COMP 640: Graduate Seminar in Machine Learning
Responsible AI for Society: Foundations, Incentives, and Policy

Instructor: Alireza Fallah

Fall 2025: Wednesdays 2:00–3:15 PM

Office: Duncan Hall, Office 3054 *E-mail*: afallah@rice.edu
Office Hours: By appointment (email to schedule) *Webpage*: afallah.com/teaching

Course Overview and Objectives

This seminar explores the theoretical foundations and interdisciplinary challenges involved in building responsible AI systems and algorithms. Topics include data privacy, fairness, incentive-aware algorithm design, strategic data sharing, and the legal and policy dimensions of AI, such as intellectual property and regulatory frameworks. By bridging ideas from machine learning and economics, the course aims to equip students with a critical understanding of the technical and societal implications of AI systems.

Course Format

During the first two weeks, the instructor will give lectures covering preliminary results on various topics, including privacy, game theory, and mechanism design, that will be needed for later lectures. Beginning in the third week, each session will focus on one or two papers (as listed below and on the course webpage).

Grading Policy

This course has no exam. It is a one-credit-hour class graded on a Satisfactory/Unsatisfactory basis, with the option of taking it as a three-credit-hour course for students interested in completing a project.

All students taking the course for credit must meet two requirements: First, each student taking the course for credit will sign up (via a link provided in the first session) to present the key ideas from one paper during a session in the term. The presentation should last approximately 40 minutes (or 25 minutes if two papers are being covered in that session) and should address the general problem

posed by the paper, its main technical contributions, and its primary results. Second, all students are expected to review the assigned materials in advance to ensure a productive seminar.

Students taking the course for three credits are additionally required to complete a project. The project may be either a short research project or a high-quality summary of three or four relevant papers. Students opting for the project must contact the instructor by the end of September to schedule a meeting in which the project will be defined, goals will be set, and milestones established.

Prerequisites

This is a theory-focused course. Students should have a solid grasp of mathematical proofs and be familiar with basic algorithms and probability.

Absence Policies

This seminar will be most productive when all students actively participate in the discussion. If a student enrolled for credit cannot attend a session, they must notify the instructor in advance by email.

Course Materials

Here is a tentative course schedule, including the topics and papers to be covered. This list is subject to change with advance notice.

Date	Topic	Material
08/27/2025	Introduction to Privacy	
09/03/2025	Introduction to Game Theory and Mechanism Design	
09/10/2025	Data Markets & Privacy (I): Data Acquisition Mechanism	Selling Privacy at Auction Optimal and Differentially Private Data Acquisition: Central and Local Mechanisms
09/17/2025	Data Markets & Privacy (II): The Role of Data Externality	Too Much Data: Prices and Inefficiencies in Data Markets
09/24/2025	Data Markets & Privacy (III): Data Monetization & Regulations	On Three-Layer Data Markets Data, Privacy Laws and Firm Production: Evidence From the GDPR
10/01/2025	Data Valuation	Towards Efficient Data Valuation Based on the Shapley Value Data Shapley in One Training Run

Date	Topic	Material
10/08/2025	Fairness: An Overview of Definitions	Equality of Opportunity in Supervised Learning Inherent Trade-Offs in the Fair Determination of Risk Scores
10/15/2025	Fairness-Accuracy Trade-offs	Algorithm Design: A Fairness-Accuracy Frontier The Statistical Fairness-Accuracy Frontier
10/22/2025	Fair Resource Allocation (I)	Best of Both Worlds: Ex Ante and Ex Post Fairness in Resource Allocation
10/29/2025	Fair Resource Allocation (II)	Fair Allocation in Dynamic Mechanism Design
11/05/2025	Strategic Manipulation in ML	Strategic Classification
11/12/2025	Human vs. AI in Decision Making	Predict Responsibly: Improving Fairness & Accuracy by Learning to Defer The Value of Context: Human versus Black Box Evaluators
11/19/2025	Copyright for Generative AI	On Provable Copyright Protection for Generative Models
11/26/2025	**No class (Thanksgiving holiday)**	
12/03/2025	Learning Preferences From Data	Recovering Preferences From Finite Data A General Framework for Estimating Preferences Using Response Time Data

Rice Honor Code

In this course, all students will be held to the standards of the Rice Honor Code, a code that you pledged to honor when you matriculated at this institution. If you are unfamiliar with the details of this code and how it is administered, you should consult the Honor System Handbook at `http://honor.rice.edu/honor-system-handbook/`. This handbook outlines the University's expectations for the integrity of your academic work, the procedures for resolving alleged violations of those expectations, and the rights and responsibilities of students and faculty members throughout the process.

Disability Resource Center

If you have a documented disability or other condition that may affect academic performance you should: 1) make sure this documentation is on file with the Disability Resource Center (Allen

Center, Room 111 / adarice@rice.edu / x5841) to determine the accommodations you need; and 2) talk with me to discuss your accommodation needs.

Mental Health Statement

The wellbeing and mental health of students is important; if you are having trouble completing your coursework, please reach out to the Wellbeing and Counseling Center. Rice University provides cost-free mental health services through the Wellbeing and Counseling Center to help you manage personal challenges that threaten your personal or academic well-being. If you believe you are experiencing unusual amounts of stress, sadness, or anxiety, the Student Wellbeing Office or the Rice Counseling Center may be able to assist you. The Wellbeing and Counseling Center is located in the Gibbs Wellness Center and can be reached at 713-348-3311 (available 24/7).

Title IX Responsible Employee Notification

At Rice University, unlawful discrimination in any form, including sexual misconduct, is prohibited under Rice Policy on Harassment and Sexual Harassment (Policy 830) and the Student Code of Conduct.

Please be aware that all employees of Rice University are "mandatory reporters," which means that if you tell me about a situation involving discrimination, sexual harassment, sexual assault, dating violence, domestic violence, or stalking, I must share that information with someone, including the University's Title IX Coordinator (titleix@rice.edu). Although I have to make that notification, you will control how your case will be handled, including whether or not you wish to pursue a formal complaint. Our goal is to make sure you are aware of the range of options available to you and have access to the resources you need.

To explore supportive measures and other resources that are available to you, please visit the Office if Interpersonal Misconduct Prevention and Support at safe.rice.edu.