

# SIADS 503: Data Science Ethics

## *Syllabus*

School of Information, University of Michigan  
Prof. Christian Sandvig  
Winter 2020

## **Part I. Course Basics**

### **Course Overview**

This class will teach you to recognize where ethical issues can arise when applying data science to real world problems. It will bring more analytic precision to ethical 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. Largely through discussion of case studies, we will focus on ways to conceptualize, measure, and mitigate harm in data-driven decision-making. You will learn to think critically about how to plan and evaluate a data science project with ethical 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 learn key technical, ethical, policy, and legal terms and concepts that are relevant to ethical assessment in data science; learn about some of the common approaches and emerging tools for mitigating or managing these ethical concerns; and gain exposure to readings that will help you understand the current ethical and regulatory environment and to 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.

### **Instructor and Course Assistants**

- Primary Instructor: Prof. Christian Sandvig
- Co-Instructor: Dr. Melissa Chalmers
- Graduate Student Instructor: Jaleesa Turner
- Instructional Assistants:
  - Amanda Stanhaus
  - Yuchen Chen
  - Danny Meyerend
  - Imani Cooper Mkandawire

### **Course Credit, Effort, and Prerequisites**

- This course provides one unit of graduate credit.
- This is an online course taught via Coursera.

- This course is structured in what the University of Michigan sometimes calls a “mini-class” format, so one unit of effort is arranged across only four weeks and not an entire semester. This results in about eight hours of effort expected per week.
- The pre-requisites for this course are admission into the M.S. program in Applied Data Science and successful completion of SIADS 501, “Being a Data Scientist.”

### Communication Expectations

Group Office Hours: Tuesdays, 10-11 a.m. Eastern Time on Zoom.

- **Contacting your instructor and course assistant(s):**  
Course channel in Slack = siads503\_Wi20
- **Email response time:** N/A (please communicate via Slack)
- **Slack response time:** 24 - 48 hours

### How to Get Help

- If you have questions concerning the degree program, encounter a technical issue with Coursera, or issues using Slack, please submit a report to the ticketing system at [umsimadshelp@umich.edu](mailto:umsimadshelp@umich.edu).
- If you have an issue specific to the Coursera environment, you can also begin a [live chat session](#) with Coursera Technical Support (24/7) or [view Coursera troubleshooting guides](#). (you may be asked to log in to your Coursera account).
- For questions regarding course content, refer to the **Communications Expectations** section above.

### Required Textbook

You will be given a list of required and recommended readings within the course. Online access to these readings are provided through the University of Michigan Library or through approved online sources. For resources provided through the library, you will be asked to sign in with your UMich uniqueness and password to access these materials.

There is no required textbook for this course that needs to be purchased separately from a bookstore (although we will read almost all of the book *Ethics and Data Science* by Loukides et al., linked below).

### Learning Outcomes

1. Achieve **literacy** on the potential harms of data collection, aggregation, and analysis typically found in applied data science contexts.
2. Achieve **literacy** in the most important terminology of ethics that applies to data science.

3. Achieve **literacy** in writing ethical assessments (e.g., a memorandum) of a data science analysis or an automated system incorporating data science.
4. Achieve **competency** in articulating the reasoning behind the most important ethical challenges of data science as applied to course domains of privacy, bias/classification, provenance/aggregation and accountability/consequences.

### Course Structure

Each week, this course consists of recorded **lectures** and an **office hour** produced by the teaching team. Lectures consist of overview material about **concepts** as well as the discussion of **case studies**. In addition, a **guest speaker** will visit or we will take a **field trip**. This will provide one or more additional recorded lectures, interviews, or conversations. Each week there will be both **required readings** and recommended ones.

A low-stakes, open-book weekly **quiz** will provide an incentive to keep up with the readings/viewing. The primary work of the course is one **writing assignment** each week.

Lectures supplement but do not always review or duplicate the readings; readings supplement but do not always duplicate the lectures. That means some of the course content is available only from a lecture or a reading. For instance a concept may not be mentioned in lecture, but it may be the key point of a reading. Students are still responsible for that material.

### Course Deadlines

This course **begins on Wednesday, March 4, 2020** and **ends on Tuesday, March 31, 2020**.

Weekly **Quizzes** and **Writing Assignments** will be due on **Tuesdays at 11:59 pm** (time zone is Ann Arbor, Michigan; Eastern Time).

### Grading

<u>Assignment</u>	<u>% of Final Grade</u>
Weekly Quizzes (4 x 2.5%/ea.)	10%
Weekly Writing Assignments	90%
Week 1: Memo about a privacy concern (15%)	
Week 2: Evaluation of the What-If Tool (25%)	
Week 3: Perform an algorithmic impact assessment (25%)	
Week 4: Design an ethics oath, pledge, or checklist (25%)	

Note: All assignments are required to earn credit for this course.

