



## From biased robots to race as technology

Inga Luchs

### review

Benjamin, R. (2019) *Race after technology: Abolitionist tools for the New Jim Code*. Cambridge, UK/Medford, MA: Polity Press.

Ranking search results, filtering spam e-mails, recommending movies and books, evaluating credit card fraud, diagnosing malignant cells in cancer research, selecting job applicants – more and more tasks are being carried out by new media technologies such as machine learning algorithms. Their logic does not only simplify our daily approach to large masses of information but also applies at a higher level in the observation and regulation of population flows. They are for instance employed for policing purposes (e.g., in predictive policing), or in the economic sector, where they gain significant attention in the realm of data mining and big data analytics. With *Race after technology: Abolitionist tools for the New Jim Code*, Associate Professor of African American Studies, Ruha Benjamin, issues a critique of this contemporary situation by demonstrating how particularly new technologies reproduce and increase social inequalities under the guise of apparent objectivity and efficiency [5f.]. Benjamin centralizes her research around what she identifies as the ‘New Jim Code’, comprising ‘a range of discriminatory designs – some that explicitly work to amplify hierarchies,

many that ignore and thus replicate social divisions, and a number that aim to fix racial bias but end up doing the opposite' [8].

With her research, Benjamin follows a long line of US-American publications in the last decade that point out patterns of discrimination caused by specific advances in technology. In *Algorithms of oppression* (2018), Safiya Noble, for instance, illustrates how Google's search engine shows biases against Women of Color and other marginalized groups (2018: 6, 15). With her book, *Automating inequality* (2017), Virginia Eubanks traces algorithmic decision-making systems and their social consequences, for instance in the management of homelessness in Los Angeles or the detection of child abuse and neglect in Pennsylvania. Cathy O'Neil provides a more in-depth technical analysis in her book, *Weapons of math destruction* (2016). With years of work experience as a data scientist in financial markets and big data analytics, she focuses on the statistical models which are part of the machine learning technologies that form the basis for automated decisions. O'Neil conceives these models as particularly troubling, given that they are 'opaque, unquestioned, and unaccountable, and...operate at a scale to sort, target, or "optimize" millions of people' (O'Neil, 2016: 12).

Ruha Benjamin adds to this previous research by combining the analysis of technologies from a Science and Technology Studies (STS) perspective with a thorough examination of its interconnection with critical race studies. For her newly devised field of research, she coins the term 'race critical code studies' [44]. Building on previous work by Safiya Noble (2018) and Simone Browne (2015), Benjamin centers this research around the historical connections between the logic of racism and technologies. Here, she follows a chronological line, referring with her notion of the 'New Jim Code' to *The New Jim Crow* (2012), a book by Michelle Alexander on the US carceral system. Alexander shows how the conditions of the contemporary US carceral system can be traced back to the 'Jim Crow era', which mandated racial segregation and manifested white supremacy from the 1870s to the 1960s (Alexander, 2012: 91). Similarly, Benjamin shows how algorithmic discrimination neither results from technologies that are inherently racist, nor from programmers who deliberately program racist algorithms [52f.]. Instead, as she points out, structural racism conditions contemporary technological classification systems, perpetuating already separated and

stratified societies along racialized lines [36]. In this vein, race has become ‘one of our most powerful tools – developed over hundreds of years, varying across time and place, codified in law and refined through custom, and, tragically, still considered by many people to reflect immutable differences between groups’ [36]. With this theoretical framework, Benjamin builds upon previous work by media researcher Wendy Chun (2009), considering race as a kind of technology: ‘[This] is an invitation to consider racism in relation to other forms of domination as not just an ideology or history, but as a set of technologies that generate patterns of social relations, and these become Black-boxed as natural, inevitable, *automatic*’ [44-45].

In her 2009 introduction to the special issue ‘Race and/as technology’ of *Camera obscura*, Chun further identifies, how regarding race as technology helps a move from the ontological question of *race* – which remains important nonetheless – to the question of its function, ‘regardless of its alleged essence’ (Chun, 2009: 9). Through her argument, Chun exposes how race ‘functions as the “as”, how it facilitates comparisons between entities classed as similar or dissimilar’ (*ibid.*: 8). Here, Chun refers to Michael Omi and Howard Winant’s argument of race as ‘a fundamental organizing principle of social relationships’ (Omi and Winant, 1986: 61-62, cited after Chun, 2009: 14). Chun clarifies how race has historically been ‘a tool of subjugation’, in the way that ‘supposedly objective scientific categorizations of race have been employed to establish hierarchical differences between people, rendering some mere objects to be exploited, enslaved, measured, demeaned, and sometimes destroyed’ (Chun, 2009: 10).

