USC Viterbi
School of Engineering

CSCI 699: Privacy-Preserving Machine Learning

Units: 4.0
Fall 2024 — Fridays —1:00-4:20PM

Location: DMC 200
Location: https://spkreddy.org/ppmlfall2023.html

Instructor: Sai Praneeth Karimireddy
Office: TBD
Office Hours: By appointment
Contact Info: karimire@usc.edu ; I will reply within 48 hours. Please include "CSCI 699" in your email subject.

Catalogue Description
Foundations of privacy in machine learning: differential privacy; private training of ML models; privacy attacks and audits; federated and decentralized machine learning.

Course Description
This course focuses on the foundations of privacy-preserving machine learning. Extremely personal data is being collected at an unprecedented scale by ML companies. While training ML models on such confidential data can be highly beneficial, it also comes with huge privacy risks. This course addresses the dual challenge of maximizing the utility of machine learning models while protecting individual privacy. We will cover the following topics: differential privacy; private training of ML models; privacy attacks and audits; federated and decentralized machine learning.

Learning Objectives
This course will prepare you to rigorously identify, reason about, and manage privacy risks in machine learning. You will learn to design algorithms that protect sensitive information, and to analyze the privacy leakage of any ML system. Additionally, the course will introduce you to cutting-edge research and practical applications. By the end of the course, you will be well-equipped to undertake research and address real-world privacy challenges in machine learning.

Recommended Preparation
Knowledge of advanced probability (at the level of MATH 505a), linear algebra and multi-variable calculus (at the level of MATH 225), analysis of algorithms (at the level of CSCI 570), introductory statistics and hypothesis testing (at the level of MATH 308), and machine learning (at the level of CSCI 567).

Course Notes
Grading type: Letter or Credit No/Credit

Technological Proficiency and Hardware/Software Required
You will need a laptop which can run PyTorch for this course – bring your laptop to class. Link information for the USC Computing Center Laptop Loaner Program.

Required Readings and Supplementary Materials
There are no required textbooks. The following writeups are excellent supplemental readings and may be used as references.

- C. Dwork and A. Roth, The Algorithmic Foundations of Differential Privacy, Foundations and Trends in Theoretical Computer Science, 2014. Reference for DP.
- Nissim et al., Differential Privacy: A Primer for a Non-technical Audience, Journal of Entertainment & Technology Law, 2018. Great read with many examples tying legal definitions and privacy in practice.
- Kairouz et al. Advances and Open Problems in Federated Learning. Community survey on federated learning.

Optional Readings and Supplementary Materials
This course builds on several related courses, which can serve as valuable additional references:

- Privacy-Preserving Machine Learning, by Aurelien Bellet at Inria (link)
- Trustworthy Machine Learning, by Reza Shokri at NUS (link)
- Federated and Collaborative Learning, by Virginia Smith at CMU (link)
- Large Scale Optimization for Machine Learning (ISE 633), by Meisam Razaviyayn at USC (link)
- Digital Privacy, by Vitaly Shmatikov at Cornell (link)
- More resources on Differential Privacy (link)

Description of Assignments and How They Will Be Assessed

1. 3 assignments worth 30% of the grade. Collaboration is allowed but must be stated. Grades are based on correctness. The theory part should be written in Latex and coding part in Jupyter python notebooks.
2. Course Presentation and Project (55% of the grade): The course includes a major component comprising a presentation (25%) and a project (30%) on a topic in privacy-preserving machine learning.
    a. Presentations (25%): These are crucial as they will expose you to cutting-edge ML theory research. You will be assigned a paper based on your interest and will present it in class for 30 minutes.
    b. Project (30%): You will write a 4 page report on 1-2 papers. This could either be on the paper you presented, supplemented by related readings, or on a different paper(s) of your choice. Pursuing your own research topic is also a strongly encouraged option; please discuss this with me if you are interested.
3. Discussions and participation will count for 15%. You will review, comment, and discuss each other's presentations and projects. We will use the role-playing reading group format introduced by Alec Jacobson and Colin Raffel.

