UNIVERSITY OF NOTRE DAME

MENDOZA COLLEGE OF BUSINESS

ITAO 40510 Ethics of Data Analytics
Spring 2023

Professor: Cam Kormylo
L018 Mendoza College of Business
ckormylo@nd.edu
Cell: 262-818-0444

Class Meets: Section 3 M/W 2-3:15 DeBart 242
Section 4 M/W 3:30-4:45 DeBart 244

Office Hours: **OFFICE: M/W 10:30-11:30am** or email me for a time to meet.
VIRTUAL: Th 2:00-3:00pm.

Course Overview and Objectives

Course Overview

Every day, in a business close to you, someone furrows their eyebrows and thinks, "This just seems wrong." Sometimes that moment turns into a larger discussion. Sometimes that moment gets buried and ignored – at least until the problem manifests into a full-blown dilemma. With data analytics, increasingly that problem finds its way to the press.

In this course, we will explore the ethical implications of data analytics. We will marry old ideas – privacy, surveillance, power, justice, accountability, corporate responsibility, stakeholder theory – with new technologies and cases, such as the use of machine learning to predict crime. The goal is to tie our discussions around data analytics with ethical concepts that transcend our current issues. This way, you will be equipped to deal with newer, ambiguous capabilities once you graduate. Technology changes, but issues of power, privacy, justice, surveillance, and marginalization withstand the test of time.

The course is designed so that each day, we analyze a current 'case' (in the form of an article or two) with a theory or concept. Our readings will include the original works of philosophers, economists, law professors, and computer scientists. Tackling the ethical issues of data science requires a wholistic approach, and no one discipline has the answer.

We are one of a few business schools to offer a course in the ethics of data analytics. While the computer scientists you may work with in the future study these topics, most business schools do not. And because we are in a business school, the course is also designed to explicitly link the ethics of data analytics to theories in business ethics (corporate responsibility, stakeholder theory, economics of trust, etc). The goal is to provide you the tools to be able to make strategic arguments about the ethics of data analytics using the language and concepts of business.

Course Objectives

In this course, you will:
1. Develop the capacity to have constructive conversations about ethical dilemmas of data analytics.
2. Identify types of ethical issues within each stage of data analytics.
3. Examine a particular ethical issue in data analytics and apply multiple theories to understand possible solutions.
4. Explain the current issues in AI/ML/Predictive Analytics in the language of business and business ethics.
5. Develop a series of questions to ask of any data analytics program in order to identify the possible moral quandaries and ethical implications.

Required Material

Textbook: *None*

Articles: All required readings are provided on Canvas within the 'Module' labeled for the day the reading is due. Each day includes readings labeled "Case" and readings labeled "Concept" You are expected to have read the required readings prior to class.

Course Format, Assignments, Evaluation and Grading

Format

We will run the classroom as a case discussion. We will rely upon you and your peers to discuss and question the cases and examples in order to identify lessons, develop moral awareness and empathy, and practice arguing about ethical issues. In order to get the most out of a case discussion, *everyone must be ready, attentive, and participating.* Please be sure to read and study the cases and readings that are assigned. Students are expected to attend each class and be prepared and ready to contribute.

Performance Measurements

- Participation 30%
- Pre-Class Qs 20%
- Memo 10%
- Black Mirror Writers' Room 10%
- Final Report 30%

Participation (30%).
Your participation will make this class what it is and will help your colleagues think through the ethical issues. If you have questions about participating, please ask and we can work on ways to make sure you participate. You will receive feedback (from me and anonymously from your peers) on your participation in order to ensure there are no surprises. The cases and assignments are designed to reinforce the importance of participation.

If you are not comfortable participating – please come see me or reach out to me ahead of time. The time we have together is short and case discussion start on the first day. I have ideas as to how to make your participation easier.

Each class will assume that you have done all the readings, are prepared to discuss with your peers, and apply the materials during class learning activities. At a minimum, I expect you to attend class and prepare readings in advance; that would constitute an entry-level "C." In order to receive a better grade, you must participate. Participation is graded based on attendance, your willingness to participate, demonstration of having prepared for class, and thoughtfulness of class participation.

I provide a midpoint grade and feedback for participation the week of April 3rd. You will provide feedback to me in an assignment – where your contribution was particularly helpful in moving the class forward and who you look forward to hearing from in class. This is 1-2 sentences and should take 30 seconds. My grade of that assignment is your first half grade in participation.

