7. Synthetic data
    - Benefits and challenges of using synthetic data
    - Private algorithms for generating synthetic data

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

- Homework assignments (40%: 4 assignments at 10% each)
- Participation and engagement (25%)
- Course project (20%)
- Final exam (15%)

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

Homework:
There will be four case-based assignments, one for each of the major topics covered in the course (Anonymization, Differential privacy, Cryptography, and Regulation). In each assignment, you will be faced with a policy decision related to the topic, and will be asked to use a combination of technical and policy tools to discuss the perspectives and challenges, make your decision, and convincingly argue the wisdom of your decision. These assignments will be done with a partner, and are intended to mimic real-world policy decisions related to applications of privacy technologies. You will be encouraged to use external academic references (e.g., textbooks, research papers, legal texts) as needed, but you must cite all of your sources. Each assignment will specify the due date and the submission method.

Participation:
This course will be highly interactive and discussion-based. This is an in-person course. There will not be an option to participate via Zoom (except in special circumstances), and lectures will not be recorded. While attendance is not mandatory, it is strongly encouraged, as much of the course content cannot be replaced with external readings. Students will also be better equipped to engage in class discussions if they have completed the assigned readings before each lecture. After each homework assignment is submitted, one lecture will be dedicated to discussion of the policy decision presented in the assignment. Students are strongly encouraged to discuss their reasoning and final decision, and to debate (collegially!) with others who advocated for different decisions. Students may also be asked to complete "mini-assignments" including, but not limited to, peer-evaluations, self-reflections, and other pedagogic activities.

Course project:
The course project will be similar in style to the assignments, but of a larger scope that enables students to engage with and apply multiple privacy technologies simultaneously. The project will present a case-based scenario inspired by real-world privacy applications, and students will work in groups to apply course content to develop a policy recommendation supported by technical and legal arguments. The project will include a written component and a group presentation in class at the end of the semester. More details will be posted early in the semester about the project topics and requirements.

Final exam:
The final exam for this course is tentatively scheduled for Thursday December 16, 2021 at 1:10-4pm, according to the Columbia University Final Exam Schedule. The exam will be written and will be comprehensive of all material covered in the course. Missing the exam will be accommodated only in case of a pre-arranged, University-approved absence.

Textbook:
There will not be any textbooks for this course. Assigned readings will be provided by the instructor, and may include book chapters, research papers, news articles, legal documents, policy briefs, and videos.

COVID-19 Policy:
All students and instructors are expected to comply with Columbia University guidelines related to the COVID-19 pandemic, to ensure the health and safety of all class participants. This

includes, but is not limited to, face coverings, social distancing, vaccination, testing, staying home if you feel sick, and using the ReopenCU App. These policies may evolve throughout the semester based on CDC and New York State guidance. Students are encouraged to review the COVID-19 Resource Guide for the Columbia Community for up-to-date policies.

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, make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Course Schedule

Date	Lecture topic	Readings	Assignments
Thurs Sept 9	Intro to privacy policy	The Ethics of Privacy Protections	
Tues Sept 14	Anonymization 1: Definitions and legal foundations	Simple Demographics Often Identify People Uniquely	HW 1 out
Thurs Sept 16	*No class – Yom Kippur*		
Tues Sept 21	Anonymization 2: Linkage attacks	How To Break Anonymity of the Netflix Prize Dataset	
Thurs Sept 23	Anonymization 3: Reconstruction attacks	Linear Program Reconstruction in Practice	
Tues Sept 28	Anonymization discussion		HW 1 due
Thurs Sept 30	Differential Privacy 1: Definition and properties	The Algorithmic Foundations of Differential Privacy, Ch 1 and 2	HW 2 out
Tues Oct 5	Differential Privacy 2: DP algorithms	The Algorithmic Foundations of Differential Privacy, Ch 3.3 and 3.4	
Thurs Oct 7	Differential Privacy 3: Local vs central model	The Algorithmic Foundations of Differential Privacy, Ch 3.2 (easy) What Can We Learn Privately? (advanced) and links in slides	
Tues Oct 12	Differential Privacy 4: Choosing epsilon, explaining DP	"I need a better description": An Investigation Into User Expectations For Differential Privacy	
Thurs Oct 14	Differential Privacy 5: DP and the Census	See links in slides	
Tues Oct 19	DP discussion		HW 2 due
Thurs Oct 21	Cryptography 1: Private-key crypto	See links posted on Ed Discussion	HW 3 out
Tues Oct 26	Cryptography 2: Multi-party computation	See links posted on Ed Discussion	
Thurs Oct 28	Cryptography 3: Public-key crypto	See links posted on Ed Discussion Watch extra lecture video	
Tues Nov 2	*No class – Election Day*		
Thurs Nov 4	Cryptography 4: 5th Amendment (Guest Lecture by Aloni Cohen)	Compelled Decryption and the Fifth Amendment: Exploring the Technical Boundaries	

Tues Nov 9	Crypto discussion		HW 3 due
Thurs Nov 11	Regulation 1: Legal definitions of privacy (Guest Lecture by Alex Wood)	Bridging the Gap between Computer Science and Legal Approaches to Privacy Section 4	HW 4 out
Tues Nov 16	Regulation 2: GDPR and Singling out	Towards Formalizing the GDPR Notion of Singling Out	
Thurs Nov 18	Regulation 3: Aggregates, memorization, deletion	The Role of Differential Privacy in GDPR Compliance (easy) The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks (advanced)	
Tues Nov 23	Contextual Integrity	Contextual Integrity Wikipedia page (easy) Contextual Integrity through the Lens of Computer Science Sections 1 and 2	
Thurs Nov 25	*No class – Thanksgiving*		
Tues Nov 30	Regulation discussion		HW 4 due
Thurs Dec 2	Data minimization / Federated learning	Advances and Open Problems in Federated Learning Sections 1 and 4-4.2	
Tues Dec 7	Student presentations		
Thurs Dec 9	Student presentations		Project due
Thurs Dec 16, 1:10-4pm	Final Exam		Final exam