an impact if a male family member gave it to you? Or, for a more recent example, how would you feel if you took time off from your hourly job to go cast your vote, only to discover when you got there that your name had been purged from the official voting roll or that there was a line so long that it would require that you miss half a day's pay, or stand for hours in the cold, or ... the list could go on. These are examples of how it *feels* to know that systems of power are not on your side and, at times, are actively seeking to take away the small amount of power that you do possess.[19]

The matrix of domination works to uphold the undue privilege of *dominant* groups while unfairly oppressing *minoritized* groups. What does this mean? Beginning in this chapter and continuing throughout the book, we use the term *minoritized* to describe groups of people who are positioned in opposition to a more powerful social group. While the term *minority* describes a social group that is comprised of fewer people, *minoritized* indicates that a social group is actively devalued and oppressed by a dominant group, one that holds more economic, social, and political power. With respect to gender, for example, men constitute the dominant group, while all other genders constitute minoritized groups. This remains true even as women actually constitute a majority of the world population. *Sexism* is the term that names this form of oppression. In relation to race, white people constitute the dominant group (racism); in relation to class, wealthy and educated people constitute the dominant group (classism); and so on.[20]

Using the concept of the matrix of domination and the distinction between dominant and minoritized groups, we can begin to examine how power unfolds in and around data. This often means asking uncomfortable questions: who is doing the work of data science (and who is not)? Whose goals are prioritized in data science (and whose are not)? And who benefits from data science (and who is either overlooked or actively harmed)?[21] These questions are uncomfortable because they unmask the inconvenient truth that there are groups of people who are disproportionately benefitting from data science, and there are groups of people who are disproportionately harmed. Asking these *who questions* allows us, as data scientists ourselves, to start to see how privilege is baked into our data practices and our data products.[22]

Data Science by Whom?

It is important to acknowledge the elephant in the server room: the demographics of data science (and related occupations like software engineering and artificial intelligence research) do not represent the population as a whole. According to the most recent data from the US Bureau of Labor Statistics, released in 2018, only 26 percent of those in “computer and mathematical occupations” are women.[23] And across all of those women, only 12 percent are Black or Latinx women, even though Black and Latinx women make up 22.5 percent of the US population.[24] A report by the research group AI Now about the diversity crisis in artificial intelligence notes that women comprise only 15 percent of AI research staff at Facebook and 10 percent at Google.[25] These numbers are probably not a surprise. The more surprising thing is that those numbers are getting worse, not better. According to a research report published by the American Association of University Women in 2015, women computer science graduates in the United States peaked in the mid-1980s at 37 percent, and we have seen a steady decline in the years since then to 26 percent today (figure 1.2).[26] As “data analysts” (low-status number crunchers) have become rebranded as “data scientists” (high status researchers), women are being pushed out in order to make room for more highly valued and more highly compensated men.[27]

