Brandeis Hill Marshall

DATA CONSCIENCE

ALGORITHMIC S1EGE ON OUR HUM4N1TY

Foreword by Timnit Gebru

Founder and Executive Director, Distributed Artificial Intelligence Research Institute

WILEY

Table of Contents

- Cover
- Title Page
- Foreword
- Introduction
- Part I: Transparency
  - Note
  - CHAPTER 1: Oppression By...
    - The Law
    - The Science
    - Summary
    - Notes
    - Recommended Reading
  - CHAPTER 2: Morality
    - Data Is All Around Us
    - Morality and Technology
    - Misconceptions of Data Ethics
    - Limits of Tech and Data Ethics
    - Summary
    - Notes
  - CHAPTER 3: Bias
    - Types of Bias
    - Before You Code
    - Bias Messaging
    - Summary
    - Notes

- CHAPTER 4: Computational Thinking in Practice
  - Ready to Code
  - Algorithmic Justice Practice
  - Code Cloning
  - Summary
  - Notes

Part II: Accountability

- Note
- CHAPTER 5: Messy Gathering Grove
  - Ask the Why Question
  - Collection
  - Reformat
  - Summary
  - Notes
- CHAPTER 6: Inconsistent Storage Sanctuary
  - Ask the “What” Question
  - Files, Sheets, and the Cloud
  - Modeling Content Associations
  - Manipulating with SQL
  - Summary
  - Notes
- CHAPTER 7: Circus of Misguided Analysis
  - Ask the “How” Question
  - Misevaluating the “Cleaned” Dataset
  - Overautomating k, K, and Thresholds
  - Not Estimating Algorithmic Risk at Scale
  - Summary
  - Notes

- CHAPTER 8: Double-Edged Visualization Sword
  - Ask the “When” Question
  - Critiquing Visual Construction
  - Pretty Picture Mirage
  - Summary
  - Notes
- Part III: Governance
  - Note
  - CHAPTER 9: By the Law
    - Federal and State Legislation
    - International and Transatlantic Legislation
    - Regulating the Tech Sector
    - Summary
    - Notes
  - CHAPTER 10: By Algorithmic Influencers
    - Group (Re)Think
    - Flyaway Fairness
    - Moderation Modes
    - Summary
    - Notes
  - CHAPTER 11: By the Public
    - Freeing the Underestimated
    - Learning Data Civics
    - Condemning the Original Stain
    - Tech Safety in Numbers
    - Summary
    - Notes
- APPENDIX A: Code for *app.py*

- A
- B
- C
- D

APPENDIX B: Code for *screen.py*
- A
- B
- C

APPENDIX C: Code for *search.py*
- A
- B
- C
- D

APPENDIX D: Pseudocode for *faceit.py*

APPENDIX E: The Data Visualisation Catalogue's Visualization Types

APPENDIX F: Glossary

Index

Copyright

Dedication

About the Author

About the Technical Editor

Acknowledgments

End User License Agreement

List of Tables

Chapter 1

- Table 1.1: Egyptian numeral system

Chapter 2

Table 2.1: Similarities between tech ethics and human ethics

Table 2.2: Similarities among data, tech, and human ethics

Chapter 3

Table 3.1: Data ethics principles and its motivating questions

Table 3.2: Synthetic (fake) employee applicant data from résumés

Table 3.3: A skills listing in order of greater to lesser experience

Table 3.4: A skills listing with no further indicators of proficiency (by de...

Chapter 4

Table 4.1: Word frequency count for *Computational Thinking*

Chapter 5

Table 5.1: Data sourcing questions

Table 5.2: Handling missing or unknown data

Chapter 6

Table 6.1: Data manipulation questions, part 1

Table 6.2: Synthetic (fake) résumé data from Chapter 3

Table 6.3: Hashtags, Blacktags, and keywords for the Black Twitter Project

Chapter 7

Table 7.1: Data manipulation questions

Table 7.2: Output after `fullMovies.describe()` code execution

Chapter 8

Table 8.1: Data interpretation questions

Chapter 10

Table 10.1: Rocchio's algorithm explained (simply?)

Chapter 11

Table 11.1: Data civics stages

Table 11.2: Actual data industry representation by ethnicity/gender demograp...

Table 11.3: U.S. population representation by ethnicity/gender demographic

Table 11.4: Troll subversion recommendations by social platforms

List of Illustrations

Chapter 1

Figure 1.1: A quoted tweet referencing the ‘Negro Women To Be Put To Work’ a...

Figure 1.2: (a) Single image of face from Turk and Pentland's 1991 facial re...

Figure 1.3: Mugshot of Alphonse Bertillon (1853–1914) with a reference card ...

Chapter 2

Figure 2.1: ACM 2020 Computer Science Curriculum Guide's recommendation on s...