While this does not replace your participation grades in class, you can add to your participation score by submitting relevant news articles to the Google Sheet linked below. The article should be about data analytics and discuss an issue that is ethical in nature. These links will serve as inspiration for the course for their final report.

Google Sheet

Recommended News Sources:
https://www.wired.com/category/ideas/ (Particularly Relevant)
https://techcrunch.com/
https://gizmodo.com/tech
https://www.theverge.com/tech

Pre-Class Qs (20%) – due before each class.
Prior to each class, you will submit a question you would ask of any future data analytics project given the case/concepts read for that day with 1-2 sentences explaining why. It should be clear from your question and explanation that you read the material. This is lightly graded (0, 1, 2) with most everyone getting a 2. This is due BEFORE class and no late submissions will be accepted. The goal of this assignment is to have you thinking about more generalized yet practical implications of the specific case you read for class.

The link to this 'quiz' is in the module for a given day.

Memos (10% total) – due by 4/3 11:59pm

Each student is expected to submit one written case study memo by the midpoint of the class, April 3rd. The memo is 1-2 pgs and will analyze the case assigned for that day using the theory also assigned for that day. You are to give me your assessment of the issue using reasoning from the course. I will grade you on your ability to understand and apply the ethical framework. The write-up is due *before* the class discussion that day – in other words you do not get the benefit of the class discussion before completing the write-up. A question designated 'memo' below cases in the reading section provides the prompt for your case memo.

Memo Rubric – The emphasis is on the theory being applied.

	No credit	**Partial Credit**	**Full Credit**
Structure 20%	No structure	Structure but not around theory	Structure based on theory
Theory 80%	No theory	Parts of theory but not all	All of the theory correctly applied
Case	No case mentioned	Case generally mentioned with no specifics	Case facts used in analysis

Black Mirror Writers' Room (10% total) – In Class Assignment April 5th

As Alec Nevala-Lee says, "Science fiction has been closely entwined with military and technological development from the very beginning." The TV show Black Mirror is a perfect example of *speculative fiction* in which each of its episodes is set in a near-future dystopia highlighting a technology that could realistically exist in the next 10, 50, 100 years... For this assignment, we will use this concept to encourage speculative thinking and think abstractly about what the ethics of data analytics will look like in the years to come.

On the day of the assignment, each of you will be put into small groups and will develop your own episode of Black Mirror. The general outline of the class will be:

Step 1: Each group is assigned one of the course topics we have discussed up to this point in the class. You will brainstorm ideas related to the topic, especially: (1) current events and news that seem relevant, (2) things that currently worry you related to data analytics, and (3) where you think this issue might be headed in the future.

Step 2: Come up with a story. Think about the issue your team wants to create a cautionary tale for. What fictional future person could help illustrate this caution? What is their story?

Step 3: Create your episode pitch. Come up with a title and condense your story into a 100-word blurb. Also find an image that illustrates it. Put your materials into the slide template I provide.

Step 4: Present! Each team will be given a few minutes to present their episode to the class.

Step 5: After class, you will individually (no longer a group assignment) reflect on how the future envisioned by your Black Mirror episode could be avoided. In 200-300 words, choose ONE theory from class to discuss how we, as a society, could work towards preventing the negative consequences of the technology used in your pitch. Due before the next class on 4/12 (you will have Easter Break to reflect on this).

Rubric:

	No credit	Partial Credit	Full Credit
Group Episode Pitch 50%	Not present in class.	Ethical issue currently being faced (little speculative thinking)	Original idea, realistic but some years off
Individual Reflection 50%	No submission.	Incorrectly utilize theory	All of the theory correctly applied

***** Prior to this class, I will host an optional screening of the Black Mirror episode "Nosedive" for anyone who hasn't seen the show or would just like a refresher. Pizza and dessert provided!
****** This assignment is adopted from Casey Fiesler of the Internet Rules Lab.

Final Report (30%) – due May 2nd

You are assessed on a final case report (there is no final exam). Each student will write a report on an issue in data analytics of their choice. *This assignment will allow you to connect classic ethical cases, problems, and theories to current dilemmas and offer solutions.* The 5-6 page report must have headings and use of at least two theories from the class to analyze an issue of your choice (but not a case we covered in class). All papers will have explicit use of theory applied to a real case. I will provide a list to all news articles submitted by the class for the "Current Events" assignments. This can be used to spur ideas for the final report.

