Background and Context for the *Our Reality* Novella

Tadayoshi Kohno

Security & Privacy Research Lab and Tech Policy Lab

Paul G. Allen School of Computer Science & Engineering

University of Washington

https://homes.cs.washington.edu/~yoshi/OurReality.html

June 6, 2021

This is a companion document to the *Our Reality* novella. This document provides background and context for the story. This document is meant for computer scientists—including computer science students and educators—as well as the general public.

I have embedded a significant amount of content in the endnotes of this document. I encourage you to read this document twice, first to absorb the overall narrative, and second to study the endnotes.

Educators are welcome to use any of the text that I have authored for this document regardless of whether or not they chose to read *Our Reality*. Since this document is a collection of references to authoritative works, educators are additionally encouraged to read and cite from those source materials as well.

Introduction

I have spent many years considering the effects of technology and technological advances on the quality of people's lives as individuals and as members of communities. As a professor at the Paul G. Allen School of Computer Science & Engineering at the University of Washington, I co-direct both the Security and Privacy Research Lab and the Tech Policy Lab, and I am the school's associate director for diversity, equity, and inclusion. I am on the advisory board for the Electronic Frontier Foundation, which focuses on issues arising from technology's impact on society. I was also an inaugural member of The National Academies of Sciences, Engineering, and Medicine's Cyber Resilience Forum, which considers the interrelationships among technology, policy, national security, and the law.

In my professional activities, I am constantly reminded about the importance of maintaining a broad context when examining technology's changing impact on how we live and work, and when examining society's impact on how technologies are designed. I wrote *Our Reality* to catalyze conversations among computer science high school, undergraduate, and graduate students (and colleagues, as well) about the roles and responsibilities of technologists as we make decisions that have many known, and—significantly—many unknown, repercussions. Through the *Our Reality* story and this companion document, I hope to instill in my computer science readers a desire to learn more about the broad downstream effects that technologies can have, and to think broadly about the role that society plays in the creation of just and unjust technologies. Through this process, I hope that we, as computer scientists, can be part of creating a better, more just world.

I also wrote the *Our Reality* story, and this companion document, with a general audience in mind—an audience ranging from middle school social studies students to adults. I hope that this companion document will provide useful background material and context, and assist the reader in thinking about some of the issues that the *Our Reality* story raises. I hope that my story and this document will be thought-provoking, motivating, and useful to readers.

I have long realized that educating the next generation of computer scientists involves far more than helping students learn how to code or to design and debug computer systems. It means helping students understand and appreciate the deep interconnections between society and computing. Computing systems can make the world a far better place for some, but a far worse one for others. A computing system can even concurrently impact an individual user *both* positively *and* negatively, at the same time. To account for these potential negative impacts, sometimes the designs of computing systems must change. Other times the most responsible action is to not build a computing system at all.

No single fictional (or factual) story can capture the full spectrum of issues that computer scientists must consider when designing computing technology for the real world. But a story can provide the backdrop for a conversation about such topics, and a starting point to learn more. *Our Reality* focuses on the relationship between people, society, and technology. Fiction, more than traditional academic computer science texts and papers, can highlight the importance, nuance, and delicacy of that relationship. Stories can be personal, intimate, and emotional—and hence thought-provoking and memorable—in ways that textbooks cannot.

Further, science fiction, as we know from its rich history of examining the embedding of science and invention in daily life, gives readers the freedom to explore the implications of alternate future worlds.[1] For example, as a reader of *Our Reality*, you might ask yourself: What different decisions could the designers of any of the technologies in *Our Reality* have made, and what would have been the implications of taking these different paths?

More on Mixed-Reality and Other Technologies in *Our Reality*

The main technology featured in this narrative is *mixed-reality*. In the story, people who can afford to do so purchase special mixed-reality Goggles to wear on their heads, over their eyes. When wearing these Goggles, they can still see the real world, but they can also see virtual (computer-generated content) overlaid on top of it. This interleaving of virtual content with the real world is the definition of mixed-reality.

Parts of the population—though only those with the financial resources to do so—have moved entirely online, into a mixed-reality world called *Our Reality*.[2] Parents who buy this virtual, alternative reality for their children believe that the educational opportunities it provides exceed those that a physical-world school can offer. And these parents also believe that online schools will keep their children safe in the event of another pandemic or from other harsh realities of the physical world.

Mixed-reality technologies, like Goggles, can deeply impact how people interact with others and the world around them. In collaboration with Professor Franziska Roesner, a colleague at the University of Washington and co-director of the Security and Privacy Research Lab, we and our students and colleagues have spent years considering the potential benefits and harms of mixed-reality systems (see https://ar-sec.cs.washington.edu/ for our publications in this space). *Our Reality* uses fiction to bootstrap conversation about these harms-benefits tradeoffs.

In addition to mixed-reality systems, this story features numerous other technologies that might similarly have pros and cons for individuals and society, including: (1) contact-tracing cameras in streetlights,[3] (2) online advertisements and the associated tracking (surveillance) of online activities that enables targeted advertising, (3) technologies for surveillance in schools, (4) technologies that parents can use to monitor their children, and (5) autonomous drones with some level of artificial intelligence. Many of these technologies may sound familiar and exist in some form today. Our Reality explores aspects of what they may look like in 2034.[4]

Elements of this story also invite conversations about other key issues, like the accessibility of future technologies for people with disabilities, and the future of the so-called digital divide, where some people have access to technology and the means to use it while others do not or cannot. Other important issues exist but are not covered by this story, including geological and climatological impacts associated with technology production (Crawford, 2021, pp. 23-51),[5] issues associated with the human labor involved in sustaining technological capabilities (Crawford, 2021, pp. 63-69),[6] and the harvesting of data of and about people toward the training of artificial intelligence systems (Crawford, 2021, pp. 89-121).[7]

More on Racism in *Our Reality*

The computer science community has, like society at large, been deficient in its understanding of and focus on racism. I, like much of the community, needed to learn more about racism and technology, and about racism in general, to improve as an educator and to write elements of this story.

I focus much of the remainder of this document on racism and racism in technology because I consider them to be areas of significant and critical need for growth and maturation within the computer science field.

The two protagonists in *Our Reality* are Emma, a 16-year old Black cisgender[8] woman, and Liam, a 16-year old white cisgender man. The antagonist in *Our Reality* is the world as it appears in 2034. My own identity is majority Japanese and white, minority Native, and cisgender, and I felt mostly confident in my ability to write Liam's character. To bring Emma to the page, I read and was inspired by an excellent book, *Writing the Other: A Practical Approach*, by Nisi Shawl and Cynthia Ward (Shawl and Ward, 2005). I subsequently enrolled in a Hugo House course offered by Nisi Shawl and then used her book and course as guides to developing Emma's character and surfacing other issues in *Our Reality*. I know that as a Japanese, white, and Native cisgender man, I have inevitably failed in some of the ways that I wrote Emma and her family. With the guidance of Nisi Shawl's and Cynthia Ward's book, and Nisi Shawl's course, I hope that I have not failed entirely.

For the remainder of this preface, I use the definition of racism provided in Beverly Daniel Tatum's book, *Why Are All the Black Kids Sitting Together in the Cafeteria?* Quoting from earlier work by David Wellman, this book's definition of racism is "a 'system of advantage based on race'" (Tatum, 2017, p. 87). Tatum then writes, "This definition of racism is useful because it allows us to see that racism, like other forms of oppression, is not only a personal ideology based on racial prejudice but a system involving cultural messages and institutional policies and practices as well as the beliefs and actions of individuals." This definition can thus apply to technologies and policies, in addition to people. Ruha Benjamin, in the book *Race After Technology*, adds, "Too often people assume that racism and other forms of bias must be triggered by an *explicit* intent to harm" (Benjamin, 2019, p. 59). Racism under this definition can manifest both with and without such an intent.[9]

Consider, as an example, a *machine learning* (ML) system designed to detect faces in photos. Such an ML system is an example of an *artificial intelligence* (AI) system. All ML systems must be *trained*, and this ML system must be trained on a set of photos, some with faces and some without. The ML system "learns" from this training. Now suppose that the creators of this ML system found a conveniently available photo dataset to use for training. Further suppose that most or all of the photos of people in this dataset are of white people. As such, the ML system, when deployed, will likely do poorly at detecting the faces of Black people in photos even if it does a very good job at detecting those of white people.[10]

In this example, the ML system has resulted in a "system of advantage based on race." In other words, the resulting ML system is racist under the definition provided above. The previous statement is true *even if the designers of the system had no explicit intent to harm and were not trying to be racist*—they used a convenient dataset, and the fact that the training dataset did not include Black people could have been an accidental oversight. If such an oversight is common across multiple systems, then the oversight becomes *systemic*—i.e., there is systemic racism in those technologies. Again, the previous statement is true even if there was no explicit intent to harm.[11] It is important to note, however, that the omission of Black people is not always accidental; it is sometimes a conscious decision.[12]

Books like Ruha Benjamin's *Race After Technology* (Benjamin, 2019), Sarah Brayne's *Predict and Surveil* (Brayne, 2021), Meredith Broussard's *Artificial Unintelligence* (Broussard, 2019), Kate Crawford's *Atlas of AI* (Crawford, 2021), Virginia Eubanks's *Automating Inequality* (Eubanks, 2018), Safiya Umoja Noble's *Algorithms of Mass Oppression* (Noble, 2018), Cathy O'Neil's *Weapons of Math Destruction* (O'Neil, 2016), and Sara Wachter-Boettcher's *Technically Wrong* (Wachter-Boettcher, 2017) make it clear that there is systemic racism in today's technologies.

Unlike the previous example, systemic racism in technology is not always so easy to spot or understand. In the face detection example, the designers trained the ML system with a dataset that included only white people. Other algorithms might be trained with different datasets, where the inequities in or associated with those datasets may be less obvious, at least at first blush.

For example, consider software that police might use to predict which regions of a city have a high likelihood of future crime. If those predictions are based on datasets of historical crime statistics, then those predictions will embody the systemic biases and inequities that led to those historical crime statistics in the first place (Benjamin, 2019, pp. 80-84).[13] As another example, consider an ML system that attempts to predict beauty but that is trained on people's biased perceptions of beauty (Benjamin, 2019, p. 51).[14] Consider also a system that computes a credit risk score based on historical information about other people, where that historical information embodies historical biases and injustices (Benjamin, 2019, p. 60).[15] As yet another example, in discussing robotics, Ruha Benjamin in the book *Race After Technology* writes, "To the extent that machine learning relies on large, 'naturally occurring' datasets that are rife with racial (and economic and gendered) biases, the raw data that robots are using to learn and make decisions about the world reflect deeply ingrained cultural prejudices and structural hierarchies." (Benjamin, 2019, p. 59). In short, there are countless ways in which biases can manifest in training sets and the resulting systems.

Problems can also arise when the data being used as input to make predictions is biased. Consider a software tool designed to predict future child abuse and neglect. Suppose that this software uses community reports (e.g., someone calling in to report suspected child abuse or neglect) as input to its prediction algorithm. These community reports, which may reflect racial biases or may be intentionally false, can inject significant racial biases into the system (Eubanks, 2018, pp. 152-155).

Problems can be present for other reasons, as well. For example, if the results of an automated system are unjust but there is no feedback loop to detect the injustice and correct the system, then the system will remain unjust (O'Neil, 2016, pp. 97-100).[16] If the outputs of the system are used as subsequent inputs to the system, there can be negative feedback loops that perpetuate or increase biases (Brayne, 2021, pp. 107-109).[17] Sometimes, the design goals of a system may, themselves, be unjust; in this situation, the resulting system will be unjust *regardless* of how the system is built (O'Neil, 2016, pp. 129-130).[18] Relatedly, the criteria upon which a training set is created may be unjust (Crawford, 2021, pp. 131-133).[19]

Further, even the mere *existence* of a digital system can create injustices for people who do *not* interact with the system. First, inequities can arise as people try to avoid the system. For example, "individuals wary of police surveillance" technologies may "deliberately and systematically [avoid] institutions that keep formal records, such as hospitals, banks, schools, and employment, to avoid coming under heightened police surveillance" (Brayne, 2021, pp. 114-115).[20] Second, injustices can also arise when a system influences what other people think or believe about an individual, even when that individual never uses the system, e.g., when a search engine's results perpetuate stereotypes (Noble, 2018, p. 155) or contributes to people forming a false impression about the world (Noble, 2018, pp. 110-111).[21]

Additionally, injustices can also arise when a system's outputs are highly accurate (by some definition of accurate); a high level of (perceived) accuracy can result in people trusting and believing (and then acting upon) the outputs of a system even when the outputs are incorrect (Brayne, 2021, p. 124).[22]

Our Reality provides a basis for continued discussions about the following aspects of racism: (1) microaggressions,[23] (2) intersectionality,[24] (3) tokenism[25], (4) the existence of a white and Asian male majority, and the lack of diverse perspectives, when designing technical systems,[26] (5) the resulting presence of biases and injustices in the design of technical systems,[27] (6) that, due to the systemic nature of some of the biases and injustices in technical systems, increasing diversity within technology design teams is not the complete solution,[28] (7) subconscious or implicit biases,[29] (8) assuming that members of marginalized populations have their positions (e.g., in a university or a company) because of their race or gender rather than because they have earned them, (9) the perception and implications of white as the default user or persona,[30] (10) the difference between intention and impact,[31] (11) the importance of understanding existing racial inequities and the needs and values of marginalized populations when designing technologies,[32] (12) racism in healthcare and healthcare inequities,[33] (13) how people often assume that technologies are neutral and objective even when they are not,[34] (14) how the creators of systems may not reveal the details of their algorithms to the public and to those negatively impacted by those technologies,[35] (15) racism and the excessive use of force by law enforcement,[36] (16) how excessive force is often blamed on victims rather than on perpetrators, (17) risks with shifting law enforcement activities to biased and unjust technologies,[37] (18) how police are using technologies controlled by the private sector,[38] and (19) that too often technologies are designed and deployed without fully anticipating and understanding the potential unintended consequences.[39]

I also use *Our Reality* to intentionally counter some common injustices. For example, Emma's mother is a leader in the technology field despite the current field being majority male and white and Asian.

As with my coverage of potential future technologies, my coverage of important aspects of racism is incomplete. Though sufficiently deep to encourage further reflection, discussion, and inquiry by an interested reader, *Our Reality* merely scratches the surface of these topics. Many important aspects of racism are not covered by this narrative, and the negative impacts of racism are significantly greater than what this story captures.

