



Review

Book Review: Data Feminism

Katherine A. Mika

Harvard University, Cambridge, MA, USA

Data Feminism

Catherine D'Ignazio and Lauren F. Klein The MIT Press (2020)

Data Feminism combines intersectional feminism and critical data studies to invite the reader to consider: "How can we use data to remake the world?" As non-profit organizations with a mandate to provide equitable access to non-neutral information and services, libraries and library workers are uniquely positioned to advance the principles laid out in *Data Feminism*.

Correspondence: Katie Mika: katherine_mika@harvard.edu

Received: February 2, 2021 **Accepted:** April 27, 2020 **Published:** May 6, 2021

Copyright: © 2021 Mika. This is an open access article licensed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/).

Disclosures: The author reports no conflict of interest.

In late 2020 more than 400 Google and Alphabet¹ workers organized to form a union in order to improve labor conditions; hold the company accountable for harassment, abuse, discrimination, and retaliation; and defend freedom for workers to decline work on ethically questionable projects (Conger 2021). These tenets of the Alphabet Workers Union (AWU)² exemplify the principles Catherine D’Ignazio and Lauren F. Klein describe in their recent book, *Data Feminism*.

Data Feminism combines intersectional feminism and critical data studies to invite the reader to consider: “How can we use data to remake the world?” (D’Ignazio and Klein 2020, 5). As non-profit organizations with a mandate to provide equitable access to non-neutral information and services, libraries and library workers, referred to collectively as ‘librarians,’ are uniquely positioned to teach, support, and inspire researchers and members of the public to think critically about their complicity in injustice and exploitation.

Data Feminism frames its central contribution as a set of applied tools designed to untangle some of the big problems in modern society. The authors illustrate seven principles (across seven chapters) to confront inequality and challenge the distribution of power:

- Examine power
- Challenge power
- Elevate emotion and embodiment
- Rethink binaries and hierarchies
- Embrace pluralism
- Consider context
- Make labor visible

The first two chapters introduce how powerful individuals’ and organizations’ data science applications often aggravate inequality and compound injustice and oppression. Here, the principles complement a growing body of literature on algorithmic bias and the inequities of data science. Assessment algorithms, databases, mapping, and modeling will continue to perpetuate and reinforce oppression because each are products of structurally unequal conditions. In order to identify and address power discrepancies, we must ask who benefits from (and whose goals are prioritized by) a given data project or product (D’Ignazio and Klein 2020, 26). For example, Google recently forced out a well-known Artificial Intelligence (AI) researcher, Timnit Gebru—whose work with Joy Buolamwini on facial recognition is profiled in *Data Feminism*—allegedly in retaliation for writing a paper exposing significant risks in training very large language models, a

1 Alphabet is the parent company of Google

2 <https://alphabetworkersunion.org>

foundation of Google's business model (Hao 2020). Established or creating labor unions have the potential to offset the risk of challenging powerful organizations that perpetuate injustice and could become an important tool of data feminists.

The next three principles: elevate emotion, rethink binaries, and embrace pluralism relate to a more holistic understanding of the diversity of experiences that good data science hopes to decipher and good datasets should aim to capture. These chapters build from the flaws identified in the first part of the book. For example, if a major problem with assessment algorithms is that aggregated information is used to make predictions about unique individuals, adding essential context to and clarifying the limitations of inputs and training datasets may provide essential in-depth information for decision makers. Knowledge is created by synthesizing multiple perspectives and sources of information. Including and giving priority to local, Indigenous, and experiential ways of knowing (D'Ignazio and Klein 2020, 125; Cochran et al. 2008) can help dominant cultural groups, those who are usually in power, be aware and cautious of inflicting and/or perpetuating epistemic violence.

The chapter on considering context is a favorite because it collects elements of the other principles into a simple shibboleth: *Data are not neutral*. Power, emotion, plurality, and labor are vital context that affect datasets and data projects. Emphasizing context allows researchers to better understand the functional limitations, associated ethical obligations, and the role that power and privilege play in obscuring the truth and preserving structural inequities in a dataset or data project (D'Ignazio and Klein 2020, 153). Here, missing data becomes less of a problem to clean up, but an opportunity to look for individuals who may likely be members of less-visible and less powerful groups (Taylor 2020). Without important context and theory to effectively apply context, data scientists and information consumers rely on their common sense and intuition to interpret results, which easily leads to bias.

