INFO 4270: ETHICS AND POLICY IN DATA SCIENCE

Fall 2017

Solon Barocas (Professor)

Brian McInnis (Teaching Assistant)

COURSE DESCRIPTION AND OBJECTIVES

This class will teach you to recognize where and understand why ethical issues and policy questions can arise when applying data science to real world problems. It will bring analytic and technical precision to normative debates about the role that data science, machine learning, and artificial intelligence play in consequential decision-making in commerce, employment, finance, healthcare, education, policing, and other areas. We will focus on ways to conceptualize, measure, and mitigate bias in data-driven decision-making, to audit and evaluate models, and to render these analytic tools more interpretable and their determinations more explainable. You will learn to think critically about how to plan, execute, and evaluate a project with these concerns in mind, and how to cope with novel challenges for which there are often no easy answers or established solutions.

To do so, you will develop fluency in the key technical, ethical, policy, and legal terms and concepts that are relevant to a normative assessment of data science; learn about some of the common approaches and emerging tools for mitigating or managing these ethical concerns; and gain exposure to legal scholarship and policy documents that will help you understand the current regulatory environment and anticipate future developments. Ultimately, the class will teach you how to reason through these problems in a systematic manner and how to justify and defend your approach to dealing with them.

COURSE MATERIALS

All course materials will be available on Blackboard.

We will read critical commentary and thoughtful reflections by seasoned practitioners, important and illustrative research from computer scientists, an interesting mix of legal scholarship, moral philosophy, and policy analysis, and a host of government documents. All along the way, we will

rely on case studies, recent controversies, and current events to ground our discussion.

The appropriate response to many of the problems that we will address in the course is far from settled. This is, consequently, a reading-heavy course. Even so, the assigned readings frequently do not present all sides of the debate. I have therefore selected materials that tend to offer a more critical—and sometimes less familiar—perspective with the goal of provoking productive debate during our class and strong reactions in your assignments. I expect you to stake out conflicting—informed and carefully reasoned—positions on the issues, and you should not shy away from doing so.

The lecture, discussion, and in-class activities will cover most, but not all of the issues raised by the readings. Given the nature of the issues and material under consideration, I expect lively debate and plan to follow the natural flow of discussion as much as possible. As such, I am certain that class will cover some important ideas that do not appear in the readings. Active listening and participation is therefore crucial.

ASSIGNMENTS

- Answer a brief question at the start of class — Ongoing

  When you arrive in class, you will find a question on the chalkboard. You will answer the question on Piazza. While I encourage you to take a moment to carefully consider the question and give a thoughtful answer, your response can be brief. The goal is to jumpstart your thinking. You must, however, submit your response before you leave class, as your submission will be a way to document your attendance.

- Post to the Blackboard discussion board — Ongoing

  As a matter of course, you should post any interesting news items that you happen to read to the Blackboard discussion board—and take a moment to reflect and comment on their significance and relevance to our class. Items directly related to the reading assigned at that time are especially welcome. While this is voluntary, I will make sure that your contributions to the discussion board are reflected in your participation grade.

- Critically assess a proposed data science project — Due September 29

  Drawing on the readings from the first third of the class, you will critically assess a proposed data science project. I will provide you with a brief description of the project and you will identify 3 potential problems with the proposed application that could raise concerns with fairness. Your answer should take the form of 3 bullet points, each comprising 2-4 sentences: 1-2 sentences identifying the source of the problem, and 1-2 sentences explaining how fairness is at stake. The problems that you identify should be as distinct as possible to receive maximum credit.

- Respond to the Consumer Financial Protection Bureau's Request for Information Regarding Use of Alternative Data and Modeling Techniques in the Credit Process — Due November 10

  The Consumer Financial Protection Bureau is currently considering a number of policy questions raised by novel forms of credit scoring that rely on new sources of data and more sophisticated learning methods. The Bureau has solicited input from outside experts and the broader public, with the aim of better understanding how to deal with issues ranging from privacy to non-discrimination and the ability to explain credit decisions. Drawing on the course readings and ideas discussed in class, you will draft a 3-5 page, double-spaced response to the Bureau's request, staking out and advocating in favor of a particular policy position. You should further support your position by explaining the strategies and tools currently available to address the Bureau's concerns.

