Ethics of Data Science
APSTA-GE 2062-001, 4 units

Time: Thursdays, 5:00 – 7:10 pm
Location: Waverly Building, 24 Waverly Place, Room 433
Office Hours: Tuesdays, 3:30 – 5:30 (drop me a line if you intend to attend)
Course Instructor: Laura Norén, laura.noren@nyu.edu

Course Description:
Ethics of Data Science is designed to build students' ethical imaginations and skills for collecting, storing, sharing and analyzing data derived from human subjects including data used in algorithms. The course provides historical background to understand the tenets of informed consent, discrimination, and privacy. Using case study design, students will explore current applications of quantitative reasoning in organizations, algorithmic transparency, and unintended automation of discrimination via data that contains biases rooted in race, gender, class, and other characteristics.

Course Overview
The power of data analytics to help us understand our world increases daily due to technological advances in strategies for collecting data (passively), implementing studies (randomized experiments), analyzing data (algorithms divorced from theory, history, or a fully contextualized understanding of the consequences of data-driven decisions), and disseminating findings both to broad public audiences and to narrow groups who are disproportionately impacted. Algorithmic decision-making has not been accompanied by a commensurate increase in our understanding of the consequences of choices made by human and machine actors and assemblages of human-in-the-loop sociotechnical systems. Legal rules lag ethical frameworks. In data science, data capture, storage, and model deployment are fragmented across national and cultural boundaries. This course focuses on understanding the ethical implications of empirical research including ethical considerations specific to data collection, study design, data analysis, and the dissemination and application of findings. It provides practical guidance about how to uncover ethical weaknesses in existing protocols and undertake constructive, effective, fair data scientific research and application of automated processes.

Learning Objectives:
After satisfactorily completing this course students will be able to:
1) identify and assess the ethical impacts of a given course of action in data-driven organizations
2) describe techniques for protecting privacy, sharing data ethically, and minimizing both collective and individual harm associated with data-driven organizational processes.
3) Perform an ethical audit of data-driven processes in a given organizational context.

Course Format:
The course will meet weekly for 2.25 hours. The course will include a mix of lectures, group discussion, guest lectures, and small group in-class activities. The professor will begin each class by reinforcing key tenets of the readings as well as historical background absent from the readings via lecture. Then, along with a student assigned on a rotating basis, the professor will lead a discussion about the required readings. Students will all be expected to provide discussion questions, but only the rotating student lead will be expected to guide the discussion. There will be two or three guest lectures from legal and philosophical experts from NYU Law and the Data & Society Institute (located in New York). Students will be expected to be active, respectful, and intellectually additive classroom participants.

Course Requirements

The grade for this course will be determined as follows:

- 10% assigned discussion leadership
- 15% class participation, including attendance
- 20% weekly responses
- 25% first ethical audit
- 30% second ethical audit

Students will be required to participate in class activities by asking questions or making comments a minimum of 20 times during the course of the semester. Participation and attendance will be documented by the students in a log they will be asked to maintain and share with the professor (who has the discretion to amend, as applicable).

First ethical audit of the Northpointe case:

The Northpointe case will be assessed by the instructor using a standardized rubric designed specifically for this assignment on a 100 point scale. The rubric will contain sections for writing techniques and communication skill; understanding of the Northpointe case and criminology; proper application of theoretical constructs and precedents; and the identification and presentation of techno-ethical findings of the audit. Students will also be provided unique written feedback addressing their strengths and possible shortcomings.

Second ethical audit:

The second ethics audit will use a rubric, written feedback strategy, and 100 point scale similar to the first ethics audit. In this case, pairs of students will be asked to provide a three-way assessment of the overall project, their contribution, and their partners' contribution. The professor will also provide an assessment of the overall project.

Reading responses and discussion questions:

Weekly reading responses will be assessed on a 5 point scale.

