

A DATA FEMINISM NETWORK &  
TORONTO WOMXN IN DATA SCIENCE  
COLLABORATION

# Data Feminism Study Guide

Introduction and Chapter 1-2

---

## Meet the Authors



**Catherine D'Ignazio**

 @kanarinka



**Lauren F. Klein**

 @laurenfklein

---

## Seven Principles of Data Feminism:

*Examine Power*

*Challenge Power*

*Elevate Emotion and Embodiment*

*Rethink Binaries and Hierarchies*

*Embrace Pluralism*

*Consider Context*

*Make Labor Visible*

# Introduction

## Key Definitions

- **Feminism** → a shorthand for the diverse and wide-ranging projects that name and challenge sexism and other forces of oppression, as well as those which seek to create more just, equitable, and livable futures. (6)
- **Intersectionality** → the idea of intersectionality describes not only the intersecting aspects of any particular person's identity, but it also describes the intersecting forces of privilege and oppression at work in a given society. (7)
  - The term was coined by legal theorist Kimberlé Crenshaw in the late 1980s
- **Oppression** → the systematic mistreatment of certain groups of people by other groups which happens when power is not distributed equally. (8)
- **Data Feminism** → a way of thinking about data, both their uses and their limits, that is informed by direct experience, by a commitment to action, and by intersectional feminist thought. (8)
- **Co-liberation** → the idea that oppressive systems of power harm all of us, that they undermine the quality and validity of our work, and that they hinder us from creating true and lasting social impact with data science. (9)
- **Datafication** → a technological trend turning many aspects of our life into data which is subsequently transferred into information and monetized. (12)

## Key Themes

### ***DATA FEMINISM IS FOR EVERYBODY***

- *Data feminism isn't only about women:* It takes more than one gender to have gender inequality and more than one gender to work toward justice. (14)
- *Data feminism isn't only for women:* Men, non-binary, and genderqueer people are proud to call themselves feminists and use feminist thought in their work. (14)
- *Data feminism isn't only about gender:* Intersectional feminists have keyed us into how race, class, sexuality, ability, age, religion, geography, and more, are factors that together influence each person's experience and opportunities in the world. (14)
- *Data feminism is about power - about who has it and who doesn't:* Intersectional feminism examines power. And in our contemporary world, data is power too. Because the power of data is wielded unjustly, it must be challenged and changed. (14)

---

# Data Feminism in Action

## **CHRISTINE MANN DARDEN, DATA ANALYST AT NASA**

Darden was faced with the fact that in her field, men with math credentials were promoted while women with the same credentials were sent to computing pools where they would languish until they quit or retired. She faced intersecting oppressions as a Black woman in a white male dominated sector. Darden, however, championed for herself and went on to be the first Black woman to advance to the top rank in the federal civil service. Her story reflects several decades of feminist activism and critical thought. You may know her story from the movie "Hidden Figures". (8)

→ <https://www.quantamagazine.org/the-nasa-engineer-whos-a-mathematician-at-heart-20210119/>

## **YESHIMABEIT MILNER, FOUNDER OF DATA FOR BLACK LIVES (D4BL)**

This organization is dedicated to "using data science to create concrete and measurable change in the lives of Black Communities." D4bl uses explicitly feminist and antiracist methods to quantify and challenge invasive data collection by law enforcement. (13)

→ <https://d4bl.org/>

## **PREDPOL MODEL, EXAMINED BY CATHY O'NEIL IN *WEAPONS OF MATH DESTRUCTION***

Predpol is a predictive policing company founded in California that is used to determine which neighborhoods to patrol more heavily and which neighborhoods to largely ignore. This model is based on historical crime data and reflects policies that disproportionately surveilled and monitored neighborhoods of color. This is one example of what O'Neil describes as a pernicious feedback loop which amplifies the effect of racial bias and of the criminalization of poverty that are already endemic to the U.S. (13)

→ <https://ideas.ted.com/justice-in-the-age-of-big-data/>



# Chapter 1 - The Power Chapter

Principle: **Examine Power**

→ Data feminism begins by analyzing how power operates in the world.