- Write a final paper that revisits a recent controversy — Due December 12

  In a 8-10 page, double-spaced paper, you will revisit a recent controversy involving different course themes. You will choose from a set of 5 predetermined cases. Your paper should draw extensively from the course materials, lectures, and in-class discussions, and present a comprehensive plan for undertaking the project in a way that addresses fairness, respects privacy (as a legal and broader normative matter), and comports with other pertinent ethical principles. The paper should not shy away from pointing out difficult tensions or unavoidable trade-offs. You should instead explain why it is not possible to "have it all," and then provide a thoughtful justification for the specific trade-off that you suggest.

SUBMITTING ASSIGNMENTS

- All assignments must be submitted through Blackboard. Do not email or physically hand in any assignments. Always confirm that your assignment has uploaded correctly after submission.
- Should you encounter a problem with Blackboard, please email the TA before the deadline with (1) your completed assignment, (2) a screenshot of the problem, and (3) the time of your attempted submission.
- You will incur a 20% penalty if you submit your work within 24 hours after the deadline. You will receive no credit thereafter.
- There are no exceptions to this late submission policy, except university-approved excuses.
- Upon receiving your graded assignment, please take at least 24 hours to consider the feedback you have received as well as the original assignment instructions. If, at that time, you feel that you deserved a better grade, you may submit a formal written request to the TA by email. Your request must explain exactly where you believe there was a

mistake in grading or why you object to a specific piece of feedback. The TA will consider well justified requests and re-grade assignments as appropriate.

GRADING

20% Participation (both in-class and on Blackboard)
15% Critical review of proposed data science project
25% Response to Consumer Financial Protection Bureau's Request for Information
40% Final paper revisiting a recent controversy

ACADEMIC INTEGRITY

I expect you to abide by Cornell's Code of Academic Integrity at all times. Please note that the Code specifically states that a "Cornell student's submission of work for academic credit indicates that the work is the student's own. All outside assistance should be acknowledged, and the student's academic position truthfully reported at all times."

Please contact me or the TA if you have any questions or concerns about appropriately acknowledging others' work in your submitted assignments. You should expect that I will rigorously enforce the Code and may use software to check for plagiarism.

SCHEDULE AND READINGS

I expect you to complete all assigned readings prior to class. Unless I've noted particular parts, sections, or pages for you to read, you should read the assigned text in its entirety. For some classes, I have listed *recommended* readings that you may choose to complete, if you are so inclined. These are optional, and I will not expect that you have read them.

The schedule and readings are subject to change as we progress through the semester. Please always refer to the syllabus posted to Blackboard before you begin reading for the next class.

Background Reading [Optional]

- Boyd and Crawford, "Critical Questions for Big Data"
- Zarsky, "The Trouble with Algorithmic Decisions"
- O'Neil, *Weapons of Math Destruction*
- Pasquale, *The Black Box Society*
- The White House Office of Science and Technology Policy, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*

August 23 — Welcome
August 28 — *Data*, the givens

- Gitelman and Jackson, *Raw Data is an Oxymoron* [Introduction]

● Agre, “Surveillance and Capture: Two Models of Privacy”

Recommended

● Bowker and Star, *Sorting Things Out*
● Auerbach, “The Stupidity of Computers”

August 30 — What problem are we solving?