The U.S. National Science Foundation states that "Women, persons with disabilities, and underrepresented minority groups—blacks or African Americans, Hispanics or Latinos, and American Indians or Alaska Natives—are underrepresented in science and engineering (S&E)."[40] This quote should serve as a reminder that the inequities and injustices in the computing field, in science and engineering in general, and in the world at large are also much broader than a single story can cover. For example, racism can manifest in different ways for different populations.[41] Further, the preceding list omits the full spectrum of sexuality and gender. It also does not consider the roles of culture, ethnicity, religion, and numerous other elements of identity, nor does it underscore the importance of considering intersectionality.[42]

On Story Structure

I wrote *Our Reality* following a traditional Japanese four-act story structure known as *kishōtenketsu* (起承転結). Quoting from Wikipedia, a story under this structure has four parts (italics added):

- Introduction (*ki*): introducing characters, era, and other important information for understanding the setting of the story.
- Development (*shō*): follows leads towards the twist in the story. Major changes do not occur.
- Twist (*ten*): the story turns toward an unexpected development. This is the crux of the story, the yama (ヤマ) or climax. In case of several turns in the narrative, this is the biggest one.
- Conclusion (*ketsu*), also called ochi (落ち) or ending, wraps up the story. [43]

In *Our Reality*, *ki* and *shō* correspond to Chapters 1-2 and Chapters 3-6, respectively. *Ten* corresponds to Chapter 7, and *ketsu* corresponds to Chapter 8.

Most of the technologies mentioned in the "More on Mixed-Reality and Other Technologies in *Our Reality*" section above are surfaced and featured in Chapters 1-6. While visible in Chapters 1-6, racism and racism in technology take center stage in Chapter 7. Chapter 8 embeds a significant amount of educational content, much which is discussed in more detail in this document; see especially the "More on Racism in *Our Reality*" section above and the associated endnotes.

Additional Resources

I include a "References and Suggested Readings" section later in this document. I encourage interested readers to consume as many of the books in the "References and Suggested Readings" section as possible. I also encourage readers to seek out other sources of information, e.g., I encourage the reader to watch the film *Coded Bias* (https://www.codedbias.com/about). There are also numerous other excellent materials available (e.g., other books beyond what I list in the "References and Suggested Reading" section, online talks, online videos, podcasts, and more). For example, I encourage readers to review the web pages for the AI Now Institute (https://ainowinstitute.org/), the Algorithmic Justice League (https://www.ajl.org/), Black in AI (https://blackinai.github.io/), and Data and Society (https://datasociety.net/), among many other online resources.

Since there are many books listed in the "References and Suggested Readings" section of this document, a reader may wonder: which book should I read first? My philosophy is that any additional learning is valuable learning, and hence one answer is: whichever book interests you the most. Still, if you would like a suggested reading list, in-order, I make the following suggestions. I recommend reading Ijeoma Oluo's *So You Want to Talk about Race* book first (Oluo, 2019). It is highly accessible and informative, which may help you to interpret the books that you subsequently pursue. A computer scientist might be tempted to read about racism and technology first rather than focus on racism in general. I believe that computer scientists must understand and be informed about the world to do their jobs well. Hence—to the computer scientists who are interested in eventually reading a book about racism and technology—I still encourage you to start with Oluo.

After reading Oluo, if you are interested in a book focused specifically on racism and technology, I recommend Ruha Benjamin's *Race After Technology* (Benjamin, 2019). For a book broadly focused on injustices with computational systems and algorithms, I recommend Cathy O'Neil's *Weapons of Math Destruction* (O'Neil, 2016). I also recommend the book *Technically Wrong* by Sara Wachter-Boettcher (Wachter-Boettcher, 2017). I consider the Oluo, Benjamin, O'Neil, and Wachter-Boettcher sequence to be reasonable texts for a self-guided "introductory course" for computer scientists on racism, justice, and technology. I also recommend watching the film *Coded Bias*, which is directed and produced by Shalini Kantayya and which centers MIT Media Lab researcher Joy Buolamwini (https://www.codedbias.com/about).

If you are looking for recommendations of books to read next, for your self-guided "second course," then I make the following additional recommendations: Beverly Daniel Tatum's *Why Are All the Black Kids Sitting Together in the Cafeteria?* (Tatum, 2017), Caroline Criado Perez's *Invisible Women* (Perez, 2019), Alice Wong's (editor) *Disability Visibility* (Wong, 2020), Virginia Eubanks's *Automating Inequality* (Eubanks, 2018), Sarah Brayne's *Predict and Surveil* (Brayne, 2021), and Safiya Umoja Noble's *Algorithms of Oppression* (Noble, 2018). I have selected these texts for the "second course" because, together, they provide additional depth on racism and breadth on inequity. Specifically, Tatum's book complement Oluo's book and is an excellent reading on the topic of racism, Perez's book provides an extensive discussion of the inequities that women face, Wong's book provides an excellent introduction to intersectionality and the inequities that people with disabilities face, Eubanks's book provides rich case studies of the injustices facilitated by automated systems as well as the historical contexts in which modern automated systems exist, Brayne's book provides a thorough study of injustices and inequities related to law enforcement technologies, and Noble's book provides a deep study of the injustices and inequities on the web. If I were to add two more books to this "second course," they would be: Michelle Alexander's *The New Jim Crow* (Alexander, 2020) and Isabel Wilkerson's *Caste* (Wilkerson, 2020).

In preparing the above self-guided syllabus, I struggled because all of the books in the "References and Suggested Readings" section have so much to offer. In many cases, I thought "this should be one of the first books that readers read!" Rather than simply suggest that readers read all the books in the "References and Suggested Readings" section, I have provided one possible syllabus, which unfortunately omits excellent books (and hence omits other important topics). I encourage you to use my suggested reading list as a guide in selecting the books that you read, not as the definitive order in which you should read these books. I am also sure that I have unintentionally omitted many other excellent books from the "References and Suggested Readings" list. I encourage you to continually strive to find more resources, and to never stop striving to learn more.

Additionally, if you are a computer science educator, from any country in the world, then I strongly encourage you to apply to the Cultural Competence in Computing (3C) Fellows Program, created by Dr. Nicki Washington, Dr. Shaundra B. Daily, and graduate assistant Cecilé Sadler at Duke University. I am a member of the first 3C Fellows cohort (2020-2021). All aspects of the program—the structure, the preparatory materials, the guest lectures, the group discussions—are invaluable. I consider enrolling in the program to be a service to oneself, their students, and—through their students—society as a whole. Information about their program is available online at: https://identity.cs.duke.edu/fellows.html.

Note to Educators

This document is less of an original piece, and more of a compilation of references to many excellent works by others (see the "Additional Resources" section above and the "References and Suggested Readings" section below). Educators are welcome to use of any of content that I have authored for this document regardless of whether or not they (or their students) read *Our Reality*. Much of the educational content in this document is embedded in the endnotes. One usage model that I envision for educators: (1) read this document in its entirety, including the endnotes, to develop an overall understanding of the content herein, (2) formulate learning objectives, (3) identify the topics surfaced in the main body of this document that one wishes to discuss with their students, (4) refer to the endnotes for detailed examples and pointers to additional resources, and (5) review and reference the original source materials for additional information.

If you choose to read the *Our Reality* novella with your students, then please also see the guiding questions at the end of the story. An understanding of the content in this document will help with the interpretation of the material in the *Our Reality* novella.

Summary

There is systemic racism (and other isms) in today's technologies, as there is in today's society. These isms can result in and perpetuate significant injustices. *Our Reality* is intended to provide a vehicle for readers to explore how these isms might manifest in a future world where we have not sufficiently challenged and overcome them. As individuals and as society, we must challenge and overcome these isms. I hope that you find the *Our Reality* story, this companion document, and the suggested readings to be thought-provoking, educational, and useful as you chart and navigate our path forward.

Acknowledgements

So many people contributed in so many ways to making the *Our Reality* novella and this companion document possible. I thank all of you sincerely: Helen Anderson, Mariko Blessing, Brian Blickenstaff, Alex Bolton, Lauren Bricker, Ryan Calo, Megan Christy, Kelly Clark, Kate Cohen, Kimiko Coleman, Alexei Czeskis, Cory Doctorow, Erica Erickson, Ivan Evtimov, Gennie Gebhart, Sue Glueck, Leah H., Elise Hooper, Jesse Edward Johnson, Sandy Kaplan, Amy J. Ko, David Kohlbrenner, Sean Kohno, Seth Kohno, Taryn Kohno, Kiron Lebeck, Kevin Lin, Nick Logler, Lindsay Lopez, Lassana Magassa, Sarah E. McQuate, Zoë Mertz, David M. Mills, Kristin Osborne, Mark Pearson, Franziska Roesner, Nisi Shawl, Brett Werenski, and Wendy Yim. I also thank Dr. Nicki Washington, Dr. Shaundra B. Daily, and graduate assistant Cecilé Sadler at Duke University for creating the Cultural Competence in Computing (3C) Fellows Program, and for giving me the opportunity to continue to learn and grow as an educator and as a member of the computing field. Thank you also to the 3C guest lecturers. And thank you to all the other members of the first 3C cohort for the amazing discussions and for the knowledge, insights, experiences, and expertise that you have shared. I highly encourage other computer science educators to apply to be members of future cohorts.

References and Suggested Readings

[1] Michelle Alexander. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. The New Press, 10th anniversary edition, 2020.

[2] Ruha Benjamin. *Race After Technology*. Polity Press, 2019.

[3] Eduardo Bonilla-Silva. *Racism Without Racists: Color-blind Racism and the Persistence of Racial Inequality in America*. Rowman & Littlefield, fifth edition, 2017.

[4] Sarah Brayne. *Predict and Surveil: Data, Discretion, and the Future of Policing*. Oxford University Press, 2021.

[5] Meredith Broussard. *Artificial Unintelligence: How Computers Misunderstand the World*. MIT Press, 2019.

[6] Ryan Calo, Batya Friedman, Tadayoshi Kohno, Hannah Almeter, and Nick Logler, editors. *Telling Stories: On Culturally Responsive Artificial Intelligence*. University of Washington Tech Policy Lab, 2020.

[7] Sasha Costanza-Chock. *Design Justice: Community-led Practices to Build the Worlds We Need*. The MIT Press, 2020.

[8] Kate Crawford. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.

[9] Robin DiAngelo. *White Fragility: Why It's So Hard for White People to Talk about Racism*. Beacon Press, 2018.

[10] Virginia Eubanks. *Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press, 2018.

[11] Ibram X. Kendi. *How to be an Antiracist*. One World, 2019.

[12] Audre Lorde. *Sister Outsider: Essays and Speeches*. Ten Speed Press, 1984.

[13] Charlton D. McIlwain. *Black Software: The Internet and Racial Justice, from the AfroNet to Black Lives Matter*. Oxford University Press, 2020.

[14] Safiya Umoja Noble. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press, 2018.

[15] Ijeoma Oluo. *So You Want to Talk about Race*. Seal Press, 2018.

[16] Cathy O'Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown Publishers, 2016.

[17] Caroline Criado Perez. *Invisible Women: Data Bias in a World Designed for Men*. Abrams Press, 2019.

[18] Nisi Shawl and Cynthia Ward. *Writing the Other: A Practical Approach*. Aqueduct Press, 2005.

[19] Kamal Sinclair and Jessica Clark. *Making a New Reality: A Toolkit for Inclusive Media Futures*. 2020.

[20] Beverly Daniel Tatum. *Why Are All the Black Kids Sitting Together in the Cafeteria? And Other Conversations About Race*. Basic Books, 2017.

[21] Sara Wachter-Boettcher. *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech*. W. W. Norton & Company, 2017.

[22] Isabel Wilkerson. *Caste: The Origins of Our Discontents*. Random House, 2020.

[23] Alice Wong, editor. *Disability Visibility: First-Person Stories from the Twenty-First Century*. Vintage Books, 2020.

[24] Shoshana Zuboff. *The Age of Surveillance Capitalism*. PublicAffairs, 2019.

End Notes

[1] Ruha Benjamin, in *Race After Technology*, writes "If, as many have argued, the rhetoric of human betterment distorts an understanding of the multifaceted interplay between technology and society, then a thoroughgoing commitment to justice has the potential to clarify and inspire possibilities for designing this relationship anew. Justice, in this sense, is not a static value but an ongoing methodology that can and should be incorporated into tech design. For this reason, too, it is vital that people engaged in tech development partner with those who do important sociocultural work honing narrative tools through the arts, humanities, and social justice organizing." (Benjamin, 2019, pp. 192-193). Benjamin then offers a 2018 quote from Kamal Sinclair. The following is a thematically similar and more recent quote from Kamal Sinclair and Jessica Clark, from their book *Making a New Reality*: "Story and narrative inform how we design everything from technology to social systems. They shape the norms in which we perform our identities, even perhaps the mutations of our DNA and perceptions of reality. Stories are the first step in the process of how we imagine our reality, and help us understand how it is shaped—and whether we can help shape it." (Sinclair and Clark, 2020, p. 28).

Meredith Broussard in *Artificial Unintelligence* writes that "A great deal of what people dream about making in tech is shaped by the images they see in movies, TV programs, and books." (Broussard, 2019, p. 10). I hope that *Our Reality* contributes positively to such shaping.

[2] The fictitious product name "Our Reality" derives from the terms "augmented reality," "mixed-reality," and "virtual reality." The name "Our Reality" is also meant to raise the question of what

"our" in "Our Reality" means, and to whom something can be "our" if not everyone has access to the technology and if someone else (not the individual user) designs the world.

[3] If a person has a highly infectious disease, health officials could use the videos recorded by these cameras for contact tracing, i.e., to determine who that person might have infected. As envisioned in this story, the contact-tracing camera system employs a face recognition subsystem.

[4] I base my descriptions of these technologies on some of my past and present research interests and results. I provide links to some of these research results below, all of which were done in collaboration others.

The discussions in *Our Reality* surrounding *contact-tracing cameras in streetlights* derives from my past research studying contact-tracing and privacy (https://arxiv.org/abs/2012.01553), privacy risks with face recognition technologies (https://arxiv.org/abs/2012.08588), and city-wide detection of IMSI catcher-based surveillance technologies (https://seaglass.cs.washington.edu/).