Figure 2.2: ACM 2020 Computer Science Curriculum Guide's prioritization of s...

Figure 2.3: Examples of high-resolution images that are downscaled to low-re...

Chapter 3

Figure 3.1: Timeline of Amazon's AI hiring software product

Figure 3.2: Summary of Margaret Mitchell's “Bias in the Vision and Language ...

Figure 3.3: Parts of a Car Wheel

Figure 3.4: Illustration of how a wheel infrastructure transfers impact

Figure 3.5: Four potential bias wheels covering business, human, data and al...

Figure 3.6: Sample résumés returned from a Google image search

Figure 3.7: Snapshot of online résumé entity-relationship diagram

Chapter 4

Figure 4.1: Concept map of how I wash my locs

Figure 4.2: Sampling of computing services and platforms costs by max data c...

Figure 4.3: ML-based résumé screening system on GitHub

Figure 4.4: Skeleton structure of ML-based résumé screening system.

Chapter 5

Figure 5.1: Example of website cookie acceptance request (top) and cookie se...

Figure 5.2: Sample Google engine result of phrase “open source dataset”

Figure 5.3: PeopleFinders results for Brandeis Hill Marshall

Figure 5.4: Snippet of `Data.gov` dataset (uncleaned)

Figure 5.5: Snippet of retooled `Data.gov`'s dataset (cleaned)

Figure 5.6: Snippet of Twitter data scraped using Tweepy API

Figure 5.7: Snippet of streamlined Twitter code in CSV format

Chapter 6

Figure 6.1: Snapshot of online résumé ERD

Figure 6.2: Twitter data usage for research purposes (June 2015)

Figure 6.3: Applicant demographic identities

Figure 6.4: (Left) A DBA's view and (right) an applicant's view of demograph...

Figure 6.5: An expanded ERD representation

Figure 6.6: General syntax of the `SELECT` statement used in DML

Figure 6.7: SQL statement processing order

Chapter 7

Figure 7.1: AI, ML, and DL quick explanations

Figure 7.2: Descriptive statistics of ratings column from Kaggle

Figure 7.3: Linear regression of UserID and MovieID

Figure 7.4: Linear regression of Rating and MovieID

Figure 7.5: Linear regression of Timestamp and MovieID

Figure 7.6: Ordinary least squares outputs using Timestamp and MovieID

Figure 7.7: Translucent Brandeis and Visible Brandeis using Zoom Background ...

Figure 7.8: Examples of deepfake technology

Figure 7.9: Deepfake of Jordan Peele as former U.S. President Barack Obama...

Figure 7.10: Princess Leia remastered using deepfake technology in the film ...

Figure 7.11: Deep Nostalgia animates the faces in your family photos with am...

Chapter 8

Figure 8.1: Content consumes our lives.

Figure 8.2: Gaping Void's representation of data, information, knowledge, in...

Figure 8.3: My January 5, 2022 Twitter post

Figure 8.4: A choropleth map of the United States indicating the states with...

Chapter 9

Figure 9.1: State data privacy legislation status (as of March 2022)

Figure 9.2: Black population distribution according to U.S. Census 2020

Figure 9.3: Data, some of the industries it affects, and the AI-based tools ...

Chapter 10

Figure 10.1: False narration of the Black tech pipeline problem

Figure 10.2: How content moderation works

Figure 10.3: Original tweet post by CNN's Rick Santorum saying that “We birt...

Figure 10.4: Twitter response to alleged community code violation: “For thos...

Chapter 11

Figure 11.1: Data civics Stage 1: people's relationship to data

Figure 11.2: The data life cycle and surrounding data ecosystem

Figure 11.3: Profit Without Oppression's Guiding Principles, by Kim Crayton...

Figure 11.4: Adapted from “The Chronicle of the Problem Woman of Colour in a...

Appendix E

Figure E.1: The following data visualizations are displayed: error bars, flo...

Figure E.2: The following data visualizations are displayed: pie chart, poin...

Figure E.3: The following data visualizations are displayed: arc diagram, ar...

Figure E.4: The following data visualizations are displayed: timetable, tree...

Data Conscience

Algorithmic Siege on Our Humanity

Brandeis Hill Marshall

WILEY

Foreword

There is an assumption that we need to code switch, straighten our hair, get rid of that accent, get rid of those big earrings, not be too much of a Black woman, but instead assume an identity that is considered “the norm” in computer science spaces in the United States. This undertone exists in every single text book containing code snippets that I have read. There is a pretense that these books were written from no one's point of view, exuding an air of neutrality that we know doesn't exist, what feminist scholars have dubbed “the view from nowhere.”