## II. Course Schedule

### Week 1: Introduction and Data Privacy

Framing Questions:

- What is Data Science Ethics?
- Does Data Science Hate Privacy?
- Is Privacy-Respecting Data Science Even Possible?

Major Class Topics This Week:

- The 15 Top Misconceptions About Data Science Ethics
- What is a “patient” or “stakeholder”?
- **Special Guest: Amanda Stanhaus**, School of Public Health, University of Michigan and Fellow, US National Institutes of Health

Additional Key Concepts:

ethics vs. law/compliance/public relations; cultural relativism; “professional” ethics in data science; individuals vs. collectives; the “nothing to hide” argument; the “Fair Information Practice Principles” and their problems; The Belmont Report principles; informed consent; privacy dependencies

Required Readings:

1. Loukides, Mike, Hilary Mason, and DJ Patil. 2018. [Ethics and Data Science](#). Sebastopol, CA: O’Reilly Media. Chapter 1, “Doing Good Data Science”
2. Loukides, Mike, Hilary Mason, and DJ Patil. 2018. [Ethics and Data Science](#). Sebastopol, CA: O’Reilly Media. Chapter 3, “The Five Cs.”
3. Solove, Daniel J. 2007. [“I’ve Got Nothing to Hide’ and Other Misunderstandings of Privacy.”](#) San Diego Law Review 44(May):1-23. Read the following sections: II. p 748- 753; IV. 764-772.
4. Cate, Fred H. 2006. [“The Failure of Fair Information Practice Principles.”](#) From Consumer Protection in the Age of the “Information Economy” by Jane K. Winn and Professor Geraint Howells. Read ONLY the following sections: pp. 345-348; pp. 352-353; pp. 356-365; pp. 369-374.
5. VIDEO: The New York Times - [Blood Journey by Kassie Bracken & Amy Harmon](#). April 2010. (11 minutes)
6. Solove, Daniel J. 2013. [“Introduction: Privacy Self-Management and the Consent Dilemma.”](#) Harvard Law Review 126(7):1880-1903. Read the following section ONLY: I. pp. 1882 - 1893.

7. Barocas, Solon, Karen Levy and Washington Law Review. 2018. "[Privacy Dependencies.](#)" [Washington Law Review 2-61](#). Read the following section ONLY: III. pp. 50-60.

Assessments Due:

- Week #1 Quiz
- Whistleblower Memo About a Privacy Concern

### Week 2: Bias and Classification

Framing Questions:

- Is Data Science Backward-Looking?
- Is it Inherently Discriminatory?

Major Class Topics This Week:

- Classification
- Cumulative Disadvantage and Protected Classes
- What is a "harm"?
- Field Trip: **Project Green Light**: University of Michigan Detroit Center; Detroit's Corktown Green Light District
- Special Guest: **Tawanna Petty**, Director of Data Justice Programming for the Detroit Community Technology Project.

Additional Key Concepts:

harm without intent; objectivity/ neutrality; bias toward the majority; error analysis; team diversity

Required Readings:

1. American Civil Liberties Union. (2004). "Scary Pizza." (video.) 01:42. Available online: <https://www.youtube.com/watch?v=33CIVjvYyEk>
2. Gitleman, Lisa, and Virginia Jackson. 2013. "[Introduction](#)" in "[Raw Data](#)" Is An [Oxymoron](#), 1-15. The MIT Press.
3. Hanna Wallach. [Big Data, Machine Learning, and the Social Sciences: Fairness, Accountability, and Transparency](#). December 14, 2014. Medium.com.
4. Angwin, Julia; Larson, Jeff; Mattu, Surya, Kirchner, Lauren. [Machine Bias: Two Drug Possession Arrests](#). The Louisiana Weekly; New Orleans, LA. 06 June 2016: 1,6.
5. Amy Harmon. [As Cameras Track Detroit's Residents, a Debate Ensues Over Racial Bias](#). New York Times.

6. Citron, Danielle Keats, and Frank A. Pasquale. (2014). "[The Scored Society: Due Process for Automated Predictions.](#)" Washington Law Review 89.

