

Professor Maria Dikcis, PhD ([mariadikcis@fas.harvard.edu](mailto:mariadikcis@fas.harvard.edu))

Office Hours: Tuesdays 12-2 PM ([Book Appointment](#))

Office Location: Barker Center 015

English 195BD



# The Dark Side of Big Data

Spring 2024

Sever Hall 214

Tu/Th 10:30-11:45 AM

+ weekly section

## DESCRIPTION

Does it sometimes feel like Instagram ads are listening a little too closely to your conversations? Have you ever wondered if certain corporations might own images of your face? Today, fears abound that algorithms are not only populating our lives with annoying targeted advertisements but might also be creating the most unequal societies that have ever existed. In this interdisciplinary seminar, we will explore key methodological overlaps and differences between humanistic and scientific approaches to the phenomenon known as Big Data, or enormously large data sets that are analyzed by computer software to reveal patterns associated with human behavior and communications. In particular, we will focus our attention on the dark side of Big Data, which is increasingly embedded with harmful biases against women, people of color, immigrants, and low-socioeconomic status communities. Our inquiries will thus concern a wide array of issues that stem from the misapplication of Big Data, such as data discrimination, biased artificial intelligence, search engines that reinforce racism, predictive policing, and surveillance capitalism, as well as how these issues intersect with race, class, gender, and citizenship. We will ground these discussions about contemporary theories of Big Data in engagements with a number of literary texts, films, and new media artworks. These cultural case studies range from a poetry collection exploring anti-Blackness and the carceral state, a documentary on social media data scandals, a glitch feminism manifesto, a memoir about working at an Amazon fulfillment center, queer video games, and robot love poems.

## COURSE OBJECTIVES

1. Think critically about contemporary computing culture and synthesize ideas about related topics from diverse disciplinary perspectives, including data science, literary studies, media theory, critical race and ethnic studies, aesthetics, and gender and sexuality.
2. Conduct careful readings of both primary texts/media and theoretical sources by gathering concrete evidence to support your observations and claims.
3. Engage in conversation with your peers during weekly discussion sections and practice communicating original ideas and informed stances through spoken dialogue.
4. Develop your writing and analytical skills by conducting interpretations of primary texts/media and producing either a final research paper *or* creative data visualization project.

## **POLICIES**

1. Arrive to class on time and attend all lectures and discussion sections. Attendance will be taken at the beginning of each class. Having a few unexcused absences here and there will not impact your participation grade. Four or more *unexcused* absences, however, will seriously jeopardize your ability to receive a passing grade for Attendance and Participation (20% of your total grade). If you must be excused from class for a valid reason, illness, or emergency, please try to communicate with me as soon as possible so we can plan accordingly.
2. Complete all the readings before the class session for which they are assigned and have them readily available during class (online or printed out, if applicable).
3. Be prepared to participate in a respectful manner with lively comments, questions, and responses to your peers during lecture and section. Giving voice to your own thoughts and having conversations with one another will undoubtedly help you to gain a better understanding of the material discussed in this course. If you anticipate that you will have difficulties speaking in class, please meet with me early on so we can discuss possible strategies for participation.
4. Submit all assignments by their due dates. Late submissions without an approved extension will lose one point for each day late. If you foresee that you will need an extension for an assignment, please communicate with me in advance of its due date.
5. Cell phones must be silenced and put away during class. Laptops and other electronic devices may only be used to take notes and access course materials. Any violations of this policy (browsing social media, emailing, etc.) will result in a lower participation grade.

## **REQUIRED TEXTS**

You do not need to purchase anything for this class. All readings, films, and other media materials are posted on Canvas or linked on our [digital syllabus](#). If you would like to purchase physical copies of *Just Us*, *Seasonal Associate*, or *Glitch Feminism*, you may do so at [The Coop](#).

## **GRADING**

Attendance and Participation: 20%

Paper 1: 20%

Paper 2: 20%

Final Project: 40%

## **ACCESSIBILITY**

I am committed to creating a class environment that is accessible to all student needs. Any students requiring reasonable accommodations related to a disability or other condition must register with the Disability Access Office ([dao@fas.harvard.edu](mailto:dao@fas.harvard.edu), 617-496-8707) and provide the Professor with a notification. All correspondence with me will remain strictly confidential. Please visit the [Disability Access Office](#) for more information on requesting an accommodation.

## **ACADEMIC INTEGRITY**

Certain assignments in this course will permit the use of generative artificial intelligence (GAI) tools. The default is that such use is disallowed unless otherwise stated. Any such use must be appropriately acknowledged. Violations of this policy will be considered academic misconduct. As outlined in Harvard College's [Student Handbook](#) and GSAS's [Policies](#), plagiarism is a serious offense that can result in serious consequences, up to and including withdrawal from the University. If you ever have questions about how to properly cite someone else's work, just ask me! Please visit [The Honor Code website](#) for helpful guides on writing and using sources.