Participation

Discussions and participation will count for 15%. You will review, comment, and discuss each other's presentations. We will use the role-playing reading group format originally introduced by Alec Jacobson and Colin Raffel.

Grading Breakdown

Assessment Tool (assignments)	% of Grade
Assignments	30%
Final project presentation	25%
Final project report	30%
Discussions and participation	15%
TOTAL	

Assignment Submission Policy

Via Gradescope.

Attendance

There is no explicit need for attendance, but participation and discussion accounts for 15% of the grade.

Academic Integrity

Unless otherwise noted, this course will follow the expectations for academic integrity as stated in the USC Student Handbook. The general USC guidelines on Academic Integrity and Course Content Distribution are provided in the subsequent "Statement on Academic Conduct and Support Systems" section. For this class Collaboration is allowed but must be explicitly acknowledged. Please ask the instructor [and/or TA(s)] if you are unsure about what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

You may not record this class without the express permission of the instructor and all other students in the class. Distribution of any notes, recordings, exams, or other materials from a university class or lectures —

other than for individual or class group study — is prohibited without the express permission of the instructor.

Use of Generative AI in this Course

Generative AI is not permitted: Since creating, analytical, and critical thinking skills are part of the learning outcomes of this course, all assignments should be prepared by the student working individually or in groups as described on each assignment. Students may not have another person or entity complete any portion of the assignment. Developing strong competencies in these areas will prepare you for a competitive workplace. Therefore, using AI-generated tools is prohibited in this course, will be identified as plagiarism, and will be reported to the Office of Academic Integrity.

Course Evaluation

Course evaluation occurs at the end of the semester university-wide. In addition, we will use continuous anonymous feedback to help shape the course to your best interests. Remember, this is an advanced level course and will be what you make of it.

Course Schedule

	Topics/Daily Activities	Deliverables/due dates
Week 1	Theory: Introduction to anonymity and data privacy; Data anonymization techniques; De-anonymization attacks; Linkage and Reconstruction attacks; Practical: Implement some linkage attacks **(bring laptop)**	
Week 2	Theory: Differential Privacy; Randomized response; Laplace mechanism; Hypothesis testing interpretation Practical: Implement differential privacy defenses **(bring laptop)** **HW 1 out.**	
Week 3	Theory: ML training; gradient descent; SGD Practical: SGD vs. Adam vs. DP-SGD on PyTorch **(bring laptop)**	**HW 1 due before class.**
Week 4	Theory: Private ML training; DP-SGD; Gaussian DP; Sub-sampling; Composition Practical: Opacus Library for private deep learning **(bring laptop)** **HW 2 out.**	
Week 5	Theory: Practical Privacy auditing; Designing powerful membership inference attacks; Measuring the influence of training data Practical: DP-auditing **(bring laptop)**	**HW 2 due before class.**
Week 6	Theory: Privacy in LLMs; RLHF/prompt engineering for privacy; Data stealing attacks; private in-context learning; reconstruction attacks Practical: Privacy defending and attacking prompts **(bring laptop)** **HW 3 out.**	
Week 7	Theory: Unlearning algorithms; guarantees; data washing; Concept forgetting Practical: implement unlearning **(bring laptop)**	**HW 3 due before class.**
Week 8	Theory: Decentralized privacy; Local DP; Multi-party computation; Trusted execution environments Practical: Comparing local vs. central DP **(bring laptop)**	
Week 9	Theory: Federated learning; challenges due to data heterogeneity, communication compression; Privacy attacks in FL Practical: federated learning on hospital data **(bring laptop)**	
Week 10	Theory: Privacy in FL; Secure aggregation; Quantized DP; Practical: DPFL vs. Local DP **(bring laptop)**	

Week 11	Privacy in Practice; Technical vs. legal notions; Incentives; Relation to Copyright law and Fourth Amendment; Case study: Genomic privacy	
Week 12	Student presentations	In class presentations. Option to schedule it earlier in the semester as well.
Week 13	Student presentations	In class presentations. Option to schedule it earlier in the semester as well.
Week 14	Student presentations	In class presentations. Option to schedule it earlier in the semester as well.
Week 15	Student presentations	In class presentations. Option to schedule it earlier in the semester as well.
FINAL	Final project report	Report due on the university-scheduled date of the final exam.