The focus in *Our Reality* on *online advertisements and the associated tracking (surveillance) of online activities that enables targeted advertising* derives from the research that Professor Franziska Roesner and I conducted, with our students and colleagues, studying the online advertising ecosystem (https://homes.cs.washington.edu/~yoshi/papers/webtracking-NSDI2012.pdf, https://trackingobserver.cs.washington.edu/, https://trackingexcavator.cs.washington.edu/, https://adint.cs.washington.edu/, https://badads.cs.washington.edu/). Our research group has also considered how online tracking and advertising might manifest in augmented reality environments; see https://ar-sec.cs.washington.edu/ for our publications in the space of augmented reality computer security and privacy.

The focus in *Our Reality* on *technologies that parents can use to monitor their children* derives from some of my past work on parent-child safety technologies (https://homes.cs.washington.edu/~yoshi/papers/SOUPS/soups2010.pdf), as well as on my past research on how family members and friends communicate and what information is revealed by the underlying communications technologies (https://homes.cs.washington.edu/~yoshi/papers/CHI_OSIs.pdf and https://homes.cs.washington.edu/~yoshi/papers/PETS_OSIs.pdf).

More generally, much of my research focuses on assessing, understanding, and mitigating risks with emerging technologies. I applied this mindset as I tried to envision the technologies that might appear in 2034, and what their risks might be. For example, I have conducted research on computer security and privacy risks with automobiles (http://www.autosec.org/), wireless medical devices (e.g., see https://homes.cs.washington.edu/~yoshi/papers/IMD/NEJM-Maisel-Kohno.pdf for an overview and https://homes.cs.washington.edu/~yoshi/papers/icd-study.pdf for one of our first papers), DNA synthesis, sequencing, and applications (http://cybio.cs.washington.edu/), computer vision systems (https://arxiv.org/pdf/1707.08945.pdf), brain-machine interface technologies (https://thejns.org/focus/view/journals/neurosurg-focus/27/1/article-pE7.xml), and smarthome technologies (e.g., see https://cacm.acm.org/magazines/2013/1/158768-computer-security-and-the-modern-home/fulltext for one our early works in this field). Additionally, much of my research focuses on the relationship between people, society, policy, and technology, e.g., through studies at the intersection of the human-computer interaction and computer security research fields.

[5] In *Atlas of AI*, Kate Crawford writes that "Computational media now participate in geological (and climatological) processes, from the transformation of the earth's materials into infrastructures and devices to the powering of these new systems with oil and gas reserves. Reflecting on media and technology as geological processes enables us to consider the radical depletion of non-renewable resources required to drive the technologies of the present moment. Each object in the extended network of an AI system, from network routers to batteries to data centers, is built using elements that required billions of years to form inside the earth." (Crawford, 2021, p. 31).

Crawford then writes that "From the perspective of deep time, we are extracting Earth's geological history to serve a split second of contemporary technological time, building devices like the Amazon Echo and the iPhone that are often designed to last for only a few years." (Crawford, 2021, p. 31).

Toward explaining the phenomenon of overlooking the geological and climatological impacts of technology production, Crawford observes that "It is a common practice of life to focus on the world immediately before us, the one we see and smell and touch every day. It grounds us where we are, with our communities and our known corners and concerns. But to see the full supply chains of AI requires looking for patterns in a global sweep, a sensitivity to the ways in which the histories and specific harms are different from place to place and yet are deeply interconnected by the multiple forces of extraction." (Crawford, 2021, p. 38).

[6] Kate Crawford, in *Atlas of AI*, writes that "Many forms of work are shrouded in the term 'artificial intelligence,' hiding the fact that people are often performing rote tasks to shore up the impression that machines can do the work. But large-scale computation is deeply rooted in and running on the exploitation of human bodies." (Crawford, 2021, p. 57).

Crawford later writes that "Exploitative forms of work exist at all stages of the AI pipeline, from the mining sector, where resources are extracted and transported to create the core infrastructure of AI systems, to the software side, where distributed workforces are paid pennies per microtask. Mary Gray and Sid Suri refer to such hidden labor as 'ghost work.' Lilly Irani calls it 'human-fueled automation.' These scholars have drawn attention to the experiences of crowdworkers or microworkers, who perform the repetitive digital tasks that underlie AI systems, such as labeling thousands of hours of training data and reviewing suspicious or harmful content. Workers do the repetitive tasks that backstop claims of AI magic—but they rarely receive credit for making the systems function." (Crawford, 2021, p. 63).

The following is among the examples that Crawford provides: "Sometimes workers are directly asked to pretend to be an AI system. The digital personal assistant start-up x.ai claimed that its AI agent, called Amy, could 'magically schedule meetings' and handle many mundane daily tasks. But a detailed Bloomberg investigation by journalist Ellen Huet revealed that it wasn't artificial intelligence at all. 'Amy' was carefully being checked and rewritten by a team of contract workers pulling long shifts. Similarly, Facebook's personal assistant, M, was relying on regular human intervention by a group of workers paid to review and edit every message." (Crawford, 2021, p. 65). Crawford continues, "Faking AI is an exhausting job. The workers at x.ai were sometimes putting in fourteen-hour shifts of annotating emails in order to sustain the illusion that the service was automated and functioning 24/7. They couldn't leave at the end of the night until the queues of emails were finished. 'I left felling totally numb and absent of any sort of emotion,' one employee told Huet." (Crawford, 2021, p. 65).

[7] In *Atlas of AI*, Kate Crawford writes that "I've looked at hundreds of datasets over years of research into how AI systems are built, but the NIST mug shot databases are particularly disturbing because they represent the model of what was to come. It's not just the overwhelming pathos of the images themselves. Nor is it solely the invasion of privacy they represent, since suspects and prisoners have no right to refuse being photographed. It's that the NIST databases foreshadow the emergence of a logic that has now thoroughly pervaded the tech sector: the unswerving belief that everything is data and is there for the taking. It doesn't matter where a photograph was taken or whether it reflects a moment of vulnerability or pain or if it represents a form of shaming the subject. It has become so normalized across the industry to take and use whatever is available that few stop to question the underlying politics." (Crawford, 2021, p. 93). Turning to another dataset (ImageNet), Crawford writes that "Over the course of a decade, ImageNet grew into a colossus of object recognition for machine learning and a powerfully important benchmark for the field. The approach of mass data extraction without consent and labeling by underpaid crowdworkers would become standard practice, and hundreds of new training datasets would follow ImageNet's lead." (Crawford, 2021, p. 109).

Crawford observes that “Machine learning systems are trained on images like these every day—images that were taken from the internet or from state institutions without context and without consent. They are anything but neutral. They represent personal histories, structural inequities, and all the injustices that have accompanied the legacies of policing and prison systems in the United States.” (Crawford, 2021, p. 94). Crawford later writes that “The AI industry has fostered a kind of ruthless pragmatism, with minimal context, caution, or consent-driven data practices while promoting the idea that the mass harvesting of data is necessary and justified for creating systems of profitable computational ‘intelligence.’ This has resulted in a profound metamorphosis, where all forms of image, text, sound, and video are just raw data for AI systems and the ends are thought to justify the means. But we should ask: Who has benefited most from this transformation, and why have these dominant narratives of data persisted? And as we saw in the previous chapters, the logic of extraction that has shaped the relationship to the earth and to human labor is also a defining feature of how data is used and understood in AI. By looking closely at training data as a central example in the ensemble of machine learning, we can begin to see what is at stake in this transformation.” (Crawford, 2021, p. 95).

[8] The term “cisgender” here means that Emma’s personal identity and gender corresponds to her birth sex.

[9] The literature offers other definitions of racism. For example, Audre Lorde in *Sister Outsider* defines racism as “The belief in the inherent superiority of one race over all others and thereby the right to dominance.” (Lorde, 1984, p. 45). Ijeoma Oluo’s book *So You Want to Talk about*

Race defines racism as “any prejudice against someone because of their race, when those views are reinforced by systems of power.” (Oluo, 2018, p. 26). Ibram X. Kendi, in *How to be an Antiracist*, defines racism as “a marriage of racist policies and racist ideas that produces and normalizes racial inequities.” (Kendi, 2019, pp. 17-18). Kendi also notes the significant value gained by having precise definitions. Kendi writes “Definitions anchor us in principles” and “Some of my most consequential steps toward being an antiracist have been the moments when I arrived at basic definitions.” (Kendi, 2019, p. 17). Kendi also provides an array of definitions, such as “biological racist,” “ethnic racism,” and “cultural racist;” the precision of these terms can help focus a conversation around the different ways in which racism can manifest (Kendi, 2019). Eduardo Bonilla-Silva also discusses different definitions of racism in the book *Racism without Racists*. Bonilla-Silva observes that “One reason why, in general terms, whites and people of color cannot agree on racial matters is because they conceive terms such as ‘racism’ very differently. Whereas for most whites racism is prejudice, for most people of color racism is systemic or institutionalized.” (Bonilla-Silva, 2017, p. 8). Bonilla-Silva also discusses how racism and terminology have evolved over time (Bonilla-Silva, 2017). Robin DiAngelo, in *White Fragility*, discusses different manifestations of racism after the civil rights movement (DiAngelo, 2018, pp. 39-50). DiAngelo provides text that can help the reader understand how the definition of racism in *White Fragility* (similar to the one that I use in this writing) is broader than the definition that many people hold (DiAngelo, 2018, p. 13). Isabel Wilkerson in *Caste* also discusses how the definition of racism has changed over time (Wilkerson, 2020, pp. 68-72). Wilkerson provides an additional term, “casteism,” relates the definitions of racism and casteism

(Wilkerson, 2020, p. 70), and notes that the "issue of caste was, to my mind, the basis of every other *ism*" (Wilkerson, 2020, p. 171).

For those seeking to learn more about racism, it is valuable to study, learn from, and reflect deeply on these and other sources. For this writing, I have chosen a definition of racism that applies to (at least) individuals, groups of people, societies, policies, and technologies and that acknowledges that racism can manifest without an explicit intent to harm.

[10] This example is based on incidents with real, deployed technologies. In the book *Algorithms of Oppression*, Safiya Umoja Noble writes that "These data aberrations have come to light in various forms. In 2015, *U.S. News and World Report* reported that a 'glitch' in Google's algorithm led to a number of problems through auto-tagging and facial recognition software that was apparently intended to help people search through images more successfully. The first problem for Google was that its photo application had automatically tagged African Americans as 'apes' and 'animals.'" (Noble, 2018, p. 6). See also the 2020 film *Coded Bias*, directed and produced by Shalini Kantayya. Quoting from https://www.codedbias.com/about (accessed December 2020): "CODED BIAS follows MIT Media Lab researcher Joy Buolamwini's startling discovery that many facial recognition technologies fail more often on darker-skinned faces, and delves into an investigation of widespread bias in artificial intelligence." See also the resources provided by the Algorithmic Justice League, founded by Joy Buolamwini: https://www.ajl.org/ (accessed December 2020).

[11] In *Race After Technology*, Ruha Benjamin writes that "it is important for us to assess how technology can reinforce bias by what it does, regardless of marketing or intention." (Benjamin, 2019, p. 22-23). Benjamin also writes that "Still, the view that ill intent is always a feature of racism is common: 'No one at Google giggled while intentionally programming its software to mislabel black people.' Here McWhorter is referring to photo-tagging software that classified dark-skinned users as 'gorillas.' Having discovered no bogeyman behind the screen, he dismisses the idea of 'racist technology' because that implies 'designers and the people who hire them are therefore "racists."' But this expectation of individual intent to harm as evidence of racism is one that scholars of race have long rejected." (Benjamin, 2019, p. 61).

In writing about the difference between intention and impact, Sasha Costanza-Chock in the book *Design Justice* writes that "Most designers today do not intend to systematically exclude marginalized groups of people. However, power inequalities as instantiated in the affordances and disaffordances of sociotechnical systems may be intentional or unintentional, and the consequences may be relatively small, or they may be quite significant." (Costanza-Chock, 2020, p. 40). Costanza-Chock then writes that "Most designers, most of the time, do not think of themselves as sexist, racist, homophobic, xenophobic, Islamophobic, ableist, or settler-colonialist. Some may consider themselves to be capitalist, but few identify as part of the ruling class. Many feel themselves to be in tension with capitalism, and many may even identify as socialist. However, design justice is not about intentionality; it is about process and outcomes. Design justice asks whether the affordances of a designed object or system disproportionally reduce opportunities for already oppressed groups of people while enhancing the life opportunities of dominant groups, independently of whether designers intend this outcome." (Costanza-Chock, 2020, p. 41).

[12] Ruha Benjamin recounts the following story in *Race After Technology*: "A former Apple employee who noted that he was 'not Black or Hispanic' described his experience on a team that was developing speech recognition for Siri, the virtual assistant program. As they worked on different English dialects—Australian, Singaporean, and Indian English—he asked his boss: 'What about African American English?' To this his boss responded: 'Well, Apple products are for the premium market.'" (Benjamin, 2019, p. 28).

[13] In the book *Race After Technology*, Ruha Benjamin writes that "As a form of social technology, institutional racism, past and present, is the precondition for the carceral technologies that underpin the US penal system. At every stage of the process—from policing, sentencing, and imprisonment to parole—automated risk assessments are employed to determine people's likelihood of committing a crime. They determine the risk profile of neighborhoods in order to concentrate police surveillance, or the risk profile of individuals in order to determine whether or for how long to release people on parole." (Benjamin, 2019, p. 81).

Benjamin later writes that "Likewise, predictive policing software will always be more likely to direct police to neighborhoods like the one I grew up in, because the data that this software is drawing from reflect ongoing surveillance priorities that target predominantly Black neighborhoods. Anti-Blackness is no glitch. The system is accurately rigged, we might say, because, unlike in natural weather forecasts, the weathermen are also the ones who make it rain." (Benjamin, 2019, p. 82).

Benjamin continues, noting that "Even those who purportedly seek 'fairness' in algorithmic decision-making are not usually willing to assert that the benchmark for whether an automated prediction is 'unwarrented' is whether it strays from the proportion of a group in the larger population. That is, if a prediction matches the current crime rate, it is still unjust! Even so, many who are grappling with how to enact ethical practices in this arena still use the crime rate as the default measure of whether an algorithm is predicting fairly, when that very measure is a byproduct of ongoing regimes of selective policing and punishment." (Benjamin, 2019, p. 82).