The goal of this assignment is to use theories from the class to analyze a real data analytics dilemma. I have provided time in class for peer feedback on the proposal and outline. The proposal (4/3) and outline (4/17) will also be submitted for feedback from me. The project is due during the last week of class (5/3).

20% of this assignment is your ability to link your analysis to four examples from the class discussion.

The paper or project must include the following (% corresponds to portion of the final report grade).

1. 10%. Explain an ethical issue/case in data analytics. This can be from a few news articles. 1-2 pages maximum. This should read similar to a newspaper article without any ethical analysis.
2. 60%. Apply a minimum of 2 concepts from class to analyze what (if anything) is wrong with the case. The concepts must be clearly labeled (headings, bolded, underlined, etc). 4 pages minimum.
3. 10%. Provide a possible alternative or solution to any problem identified in #2 above. This should include a technical alternative such as a different set of training data used, different algorithm or approach in data analytics, or

different type of governance approach (just as examples). Half page approximately.

4. 20%. The report should include four references to examples or cases from class. These can be either assigned cases or examples that come up in the class discussion. This would be as short as "this company violated the privacy expectations of their users by using the commercial data set similar to the case of ad tech in class, both….."

There are three steps to this process. You will submit a proposal and outline as attempts in Canvas for the final report. I will provide feedback on both. You will be given time to work on the proposal and outline in class to get feedback from your peers.

- Proposal: April 5th.
- Simple Outline ("The goal of this paper is to…" + major sections): April 17th
- Final Report: May 3rd

Grading

Grading criteria for assignment of final course grade is based on:

Percentage of Course Points Earned	Grade Earned
93.0 % and above	A
from 90.0 % but less than 93.0 %	A-
from 87.0 % but less than 90.0 %	B+
from 83.0 % but less than 87.0%	B
from 80.0 % but less than 83.0%	B-
from 77.0 % but less than 80.0 %	C+
from 73.0 % but less than 77.0 %	C
from 70.0 % but less than 73.0 %	C-
from 60.0 % but less than 70.0 %	D
less than 60.0%	F

*The professor reserves the right to 'curve' grades to comply with the Mendoza College of Business' grade point average range of 3.0 to 3.6.

The students are expected to adhere to the University Honor Code, *Student Guide to Academic Code of Honor* (www.nd.edu/~hnrcode). Any violations of the Honor Code will be referred to the appropriate committee.

Course Schedule

Class	Date	Topic	Readings	Assignments
1	Wed 3/8	Intro to Algorithm Bias	1. Stanford Vaccine 2. Friedman/Nissenbaum 3. Winner	None
	3/12-3/18	Spring Break	None	None
2	Mon 3/20	Privacy	1. DNA Data 2. Tracking 3. Nissenbaum	• Pre Class 2 Q
3	Wed 3/22	Biased Data Sets	1. Wrongfully Accused 2. AI Slurs 3. Gebru	• Pre Class 3 Q
4	Mon 3/27	Purpose of a Corporation/A.I.	1. Race Detection 2. Face Scanning 3. Freidman 4. Freeman	• Pre Class 4 Q
5	Wed 3/29	Fairness & Justice	1. Machine Bias 2. Criminal Risk Score 3. Rawls 4. Walzer (Video)	• Pre Class 5 Q
6	Mon 4/3	Surveillance & Power	1. Twelve Mil Phones 2. Lyon	• Participation F/B • Memo • Pre Class 6 Q
	Tue 4/4	OPTIONAL: Black Mirror Screening - 8pm	None	None
7	Wed 4/5	Black Mirror Writers' Room	BM Season 3, Episode 1: Nosedive	• Final Proposal • Pitch (In Class)
	4/7-4/10	Easter Break	None	None
8	Wed 4/12	Disparate Treatment & Impacts	1. Secret AI 2. Credit Scores 3. Barocas/Selbst	• Black Mirror Reflection • Pre Class 8 Q
9	Mon 4/17	Economics of Trust & Stakeholders	1. Insurrection 2. Frank	• Pre Class 9 Q • Final Outline
10	Wed 4/19	Gamification & Manipulation	1. Uber Tricks 2. Deepfakes 3. Kim/Werbach	• Pre Class 10 Q
11	Mon 4/24	Transparency & Accountability	1. Secret Evals 2. Hao 3. Mulligan/Kluttz	• Pre Class 11 Q
12	Wed 4/26	Creating & Measuring Accuracy	1. Criminal Schoolchildren 2. Metrics	• Pre Class 12 Q
13	Mon 5/1	Summary Class	1. Predicting Cheaters	None
14	Wed 5/3	No Class - Final Exam Due at 11:59pm	None	• Final Report