1 = submission received
2 = student demonstrates understanding of at least one reading
3 = student demonstrates understanding of at least one reading or part of a reading, provides a thoughtful discussion question for at least one reading or part of a reading
4 = student demonstrates understanding of the full reading assignment
5 = student demonstrates understanding of the full reading assignment, provides meaningful analysis, insight, and examples during discussion

Class Participation:

This course is highly interactive, both in terms of working and learning in teams and as a classroom. Interaction takes a variety of forms, ranging from small group discussions, seminar participation, and leading discussions of particular readings. Different skills are emphasized at different times. The evaluation of class participation uses a flexible scale so that everyone can achieve the highest measure. For each class meeting students will provide a self-assessment and receive an assessment from a randomly selected peer: 1=present, 2=responsive, 3=active. During sessions in which the student is co-leading the discussion, they will receive three assessments: a self-assessment, a peer assessment, and an instructor assessment for preparedness, grasp of the material, and capacity to spark discussion.

The overall participation grade is obtained by averaging over the class sessions.

Letter grades will be assigned using the following criteria:
A= Excellent [90-100% of points available in a given assignment category]
This work is comprehensive and detailed, integrating themes and concepts from discussions, lectures and readings and offering valuable original insight. Writing is clear, analytical and organized. Arguments offer specific examples, incorporate relevant literature, and concisely evaluate evidence. Students who earn this grade are prepared for class, synthesize course materials, contribute insightfully, and craft salient techno-ethical scholarly contributions.

B=Good [80-89% of points available in a given assignment category]
This work is complete and accurate, offering insights and competent understanding. Writing is clear, uses examples properly and tends toward broad analysis. Classroom participation is thoughtful and frequent.

C=Average [70-79% of points available in a given assignment category]
This work is correct but is largely regurgitates readings, lacking synthetic analysis. Writing is vague and fails to completely address the key questions. Arguments are unorganized, without specific analysis. Classroom participation is lacking or inarticulate.

D= Unsatisfactory [60-69% of points available in a given assignment category]
This work is incomplete, and evidences little understanding of the readings or discussions. Arguments demonstrate inattention to detail, misunderstand course material and overlook significant themes. Classroom participation is spotty, unprepared and off topic.

F=Failed [<60% of points available in a given assignment category]
This grade indicates a failure to attend class, participate in class, and/or complete assignments.

Assignments

1) *Weekly responses:* This is a reading heavy course. In order to make sure it is also a thinking heavy course, students will be asked to engage with the texts through written responses to be submitted via email as well as in-class discussions. Ethics is a practice. Regular, ongoing, deep engagement with the ideas in the texts is a key strategy for developing an ethical imagination. Each week students will prepare 400-500 word responses to the readings, due before class via email. Responses will crystallize the authors' key points and offer a question or set of questions based on the text and any relevant previous readings suitable for in-class discussion.

2) *Co-leading class discussion*: Depending on enrollment, each student will be expected to co-lead the discussion portion of the course at least twice with a student partner. They will ideally co-lead discussions with a single partner, though with high enrollment there may be groups of three. Discussions will be assessed based on ability to identify key questions, present a balanced overview (include all the readings and the viewpoints within them), and tie in key

ideas from previous weeks.

3) *Perform an ethical audit of the COMPAS/Northpointe bail setting case:* Jails and prisons are overcrowded; judges' schedules are overbooked, threatening the constitutional right to a *speedy* trial and putting municipal budgets under pressure. Cities and counties are challenged to maintain public safety without expanding their physical capacity to house convicted criminals. One of the ways they hope to minimize costs and social harms is to release as many accused people out on bail to await trial in the community as they can without reducing safety levels in the community. This allows those who are released to keep their jobs, maintain childcare responsibilities (if any), and otherwise continue to support themselves, their families, and their communities. These benefits must be balanced against the potential harms that occur if accused criminals are released into situations where they commit additional crimes. Judges already make these kinds of bail decisions every day. In many situations, their dockets are so full they have very little time to adjudicate any particular case. Now data-analytics companies are offering to use predictive models to give each arrestee a recidivism score that time-strapped judges can use in their rapid-fire decision making context.

   We look at the company Northpointe which offers a recidivism prediction score in roughly a dozen states. Investigative journalists at ProPublica have denounced the company for producing racial bias as measured by the differential rate of false positives between whites and blacks. Northpointe has countered this accusation, arguing that their algorithm is fair, focusing attention on community protection true positives and true negatives
   Students are asked to perform an audit of this case using ethical theory from weeks 2 and 3 as well as a socio-statistical assessment. How do we balance harms to society, rights to fair trials, budgets, and communities? Is any predictive analytics platform likely to lead to ethical lapses? Is a data-driven decision making strategy ethically acceptable if it outperforms the humans and organizational processes previously in place or can we develop a new ethical template for assessing the impact of data-driven decision making tools?