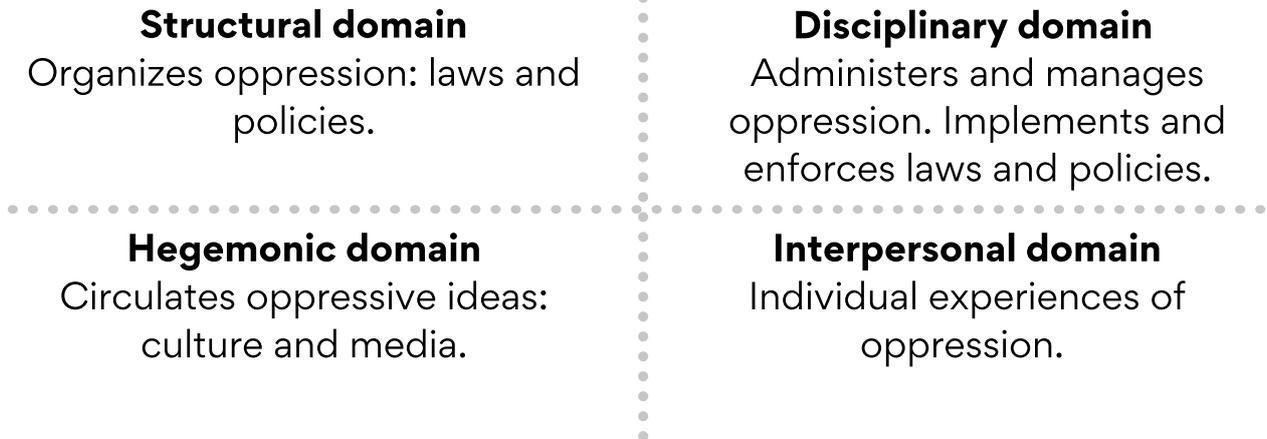
## Key Definitions:

- **Examining power** → naming and explaining the forces of oppression that are so baked into our daily lives - and into our datasets, our databases, and our algorithms - that we often don't even see them. (24)
- **Power** → the current configuration of structural privilege and structural oppression, in which some groups experience unearned advantages - because various systems have been designed by people like them and work for people like them - and other groups that experience systematic disadvantages - because those same systems were not designed by them or with people like them in mind. (24)
- **Minoritized** → describes groups of people who are positioned in opposition to a more powerful social group. (26)
- **Privilege hazard** → the phenomenon that makes those who occupy the most privileged positions among us - those with good educations, respected credentials, professional accolades - so poorly equipped to recognize instances of oppression in the world. Their lack of lived experience profoundly limits their ability to foresee and prevent harm, to identify existing problems in the world, and to imagine possible solutions. (29)
- **Scarcity bias** → the idea that there are not enough resources for everyone so we should think small and allow technology to fill the gaps. (40)

## Key Themes

- *Data science by whom?* (26)
  - Who is doing the work of data science (and who is not)?
- *Data science for whom?* (33)
  - Who benefits from data science (and who is either overlooked or actively harmed?)
- *Data science with whose interests and goals?* (39)
  - Whose goals are prioritized in data science (and whose are not)?

# The Four Domains of the Matrix of Domination



*Power and the Matrix of Domination* → The matrix of domination works to uphold the undue privilege of dominant groups while unfairly oppressing minoritized groups. (25)

## Data Feminism in Action

### TENNIS STAR SERENA WILLIAMS' PREGNANCY COMPLICATIONS

Williams faced life threatening pregnancy complications and her self-advocacy played a crucial role in her survival. Her doctors did not believe her when she said she was in pain and thought something was wrong. Through social media, Williams found many other women faced similar experiences. She posted that "Black women are over 3 times more likely than white women to die from pregnancy or childbirth related causes". The privilege she experiences as a tennis star intersected with the oppression she experienced as a Black woman, enabling her to avoid becoming a statistic herself. As Williams asserted, "that's not fair". (23-24)

→ These disparities are well known to Black-women led reproductive justice groups like Sister Song (<https://www.sistersong.net/>), Black Mamas Matter Alliance (<https://blackmamasmatter.org/>), and Raising Our Sisters Everywhere (<http://rose.kellimcnair.com/>)



## **JOY BUOLAMWINI, A GHANIAN-AMERICAN GRAD STUDENT AT MIT, FACIAL ANALYSIS SOFTWARE PROJECT**

The software Boulamwini was working with couldn't see her dark skinned face even though it had no problem seeing her lighter-skinned collaborators. Once she put on a white mask, the software recognized her. This story exemplifies the privilege hazard, as only 4% of faces in the data set for this software were women and dark skinned thus the problem was simply not identified by the coders. (30)

→ <https://www.ajl.org/>

## **MIMI ONUOHA, ARTIST, DESIGNER, AND EDUCATOR, AND HER PROJECT, *THE LIBRARY OF MISSING DATASETS***

A list of datasets one might expect to already exist in the world, because they help to address pressing social issues, but have never actually been created. (33)