The final chapter on making labor visible describes strategies for illuminating the invisible work required by many people to produce successful data science projects. However, visibility is often insufficient to generate value and appropriate recognition. Volunteer labor, micro-payments, and precarious employment practices are the backbone of academia and the technology industry, including galleries, libraries, archives, and museums (GLAMs) (Digital Library Federation Working Group on Labor in Digital Libraries 2018; Mayer 2016; Brons et al. 2020). Not only are these types of undercompensated and unprotected labor practices exploitative, they also reinforce discrimination and further entrench unequal power relations (D'Ignazio and Klein 2020, 200). As illustrated by AWU, collective action that pushes back against the forces of power that create injustice and inequality can alter the structure of society to support and promote activities that encourage co-liberation and use data to remake the world.

Data Feminism is an exceptional work of data activism. The authors connect important critiques of modern data science to concrete examples that demonstrate

how to create better, more equitable research. However, the authors are less successful at delivering strategies for addressing the systemic problems that incentivize oppression and unequal power relationships that penetrate all aspects of society. For example, in an example featuring the New York Times election needle, the discussion centers around the difficulty in visualizing uncertainty (D'Ignazio and Klein 2020, 90-91). But, the success of the election needle is not defined by how effectively it communicates accurate information, it is measured by how successful it is at driving traffic. The anxiety it produced made the New York Times more money, which is why it continues to pop up every election cycle. Data feminism is good praxis, but capitalism, by valuing profit over human risk and cost, is a major barrier that largely goes untreated.

Strengthening public institutions, like libraries, is one way to build social infrastructure that can support the work of reshaping the data economy. Libraries and librarians sit at the intersection of instruction, activism, and public service, and as such are well positioned to coordinate data feminist activities across domains. Indeed, many of us are already engaging in elements of data feminism in our day-to-day jobs. For example, data librarians offer consultation and instruction to improve and ensure that data assets are properly stewarded. Support technologists, technical services, metadata, collection development, copyright, and digital collections librarians create and maintain the infrastructure for accessing non-neutral data and information sources. Metadata and cataloging librarians interrogate classification systems to better describe, categorize, find, and access information. Libraries host trainings and provide daily reference on how to use computers, and search for and crucially evaluate the information returned. Digital collection and visualization specialists provide consultation on and computational access to collection data and metadata for distant reading, machine learning, and other, algorithmically-based research methods. And there are still so many more examples that I have left out. All of these responsibilities and activities are locations in which librarians can apply the principles of data feminism. Library work is fundamentally, if imperfect, care work that emphasizes people and their information needs. Librarians care for collections and communities in order to contribute to the public good. Wouldn't it be nice if data science had the same mandate?

References

Brons, Adena, Chloe Riley, Ean Henninger, and Crystal Yin. 2018. "Precarity in Libraries." April. <https://osf.io/ybczx>

Cochran, Patricia A. L., Catherine A. Marshall, Carmen Garcia-Downing, Elizabeth Kendall, Doris Cook, Laurie McCubbin, and Reva Mariah S. Gover. 2008. "Indigenous Ways of Knowing: Implications for Participatory Research and Community." *American Journal of Public Health* 98(1): 22-27. <https://doi.org/10.2105/AJPH.2006.093641>

Conger, Kate. 2021. "Hundreds of Google Employees Unionize, Culminating Years of Activism." *New York Times*[online] 4. <https://www.nytimes.com/2021/01/04/technology/google-employees-union.html>

Digital Library Federation Working Group on Labor in Digital Libraries. 2018. "Research Agenda: Valuing Labor in Digital Libraries." DLF Wiki.
<https://wiki.diglib.org/Labor/Valuing-Labor/Research-Agenda>

D'Ignazio, Catherine, and Lauren F. Klein. 2020. *Data Feminism*. Cambridge, MA: MIT Press.
<https://mitpress.mit.edu/books/data-feminism>

Hao, Karen. 2020. "We Read the Paper That Forced Timnit Gebru out of Google. Here's What It Says." *MIT Technology Review* December 4, 2020.
<https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru>

Mayer, Allana. 2016. "Crowdsourcing, Open Data and Precarious Labour." *Model View Culture* 33.
<https://modelviewculture.com/pieces/crowdsourcing-open-data-and-precarious-labour>

Taylor, Linnet. 2020. "The Price of Certainty: How the Politics of Pandemic Data Demand an Ethics of Care." *Big Data & Society* 7(2): 2053951720942539. <https://doi.org/10.1177/2053951720942539>