   Essays will provide historical background related to predictive policing, American attitudes towards punitive not rehabilitative policing, and carceral budgeting. They will then present an original ethical audit of the Northpointe/COMPAS algorithm ad its application.

   Essay length will range from 12-15 pages.

4) *Perform an ethical audit of your own research partnership (or a research project from outside of class):* In teams of two, students will perform an ethical audit of a project they are working on in an internship or for another course. If neither partner has worked on a suitable project, they may select from a submitted list of faculty research projects. In pairs, students will gather relevant case material, examine research protocols, data capture and storage, and investigate the impact of ongoing or future applications (if any). Students will provide an analysis of ethical adherence, lapses, discuss likely impacts on the organization, the subjects, the data, and key stakeholder groups throughout the community.

Essay length will range from 12-15 pages.

Required Readings and/or Text

The readings are all articles that are referenced in the course outline. All are posted on NYU Classes.

Course Outline

Week 1: Overview of ethical issues in data-driven organizations

<u>Overview of data science as an ethical practice</u>

O'Neil, Cathy. (2013) On being a data skeptic. p. 1-19. Sebastopol, CA: O'Reilly Media. http://www.oreilly.com/data/free/files/being-a-data-skeptic.pdf

boyd, danah. (2017) Toward accountability: Data, Fairness, Algorithms, Consequences. *Data and Society: Points*. [blog post] Accessed online: https://points.datasociety.net/toward-accountability-6096e38878f0

<u>Introduction to the unique ethical challenges of 'big data'</u>

Crawford, Kate. (2013, April 1) The hidden biases in big data. *Harvard Business Review.* https://hbr.org/2013/04/the-hidden-biases-in-big-data

Lerman, Jonas. (2013, September) Big data and its exclusions. Stanford Law Review. Accessed online: https://www.stanfordlawreview.org/online/privacy-and-big-data-big-data-and-its-exclusions

Week 2: Ethical Theory - Philosophical frameworks for assessing fairness

<u>Early theories of fairness</u>

Rousseau, Jean-Jacques. (1754) *Discourse on the origin and basis of inequality among men.* [Many print versions of this text have been published. Available in full here: https://www.aub.edu.lb/fas/cvsp/Documents/DiscourseonInequality.pdf879500092.pdf]

Mill, John Stuart. (1861) *Utilitarianism.* [There are multiple published versions of this treatise - it is also available in full here: https://www.utilitarianism.com/mill1.htm]

<u>Moving towards contemporary theories of fairness</u>

Blackburn, Simon. (2001) Selections from *Ethics: A Very Short Introduction.* pp.75-92. Oxford University Press.

Rawls, John. (1971) A Theory of Justice: Revised Edition. Cambridge, MA: Belknap Press. [excerpts including the Veil of Ignorance]

Voorhoeve, Alex. (2009) Interview with Frances Kamm in *Conversations on Ethics.* Oxford University Press. Accessed online: http://personal.lse.ac.uk/voorhoev/frances%20kamm%20chapter.pdf

Week 3: Research ethics for data science

<u>Ethical side effects of the publish or perish system: p-hacking and small sample size</u>

Faden, Ruth R. and Beauchamp, Tom L. (1986) "The Concept of Autonomy" and "The Concepts of Informed Consent and Competence" in *A History and Theory of Informed Consent.* Pp. 235-297. [SKIM]

Gelman, Andrew and Loken, Eric. (2014) The statistical crisis in science. *American Scientist.* Accessed online: http://www.stat.columbia.edu/~gelman/research/published/ForkingPaths.pdf

Ioannidis, John P. A. (2005) Why most published research findings are false. *PLOS Medicine*. https://doi.org/10.1371/journal.pmed.0020124

Vitak, Jessica; Proferes, Nicholas; Shilton, Katie; and Ashktorab, Zahra. (2017) Ethics Regulation in Social Computing Research: Examining the Role of Institutional Review Boards. *Journal of Empirical Research on Human Research Ethics.* Vol. 12(5): 3722-382. DOI: 10.1177/1556264617725200

In-class: the Belmont Report, history of research ethics guidelines and protocols.