While Benjamin offers a compelling illustrative account, dense with examples of contemporary technologies and their intersection with race, a more substantial historical classification and theoretical examination unfortunately often disappears behind footnotes. Yet, her specific approach reads as a conscious choice: with her method of ‘thin description’, she explicitly opposes the ‘New Jim Code’, which ‘seeks to penetrate all areas of life’, favoring instead ‘a much needed discretion, pushing back against the all-knowing, extractive, monopolizing practices of coded inequity’ [46]. Understanding race as technology helps Benjamin open the discussion of racial discrimination at all technological levels, from naming as technology to concrete algorithmic systems. By choosing this mode of analysis,

however, Benjamin's book appears to remain on the surface of things, not clarifying whether she is only illustrating existing theoretical work or developing her own contribution to the existing literature. This raises the question of whether a deeper engagement with the theoretical concepts addressed, as well as the realities of the various technologies, would be more productive. This will become particularly clear in the following discussion of the chapters.

### **Everyday apps and complex algorithms:<sup>1</sup> The importance of technical specificities**

In Chapter 1, *Engineered inequity*, Benjamin starts with the example of Beauty AI, which was an Australian-Chinese application, where contestants worldwide could take part in a beauty contest by uploading pictures of themselves. What was peculiar about the project was that the winners were not chosen by a human judge but a supervised machine learning algorithm. The 'robot jury' of Beauty AI processed the uploaded images according to certain criteria such as 'wrinkles, face symmetry, skin color, gender, age group, ethnicity and "many more parameters"' [50] and compared them to an already existing, human-classified database. Not surprisingly, the algorithm ultimately turned out to be discriminatory, predominantly declaring white women winners of the contest and thus reflecting societal biases (*ibid.*). In Chapter 2, *Default discrimination*, Benjamin draws on the, by now, well-known example of the recidivism algorithm, COMPAS, implemented in several US states to predict the probability of criminals committing future crimes. The program developed by the for-profit company, Northpointe, assigned scores to individuals who committed a felony. This so-called 'risk assessment' was then used to estimate a person's potential recidivism, influencing his or her time of release ([81]; Angwin et al., 2016). In the first two chapters, Benjamin aptly illustrates what race as technology looks like in relation to contemporary algorithmic systems – showing that it is not malfunctioning technologies or biased programmers

---

<sup>1</sup> 'From everyday apps to complex algorithms, *Race after technology* aims to cut through industry hype to offer a field guide into the world of biased bots, altruistic algorithms, and their many coded cousins' [7].

that lead to discriminatory systems, but a structural form of racial discrimination which runs consistently through the technologies.

In Chapter 3, *Coded exposure*, Benjamin traces the duality of creating (in)visibility through technologies and the consequences of inclusion and exposure for racialized groups. Here, she investigates the implications of technologies that are ‘presenting partial and distorted visions as neutral and universal’ [100]. With Kodak’s so-called ‘Shirly Card’, the image of a white woman that was used from the 1950s to the 1990s for the standardization of light exposure in photography, Benjamin exemplifies how whiteness was set as default. Consequently, dark-skinned people were regularly underexposed [103f.]. Yet, also contemporary technologies, applied in algorithmic decision-making – such as in the granting of credit loans or in the automatic pre-selection of applicants for a job – are often advertised as neutral, even though they, too, evidently ‘...reproduce long-standing forms of structural inequality and colorblind racism’ [64]. As Benjamin further elaborates: ‘[These] default settings, once fashioned, take on a life of their own, projecting an allure of objectivity that makes it difficult to hold anyone accountable’ [64]. On the other side, Benjamin shows how efforts at inclusion and increased visibility of Black people within technologies should be met with caution, as ‘we do not all experience the dangers of exposure in equal measure’ [111]. Particularly when it comes to surveillance measures and the application of face recognition, Black people are frequently overly patrolled and discriminated against, as examples of predictive policing show [112]. Consequently, Black people stand under enhanced scrutiny but are at the same time also more often falsely accused, which can lead to life-impacting, even deadly results. As Benjamin states, ‘[who] is seen and under what terms holds a mirror onto more far-reaching forms of power and inequality’ [99].

Likewise, attempts for more diverse datasets might equally lead to troubling effects. In Chapter 4, *Technological benevolence*, Benjamin challenges technological products that aim for more diversity, but which merely offer temporary fixes, sometimes even deepening the issue. In this chapter, Benjamin turns her attention to the company, *Diversity Inc.*, which supports companies in their targeted advertising through ‘ethnic targeting’. As the companies themselves are legally not allowed to collect data on race and

ethnicity, *Diversity Inc.* offers them to predict these based on the names of individuals. However, since names are not sufficient for an accurate prediction, as *Diversity Inc.* states, they also use attributes such as socioeconomic status, location, or gender. Thus, similar to the example of the recidivism algorithm, the diverse attributes function as proxies for the ethnicity of the companies' customers [144ff.]. In this way, historic forms of redlining that were used to separate neighborhoods by race 'are now a source of profit for those who want to gear their products and services to different racial-ethnic groups' [147].