For a broad discussion of the relationship between racism and law enforcement, I suggested Michelle Alexander's *The New Jim Crow* (Alexander, 2020). I additionally suggest the "Is Police Brutality Really about Race?" chapter in Ijeoma Oluo's book *So You Want to Talk About Race* (Oluo, 2018, pp. 83-98). For additional treatment of algorithms and policing, see the "Civilian Casualties: Justice in the Age of Big Data" chapter in Cathy O'Neil's book *Weapons of Math Destruction* (O'Neil, 2016, pp. 84-104). For a text on the historical evolution of technology in policing, see the "Book Two" section of Charlton D. McIlwain's book *Black Software* (McIlwain, 2020). For a book dedicated to the topic of technology and policing, see *Predict and Surveil* by Sarah Brayne (Brayne, 2021).

Regarding biases in data, Sarah Brayne in *Predict and Surveil* writes that "Data is increasingly positioned as *the* answer to policing challenges. However, the police use of data is as inextricably linked to race as the history and present of policing itself. Media, policy, and academic discourse around data analytics may treat data as objective and color-blind, but one would have to ignore the social dynamics that shape the data to assume that *police data* are color-blind.” (Brayne, 2021, p. 27). The following quote from Brayne provides an example of racial disparities in policing: “Researchers have gained access and analyzed the magnitude of racial disparities in hit rates under stop-and-frisk, finding that 80-90 percent of individuals were released without charge and hit rates were higher for whites than Blacks. Specifically, this confirmed that, controlling for precinct variables and race-specific baseline crime rates, Blacks and Hispanices were stopped more frequently than whites, even though whites who were stopped were more likely to be carrying weapons or contraband than were Blacks.” (Brayne, 2021, p. 104).

The following is a concrete example, from *Weapons of Math Destruction* (O’Neil, 2016), of how selective policing and punishment can manifest in technologies. O’Neil writes that courts in many states used “complicated and mathematical” recidivism models to predict whether someone convicted of a crime might commit a crime in the future. One of the models used by the courts “includes a lengthy questionnaire for the prisoner to fill out.” The questionnaire asks the respondent to provide “the first time you were ever involved with the police”. O’Neil observes that “a criminal who grew up in comfortable suburbs” “might not have a single incident to report other than the one that brought him to prison.” On the other hand, O’Neil observes that “Young black males, by contrast, are likely to have been stopped by police dozens of times, even when they’ve done nothing wrong.” After presenting relevant data about the likelihood of being stopped by police, and the likelihood of being innocent (or what crime might have been committed, if not innocent), O’Neil concludes that “if early ‘involvement’ with the police signals recidivism, poor people and racial minorities look far riskier.” (O’Neil, 2016, pp. 25-26).

O’Neil notes that “The questions [in the questionnaire] hardly stop there” and that “This is unjust. This questionnaire includes circumstances of a criminal’s birth and upbringing, including his or her family, neighborhood, and friends.” The output of the recidivism model is a score, based on the answers to these questions. O’Neil notes that, in some states, “judges use the scores to guide their sentencing.” (O’Neil, 2016, pp. 25-26).

In this text, O’Neil provides a concrete example of how selective policing and punishment might manifest. An individual’s life circumstances—including factors such as their neighborhood, which might be racially segregated—contribute to a score that judges may use to guide their sentencing decisions. Elaborating on geography, Ruha Benjamin in *Race After Technology* writes that “In a country as segregated as the United States, geography is a reliable proxy for race. If zip codes are a relatively low-tech device for instituting racism, how might we apply this insight to computer codes? How do they reinforce racist norms and structures without explicitly invoking race?” (Benjamin, 2019, p. 35).

[14] In *Race After Technology*, Ruha Benjamin discusses a contest from Beauty AI—the “first ever beauty contest judged by robots” (Benjamin 2019, p. 49). Benjamin writes that “On August 2, 2016, the creators of Beauty AI expressed dismay at the fact that ‘the robots did not like people with dark skin.’ All 44 winners across the various age groups except six were White, and ‘only one finalist had visibly dark skin.’ The contest used what was considered at the time the most advanced machine-learning technology available. Called ‘deep learning,’ the software is trained

to code beauty using pre-labeled images, then the images of contestants are judged against the algorithms' embedded preferences. Beauty, in short, is in the trained eye of the algorithm." (Benjamin, 2019, p. 50).

Benjamin then writes that "As one report about the contest put it, '[t]he simplest explanation for biased algorithms is that the humans who create them have their own deeply entrenched biases. That means that despite perceptions that algorithms are somehow neutral and uniquely objective, they can often reproduce and amplify existing prejudices. Columbia University professor Bernard Harcourt remarked: 'The idea that you could come up with a culturally neutral, racially neutral conception of beauty is simply mind-boggling.' Beauty AI is a reminder, Harcourt notes, that humans are really doing the thinking, even when 'we think it's neutral and scientific.' And it is not just the human programmers' preferences for Whiteness that is encoded, but the combined preferences of *all* the humans whose data are studied by machines as they learn to judge beauty and, as it turns out, *health*." (Benjamin, 2019, p. 51).

Meredith Broussard in *Artificial Unintelligence* writes about a similar example: "How can you take a 'good' selfie? In 2015, several prominent American media outlets covered the results of an experiment that purported to answer this question using data science." (Broussard, 2019, p. 149). Broussard then writes that "almost all the 'good' photos were of young white women, despite the fact that older women, men, and people of color were included in the original pool of selfies. Karpathy used a measure of *popularity*—the number of 'likes' each photo garnered on social media—as the metric for what constituted *good*. This type of mistake is quite common among computational researchers who do not critically reflect on the social values and human behaviors that lead to statistics being produced. Karpathy assumed that the photos were popular, and therefore they must be good. By selecting for popularity, the data scientist created a model that had significant bias: it prioritized young, white, cisgender images of women that fit a narrow, heteronormative definition of attractiveness. Let's say that you are an older black man, and you give your selfie to Karpathy's model to be rated. The model will not label your photo as good, no matter what. You are not white and you are not a cisgender woman and you are not young; therefore you do not satisfy the model's criteria for 'good.' The social implication for a reader is that unless you look a certain way, your picture cannot possibly be good. This is not true. Also, no kind of reasonable person would say this to another person!" (Broussard, 2019, p. 149).

[15] Ruha Benjamin, in Race After Technology, writes "An indifferent insurance adjuster who uses the even more disinterested metric of a credit score to make a seemingly detached calculation may perpetuate historical forms of racism by plugging numbers in, recording risk scores, and 'just doing her job.'" (Benjamin, 2019, p. 60).

[16] In the book *Weapons of Math Destruction*, in a discussion of algorithmic approaches for predicting recidivism, Cathy O'Neil writes "Such a finding would undermine the very basis of the recidivism sentencing guidelines. But prison systems, which are awash in data, do not carry out this highly important research. All too often they use data to justify the workings of the system but not to question or improve the system." (O'Neil, 2016, pp. 97-98).

[17] Sarah Brayne, in *Predict and Surveil*, says, “Consider the person-based points system the LAPD uses to identify people who are high risk. Recall that individuals receive five points if they are on parole or probation, five points if they are documented as having a gang affiliation, five points for a violent criminal history, five points for a prior arrest with a handgun, and one point for every police contact. Officers are instructed to find reasons to stop the highest-point people in their patrol areas. But this process obviously leads to a feedback loop: if individuals have a high point value, they are under heightened surveillance and therefore have a greater likelihood of being stopped. Because they gain points for police contact, each time they are stopped, their point value rises. In that sense, the LAPD’s predictive models have created a behavioral loop: they not only predict events (e.g., crime or police contact), they also actually contribute to those events’ future occurrence.” (Brayne, 2021, p. 107).

Regarding place-based predictive policing, rather than person-based predictive policing, Brayne similarly writes that “if historical crime data are used as inputs in a location-based predictive policing algorithm, the algorithm will identify areas with historically higher crime rates as high risk for future crime, officers will be deployed to those areas, and will thus be more likely to detect crimes in those areas, creating a self-fulfilling statistical prophecy while obscuring the role of the law enforcement in shaping crime statistics.” (Brayen, 2021, p. 109).

[18] Cathy O’Neil describes this problem well in the book *Weapons of Math Destruction*. O’Neil writes that “The root of the trouble, as with so many other [Weapons of Math Destruction], is the modelers’ choice of objectives. The model is optimized for efficiency and profitability, not for justice or the good of the ‘team.’” (O’Neil, 2016, pp. 129-130).

Ruha Benjamin, in the book *Race After Technology*, writes that “Finally, the New Jim Code is part of a broader push toward privatization where efforts to cut costs and maximize profits, often at the expense of other human needs, is a guiding rationale for public and private sectors alike. Computational approaches to a wide array of problems are seen as not only good but necessary, and a key feature of cost-cutting measures is the out-sourcing of decisions to ‘smart’ machines. Whether deciding which teacher to hire or fire or which loan applicant to approve or decline, automated systems are alluring because they seem to remove the burden from gatekeepers, who may be too overworked or too biased to make sound judgements. Profit maximization, in short, is rebranded as bias minimization.” (Benjamin, 2019, p. 30). Benjamin then writes that “But the outsourcing of human decisions is, at once, the insourcing of coded inequity.” (Benjamin, 2019, p. 30).

In *Race After Technology*, Benjamin discusses how the Google Maps narrator said “turn right on Malcolm Ten Boulevard” to “Princeton University media specialist Allison Bland” as they were “driving through Brooklyn,” “verbally interpreting the X in the street name as a Roman numeral rather than as referring to the Black liberation leader who was assassinated in the New York City in 1965” (Benjamin, 2019, p. 77). Benjamin discusses how this represents a success, from the perspective of a certain design goal. Specifically, Benjamin writes that “Ironically, this problem of misrecognition actually reflects a solution to a difficult coding challenge. A computer’s ability to parse Roman numerals, interpreting an ‘X’ as ‘ten,’ was a hard-won design achievement. That is, from a strictly technical standpoint, ‘Malcolm Ten Boulevard’ would garner cheers.” (Benjamin, 2019, p. 79).

Benjamin further writes that “what we need is greater investment in socially just imaginaries. This, I think, would have to entail a socially conscious approach to tech development that would require prioritizing equity over efficiency, social good over market imperatives.” (Benjamin, 2019, p. 183). Benjamin later writes that “The key is that all this takes time and intention, which runs against the rush to innovate that pervades the ethos of tech marketing campaigns. But if we are not simply ‘users’ but people committed to building a more just society, it is vital that we demand a slower and more socially conscious innovation.” (Benjamin, 2019, p. 183).

Safiya Umoja Noble, in *Algorithms of Oppression*, writes that “At the core of my argument is the way in which Google biases search to its own economic interests—for its profitability and to bolster its market dominance at any expense.” (Noble, 2018, p. 28).

As another example, in discussing a U.S. state’s (Indiana’s) efforts to automate social service programs, Virginia Eubanks in the book *Automating Inequality* writes, “The goals of the project were consistent throughout the automation experiment: maximize efficiency and eliminate fraud by shifting to a task-based system and severing caseworker-to-client bonds. They were clearly reflected in contract metrics: response time in the call centers was a key performance indicator; determination accuracy was not. Efficiency and savings were built into the contract; transparency and due process were not.” (Eubanks, 2018, p. 74).

Meredith Broussard in *Artificial Unintelligence* writes that “There are also limits to what we *should* do with technology. When we look at the world only through the lens of computation, or we try to solve big social problems using technology alone, we tend to make a set of the same predictable mistakes that impede progress and reinforce inequality.” (Broussard, 2019, p. 7).

Broussard then writes that "One of the red flags I want to raise in this book is a flawed assumption that I call *technochauvinism*. Technochauvinism is the belief that tech is always the solution." (Broussard, 2019, pp. 7-8).

[19] In *Atlas of AI*, Kate Crawford writes that "Again, the practice of classification is centralizing power: the power to decide which differences make a difference." (Crawford, 2021, p. 132). The example that Crawford gives immediately prior to that quote is the following: "The IBM team claimed that their goal was to increase diversity of facial recognition data. Though well intentioned, the classifications they used reveal the politics of what diversity meant in this context. For example, to label the gender and age of a face, the team asked crowdworkers to make subjective annotations, using the restrictive model of binary gender. Anyone who seemed to fall outside of this binary was removed from the dataset. IBM's vision of diversity emphasized the expansive options for cranial orbit height and nose bridges but discounted the existence of trans or gender nonbinary people. 'Fairness' was reduced to meaning higher accuracy rates for machine-led facial recognition, and 'diversity' referred to a wider range of faces to train the model. Craniometric analysis functions like a bait and switch, ultimately depoliticizing the idea of diversity and replacing it with a focus on *variation*. Designers get to decide what the variables are and how people are allocated to categories." (Crawford, 2021, p. 132).

Crawford also writes that "Bowker and Star also underscore that once classifications of people are constructed, they can stabilize a contested political category in ways that are difficult to see. They become taken for granted unless they are actively resisted. We see this phenomenon in the AI field when highly influential infrastructures and training datasets pass as purely technical, whereas in fact they contain political interventions within their taxonomies: they naturalize a particular ordering of the world which produces effects that are seen to justify their original ordering." (Crawford, 2021, p. 139).

Additionally, Crawford notes that "The politics are baked into the classificatory logic, even when the words aren't offensive." (Crawford, 2021, p. 142).

[20] In *Predict and Surveil*, Sarah Brayne writes that "the rise in surveillance—and, more importantly, individuals' *perceptions* of pervasive surveillance—may be met with a concomitant increase in individuals' efforts to evade it. I used nationally representative data from the National Longitudinal Study of Adolescent Health and the National Longitudinal Survey of Youth to test the hypothesis that individuals wary of police surveillance engage in what I have termed 'system avoidance,' or deliberately and systematically avoiding institutions that keep formal records, such as hospitals, banks, schools, and employment, to avoid coming under heightened police surveillance. Results from a range of cross-sectional and longitudinal models suggest that individuals with criminal justice contact systematically avoid interacting with important institutions where they would leave a digital trace. More specifically, individuals who have been stopped by police, arrested, convicted, or incarcerated are more likely to avoid surveilling institutions such as medical, financial, educational, and labor market institutions that keep formal records (i.e., put them 'in the system'). For example, net of sociodemographic and behavioral characteristics, individuals with criminal justice contact have 31 percent higher odds of reporting not obtaining medical care when they thought they needed it, compared to those without criminal justice contact (see Appendix C for full results)." (Brayne, 2021, p. 114).

