

School of Communication & Information

Data in Context (04:189:220:02)

Fall 2024 (10516)

CLASS TIME: Tuesdays, 10:20am-1:20pm

LOCATION: HH-A7

INSTRUCTOR: Yehuda Perry

EMAIL: Yehuda.Perry@rutgers.edu

OFFICE HOURS: By appointment on Zoom

INSTRUCTIONAL SUPPORT: Liyang Xue

EMAIL: lx109@scarletmail.rutgers.edu

OFFICE HOURS: By appointment on Zoom

COURSE DESCRIPTION

This course introduces students to the interplay between data and the societal context surrounding them. We will cover the basics of data science, focusing on data and algorithms' ethical, legal, and social implications. Students will learn to conceptualize and evaluate practical applications of data science in communication, information, and media contexts.

LEARNING OUTCOMES

1. Develop the ability to assess and articulate the relevance of data for a particular organizational or societal problem
2. Identify frameworks for understanding the impact of data on society and that of society on data
3. Explain the background of quantification and data-fication, including its social and political purposes
4. Analyze and critique the ethical, legal, and social implications of data collection, data processing, and algorithm development
5. Design practical applications of data science in communication, information, and media contexts in ways that are sensitive to social, structural and political economic concerns
6. Communicate data science outputs to relevant audiences with attention to their Contexts

REQUIRED READINGS

There is no required textbook for this class. All readings are available on the course's Canvas website (canvas.rutgers.edu). Log in using your Rutgers NetID, navigate to the course site, and browse the Modules page. The reading materials for each week of class are also listed in the Weekly Topics section of this syllabus.

COURSE OUTLINE

Week	Date	Topic	Assignment Due
1	09.09.2024	Introduction	Mon, September 9 th , 2024
2	09.16.2024	Technology and Society	Mon, September 16 th , 2024
3	09.23.2024	Ethics of Data Science	Mon, September 23 rd , 2024 (Case study 1, Due Sept 30 th , 2024)
4	09.30.2024	Datafication & Quantification	Mon, September 30 th , 2024
5	10.07.2024	Science of Data Science	Mon, Oct 7 th , 2024
6	10.14.2024	Data Collection	Mon, Oct 14 th , 2024
7	10.21.2024	Experiment Design	Mon, Oct 21 st , 2024 (Case study 2, Due Oct. 28 th , 2024)
8	10.28.2024	Algorithm Design (Guest Lecture by Liyang)	Mon, Oct 28 th , 2024
9	11.04.2024	Algorithm Auditing	Mon, Nov 4 th , 2024
10	11.11.2024	Communicating Findings	Mon, Nov 11 th , 2024 (Midterm Project, Due Nov 18 th , 2024)
11	11.18.2024	Data Science and Discrimination	Mon, Nov 18 th , 2024 (Case Study 3, Due Nov 25 th , 2024) + (Final Project Part I, Due Dec 02 nd , 2024)
12	11.25.2024	Surveillance and Privacy	Mon, Nov 25 th , 2024
13	12.02.2024	Data Activism and Counter Data Science	Mon, Dec 02 nd , 2024 (Case Study 4, Due Dec 09 th , 2024)
14	12.09.2024	Physical Infrastructure and Political Economy of Data Science	Mon, Dec 09 th , 2024 (Final Project Part II, Due Dec 16 th , 2024)

WEEKLY TOPICS

* Reading lists are subject to change. Please refer to the Canvas Module for the most up-to-date readings.

MODULE 1: Laying the Groundwork

WEEK 1 (Sept 09-13): Introduction: What Is Data/Data Science? Why in Context?