Wed 3/8 Introduction to Algorithmic Bias Stanford Vaccine

The idea that computer programs or technology has bias - preferences or a forcing function to guide users - has long been understood. The question for us is how do biases 'look' in the area of AI or data analytics? We will read two articles on technological biases and discuss the possible biases in the Stanford Vaccine algorithm.

The goal of the class is be able to identify types of bias in analytics, sources of those biases, and how those biases could be ethical or unethical.

Case:

1. Guo, Eileen and Karen Hao 2020. "This is the Stanford vaccine algorithm that left out frontline doctors" MIT Technology Review https://www.technologyreview.com/2020/12/21/1015303/stanford-vaccine-algorithm/

Questions:

1. *Is there anything wrong with using a computer program to determine who gets a vaccine? Why or why not?*
2. *If the program is wrong, who is at fault?*
3. *What type of bias do you see in the Stanford algorithm?*

Concepts:

- Pg. 330—336 only. Friedman and Nissenbaum. Bias in Computer Systems. https://nissenbaum.tech.cornell.edu/papers/Bias%20in%20Computer%20Systems.pdf
- Pg. 121-128 only. Winner, Langdon. 1980. "Do Artifacts Have Politics?" Daedalus 109(1): 121-136.

Mon 3/20 Privacy & Shared Responsibility Ad Tech

Privacy has been defined in three ways: (1) as the degree you control your information, (2) as whether others have access to you and your information, and (3) the negotiated norms around how information should be gathered, shared, used (and by whom). We will focus on the third definition and discuss the limitations of the first two. Our questions focus on whether privacy is violated with the cross referencing of DNA data and who is responsible for any privacy violations, harms, or diminishment of value/rights.

Case:

- Use of rape-kit DNA to probe other crimes shocks prosecutors
- Tracking Overview: Angwin. J. 2020. "What They Know … Now" The Markup. https://themarkup.org/blacklight/2020/09/22/what-they-know-now

Concepts:

- **Pg. 32-36 only**. Nissenbaum, Helen. 2011. "A Contextual Approach to Privacy Online." Daedalus 140(4): 32-48. http://www.cs.cornell.edu/~shmat/courses/cs5436/contextualapproach.pdf

1. *MEMO: When does DNA cross referencing violate consumers' privacy expectations? Why?*
2. *Who is responsible for any privacy violations with hyper targeted advertising?*

Wed 3/22 Biased Data Sets. Facial Recognition & NLP

For analytics that relies on 'learning' from data, how does the choice of data impact the eventual model? Data scientists, who choose the data, have a role in assessing the appropriateness of the data as representative of the 'real world' and as impacting the eventual model used for categorizing and predicting. For us, we will focus on the ways data could be biased, how that bias is different (if at all) from biases in the experienced world, and who is responsible for the moral implications of that bias.

Case:

- Hill, K. 2020. Wrongfully accused By an Algorithm. NY Times. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html?action=click&module=RelatedLinks&pgtype=Article
- SKIM SITE. Buolamwini, Joy, and Timnit Gebru How well do IBM, Microsoft, and Face++ AI services guess the gender of a face? http://gendershades.org
- Actual Data. Quach, K. 2020. MIT apologizes, permanently pulls offline huge dataset that taught AI systems to use racist, misogynistic slurs. https://www.theregister.com/2020/07/01/mit_dataset_removed/

Concepts:

- SKIM. Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford**.** "Datasheets for datasets." arXiv preprint arXiv:1803.09010 (2018). https://arxiv.org/pdf/1803.09010.pdf

1. *MEMO: Was the arrest morally wrong? Why or why not? Who is at fault?*
2. *Do corporations have an obligation to vet data sets used in creating AI? Why or why not?*

Discuss In Class.