This tone is that of a cishet white man, seen as the norm while everyone else is considered a deviation. Dr. Marshall shatters that expectation. With DataedX, Black women in Data, the Rebel Tech newsletter, her many educational events and this book, Data Conscience, Dr. Marshall brings her full self to work and encourages us to do the same. Reading this book feels like we're having an honest conversation about data and code over our favorite food or drink.

Exclusionary views of who is considered a data expert, are one of the many ways that Black people are gate kept from a field that profits off of our datafication and criminalization. As Prof. Ruha Benjamin, sociologist and associate professor of African American Studies at Princeton University, notes, technology allows racism to enter through the backdoor: the false assumption of neutrality masks ingrained, discriminatory, systems, while at the same time we are told, implicitly and explicitly, that we do not fit into the computer science archetype.

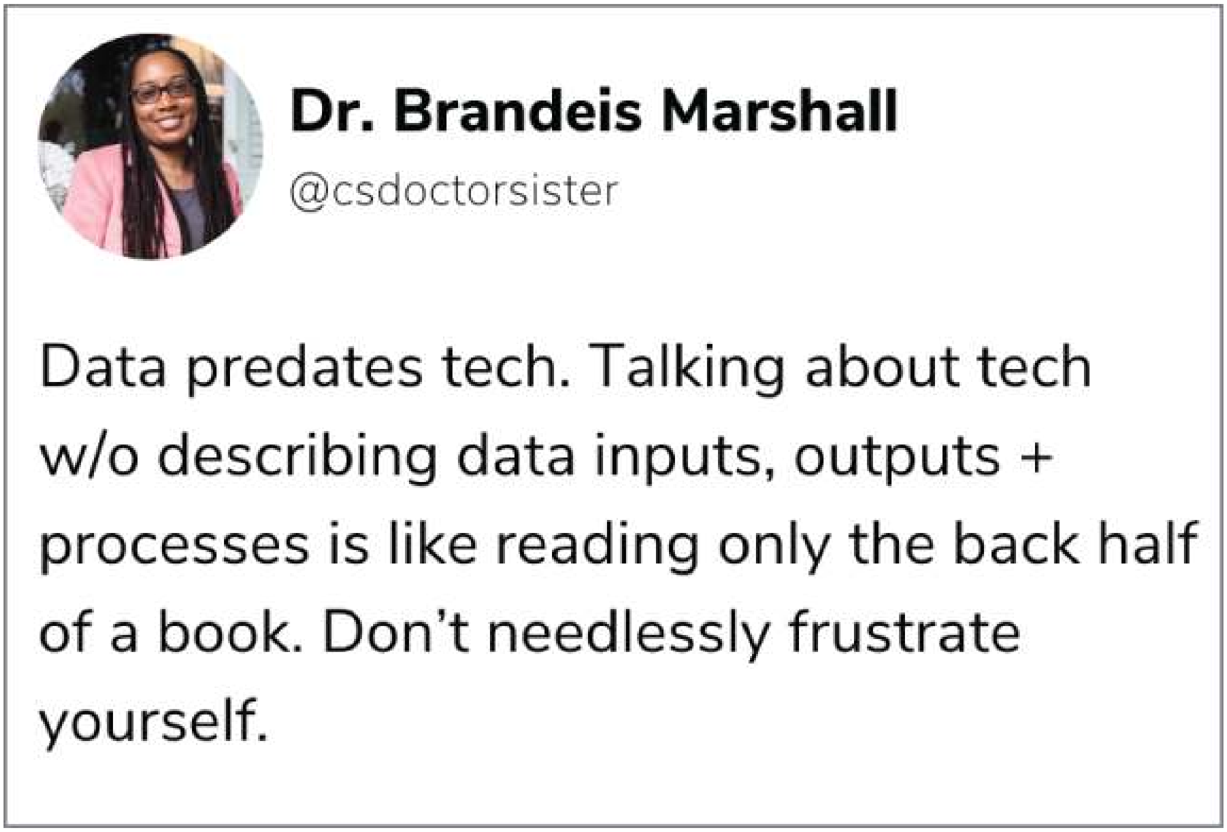
This book tells us that we do fit, and helps us develop toolkits for responsible computing. It goes through all aspects of the data lifecycle including the processes of gathering, labeling, cleaning, storing, and governing data, explaining how each of these steps affects us. Dr. Marshall moves us away from the notion that everyone should be datafied and every problem should be solved by technology. Instead, she explains the complex societal layers that we lose when we attempt to convert everything into discrete data points. She asks us to consider, and resist, the many ways in which we are being taught to act like machines, and gives us tangible tips for how to actively push back to retain our humanness and Black identities in this world of data.

For those of us looking to hone our data skills, this book helps us do that with examples of what could go wrong, and how to avoid those paths. Contrary to textbooks that abstract out one aspect of the data lifecycle without any societal context, Dr. Marshall reminds us that “coding requires a 360 degrees of panoramic view that requires more than coders in order to see, understand, capture and partially address the social, technical and ethical considerations.”