Throughout her book, Benjamin draws upon various examples that cut across different technological levels. Next to these mentioned here, which primarily rely on machine learning algorithms, Benjamin also includes other cases such as autocorrection in text editors or the idea of a social credit system in an episode of the TV show, *Black Mirror*. While Benjamin summarizes the core arguments around the debate of algorithmic discrimination, they tend to get lost in a stream of illustrative examples, resulting in an associative analysis. This also weakens her argumentative structure throughout the chapters, making it hard to understand her principal argument and how each chapter relates.

Furthermore, the differentiation of technologies in their technical characteristics lacks precision. Particularly Benjamin's decision to frame her research objects – 'from everyday apps to complex algorithms' [7] – under the loaded term of the 'robot' as 'any machine that can perform a task, simple or complex, directed by humans or programmed to operate automatically' [54-55] hinders a deeper understanding of the technologies and their underlying logics. Likewise, the notion of the 'robot' further reinforces the image of technologies as automatic and independent agents, which contradicts her explicit stance against technological determinism: 'We observe not only that any given social order is impacted by technological development, as determinists would argue, but that social norms, ideologies, and practices are a constitutive part of technical design' [41]. Despite this human impact in their design, however, technologies themselves should also be taken seriously in their ability to create meaning and thus bring political relevance (Rieder, 2020: 53ff.). Only this perspective enables us to shift our attention to the inner logic of these systems in their

design and understand, why it is precisely these social effects that they bring about, and which Benjamin aptly highlights. This outlook is vital when it comes to the formulation of alternatives.

### **So, what now? A call for action**

With her book, Benjamin offers an impressive account of a range of discriminatory technological designs and their entanglement with structural racism inherent in society. She insightfully brings together these technological systems and their social effects with her wealth of knowledge in the field of critical race studies. The question that remains is what we can do to address these problematics. Or, in the words of the author: ‘What does an emancipatory approach to tech entail?’ [173].

In Chapter 5, *Retooling solidarity, reimagining justice*, Benjamin illustrates with the cases of the apps, *Appolition* and *Promise*, to what end a technological system is developed and how values and interests inscribed in that goal inevitably affect society. Both applications address the problem of ‘pretrial detention, which impacts disproportionately Black and Latinx people who cannot afford bail’ [164]. *Appolition* works by converting its users’ daily change into bail money and is founded in an abolitionist belief [162]. *Promise*, on the other hand, costs \$17 a day, and it functions by ‘tracking individuals via the app and GPS monitoring’ [164], thus making pretrial detention obsolete. This application receives large capital payments from investors. According to Benjamin, *Appolition* is a ‘technology with an emancipatory ethos, a tool of solidarity that directs resources to getting people literally free’ [163]. On the other hand, *Promise* further manifests the prison industrial complex by providing new forms of digital imprisonment [164].

With this comparison, Benjamin underlines the need for different values than economic interests, demanding ‘a socially conscious approach to tech development that would require prioritizing equity over efficiency, social good over market imperatives’ [183]. She further points out the need for new narratives and demands surrounding technology that work towards overcoming structural oppression and racism [197]. The reshaping of the

narratives that accompany technologies goes hand in hand with the call for their active re-design. The book thus functions as a ‘field guide’ [36], or ‘conceptual toolkit’ [41], supporting the fight for ‘justice-oriented design practices’ [48].<sup>2</sup> Here, she states that ‘[efforts] to combat coded inequity cannot be limited to industry, nonprofit, and government actors, but must include community-based organizations that offer a vital set of counternarratives about the social and political dimensions of the New Jim Code’ [188].<sup>3</sup>

While this stance is significant, it does not clarify what the redesign of technologies should look like in concrete terms, aside from community involvement. Instead, the technological system of the ‘robot’ remains the unattainable other. Benjamin for instance describes, how ‘the way robots can be racist often remains a mystery’ [53] and that ‘there is an enormous mystique around computer codes, which hides the human biases involved in technical design’ [78]. Further, when Benjamin draws on the example of *Appolition* as a positive example for her demand of ‘abolitionist tool-making’, she falls back on a view of technologies as tools for a specific purpose. However, in order to gain a deeper understanding of its societal impact, as I want to argue, it is necessary to not only think from the outside of the technological artifact in terms of its effects, but rather to gain a deeper understanding of it by looking inside the supposed ‘black box’. In fact, recent research shows evidence that the technical principles behind algorithmic systems are actually accessible (e.g. Bechmann and Bowker,

---

<sup>2</sup> Benjamin grounds her call for ‘design justice’ on work done by researcher Sasha Costanza-Chock (2018) and a network of designers, software developers, activists and researchers that aim to think ‘more about the *process* and power dynamics of design across multiple axes of oppression’ [175]. Catherine D’Ignazio and Lauren F. Klein, authors of the book *Data feminism*, similarly demand to ‘understand and design systems that address the source of the bias: structural oppression’ (D’Ignazio and Klein, 2020: 63). Hence, they state that ‘[s]tarting from the assumption that oppression is the problem, not bias, leads to fundamentally different decisions about what to work on, who to work with, and when to stand up and say that a problem cannot and should not be solved by data and technology’ (*ibid.*).

