

ECE 285. Special Topics: Ethics and Economics of AI

Fall 2025

Lectures: Tuesdays and Thursdays, 3:30–4:50pm.

Location: CENTR 212.

Instructor: Parinaz Naghizadeh

Email: parinaz@ucsd.edu. (Please use my ucsd email to contact me; I do not check the Canvas inbox.)

Office hours: Tuesdays 11am-noon and Thursdays noon-1pm.

Location: Jacobs Hall 6403.

TA: Yiqiao Liao

Email: yiqiaol@ucsd.edu

Office hours: Friday 1-3pm

Location: Jacobs Hall 5101E

TA: Xiaojun Shan

Email: xishan@ucsd.edu

1 Course Overview

Course topics and goals: This class will explore the impacts of the recent advances in AI, particularly their use in automated decision making systems, from two complementary perspectives: *ethical* and *economic*. From the ethics of AI angle, we will focus on three primary topics: fairness, privacy, and explainability. We will explore the most common definitions and relevant algorithmic solutions in each area, as well as recent research advances in each direction. We will then look at these same ethical questions, and also at the general use of AI, from an economic theory lens. This will include microeconomic questions such as gaming of algorithmic systems by human users, and macroeconomic questions such as the impacts of AI use on labor markets. By the end of this course, students will be able to identify, and critically reflect on, some of the ethical and economic considerations that arise in the design of an algorithmic decision system, and they will be familiar some tools and algorithmic techniques for modeling, measuring, and addressing these ethical and economic concerns in the design of AI systems.

Class format: This class is an advanced topics, graduate seminar, so while the material will be delivered through lectures, active participation and independent work are an important element of this course. Participants will be expected to independently complete readings and reflections on seminal and new research papers as part of the coursework, and participate in pre-class and in-class reflections.

Assumed prerequisites: The official prerequisite is “graduate standing” in ECE. This entails familiarity with what is covered in an ECE undergraduate and early graduate curriculum. For instance, in terms of mathematics background, exposure to probability and statistics, linear algebra, and (convex) optimization is assumed. Some basic prior exposure to Machine Learning is also assumed; for instance familiarity with

the concepts of classification, regression, training and testing datasets, deep neural networks, hyperparameter tuning, etc. In terms of coding/programming skills, you may need basic familiarity with a programming language like Python (as there may be one or two coding exercises). As you may note from the topic list below, neither of the math, ML, or programming backgrounds are a core focus of the course; they will be needed to understand the course material and readings/papers.

2 Textbooks and Readings

Readings and reference books: Lecture slides will be posted on Canvas. These are intended to provide an overview of all topics covered in each lecture, but will not be a substitute to lectures, as they may not include all details such as some of the discussions or examples from class. In addition to the selected set of research papers and articles, you may find the following references useful (readings will be assigned from some of the following references):

- [Fairness textbook] “Fairness and Machine Learning: Limitations and Opportunities”, by Solon Barocas, Moritz Hardt, Arvind Narayanan. (*available online at <https://fairmlbook.org>*)
- [Differential privacy textbook] “The Algorithmic Foundations of Differential Privacy” by Cynthia Dwork and Aaron Roth. (*available online at <https://www.cis.upenn.edu/~aaroht/Papers/privacybook.pdf>*)
- [Explainability textbook] “Interpretable Machine Learning: A Guide for Making Black Box Models Explainable”, by Christoph Molnar. (*available online at <https://christophm.github.io/interpretable-ml-book/>*)
- [Game theory textbook] “A Course in Game Theory”, by Martin J. Osborne and Ariel Rubinstein. *available online at <https://books.osborne.economics.utoronto.ca>*

All of these main references are available online on the publisher or authors’ webpages. They are also available through the library.

If you are interested in reading more, some (optional) additional references include:

- “The Ethical Algorithm: The Science of Socially Aware Algorithm Design”, by Aaron Roth and Michael Kearns.
- “Justice: What’s the Right Thing to Do?”, by Michael Sandel.
- “The Fundamentals of Ethics”, by Russ Shafer-Landau.
- “Algorithms and Autonomy: The Ethics of Automated Decision Systems”, by Adam Pham, Alan Rubel, and Clinton Castro.

Copyright disclaimer: The materials used in connection with this course may be subject to copyright protection and are only for the use of students officially enrolled in the course for the educational purposes associated with the course. Copyright law must be considered before copying, retaining, or disseminating materials outside of the course.

3 Grading

Grading: 15% pre-class work, 15% in-class quizzes, 30% midterm exam, 40% final exam. Letter grades will be determined based on the standard scheme ($A+ \geq 97\%$, $A \geq 93\%$, $A- \geq 90\%$, $B+ \geq 87\%$, ...).