- Course Syllabus

WEEK 2 (Sept 16-20): Technology and Society

- Baym, N. K. (2015). Making New Media Make Sense. Ch.2 in Personal Connections in the Digital Age (2nd Ed.). Malden, MA: Polity.
- Christin, A. (2020). What Data Can Do. International Journal of Communication 14: 1115-1134.
- Sacasas, L.M. (2014). Do Artifacts Have Ethics?
<https://thefrailestthing.com/2014/11/29/do-artifacts-have-ethics/>
- OPTIONAL:
 - Winner, Langdon. 1981. Do Artifacts Have Politics? Daedalus 109(1):121- 136

WEEK 3 (Sept 23-27): The Ethics of Data Science

- Donovan, J., Caplan, R., Matthes, J., & Hanson, L. (2018). Algorithmic Accountability: A Primer. Data & Society Research Institute.
<https://datasociety.net/library/algorithmic-accountability-a-primer/>
- Eubanks, Virginia. (2018.) A Hippocratic Oath for Data Science.
<https://virginia-eubanks.com/2018/02/21/a-hippocratic-oath-for-data-science/>
- Costanza-Chock, Sasha. (2020). Design Justice. Cambridge, MA: MIT Press. Introduction (pp. 1-24 – you can stop reading when you get to the sentence: “The book is organized as follows”).
- Principles for Accountable Algorithms and a Social Impact Statement for Algorithms:
<http://www.fatml.org/resources/principles-for-accountable-algorithms>

WEEK 4 (Sept 30 Oct 4): Datafication & Quantification in Historical Context

- Scott, J. (1998). Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed. Yale University Press. Chapter 1: Nature & Space, pp. 11-52.
- Igo, S. (2007). The Averaged American: Surveys, Citizens, and the Making of a Mass Public. Harvard University Press. Chapter 3, “Polling the Average Populace,” pp. 103-149.
- Hacking, I. (2006). “Making Up People.” London Review of Books.
- OPTIONAL:
 - Lippold, J.-C. (2017). We Are Data: Algorithms and the Making of Our Digital Selves. NYU Press. Introduction, pp. 1-36.
 - Porter, T. (1995). Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton University Press. Chapter 2: How Social Numbers Are Made Valid, pp. 33-48.
 - Hacking, I. “How Should We Do the History of Statistics?” The Foucault Effect, pp. 181-196.

WEEK 5 (Oct 7-11): Science of Data Science / Research Design

- Hayes, A. F. (2005). Statistical Methods for Communication Science. Mahwah, N.J.: Lawrence Erlbaum Associates. Ch 1-2.
- Salganik, M. J. (2017). Bit by Bit: Social Research in the Digital Age. Princeton, NJ: Princeton University Press. Ch.2 Observing behavior

MODULE 2: Designing & Auditing a Data Science Project

WEEK 6 (Feb 21-27): Collecting Data / Data Science Applications

- Salganik, M. J. (2017). Bit by Bit: Social Research in the Digital Age. Princeton University Press. Ch. 6 Ethics
- Mislove, A., & Wilson, C. (2018). A Practitioner's Guide to Ethical Web Data Collection. In The Oxford Handbook of Networked Communication.
- Siegel, E. (2016). Predictive analytics: The power to predict who will click, buy, lie, or die. Wiley. Appendix: 147 examples of Predictive Analytics.

WEEK 7 (Oct 14-18): Designing Experiments

- Kohavi, R., Tang, D., & Xu, Y. (2020). Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing. Cambridge University Press. Ch.1 Introduction and Motivation.
- Salganik, M. J. (2017). Bit by Bit: Social Research in the Digital Age. Princeton University Press. Ch. 4 Running experiments
- Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788–8790.
- OPTIONAL:
 - McDermott, R., & Hatemi, P. K. (2020). Ethics in field experimentation: A call to establish new standards to protect the public from unwanted manipulation and real harms. Proceedings of the National Academy of Sciences, 117(48), 30014–30021.

WEEK 8 (Oct 21-25): Designing Algorithms

- Kelleher, J., D., & Tierney, B. (2018). Data Science. MIT Press. Ch.5 Standard Data Science Tasks
- Kearns, M., & Roth, A. (2019). The Ethical Algorithm: The Science of Socially Aware Algorithm Design. New York: Oxford University Press. Ch. 2 Algorithmic Fairness.
- Fleming, G., & Bruce, P. C. (2021). Responsible Data Science. Wiley. Ch. 4 The Responsible Data Science Framework
- OPTIONAL:
 - Siegel, E. (2016). Predictive analytics: The power to predict who will click, buy, lie, or die. Wiley. Ch.4: The Machine That Learns: A Look Inside Chase's Prediction of Mortgage Risk.

