Philosophy 290 / CS 294: The Philosophy and History of Automated Decision Making

Spring 2024

Ben Recht
brecht@berkeley.edu
Office hours: Thursdays 4-5 PM

Shamik Dasgupta
shamikd@berkeley.edu
Office Hours: Fridays 3-4:30pm

Seminar overview

This course will survey automated decision making, focusing on its historical development during the Cold War and its philosophical foundations. We will consider the parallel development of modern optimization, game theory, randomized clinical trials, and machine learning between 1944 and 1970 and how seminal data science was forged on the first computers. This history will let us ask what one needs to assume about the future in order to automate decisions based on past experience. Along the way, we will examine various philosophical issues that arise in deploying these decision-making systems, including idealization in scientific models and rational choice theory, the role of chance and probability in guiding decisions, and the nature of pattern recognition. Through this philosophical lens, we'll probe why these particular decision-making frameworks have become so entrenched over the past 50 years.

Readings

Weekly reading selections will be made available as PDF documents through the bCourses site. Several of the readings are works in progress by Ben. Each week, please read the selections and post a question or comment on the material in bCourses by Wednesday evening before the seminar.

We will also maintain a living bibliography of associated readings. There will be a dropbox folder with readings. If you find something worth sharing or reading, please add it to the dropbox and list it in the google doc.

Coursework

A substantial 5000-7000 word research paper OR two 1500-2500 word articles aimed at a popular audience with no academic jargon. All papers due by Friday May 10th (last day of the semester).

Syllabus

No one really knows what this class is going to be like. We have a fairly unique set of backgrounds for a graduate seminar and the topic doesn't live neatly in any one academic category. We're writing this vision down in this document as a rough guide, but it will inevitably change as we progress.

January 18	Overview of the course: frameworks for decisions, computation and probability.
January 25	Discovery and decision making before data science. ● Recht, "Steampunk Data Science"
February 1	Optimization, idealization, and policy. ● Recht, "Controlling The Stable"
February 8	Idealization I: Models and Values ● Selections from Winsberg and Harvard, *Scientific Models and Decision-Making*
February 15	Computational Game Theory ● Recht, "The History of Computational (Parlor) Game Theory"
February 22	Idealization II: Rational choice theory ● Selections from Greco, *Idealization in Epistemology*
February 29	Potential outcomes and counterfactuals ● TBD
March 7	History of Randomized Clinical Trials ● Recht, "Drugs, Vaccines, and Screens - regulation by random trials"
March 14	Chance and decision ● TBD
March 21	Chance and decision ● TBD

April 4	Pattern recognition, artificial intelligence, and data ● Recht, “The resilience of pattern recognition in artificial intelligence”
April 11	Patterns and induction ● Goodman, “The New Riddle of Induction”
April 18	Patterns and language ● Selections from Kripke, “Wittgenstein on Rules and Private Language”
April 25	Patterns and ethical alignment ● TBD