

Brandeis Hill Marshall



DATA CONSCIENCE

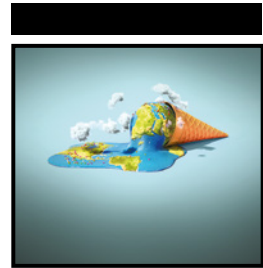
ALGORITHMIC SIEGE ON OUR HUM4N1TY

Foreword by Timnit Gebru

Founder and Executive Director, Distributed Artificial Intelligence Research Institute

WILEY

Data Conscience



Data Conscience

Algorithmic Siege on Our Humanity

Brandeis Hill Marshall

WILEY

Copyright © 2023 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.

Published simultaneously in Canada and the United Kingdom.

ISBN: 978-1-119-82118-2

ISBN: 978-1-119-82119-9 (ebk)

ISBN: 978-1-119-82120-5 (ebk)

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at www.wiley.com/go/permission.

Trademarks: WILEY and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc. is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

If you believe you've found a mistake in this book, please bring it to our attention by emailing our reader support team at wileysupport@wiley.com with the subject line "Possible Book Errata Submission."

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Control Number: 2022942937

Cover image: © sveta/Adobe Stock Photos

Cover design: Wiley

To my husband, my parents and my ancestors



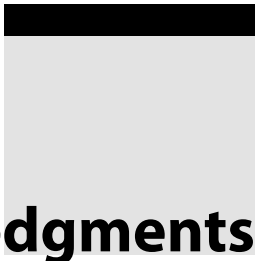
About the Author

Brandeis Hill Marshall teaches, speaks, and writes about the impact of data practices on technology and society. Her work contributes to data education, computer science education, and data science fields. She leads and runs a social impact business that provides educational resources and training designed to help data educators, scholars, and practitioners humanize data operations. She has advised and educated thousands in the public and corporate sectors. She also consults with organizations that are proactive in mitigating harmful effects of their data pipelines. Brandeis holds a PhD and master's degree in computer science from Rensselaer Polytechnic Institute and a BS in computer science from the University of Rochester.



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. Kears, 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.



Contents at a Glance

Foreword		xix
Introduction		xxi
Part I	Transparency	1
Chapter 1	Oppression By. . .	3
Chapter 2	Morality	27
Chapter 3	Bias	61
Chapter 4	Computational Thinking in Practice	87
Part II	Accountability	117
Chapter 5	Messy Gathering Grove	119
Chapter 6	Inconsistent Storage Sanctuary	143
Chapter 7	Circus of Misguided Analysis	163
Chapter 8	Double-Edged Visualization Sword	191
Part III	Governance	213
Chapter 9	By the Law	215
Chapter 10	By Algorithmic Influencers	231
Chapter 11	By the Public	263
Appendix A	Code for <i>app.py</i>	287
Appendix B	Code for <i>screen.py</i>	291
Appendix C	Code for <i>search.py</i>	297

Appendix D Pseudocode for <i>faceit.py</i>	305
Appendix E The Data Visualisation Catalogue's Visualization Types	309
Appendix F Glossary	313
Index	315



Contents

Foreword	xix
Introduction	xxi
Part I Transparency	1
Chapter 1 Oppression By . . .	3
The Law	4
Slave Codes	5
Black Codes	5
The Rise of Jim Crow Laws	8
Breaking Open Jim Crow Laws	11
Overt Surveillance	12
Surveillance at Scale	13
The Science	16
Numbers	16
Anthropometry	18
Eugenics	19
Summary	23
Notes	23
Recommended Reading	25
Chapter 2 Morality	27
Data Is All Around Us	29
Morality and Technology	33
Defining Tech Ethics	33
Mapping Tech Ethics to Human Ethics	39
Squeezing in Data Ethics	45
Misconceptions of Data Ethics	49
Misconception 1: Goodness of Data, and Tech by Proxy, Is Apolitical or Bipartisan	49

