# Ethics & Policy of Data Analytics

(94-836)

David Danks ddanks@cmu.edu

Sina Fazelpour sinaf@andrew.cmu.edu

**Class meetings:** Tuesdays & Thursdays, 3:10 pm – 4:30 pm (Pittsburgh time)

**Office hours:** Just email either (or both) of us for Zoom or in-person meetings

**Course materials**: Everything will be on the course Canvas site

## Course description

We live in an increasingly data-intensive and algorithmic society, and so we must consider the ethical, societal, and personal impacts of these technologies. This course will explore ethical and policy aspects of data analytics, including issues of privacy, bias, trust, and more. We will consider both ethical questions about what we *ought* to do with data analytics, and also policy questions about what we are *permitted* or *required* to do by law and regulation (in both US and non-US contexts). This course will thus provide an important complement to more statistics/technology-centric courses that emphasize what we *can* do.

## Learning objectives

At the end of this course, students will be able to:

- Understand the key concepts of privacy, fairness, bias, explainability, and trust
- Determine the ethical impacts (along these dimensions) of various standard data analysis practices, methods, and products
- Derive relevant, key policy and legal constraints on data analytic practices and products
- Apply both ethical and policy considerations to an analysis of the permissibility and/or legitimacy of different data analytics

## Course requirements

**Assignments**: Grades in this course will be based on two different types of assignments:

- *Daily writing assignments (10 × 4%)*: For every class (except the first one), there will be a short writing assignment. Your ten best scores (of the eleven assignments) will be counted for this portion of your grade.
    - Due: 9:00 am (Pittsburgh time) on the day of class
- *Integrative analysis (3 × 20%)*: For each of the three main concepts—privacy, bias/fairness, and explainability—there will be an integrative assignment in which you analyze a data analytic practice in terms of that concept, from both ethical and policy perspectives.
    - Due: 5:00 pm (Pittsburgh time) on February 24; March 10; and March 17.

**Extensions/late submission**: This course will move quite quickly, so it is important that students do not fall behind. For that reason, we have a relatively strict late submission policy: 50% penalty for assignments submitted 0-6 hours after the deadline; 100% penalty (i.e., zero for the assignment) for anything submitted 6+ hours after the deadline. Please note that the deadline times are all **Pittsburgh time**. Also, we will not grant extensions except in truly exceptional circumstances. (If you find yourself in such a circumstance, then please reach out to both of us as soon as possible, so we can work with you to find a path forward that supports you.)

**Group work**: Collaboration and discussion with others is *strongly encouraged* in this course; several of the assignments will be much easier if you work in a group. However: **Unless explicitly stated on the assignment instructions, you must submit your own answers/writing**. Identical language used by multiple individuals will be regarded as plagiarism (again, unless explicitly stated otherwise on the assignment instructions).

## Plagiarism

Cheating—and plagiarism specifically—is a very serious violation of both academic integrity and CMU policy. Plagiarism occurs *whenever* you present someone else's words or ideas as your own. *This includes material taken from the Internet.* For proper citation, you should cite any ideas and quotes that are due to other people. (Your choice about citation format, but please be consistent.) Basically, you should cite anything that (a) you got from someone else; and (b) a reasonable person would not know ahead of time. We realize that the vast majority of you will never consider cheating. However, a few of you may (for a variety of reasons) be tempted to plagiarize others' work. Do not take chances with plagiarism: **if you are uncertain whether you are doing something acceptable, please just ask**. We are happy to answer questions about whether something constitutes plagiarism.

## Accommodations for students with disabilities

If you have a disability and have an accommodations letter from the Disability Resources office, we encourage you to discuss your accommodations and needs with us as early in the semester as possible. We will work with you to ensure that accommodations are provided as appropriate. (Note that there are no exams in this course.) If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, then we encourage you to contact them at access@andrew.cmu.edu.

## Statement of support for your well-being

We as a community have to support one another. If you, or anyone you know, experiences academic stresses, difficult life events, or feelings of anxiety or depression, then we strongly encourage you to seek support. Take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep, and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is almost always helpful. Counseling and Psychological Services (CaPS) can provide assistance at 412-268-2922 or http://www.cmu.edu/counseling/ Consider reaching out to a friend, faculty or family member you trust for help getting connected to support that can help.

## Schedule of Topics

**[Tentative & subject to change]**

**Feb. 2** **Introduction & course overview**

**Feb. 4** **What is "ethics & policy of data analytics"? (via case study)**
*Reading*: Jobin, *et al.*, "The global landscape of AI ethics guidelines"
*Optional reading*: Aizenberg & van den Hoven, "Designing for human rights in AI"

**Feb. 9** **Privacy: Core concepts**
*Reading*: "Privacy" entry on *Stanford Encyclopedia of Philosophy*
*Reading*: Acquisti, *et al*., "Privacy and human behavior in the age of information"

**Feb. 11** **Privacy: Ethical considerations**
*Reading*: Mittelstadt, "From Individual to Group Privacy in Big Data Analytics"

**Feb. 16** **Privacy: Policy constraints & guidelines**
*Reading*: Crawford & Schultz, "Big data and due process…"

**Feb. 18** **Privacy: Integrated analysis**
*Reading*: Chander, *et al*., "Catalyzing Privacy Law"
*Guest speaker* (hopefully): JoAnn Stonier, Mastercard

**Feb. 23** **NO CLASS (CMU break day)**

**Feb. 25** **Fairness & bias: Core concepts & measures**
*Reading*: Mitchell, *et al*., "Prediction-based decisions and fairness"
*Optional reading:* Kusner & Loftus, "The long road to fairer algorithms"
*Optional reading:* Selbst, *et al.,* "Fairness and Abstraction in Sociotechnical Systems"
*Optional reading*: Obermeyer, *et al*., "Dissecting racial bias in an algorithm..."

**Mar. 2** **Fairness & bias: Ethical considerations**
*Reading*: Hellman, "Measuring algorithmic fairness"

*Optional reading*: Herington & Glymour, "Measuring the Biases that Matter"
*Optional reading*: Binns, "Fairness in machine learning..."

**Mar. 4** **Fairness & bias in policy & practice**
*Reading*: Wachter, *et al*., "Why fairness cannot be automated" (focus on Sec. III)
*Guest speaker*: Kathy Baxter, Salesforce.com

**Mar. 9** **Explainability: Core concepts**
*Reading*: Lipton, "The Mythos of Model Interpretability"
*Reading:* Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms"
*Optional reading*: Coyle & Weller, ""Explaining" machine learning reveals policy challenges" [This is quite short, and worth considering]
*Optional reading:* Kaur, *et al.*, "Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning"

**Mar. 11** **Explainability: Ethical & policy considerations**
*Reading:* Wachter, *et al.*, "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR" [Perhaps only particular sections]
*Guest speaker* (hopefully): Bayan Bruss, Capital One

**Mar. 16** **Trust: A unifying approach?**
*Reading*: TBD (depending on how the class evolves…)
*Guest speaker*: Jim Guszcza, Deloitte (on sabbatical)