WEEK 9 (Oct 28 Nov 01): Auditing Algorithms

- Shen, H., DeVos, A., Eslami, M., & Holstein, K. (2021). Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors. *Proceedings of the ACM on Human-Computer Interaction*, 5 (CSCW2), 1–29.
- Brown, S., Davidovic, J., & Hasan, A. (2021). The algorithm audit: Scoring the algorithms that score us. *Big Data & Society*, 8(1).
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency* (pp. 77-91). PMLR.
- Vincent, James. (2020). What a machine learning tool that turns Obama white can (and can't) tell us about AI bias. *The Verge*.
<https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias>
- OPTIONAL:
 - Metaxa, D., Park, J. S., Robertson, R. E., Karahalios, K., Wilson, C., Hancock, J., & Sandvig, C. (2021). Auditing Algorithms: Understanding Algorithmic Systems from the Outside. In *Foundations and Trends in Human-Computer Interaction*, 14(4), 272–344.

WEEK 10 (Nov 4–8): Communicating Findings / Data visualization

- Healy, K. (2018). *Data Visualization: A Practical Introduction*. Princeton University Press. Ch.1 Look at Data.
- Engebretsen, M., & Kennedy, H. (Eds.). (2020). *Data Visualization in Society*. Amsterdam University Press. Ch.22 Visualizing diversity: Data deficiencies and semiotic strategies
- Riche, N. H., Hurter, C., Diakopoulos, N., & Carpendale, S. (2018). *Data-Driven Storytelling*. CRC Press. Ch. 9 Communicating Data to an Audience
- OPTIONAL:
 - Fox, P., & Hendler, J. (2011). Changing the Equation on Scientific Data Visualization. *Science*, 331(6018), 705–708

MODULE 3: Social Implications of Data Science and Algorithms

WEEK 11 (Nov 11-15): Data Science and Discrimination

- Hoffman, A. L. (2019). When Fairness Fails: Data, Algorithms, and the Limits of Anti-Discrimination Discourse. *Information, Communication & Society* 22(7): 900- 915.
- Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge, UK: Polity. Chapter 2: Engineered Inequity, pp. 33-52.
- Davis, J., Williams, A., & Yang, M. 2021. Algorithmic Reparation. *Big Data & Society* 8(2).
- OPTIONAL:
 - Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NY: NYU Press. Ch.1 A Society, Searching
 - Wachter-Boettcher, S. 2017. *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*. Chapter 7: Algorithmic Inequity, pp. 119-146.
 - Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist—It's time to make it fair. *Nature*, 559(7714), 324.
 - Ruha Benjamin *Data & Society Databite* talk:
https://www.youtube.com/watch?v=zZEVAVi6_Ak

WEEK 12 (Nov 18-22): Surveillance, Privacy, Anonymization and Re-identification

- Kearns, M., & Roth, A. (2019). Ch. 1 Algorithmic Privacy. In *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press.
- Trotter, D. (2018). Ch. 25 Privacy and Surveillance. In J. Burgess., A. E. Marwick., & T. Poell. (Eds.) *The SAGE Handbook of Social-Media*. SAGE Publications.
- Singh, R., & Jackson, S. (2021). Seeing Like an Infrastructure: Low-resolution Citizens and the Aadhaar Identification Project. *PACM on Human-Computer Interaction*, 5(315). CSCW2.
- Lomas, Natasha. (2019). Researchers Spotlight the Lie of 'Anonymous' Datasets. TechCrunch.
<https://techcrunch.com/2019/07/24/researchers-spotlight-the-lie-of-anonymous-data/>
- OPTIONAL:
 - Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press. Ch. 2: Digital Reputation in an Era of Runaway Data.