● Moor, “What is Computer Ethics?”
● Hand, “Deconstructing Statistical Questions”

September 4 — Labor Day — No class

September 6 — Cultivating a critical disposition

● O’Neil, *On Being a Data Skeptic*
● Domingos, “A Few Useful Things to Know About Machine Learning”

Recommended

● Luca, Kleinberg, and Mullainathan, “Algorithms Need Managers, Too”

September 11 — Bias and exclusion

● Friedman and Nissenbaum, “Bias in Computer Systems”
● Lerman, “Big Data and Its Exclusions”

Recommended

● Hand, “Classifier Technology and the Illusion of Progress” [Sections 3 and 4]

September 13 — The social science of discrimination

● Pager and Shepherd, “The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets”
● Goodman, “Economic Models of (Algorithmic) Discrimination”

September 18 — How machines learn to discriminate

● Hardt, “How Big Data Is Unfair”
● Barocas and Selbst, “Big Data’s Disparate Impact” [Parts I and II]

Recommended

● Gandy, “It’s Discrimination, Stupid”
● Dwork and Mulligan, “It’s Not Privacy, and It's Not Fair”

September 20 — Auditing algorithms

- Sandvig, Hamilton, Karahalios, and Langbort, “Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms”
- Diakopoulos, “Algorithmic Accountability: Journalistic Investigation of Computational Power Structures”

Recommended

- Lavergne and Mullainathan, “Are Emily and Greg more Employable than Lakisha and Jamal?”

September 25 — Algorithms audited

- Sweeney, “Discrimination in Online Ad Delivery”
- Datta, Tschantz, and Datta, “Automated Experiments on Ad Privacy Settings”

September 27 — Formalizing and enforcing fairness

- Dwork, Hardt, Pitassi, Reingold, and Zemel, “Fairness Through Awareness” ● Feldman, Friedler, Moeller, Scheidegger, and Venkatasubramanian, “Certifying and Removing Disparate Impact”

Recommended

- Žliobaitė and Custers, “Using Sensitive Personal Data May Be Necessary for Avoiding Discrimination in Data-Driven Decision Models”

October 2 — Accounting for disparities in accuracy and error rates [Manish Raghavan, a doctoral student in computer science at Cornell and co-author of one of the assigned readings, will join us for this class]

- Angwin, Larson, Mattu, and Kirchner, “Machine Bias”
- Kleinberg, Mullainathan, and Raghavan, “Inherent Trade-Offs in the Fair Determination of Risk Scores”

Recommended

- Northpointe, *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*
- Chouldechova, “Fair Prediction with Disparate Impact”
- Berk, Heidari, Jabbari, Kearns, and Roth, “Fairness in Criminal Justice Risk Assessments: The State of the Art”

October 4 — Competing notions of fairness

● Hardt, Price, and Srebro, “Equality of Opportunity in Supervised Learning” ○ Wattenberg, Viégas, and Hardt, “Attacking Discrimination with Smarter Machine Learning”
● Friedler, Scheidegger, and Venkatasubramanian, “On the (Im)possibility of Fairness”

Recommended
● Tene and Polonetsky, “Taming the Golem: Challenges of Ethical Algorithmic Decision Making”

October 9 — Fall break — No class

October 11 — Feedback loops and fairness

● Lum and Isaac, “To Predict and Serve?”
● Joseph, Kearns, Morgenstern, and Roth, “Fairness in Learning: Classic and Contextual Bandits”

October 16 — The fairness of different factors

● Barocas, “Data Mining and the Discourse on Discrimination”
● Grgić-Hlača, Zafar, Gummadi, and Weller, “The Case for Process Fairness in Learning: Feature Selection for Fair Decision Making”

October 18 — Profiling and particularity

● Vedder, “KDD: The Challenge to Individualism”
● Lippert-Rasmussen, “‘We Are All Different’: Statistical Discrimination and the Right to Be Treated as an Individual”

Recommended
● Schauer, *Profiles, Probabilities, And Stereotypes*

October 23 — From allocative to representational harms

● Caliskan, Bryson, and Narayanan, “Semantics Derived Automatically from Language Corpora Contain Human-like Biases”
● Zhao, Wang, Yatskar, Ordonez, and Chang, “Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints”

Recommended
● Bolukbasi, Chang, Zou, Saligrama, and Kalai, “Man Is to Computer Programmer as Woman Is to Homemaker?”