- NLP for class discussion. Ghaffary, S. 2019. Vox. The algorithms that detect hate speech online are biased against black people. https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter

Mon 3/27 Inferences/Purpose of Corp Facial Recognition

Facial recognition technology can be used not only for identification purposes (or mis-identification), but also to predict or guess something about the person. The idea is that the structure of the face 'reveals' hidden attributes about the person. Here we will focus on the use of facial recognition programs to identify race or emotions -- and whether the purpose of the program fits with what we understand to be the purpose of the corporation.

Cases Facial Recognition

- Olson, Parmy. 2020. The quiet growth of race detection software. The Wall Street Journal
- Harwell, D. 2019. A face-scanning algorithm increasingly decides whether you deserve the job. The Washington Post

Concepts. Purpose of the firm.

- Freidman – Corporate Responsibility of Corporations is to Increase Profits.
- Freeman et al. Chapter on Stakeholders.

Questions:

1. *<u>MEMO:</u> Should a company develop AI software that purports to identify ethnicity? Why or why not?*
2. *Should a company use AI software that purports to identify 'emotions'? Why or why not?*
3. *What is a good use of technology that would identify ethnicity? emotions?*

Wed 3/29 Fairness. (Rawls, Nozick, Walzer). Compas

Fairness, Justice, and COMPAS Sentencing Program

The question of fairness is important for us since analytics and AI are frequently used to allocate 'things' to people (employment, vaccines, ads, medical care, punishment, sentencing, money, etc). Who gets what and whether the allocation is fair will impact the legitimacy of the analytics program. Here we study two different philosophers of justice - Rawls and Walzer - to analyze whether the COMPAS sentencing algorithm is fair.

Rawls is always concerned with inequalities and finds inequalities suspect and needing a justification. Rawls wants to know if the position to which the inequality is attached was open to everyone and whether the entire system works out for everyone -- but particularly the least fortunate in society. Rawls also gives us the tool of the veil of ignorance to design the system of allocation, where we should pretend to not know who we are and what goods we have when we design the system of allocation. Rawls gives us an argument when we find a system reinforces existing inequalities or advantages one group over others.

Walzer introduces the idea of spheres of justice and provides us with a way of seeing 'winners' in the allocation in one sphere of life (e.g., education) should not automatically become winners in another sphere of life (e.g., health care). Walzer provides language and an argument when we see someone 'unjustly' winning or losing in a sphere such as sentencing a prisoner based on attributes from another sphere (income of father).

Case:

- Compas Algorithm. Julia Angwin, Jeff Larson, Surya Mattu, Lauren Kirchner, "Machine Bias," ProPublica (May 23, 2016) available at https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- Angwin, J. Larson, J. 2016. Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say. ProPublica. https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say

1. *MEMO: Is using Compas to make sentencing or parole decisions fair? Why or why not?*

Concepts:

- Rawls: https://eportfolios.macaulay.cuny.edu/thorne15/files/2015/03/Rawls-JUSTICE-AS-FAIRNESS.pdf
- Walzer. Video: https://bigthink.com/videos/michael-walzer-what-is-justice Spheres of Justice. (full). http://fs2.american.edu/dfagel/www/Philosophers/Walzer/SpeheresofJustice-Chapters%201-3.pdf

Mon 4/3 Surveillance/Power and Shared Responsibility.

Data on where we go, who we see, and what we do is collected not only by social network sites and trackers on our websites. Location data aggregators collect detailed GPS location data on individuals and are able to track individual's behavior offline. Detailed data is not only used in advertising, alternative credit or trust scores, based on where we go and who we see, can be used to assess the trustworthiness of individuals who are not known to the company for employment, lending, education, or just matching with other users. Here we will analyze whether detailed location data should be used for trust scores and under what conditions companies should use trust scores to assess users, applicants, or employees. We will use David Lyon thoughts on surveillance to examine who is impacted by the use of these scores. People that study surveillance are concerned about the power dynamic between the watcher and the surveilled as well as how the surveilled may change their thoughts and behavior merely by thinking they could be surveilled.