WEEK 13 (Nov 25-29) Data Activism & Counter Data Science

- D'Ignazio, C. (2022). *Counting Feminicide: Data Feminism in Action*. MIT Press. "Introduction," pp. 1-25.
- Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, 4(2): 1-14.
<https://doi.org/10.1177/2053951717736335>
- Milan, S., & Van der Velden, L. (2016). The Alternative Epistemologies of Data Activism. *Digital Culture & Society*, 2(2): 57-74.
<https://doi.org/10.14361/dcs-2016-0205>
- OPTIONAL:
 - Milan, S. & Treere, E. (2020). The Rise of the Data Poor: The COVID-19 Pandemic Seen from the Margins. *Social Media + Society*.
 - D'Ignazio, C., & Klein, L. (2020). Seven Intersectional Feminist Principles for Equitable and Actionable COVID-19 Data. *Big Data & Society*, 7(2).
<https://doi.org/10.1177/2053951720942544>

WEEK 14 (Dec 2-6): Physical Infrastructures & Political Economy of Data Science

- Holt, J., & Vanderau, P. (2015). "'Where the Internet Lives': Data Centers as Cloud Infrastructure. In L. Parks., & N. Starosielski (Eds.). *Signal Traffic: Critical Studies of Media Infrastructures* (pp.71-93). University of Illinois Press.
- Gray, M., & Siddharth, S. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Introduction.
- Au, Y. (2022). Data centres on the Moon and other tales: A volumetric and elemental analysis of the coloniality of digital infrastructures. *Territory, Politics, Governance*, DOI: 10.1080/21622671.2022.2153160
- OPTIONAL:
 - Starosielski, N. (2022). "Introduction: Media Hot and Cold." In *Media Hot & Cold*. Duke University Press.
 - Griffiths, Sarah. (2020). Why your internet habits are not as clean as you think. BBC.
<https://www.bbc.com/future/article/20200305-why-your-internet-habits-arent-as-clean-as-you-think>
 - Bogost, I. (2015). The Cathedral of computation. *The Atlantic*.
<https://www.theatlantic.com/technology/archive/2015/01/the-cathedral-of-computation/384300/>

STUDENT ASSESSMENT

1. Case Studies (30%): 4 assignments, students pick 3 of them to complete. Each assignment is worth 10% of the final grade)

Throughout the semester, students will write three reflection memos (800 to 1,000 words each) on a contemporary case study selected by the instructors. Students will select a small number of sources (popular as well as scholarly, where applicable), and will then draw on concepts, theories, and ideas covered in class to analyze the case in their memos, with particular attention to its ethical, social, and legal ramifications. In particular, you will be asked to reflect on the question of "should we build X" for the various case studies.

(1) **Case Study 1 (Algorithm accountability case study):** This case study will ask students to reflect on thorny issues of accountability and culpability (both moral and legal) when semi-autonomous algorithms are involved in producing a damaging outcome.

DUE September 30 (Monday) by Midnight

(2) **Case Study 2 (Data collection & consent case study):** This case study will ask students to consider the ethics of a particular instance of data collection, taking into account issues of privacy, consent, and consequences.

DUE October 28 (Monday) by Midnight

(3) **Case Study 3 (Algorithm ethics case study):** This case study will ask students to weigh the ethical benefits and drawbacks of a particular algorithm.

DUE November 25 (Monday) by Midnight

(4) **Case Study 4 (Policy case study):** This case study will focus on policy issues arising from data science and AI.

DUE December 09 (Monday) by Midnight

2. Midterm Project: Critiquing a dataset (25%)

In this assignment, we will learn how to critically analyze a dataset. I will provide a dataset and ask you to examine it critically.

Consider the following aspects and connect them to the readings and concepts discussed in the class where appropriate. There is more than one possible answer for every prompt; try not to think of yes/no answers but rather describe your thought process for coming up with certain responses.