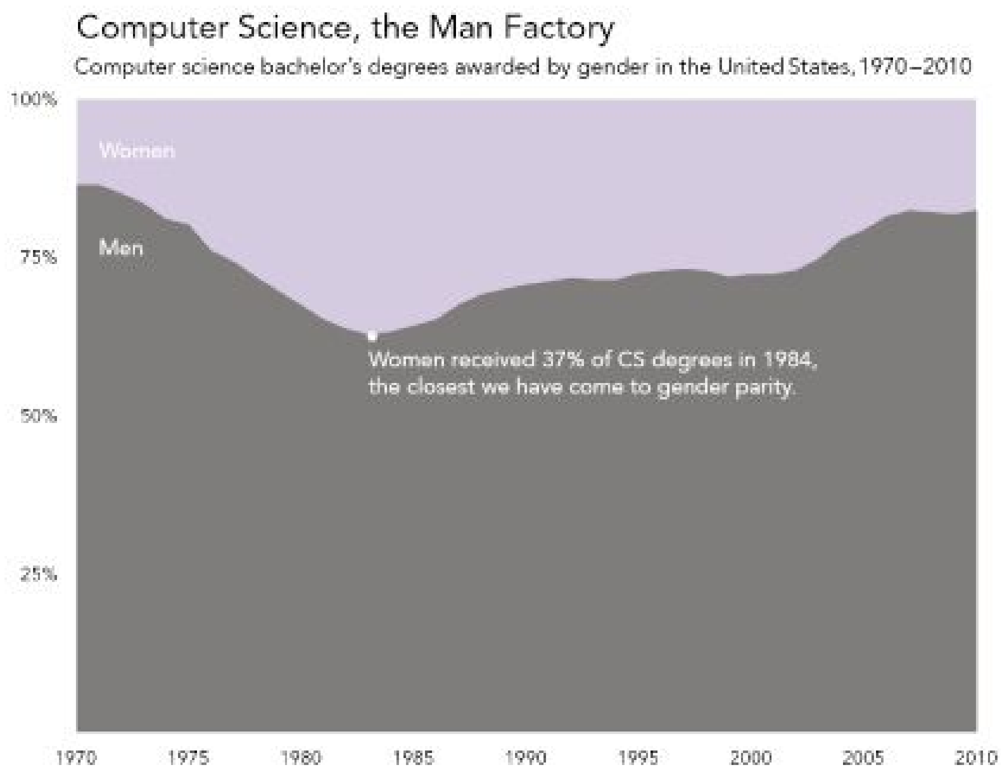

Figure 1.2: Computer science has always been dominated by men and the situation is worsening (even while many other scientific and technical fields have made significant strides toward gender parity). Women awarded bachelor's degrees in computer science in the United States peaked in the mid-1980s at 37 percent, and we have seen a steady increase in the ratio of men to women in the years since then. This particular report treated gender as a binary, so there was no data about nonbinary people. Graphic by Catherine D'Ignazio. Data from the National Center for Education Statistics. *Source:* Data from Christianne Corbett and Catherine Hill, *Solving the Equation: The Variables for Women's Success in Engineering and Computing* (Washington, DC: American Association of University Women, 2015). *Credit:* Graphic by Catherine D'Ignazio.

There are not disparities only along gender lines in the higher education pipeline. The same report noted specific underrepresentation for Native American women, multiracial women, white women, and all Black and Latinx people. So is it really a surprise that each day brings a new example of data science being used to disempower and oppress minoritized groups? In 2018, it was revealed that Amazon had been developing an algorithm to screen its first-round job applicants. But because the model had been trained on the resumes of prior applicants, who were predominantly male, it developed an even stronger preference for male applicants. It downgraded resumes with the word *women* and graduates of women's colleges. Ultimately, Amazon had to cancel the project.[28] This example reinforces the work of Safiya Umoja Noble, whose book, *Algorithms of Oppression*, has shown how both gender and racial biases are encoded into some of the most pervasive data-driven systems—including Google search, which boasts over five billion unique web searches per day. Noble describes how, as recently as 2016, comparable searches for “three Black teenagers” and “three white teenagers” turned up wildly different representations of those teens. The former returned mugshots, while the latter returned wholesome stock photography.[29]

The problems of gender and racial bias in our information systems are complex, but some of their key causes are plain as day: the data that shape them, and the models designed to put those data to use, are created by small groups of people and then scaled up to users around the globe. But those small groups are not at all representative of the globe as a whole, nor even of a single city in the United States. When data teams are primarily composed of people from dominant groups, those perspectives come to exert outsized influence on the decisions being made—to the exclusion of other identities and perspectives. This is not usually intentional; it comes from the ignorance of being on top. We describe this deficiency as a *privilege hazard*.

How does this come to pass? Let’s take a minute to imagine what life is like for someone who epitomizes the dominant group in data science: a straight, white, cisgender man with formal technical credentials who lives in the United States. When he looks for a home or applies for a credit card, people are eager for his business. People smile when he holds his girlfriend’s hand in public. His body doesn’t change due to childbirth or breastfeeding, so he does not need to think about workplace accommodations. He presents his social security number in jobs as a formality, but it never hinders his application from being processed or brings him unwanted attention. The ease with which he traverses the world is invisible to him because it has been designed for people just like him. He does not think about how life might be different for everyone else. In fact, it is difficult for him to imagine that at all.