	Misconception 2: Data Ethics Is Focused Solely on Laws Protecting Confidentiality and Privacy	50
	Misconception 3: Implementing Data Ethics Practices Will Make Data Objective	52
	Notable Misconception Mentions: Ethics and Diversity, Equity, and Inclusion (DEI) Are Interchangeable	53
	Another Notable Mention: Software Developers Are Only Responsible for Societal Outcomes Stemming from Their Code	54
	Limits of Tech and Data Ethics	55
	Summary	57
	Notes	57
Chapter 3	Bias	61
	Types of Bias	62
	Defining Bias	63
	Concrete Example of Biases	65
	The Bias Wheel	70
	Before You Code	73
	Case Study Scenario: Data Sourcing for an Employee Candidate Résumé Database	77
	Case Study Scenario: Data Manipulation for an Employee Candidate Résumé Database	78
	Case Study Scenario: Data Interpretation for an Employee Candidate Résumé Database	82
	Bias Messaging	83
	Summary	83
	Notes	84
Chapter 4	Computational Thinking in Practice	87
	Ready to Code	88
	The Shampoo Algorithm	89
	Computational Thinking	91
	Coding Environments	93
	Algorithmic Justice Practice	95
	Code Cloning	97
	Socio-Techno-Ethical Review: <i>app.py</i>	101
	Socio-Techno-Ethical Review: <i>screen.py</i>	103
	Socio-Techno-Ethical Review: <i>search.py</i>	109
	Summary	114
	Notes	114
Part II	Accountability	117
Chapter 5	Messy Gathering Grove	119
	Ask the Why Question	120
	Collection	124
	Open Source Dataset Example: Deciding Data Ownership	127
	Open Source Dataset Example: Considering Data Privacy	129

	Reformat	133
	Summary	139
	Notes	139
Chapter 6	Inconsistent Storage Sanctuary	143
	Ask the “What” Question	144
	Files, Sheets, and the Cloud	146
	Decisions in a Vacuum	149
	Case Study: Black Twitter	150
	Modeling Content Associations	153
	Manipulating with SQL	158
	Summary	160
	Notes	161
Chapter 7	Circus of Misguided Analysis	163
	Ask the “How” Question	164
	Misevaluating the “Cleaned” Dataset	169
	Overautomating k, K, and Thresholds	177
	Deepfake Technology	179
	Not Estimating Algorithmic Risk at Scale	185
	Summary	187
	Notes	187
Chapter 8	Double-Edged Visualization Sword	191
	Ask the “When” Question	192
	Critiquing Visual Construction	197
	Disabilities in View	201
	Pretty Picture Mirage	204
	Case Study: SAT College Board Dataset	207
	Summary	208
	Notes	209
Part III	Governance	213
Chapter 9	By the Law	215
	Federal and State Legislation	216
	International and Transatlantic Legislation	219
	Regulating the Tech Sector	221
	Summary	228
	Notes	228
Chapter 10	By Algorithmic Influencers	231
	Group (Re)Think	232
	Flyaway Fairness	238
	Algorithmic Fairness	239
	Broadening Fairness	241
	Moderation Modes	245
	Double Standards	246
	Calling Out Algorithmic Misogynoir	252

	Data and Oversight	254
	Summary	256
	Notes	256
Chapter 11	By the Public	263
	Freeing the Underestimated	264
	Learning Data Civics	267
	The State of the Data Industry	271
	Living in the 21st Century	273
	Condemning the Original Stain	277
	Tech Safety in Numbers	279
	Summary	283
	Notes	283
Appendix A	Code for <i>app.py</i>	287
	A	287
	B	288
	C	288
	D	289
Appendix B	Code for <i>screen.py</i>	291
	A	291
	B	294
	C	295
Appendix C	Code for <i>search.py</i>	297
	A	297
	B	300
	C	301
	D	303
Appendix D	Pseudocode for <i>faceit.py</i>	305
Appendix E	The Data Visualisation Catalogue's Visualization Types	309
Appendix F	Glossary	313
	Index	315



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 exist. 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

Proffitt, Dennis, 194
 Profit Without Oppression guiding
 principles, 280
 programming, FORTRAN
 language, 29
 pseudocode for FaceIt, 305–307
 PULSE vision model, 49–50
 PVF (public value
 framework), 254–255