→ <https://mimionuoha.com/the-library-of-missing-datasets>

## **MARÍA SALGUERO'S COMPREHENSIVE DATA SET ON FEMICIDES IN MEXICO - GENDER RELATED KILLINGS OF WOMEN AND GIRLS**

She has logged more than 5,000 femicides since 2016. Her work provides the most accessible information on the subject for journalists, activists, and victim's families seeking justice. (35)

→ <https://iqlatino.org/2019/one-only-woman-is-mapping-the-femicides-of-all-of-mexico-her-name-is-maria-salguero/>

## **SEALS ALLERS AND HER SON, MICHAEL'S APP, *IRTH***

Irth stands for birth, but with the b for bias removed. This app is a data driven contribution to the maternal and infant health conversation. One of the major contributing factors to poor birth outcomes, as well as maternal and infant mortality, is biased care. Hospitals, clinics, and caregivers routinely disregard Black women's expressions of pain and wishes for treatment. (46)

→ <https://irthapp.com/>

# Chapter 2 - Collect, Analyze, Imagine, Teach

*Principle: Challenge Power*

→ Data feminism commits to challenging unequal power structures and working toward justice

## Key Definitions:

- **Redlining** → a term used to describe how banks rated the risk of granting loans to potential homeowners on the basis of neighborhood demographics (specifically race and ethnicity), rather than individual credit worthiness. (51)
  - Redlining initially entailed the literal drawing of red lines on a map
  - Redlining maps were very similar to “big data” approaches today where it is ensured that wealth remain attached to the racial category of whiteness. (52)
- **Racial capitalism** → coined by philosopher Cedric Robinson, describes the implementation of data driven methods for granting (or denying) policies to customers based on their demographics.
- **Challenging power** → requires mobilizing data science to push back against existing and unequal power structures and to work toward more just and equitable futures. (53)
- **New Jim Code** → coined by sociologist Ruha Benjamin, where software code and a false sense of objectivity come together to contain and control the lives of Black people and other people of color. (55)
  - The danger of the New Jim Code is that these findings are actively promoted as objective, and they track individuals and groups through their lives and limit their future potential. (55)
- **Auditing algorithms** → using data science to challenge data science
  - Collecting missing data and reverse-engineering the algorithm to prove systemic bias. (57)

- **Deficit narratives** → reduce a group or culture to its “problems,” rather than portraying it with the strengths, creativity, and agency that people from those cultures possess. (58)
  - While it is important to collect counterdata and analyze data to provide proof of oppression, it is equally important to remain aware of how the subjects of oppression are portrayed. (59)
  - Key too is a sustained attention to the ways in which communities themselves are already addressing the issues. (59)
- **Imagined objectivity** → coined by Cedric Benjamin in *Race After Technology: Abolitionist Tools for the New Jim Code*, emphasizes the role that cultural assumptions and personal perceptions play in upholding this false belief: one imagines (wrongly) that data sets and algorithms are less partial and less discriminatory than people and thus more “objective”. (59-60)
- **Data ethics** → growing interdisciplinary effort- both critical and computational- to ensure that the ethical issues brought about by our increasing reliance on data-driven systems are identified and addressed. (60)
- **Equity** → equality assumes everyone at the same starting point and can actually be systematically unfair (for example in the U.S. where Black babies are dying at twice the rate of white babies). Working toward a world in which everyone is treated equitably, not equally, means taking into account these different power differentials and distributing (or redistributing) resources accordingly. (62)
- **Co-liberation** → requires a commitment to and a belief in mutual benefit, from members of both dominant groups and minoritized groups.
  - Needs not only transparency but reflexivity: the ability to reflect on and take responsibility for one’s own position within the multiple, intersecting dimensions of the matrix of domination. (64)