I hope you enjoy this book as much as I have. As Dr. Marshall says, every one of you is a data person, and you are enough. We need you to enter, and if you're already in, stay and thrive in this field so that we have a technological future that works for all of us and not only the select few.

—Dr. Timnit Gebru,
Founder and Executive Director of the Distributed Artificial Intelligence Research Institute (DAIR)

Introduction



Source: https://twitter.com/csdoctorsister/status/1536343596336418821

My journey to data conscientiousness started when I was a kid as I rolled coins with my mom and helped my dad organize his job's employee resource group's annual membership rosters. My mom would bring out the big Welch's jar, about half full of loose change. Sitting on the living room floor, she'd dump all the coins on the carpet and we'd start separating them by denominations. Mom would bring out all the coin wrapper rolls she'd gotten from the bank. We'd stack the pennies, nickels, dimes, and quarters—and talk about whatever moms and daughters talk about. She taught me how many of each denomination goes into each coin wrapper roll: 50 pennies gives us 50 cents, 40 nickels gives us $2, 40 quarters gives us $10, and 50 dimes gives us $5. When we'd filled as many of the rolls as we could, we'd count up our earnings. Sometimes it would be $30, and other times it would be closer to $100.

At first, I simply liked the counting, the talking, and stuffing the coins in those small paper wrappings. As I grew up, I started to associate these coins as a resource to get what I wanted. That 50 cents could be put to excellent use to get some SweeTARTS. Two dollars in nickels would keep my candy stash stocked for a week. Five dollars would pay for my favorite order at Swenson's and I'd have change left over. In college, $10 in quarters was gold because no laundry would have been done otherwise. Looking back, I realize it was her way of having me practice my counting, learning the many ways to make a dollar using coins and the value of saving.

My dad, for more than a few years, had this annual huge task of verifying each chapter's membership rosters for his job's nationwide employee resource group. The first year or two or three, Mom and I watched him. Somewhere along the way, I started to help out when asked at first and even volunteered, maybe once. My recollection is fuzzy. What I remember vividly was that the amount of mail he received took down a few forests. There were endless printouts on standard perforated paper from those old dot-matrix printers. The basement became overrun with boxes of unopened envelopes of various sizes, from letter-sized to overstuffed legal-sized.

While my dad was figuring out which pile to tackle, opening each envelope started my supply chain of organization and sorting: record the chapter location and region, clip the membership roster to the envelope, highlight the number of chapter members, leave the chapter's membership dues checks in the envelope, and add this new envelope to the other envelopes in that region pile. And yes, each chapter printed and snail-mailed their membership rosters. The cross-checking of the mailed rosters year after year was dizzying. Some chapters used a great printer and had access to plenty of printer ink. Other chapters weren't so fortunate. My young eyes were called upon to read the smeared and faded letters.

Reconciling these membership rosters took weeks of shifting from one pile to the next. Some people changed chapters due to job relocations but didn't update their membership affiliation. Other people decided to not be part of the employee resource group anymore and didn't follow the member pause/deactivate process. The employee resource group finally created an online database—my dad had something to do with this, I'm sure.

Rolling coins introduced me to data as numbers, math, and financial literacy without being intimidating. Organizing chapter membership rosters introduced me to data as people, context. and guideposts to decisions. Armed with this understanding, I found that school was just one cool place where reading, math, and general exploration of data things happened.

But while roaming the computer science hallways at the University of Rochester I came to recognize how data was viewed by the world. The Year 2000 problem, the Y2K bug, dominated the headlines my junior year. Everyone seemed so concerned about the computing infrastructure and whether systems would “hold up” after the clock struck 12 a.m. on January 1st, 2000. Businesses were desperately trying to back up their data on file servers, zip drives. and 3.5-inch disks. My classmates began signing big money employment contracts with signing bonuses by that fall. They were focused on refining their computer systems and networking skills. That's what employers wanted. I predicted that this dotcom boom was about to bust, so I elected to pursue graduate studies.

I saw the additional concern that businesses feared of not having their data. Everything I could think of had a critical connection to data, particularly why, how, and what we digitally house in systems. And society was singularly focused on the systems themselves. I believed then, as I do now, that data runs the world. I decided to go all in on data in graduate school and my career.

Data gets a bad reputation as a pseudo demon spirit creature because all the numbers and math are deemed complicated, confusing, and not relatable. Data is not a tangible concept to many people—those in computing, tech, and data spaces and those who are not in those spaces. We all, to some degree, are in digital spaces where data lives. Critiquing data uses involves all of us, but for those of us in the data trenches, there's a bigger pressure to suss out the issues and course-correct before the tech product goes public.