Statement on Academic Conduct and Support Systems

Academic Integrity:
The University of Southern California is a learning community committed to developing successful scholars and researchers dedicated to the pursuit of knowledge and the dissemination of ideas. Academic misconduct, which includes any act of dishonesty in the production or submission of academic work, comprises the integrity of the person who commits the act and can impugn the perceived integrity of the entire university community. It stands in opposition to the university's mission to research, educate, and contribute productively to our community and the world.

All students are expected to submit assignments that represent their own original work, and that have been prepared specifically for the course or section for which they have been submitted. You may not submit work written by others or "recycle" work prepared for other courses without obtaining written permission from the instructor(s).

Other violations of academic integrity include, but are not limited to, cheating, plagiarism, fabrication (e.g., falsifying data), collusion, knowingly assisting others in acts of academic dishonesty, and any act that gains or is intended to gain an unfair academic advantage.

The impact of academic dishonesty is far-reaching and is considered a serious offense against the university. All incidences of academic misconduct will be reported to the Office of Academic Integrity and could result in outcomes such as failure on the assignment, failure in the course, suspension, or even expulsion from the university.

For more information about academic integrity see the student handbook or the Office of Academic Integrity's website, and university policies on Research and Scholarship Misconduct.

Please ask your instructor if you are unsure what constitutes unauthorized assistance on an exam or assignment, or what information requires citation and/or attribution.

Course Content Distribution and Synchronous Session Recordings Policies
USC has policies that prohibit recording and distribution of any synchronous and asynchronous course content outside of the learning environment.

Recording a university class without the express permission of the instructor and announcement to the class, or unless conducted pursuant to an Office of Student Accessibility Services (OSAS) accommodation. Recording can inhibit free discussion in the future, and thus infringe on the academic freedom of other students as well as the instructor. (Living our Unifying Values: The USC Student Handbook, page 13).

Distribution or use of notes, recordings, exams, or other intellectual property, based on university classes or lectures without the express permission of the instructor for purposes other than individual or group study. This includes but is not limited to providing materials for distribution by services publishing course materials. This restriction on unauthorized use also applies to all information, which had been distributed to students or in any way had been displayed for use in relationship to the class, whether obtained in class, via email, on the internet, or via any other media. (Living our Unifying Values: The USC Student Handbook, page 13).

Students and Disability Accommodations:
USC welcomes students with disabilities into all of the University's educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each

course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Support Systems:

Counseling and Mental Health - (213) 740-9355 – 24/7 on call
Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

988 Suicide and Crisis Lifeline - 988 for both calls and text messages – 24/7 on call
The 988 Suicide and Crisis Lifeline (formerly known as the National Suicide Prevention Lifeline) provides free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week, across the United States. The Lifeline is comprised of a national network of over 200 local crisis centers, combining custom local care and resources with national standards and best practices. The new, shorter phone number makes it easier for people to remember and access mental health crisis services (though the previous 1 (800) 273-8255 number will continue to function indefinitely) and represents a continued commitment to those in crisis.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL) – 24/7 on call
Free and confidential therapy services, workshops, and training for situations related to gender- and power-based harm (including sexual assault, intimate partner violence, and stalking).

Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - (213) 740-5086
Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298
Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

The Office of Student Accessibility Services (OSAS) - (213) 740-0776
OSAS ensures equal access for students with disabilities through providing academic accommodations and auxiliary aids in accordance with federal laws and university policy.

USC Campus Support and Intervention - (213) 740-0411
Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity, Equity and Inclusion - (213) 740-2101
Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call
Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-1200 – 24/7 on call
Non-emergency assistance or information.

Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)

A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

Occupational Therapy Faculty Practice - (323) 442-2850 or otfp@med.usc.edu
Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.