Practice materials for this week's assignment:

- **The What-If tool.**
  - Specifically, the COMPAS Recidivism Classifier "Notebook Demo With Attributions" dataset for the What-If Tool:  
[https://colab.research.google.com/github/pair-code/what-if-tool/blob/master/WIT\\_COMPAS\\_with\\_SHAP.ipynb](https://colab.research.google.com/github/pair-code/what-if-tool/blob/master/WIT_COMPAS_with_SHAP.ipynb)
  - See also the What-If Tool documentation, e.g., the walkthrough <https://pair-code.github.io/what-if-tool/walkthrough.html> (optional), or introductory videos about the tool on YouTube [https://www.youtube.com/results?search\\_query=%23whatiftool](https://www.youtube.com/results?search_query=%23whatiftool) (optional).

Assessments Due:

- Week #2 Quiz
- Memo Evaluating the What-If Tool

### Week 3: Accountability and Governance

Framing Questions:

- Can large, automated systems be effectively controlled?
- When are YOU responsible?

Major Class Topics This Week:

- How Transparency Works and Doesn't Work
- Algorithm Auditing, External Auditing, and Reverse Engineering
- How should we consider "power" and regressing/progressive acts in data science ethics?
- **Special Guest: J. M. Porup**, a freelance cybersecurity reporter whose work has appeared in *The Economist*, *Vice*, *Ars Technica*, *Motherboard*, and elsewhere.
- **Special Guest: Sol Bermann**, Former Director of International Privacy, Walmart Corporation; current Chief Information Security Officer, University of Michigan.

Additional Key Concepts:

governance; accountability workflows; consequences predictable by experts vs. discoverable by users; approaches to auditing; algorithmic accountability reporting; reverse engineering.

## Required Readings:

1. Mike Ananny and Kate Crawford, "[Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability.](#)" *New Media & Society* 20.3 (2018): 973-989.
2. Sandvig, Christian, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. forthcoming. "[Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms.](#)" *Computational Culture*.
3. Diakopoulos, Nicholas. (2015). "[Algorithmic Accountability.](#)" *Digital Journalism* 3 (3): 398-415.
4. Christian Grothoff & J.M. Porup. [The NSA's SKYNET program may be killing thousands of innocent people.](#) February 16, 2016. *Ars Technica*.
5. Drew Harwell. [Colleges are turning students' phones into surveillance machines, tracking the locations of hundreds of thousands.](#) *The Washington Post*.

## Assessments Due:

- Week #3 Quiz
- Memo Evaluating the What-If Tool

## ***Week 4: Data Provenance and Aggregation; Course Conclusion***

### Framing Questions:

- Does data science always leave something out?
- Are data ever truly portable?

### Major Class Topics This Week:

- Challenges of "public" data
- Sampling as an ethical problem
- What is a "norm"? What are grounds by which we make ethical claims?
- **Special Guest: Kathy Pham**, founder, Ethical Tech Collective; former founding member, Product and Engineering Team, United States Digital Service, The White House.

### Additional Key Concepts:

provenance (a.k.a. origination); sampling bias; aggregation; retention; disposition; forgetting / erasure; building ethics into a data science culture

### Required Readings:

1. Rieder and Simon (2016) [Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data.](#) *Big Data & Society* 3(1): 1-6.

2. Loukides, Mike, Hilary Mason, and DJ Patil. 2018. [Ethics and Data Science](#). Sebastopol, CA: O'Reilly Media. Chapter 2, "Of Oaths and Checklists"
3. Loukides, Mike, Hilary Mason, and DJ Patil. 2018. [Ethics and Data Science](#). Sebastopol, CA: O'Reilly Media. Chapter 4, "Data's Day of Reckoning"

Assessments Due:

- Week #4 Quiz
- Design and Ethics Oath, Pledge, or Checklist

### III. Additional Course Policies

#### Guidelines for Dialogue

This section primarily refers to discussion on our Slack channel, but it could apply to any communication among students or between students and instructors. This course practices the "Guidelines for Dialogue" developed by students and faculty from the University of Michigan [Program on Intergroup Relations](#). That means that in discussions (including written online discussion forums) we will do our best to:

1. Maintain confidentiality. We want to create an atmosphere for open, honest exchange.
2. Commit to learning from each other. We will listen to other and not talk at each other. We acknowledge differences among us in backgrounds, skills, interests, identities and values. We realize that it is these very differences that will increase our awareness and understanding through this process.
3. Not demean, devalue, or "put down" people for their experiences, lack of experiences, or difference in interpretation of those experiences.
4. Trust that people are always doing the best they can. We will give each other the benefit of the doubt. We will assume we are all trying our hardest and that our intentions are good even when the impact is not.
5. Challenge the idea and not the person. If we wish to challenge something that has been said, we will challenge the idea or the practice referred to, not the individual sharing this idea or practice.
6. Speak our discomfort. If something is bothering us, we will share this with the group. Often our emotional reactions to this process offer the most valuable learning opportunities.
7. Step Up, Step Back. We will be mindful of taking up much more space than others. On the same note, empower ourselves to speak up when others are dominating the conversation.