This book is for the rebel tech talent, those who acknowledge and are ready to address the limitations of software development. They recognize that tech's philosophy and practice of “move fast, break things” is inherently problematic, and needs to be changed, and they want to pinpoint the ways discrimination exists in this digital data space. The primary reader for this book, however, is the entry-level software developer or data analyst. But frankly, it should be considered a reference guide to making more responsible and equitable data connections.

Data Conscience translates theory to practice. The gaps in our current data infrastructure are spotlighted so that data practitioners know more precisely where issues exists. And I'm centering the most vulnerable, ethical issues and resolutions to address social, political, and economic implications and not just computational ones like optimization, load balancing, and latency.

What you will read in this book is a blend of social sciences, humanities, and data management with tangible, real-world examples. Consider it a modern antemortem describing specific instances of where ethical flags are raised and how data structures help or hinder ethics resolutions. I focus on being preemptive in handling data operation for inclusion rather than conducting conversational (generic) autopsies of case studies and algorithmic audits.

The book is divided into three parts. Part I, “Transparency” (Chapters 1–4), takes you on the rollercoaster of how outcomes and impacts of data, code, algorithms, and systems are revealed to all of us by companies, organizations, and groups. Part II, “Accountability” (Chapters 5–8), covers ways in which data and software teams can critique and explore interventions to make responsible data connections during the tech building phase. And lastly, Part III, “Governance” (Chapters 9–11), reviews the action steps taken thus far and ends as a public accountability manifesto on what all of us can do to humanize our relationship to data.

Here's a brief chapter-by-chapter overview:

- Chapter 1 explores the role data has played in our society, particularly in the United States—how we've handled it and our relationship to handling it well. Oppression tactics, in the law and in the sciences, are mere social controls to enforce a hierarchy positioning that doesn't exist.
- Chapter 2 describes for those of us on the “inside” of tech how we're torn by this realization that the code we write is likely contributing to a cycle of harm that we

don't know how to curtail, stop, or dislodge ourselves from. Reconciling—and more to the point accepting—imperfection in data and tech needs a place in tech. The choice between error or no error doesn't exist anymore. There's a third choice: nontech-solvable.

- Chapter 3 tackles the term “bias” and its multitude of interpretations head on. I describe how bias shows up and ways to shift our mindset on how we recognize and handle it, even before we write a single line of code. Getting overwhelmed and disengaging in combatting bias efforts is no longer an option.
- Chapter 4 stretches our minds about what we've accepted as computational thinking and standard discussion points to fold in a more intentional socio-ethical tech understanding. Coding requires a 360-degree panoramic view that requires more than coders in order to see, understand, capture, and partially address the social, technical, and ethical considerations.
- Chapter 5 focuses on asking the “why” questions, especially as part of data collection and reformat practices. Tech does a poor job of handling data collection and reformat. Learning to ask questions early, often, and with real people in mind streamlines how we manage data operations as a data, computing, and larger tech community.
- Chapter 6 focuses on asking the “what” questions, especially as part of data storage and management procedures. We must come to grips with the fact that the data storage landscape is a culmination of intentional, yet sometimes harmful, decision-making exercises with social, computational, and morality implications.

- Chapter 7 focuses on asking the “how” questions, especially as part of data analytics. The algorithms, systems, and platforms are taken from the same playbook, by the same homogeneous people. There's much we can learn about data from our comrades in the humanities and social sciences.
- Chapter 8 focuses on asking the “when” questions, especially as part of data visualization. The allure of data visualizations is enticing, so proceed with caution with every chart, graph, and dashboard you encounter.
- Chapter 9 snaps us back to reality as tech moves fast while the law moves slow. Juggling the usefulness versus persnickety hindrances of the law dominates this chapter. Focusing on the fundamental building blocks of tech, like data and the algorithms that use it, has traction and momentum rather than constructing legislation that's fixated on applications of technologies.
- Chapter 10 discusses tech's dominance in our society and, in particular, as a culture that excludes other solution options. The rockstar tech workers, or algorithmic influencers, guide every industry one software update, code version release, or code library at a time. But clearly, we've hit a ceiling in what tech should do.
- Chapter 11 is my public accountability manifesto. We're each battling for our dignity in digital spaces, and we must do so with the same indignant veracity as we do in physical spaces. The tech industry's ability to operate, for the most part, without impunity puts more onus on us, as global digital citizens, to maintain intense pressure for transparency, accountability and governance in all spaces where data resides. Algorithmic processes, systems, platforms, and

institutions won't become responsible or equitable without us making it a requirement, rather than a choice.