This is the *privilege hazard*: the phenomenon that makes those who occupy the most privileged positions among us—those with good educations, respected credentials, and professional accolades—so poorly equipped to recognize instances of oppression in the world.[30] They lack what Anita Gurumurthy, executive director of IT for Change, has called “the empiricism of lived experience.”[31] And this lack of lived experience—this evidence of how things truly *are*—profoundly limits their ability to foresee and prevent harm, to identify existing problems in the world, and to imagine possible solutions.

The privilege hazard occurs at the level of the individual—in the interpersonal domain of the matrix of domination—but it is much more harmful in aggregate because it reaches the hegemonic, disciplinary and structural domains as well. So it matters deeply that data science and artificial intelligence are dominated by elite white men because it means there is a collective privilege hazard so great that it would be a profound surprise if they could actually identify instances of bias prior to unleashing them onto the world. Social scientist Kate Crawford has advanced the idea that the biggest threat from artificial intelligence systems is not that they will become smarter than humans, but rather that they will hard-code sexism, racism, and other forms of discrimination into the digital infrastructure of our societies.[32]

What's more, the same cis het white men responsible for designing those systems lack the ability to detect harms and biases in their systems once they've been released into the world.[33] In the case of the "three teenagers" Google searches, for example, it was a young Black teenager that pointed out the problem and a Black scholar who wrote about the problem. The burden consistently falls upon those more intimately familiar with the privilege hazard—in data science as in life—to call out the creators of those systems for their limitations.

For example, Joy Buolamwini, a Ghanaian-American graduate student at MIT, was working on a class project using facial-analysis software.[34] But there was a problem—the software couldn't "see" Buolamwini's dark-skinned face (where "seeing" means that it detected a face in the image, like when a phone camera draws a square around a person's face in the frame). It had no problem seeing her lighter-skinned collaborators. She tried drawing a face on her hand and putting it in front of the camera; it detected that. Finally, Buolamwini put on a white mask, essentially going in "whiteface" (figure 1.3).[35] The system detected the mask's facial features perfectly.

Digging deeper into the code and benchmarking data behind these systems, Buolamwini discovered that the dataset on which many of facial-recognition algorithms are tested contains 78 percent male faces and 84 percent white faces. When she did an intersectional breakdown of another test dataset—looking at gender and skin type together—only 4 percent of the faces in that dataset were women and dark-skinned. In their evaluation of three commercial systems, Buolamwini and computer scientist Timnit Gebru showed that darker-skinned women were up to forty-four times more likely to be misclassified than lighter-skinned males.[36] It's no wonder that the software failed to detect Buolamwini's face: both the training data and the benchmarking data relegate women of color to a tiny fraction of the overall dataset.[37]

Figure 1.3: Joy Buolamwini found that she had to put on a white mask for the facial detection program to “see” her face. Buolamwini is now founder of the Algorithmic Justice League. Courtesy of Joy Buolamwini. *Credit:* Courtesy of Joy Buolamwini.

This is the privilege hazard in action—that no coder, tester, or user of the software had previously identified such a problem or even thought to look. Buolamwini’s work has been widely covered by the national media (by the *New York Times*, by CNN, by the *Economist*, by *Bloomberg BusinessWeek*, and others) in articles that typically contain a hint of shock.[38] This is a testament to the social, political, and technical importance of the work, as well as to how those in positions of power—not just in the field of data science, but in the mainstream media, in elected government, and at the heads of corporations—are so often surprised to learn that their “intelligent technologies” are not so intelligent after all. (They need to read data journalist Meredith Broussard’s book *Artificial Unintelligence*).[39] For another example, think back to the introduction of this book, where we quoted Shetterly as reporting that Christine Darden’s white male manager was “shocked at the disparity” between the promotion rates of men and women. We can speculate that Darden herself wasn’t shocked, just as Buolamwini and Gebru likely were not entirely shocked at the outcome of their study either. When sexism, racism, and other forms of oppression are publicly unmasked, it is almost never surprising to those who experience them.

