ORCS 4201 – Policy for Privacy Technologies, Spring 2024

Instructor: Prof. Rachel Cummings
Times: Mon/Wed, 5:40-6:55pm
Format: In-person
Location: Fayerweather 313
Office Hours: Tues 2:15-3pm & Thurs 4:15-5pm
Office: Mudd 535E
Email: rac2239@columbia.edu
TA/CA: Tingting Ou (TA) to2372@columbia.edu OH Wed 2:30-3:30pm & Thurs 2-3pm in Mudd 333 and Hari Bhimaraju (CA) mhb2189@columbia.edu Mon 2:30-3:30pm & Wed 10:45-11:45am virtual (Zoom link on Courseworks)
Course Website: Courseworks
Prerequisites: None. Mathematical maturity is expected. 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. **Mathematically mature 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.

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
    - Private-key and public-key encryption schemes
    - Multi-party computation
    - 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. Contextual integrity
    - Formal definition
    - Application to privacy policies
6. Synthetic data
    - Benefits and challenges of using synthetic data
    - Private algorithms for generating synthetic data
7. Federated learning and data minimization
    - Data flow model
    - Relationship with DP and crypto

Grading:

The main graded components of the course are as follows:

- Participation and engagement (20%)
- Smaller technical assignments (15%)
- Homework assignments (32%: 3 assignments respectively at 8%, 12%, 12%)
- Final project (18%)
- Final exam (15%)

Participation, engagement, and attendance:
This course will be highly interactive and discussion-based. There will be "mini-assignments" accompanying each class period, to be completed before, during, and after class. Students must complete these assignments on time to fully participate in class discussions and activities.

Regular class attendance is expected, especially because learning depends upon the preparedness and participation of us all. This is an in-person course. There will not be an option to participate via Zoom (except in special circumstances), and lectures will typically not be recorded.

Beyond these mini-assignments and class attendance, participation also includes adhering to classroom expectations, such as engaging in discussions, coming to class prepared, not speaking when someone else is presenting, arriving on time to class, not disturbing others around you, and so on. Violation of these classroom expectations may result in reduced participation scores.

To make up in-class assignments if you have an excused absence, you must do the following: (1) Submit all assignments on time that are due before the class period which will be missed, (2) Email Prof. Cummings NO LATER than 10am on the day you will miss class, and (3) Make-up the in-class assignments (when applicable) on your own and submit this work within 48 hours. Due to the participatory nature of this course, not all in-class activities can be made up. These determinations will be made at the discretion of the instructor.

Technical assignments:
The course will also contain several small technical assignments to ensure comprehension of the technical concepts covered in class. These will be of the same size as the participation "mini-assignments" but will be graded on correctness rather than completeness/effort.

Homework assignments:
There will be three case-based assignments, one for each of the first three major topics covered in the course (Anonymization, Differential Privacy, and Cryptography). These three assignments will be substantially larger than the other graded components, and will be broken down into multiple sub-parts, each with their own due dates that will span multiple class periods. See the Course Schedule at the end of this Syllabus for projected deadlines.

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 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.

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 course will have an in-class final exam that tests both understanding of the technical concepts covered in the course, and the application of technical concepts to real-world business and policy settings. The content of the final exam will follow closely to material covered in class and in all graded activities throughout the semester. The final exam for this course will be held during the course's assigned final exam slot, which is tentatively scheduled for Wednesday May 8, 2022 at 4:10-7pm, according to the Columbia University Final Exam Schedule. No make-up exams or alternative exam times will be given.

Textbook and readings:
There will not be any textbooks for this course. Assigned required readings will be provided by the instructor, and may include book chapters, research papers, news articles, legal documents, policy briefs, and videos. Most lectures will also have optional "beyond-the-scope" readings as an additional reference for those who want to learn more about the topic. It is emphasized that these are not required readings, as they are beyond the scope of what we cover in class

COVID-19 Policy:
Everyone is expected to comply with Columbia University guidelines related to the COVID-19 pandemic, to ensure the health and safety of all class participants. Current CDC guidelines for fully vaccinated individuals who are exposed to COVID-19 but are asymptomatic no longer recommend isolation. For fully vaccinated individuals who are symptomatic or test positive for COVID-19 are recommended to isolate for 5 days. 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 – subject to change

	Date	Lecture Topic	In-class	Homework
1	Wed Jan 17	Intro to privacy policy	D,C	S,D
2	Mon Jan 22	Anonymization 1: Legal foundations and k-anonymity	C,D,T	T,M,R
3	Wed Jan 24	Anonymization 2: Linkage attacks	D,D,C,D	C,D,R,C,V
4	Mon Jan 29	Anonymization 3: Reconstruction attacks	P,P,P	HW1.1*
5	Wed Jan 31	Anonymization 4: Homework 1 discussion	D*	HW1.2*
6	Mon Feb 5	Contextual Integrity	D	V,C
7	Wed Feb 7	DP 1: Intro to DP, RR	D,C,D	V,R,D,M
8	Mon Feb 12	DP 2: Laplace and Exponential	P,T	R,T,T,V
9	Wed Feb 14	DP 3: Local/central models, DP in practice	D,C,D	D,C,R
10	Mon Feb 19	DP 4: DP at the Census	D*,C	C,HW2.1*
11	Wed Feb 21	DP 5: DP at LinkedIn (A) and (B)	D*,D*,C	HW2.2*
12	Mon Feb 26	DP 6: DP at LinkedIn (C)	D*	HW2.3*,S
13	Wed Feb 28	Synthetic data	C	
14	Mon Mar 4	Crypto 1: Security defs, private key crypto	T,T,C	
15	Wed Mar 6	(Virtual) Crypto 2: public key crypto	T	T
16	Mon Mar 18	Crypto 3: Cryptography's greatest hits	T,D,C	D
17	Wed Mar 20	Crypto 4: Secure MPC	T,C	C,R,F.1*
18	Mon Mar 25	Crypto 5: Legal implications of crypto	D,C	HW3.1*
19	Wed Mar 27	Crypto 6: MPC at BWWC	D*	HW3.2*
20	Mon Apr 1	Federated learning	D,D,C	
21	Wed Apr 3	Law 1: Law & Cryptography	D*	R,C
22	Mon Apr 8	Law 2: 4th and 5th Amendment (guest lecture)		F.2*
23	Wed April 10	Law 3: GDPR & machine learning	D,C	D
24	Mon April 15	Law 4: Privacy laws in the U.S.	C	T
25	Wed April 17	Law 5: ADPPA	D*	
26	Mon April 22	(Optional) In-class workday		
27	Wed April 24	Student presentations	F.3*	
28	Mon April 29	Final exam review		S,F.4*

Key:

In-class:
D=Discussion with partner or small group
C=Courseworks quiz
P=PollEverywhere
T=Technical problem solving

*=Larger assignment

Homework:
S=Survey
D=Discussion board
M=Muddiest point
C=Courseworks quiz
T=Technical assignment
V=Video
R=Required reading
HW=Homeworks 1,2,3 and their sub-parts