- **Pre-class work:** You will be assigned readings prior to most lectures. These include seminal or recent research papers in the field, news articles, and chapters from our course textbooks. You will be asked to fill out a short report (through a Canvas quiz) based on these readings prior to class. For analytical lessons, these pre-class homework may also include short problems, which you will also solve and answer on the Canvas quiz. There may also be one or two short programming assignments, which you will also turn in through the Canvas quiz.
- **In-class quiz:** During 6 randomly selected lectures, you will be asked to turn in a short critical reflection and/or some analytical work on the papers/topics discussed that day (these may also be related to the pre-class readings, if any). Prompts for the critical reflection will be given during class. The time allotment will be 10 minutes. You must have pen and paper with you to complete this. Each assessment is worth 3% of your grade. This means that you can miss one of these assessments and still gain full credit for this component.
- **Midterm exam:** There will be an in-class midterm exam on Tuesday, 10/28. The midterm exam will be based on lectures 1-9. This will be a proctored exam.
- **Final exam:** There will be a final exam on Monday, December 8th, 2025, 3:00-5:59pm, at CENTR 212. The final exam is comprehensive. This will be a proctored exam.
- **Late work policy:** Late submissions or class absences will be accepted only if there is a documented reason or unexpected circumstances (verifiable by the university). No exceptions will be made for late enrollments. *No exceptions to this policy will be made; late work beyond the stated reasons will not be accepted and will get a grade of zero.*

The above are subject to change. Canvas will be used for announcements and to distribute course material. Please refer to the course announcements on Canvas to stay notified of any updates to this syllabus, including updates to homework or exam(s) assignment and/or due dates, as well as clarifications/corrections to assignment questions.

Collaboration policy: Students are welcome (and encouraged) to work together on the paper readings and on general discussion of course material. Keep in mind however that all solutions handed in by the students must be solely by them and reflect their independent understanding of the material. No collaboration, discussions, or use of any online resources, is permitted for the in-class assignments, midterm, and/or final exams.

On the use of GenAI tools: Please see Section 5 for the full description of our course policy in this regard. In short: you may use GenAI tools *only for help in the pre-class assignments (but not on in-class assessments or exams)* in the following ways: to brainstorm ideas for an assignment, help understanding some parts of the readings, and to help edit the clarity or grammar of your writing. As a way to demonstrate your honest use of these tools and your learning process, you must: keep histories of your chats and submit them when requested, and cite the content that came from the GenAI tools in your submission. We advise you to read the detailed policy in Section 5 if you are planning on using such tools.

4 Course Outline

Table 1: Lecture outline (subject to change)

Lecture	Date	Topics covered
1	9/25	Introduction: overview and course logistics
2	9/30	Fairness (I): Sources of unfairness; statistical fairness criteria
3	10/2	Fairness (II): Statistical fairness criteria, continued; impossibility results
4	10/7	Fairness (III): Satisfying fairness criteria: pre-processing, post-processing
5	10/9	Fairness (IV): Satisfying fairness criteria: in-processing
6	10/14	Fairness (V): Wrap up, and an overview of some other considerations
7	10/16	Privacy (I): Examples, approaches, and introduction to differential privacy
8	10/21	Privacy (II): Global and local differential privacy
9	10/23	Privacy (III): DP in Deep Learning; FL+DP; Fairness-privacy tradeoffs
10	10/28	Midterm exam (in-class)
11	10/30	Explainability (I): Overview of existing approaches; feature-based explanations
12	11/4	Explainability (II): Post-hoc explanations, continued: Shapley values and SHAP
13	11/6	Explainability (III): Other methods; evaluating explanations, concluding thoughts
	11/11	<i>Veterans Day Holiday (no class)</i>
14	11/13	Microeconomics (I): Examples of gaming of algorithmic systems; intro to game theory
15	11/18	Microeconomics (II): Strategic machine learning; fairness implications
16	11/20	Microeconomics (III): Implications for fairness and explainability; concluding thoughts
17	11/25	Macroeconomics: Labor markets; labor management; and advertising markets
	11/27	<i>Thanksgiving Holiday (no class)</i>
18	12/2	Ethics and economics of GenAI
19	12/4	Other ethical and economic considerations; conclusion
	12/8	Final exam (Monday, December 8th, 2025, 3:00-5:59pm)

5 Other Course Policies

Academic Integrity: Academic Integrity is expected of everyone at UC San Diego. This means that you must be honest, fair, responsible, respectful, and trustworthy in all of your actions. Lying, cheating or any other forms of dishonesty will not be tolerated because they undermine learning and the University’s ability to certify students’ knowledge and abilities. Thus, any attempt to get, or help another get, a grade by dishonesty will be reported to the Academic Integrity Office and will result in sanctions. Sanctions can include an F in this class and suspension or dismissal from the University. So, think carefully before you act by asking yourself:

a) is what I’m about to do or submit for credit an honest, fair, respectful, responsible and trustworthy representation of my knowledge and abilities at this time and,

b) would my instructor approve of my action?

You are ultimately the only person responsible for your behavior. So, if you are unsure, don’t ask a friend—ask your instructor, instructional assistant, or the Academic Integrity Office. You can learn more about academic integrity at <http://academicintegrity.ucsd.edu>.