<u>The misapplication of informed consent in dataveillance practices</u>

Bishop, Libby. (2017). Big data and data sharing: Ethical issues. UK Data Service, UK Data Archive. Accessed online: https://bigdata.ukdataservice.ac.uk/media/604711/big-data-and-data-sharing_ethical-issues.pdf

Nunan, Daniel and Yenicioglu, B. (2013) Informed, uninformed and participative consent in social media research. *International Journal of Market Research* 55 (6), pp. 791-808. ISSN 1470-7853.

Week 4: Data ethics 101

Bainbridge, L. (1983) Ironies of Automation. *Automatica*. 19: 775-779.

Elish, M. C., Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction (We Robot 2016) (March 20, 2016). We Robot 2016 Working Paper. Available at SSRN: https://ssrn.com/abstract=2757236 or http://dx.doi.org/10.2139/ssrn.2757236

Narayanan, Arvind and Shmatikov, Vitaly. (2010, June). Privacy and security: Myths and fallacies of "Personally Identifiable Information". *Communications of the ACM.* Vol. 53(6): 24-26. Accessed online: https://www.cs.utexas.edu/~shmat/shmat_cacm10.pdf

Sandvig, Christian; Hamilton, Kevin; Karahalios, Karrie; and Langbort, Cedric. (2014) Auditing Algorithms: Research methods for detecting discrimination on internet platforms. Paper presented to "Data and discrimination: converting critical concerns into productive inquiry" a pre-conference at the 64th Annual Meeting of the International Communication Association; Seattle, WA.

Vedder, Anton. (1999) KDD: The challenge to individualism. Ethics and information technology. Vol. 1(4): 275-281. Springer. doi:10.1023/A:1010016102284

<u>Internet traces and Geofingerprints.</u>

Golle, Philippe and Partridge, Kurt. (2009) On the anonymity of home/work location pairs p. 390-397 in H. Tokuda et al. (Eds.): Pervasive 2009, LNCS 5538. Berlin: Springer-Verlag. Accessed online http://link.springer.com/chapter/10.1007%2F978-3-642-01516-8_26?LI=true

Kosinski, Michal, David Stillwell, and Thore Graepel. "Private Traits and Attributes Are Predictable From Digital Records of Human Behavior." *Proceedings of the National Academy of Sciences* 110, no. 15 (April 9, 2013): 5802–5805. doi:10.1073/pnas.1218772110.

Week 5: All data are human data: On the discriminatory trouble with training data

Barocas, Solon and Selbst, Andrew. (2016) Big data's disparate impact. California Law Review. Vol. 104(3). Accessed online: http://scholarship.law.berkeley.edu/californialawreview/vol104/iss3/2

Douglas, Laura. (2017) AI is not just learning our biases; it is amplifying them. Blog post. Medium. https://medium.com/@laurahelendouglas/ai-is-not-just-learning-our-biases-it-is-amplifying-them-4d0dee75931d

Grush, Loren. (2015) Google engineer apologizes after Photos app tags two black people as gorillas. The Verge. http://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black- people-gorillas

Roth, Lorna. (2009) Looking at Shirley, the Ultimate Norm: Colour Balance, Image Technologies, and Cognitive Equity. Canadian Journal of Communication. Vol 34(1) fotopodcast https://fotopodcast.de/wp-content/uploads/2015/12/2196-5220-1-PB.pdf

Week 6: Cases of discrimination and data-driven decision making

In some cases, algorithms obscure unintentional bias. In other cases, algorithmic bias is, if not intended then at least condoned/obscured, by company's refusal to share data with would-be third party auditors.

Eubanks, Virginia. (2018) Selections TBD from *Automating Inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

The ethics of price discrimination

Elegido, Juan. (2011, October) The ethics of price discrimination. *Business Ethics Quarterly*. Vol. 21(4): 633-660.