Although not explicitly about surveillance software, Michelle Alexander in *The New Jim Crow* makes a related observation: “Four years later, voter registration workers in the South encountered scores of people with criminal records who were reluctant to register to vote, even if they were technically eligible, because they were scared to have any contact with governmental authorities.” (Alexander, 2020, p. 200).

[21] In *Algorithms of Oppression*, Safia Umoja Noble writes that “Media stereotypes, which include search engine results, not only mask the unequal access to social, political, and economic life in the United States as broken down by race, gender, and sexuality; they maintain it.” (Noble, 2018, p. 155).

Noble also writes that “On the evening of June 17, 2015, in Charleston, South Carolina, a twenty-one-year-old White nationalist, Dylann ‘Storm’ Roof, opened fire on unsuspecting African American Christian worshipers at ‘Mother’ Emanuel African Methodist Episcopal Church in one of the most heinous racial and religious hate crimes of recent memory.” (Noble, 2018, p. 110). Noble then provides the following information about Roof’s history, and the role of search engines: “According to the manifesto, Roof allegedly typed ‘black on White crime’ in a Google search to make sense of the news reporting on Trayvon Martin, a young African American teenager who was killed and whose killer, George Zimmerman, was acquitted of murder. What Roof found was information that confirmed a patently false notion that Black violence on White Americans is an American crisis.” (Noble, 2018, p. 111).

[22] In the book *Predict and Surveil*, Sarah Brayne writes that "The risk of false positives is of particular concern when the risk concentrates among particular groups. And, as Bolinger points out, that is precisely what statistical prediction does—it reinforces concentrations. In that sense, it is possible for statistical generalizations to be both accurate and unjust. For example, say there were different baseline rates of crime commission by gender, such that 90 percent of people committing crimes are men. You would be statistically justified in treating all men as criminals, but you would also constantly treat innocent men as criminals and subject them to ongoing mistreatment. Ironically, then, the more accurate a prediction, the more unjust it can be in implementation: the more accurate, the more likely police are to act on the prediction and treat people unjustly (with, as we saw in Chapter 6, a host of negative knock-on consequences). So, from a policy perspective, we need to consider not just the *efficacy* of predictive algorithms, but also the *chilling effects* of unevenly applied police contact." (Brayne, 2021, p. 124).

In *Atlas of AI*, Kate Crawford writes that "The tendency to focus on the issue of bias in artificial intelligence has drawn us away from assessing the core practices of classification in AI, along with their attendant politics. To see that in action, in this chapter we'll explore some of the training datasets of the twenty-first century and observe how their schemas of social ordering naturalize hierarchies and magnify inequalities. We will also look at the limits of the bias debates in AI, where mathematical parity is frequently proposed to produce 'fairer systems' instead of contending with underlying social, political, and economic structures. In short, we will consider how artificial intelligence uses classification to encode power." (Crawford, 2021, p. 128).

Crawford additionally writes that "Technical designs can certainly be improved to better account for how their systems produce skews and discriminatory results. But the harder questions of why AI systems perpetuate forms of inequity are commonly skipped over in the rush to arrive at narrow technical solutions of statistical bias as though that is a sufficient remedy for deeper structural problems. There has been a general failure to address the ways in which the instruments of knowledge in AI reflect and serve the incentives of a wider extractive economy. What remains is a persistent asymmetry of power, where technical systems maintain and extend structural inequality, regardless of the intention of the designers." (Crawford, 2021, p. 135).

[23] In the book *So You Want to Talk about Race*, Ijeoma Oluo writes that "Microaggressions are small daily insults and indignities perpetrated against marginalized or oppressed people because of their affiliation with that marginalized or oppressed group, and here we are going to talk about racial microaggressions—insults and indignities perpetrated against people of color. But microaggressions are more than just annoyances. The cumulative effect of these constant reminders that you are 'less than' does real psychological damage." (Oluo, 2018, p. 169). Oluo then elaborates on how serious they are. Oluo also offers illustrative, educational examples.

Like Oluo, Ibram X. Kendi, in the book *How to be an Antiracist*, also explains how serious microaggressions can be. Kendi observes that microaggressions are more than "micro" and more than "aggressions" (Kendi, 2019, p. 47). Kendi writes that "A persistent daily low hum of racist abuse is not minor. I use the term 'abuse' because aggression is not as exacting a term." (Kendi, 2019, p.47). Kendi then writes that "What other people call racial microaggressions I call racist abuse." (Kendi, 2019, p. 47). In this writing, I use the term "microaggression," as this is the term that the reader will likely encounter again as they continue to learn about racism, but I encourage the reader to recognize that microaggressions are abusive, as Kendi notes.

In the book *Technically Wrong*, Sara Wachter-Boettcher observes how microaggressions can manifest in technologies. After discussing computer interfaces that ask users to enter personal information, e.g., by selecting a gender or racial category from a set of options, Wachter-Boettcher writes: "Is being forced to use a gender you don't identify with (or a title you find oppressive, or a name that isn't yours) the end of the world? Probably not. Most things aren't. But these little slights add up—day after day, week after week, site after site—making assumptions about who you are and sticking you into boxes that just don't fit. Individually, they're just a paper cut. Put together, they're a constant thrumming pain, a little voice in the back of your head: *This isn't for you. This will never be for you.*" (Wachter-Boettcher, 2017, pp. 71-72).

Later, Wachter-Boettcher writes, quoting from Aimee Gonzalez-Cameron, a software engineer at Uber, "'"You don't fit on a form" after a while starts to feel like, "you don't fit in a community,"' she told me. 'It chips away at you. It works on you the way that water works on rock.'" (Wachter-Boettcher, 2017, p. 72). Wachter-Boettcher then writes, "This is why I think of noninclusive forms as parallel to microaggressions: the daily little snubs and slights that marginalized groups face in the world" (Wachter-Boettcher, 2017, p. 72).

Wachter-Boettcher continues, noting that "Lots of people think caring about these microaggressions is a waste of time: *Stop being so sensitive! Everyone's a victim these days!* Those people also tend to be the ones least likely to experience them: white, able-bodied, cisgender men—the same people behind the majority of tech products. As Canadian web developer Emily Horseman puts it, forms 'reflect the restricted imagination of their creators: written and funded predominantly by a privileged majority who have never had components of their identity denied, or felt a frustrating lack of control over their representation.'" (Wachter-Boettcher, 2017, pp. 72-73).

In *Design Justice*, Sasha Costanza-Chock writes that "Looking at biased systems through the lens of microaggressions means trying to understand the impact on individuals from marginalized groups as they encounter, experience, and navigate these systems daily. For example, a Black person might experience a microaggression if their hands do not trigger a hand soap dispenser that has been (almost certainly unintentionally) calibrated to work only, or better, with lighter skin tones. This minor interruption of daily life is nevertheless an instantiation of racial bias in the specific affordances of a designed object: the dispenser affords hands-free soap delivery, but only if your hands have white skin. The user is, for a brief moment, reminded of their subordinate position within the matrix of domination." (Costanza-Chock, 2020, p. 45).

In *Race After Technology*, Benjamin also notes that "glitches" can suggest the existence of greater, systemic issues. Specifically, Benjamin writes that "Glitches are generally considered a fleeting interruption of an otherwise benign system, not an enduring and constitutive feature of social life. But what if we understand glitches instead to be a slippery place (with reference to the possible Yiddish origin of the word) between fleeting and durable, micro-interactions and macro-structures, individual hate and institutional indifference? Perhaps in that case glitches are not spurious, but rather a kind of signal of how the system operates. Not an aberration but a form of evidence, illuminating underlying flaws in a corrupted system." (Benjamin, 2019, p. 80).

For more of a discussion of microaggressions, I suggest the entire "What are Microaggressions?" chapter in Ijeoma Oluo's book *So You Want to Talk About Race* (Oluo, 2018, pp. 162-178).

[24] In the book *So You Want to Talk about Race*, Ijeoma Oluo defines intersectionality as "the belief that our social justice movements must consider all the intersections of identity, privilege, and oppression that people face in order to be just and effective" (Oluo, 2018, p. 74). Oluo further writes that "Each of us has a myriad of identities—our gender, class, race, sexuality, and so much more—that inform our experiences in life and our interactions with the world. As we saw when we were checking our privilege, the different hierarchies, privileges, and oppressions assigned to these identities affect our lives in many ways. These privileges and oppressions do not exist in a vacuum, however, and can combine with each other, compound each other, mitigate each other, and contradict each other." (Oluo, 2018, p. 75).

Safiya Umoja Noble, in *Algorithms of Oppression*, writes that "Black feminist thought offers a useful and antiessentializing lens for understanding how both race and gender are socially constructed and mutually constituted through historical, social, political, and economic processes, creating interesting research questions and new analytical possibilities. As a theoretical approach, it challenges the dominant research on race and gender, which tends to universalize problems assigned to race or Blackness as 'male' (or the problems of men) and organizes gender as primarily conceived through the lenses and experiences of White women, leaving Black women in a precarious and understudied position." (Noble, 2018, p. 33).

For more of a discussion of intersectionality, I suggest the entire "What is Intersectionality and Why Do I Need it?" chapter in Ijeoma Oluo's book *So You Want to Talk About Race* (Oluo, 2018, pp. 70-82). I further suggest the "Gender" and "Sexuality" chapters in Ibram X. Kendi's *How to be an Antiracist* (Kendi, 2019, pp. 181-200). I additionally recommend

Sister Outsider by Audre Lorde (Lorde, 1984) and *Disability Visibility* by Alice Wong (editor) (Wong, 2020).

Sasha Costanza-Chock, in *Design Justice*, observes that "intersectionality is an absolutely crucial concept for the development of AI. Most pragmatically, single-axis (in other words, nonintersectional) algorithmic bias audits are insufficient to ensure algorithmic fairness (let along justice)." (Costanza-Chock, 2020, p. 19).

[25] Quoting from https://en.wikipedia.org/wiki/Tokenism: "Tokenism is the practice of making only a perfunctory or symbolic effort to be inclusive to members of minority groups, especially by recruiting people from underrepresented groups in order to give the appearance of racial or gender equality within a workforce. The effort of including a token employee to a workforce is usually intended to create the impression of social inclusiveness and diversity (racial, religious, sexual, etc.) in order to deflect accusations of discrimination." (Accessed April 2021).

[26] In *Black Software*, Charlton D. McIlwain discusses ALERT II, a law enforcement application. As part of that discussion, McIlwain writes that "ALERT II—and systems like it—exemplified black people's relationship to computer technology. The contours of this relationship were wire framed by the late 1960s. They were fully structured and sedimented by the mid-1970s. What was the computer to black people? Black people, by and large, did not have access to the technology being used to profile, target, and forecast their tendency toward criminality. Black people were not hired as technicians to process the data input into the machine. Black people certainly did not design the systems. Black people were not at the table to contribute to

conversations about how to deploy the outputs. Black people were not represented among the industry consultants who showcased the computer's capabilities, developed its use cases, or had the technical skills to know how to build in necessary constraints. Black people were scarce in the ranks of students at higher-education institutions that provided the pipeline to government agencies and industries that were so invested in this work." (McIlwain, 2020, pp. 244-245). McIlwain continues, writing, "What was the computer to black people throughout the 1960s and early 1970s? They certainly never saw themselves pictured in an IBM advertisement utilizing one of the new machines for all kinds of business purposes. To black people by and large the computer was an alien technology destined to go to work on them the way all prior US technologies had—to grind them into submission and exert racial power over their entire existence." (McIlwain, 2020, p. 245).

In *Technically Wrong*, Sara Wachter-Boettcher writes "So how do these alienating, unethical, and downright offensive decisions unfold—over and over again? We can see a common example in the story of Fatima, a Middle-Eastern American design strategist based in the Bay Area." (Wachter-Boettcher, 2017, p. 13). After describing some of Fatima's experiences, Wachter-Boettcher writes that "Fatima's story is over the top: her company ignored her input, made sexist assumptions, and launched a product that failed. But this mind-set—where someone assumes they have all the answers about a product, and leaves out anyone with a different perspective—isn't rare. Scratch the surface at all kinds of companies—from Silicon Valley's 'unicorns' (startups with valuations of more than a billion dollars) to tech firms in cities around the world—and you'll find a culture that routinely excludes anyone who's not young, white, and male." (Wachter-Boettcher, 2017, p.16).

Later, Wachter-Boettcher writes, "With these examples in mind, the racism, sexism, and insensitivity of so many tech products suddenly makes a lot more sense. This is an industry that can look around at a bunch of young white men who plank together in the mornings and get drunk together in the evenings and think, *This is great. This is what a healthy workplace looks like*. If tech culture doesn't notice how its culture excludes others—if it can't even bother to listen to a woman in a meeting—why would it notice when its products do the same? Until the tech industry becomes more representative of the people it's trying to serve, these problems will persist—and our products will be the worse off because of it." (Wachter-Boettcher, 2017, p. 18).

In *Design Justice*, Sasha Costanza-Chock writes that "In 2016, several technology firms, under pressure from mobilized publics, released diversity data about their employment practices. Unsurprisingly, this data does not paint a flattering picture of progress toward gender and racial equality in the tech sector. White and Asian cis men dominate technology jobs. For example, in the United States, women overall hold 26 percent of tech jobs, Black women hold just 3 percent of computer programming jobs, and Latinas hold 2 percent. As feminist media anthropologist Christina Dunbar-Hester notes, gender disparity in the software industry is far worse within the supposedly 'open' arena of free/libre and open-source software (F/LOSS): just 2 percent of F/LOSS developers are women, compared to 30 percent of developers who work on proprietary software. A 2016 report by Intel found that nearly two-thirds of tech workers are white. Sector-wide employment trends are not steadily advancing toward increasing diversity; instead, women and/or B/I/PoC sometimes gain ground, sometimes lose ground." (Costanza-Chock, 2020, pp. 73-74).

Costanza-Chock then writes that "Even when women and/or B/I/PoC are employed in technology design, development, and product management, only a handful have positions at the top of these extremely hierarchical organizations. Gender diversity on the boards of top tech companies tends to range from just 10 percent to 25 percent (almost exclusively white) women. For example, Apple's board currently has six men and two women, Google (Alphabet) has nine and two, Facebook has seven and two, and so on. Yahoo, with a board composed of six men and three women, is the top-tier tech firm that comes closest to gender parity at the highest decision-making level." (Costanza-Chock, 2020, p. 74).