## SCHEDULE

### Week 1: The Dark Side of Big Data

Tu 1/23 [Big Data](#), Bonilla and Mata

Th 1/25 “Introduction,” *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, O’Neil  
[“Big Data: Inside the Algorithm,”](#) *Do Not Track*, Gaylor

### Week 2: What’s in a Face?

Tu 1/30 [This Person Does Not Exist](#), Wang (refresh the page a few times)  
[Which Face is Real?](#), West and Bergstrom  
“Uncertain Archives: Approaching the Unknowns, Errors, and Vulnerabilities of Big Data through Cultural Theories of the Archive,” Agostinho et al.

Th 2/1 [Facial Weaponization Suite](#), Blas  
“Recognizing Bias in Facial Recognition,” *More Than a Glitch*, Broussard

### Week 3: White Feelings, Algorithmic Racism

Tu 2/6 *Just Us: An American Conversation* (excerpts), Rankine  
[“A Simplified Political History of Big Data,”](#) [“Private Internet and Public Streets,”](#) [“Surveys, Vigilance and the Myth of Neutral Data,”](#) [“When White Fears Become Big Data,”](#) Dzodan

Th 2/8 “Introduction: The Power of Algorithms,” *Algorithms of Oppression: How Search Engines Reinforce Racism*, Noble

### Week 4: Policing and Predictive Analytics

Tu 2/13 [In the Robot Skies](#), Young and Maughan  
[2015](#), American Artist  
“Big Data Policing” and “Data is the New Black,” *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, Ferguson

Th 2/15 [“Predicting Financial Crime: Augmenting the Predictive Policing Arsenal”](#) and [White Collar Crime Risk Zones](#), Clifton et al.

### Week 5: Social Media and Surveillance Capitalism

Tu 2/20 [The Great Hack](#), Amer and Noujaim (also on [Netflix](#))  
“Home or Exile in the Digital Future,” *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Zuboff

Th 2/22 “The Story of a Generation in Seven Scams,” *Trick Mirror: Reflections on Self-Delusion*, Tolentino

### Week 6: Labor Rights as Human Rights: The Problem with Amazon.com, Inc.

Tu 2/27 *Seasonal Associate*, chapters 1-6, Geissler

Th 2/29 *Seasonal Associate*, chapters 7-11, Geissler

Fr 3/1 Paper 1 due by 11:59pm

### **Week 7: Making Feminist Data Visible**

- Tu 3/5 "The Default Male," *Invisible Women*, Criado Perez  
[Glitch Feminism](#) (excerpts), Russell
- Th 3/7 [Gender Shades](#), Buolamwini and Gebru  
"Why Data Science Needs Feminism," *Data Feminism*, D'Ignazio and Klein

**Spring Recess:** Tu 3/12 and Th 3/14 (no lectures or sections)

### **Week 8: Information is Beautiful: Data Meets Art**

- Tu 3/19 [The Library of Missing Datasets 2.0](#), Onuoha
- Th 3/21 [Data Art](#), *Nightingale* (explore at least 2 projects of your choice)

### **Week 9: Queering Data: LGBTQ Demographics and Trans Video Games**

- Tu 3/26 "Towards a Queer Futurity of Data," *Journal of Cultural Analytics*, Zeffiro  
[Dys4ia](#) (⚠ contains flash) and [Queers in Love at the End of the World](#), Anthropy
- Th 3/28 ["Data for Queer Lives,"](#) *Big Data & Society*, Ruberg and Ruelos

### **Week 10: Asian American Gender Politics and Robot Love Poems**

- Tu 4/2 "Algorithm Light," "Machine Testimonials," Interview; *Love, Robot*, Rhee  
"Turing Test" poems, *Soft Science*, Choi
- Th 4/4 "The Complexity of Otherness: Anthropological Contributions to Robots and AI,"  
*Oxford Handbook of AI Ethics*, Richardson
- Fr 4/5 Paper 2 due by 11:59pm

### **Week 11: Data at the Border: Latinx Migration and Cyborg Labor**

- Tu 4/9 *Sleep Dealer*, Rivera
- Th 4/11 ["How the AI Industry Profits from Catastrophe,"](#) Hao and Hernández  
Explore [Mapping Border Tech](#), Migration + Tech Monitor