For people in positions of power and privilege, issues of race and gender and class and ability—to name only a few—are OPP: other people's problems. Author and antiracist educator Robin DiAngelo describes instances like the "shock" of Darden's boss or the surprise in the media coverage of Buolamwini's various projects as a symptom of the "racial innocence" of white people.[40] In other words, those who occupy positions of privilege in society are able to remain innocent of that privilege. Race becomes something that only people of color have. Gender becomes something that only women and nonbinary people have. Sexual orientation becomes something that all people *except* heterosexual people have. And so on. A personal anecdote might help illustrate this point. When we published the first draft of this book online, Catherine told a colleague about it. His earnestly enthusiastic response was, "Oh great! I'll show it to my female graduate students!" To which Catherine rejoined, "You might want to show it to your other students, too."

If things were different—if the 79 percent of engineers at Google who are male were specifically trained in structural oppression before building their data systems (as social workers are before they undertake social work)—then their overrepresentation might be very slightly less of a problem.[41] But in the meantime, the onus falls on the individuals who already feel the adverse effects of those systems of power to prove, over and over again, that racism and sexism exist—in datasets, in data systems, and in data science, as in everywhere else.

Buolamwini and Gebru identified how pale and male faces were overrepresented in facial detection training data. Could we just fix this problem by diversifying the data set? One solution to the problem would appear to be straightforward: create a more representative set of training and benchmarking data for facial detection models. In fact, tech companies are starting to do exactly this. In January 2019, IBM released a database of one million faces called Diversity in Faces (DiF).[42] In another example, journalist Amy Hawkins details how CloudWalk, a startup in China in need of more images of faces of people of African descent, signed a deal with the Zimbabwean government for it to provide the images the company was lacking.[43] In return for sharing its data, Zimbabwe will receive a national facial database and "smart" surveillance infrastructure that it can install in airports, railways, and bus stations.

It might sound like an even exchange, but Zimbabwe has a dismal record on human rights. Making things worse, CloudWalk provides facial recognition technologies to the Chinese police—a conflict of interest so great that the global nonprofit Human Rights Watch voiced its concern about the deal.[44] Face harvesting is happening in the US as well. Researchers Os Keyes, Nikki Stevens and Jacqueline Wernimont have shown how immigrants, abused children, and dead people are some of the groups whose faces have been used to train software—without their consent.[45] So is a diverse database of faces really a good idea? Voicing his concerns in response to the announcement of Buolamwini and Gebru's 2018 study on Twitter, an Indigenous Marine veteran shot back, "I hope facial recognition software has a problem identifying my face too. That'd come in handy when the police come rolling around with their facial recognition truck at peaceful demonstrations of dissent, cataloging all dissenters for 'safety and security.'"[46]

Better detection of faces of color cannot be characterized as an unqualified good. More often than not, it is enlisted in the service of increased oppression, greater surveillance, and targeted violence. Buolamwini understands these potential harms and has developed an approach that works across all four domains of the matrix of domination to address the underlying issues of power that are playing out in facial analysis technology. Buolamwini and Gebru first quantified the disparities in the dataset—a technical audit, which falls in the disciplinary domain of the matrix of domination. Then, Buolamwini went on to launch the Algorithmic Justice League, an organization that works to highlight and intervene in instances of algorithmic bias. On behalf of the AJL, Buolamwini has produced viral poetry projects and given TED talks—taking action in the hegemonic domain, the realm of culture and ideas. She has advised on legislation and professional standards for the field of computer vision and called for a moratorium on facial analysis in policing on national media and in Congress.[47] These are actions operating in the structural domain of the matrix of domination—the realm of law and policy. Throughout these efforts, the AJL works with students and researchers to help guide and shape their own work—the interpersonal domain. Taken together, Buolamwini's various initiatives demonstrate how any "solution" to bias in algorithms and datasets must tackle more than technical limitations. In addition, they present a compelling model for the data scientist as public intellectual—who, yes, works on technical audits and fixes, but also works on cultural, legal, and political efforts too.

While equitable representation—in datasets and data science workforces—is important, it remains window dressing if we don't also transform the institutions that produce and reproduce those biased outcomes in the first place. As doctoral health student Arrianna Planey, quoting Robert M. Young, states, "A racist society will give you a racist science."[48] We cannot filter out the downstream effects of sexism and racism without also addressing their root cause.