<sup>3</sup> Throughout her book and in the appendix, Benjamin assembles a variety of US-based tech and social justice institutes as points for further research [235].



2019; Mackenzie, 2017; Rieder, 2020), thus offering precisely the analytical depth necessary for redesigning technologies.

At this point, concepts such as representation, classification and discrimination are not only captivating in their cultural significance regarding knowledge production and the filtering of information, but they are also specific technological operations that require investigation. What, for instance, does it mean for the design of alternative technologies, when a significant and inevitable part of their operations is discrimination (Apprich et al., 2019)? Or, on a more specific level, what happens exactly when a machine learning algorithm performs a classification? What assumptions are leading the process and how are certain entities represented through specific categories and labels? How are processes of inclusion and exclusion taking place in detail, and who is taking these decisions? In this line of thought, Wendy Chun (2021) shows how, fundamentally, machine learning algorithms operate based on the homophily principle, which means that algorithms discriminate data by referring to notions of similarity. Consequently, segregation emerges in online networks, resulting in the creation of echo chambers and discriminatory effects. Precisely at this point, one could start and ask, with regard to concrete technological structures, how algorithmic systems could look that take heterophily or diversity as their starting point. Then, of course, as Safiya Noble writes, ‘an app will not save us’ (Noble, 2018: 165, cited after Benjamin, 2019: 179). However, this does not mean that we should not engage with the intricate details of the technological systems that we are facing today. A comprehensive understanding of their inner workings might lead us to a better understanding of how they function and what assumptions they rely on. Starting from there, we can think of how we might redesign technologies, integrating the values that Benjamin, and the data justice projects she refers to, so compellingly demand.

## references

Alexander, M. (2012) *The new Jim Crow: Mass incarceration in the age of colorblind racism*. New York: New Press.

- Angwin, J., J. Larson, S. Mattu and L. Kirchner (2016) 'Machine bias. There's software used across the country to predict future criminals. And it's biased against Blacks', *ProPublica* [<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>].
- Apprich, C., W.H.K. Chun, F. Cramer and S. Steyerl (eds.) (2019) *Pattern discrimination*. Lüneburg/Minneapolis: meson press/Minnesota Press.
- Bechmann, A. and G.C. Bowker (2019) 'Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media', *Big Data & Society*, 6(1): 1-11.
- Benjamin, R. (2019) *Race after technology: Abolitionist tools for the New Jim Code*. Cambridge/Medford: Polity Press.
- Berry, D.M. (2017) 'Prolegomenon to a media theory on machine learning: Compute-computing and compute-computed', *Media Theory*, 1(1): 74-87.
- Browne, S. (2015) *Dark matters: On the surveillance of blackness*. Durham: Duke University Press.
- Chun, W.H.K. (2009) 'Introduction: Race and/as technology: Or how to do things with race', *Camera Obscura*, 24(1/70): 7-35.
- Chun, W.H.K. (2021) *Discriminating data: Correlation, neighborhoods, and the new politics of recognition*. Cambridge: MIT Press.
- Costanza-Chock, S. (2018) 'Design justice: Towards an intersectional feminist framework for design theory and practice', *Proceedings of the Design Research Society* 2018.
- D'Ignazio, C. and L.F. Klein (2020) *Data feminism*. Cambridge: MIT Press.
- Eubanks, V. (2017) *Automating inequality*. New York: St. Martin's Press.
- Mackenzie, A. (2017) *Machine learners: Archaeology of a data practice*. Cambridge: MIT Press.
- Noble, S.U. (2018) *Algorithms of oppression*. New York: New York University Press.
- O'Neil, C. (2016) *Weapons of math destruction*. New York: Penguin Random House.

Omi, M. and H. Winant (1986) *Racial formation in the United States: From the 1960s to the 1980s*. New York: Routledge.

Rieder, B. (2020) *Engines of order: A mechanology of algorithmic techniques*. Amsterdam: Amsterdam University Press.

## **the author**

Inga Luchs is PhD Candidate in the Department for Media and Journalism Studies at the Rijksuniversiteit Groningen. Her PhD project revolves around the topic of algorithmic discrimination. From a decidedly technical perspective, she critically investigates the key principles of machine learning.

Email: [i.m.luchs@rug.nl](mailto:i.m.luchs@rug.nl)