8. Not to freeze people in time. We are all works in progress. We will be willing to change and make space for others to do so. Therefore we will not assume that one comment or one opinion made at one time captures the whole of a person's character.

--The Program on Intergroup Relations, University of Michigan, 2012

### **Letter Grades, Course Grades, Late Submission Policy, and Appeals**

Refer to the [MADS Assignment Submission and Grading Policies](#) section of the UMSI Student Handbook (access to Student Orientation course required)

Final letter grades for the course will be calculated using the following scale: A 93%+; A- 90-92%; B+ 87-89%; B 83-86%; B- 80-82%; C+ 77-79%; C 73-76%; C- 70-72%; D+ 67-69%; D 63-66%; D- 60-62%; E 59% or below.

When a writing assignment is given a letter grade, it is represented in the gradebook as the middle of the range listed above (e.g., B+ is recorded as 88%). If the range contains an odd number of values the higher median is used (e.g., a C is recorded as 75%). An A+ is recorded as 100%.

For this course, **late work results in a reduction of one letter grade per day (10% on the above scale).**

If you feel one of your writing assignments has been mis-graded, you may appeal to the grader who issued the grade by direct messaging them on Slack within three days (72 hours) after receiving the assignment grade. Please write a professionally-worded statement that begins "GRADE APPEAL" (in caps please), identifies the assignment, and includes specific references to the Grading Rubric for Written Work (e.g., explain how it has been misapplied). Note that because your written work is being graded again, it is possible that your new grade goes down and not up -- for instance, the grader notices an additional problem they did not consider before.

### **Accommodations**

Refer to the [Accommodations for Students with Disabilities](#) section of the UMSI Student Handbook (access to the Student Orientation course required).

Use the [Student Intake Form](#) to begin the process of working with the University's Office of Services for Students with Disabilities.

### **Library Access**

Refer to the [U-M Library's information sheet](#) on accessing library resources from off-campus. For more information regarding library support services, please refer to the [U-M Library Resources](#) section of the UMSI Student Handbook (access to the Student Orientation course required).

## Student Mental Health

Refer to the University's [Resources for Stress and Mental Health website](#) for a listing of resources for students.

## Student Services

Refer to the [Introduction to UMSI Student Life](#) section of the UMSI Student Handbook (access to the Student Orientation course required).

## **IV: Bibliography of Additional Readings**

These are additional *recommended* readings that may be of interest to you. (Listed by week and topic.)

### Overall

- O'Neil, Cathy. 2014. [On being a data skeptic](#). O'Reilly Media, Inc.
- Zook, Matthew, Solon Barocas, Kate Crawford, Emily Keller, Alyssa Goodman, Rachelle Hollander, Barbara A. Koenig, Jacob Metcalf, Arvind Narayanan, Alondra Nelson, and Frank Pasquale. 2017. ["Ten Simple Rules for Responsible Big Data Research."](#) *PLOS Computational Biology* 13(3):1-11.
- Cornell Tech - Good Code Podcast. [Episode 8: Solon Barocas on Teaching Ethics in Data Science](#).
- Ananny, Mike. 2016. ["Toward an Ethics of Algorithms : Convening Observation, Probability , and Timeliness."](#) *Science, Technology, & Human Values* 41(1):93-117.
- Natasha Singer. [Tech's Ethical 'Dark Side': Harvard, Stanford and Others Want to Address It](#). The New York Times. February 12, 2018.
- Moor, James. H. 2005. ["Why We Need Better Ethics for Emerging Technologies,"](#) *Ethics and Information Technology* 7:111-119. [8 pages]
- Brey, Philip A. E. 2012. ["Anticipating Ethical Issues in Emerging IT."](#) *Ethics and Information Technology* 2012 [13 Pages].
- Metcalf, Jacob, Emanuel Moss, and danah boyd. 2019. ["Owning Ethics: Corporate Logics, Silicon Valley, and the Institutionalization of Ethics."](#) *Social Research: An International Quarterly* 82(2):449-76.

### Privacy

- Federal Trade Commission - Consumer Information. [Direct-to-Consumer Genetic Tests](#).
- Messner, Donna A. (2011). [Informed choice in direct-to-consumer genetic testing for Alzheimer and other diseases: lessons from two cases](#). *New Genetics and Society*, 30, 59-72.