Part I Transparency

In This Part

Chapter 1: Oppression By…

Chapter 2: Morality

Chapter 3: Bias

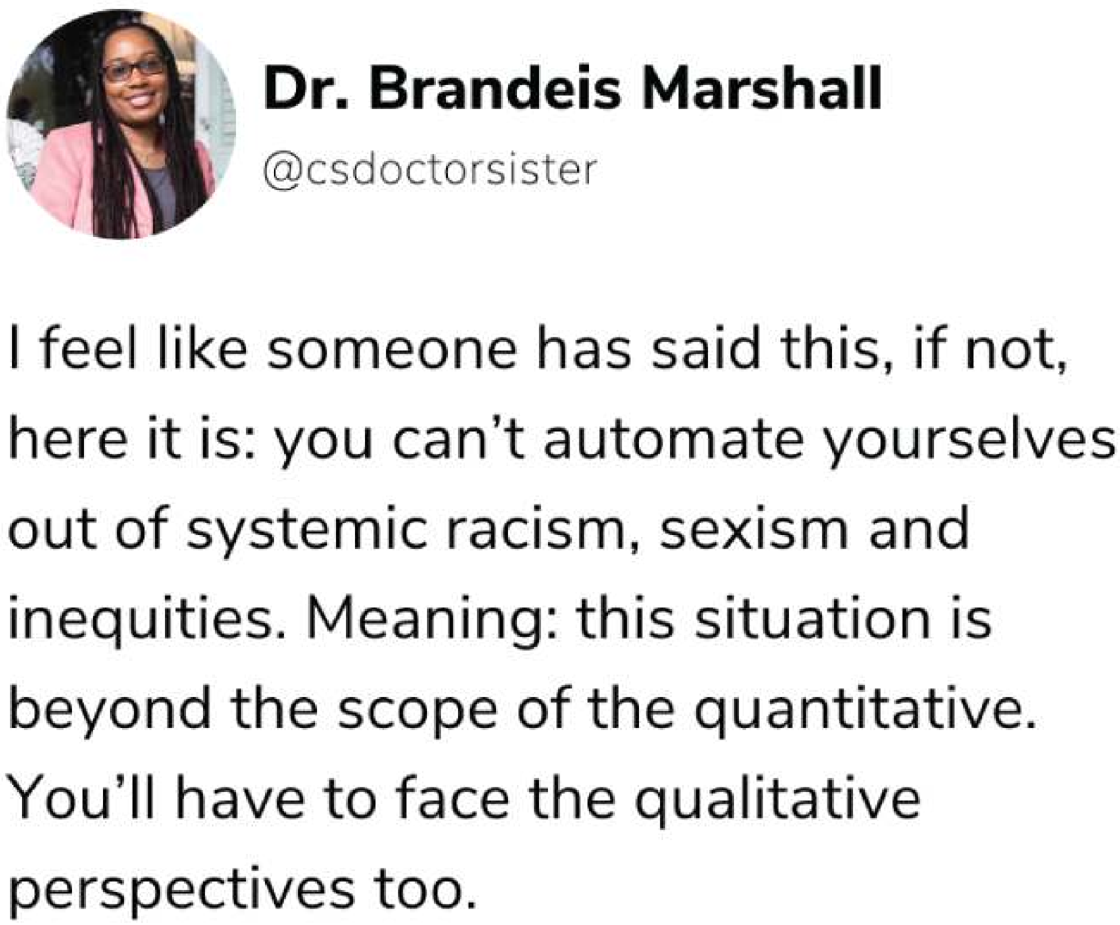
Chapter 4: Computational Thinking in Practice

Welcome to Part I, “Transparency.” Our society today is obsessed with *using* technology. The power that technology brings to people's ability in completing perceived mundane tasks is astonishing and lightning fast. The “hurry up and complete more tasks in a fixed time frame” mindset overlooks the cascading and compounding influences of those activities on the human condition. But this isn't the first time the outcomes overshadowed the people. There's an uncomfortable U.S. history of sidelining the humanity of certain people so that others can thrive. The journey starts at this uncomfortable place when technology wasn't prevalent. How we as a society are living through this technology era isn't new, but simply a reincarnation of how the world has operated in the past. Being transparent about the pre-tech era illuminates how tech systems came to be designed with so many flaws. The data, code, algorithms, systems, and platforms that make the tech industry thrive are a wicked web of assumptions, presumptions, and opaqueness.

> **Transparency (n)** – Revealing your data and code. Bad for proprietary and sensitive information. Thus really hard; quite frankly, even impossible. Not to be confused with clear communication about how your system actually works.
>
> **—Karen Hao, in “Big Tech’s Guide to Talking About AI Ethics,”**
>
> www.technologyreview.com/2021/04/13/1022568/big-tech-ai-ethics-guide[1]

> **Transparency (n)** – Revealing outcomes and impact of the data, code, algorithms, and systems by companies, organizations, and groups.
>
> —**Revised Definition by Brandeis Hill Marshall**

Data is the fuel that runs our algorithms and systems. Chapters 1–4 have us get inside data as a construct, structure, and limiting factor when digitized. We struggle to wrap our minds around data within and outside of digital infrastructures. Ultimately, it's up to the collective we to be clear-eyed in our expectations of data and its uses.