(1) Who has provided this data? How will you go about assessing the reliability and validity of the dataset? What did you find?

(2) Identify the strengths and limitations of the dataset. You can choose to focus on a subset of the following prompts.

What information can you find about the data collection process? What does each of the variables identified in the dataset mean? What are some of the limitations of the data collection process? In what ways do the data mimic the real world (or not)?

What are the biases and the assumptions built into the data collection process e.g., who has been "counted" in the dataset and who has been left out? Are there any ethical implications of this data collection process?

(3) Identify applications where this dataset could be useful and applications where the usage of these data would be problematic.

(4) Do you have any suggestions to improve this dataset (or its future versions)? The

suggestions could be for the data collection process, the interpretation of the data collected, and its application in different algorithms.

This is an individual assignment. The length of your submission should be between 1200-1500 words.

DUE November 18 (Monday) by Midnight

3. Final Project: Auditing an algorithm (30%)

In this assignment, we will practice auditing an algorithm. Specifically, we will evaluate the performance of a face image-based algorithm (e.g., face bounding box detection). You can select one of the following platforms (or, you may also choose to work with another API and/or another face-based recognition task):

- Kairos (<https://www.kairos.com/demos>)
- Face++ (<https://www.faceplusplus.com/face-detection/>)
- Everypixel (<https://labs.everypixel.com/api/demo>)
- Microsoft Facial Recognition (<https://azure.microsoft.com/en-us/products/cognitive-services/face/#demo>)

The goal of the final project is to examine the performance of these systems for faces representing different demographic groups. The final project consists of two parts:

(1) group work on data collection & analysis, and (2) an individual project report (1200-1500 words). Data collection and analysis should be undertaken in groups of three. Reports (including reflections) should be written and submitted individually. Grades will be individually assigned.

(1) Part 1: Data collection, coding, and sharing with the project group. (10%)

DUE December 02 (Monday) by Midnight

(2) Part 2: Project report (1200-1500 words) (20%)

DUE December 16 (Monday) by Midnight

4. Discussion & Class Participation (15%)

I expect students to attend every class and actively participate in the activities and conversations we will have. For each class, students will be asked to submit a short reflection at the end of the class, which will be counted towards their participation grade. If you cannot attend a class, please notify us in advance.

CLASS POLICY

1. Email Policy

Feel free to contact me via email if you have any questions/concerns about the class. I will be checking e-mail every day during the week (Monday-Friday). It is best to get in touch with me from 9am-5pm on weekdays. You can expect a response in less than 24 hours on weekdays, but you will be informed beforehand if scheduling requires a longer wait time. On weekends expect a response in 48-72 hours. Also, I will not respond to last-minute emails within 24 hours before assignment dues.

2. Late Work

Your assignments must be submitted on time. Late papers are reduced by 3% for every 24-hour period they are late. However, understanding there may be unexpected circumstances that hinder you from submitting assignments on time, you will have two opportunities (or what we call, "tickets") to extend & reschedule the deadline of a given assignment. If you choose to use the ticket, you should pick the new due date that you feel most confident to meet (but it should be no later than 2 weeks from the original due date).

You must notify the instructor(s) that you intend to use the ticket BEFORE the original deadline, and propose a new due date in the same email. If the assignment is submitted later than the new due date, late penalties will be imposed as usual.

You must use these tickets wisely & use them only when necessary. If you use up both tickets, you will not get an extension for late assignments that occur after that. Those who haven't used any tickets until the end of the term AND submitted all assignments on time will receive extra credit (1 point).

3. GRADE APPEALS

You can submit appeals for individual assignment grades up to 7 days after the grades are announced. Appeals submitted later than that will not be accepted. In order to be reviewed, your appeal has to be submitted in writing over e-mail to your instructors. It should present solid arguments demonstrating that you deserve a higher grade.

If you have concerns about your course grade, schedule an appointment early during the semester to discuss it. Once the course grades are announced, they are final and will only be changed in case of an error in the computation of the student's score.