- Udesky, Laurie. 2011. [“The Ethics of Direct-to-Consumer Genetic Testing.”](#) The Lancet 376(9750):1377–78.
- Cornell Tech - Good Code Podcast. [Episode 2: Helen Nissenbaum on Post-Consent Privacy.](#) March 26, 2019.
- Allyse MA et al. (2018). [Direct-to-consumer genetic testing 2.0: Emerging models of direct-to-consumer genetic testing.](#) Mayo Clinic Proceedings, 93(1), 113-20
- Kowal, Emma, Joanna Radin, and Jenny Reardon. 2013. [“Indigenous Body Parts, Mutating Temporalities, and the Half-Lives of Postcolonial Technoscience.”](#) Social Studies of Science 43:465.
- Mcguire, Amy L. and Laura M. Beskow. 2010. [“Informed Consent in Genomics and Genetic Research.”](#) Annual Review of Genomics and Human Genetics 11(361–81).
- Cate, F.H. and Mayer-Schonberger, V. [Notice and consent in a world of big data.](#) International Data Privacy Law 3, 2 (May 20, 2013), 67–73.
- Narayanan, Arvind and Shmatikov, Vitaly. (2010). [Privacy and Security: Myth and Fallacies of "Personally Identifiable Information."](#) *Communications of the ACM* 53(6): 26.
- Barocas, S., & Nissenbaum, H. (2014). [Computing Ethics: Big data's End Run Around Procedural Privacy Protections.](#) *Communications of the ACM* 57(11): 31-33.
- Schermer, B. W., Custers, B. H. M., & Van der Hof, S. (2014). [The crisis of consent: How stronger legal protection may lead to weaker consent in data protection.](#)
- Helm, Paula. 2018. [“Treating Sensitive Topics Online: A Privacy Dilemma.”](#) Ethics and Information Technology 20:303–13.

### **Bias and Classification**

- boyd, danah, Karen Levy, and Alice Marwick. (2014). [“The Networked Nature of Algorithmic Discrimination.”](#) Open Technology Institute.
- Bowker, Geoffrey C. and Star, Susan Leigh. [“Some Tricks of the Trade in Analyzing Classification,” in Sorting Things Out: Classification and Its Consequences,](#) pp. 33-50, Cambridge, MA: MIT Press, 1999
- Sandvig, Christian. (2015). [“Seeing the Sort: The Aesthetic and Industrial Defense of 'The Algorithm.'”](#) Media-N: Journal of the New Media Caucus 10(3).
- Joy Buolamwini & Timnit Gebru. (2017). [Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.](#) Proceedings of Machine Learning Research 81: 77-91.
- Chen, L., Ma, R., Hannak, A., & Wilson, C. (2018). [Investigating the Impact of Gender on Rank in Resume Search Engines.](#) Proceedings of the 2018 ACM Conference on Computer-Human Interaction (CHI'18).
- Haimson, Oliver L. and Anna Lauren Hoffmann. 2016. [“Constructing and enforcing ‘authentic’ identity online: Facebook, real names, and non-normative identities.”](#) First Monday 21 (June). [15 pages].

- 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.

### Accountability and Governance

- Karppi, Tero, and Kate Crawford. 2015. "[Social Media, Financial Algorithms and the Hack Crash.](#)" *Theory, Culture & Society* 33 (1): 73–92.
- boyd, danah. (2017) [Toward Accountability: Data, Fairness, Algorithms, Consequences.](#) *Data and Society: Points.* [blog post]
- Pasquale, Frank. (2015). [The Black Box Society: The Secret Algorithms That Control Money and Information.](#) Cambridge: Harvard University Press. Especially the following chapters: Ch. 1: "The Need to Know" and Ch. 6: "Toward an Intelligible Society."
- Pasquale, Frank. (2006). [Rankings, Reductionism, and Responsibility.](#) *Cleveland State Law Review*, 54: 115+

### Professional Pledges, Oaths, and Checklists

- Simonite, Tom. 2018, February 8. "Should Data Scientists Adhere to a Hippocratic Oath?" *Wired: Business.* <https://www.wired.com/story/should-data-scientists-adhere-to-a-hippocratic-oath/>
- United Nations. 1948. *Universal Declaration of Human Rights.* <https://www.un.org/en/universal-declaration-human-rights/>

## V. Acknowledgements

Thank you to Casey Fiesler, Natalie Garrett, and Nathan Beard's work on ethics and their ethics syllabi research project. This syllabus is in debt to the syllabi of Paul Conway. Thanks for specific suggestions from Solon Barocas and Karrie Karahalios. Thank you to our generous beta-testers and guest speakers, as well as to James Wexler, Fernanda Viegas, and Google PAIR (People + AI Research). Finally, general thanks to the Ethical Tech Collective and the FAT\* community.