Case – Location and Trust Scores

- Thompson, S. Warzel, C. Twelve Million Phones, One Dataset, Zero Privacy. 2019 NY Times: https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html

Concepts

- Pg. 57-67 Only. Lyon, D. From Big Brother to Electronic Panopticon**.** *The Electronic Eye: The Rise of Surveillance Society* (Minneapolis: University of Minnesota Press, 1994): 57-80. https://home.fnal.gov/~annis/digirati/otherVoices/Lyon.html or https://www.jstor.org/stable/pdf/10.5749/j.ctttsqw8.7.pdf?refreqid=excelsior%3A72520fa63646e674483b1b1918203b34

1. *MEMO: What is worrisome about the collection or use of information in these articles? Why?*
2. *Should a company use a trust score for potential employees that tracks their location? for the promotion of current employees? Customers? Why or why not?*
3. *Who is impacted by the decision to use trust scores with location data?*

Wed 4/12 Disparate Treatment/Impact Hiring

Within both employment and housing law, companies are responsible for both the disparate treatment (intentionally treating people differently based on special classes) as well as disparate impact (having an adverse impact on people based on their protected class). Protected classes can include race, color, religion, ethnicity, sex, national origin, disability status, as well as others. While disparate treatment may seem obvious (i.e. you can't treat someone differently just because they are female or because they are from a specific country you don't like), disparate impact is the use of a 'facially neutral' practice that has an unjustified adverse impact on members of a protected class. The easiest example is if someone had a height requirement for admittance into a school of 6' tall. While 'facially neutral' (it doesn't say anything about a protected class), the rule would have an adverse impact on women.

Here we need to better understand how algorithms, analytics, and AI could run into the rules around disparate treatment and impact since so many AI programs are used in areas where these laws apply.

Case: Hiring

- Amazon scraps secret AI recruiting tool that showed bias against women | Article. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G .
- Heaven, Will Douglas. 2021. "Bias isn't the only problem with credit scores—and no, AI can't help." *MIT Technology Review.*
- FOR CLASS EXERCISE: 2019. Schools are using software to help pick who gets in. What could go wrong? Fast Company. https://www.fastcompany.com/90342596/schools-are-quietly-turning-to-ai-to-help-pick-who-gets-in-what-could-go-wrong

Concepts:

- Selection only. Barocas Solon, and Selbst Andrew D. 2016. "Big Data's Disparate Impact." California Law Review. P. 672-712.

1. *Is hiring via algorithm fair? What could go wrong given Barocas and Selbst's paper?*
2. *What would Walzer, Rawls, or Nozick think about Amazon's program?*
3. *IN CLASS: If we were to design a project for admissions to college, how would you design it? We will work on this in class.*

Mon 4/17 Economics of Trust Content Moderation

How much to modify in an automated system is value-laden; we decide when we override a programmed decision or when/where we insert human judgement into an AI enabled decision. The governance of automated systems, to trim back problematic outcomes, is a constant problem for companies whether curating videos, comments, user content, or music. Here we examine how to design a content moderating program that aligns with the goals of the firm. We will ask *which stakeholders' concerns matter in designing this program?*

Robert Frank is an economist at Cornell who argues (in the assigned reading) that acting in a trustworthy manner, taking into consideration stakeholders, being a responsible corporation, etc is economically advantageous. Such acts, if visible, allows stakeholders - such as employees, customers, suppliers, investors - to more easily find companies with matching values and allows others to make commitments faster and with fewer safeguards (because they trust you).

For us, we are concerned with how to assess data analytics where the incentives may not align with the long term value of the firm or with the interests of the company's stakeholders.

Case – Content Moderation

- CASE: Recommending an Insurrection

1. *Should Facebook do more to control their content? Why? Who should they take into consideration when making this decision?*
2. *What should be the end goal of a recommender system? Of a content moderation system?*
3. *Who should matter when making a recommender system? Content moderation system? Give an example of content and who should be considered.*

Concepts. Economics of Trust

- Frank, Robert H. "Can socially responsible firms survive in a competitive environment." Codes of Conduct: Behavioral Research into Business Ethics (Russell Sage Foundation, New York) (1996): 252-261.

Wed 4/19 Gamification, Manipulation, and Analytics

Gamification is a part of a suite of data analytics tactics designed to influence decision making, which includes the use of dark patterns, manipulative advertising, and deepfakes. All seek to influence an individual – their beliefs, their behaviors, their decisions – in a manner that is not obvious to the target. When employed in their best possible use, these tactics act for the betterment of the individual (the target) and society. However, when employed in alternative uses, these data analytics tactics can be exploitive and undermine individuals' decision making.