### **Week 12: Indigenous Circuits: Data Sovereignty and Governance**

- Tu 4/16 ["Indigenous Circuits,"](#) *American Quarterly*, Nakamura  
Circuit Board Art, Martinez
- Th 4/18 "The CARE Principles for Indigenous Data Governance," Carroll et al.  
[The Quipu Project](#), Chaka Studio

### **Week 13: #Justice: Big Data for Social Good**

- Tu 4/23 "Retooling Solidarity, Reimagining Justice," *Race After Technology*, Benjamin

**Reading Period:** Th 4/25 and Tu 4/30 (no lectures or sections)

- W 5/8 Final Project due by 11:59pm

## ASSIGNMENT INSTRUCTIONS

### Note for Graduate Students

You are welcome to use this class as an opportunity to develop a project that is beneficial to your graduate studies, which would take the place of Papers 1 and 2 and the Final Project. For example, you might write a seminar paper, journal article draft, or portion of your Master's thesis or a dissertation chapter, or develop a more extensive media or data visualization project. If you choose this option, the final paper should be about 15-18 pages and meaningfully relate to one of our course themes. Please try to meet with me early in the semester (ideally between weeks 1-5) so we can discuss. I'm happy to provide the flexibility to accommodate your goals and schedule.

### Paper 1: Me vs. ChatGPT

**(4 to 5 pages double-spaced + ChatGPT transcript, due Friday 3/1 by 11:59pm)**

There are three required components of this assignment:

1. First, you will write an essay on one of the primary texts or media objects that we discussed during weeks 1-6. You may choose from the following: *Big Data*, *This Person Does Not Exist*, *Which Face is Real?*, *Facial Weaponization Suite*, *Just Us: An American Conversation*, *In the Robot Skies*, *2015*, "Predicting Financial Crime," *White Collar Crime Risk Zones*, *The Great Hack*, "The Story of a Generation in Seven Scams," or *Seasonal Associate*. This portion of the assignment is meant to give you an opportunity to develop your analytical skills by focusing closely on the features of one of these works. The goal is to move beyond simply summarizing what the work is *about* and instead consider *how* it is representing, critiquing, or altering our understanding of Big Data, race, power, society, gender, or injustice, etc. Your essay should therefore formulate a unique thesis statement that is specific, contestable, and analytical. You must also offer formal or stylistic evidence for your thesis throughout the essay, which should use details drawn directly from the work you are analyzing to support your overarching argument. You are not expected to use any theoretical sources and, in fact, you should not. Your essay should be roughly 3 pages in length.
2. Next, you will prompt ChatGPT to write an essay on the same topic you wrote your own paper on. Start by creating an account on [OpenAI](#). On the [ChatGPT page](#), you will see a text box that says "Message ChatGPT..." Try experimenting with different prompts that you pose to ChatGPT. For example, you might state: "Write an argument-driven essay focused on data in Claudia Rankine's *Just Us: An American Conversation*" or "Produce a five-paragraph essay discussing the implications of bias in predictive policing as seen in American Artist's *2015*." Usually, the more specific your prompt is, the better the results. Feel free to get creative! Don't worry if the essay that ChatGPT produces is only about 600 words. Remember to submit ChatGPT's essay—and the prompt you gave it as well—when you upload your own essay and reflection to Canvas.
3. Finally, you will write a brief reflection comparing your essay to the one produced by ChatGPT. Here are some questions you might consider: What kind of information or knowledge is ChatGPT's essay prioritizing? Does it have an arguable thesis with strong support? Are there any similarities between your own essay and ChatGPT's? Where are there significant differences? What do you feel is missing from ChatGPT's essay and where does it contain flaws? Does it effectively discuss issues related to race, power, and/or society with nuance and sensitivity? What do you think ChatGPT's analysis says about large language models or natural language processing and the role they might play in education, politics, social interactions, etc.? Your reflection essay should be roughly 1 to 2 pages in length (making the total amount of writing *you* will do for this assignment add up to about 4 to 5 pages).



## **Paper 2: The Library of Missing Datasets**

**(3 to 4 pages double-spaced or complete datasheet for datasets, due Friday 4/5 by 11:59pm)**

This assignment is inspired by [The Library of Missing Datasets 2.0](#) by Mimi Onuoha, who argues that “Black folks are both over-collected and under-represented in American datasets, featuring strongly as objects of collection but rarely as subjects with agency over collection, ownership, and power.” You will conceptualize a potential dataset that you might contribute to *The Library of Missing Datasets*. Where do you see “blank spots” in spaces that are otherwise data-saturated? What dataset does not exist, but should? What kinds of people, events, or social issues are both “over-collected” and “under-represented” in datasets? The primary goal of this assignment is for you to reflect on questions of responsible data collection with respect to the most vulnerable or under-resourced populations. Once you’ve theorized a missing dataset that you would contribute to the Library, you have two possible ways of completing the assignment:

### Option 1: Data Storytelling Essay (3 to 4 pages)

For this option, you will draw upon literary narrative techniques to tell a story about the life and possibilities of a missing dataset. Your essay can be entirely factual, autobiographical, or based in historical events, or it can be fantastical or inspired by the genre of science fiction. Why does this dataset not currently exist and what would be gained by bringing it into existence? What does this exclusion serve? Who does this exclusion serve? Who should create the dataset? Who is benefited by the existence of the dataset? What is the purpose? Who might be harmed? You might also discuss the conceptual, practical, and ethical challenges inherent to collecting this data. How can you ensure this dataset is free from abuse, unfairness, and bias? As you go about crafting your narrative, you could frame the absence of data as a mystery, prompting your reader to delve into the story by searching for clues behind the missing information. You can construct a story arc that mirrors traditional narratives—introduce the problem, build tension as you discuss the potential impact, and offer resolution or speculation on what could have been. Or you might incorporate descriptive language to breathe life into the missing dataset and paint a vivid picture of what it could reveal about our society or humanity. Feel free to go wherever your imagination takes you!

### Option 2: Datasheets for Datasets (fill out complete [template](#))

This template was created by [scientists](#) as a means of increasing transparency and accountability within the machine learning community. Datasheets document a dataset's motivation, composition, collection process, recommended uses, and so on. They have the potential to mitigate unwanted biases in machine learning systems, facilitate greater reproducibility of machine learning results, and help researchers and practitioners select more appropriate datasets for their chosen tasks. The questions are grouped into seven sections that roughly match the key stages of the dataset creation, maintenance, and distribution process. By grouping the questions in this way, dataset creators are encouraged to reflect on the process of creating, distributing, and maintaining datasets, and even to modify this process in response to that reflection.

## **Final Project: Research Paper or Data Visualization** **(Due Wednesday 5/8 by 11:59pm)**

Choose one option below. I strongly encourage you to set up a meeting with me or your TF to discuss your final project by April 26. This will give you an opportunity to brainstorm ideas, receive concrete feedback on a preliminary argument or data visualization, and work through any questions or concerns you might have related to conducting research or using digital tools.

### Option 1: Research Paper (7 to 8 pages, double-spaced)

For this option, you will identify a topic related to a course theme that interests you and write a paper that advances your own original argument on that topic and contributes in a substantial way to the existing debates around it. Now that we have considered many perspectives on Big Data, in this paper you will draw from these received notions and practice entering into current debates with your own carefully crafted argument. Whether you agree or disagree with, change the focus of, nuance, or expand on your chosen conversation, your particular argument must be clearly articulated and persuasive. Your paper should reference at least two scholarly sources about your topic (either from our syllabus or outside sources) and show how your own argument meaningfully engages with the perspective of other theorists. This paper is meant to be a culmination of the skills that you have been developing all semester. In addition to advancing your own unique thesis statement, offering close readings of literary texts, data visualizations, cultural works, historical events, and/or political case studies, and supporting your claims with evidence, you will now be integrating those skills with an analysis that aims to further develop, complicate, or enhance critical debate around your topic.

### Option 2: Data Visualization and Reflection Paper (2 to 3 pages, double-spaced)

For this option, you will create a data visualization that contributes meaningfully to the dialogue around one of our course themes. The goal of this assignment is for you to reflect on questions of responsible and ethical data visualization practices with respect to marginalized communities or underrepresented issues. Required components of this option include:

1. Data Visualization: For this component of the assignment, you will create a data visualization that relates to or comments on a contemporary issue concerning race, gender, sexuality, labor, or other power struggles such as predictive policing, racist border regimes, surveillance, capitalism, mass incarceration, refugee rights, gender inequality in the workplace, Indigenous dispossession, etc. Later in the semester, we will discuss datasets and potential digital tools you might use to develop your visualization. Though as the Nightingale [Data Art](#) projects illustrate, data visualizations can take many different forms and involve such media as paper, paint, fabric, yarn, etc. In other words, you do not need to be a data or computer scientist to choose this option!

2. Reflection on your Data Visualization: Why did you choose to create this data visualization? How do you hope it intervenes in the conversation around a particular issue related to race, gender, power, social justice, etc.? What conceptual or practical challenges did you face when developing your visualization? What do you hope audiences will ultimately take away from interacting with your data? For this component of the assignment, you must draw upon the ideas of at least one scholarly source (either from our syllabus or an outside source) and show how your visualization engages with the perspective of another theorist.