1
Engineered Inequity
Are Robots Racist?

WELCOME TO THE FIRST INTERNATIONAL BEAUTY CONTEST JUDGED BY ARTIFICIAL INTELLIGENCE.

So goes the cheery announcement for Beauty AI, an initiative developed by the Australian- and Hong Kongbased organization Youth Laboratories in conjunction with a number of companies who worked together to stage the first ever beauty contest judged by robots (Figure 1.1).[1] The venture involved a few seemingly straightforward steps:

1. Contestants download the Beauty AI app.
2. Contestants make a selfie.
3. Robot jury examines all the photos.
4. Robot jury chooses a king and a queen.
5. News spreads around the world.

As for the rules, participants were not allowed to wear makeup or glasses or to don a beard. Robot judges were programmed to assess contestants on the basis of wrinkles, face symmetry, skin color, gender, age group, ethnicity, and “many other parameters.” Over 6,000 submissions from approximately 100 countries poured in. *What could possibly go wrong?*

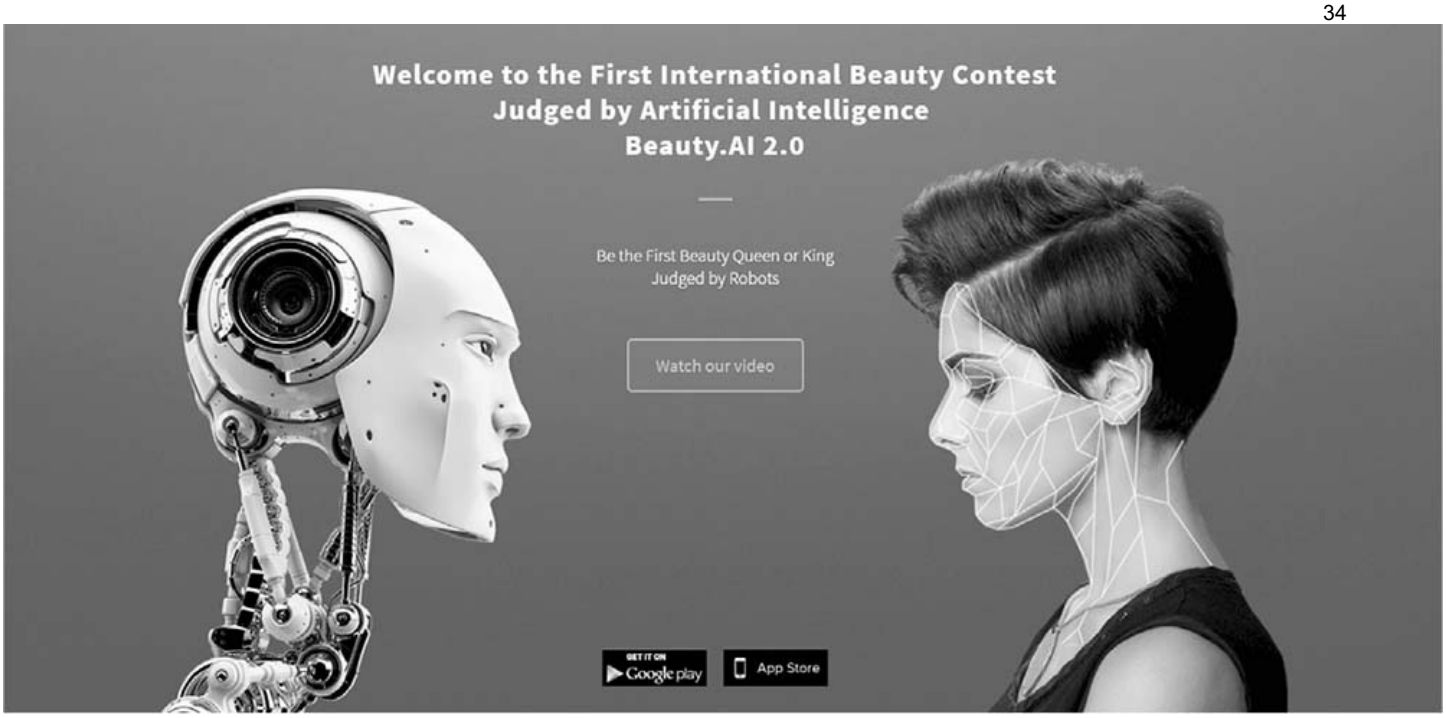

Figure 1.1 Beauty AI

Source: http://beauty.ai

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."[2] 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 algorithm's embedded preferences.[3] Beauty, in short, is in the trained eye of the algorithm.

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."[4] 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."[5] And it is not just the human programmers' preference 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*.