Dr. Brandeis Marshall

@csdoctorsister

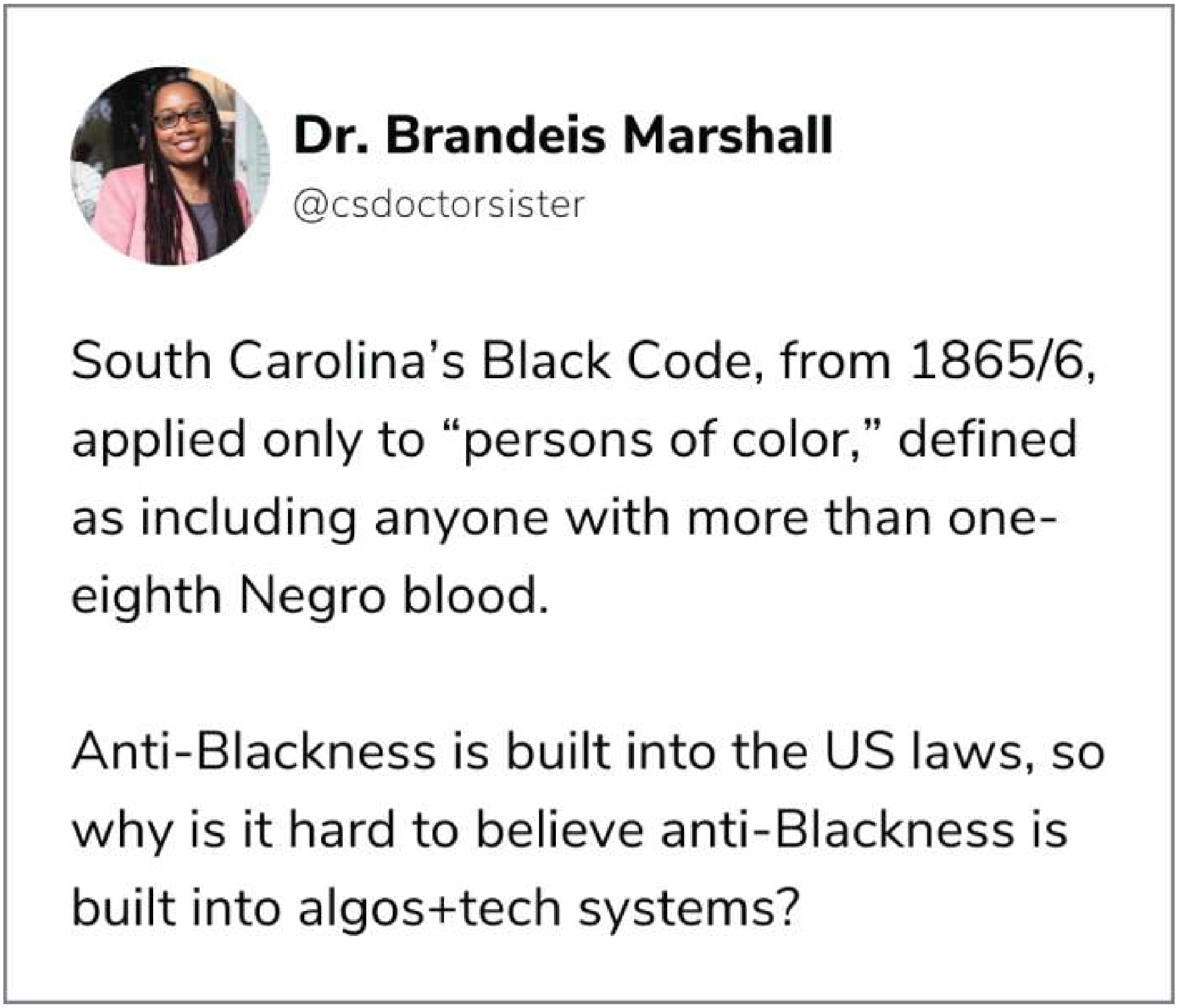
I feel like someone has said this, if not, here it is: you can't automate yourselves out of systemic racism, sexism and inequities. Meaning: this situation is beyond the scope of the quantitative. You'll have to face the qualitative perspectives too.

Source: https://twitter.com/csdoctorsister/status/1315376187632476161

Note

1. Hao, Karen. "Big Tech's Guide to Talking About AI Ethics." MIT Technology Review, 2021. www.technologyreview.com/2021/04/13/1022568/big-tech-ai-ethics-guide.

CHAPTER 1
Oppression By…



Source: Twitter, Inc.