Angwin, Julia; Larson, Jeff; Kirchner, Lauren; and Mattu, Surya. (2017, 5 April) Minority neighborhoods pay higher car insurance than white neighborhoods with the same risk. ProPublica, co-published with Consumer Reports. Accessed online https://www.propublica.org/article/minority-neighborhoods-higher-car-insurance-premiums-white-areas-same-risk

Criminal justice by algorithm: COMPAS edit

Angwin, Julia; Larson, Jeff; Mattu, Surya; and Kirchner, Lauren. (2016, May 23) Machine bias. ProPublica. Accessed online: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Larson, Jeff and Angwin, Julia. (2016) Technical response to Northpointe. ProPublica. Accessed online: https://www.propublica.org/article/technical-response-to-northpointe

Deiterch, William; Mendoza, Christina; and Brennan, Tim. (2016, July 8) COMPAS risk scales: Demonstrating Accuracy equity and predictive parity. [Report] Northpointe, Inc. Research Department. Accessed online: https://www.documentcloud.org/documents/2998391-ProPublica-Commentary-Final-070616.html

Week 7: The philosophical challenge of thinking in categories

How humans explain their social worlds through perceptions and statistics

Kahneman, Daniel. (2011) Excerpt from *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux. [See Scientific American "Of 2 minds: How fast and slow thinking shape perception and choice"]

Social processes and the impact of categorical life

Bowker, Geoffrey and Leigh Star, Susan. (2000) Sorting things out: Classification and its consequences. Chapter 1: Some Tricks of the Trade in Analyzing Classification p. 33-50; Chapter 5 Of Tuberculosis and Trajectories p. 165-194; Chapter 6: The case of race classification and reclassification under apartheid p. 195-225; Why classifications matter, p. 319-326.

Week 8: Data sharing

Health Research

The International consortium of investigators for fairness in trial data sharing. (2016) Toward fairness in data sharing. *The New England Journal of Medicine.* 375: 405-407. DOI: 10.1056/NEJMp1605654

Obermeyer, Ziad and Emanuel, Ezekiel. (2016) Predicting the future - big data, machine learning, and clinical medicine. *The New England Journal of Medicine.* 375: 1216-1219. DOI: 10.1056/NEJMp1606181

(2014) Personal Data for the Public Good. Final report of the health data exploration project. Robert Wood Johnson Foundation. Accessed online: http://www.rwjf.org/content/dam/farm/reports/reports/2014/rwjf411080

Javitt, Gail. (2010) Why Not Take All of Me? Reflections on The Immortal Life of **Henrietta Lacks** and the Status of Participants in Research Using Human Specimens. *Minnesota Journal of Law, Science, and Technology.* Vol. 11: 713.

Gross, Terri. (2010) Henrietta Lacks: A donor's immortal legacy. [interview; 37 minutes] *NPR: Fresh Air.* Accessed online: http://www.npr.org/2010/02/02/123232331/henrietta-lacks-a-donors-immortal-legacy

Educational Research

Bogle, Ariel. (2014) What the failure of inBloom means for the student data industry. *Slate*. Accessed online:

http://www.slate.com/blogs/future_tense/2014/04/24/what_the_failure_of_inbloom_means_for_the_student_data_industry.html

Slade, Sharon and Prinsloo, Paul. (2014) Student perspectives on the use of their data: Between intrusion, surveillance and care. *Proceedings of the European distance and e-learning network 2014 research workshop.* Oxford. Accessed online: http://oro.open.ac.uk/41229/1/BRPA_Slade_Prinsloo.pdf

Week 9: The ethics of data scraping and storage

Cate, Fred H. (2009) The Failure of fair information practice principles. Chapter 13 in *Consumer protection in the age of the information economy.* Routledge.

Kayyali, Dia and O'Brien, Danny. (2015) Facing the challenge of online harassment. [blog] Electronic Frontier Foundation. Accessed online: https://www.eff.org/deeplinks/2015/01/facing-challenge-online-harassment

Shiab, Nael. (2015) On the ethics of web scraping and data journalism. *Global Investigative Journalism Network.* Accessed online: http://gijn.org/2015/08/12/on-the-ethics-of-web-scraping-and-data-journalism/

Finch, Kelsey. (2016) A visual guide to practical data de-identification. Accessed online: https://fpf.org/wp-content/uploads/2016/04/FPF_Visual-Guide-to-Practical-Data-DeID.pdf

Week 10: Privacy and Surveillance

Brunton, Finn and Nissenbaum, Helen. (2015) *Obfuscation: A user's guide for privacy and protest.* Cambridge, MA: MIT Press. [This book must be purchased.]