In addition to the skewed racial results, the framing of Beauty AI as a kind of preventative public health initiative raises the stakes considerably. The team of biogerontologists and data scientists working with Beauty AI explained that valuable information about people's health can be gleaned by "just processing their photos" and that, ultimately, the hope is to "find effective ways to slow down ageing and help people look healthy and beautiful."[6] Given the

overwhelming Whiteness of the winners and the conflation of socially biased notions of beauty and health, darker people are implicitly coded as unhealthy and unfit – assumptions that are at the heart of scientific racism and eugenic ideology and policies.

Deep learning is a subfield of machine learning in which "depth" refers to the layers of abstraction that a computer program makes, learning more "complicated concepts by building them out of simpler ones."[7] With Beauty AI, deep learning was applied to image recognition; but it is also a method used for speech recognition, natural language processing, video game and board game programs, and even medical diagnosis. Social media filtering is the most common example of deep learning at work, as when Facebook auto-tags your photos with friends' names or apps that decide which news and advertisements to show you to increase the chances that you'll click. Within machine learning there is a distinction between "supervised" and "unsupervised" learning. Beauty AI was supervised, because the images used as training data were pre-labeled, whereas unsupervised deep learning uses data with very few labels. Mark Zuckerberg refers to deep learning as "the theory of the mind … How do we model – in machines – what human users are interested in and are going to do?"[8] But the question for us is, is there only *one* theory of the mind, and *whose mind* is it modeled on?

It may be tempting to write off Beauty AI as an inane experiment or harmless vanity project, an unfortunate glitch in the otherwise neutral development of technology for the common good. But, as explored in the pages ahead, such a conclusion is naïve at best. Robots exemplify how race is a form of technology itself, as the algorithmic judgments of Beauty AI extend well beyond adjudicating attractiveness and into questions of health, intelligence, criminality, employment, and many other fields, in which innovative techniques give rise to newfangled forms of racial discrimination. Almost every day a new headline sounds the alarm, alerting us to the New Jim Code:

"Some algorithms are racist"

"We have a problem: Racist and sexist robots"

"Robots aren't sexist and racist, you are"

"Robotic racists: AI technologies could inherit their creators' biases"

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.

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.[9]

For example, in *Automated Inequality* Virginia Eubanks (2018) documents the steady incorporation of predictive analytics by US social welfare agencies. Among other promises, automated decisions aim to mitigate fraud by depersonalizing the process and by determining who is eligible for benefits.[10] But, as she documents, these technical fixes, often promoted as benefiting society, end up hurting the most vulnerable, sometimes with deadly results. Her point is not that human caseworkers are less biased than machines – there are, after all, numerous studies showing how caseworkers actively discriminate against racialized groups while aiding White applicants deemed more deserving.[11] Rather, as Eubanks emphasizes, automated welfare decisions are not magically fairer than their human counterparts. Discrimination is displaced and accountability is outsourced in this postdemocratic approach to governing social life.[12]

So, how do we rethink our relationship to technology? The answer partly lies in how we think about race itself and specifically the issues of intentionality and visibility.

I Tinker, Therefore I Am

Humans are toolmakers. And robots, we might say, are humanity's finest handiwork. In popular culture, robots are typically portrayed as humanoids, more efficient and less sentimental than *Homo sapiens*. At times, robots are depicted as having human-like struggles, wrestling with emotions and an awakening consciousness that blurs the line between maker and made. Studies about how humans perceive robots indicate that, when that line becomes too blurred, it tends to freak people out. The technical term for it is the "uncanny valley" – which indicates the dip in empathy and increase in revulsion that people experience when a robot appears to be too much like us.[13]

Robots are a diverse lot, with as many types as there are tasks to complete and desires to be met: domestic robots; military and police robots; sex robots; therapeutic robots – and more. A robot is any machine that can perform a task, simple or complex, directed by humans or programmed to operate automatically. The most advanced are smart machines designed to learn from and adapt to their environments, created to become independent of their makers. We might like to think that robotic concerns are a modern phenomenon,[14] but our fascination with automata goes back to the Middle Ages, if not before.[15]

In *An Anthropology of Robots and AI*, Kathleen Richardson observes that the robot has "historically been a way to talk about dehumanization" and, I would add, *not* talk about racialization.[16] The etymology of the word robot is Czech; it comes from a word for "compulsory service," itself drawn from the Slav *robota* ("servitude, hardship").[17] So yes, people have used robots to express anxieties over annihilation, including over the massive

threat of war machines. But robots also convey an ongoing agitation about human domination over other humans![18]