Q

QA engineers, Amazon AI
 hiring, 65–70
 questions
 data interpretation, 195
 data manipulation, 167–169
 data sourcing, 122
 how, 164–169
 what, 144–146
 when, 192–197
 why, 120–124
 Quetelet, Adolphe, 18

R

racism
 platformed, 253–254
 scientific (eugenics), 19
 RAIL (Responsible AI Licenses),
 242
 Raji, Deborah, 254
 Raynham, Helen, 254
 RealTalk, 181
 recommendations
 banning, 226–227
 data brokers, 225
 malcoding, 224–225
 reformatting data, 133–139
 regression modeling, 173–174
 linear regression, 175–176
 regulations, 221–227. *See also* legal
 issues; legislation
 AI-based tools, 222–223
 CTC (Center for Technological
 Civility) example, 224–227

FTC (Federal Trade
 Commission), 216
 recommendations
 banning, 226–227
 data brokers, 225
 malcoding, 224–225
 Reign, April, 32
 Responsible AI Toolkit, 242
 results, 192
 resume processing
 applicant demographic
 identities, 153
 automated, 38–39
 biases example, 65–70
 data ethics case studies, 77–82
 ERD representation, 155
 synthetic resume data, 147
 revisionism, platform
 demographics, 32
 ROC (receiver operating
 characteristic), 107
 Rocchio algorithm, 234–236

S

SACNAS (Society for Advancement of
 Chicanos/Hispanics & Native
 Americans in Science), 125
 safety
 alternative language/spelling,
 276
 direct communication, 276
 online engagement, 276
 preemptive actions, 274–276
 sampling bias, 96
 scientific racism, 19. *See also* eugenics
 screen.py file, 103–109
 code blocks, 291–295
 extractText method, 104
 fit() method, 107
 res() class, 103
 res() method, 103
 slate module, 104
 summarize method, 105–106
 textextract.process() method, 105

- TextRank algorithm, 105
- transform() method, 107
- SDLC (software development lifecycle), 66–67
- search.py file, 109–114
 - code blocks, 297–304
 - normalize() routine, 109–110
 - semanticSearch() method, 112
- selection bias, 96
- sensitive data, 130–131
- separate but equal
 - Brown v. Board of Education of Topeka*, 11–12
 - cluster analysis and, 8
 - Plessy v. Ferguson*, 8–9
- service terms/agreement
 - acceptance, 126
- shampoo algorithm, 89–91
- SHAP (SHapley Additive exPlanations), 243
- slave codes, 5. *See also* Black codes
- social Darwinism, 19. *See also* eugenics
- social media. *See also*
 - Facebook; Twitter
 - Cambridge Analytics, 32
- socio-techno-ethical compliance, 97–98
 - Automated Résumé Screening System, 98–99
 - phpMyAdmin, 101–114
- software developers, societal
 - outcomes from code, 54–55
- solutionism, 233
- solutions, 192
- Spanish flu, Black women working, 11
- spelling alternatives, 276
- Spyder, 93
- SQL (Structured Query Language), 158–160
- state legislation, 216–219
- Steggerda, Morris, 21
- STEM (science, technology, engineering, math), 154–155
- stemming, 110–111
- sterilization, forced, 21

- stolen content, 102
- storage, databases, 148
- subverting trolls, 275–276
- surveillance
 - digital, modern, start of, 14
 - Katz v. United States*, 13
 - overt surveillance, 12–13
 - at scale, 13–16
 - wiretap surveillance, 12
- synthetic resume data, 147