October 25 — Transparency and due process

● Citron and Pasquale, “The Scored Society: Due Process for Automated Predictions”
● Ananny and Crawford, “Seeing without Knowing”

Recommended
● de Vries, “Privacy, Due Process and the Computational Turn”
● Zarsky, “Transparent Predictions”
● Crawford and Schultz, “Big Data and Due Process”
● Kroll, Huey, Barocas, Felten, Reidenberg, Robinson, and Yu, “Accountable Algorithms”

October 30 — Interpretability in machine learning

● Bornstein, “Is Artificial Intelligence Permanently Inscrutable?”
● Burrell, “How the Machine 'Thinks'”
● Lipton, “The Mythos of Model Interpretability”

Recommended
● Doshi-Velez and Kim, “Towards a Rigorous Science of Interpretable Machine Learning” ● Hall, Phan, and Ambati, “Ideas on Interpreting Machine Learning”

November 1 — The value of explanation

● Grimmelmann and Westreich, “Incomprehensible Discrimination”
● Selbst and Barocas, “Regulating Inscrutable Systems”

Recommended
● Jones, “The Right to a Human in the Loop”
● Edwards and Veale, “Slave to the Algorithm? Why a ‘Right to Explanation’ is Probably Not the Remedy You are Looking for”

November 6 — The future of scoring

● Robinson and Yu, *Knowing the Score*
● Hurley and Adebayo, “Credit Scoring in the Era of Big Data”

November 8 — The privacy implications of inference
● Duhigg, “How Companies Learn Your Secrets”
● Kosinski, Stillwell, and Graepel, “Private Traits and Attributes Are Predictable from Digital Records of Human Behavior”

Recommended

● Barocas and Nissenbaum, “Big Data's End Run around Procedural Privacy Protections”
● Chen, Fraiberger, Moakler, and Provost, “Enhancing Transparency and Control when Drawing Data-Driven Inferences about Individuals”

November 13 — Price discrimination

● Valentino-Devries, Singer-Vine, and Soltani, “Websites Vary Prices, Deals Based on Users' Information”
● The Council of Economic Advisers, *Big Data and Differential Pricing*

Recommended
● Hannak, Soeller, Lazer, Mislove, and Wilson, “Measuring Price Discrimination and Steering on E-commerce Web Sites”
● Kochelek, “Data Mining and Antitrust”

November 15 — Insurance

● Helveston, “Consumer Protection in the Age of Big Data”
● Kolata, “New Gene Tests Pose a Threat to Insurers”

Recommended
● Swedloff, “Risk Classification's Big Data (R)evolution”
● Cooper, “Separation, Pooling, and Big Data”
● Simon, “The Ideological Effects of Actuarial Practices”

November 20 — Algorithmic persuasion and manipulation

● Tufekci, “Engineering the Public”
● Calo, “Digital Market Manipulation”

Recommended
● Kaptein and Eckles, “Selecting Effective Means to Any End”

November 22 — Thanksgiving — No class

November 27 — Algorithmic publics

● Pariser, “Beware Online ‘Filter Bubbles’”
● Gillespie, “The Relevance of Algorithms”

November 29 — Rejecting certain applications of machine learning

● Buolamwini, “Algorithms Aren’t Racist. Your Skin Is just too Dark” ● Hassein, “Against Black Inclusion in Facial Recognition” ● Agüera y

Arcas, Mitchell, and Todorov, “Physiognomy’s New Clothes”

Recommended

- Garvie, Bedoya, and Frankle, *The Perpetual Line-Up*
- Wu and Zhang, “Automated Inference on Criminality using Face Images”
- Haggerty, “Methodology as a Knife Fight”