We explore how firms utilize these tactics to 1) manipulate workers into longer hours by using gamified "badges" and 2) generate hyper-personalized advertisement that increase willingness to pay.

Case – Gamification

- How Uber Uses Psychological Tricks to Push Its Drivers' Buttons
- How Deepfakes could change fashion advertising.

1. *In what contexts is the use of gamification permissible? When does it cross the line?*
2. *What issue, if any, should we have with "badges" in the labor market?*
3. *What is the most troublesome use of deepfakes? Why?*

Concepts – Gamification and Manipulation

- Kim, Tae Wan and Kevin Werbach. Ethics of Gamification. pp 1-36.

Mon 4/24 Transparency v accountability People Analytics

When decisions are deemed fair or legitimate, people care not only as to the outcomes of 'who-gets-what', but also the process by which the decision was made and how they were treated along the way. This is distributional (who gets what), procedural (how it was decided), and interactional fairness (how I was treated along the way). For decisions leveraging data analytics, explaining procedural and interactional fairness can be a challenge since explaining how the decision was made may be obscured by the designer of the program. However, much work has been done in CS to explore how AI-enabled decisions can be more transparent and how designers and users of data analytics tools should be held accountable. We will focus on two examples of people analytics (i.e., the analysis of employees, students, users based on their behavior as captured in data) and how the programs were designed for transparency, accountability, and (as we will read) 'contestability'.

Case. Uber/workers Workers AI

- Houston teachers to pursue lawsuit over secret evaluation system. https://www.houstonchronicle.com/news/houston-texas/houston/article/Houston-teachers-to-pursue-lawsuit-over-secret-11139692.php

Concepts

- Hao, Karen. 2019. When algorithms mess up, the nearest human gets the blame. Technology Review. https://www.technologyreview.com/2019/05/28/65748/ai-algorithms-liability-human-blame/
- Mulligan, Deirdre K. and Kluttz, Daniel and Kohli, Nitin. Shaping Our Tools: Contestability as a Means to Promote Responsible Algorithmic Decision Making in the Professions (July 7, 2019). https://www.cambridge.org/core/books/after-the-digital-tornado/shaping-our-tools-contestability-as-a-means-to-promote-responsible-algorithmic-decision-making-in-the-professions/311281626ECA50F156A1DDAE7A02CECB/core-reader

    1. *Do the teachers need to know about the use of AI or how it works?*
    2. *Why are the teachers of AI upset? Could this have been avoided?*
    3. *Where could you see the need for AI transparency? Accountability? Contestability?*

Wed 4/26 Creating & Measuring Accuracy. Sheriff/school/parole

Measuring whether data analytics 'works' is not clear cut. The outcome being measured ('fitness score') has a relationship with the overall goal of the project ('good employee'); and that outcome is what is measured as being 'accurate' or not. We examine two related issues with 'accuracy' and AI programs: (1) whether enough information is gathered to measure 'accuracy' including false positives and false negatives; (2) whether is labeling the outcome variable (e.g., 'possible cheater'), we then put in place steps to make that outcome more likely ('caught cheater').

We first examine accuracy in regards to algorithms that may create poverty traps. We then examine how to measure accuracy with labeling students as potential criminals. Constructing the outcome of interest and how to measure accuracy is an important step within data analytics with moral implications.

Case:

- Pasco's sheriff uses grades and abuse histories to label schoolchildren potential criminals. https://projects.tampabay.com/projects/2020/investigations/police-pasco-sheriff-targeted/school-data/

Concepts:

- 2020 The problem with metrics is a big problem for AI. https://www.fast.ai/2019/09/24/metrics/

Questions:

1. *Should schools provide student indicators for potential criminals to police? Why or why not?*

Mon 5/1 Summary – The Everything Goes Wrong Case.

The cheating detection software used to proctor online exams exemplifies many of the issues we have identified in class. We will use this as a summary class – to show the many ways design of data analytics programs can go wrong.

- Students: Harwell. 2020. Washington Post. Cheating-detection companies made millions during the pandemic. Now students are fighting back. https://www.washingtonpost.com/technology/2020/11/12/test-monitoring-student-revolt/

Questions:

1. *Why are the students upset with the cheating detection program? Use any theory from the class to make your case.*

Wed 5/3 **NO CLASS – REPORT DUE BY MIDNIGHT**