4. Religious Observances

It is the university policy to respect the religious obligations of students, faculty, and staff. Should religious observances conflict with assignments, or the course schedule, please notify me well in advance so that an accommodation can be made.

5. Academic Misconduct

Academic misconduct includes plagiarism, cheating on examinations, providing work to other students that should be authored by them, or accepting and submitting such work as your own. ***To plagiarize means to take and pass-off as one's own ideas and/or the expression of ideas of another person.*** Evidence of academic misconduct constitutes grounds for failure of this course. Students must present original work and always document material drawn from outside sources. Students are expected to research and prepare their own work when submitting written assignments.

It is academic misconduct if you:

- buy a term paper
- copy articles in whole or in part from published sources or published on the Internet
- have another student *or computer* do your work. In other words, Use of AI such as ***ChatGPT is not permitted*** in any stages of the writing process on any assignment.
- turn in a previously written assignment and misrepresent it as new work,
- use source material in an assignment, print and electronic, without giving proper credit (attribution) to sources. APA in-text citations are expected ([Purdue OWL APA guide](#)).

Ethical conduct and academic integrity are expected of all students. Any student found guilty of plagiarism, fabrication of course work, cheating on examinations and quizzes, or purchasing papers or other assignments will be reported to the chair of the department, to the College Dean, and will receive a failing grade in the course.

GRADE BREAKDOWN

The grade breakdown is as follows:

Case Studies	3 x 10%
Midterm Project	25%
Final Project	30% (Part 1: 10% + Part 2: 20%)
Discussion & Participation	15%
Total	100%

The final grade will be awarded according to the following scale:

A	90%-100%
B+	85%-89.9%
B	80%-84.9%
C+	75%-79.9%
C	70%-74.9%
D	60%-69.9%
F	Below 60%

ACCOMMODATION

This course will accommodate any student in need of assistance. Students with documented disabilities who need accommodations should contact the Rutgers Disabilities Services Office (see disabilityservices.rutgers.edu for details). You can also speak with an SC&I adviser by calling them at 848-932-7500 (dial 2 as your menu choice).

Please contact me with information about the requested assistance and present your letter of accommodation as early in the semester as possible.

ADDITIONAL RESOURCES

The university offers a number of resources that you can access if needed:

- For additional tutoring, training, or writing help, visit the Rutgers Learning Center (online at rlc.rutgers.edu) and the Writing Center (plangere.rutgers.edu).
- If you need a consultation on research materials and ways to find them, you can contact the Rutgers University subject specialist librarians.
- If you need help with class schedule or registration, visit the Student Services Office located in CI 214, and online at comminfo.rutgers.edu/student-services/contact-us.html
- SC&I IT Services can help you with various technological problems. You can find them in CI 120, by phone at 848-932-5555, or by email at help@comminfo.rutgers.edu
- If you encounter a problem with Canvas, you can contact the Rutgers Canvas help desk at help@oit.rutgers.edu or call them at 833-648-4357.
- Student wellness services are available to you at Rutgers. You can contact CAPS for mental health support at rhscaps.rutgers.edu or by phone at 848-932-7884.
- The Office for Violence Prevention and Victim Assistance provides confidential crisis intervention, counseling, and advocacy for victims of sexual and relationship violence. You can reach VPVA at vpva.rutgers.edu and 848-932-1181.
- The Office of Disability Services can be reached for help with accommodation and facilities for students with disabilities at ods.rutgers.edu, or by phone at 848-445-6800.
- On occasion, the university may have to cancel classes due to inclement weather. To check if classes are canceled, visit campusstatus.rutgers.edu or call 732-932-7799.

ACKNOWLEDGEMENTS

This syllabus has benefitted from the syllabi shared by multiple educators including: Critical Algorithm Studies, Kaitlin Costello, Rutgers University, Data Science Ethics, Elisa Celis, Yale University, Ethics in Data Science, Suresh Venkatasubramanian, University of Utah, Dr. Youngrim Kim and Dr. Jonthan Bullinger, Rutgers University.