Meredith Broussard in *Artificial Unintelligence* writes that "According to 2015 diversity figures compiled by the *Wall Street Journal*, LinkedIn is the big tech firm with the greatest percentage of women in leadership roles, with a measly 30 percent. Amazon, Facebook, and Google lag with 24, 23, and 22 percent, respectively. In general, statistics on leadership positions tend to be increased by women who rise to the top in marketing and human resources. These two departments tend to be more gender-balanced than engineering roles, as do social media teams. However, at tech firms, the real power is held by the developers and engineers, not by the marketers or HR folks." (Broussard, 2019, p. 158).

[27] In *Race After Technology*, Ruha Benjamin writes that "As one report about the contest put it, '[t]he simplest explanation for biased algorithms is that the humans who create them have their own deeply entrenched biases. That means that despite perceptions that algorithms are somehow neutral and uniquely objective, they can often reproduce and amplify existing prejudices." (Benjamin, 2019, p. 50).

Ruha Benjamin, in *Race After Technology,* gives another concrete example of how the population of developers on a team impacts the behavior of the resulting technology. Specifically, Benjamin writes that "What's more, the different global settings in which AI is taught to 'see' impacts the technical settings designed to identify individuals from various groups. It turns out that algorithms 'developed in China, Japan, and South Korea recognized East Asian faces far more readily than Caucasians. The reverse was true for algorithms developed in France, Germany, and the United States, which were significantly better at recognizing Caucasian facial characteristics.' This suggests that the political-geographic setting augments the default setting of Whiteness. The ethnoracial makeup of the software design team, the test photo databases, and the larger population of users influence the algorithms' capacity for recognition, though not in any straightforward sense." (Benjamin, 2019, p. 112).

In *Algorithms of Oppression*, Safiya Umoja Noble writes that "The political nature of search demonstrates how algorithms are a fundamental invention of computer scientists who are human beings—and code is a language full of meaning and applied in varying ways to different types of information. Certainly, women and people of color could benefit tremendously from becoming programmers and building alternative search engines that are less disturbing and that reflect and prioritize a wider range of informational needs and perspectives." (Noble, 2018, p. 26).

In *Technically Wrong*, Sara Wachter-Boettcher notes that "It's not just Google Photos that has failed to make facial recognition products work as well for people of color as they do for white people. In 2015, Flickr's automatic image tagging labeled a black man as 'ape' (not to mention, the gates of Dachau as a 'jungle gym'). Back in 2009, Nikon released a camera with a special feature: it would warn you if someone in a photo had blinked. The problem was that it hadn't been tested well enough on Asian eyes, and as a result, it routinely flagged them as blinking. The same year, there was the HP computer with a camera that used facial-recognition software to move with the user—panning and zooming to keep their face front and center. Unless the user was black—in which case the software often couldn't recognize them at all." (Wachter-Boettcher, 2017, p. 135). Wachter-Boettcher then writes, "And yet, several years later—and after huge leaps forward in machine-learning capabilities—Google Photos was making the same mistakes. Why? The answer is right back in those tech-company offices we encountered in Chapter 2. If you recall, Google reported that just 1 percent of technical staff was black in 2016. If the team that made this product looked like Google as a whole, it would have been made up almost entirely of white and Asian men. Would they have noticed if, like COMPAS, their failure rates disproportionately affected black people?" (Wachter-Boettcher, 2017, pp. 135-136).

Meredith Broussard, in *Artificial Unintelligence*, writes that "This brings us back to the fundamental problem: algorithms are designed by people, and people embed their unconscious biases in algorithms. It's rarely intentional—but this doesn't mean we should let data scientists off the hook. It means we should be critical about and vigilant for the things we know can go wrong. If we assume discrimination is the default, then we can design systems that work toward notions of equality." (Broussard, 2021, p. 150).

[28] After discussing in *Race After Technology* an employee at Apple who, when asked by someone they supervised if Siri should support African American English, replied that, "Well, Apple products are for the premium market," Ruha Benjamin writes that "the Siri example helps

to highlight how just having a more diverse team is an inadequate solution to discriminatory design practices that grow out of the interplay of racism and capitalism." (Benjamin, 2019, p. 28) Benjamin also writes that "Reflecting on the connection between workforce diversity and skewed datasets, one tech company representative noted that, 'if the training data is produced by a racist society, it won't matter who is on the team, but the people who are affected should also be on the team.'" (Benjamin, 2019, p. 59). Later, Benjamin writes that "We could expect a Black programmer, immersed as she is in the same systems of racial meaning and economic expediency as the rest of her co-workers, to code software in a way that perpetuates racist stereotypes. Or, even if she is aware and desires to intervene, will she be able to exercise the power to do so? Indeed, by focusing mainly on individuals' identities and overlooking the norms and structures of the tech industry, many diversity initiatives offer little more than cosmetic change, demographic percentages on the company pie chart, concealing rather than undoing the racist status quo." (Benjamin, 2019, pp. 61-62).

Safiya Umoja Noble, in *Algorithms of Oppression*, writes that "Filling the pipeline and holding 'future' Black women programmers responsible for solving the problems of racist exclusion and misrepresentation in Silicon Valley or in biased product development is not the answer." (Noble, 2018, p. 65.)

Noble also writes that "Hiles goes on to discuss the exclusionary practices of Silicon Valley, challenging the notion that merit and opportunity go to the smartest people prepared to innovate. Despite her being the only openly gay Black woman to raise $12 million in venture capital for her company, she still faces tremendous obstacles that her non-Black counterparts do not. By rendering people of color as nontechnical, the domain of technology 'belongs' to Whites and reinforces problematic conceptions of African Americans. This is only exacerbated by framing the problems as 'pipeline' issues instead of as an issue of racism and sexism, which extends from employment practices to product design. 'Black girls need to learn how to code' is an excuse for not addressing the persistent marginalization of Black women in Silicon Valley." (Noble, 2018, p. 66).

Sasha Costanza-Chock, in *Design Justice*, writes that "Employment diversity is important. However, ultimately, design justice challenges us to push beyond the demand for more equitable allocation of professional design jobs. Employment diversity is a necessary first move, but it is not the far horizon of collective liberation and ecological sustainability." (Costanza-Chock, 2020, p. 72). Costanza-Chock also writes that "It is tempting to hope that employment diversity initiatives in the tech sector, if successful over time, will solve this problem. Diversifying the technology workforce, as noted above, is a good move, but unfortunately, it will not automatically produce a more diverse default imagined user. Research shows that unless the gender identity, sexual orientation, race/ethnicity, age, nationality, language, immigration status, and other aspects of user identity are explicitly specified, even diverse design teams tend to default to imagined users who belong to the dominant social group." (Costanza-Chock, 2020, p. 78). Costanza-Chock additional writes that "even if design teams perfectly mirrored users in terms of standpoint within the matrix of domination, and even if the unequal costs of communicating specific user needs to decision-makers were addressed, firms would still face pressures from economies of scale to produce solutions optimized for the specifications of the most profitable group of users. As Von Hippel describes, because of economies of scale, firms have very strong incentives to foist existing solutions on all users, even where some users have different specifications. User product specifications for groups of users who are numerically a minority and/or whose purchasing power is relatively small are less likely to be met." (Costanza-Chock, 2020, p. 80).

[29] In *Why Are All the Black Kids Sitting Together in the Cafeteria?*, Beverly Daniel Tatum writes that, "With advances in our understanding of human cognition, psychologists now agree that much of human judgement and behavior is produced with little conscious thought. Our internalized stereotypes and biases are not always consciously known to us, but they can still influence our behavior." (Tatum, 2017, pp. 222-223).

In *Invisible Women*, Caroline Criado Perez explicitly names numerous types of biases, discusses how biases can manifest in data, and discusses the implications of those data biases (Perez, 2019). The following quote, which comes after a discussion of research by Molly Crockett, is illustrative of issues that technology designers should consider: "'It's just a feature of human psychology,' she explains, to assume that our own experiences mirror those of human beings in general. This is a concept in social psychology that is sometimes called 'naive realism' and sometimes called 'projection bias'. Essentially, people tend to assume that our own way of thinking about or doing things is typical. That it's just normal. For white men this bias is surely magnified by a culture that reflects their experience back to them, thereby making it seem even more typical. Projection bias amplified by a form of confirmation bias, if you like. Which goes some way towards explaining why it is so common to find male bias masquerading as gender neutrality. If the majority of people in power are men—and they are—the majority of people in power just don't see it. Male bias just looks like common sense to them. But 'common sense' is in fact a product of the gender data gap." (Perez, 2019, p. 270).

In *Caste*, Isabel Wilkerson writes that "Toward the end of the twentieth century, social scientists found new ways to measure what had transformed from overt racism to a slow boil of unspoken antagonism that social scientists called unconscious bias. This was not the cross-burning, epithet-spewing biological racism of the pre-civil-rights era, but rather discriminatory behaviors based on subconscious prejudgements by people who professed and believed in equality." (Wilkerson, 2020, p. 186). Wilkerson then notes that "By adulthood, researchers have found, most Americans have been exposed to a culture with enough negative messages about African-Americans and other marginalized groups that as much as 80 percent of white Americans hold unconscious bias against black Americans, bias so automatic that it kicks in before a person can process it, according to the Harvard sociologist David R. Williams." (Wilkerson, 2020, pp. 186-187).

In *The New Jim Crow*, Michelle Alexander writes that "Decades of cognitive bias research demonstrates that both unconscious and conscious biases lead to discriminatory actions, even when an individual does not want to discriminate." (Alexander, 2020, pp. 133-134). Alexander further writes that "Most striking, perhaps, is the overwhelming evidence that implicit bias measures are disassociated from explicit bias measures. In other words, the fact that you may honestly believe that you are not biased against African Americans, and that you may even have black friends or relatives, does not mean that you are free from unconscious bias. Implicit bias tests may still show that you hold negative attitudes and stereotypes about blacks, even though you do not believe you do and do not want to." (Alexander, 2020, p. 134). Later in that paragraph Alexander writes that “many people who think they are not biased prove when tested to have relatively high levels of bias.” (Alexander, 2020, p. 134).

Wikipedia offers an extensive list of different types of cognitive biases: https://en.wikipedia.org/wiki/List_of_cognitive_biases (accessed May 2021). Other sources provide shorter, more curated lists. For example, https://builtin.com/diversity-inclusion/unconscious-bias-examples (accessed May 2021) lists the following 16 cognitive biases: affinity bias, confirmation bias, attribution bias, conformity bias, the halo effect, the horns effect, contrast effect, gender bias, ageism, name bias, beauty bias, height bias, anchor bias, nonverbal bias, authority bias, and overconfidence bias.

[30] In *Writing the Other*, Nisi Shawl and Cynthia Ward observe that “Commonly, the unmarked state is revealed as white, male, heterosexual, single, young, and physically able. Other characteristics people have noted include possessing a mid-level income, childless, and human.” (Shawl and Ward, 2005, p. 12).

In *Technically Wrong* Sara Wachter-Boettcher writes that “organizations of all types use tools called *personas*—fictional representations of people who fit their target audiences—when designing their products, apps, websites, and marketing campaigns.” (Wachter-Boettcher, 2017, p. 27). Wachter-Boettcher later writes that “when personas are created by a homogeneous team that hasn’t taken the time to understand the nuances of its audience—teams like we saw in Chapter 2—they often end up designing products that alienate audiences, rather than making them feel at home.” (Wachter-Boettcher, 2017, p. 28).

Related to these default personas, Wachter-Boettcher also discusses default settings for applications. Wachter-Boettcher writes that “These settings are powerful, and not just because we might not notice that a checkbox is already selected (though you can bet marketers are relying on that). Defaults also affect how we perceive our choices, making us more likely to choose whatever is presented as default, and less likely to switch to something else. This is known as the *default effect*.” (Wachter-Boettcher, 2017, p. 34). Wachter-Boettcher also writes that “Default settings can be helpful or deceptive, thoughtful or frustrating. But they’re never neutral. They’re designed. As *ProPublica* journalist Lena Groeger writes, ‘Someone, somewhere, decided what those defaults should be—and it probably wasn’t you.’” (Wachter-Boettcher, 2017, p. 35).

In *Race After Technology*, Ruha Benjamin writes that “the presumed blandness of White American culture is a crucial part of our national narrative. Scholars describe the power of this plainness as the invisible ‘center’ against which everything else is compared and as the ‘norm’ against which everyone else is measured. Upon further reflection, what appears to be an absence in terms of being ‘cultureless’ works more like a superpower. Invisibility, with regard to Whiteness, offers immunity. To be unmarked by race allows you to reap the benefits but escape responsibility for your role in an unjust system.” (Benjamin, 2019, p. 4).

Safiya Umoja Noble, in *Algorithms of Oppression*, writes that “These explorations of web results on the first page of a Google search also reveal the default identities that are protected on the Internet or are less susceptible to marginalization, pornification, and commodification.” (Noble, 2018, p. 58).

The implications of white as the default user or persona, and the unmarked state in general, has direct implications to the design of technologies *and* to society as a whole through the perpetuation of racism and inequities via colorblind designs. Noble writes that “More than fifteen years later, the present research corroborates these concerns. Women, particularly of color, are represented in search queries against a backdrop of a White male gaze that functions as the dominant paradigm on the Internet in the United States. The Black studies and critical Whiteness scholar George Lipsitz, of the University of California, Santa Barbara, highlights the ‘possessive investment in Whiteness’ and the ways that the American construction of Whiteness is more ‘nonracial’ or null. Whiteness is more than a legal abstraction formulated to conceptualize and codify notions of the ‘Negro,’ ‘Black Codes,’ or the racialization of diverse groups of African peoples under the brutality of slavery—it is an imagined and constructed community uniting ethically diverse European Americans. Through cultural agreements about who subtly and explicitly constitutes ‘the other’ in traditional media and entertainment such as minstrel shows, racist films and television shows produced by Hollywood, and Wild West narratives, Whiteness consolidated itself ‘through inscribed appeals to the solidarity of White supremacy.’ The cultural practices of our society—which I argue include representations on the Internet—are part of the ways in which race-neutral narratives have increased investments in Whiteness.” (Noble, 2018, p. 59).