T

- Tableau, 93
- Tay (Microsoft chatbot), 38
- tech, as computer science, 62
- tech communities, rebel, 281
- tech errors, 41
- tech ethics, 33–39
 - automated resume processing, 38–39
 - biases example, 65–70
 - human ethics and, 39–45
 - IEEE (Institute of Electrical and Electronics Engineers), 35
 - IRB (institutional review board), 124–125
 - limits, 55–56
 - similarities with data and human ethics, 47–48
 - software development, 38
- technological solutionism, 232–233
 - iPhone, 232
 - mental health apps, 234
 - Moodbeam, 233–234
- technology
 - engagement level changes, 34
 - as a service, 33
- TF-IDF (term frequency-inverse document frequency), 106
- third-party vendors, data breaches and, 122
- Thompson, Joseph, 278
- topic drift in algorithms, 276
- top-k, 178–179

transatlantic legislation, 219–221
 transparency, 41–45
 trauma of enslavement, 4–5
 TREC (Text Retrieval Conference), 133–134
 troll subversion, 275–276
 Tweepy API, 136, 151
 Twitter, 30–31
 BlackTwitter, 32
 engagement rate, 31–32
 false news story retweets, 32–33
 FTC fine, 221–222
 Offensive Content page, 247
 #OscarsSoWhite, 32
 Transparency Report, 249
 troll subversion, 275–276
 White, Shana V., 248–252

U

underestimated persons, 264–266
 unknown data, formatting and, 138
 unpaid labor, 278–279
 U.S. Constitution, 14th
 Amendment, 7, 278
 use cases, algorithmic influences
 and, 232
 user authorization, 102–103

V

value, Big Data, 313
 variability, Big Data, 314
 variety, Big Data, 313
 velocity, Big Data, 313
 veracity, Big Data, 313
 Villermé, Louis R., 18
 Visual Studio Code, 93
 visualizations, 191–192. *See also* visuals
 case study, 207–208
 chart junk, 206
 cloropleth, 205–206
 Data Visualization Catalogue
 project, 204
 arc diagram, 311
 area graph, 311

bar chart, 311
 box and whisker plot, 311
 brainstorm, 311
 bubble chart, 311
 bubble map, 311
 bullet graph, 311
 calendar, 311
 candlestick chart, 311
 chord diagram, 311
 choropleth map, 311
 circle packing, 311
 connection map, 311
 density plot, 311
 donut chart, 311
 dot map, 311
 dot matrix chart, 311
 error bars, 310
 flow chart, 310
 flow map, 310
 Gantt chart, 310
 heat map, 310
 histogram, 310
 illustration diagram, 310
 Kagi chart, 310
 line graph, 310
 Marimekko chart, 310
 multiset bar chart, 310
 network diagram, 310
 nightingale rose chart, 310
 non-ribbon chord diagram, 310
 open-high-low-close chart, 310
 parallel coordinates plot, 310
 parallel sets, 310
 pictogram chart, 310
 pie chart, 310
 point and figure chart, 310
 population pyramid, 310
 proportional area chart, 310
 radar chart, 310
 radial bar chart, 310
 radial column chart, 310
 Sankey diagram, 310
 scatterplot, 310
 span chart, 310

- spiral plot, 310
- stacked area graph, 310
- stacked bar graph, 310
- steam graph, 310
- stem and leaf plot, 310
- sunburst diagram, 310
- tally chart, 310
- time line, 310
- timetable, 311
- tree diagram, 311
- treemap, 311
- Venn diagram, 311
- violin plot, 311
- word cloud, 311
- SAT College Board dataset, 207–208
- types, 204–206
- visuals
 - created content, 193
 - critiquing, 197–204
 - data interpretation questions, 195
 - as deflection practice, 194
 - story telling, 193
- volume, Big Data, 313

W-Z

- Warren, Earl, 11–12
- weaponized Blackness, 13
- webscraping, Black Twitter Project, 151
- well ordered, 88
- Wells-Barnett, Ida B., 279
- WFH (work-from-home) team,
 - Moodbeam and, 233–234
- what question, 144–146
- when question, 192–197
- White, Shana V., Twitter ban,
 - 248–252
- white colonialism, 4–5
 - group targeting, 12–13
- why question, 120–124
- Williams, Robert Julian-Borchak,
 - 41
- wiretap surveillance, 12–13
- WordNet, 131
- workplace problem woman of
 - color, 281–283
- Wysinger, Carolyn, 247