Sample Syllabi, cont.

2. Ethics and Technology—Proposal for Introductory Course

Course Description: Society and politics have become inextricably linked with information technologies. How our lives go and how we make political decisions are heavily influenced by how people and corporations use computers, smartphones, data, and artificial intelligence. This class explores the ethical questions that arise in how we use these technologies and how we decide as a community how they should be used. We begin with an introduction to moral philosophy in the context of technological development, and then take up key topics emerging in this field: fairness and discrimination in data use, transparency, privacy, artificial intelligence, attention and choice, and free speech.

Part I: Methods and Problems

We start off with a survey of the landscape of ethical problems in technology and then we develop the basic tools of moral philosophy. The goal is to orient ourselves with the distinctively ethical or philosophical problems posed by recent technology developments. And we want to clarify the distinction between consequentialist and deontological approaches to these issues.

Readings:
O'Neil, *Weapons of Math Destruction*, Intro, chapters 1–4, 8–10
Bentham, *Principles of Moral and Legislation*, IV, XIII
Parfit, *Reasons and Persons*, sections 130–131
Thomson, "The Trolley Problem"
Rawls, *A Theory of Justice*, section 5

Part II: Discrimination and Fairness

It is wrong for human beings to discriminate against one another, or to treat one another unfairly. But what, exactly, makes discrimination wrongful? And how might discriminatory behavior migrate from human practices into technological decision-making?

Readings:
Kearns and Roth, *The Ethical Algorithm*, chapter 2
Buolamwini, "Algorithms Aren't Racist, Your Skin is just too Dark"
Hassein, "Against Black Inclusion in Facial Recognition"
Propublica, "Machine Bias"
Kleinberg, Mullainathan, Raghavan, "Inherent Trade-Offs in the Fair Determination of Risk Scores"
Dressel and Faird, "The Accuracy, Fairness, and Limits of Predicting Recidivism"
Scanlon, *Why Does Inequality Matter?*, chapter 4

Hellman, *When is Discrimination Wrong?*, Intro. and chapter 1
Barocas and Selbst, “Big Data’s Disparate Impact”

Part III: Data, Individuality, and Explanations

Private corporations and public institutions like schools collect data about us—our identities and habits—and we need to know how they should and should not use this data. When, if ever, is it acceptable for someone to decide how to treat you based on your data rather than more direct facts about your personality? If someone uses data to determine how to treat you in a certain way, are you entitled to a transparent explanation as to how this decision was reached?

Readings: Schauer, *Profiles, Probabilities, and Stereotypes*, Intro
Piper, “The UK used a formula to predict students’ scores for canceled exams”
Eidelson, “Treating People as Individuals”
Barocas and Selbst, “The Intuitive Appeal of Explainable Machines”
Vredenburgh, “The Right to Explanation”

Part IV: Privacy

Privacy is a controversial issue in moral philosophy, and it has become even more urgent to clarify our thinking about it with the advent of information technology. Much of our lives are now lived online, and we need to know when it is wrong for someone, or a corporation, to learn things about us and share this information. This problem is especially tricky because nowadays we can learn things about one another by using massive data sets to make inferences about people to whom the set applies.

Readings: Véliz, *Privacy is Power*, chapter 1
Kosinski, Stillwell, and Graepel, “Private Traits and Attributes Are Predictable from Digital Records of Human Behavior”
Kearns and Roth, *The Ethical Algorithm*, chapter 1
Nissenbaum, “A Contextual Approach to Privacy Online”
Thomson, “The Right to Privacy”
Marmor, “What is the Right to Privacy?” and “Privacy in Social Media”

Part V: Artificial Intelligence

AI is one of the hottest topics in recent technological advances, and it poses a dizzying array of threats to human wellbeing. We want to understand what these distinctive threats are and diagnose their various ethical dimensions. We will particularly focus on the problem of aligning AI with human values, which has both practical and theoretical dimensions.

Readings: Russell, *Human Compatible*, chapters 1 and 5
Hendrycks, *Introduction to AI Safety, Ethics, and Society*, chapters 1–2
Gabriel, “Artificial Intelligence, Values, and Alignment”
Petersen, “Machines Learning Values”
Winter, Hollman, and Manheim, “Value Alignment for Advanced Artificial Intelligence”
Bostrom, “Existential Risk Prevention as Global Priority”
Thorstad, “High Risk, Low Reward”

Part V: Choice and Free Speech

Technology heavily shapes our social world. It grabs our attention—or, perhaps more dangerously, it shifts our attention in directions without us noticing. When is the manipulation of someone’s attention wrongful, and why? Technology also constitutes a huge part of our communicative infrastructure, an essential part of democratic public discourse. What norms should we use to govern these shared communicative spaces?

Readings: Castro and Pham, “Is the Attention Economy Noxious?”
Harris, “How Technology Hijacks People’s Mind—from a Magician and Google’s Design Ethicist” and “Tech Companies Design Your Life, Here’s Why You Should Care”
Frankfurt, “Freedom of the Will and the Concept of a Person”
Scanlon, *What We Owe to Each Other*, chapter 6, section 2
Sunstin and Thaler, *Nudge*
Cohen and Fung, “Democracy and the Digital Public Sphere”
Gillespie, *Custodians of the Internet*, chapters 1–2
Sunstein, *#Republic: Divided Democracy in the Age of Social Media*, chapter 1

Assignments

There will be five assignments: 3 reading responses, each worth 15% of the grade (total 45%) and 2 short papers, each worth 20% of the grade (total 40%). Participation will make up the remaining 15%.

Three 2-page reading responses on **Parts I, III, and V**.

Two 4-page papers on **Parts II and IV**.