Noble further writes that “Of course, these notions have been constantly challenged, yet they still serve as the basis for beliefs in an ideal of an unmarked humanity—nonracialized, nongendered, and without class distinction—as the final goal of human transcendence. This teleology of the abstracted individual is challenged by the inevitability of such markers and the ways that the individual particularities they signal afford differential realities and struggles, as well as privileges and possibilities. Those who become 'marked' by race, gender, or sexuality as other are deviations from the universal human—they are often lauded for 'transcending' their markers—while others attempt to 'not see color' in a failing quest for colorblindness. The pretext of universal humanity is never challenged, and the default and idealized human condition is unencumbered by racial and gender distinction. This subtext is an important part of the narrative that somehow personal liberties can be realized through technology because of its ability to supposedly strip us of our specifics and make us equal. We know, of course, that nothing could be further from the truth." (Noble, 2018, pp. 62-63).

Sasha Costanza-Chock, in *Design Justice*, writes that "Penny's critique of classed user prioritization within capitalist start-up scenes can be extended: default imagined users are raced, classed, and gendered within a worldview produced by the matrix of domination, internalized and reproduced within technology design teams. Designers most frequently assume that the unmarked user has access to several very powerful privileges, such as US citizenship, English language proficiency, access to broadband internet, a smartphone, a normatively abled body, and so on." (Costanza-Chock, 2020, pp. 76-77). Costanza-Chock also writes that "Put another way, design always involves centering the desires and needs of some users over others. The choice of *which users* are at the center of any given UCD process is political, and it produces outcomes (designed interfaces, products, processes) that are better for some people than others (sometimes very much better, sometimes only marginally so)." (Costanza-Chock, 2020, p. 77).

Costanza-Chock then writes that "In addition, designers tend to unconsciously default to imagined users whose experiences are similar to their own. This means that users are most often assumed to be members of the dominant and hence ‘unmarked’ group: in the United States, this means (cis) male, white, heterosexual, ‘able-bodied,’ literate, college educated, not a young child and not elderly, with broadband internet access, with a smartphone, and so on. Most technology product design ends up focused on this relatively small, but potentially highly profitable, subset of humanity. Unfortunately, this produces a spiral of exclusion as design industries center the most socially and economically powerful users, while other users are systematically excluded on multiple levels: their user stories, preferred platforms, aesthetics, language, and so on are not taken into consideration. This in turn makes them less likely to use the designed product or service. Because they are not among the users, or are only marginally present, their needs, desires, and potential contributions will continue to be ignored, sidelined, or deprioritized.” (Costanza-Chock, 2020, pp. 77-78).

[31] In *So You Want to Talk About Race*, Ijeoma Oluo speaks to the difference between intent and impact on multiple occasions. For example, Oluo writes that “racial oppression is even harder to see than the abuse of a loved one, because the abuser is not one person, the abuser is the world around you, and the person inflicting pain in an individual instance may themselves have the best of intentions.” (Oluo, 2018, p. 19). Oluo also makes the following recommendation: “**Reinforce that good intentions are not the point.** ‘You may not have meant to offend me, but you did. And this happens to people of color all the time. If you do not mean to offend, you will stop doing this.’” (Oluo, 2018, p. 173). Additionally: “**Don’t force people to acknowledge your good intentions.** What matters is that somebody was hurt. That should be the primary focus. The

fact that you hurt someone doesn't mean that you are a horrible person, but the fact that you meant well doesn't absolve you of guilt." (Oluo, 2018, p. 176).

Ruha Benjamin, in *Race After Technology*, also discusses the difference between intent and impact, and the criticality of focusing on impact, with technologies. For example, Benjamin writes that "Today the glaring gap between egalitarian principles and inequitable practices is filled with subtler forms of discrimination that give the illusion of progress and neutrality, even as coded inequity makes it easier and faster to produce racist outcomes. Notice that I said outcomes and not beliefs, because it is important for us to assess how technology can reinforce bias by what it does, regardless of marketing or intention." (Benjamin, 2019, pp. 22-23). As another example, Benjamin writes that "So, to understand racist robots, we must focus less on their intended uses and more on their actions." (Benjamin, 2019, p. 64). Benjamin also writes that "solutions may reinforce forms of social dismissal, regardless of the intentions of the individual programmers." (Benjamin, 2019, p. 79).

[32] In *Algorithms of Oppression*, Safiya Umoja Noble writes that "While organizing this book, I have wanted to emphasize one main point: there is a missing social and human context in some types of algorithmically driven decision making, and this matters for everyone engaged with these types of technologies in everyday life. It is of particular concern for marginalized groups, those who are problematically represented in erroneous, stereotypical, or even pornographic ways in search engines and who have also struggled for nonstereotypical or nonracist and nonsexist depictions in the media and in libraries." (Noble, 2018, p. 10).

Noble also writes that "Rather than assert that problematic or racist results are impossible to correct, in the ways that the Google disclaimer suggests, I believe a feminist lens, coupled with racial awareness about the intersectional aspects of identity, offers new ground and interpretations for understanding the implications of such problematic positions about the benign instrumentality of technologies. Black feminist ways of knowing, for example, can look at searches on terms such as 'black girls' and bring into the foreground evidence about the historical tendencies to misrepresent Black women in the media." (Noble, 2018, p. 31).

In *Race After Technology*, Ruha Benjamin writes that "The danger of New Jim Code impartiality is the neglect of ongoing inequity perpetuated by colorblind designs." (Benjamin, 2019, p. 143). Benjamin also writes that "Algorithmic neutrality reproduces algorithmically sustained discrimination." (Benjamin, 2019, p. 143).

Costanza-Chock also writes that "To make matters worse, far too often user personas are created out of thin air by members of the design team (if not autogenerated by a service like Userforge), based on their own assumptions or stereotypes about groups of people who might occupy a very different location in the matrix of domination. When this happens, user personas are literally objectified assumptions about end users. In the worst case, these objectified assumptions then guide product development to fit stereotyped but unvalidated user needs. Sometimes, they may also help designers *believe* they are engaged in an inclusive design process, when in reality the personas are representations of designers' unvalidated beliefs about marginalized or oppressed communities." (Costanza-Chock, 2020, pp. 82-83). As an alternate approach, Costanza-Chock suggests that "design justice practitioners focus on trying to ensure that community members are actually included in meaningful ways throughout the design process." (Costanza-Chock, 2020, p. 84). Costanza-Chock also writes that "design justice compels us to begin by listening to community organizers, learning what they are working on, and asking what the most useful focus of design efforts would be. In this way, design processes can be community-led, rather than designer- or funder-led." (Costanza-Chock, 2020, p. 84).

In *Atlas of AI*, Kate Crawford writes that "Making these choices about which information feeds AI systems to produce new classifications is a powerful moment of decision making: but who gets to choose and on what basis? The problem for computer science is that justice in AI systems will never be something that can be coded or computed. It requires a shift to assessing systems beyond optimization metrics and statistical parity and an understanding of where the frameworks of mathematics and engineering are causing problems. This also means understanding how AI systems interact with data, workers, the environment, and the individuals whose lives will be affected by its use and deciding where AI should not be used." (Crawford, 2021, pp. 147-148).

[33] In *Why Are All the Black Kids Sitting Together in the Cafeteria?*, Beverly Daniel Tatum discusses implicit biases and the Race IAT (implicit-association test). Tatum then writes that "It also predicts doctors' differential treatment of Black and White patients—emergency room and resident physicians recommend the optimal treatment, thrombolytic (blood-clot-dissolving) therapy, less often for a Black patient than for a White one with the same acute cardiac symptoms. (The Institute of Medicine has concluded that racial and ethnic minorities receive less-effective care even when income levels and insurance benefits are the same, pointing to implicit bias as the cause.)" (Tatum, 2017, p. 224).

[34] In *Predict and Surveil*, Sarah Brayne writes that "Contrary to popular accounts, big data is not objective or less biased than discretionary 'human' decision making. More to the point: big data is fundamentally social." (Brayne, 2021, p. 4). Brayne adds that "Algorithms do not transcend the social, but are shaped by the social world in which they are created and used. The activities that generate and the technologies that analyze data are all embedded within social contexts and power structures, so the resulting data are anything but 'natural,' detached, or purely descriptive. Rather, people situated in preexisting social, organizational, and institutional contexts *decide* what data to collect and analyze, about whom, and for what purpose. So, just as individuals carry a range of implicit biases that affect their decisions, algorithms are loaded up with a scaffolding of implicit biases affecting the data they analyze and produce. What data can measure and quantify is not a technical question, but a normative one related to institutional priorities, organizational imperatives, and individual and group preferences. Nor does everyone have an equal ability to collect data, make decisions, construct policies, and intervene in others' lives based on that data." (Brayne, 2021, pp. 4-5).

Virginia Eubanks writes in *Automating Inequality*, in the context of a software tool for predicting child welfare needs, that "We all tend to defer to machines, which can seem more neutral, more objective. But it is troubling that managers believe that if the intake screener and the computer's assessment conflict, the human should learn from the model." (Eubanks, 2018, p. 142).

Ruha Benjamin, in *Race After Technology*, writes that "I argue that tech fixes often hide, speed up, and even deepen discrimination, while appearing to be neutral or benevolent when compared to the racism of a previous era." (Benjamin, 2019, p. 8). Benjamin also writes that "the notion that tech bias is 'unintentional' or 'unconscious' obscures the reality—that there is no way to create something without some intention and intended user in mind" (Benjamin, 2019, p.28).

Writing to the "allure of objectivity without public accountability" of technologies, and biases can be "hidden from public view," Benjamin write: "Racist robots, as I invoke them here, represent a much broader process: social bias embedded in technical artifacts, the allure of objectivity without public accountability. Race as a form of technology—the sorting, establishment and enforcement of racial hierarchies with real consequences—is embodied in robots, which are often presented as simultaneously akin to humans but different and at times superior in terms of efficiency and regulation of bias. Yet the way robots can be racist often remains a mystery or is purposefully hidden from public view." (Benjamin, 2019, p. 53). Benjamin also writes that "Just as legal codes are granted an allure of objectivity—'justice is (color)blind' goes the fiction—there are enormous mystique around computer codes, which hides the human biases involved in technical design." (Benjamin, 2019, p. 78).

Highlighting the risks with "algorithmic neutrality" and "colorblind designs," Benjamin writes that "The danger of New Jim Code impartiality is the neglect of ongoing inequity perpetuated by colorblind designs. In this context, algorithms may not be just a veneer that covers historical fault lines. They also seem to be streamlining discrimination—making it easier to sift, sort, and *justify* why tomorrow's workforce continues to be racially stratified. Algorithmic neutrality reproduces algorithmically sustained discrimination." (Benjamin, 2019, p. 143).

Also writing to the problems with "neutral" and "colorblind" designs, Safiya Umoja Noble, in *Algorithms of Oppression*, writes "Central to these 'colorblind' ideologies is a focus on the inappropriateness of 'seeing race.' In sociological terms, colorblindness precludes the use of racial information and does not allow any classifications or distinctions. Yet, despite the claims of colorblindness, research shows that those who report higher racial colorblind attitudes are more likely to be White and more likely to condone or not be bothered by derogatory racial images viewed in online social networking sites. Silicon Valley executives, as previously noted, revel in their embrace of colorblindness as if it is an asset and not a proven liability. In the midst of reenergizing the effort to connect every American and to stimulate new economic markets and innovations that the Internet and global communications infrastructures will afford, the real lives of those who are on the margin are being reengineered with new terms and ideologies that make a discussion about such conditions problematic, if not impossible, and that place the onus of discriminatory actions on the individual rather than situating problems affecting racialized groups in social structures." (Noble, 2018, p. 168).

Speaking to a concrete example of how users can misinterpret designs as objective, Noble writes: "The very presence of Black women and girls in search results is misunderstood and clouded by dominant narratives of the authenticity and lack of bias of search engines. In essence, the social context or meaning of derogatory or problematic Black women's representations in Google's ranking is normalized by virtue of their placement, making it easier for some people to believe that what exists on the page is strictly the result of the fact that more people are looking for Black women in pornography than anything else. This is because the public believes that what rises to the top in search is either the most popular or the most credible or both." (Noble, 2018, p. 32).

Meredith Broussard, in *Artificial Unintelligence*, writes that "Part of the reason we run into problems when making social decisions with machine learning is that the numbers camouflage important social context." (Broussard, 2019, p. 115). Broussard later writes that "This speaks to a principle called the *unreasonable effectiveness of data*. Unless you are alert to the possibilities of discrimination and disarray, AI seems like it works beautifully." (Broussard, 2019, p. 118). Broussard additionally writes that "algorithms don't work fairly because people embed their unconscious biases into algorithms. Technochauvinism leads people to assume that mathematical formulas embedded in code are somehow better or more just for solving social problems—but that isn't the case." (Broussard, 2019, p. 156).

[35] In *Weapons of Math Destruction*, Cathy O'Neil identifies opacity—the inability to know how an algorithm works or how the results of the algorithm are used—as one of the three key elements of what she calls a weapon of math destruction (O'Neil, 2016, p. 31). O'Neil states that "Opaque and invisible models are the rule, and clear ones very much the exception." (O'Neil, 2016, p. 28). O'Neil discusses a school district that used a computer algorithm to rate teachers, and notes that teachers have been fired as a result of the algorithm's output. O'Neil writes that teachers sought to be told how the algorithm works, but were not told. O'Neil then writes, "if the people being evaluated are kept in the dark, the thinking goes, they'll be less likely to attempt to game the system. Instead, they'll simply have to work hard, follow the rules, and pray that the model registers and appreciates their efforts. But if the details are hidden, it's also harder to question the score or to protest against it." (O'Neil, 2016, p. 8).

As another reason for opacity, O'Neil observes that "One common justification is that the algorithm constitutes a 'secret sauce' crucial to their business. It's *intellectual property*, and must be defended, if need be, with legions of lawyers and lobbyists." (O'Neil, 2016, p. 29).

In *Race After Technology,* Ruha Benjamin writes: "Consider that machine-learning systems, in particular, allow officials to outsource decisions that are (or should be) the purview of democratic oversight. Even when public agencies are employing such systems, private companies are the ones developing them, thereby acting like political entities but with none of the checks and balances. They are, in the words of one observer, 'governing without a mandate,' which means that people whose lives are being shaped in ever more consequential ways by automated decisions have very little say in how they are governed." (Benjamin, 2019, p. 53).