In tech, the approach is to create a solution that works well for 80 percent of users. The remaining 20 percent have to conform, find workarounds, or suffer in silence. Those on the fringe aren't considered core demographic or essential members of the targeted population. So tech folks don't want to discuss or address racism, sexism, ableism, and otherism. Because each of these distinguishing

About the Technical Editor

Dr. Triveni Gandhi is a “jill-of-all-trades” data scientist, thought leader, and advocate for the responsible use of AI who likes to find simple solutions to complicated problems. Her journey with data science began with basic statistics in grad school, and after finishing her degree she worked as a data analyst at an educational nonprofit organization, where she saw just how transformative data technology can be. It was not until she moved into the for-profit sector that Triveni saw the numerous shortcomings of the tech world—especially the lack of socioeconomic perspectives in the development and deployment of technology. As a result, Triveni has become a champion for diversifying voices in tech and teaching data practitioners how to discover and mitigate biases in their work. In her spare time, Triveni enjoys gardening, playing D&D, and spending time with her husband and puppy.

Acknowledgments

From information gathering to impact making, my passion for putting data into a humanizing context has been an over 20-year journey. The dotcom boom and subsequent bust showed to me the immense power of data actions and our society's reliance on a digital infrastructure. While others saw databases as boring and tedious, I saw them as a bedrock of our digital globalized world. Being this bedrock comes with it a deep responsibility. We all must be more careful in how we handle and manage data online and offline. After many interweb searches, I couldn't find a book that addressed this delicate and complicated, yet important, reality. Embarking on writing this book has been a dream of mine that could not have happened without so many people.

First, to my husband and life partner, Gemez Marshall, you are my biggest supporter and advocate. You've endured listening to me talk through ideas. You brought me food and drink when you saw I was in the “writing zone.” You meet my anxiety and doubts with “one thing at a time” and “happy thoughts.” You tell me I can do anything, and with your loving encouragement I do. I am eternally grateful for your patience and constant cheerleading. I know you're ready for me to write another book, but let's just see how this one goes.

To my parents, Carlton and Magnolia Hill, you both have been my go-to models on “doing the right thing even when the right thing isn't easy.” You've been my fiercest champions and guided me through difficult situations. Thanks for the weekly pep talks, Dad. You know I needed each and every one. And Mom, thanks for always keeping it 100% real. You taught me to be candid. My love of data started with you as we rolled all your coin change on the living room floor when I was a kid.

To the Marshalls, you're my in-laws but you've always treated me like one of your biological children. Thanks for lifting me in your prayers throughout the writing journey, being quick to text an encouraging word, and checking in on me. You two were one of the first to preorder, and I already know where the book is going to reside in your house.

Iretta B. C. Kearse, thanks for being my work friend-turned-sister friend. From the many, many Instagram reel shares to the text messages, you kept me smiling and laughing this entire process, even when I wanted to stay in a sour mood. I got you and you got me. We ride together.

Kim Crayton, thanks for being the friend I never saw coming. When we met via Twitter, I would have never imagined all the powerful and insightful conversations we would have had. Our conversations go from the fun-silly to an audacious liberated society and back to fun-silly in minutes. You helped me push past my own bounds of what's possible. Thanks, once again, for allowing your guiding principles to grace this book.

Thanks to Ayodele Odubela for being the person who opened the Wiley door to me. I felt discouraged and uncertain whether I had a book proposal that a publishing company would deem publishable. For acknowledging my reservations and showing me an alternate path, I am sincerely grateful. Happy to be combatting algorithmic-based harms and fighting for equity and justice in automated systems with you.

To my acquisitions editor, Devin Lewis, you were my first point of contact at Wiley. From our initial call, I felt seen. Thanks for being transparent about the process, advocating for this book internally, and encouraging me to share my book writing journey through monthly newsletter updates.

To my development editor, Adaobi Obi Tulton, you have been my North Star. From the gentle nudge reminders to submit those latter chapter edits to the initial document formatting fixes, the care you showed me is much appreciated. I felt fully supported even when I struggled to find the words to express my thoughts.

To my technical editor, Triveni Gandhi, I'm forever grateful for your comments, thoughts, and insightful quips. You took on this project without hesitation and it has meant the world to me. I'm still amazed how effortless our first and hours-long conversation flowed at that Dataiku event. Every time we correspond, it's having a meetup with a good friend.

To the Foreword author, Timnit Gebru, I was frankly gobsmacked when you agreed to write the Foreword. I assumed you would say you were too busy. But no, you lead with a strength of kindness that continually inspires me. Thank you for lending your words here and your time always wholeheartedly.

To everyone at Wiley, thank you for helping me to fulfill my dream by contributing to the data and tech industries as the whole me—an educator-scholar-consultant who intentionally seeks ways to humanize data operations.

Finally, I'd like to thank those of you who read, share, and comment on this book. Hope this book helps you be more seen.