Green, Matthew. (2016) What is differential privacy? [blog] A few thoughts on cryptographic engineering. Accessed online: https://blog.cryptographyengineering.com/2016/06/15/what-is-differential-privacy/

Nissenbaum, Helen. (2004) Privacy as contextual integrity. *Washington Law Review.* Accessed online: https://crypto.stanford.edu/portia/papers/RevnissenbaumDTP31.pdf

Week 11: The right to be forgotten and the GDPR

Jones, Meg Leta. (2016) *Ctrl + Z: The right to be forgotten*. Selections TBD. NYU Press.

Additional readings related to the roll-out of the GDPR TBD.

Week 12: Special topics in surveillance: Adtech

Sweeney, Latanya. (2013, May) Discrimination in online ad delivery. *Communications of the ACM*. Vol. 56(5): 44-54. Accessed online:
http://cacm.acm.org/magazines/2013/5/163753-discrimination-in-online-ad-delivery/fulltext

Marwick, Alice. (2014) How your data are being deeply mined. *New York Review of Books*. Accessed online:
http://www.nybooks.com/articles/2014/01/09/how-your-data-are-being-deeply-mined/

Citron, Danielle Keats and Pasquale, Frank. (2014) The scored society: Due process for automated predictions. *The Washington Law Review.* Vol. 89(1).

Week 13: Special topics in surveillance: Employment

Traub, Amy. (2011) Give us some credit. *The American Prospect.* Accessed online: http://prospect.org/article/give-us-some-credit

Peck, Don. (2013, December) They're watching you at work. *The Atlantic.* Accessed online: https://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/

Karkazis, Katrina and Fishman, Jennifer. (2017) Tracking US Professional Athletes: The ethics of biometric technologies. *The American Journal of Bioethics.* Vol. 17(1): 45-60. http://dx.doi.org/10.1080/15265161.2016.1251633

boyd, danah; Levy, Karen; and Marwick, Alice. (2014) The networked nature of algorithmic discrimination. In *Data and discrimination: Collected Essays*. Open Technology Institute.

Barocas, Solon. (2014) Putting data to work. In *Data and discrimination: Collected Essays* Open Technology Institute.

Week 14: Codes of Ethics for doing data science and artificial intelligence

<u>Guidance for acting ethically with data</u>

ACM US Public Policy Council. (2017) Statement on Algorithmic Transparency and Accountability. Association for Computing Machinery.

CPEDS code of ethics for data scientists. (Forthcoming)

IEEE Global Initiative for Ethical considerations in AI and Autonomous Systems. (2016) Ethically Aligned Design Version 1.0.

Markham, Annette; Buchanan, Elizabeth. Ethical Decision Making and Internet Research: Recommendations from the AoIR Ethics Working Committee Version 2.0.

Zook, Matthew; Barocas, Solon; boyd, danah; Crawford, Kate; Keller, Emily; Gangadharan, Seeta Peña; Goodman, Alyssa; Hollander, Rachelle; Koenig, Barbara A.; Metcalf, Jacob; Narayanan, Arvind; Nelson, Alondra; Pasquale, Frank. (2017) Ten Simple Rules for responsible big data research. *PLoS Computational Biology*. Vol. 13(3): e1005399. DOI: 10.1371/journal.pcbi.1005399 Accessed online: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005399

Academic Integrity:

All students are responsible for understanding and complying with the New York University Steinhardt School Statement on Academic Integrity. A copy of this statement is available at: http://steinhardt.nyu.edu/policies/academic_integrity.

Students with Disabilities:

Students with physical or learning disabilities are required to register with the Moses Center for Students with

Disabilities, 726 Broadway, 2nd Floor, (212-998-4980 and online at http://www.nyu.edu/csd) and are required to present a letter from the Center to the instructor at the start of the semester in order to be considered for appropriate accommodation.