In *Predict and Surveil*, Sarah Brayne writes that "It's hard to trust what you can't understand. Predictive algorithms are opaque, shrouded in 'algorithmic secrecy.' Partially, this is because algorithms—such as Google's 'secret sauce,' risk assessment tools such as COMPAS, or PredPol's predictive policing algorithm—are usually proprietary. However, even those that are 'transparent,' in that they are publicly available, tend to be difficult to interpret. Algorithmic opacity is particularly relevant in the case of law enforcement, as it exacerbates the base-level skepticism that exists about the value of data-driven policing. For law enforcement—or anyone for that matter—to evaluate the fairness or efficacy of an algorithm, it must be understandable and interpretable." (Brayne, 2021, p. 86).

Brayne also observes that the creators of a system might operate under different legal constructs than the users of a system. Specifically, Bryane writes that "Much contemporary policing is programmatic, suspicionless, cumulative, probabilistic, and technologically mediated.

It involves not only police-civilian interactions, but also police-dataset interactions, in which the data are often collected by nonpolice actors, before legal protections guarding against police overreach come into play." (Brayne, 2021, p. 119). The algorithms used by those nonpolice actors may be hidden from the public and from oversight. As a concrete suggestion, Brayne writes that "Just as you can see and correct errors in your credit score, people should be able to argue and even change information about themselves populating the machine learning models used by the police." (Brayne, 2021, p. 143).

In *Atlas of AI*, Kate Crawford writes that "The typical structure of an episode in the ongoing AI bias narrative begins with an investigative journalist or whistleblower revealing how an AI system is producing discriminatory results. The story is widely shared, and the company in question promises to address the issue. Then either the system is superseded by something new, or technical interventions are made in the attempt to produce results with greater parity. Those results and technical fixes remain proprietary and secret, and the public is told to rest assured that the malady of bias has been 'cured.'" (Crawford, 2021, p. 129).

Crawford additionally writes that "When the matching process of AI are truly hidden and people are kept unaware of why or how they receive forms of advantage or disadvantage, a collective political response is needed—even as it becomes more difficult." (Crawford, 2021, p. 149).

[36] There is a long history of excessive use of force by law enforcement; see, for example, Michelle Alexander's book *The New Jim Crow* (Alexander, 2020), Ijeoma Oluo's book *So You Want to Talk about Race* (Oluo, 2018), and Beverly Daniel Tatum's book, *Why Are All the Black*

Kids Sitting Together in the Cafeteria? (Tatum, 2017). The book *Disability Visibility*, edited by Alice Wong, has a section specifically focused on intersectionality, racism, disabilities, and excessive use of force by law enforcement (Wong, 2020, pp. 236-242); that section is attributed to the Harriet Tubman Collective.

[37] Ruha Benjamin, in *Race After Technology*, writes that "As widespread concern over mass incarceration increases, people are turning to technological fixes that encode inequity in a different form." (Benjamin, 2019, p. 25). Benjamin also writes that "As a form of social technology, institutional racism, past and present, is the precondition for the carceral technologies that underpin the US penal system. At every stage of the process—from policing, sentencing, and imprisonment to parole—automated risk assessments are employed to determine people's likelihood of committing a crime. They determine the risk profile of neighborhoods in order to concentrate police surveillance, or the risk profile of individuals in order to determine whether or for how long to release people on parole." (Benjamin, 2019, p. 81).

As a concrete example, Benjamin writes that "Most EM [electronic monitoring] is being used in pre-trial release programs for those who cannot afford bail, employing GPS to track individuals at all times—a newfangled form of incarceration before conviction. As founder and executive director of the Center for Media Justice, Malkia Cyril stresses: 'There is increasing evidence that the algorithmic formulas used to make these decisions [about who should be assigned e-monitoring] contain deep-seated racial bias, so we must explore the extent to which EM both infringes upon core civil rights and represents a new frontier to compile surveillance data.' The very solutions to mass incarceration and prison overcrowding, in other words, give rise to innovative forms of injustice. They are, in short, racial fixes that harm even as they purport to help." (Benjamin, 2019, pp. 139-140).

In the book *Predict and Surveil*, Sarah Brayne provides a detailed study of law enforcement use of technology. The following quotes provide some (partial) context for the shift of law enforcement to new technologies and hint at underlying inequities with these technologies. Brayne writes that "An institutional perspective goes further: law enforcement adopted big data analytics not because there was empirical evidence that it *actually* improved efficiency, but because there was immense institutional pressure to conform as other institutions began marshaling big data and algorithmic predictions for decision-making. Using predictive analytics and surveillance technologies initially developed in military contexts, for example, could confer a degree of legitimacy upon police departments. Moreover, data-driven policing may provide greater accountability as departments respond to criticisms over discriminatory practices. For example, as I will discuss in more detail later, law enforcement has responded to nationwide movements around police violence, including Black Lives Matter, and their calls for police reform by holding up data-driven policing as a partial antidote. Implicitly, these responses have promoted data as impartial and objective, though as noted earlier, data are inescapably social." (Bryane, 2021, p. 23).

Brayne later writes that "I assumed that law enforcement representatives would ask surveillance company representatives how their platforms could help police achieve their goals. I quickly saw, however, that the pattern was usually the inverse: software representatives would demonstrate the use of their platform in a non-law enforcement—usually military—context, then ask local law enforcement whether they would be interested in a similar application for their local context. In other words, instead of filling analytic gaps or technical voids identified by law enforcement, software representatives helped create new kinds of institutional demand to sell lucrative platform licensing agreements." (Brayne, 2021, p. 26).

In *Black Software*, Charlton D. McIlwain writes that "The Commission made many recommendations based on the Committeemen's work. They all added up to a single conclusion: the federal government should spend hundreds of millions of dollars to build, test, and transfer new technological solutions to law enforcement agencies to help solve law enforcement problems. But what should these new systems focus on? Which of the many elements of law enforcement was the computer especially equipped to tackle? For what problems were law enforcement agencies seeking a solution? The answer to the final question was ... none." (McIlwain, 2020, p. 204).

McIlwain continues, writing that "The Committeemen launched their studies, gathered their data, crunched their numbers, plotted their systems, and redefined police operations. All the while, police chiefs and rank-and-file officers from around the country didn't give two shits about using technology to do a job they knew they had been doing so well all on their own." (McIlwain, 2020, p. 204).

Brayne also writes that "The solution proffered in response to so many of the problems of race and policing is 'more data.' Humans are subjective and biased, the reasoning goes, but data are objective and unbiased. Data carries promises of 'mechanical objectivity': in contrast to an individual's personal account of an emotionally fraught event such as a police-involved shooting, data are ostensibly stripped of subjective interpretation, and presented as objective measures. Again, we run up against the stark fact that data are, no matter how they are presented, social products of social contexts." (Brayne, 2021, p. 29). Brayne then notes that, "Thus, there is reason to be wary when data-driven policing is offered as *the* antidote to racially discriminatory practices in police departments across the country." (Brayne, 2021, p. 29). Later, after describing a particular predictive policing strategy, Brayne writes that "It also shows how 'data-driven' policing is seen as offering a legal solution for law enforcement: it does the same old thing, but faster and in a supposedly race-neutral, quantified way. Put differently, coding people as 'likely offenders' looks race-neutral and makes disparate treatment more legally defensible." (Brayne, 2021, p. 68). Brayne also writes that "despite one of its goals being to avoid legally contestable bias in police practices, Operation LASER hides both intentional and unintentional bias in policing." (Brayne, 2021, p. 69). Brayne also writes that "Scholars are only starting to bring labeling theory into the digital age and trace out how the integration of records and data across platforms may effectively make the mark of a criminal record or criminal justice contact indelible, and cascade across previously discrete institutional boundaries. Existing research underestimates the effects of involvement in the criminal justice system when that involvement is tied to impenetrable, interwoven, and proprietary data systems. We literally don't know where all the data *goes*." (Brayne, 2021, p. 141).

[38] In the book *Predict and Surveil*, Sarah Brayne writes that "local law enforcement is embracing technological tools initially developed in military contexts and that police are partnering with private companies to design tools and collect data." (Brayne, 2021, p. 17). Brayne also writes that "In some instances, it is simply easier for law enforcement to purchase privately collected data than to rely on in-house data, because there are fewer constitutional protections, reporting

requirements, and appellate checks on private sector surveillance and data collection. Therefore, purchasing data from data brokers can be a way for law enforcement agencies to circumvent privacy laws." (Brayne, 2021, p. 25). Regarding which technologies are chosen for police use, Brayne writes that "Just like other organizations, the LAPD's decisions about which technologies to adopt, which data to collect in-house, and which data to purchase from external vendors are not made in a vacuum. Rather, they are shaped by political, economic, and social factors, and by an organizational context rife with power politics." (Brayne, 2021, p. 34).

In *Race After Technology,* Ruha Benjamin writes: "Consider that machine-learning systems, in particular, allow officials to outsource decisions that are (or should be) the purview of democratic oversight. Even when public agencies are employing such systems, private companies are the ones developing them, thereby acting like political entities but with none of the checks and balances. They are, in the words of one observer, 'governing without a mandate,' which means that people whose lives are being shaped in ever more consequential ways by automated decisions have very little say in how they are governed." (Benjamin, 2019, p. 53).

Kate Crawford, in *Atlas of AI*, writes that "The dual use of AI applications in both civilian and military domains has also produced strong incentives for close collaboration and funding." (Crawford, 2021, p. 187). Crawford provides several examples of private sector technologies used by law enforcement, including the following: "To complete its public-private infrastructure of surveillance, Amazon has been aggressively marketing the Ring system to police departments, giving them discounts and offering a portal that allows police to see where Ring cameras are located in the local area and to contact homeowners directly to request footage informally without a warrant. Amazon has negotiated Ring video-sharing partnerships with more than six hundred police departments." (Crawford, 2021, p. 202). Crawford also writes that "Another example of this phenomenon is Vigilant Solutions, established in 2005. The company works on the basis of a single premise: take surveillance tools that might require judicial oversight if operated by governments and turn them into a thriving private enterprise outside constitutional privacy limits." (Crawford, 2021, p. 199).

Crawford also writes that "Despite the massive expansion of government contracts for AI systems, little attention has been given to the question of whether private vendors of these technologies should be legally accountable for the harms produced when governments use their systems." (Crawford, 2021, pp. 198-199). Crawford later writes that "With legal scholar Jason Schultz, I've argued that developers of AI systems that directly influence government decisions should be found to be state actors for the purposes of constitutional liability in certain contexts." (Crawford, 2021, p. 199).

Although not specifically related to law enforcement technologies, Safia Umoja Noble in *Algorithms of Oppression* also discusses the control that the private sector has over technologies. For example, Noble writes that "An increasingly de- and unregulated commercially driven Internet raises significant issues about how information is accessed and made available." (Noble, 2018, p. 154). Noble also writes that "It is these kinds of practices that mark the consequences of the rapid shift over the past decade from public-interest information to the corporate takeover of U.S. news media, which has made locating any kind of alternative information increasingly difficult and pushed the public toward the web. Equally, media consolidations have contributed to the erosion of professional standards such as fact checking, not misrepresenting people or situations, avoiding imposing cultural values on a group, and distinguishing between commercial and advertising interests versus editorial decisions—all of which can be applied to information provision on the web. As the search arena is consolidated under the control of a handful of corporations, it is even more crucial to pay close attention to the types of processes that are shaping the information prioritized in search engines." (Noble, 2018, p. 155).

[39] As an example of how technologies can be designed and deployed without fully anticipating and understanding their consequences, in describing Chapter 4 of the book *Race After Technology*, Ruha Benjamin writes that "I take a look at technologies that explicitly work to address different forms of discrimination, but that may still end up reproducing, or even deepening, discriminatory processes because of the narrow way in which 'fairness' is defined and operationalized." (Benjamin, 2019, pp. 47-48).

The gap between design intention and technology impact is an example of the failure to fully anticipate and understand the consequences of a technology before deployment. The following quote from Benjamin highlights this gap: "Notice that I said outcomes and not beliefs, because it is important for us to assess how technologies can reinforce bias by what it does, regardless of marketing or intention." (Benjamin, 2019, pp. 21-22).

In *Predict and Surveil*, Sarah Brayne writes that "We must find ways to curb institutions' craving for data, because that hunger too often outpaces our understanding of the intended and unintended consequences of a data binge." (Brayne, 2021, p. 16). Brayne also writes: "*Slow down*. Organizations today have a seemingly insatiable appetite for digital troves. I am not calling for wholesale rejection of data analytics or new surveillance technologies. Rather, instead of continuing with the current pattern of data collection and law enforcement intervention first, assessment and evaluation later, we need to invert the order of operations. Despite the rhetoric, the evidence is still weak. We don't actually know whether and how algorithms 'outperform' humans and what harms may come with their implementation. The onus should fall on law enforcement to justify the use of big data and new surveillance tools *prior* to mass deployment. Moreover, such evaluations need to be *independent*." (Brayne, 2021, pp. 141-142).

[40] https://ncses.nsf.gov/pubs/nsf19304/digest/introduction (accessed October 2020).

[41] See the "What is the Model Minority Myth?" chapter in Ijeoma Oluo's book *So You Want to Talk About Race* (Oluo, 2018, pp. 189-200). See also the "Beyond Black and White" section of Beverly Daniel Tatum's book *Why Are All the Black Kids Sitting Together in the Cafeteria?* (Tatum, 2017, pp. 235-328).

[42] Ijeoma Oluo, in *So You Want to Talk About Race*, writes: "How do our social justice efforts so often fail to help the most vulnerable in our populations? This is primarily a result of unexamined privilege. Because of how rarely our privilege is examined, even our social justice movements will tend to focus on the most privileged and most well represented people within those groups. Anti-racism groups will often tend to prioritize the needs of straight men of color, feminist groups will tend to prioritize the needs of white women, LGBTQ groups will tend to prioritize the needs of white gay cisgender men, disability rights groups will tend to prioritize the needs of disabled white men. Imagine where this leaves a disabled Latinx trans woman on any group's priority list. Because the needs of the most privileged are usually the ones prioritized,

they are often the only ones considered when discussing solutions to oppression and inequality. These solutions, not surprisingly, often leave the underprivileged populations in our movements behind." (Oluo, 2018, pp. 76-77).

[43] https://en.wikipedia.org/wiki/Kish%C5%8Dtenketsu (accessed May 2021).