The first cultural representation that employed the word robot was a 1920 play by a Czech writer whose machine was a factory worker of limited consciousness.[19] Social domination characterized the cultural laboratory in which robots were originally imagined. And, technically, *people* were the first robots. Consider media studies scholar Anna Everett's earliest experiences using a computer:

> In powering up my PC, I am confronted with the DOS-based text that gave me pause … "Pri. Master Disk, Pri. Slave Disk, Sec. Master, Sec. Slave." Programmed here is a virtual hierarchy organizing my computer's software operations … I often wondered why the programmers chose such signifiers that hark back to our nation's ignominious past … And even though I resisted the presumption of a racial affront or intentionality in such a peculiar deployment of the slave and master coupling, its choice as a signifier of the computer's operations nonetheless struck me.[20]

Similarly, a 1957 article in *Mechanix Illustrated*, a popular "how-to-do" magazine that ran from 1928 to 2001, predicted that, by 1965:

> Slavery will be back! We'll all have personal slaves again … [who will] dress you, comb your hair and serve meals in a jiffy. Don't be alarmed. We mean robot "slaves."[21]

It goes without saying that readers, so casually hailed as "we," are not the descendants of those whom Lincoln freed. This fact alone offers a glimpse into the implicit Whiteness of early tech culture. We cannot assume that the hierarchical values and desires that are projected onto "we" – *We, the People* with inalienable rights and not *You, the Enslaved* who serve us meals – are simply a thing of the past (Figure 1.2).

Coincidentally, on my way to give a talk – mostly to science, technology, engineering, and mathematics (STEM) students at Harvey Mudd College – that I had planned to kick off with this *Mechanix* ad, I passed two men in the airport restaurant and overheard one say to the other: "I just want someone I can push around …" So simple yet so profound in articulating a dominant and dominating *theory of power* that many more people feel emboldened to state, unvarnished, in the age of Trump. *Push around*? I wondered, in the context of work or dating or any number of interactions. The slavebot, it seems, has a ready market!

For those of us who believe in a more egalitarian notion of power, of collective empowerment without domination, how we imagine our relation to robots offers a mirror for thinking through and against race as technology.

The robots are coming! When they do, you'll command a host of push-button servants.

You'll Own

By O. O. Binder

IN 1863, Abe Lincoln freed the slaves. But by 1965, slavery will be back! We'll all have personal slaves again, only this time we won't fight a Civil War over them. Slavery will be here to stay.

Don't be alarmed. We mean robot "slaves." Let's take a peek into the future

Robots will dress you, comb your hair and serve meals in a jiffy.

Figure 1.2 Robot Slaves

Source: Binder 1957

It turns out that the disposability of robots and the denigration of racialized populations go hand in hand. We can see this when police officers use "throwbots" – "a lightweight, ruggedized platform that can literally be thrown into position, then remotely controlled from a position of safety" – to collect video and audio surveillance for use by officers. In the words of a member of one of these tactical teams, "[t]he most significant advantage of the throwable robot is that it 'allows them [*sc*. the officers] to own the real estate with their eyes, before they pay for it with their bodies.'"[22] Robots are not the only ones sacrificed on the altar of public safety. So too are the many *Black* victims whose very bodies become the real estate that police officers own in their trigger-happy quest to keep the peace. The intertwining history of machines and slaves, in short, is not simply the stuff of fluff magazine articles.[23]

While many dystopic predictions signal a worry that humans may one day be enslaved by machines, the current reality is that the tech labor force is already deeply unequal across racial and gender lines. Although not the same as the structure of enslavement that serves as an analogy for unfreedom, Silicon Valley's hierarchy consists of the highest-paid creatives and entrepreneurs, who are comprised of White men and a few White women, and the lowest-paid manual laborers – "those cleaning their offices and assembling circuit boards," in other words "immigrants and outsourced labor, often women living in the global south," who usually perform this kind of work.[24] The "diasporic diversity" embodied by South Asian and Asian American tech workforce does not challenge this hierarchy, because they continue to be viewed as a "new digital 'different caste.'" As Nakamura notes, "no amount of work can make them part of the digital economy as 'entrepreneurs' or the 'new economic men.'"[25] Racism, in this way, is a technology that is "built into the tech industry."[26] But how does racism "get inside" and operate through new forms of technology?

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.[27] 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."[28] As machines become more "intelligent," that is, as they learn to think more like humans, they are likely to become more racist. But this is not inevitable, so long as we begin to take seriously and address the matter of how racism structures the social and technical components of design.