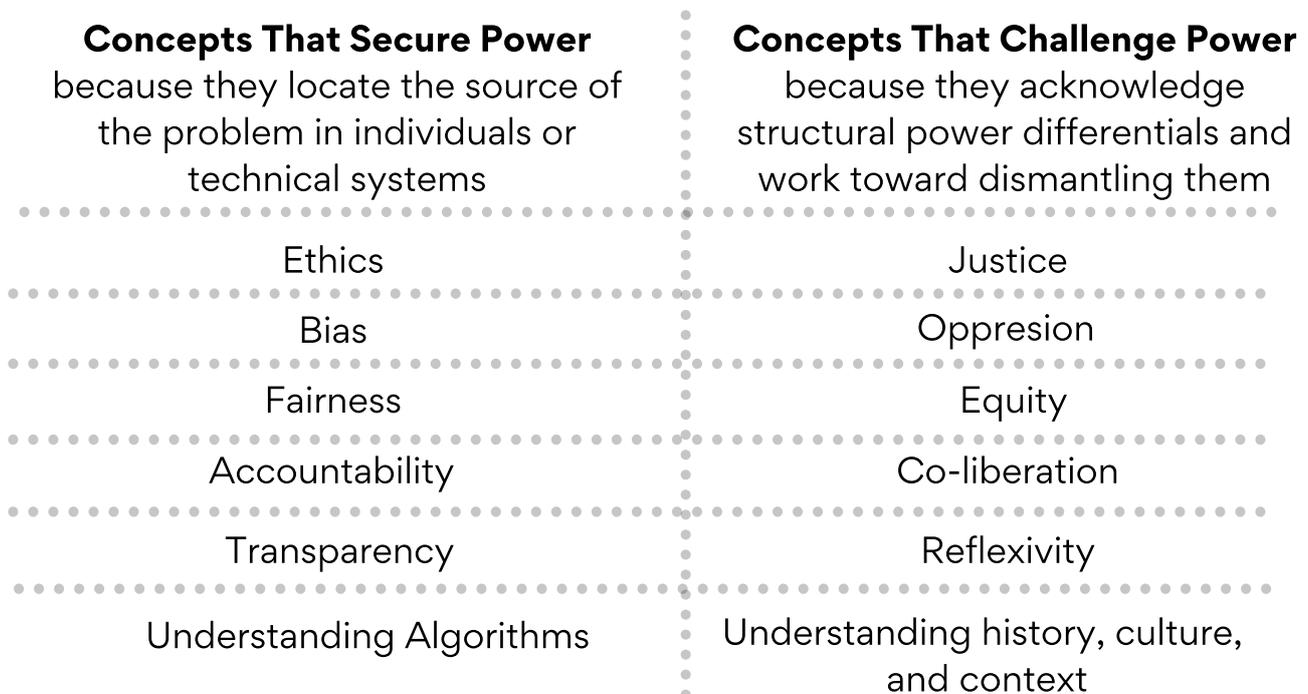


# Key Themes

## Four Ways to Challenge Power (53)

- *Collect* → Compiling counterdata - in the face of missing data or institutional neglect.
- *Analyze* → Challenging power often requires demonstrating inequitable outcomes across groups, and new computational methods are being developed to audit opaque algorithms and hold institutions accountable.
- *Imagine* → We cannot only focus on inequitable outcomes, because then we will never get to the root cause of injustice. In order to truly dismantle power, we have to imagine our end point not as “fairness” but as co-liberation.
- *Teach* → The identities of data scientists matter, so how might we engage and empower newcomers to the field in order to shift the demographics and cultivate the next generation of data feminists?

## From Data Ethics to Data Justice



---

# Data Feminism in Action

## **DETROIT GEOGRAPHIC EXPEDITION AND INSTITUTE (DGEI) MAP, WHERE COMMUTERS RUN OVER BLACK CHILDREN ON THE POINTES-DOWNTOWN TRACK**

The map uses sharp black dots to illustrate the places in the community where the children are killed. Those who lived along this route were aware of the problem and its profound impact on friends and neighbors, but gathering data turned out to be a real challenge. No one kept detailed records of the deaths and no one was making basic information of what happened publicly available. In lieu of missing data, DGEI compiled counter-data to ensure the problem be properly addressed. (52)

→ <https://medium.com/nightingale/gwendolyn-warren-and-the-detroit-geographic-expedition-and-institute-df9ee10e6ad2>

## **PROPUBLICA TEAM LED BY JULIA ANGWIN INVESTIGATES A WIDELY USED CRIMINAL BAIL ASSESSMENT ALGORITHM**

Her team found that white defendants are more often mislabeled as low risk than Black defendants and, conversely, Black defendants are mislabeled as high risk more often than white defendants. This project successfully proved systemic racial bias. (53)

→ <https://www.propublica.org/article/how-we-analyzed-the-compass-recidivism-algorithm>

## **OUR DATA BODIES (ODB) PROJECT**

Led by a group of five women, this project is a community centered initiative focused on data collection efforts that disproportionately impact minoritized people. They recently released Digital Defence Playbook, which is a set of activities, tools, and tip sheets intended to be used by and for marginalized communities to understand how data driven technologies impact their lives. (64)

→ <https://www.odbproject.org/tools/>

## **LOCAL LOTTO PROJECT**

Taught local high school students statistics and data analysis rooted in neighborhood and justice concerns. (68)

→ [http://citydigits.org/utm\\_content=buffer0f105&utm\\_medium=social&utm\\_source=linkedin.com&utm\\_campaign=buffer](http://citydigits.org/utm_content=buffer0f105&utm_medium=social&utm_source=linkedin.com&utm_campaign=buffer)