Use of GenAI tools: Generative artificial intelligence (GenAI) tools, like ChatGPT, may be used in this course in limited ways with proper documentation, citation, and acknowledgement. In accordance with the course learning outcomes, you may use GenAI tools *only on the Homework assignment (but not quizzes or exams)* in the following ways: to brainstorm ideas for an assignment, and to help edit your

writing. As a way to demonstrate your honest use of these tools and your learning process, you must: keep histories of your chats and submit them when requested, and cite the content that came from the GenAI tools in your submission. In deciding whether your use aligns with the learning goals of this course, reflect on the process of using the GenAI tool by answering these questions: Was your use with the aim to only provide a starting point for independent thinking? Are the final solutions your own? Was your use with the aim of deepening understanding of the subject matter? If your answer to any of these questions is no, then the use of GenAI would not align with the learning objectives of this course.

NOTE: GenAI is known to fabricate sources, facts, and give false information. It also perpetuates bias. You should also be aware that there are copyright and privacy concerns with these tools. You should exercise caution when using large portions of content from AI sources for these reasons. Also, you are accountable for the content and accuracy of all work you submit in this class, including any supported by generative AI.

You are encouraged to reach out to me or the TA for support, or to The Academic Integrity Office for questions about ethical use of GenAI. You can learn more about academic integrity in the use of GenAI at <https://academicintegrity.ucsd.edu/excel-integrity/gen-ai/ai-in-education.html>.

Your Mental Health: As a student you may experience a range of issues that can cause barriers to your learning, such as increased anxiety, feeling down, difficulty concentrating, and/or lack of motivation. Please feel free to talk to me if there are circumstances affecting your ability to participate in class. The University offers services to assist you with addressing these and other concerns you may be experiencing. If you find yourself feeling isolated, anxious, or overwhelmed, please know that there are resources to help: CAPS Student Health and Well-Being provides services like confidential counseling and consultations for psychiatric services and mental health programming <https://wellness.ucsd.edu/caps/>.

Accessibility Accommodations: I ask that students requesting accommodations for this course due to a disability provide a current Authorization for Accommodation (AFA) letter issued by the Office for Students with Disabilities (OSD) which is located in University Center 202 behind Center Hall. Students are required to present their AFA letters to Faculty (please make arrangements to contact me privately) and to the OSD Liaison in the department in advance so that accommodations may be arranged.

Office for Students with Disabilities (OSD)

Documents student disabilities, provides accessibility resources, and reasonable accommodations

858.534.4382 — osd@ucsd.edu — <https://disabilities.ucsd.edu/>

UC San Diego Principles of Community: The University of California, San Diego is dedicated to learning, teaching, and serving society through education, research, and public service. Our international reputation for excellence is due in large part to the cooperative and entrepreneurial nature of the UC San Diego community. UC San Diego faculty, staff, and students are encouraged to be creative and are rewarded for individual as well as collaborative achievements.

To foster the best possible working and learning environment, UC San Diego strives to maintain a climate of fairness, cooperation, and professionalism. These principles of community are vital to the success of the University and the well being of its constituents. UC San Diego faculty, staff, and students are expected to practice these basic principles as individuals and in groups. The complete UC San Diego Principles of Community in English and Spanish can be found here: <https://ucsd.edu/about/principles.html>.

Nondiscrimination and Harassment: The University of California, in accordance with applicable federal and state laws and university policies, does not discriminate on the basis of race, color, national origin, religion, sex, gender, gender identity, gender expression, pregnancy (including pregnancy, childbirth, and medical conditions related to pregnancy or childbirth), physical or mental disability, medical

condition, genetic information, ancestry, marital status, age, sexual orientation, citizenship, or service in the uniformed services (including membership, application for membership, performance of service, application for service, or obligation for service in the uniformed services). The university also prohibits harassment based on these protected categories, including sexual harassment, as well as sexual assault, domestic violence, dating violence, and stalking. The nondiscrimination policy covers admission, access, and treatment in university programs and activities. If students have questions about student-related nondiscrimination policies or concerns about possible discrimination or harassment, they should contact the Office for the Prevention of Harassment & Discrimination (OPHD) at (858) 534-8298, ophd@ucsd.edu, or <http://reportbias.ucsd.edu>. Campus policies provide for a prompt and effective response to student complaints. This response may include alternative resolution procedures or formal investigation. Students will be informed about complaint resolution options. A student who chooses not to report may still contact CARE at the Sexual Assault Resource Center for more information, emotional support, individual and group counseling, and/or assistance with obtaining a medical exam. For off-campus support services, a student may contact the Center for Community Solutions. Other confidential resources on campus include Counseling and Psychological Services, Office of the Ombuds, and Student Health Services.

CARE at the Sexual Assault Resource Center
858.534.5793 — sarc@ucsd.edu — <https://care.ucsd.edu>

Counseling and Psychological Services (CAPS)
858.534.3755 — <https://caps.ucsd.edu>