Raising Robots

So, are robots racist? Not if by "racism" we only mean white hoods and racial slurs.[29] Too

often people assume that racism and other forms of bias must be triggered by an *explicit* intent to harm; for example, linguist John McWhorter argued in *Time* magazine that "[m]achines cannot, themselves, be racists. Even equipped with artificial intelligence, they have neither brains nor intention."[30] But this assumes that self-conscious intention is what makes something racist. Those working in the belly of the tech industry know that this conflation will not hold up to public scrutiny. As one Google representative lamented, "[r]ather than treating malfunctioning algorithms as malfunctioning machines ('classification errors'), we are increasingly treating tech like asshole humans." He went on to propose that "we [programmers] need to stop the machine from behaving like a jerk because it can look like it is being offensive on purpose."[31] If machines are programmed to carry out tasks, both they and their designers are guided by some purpose, that is to say, intention. And in the face of discriminatory effects, if those with the power to design differently choose business as usual, then they are perpetuating a racist system whether or not they are card-carrying members of their local chapter of Black Lives Matter.

Robots are not sentient beings, sure, but racism flourishes well beyond hate-filled hearts.[32] 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." Thinking with Baldwin, someone who insists on his own racial innocence despite all evidence to the contrary "turns himself into a monster."[33] No malice needed, no N-word required, just lack of concern for how the past shapes the present – and, in this case, the US government's explicit intention to concentrate wealth in the hands of White Americans, in the form of housing and economic policies.[34] Detachment in the face of this history ensures its ongoing codification. Let us not forget that databases, just like courtrooms, banks, and emergency rooms, do not contain organic brains. Yet legal codes, financial practices, and medical care often produce deeply racist outcomes.

The intention to harm or exclude may guide some technical design decisions. Yet even when they do, these motivations often stand in tension with aims framed more benevolently. Even police robots who can use lethal force while protecting officers from harm are clothed in the rhetoric of public safety.[35] This is why we must separate "intentionality" from its strictly negative connotation in the context of racist practices, and examine how aiming to "do good" can very well coexist with forms of malice and neglect.[36] In fact a do-gooding ethos often serves as a moral cover for harmful decisions. 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."[37] 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.[38]

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 a company pie chart, concealing rather than undoing the racist status quo.[39]

So, can robots – and, by extension, other types of technologies – be racist? Of course they can. Robots, designed in a world drenched in racism, will find it nearly impossible to stay dry. To a certain extent, they learn to speak the coded language of their human parents – not only programmers but all of us online who contribute to "naturally occurring" datasets on which AI learn. Just like diverse programmers, Black and Latinx police officers are known to engage in racial profiling alongside their White colleagues, though they are also the target of harassment in a way their White counterparts are not.[40] One's individual racial identity offers no surefire insulation from the prevailing ideologies.[41] There is no need to identify "giggling programmers" self-consciously seeking to denigrate one particular group as evidence of discriminatory design. Instead, so much of what is routine, reasonable, intuitive, and codified reproduces unjust social arrangements, without ever burning a cross to shine light on the problem.[42]

A representative of Microsoft likened the care they must exercise when they create and sell predictive algorithms to their customers with "giving a puppy to a three-year-old. You can't just deploy it and leave it alone because it will decay over time."[43] Likewise, describing the many controversies that surround AI, a Google representative said: "We are in the uncomfortable birthing stage of artificial intelligence."[44] Zeros and ones, if we are not careful, could deepen the divides between haves and have-nots, between the deserving and the undeserving – rusty value judgments embedded in shiny new systems.

Interestingly, the MIT data scientists interviewed by anthropologist Kathleen Richardson

> were conscious of race, class and gender, and none wanted to reproduce these normative stereotypes in the robots they created … [They] avoided racially marking the "skin" of their creations … preferred to keep their machines genderless, and did not speak in class-marked categories of their robots as "servants" or "workers," but companions, friends and children.[45]

Richardson contrasts her findings to that of anthropologist Stefan Helmreich, whose pioneering study of artificial life in the 1990s depicts researchers as "ignorant of normative models of sex, race, gender and class that are refigured in the computer simulations of artificial life."[46] But perhaps the contrast is overdrawn, given that colorblind, gender-neutral, and class-avoidant approaches to tech development are another avenue for coding inequity. If data scientists do indeed treat their robots like children, as Richardson describes, then I propose a race-conscious approach to parenting artificial life – one that does not feign colorblindness. But where should we start?

Automating Anti-Blackness

As it happens, the term "stereotype" offers a useful entry point for thinking about the default settings of technology and society. It first referred to a practice in the printing trade whereby a solid plate called a "stereo" (from the ancient Greek adjective *stereos*, "firm," "solid") was used to make copies. The duplicate was called a "stereotype."[47] The term evolved; in 1850 it designated an "image perpetuated without change" and in 1922 was taken up in its contemporary iteration, to refer to shorthand attributes and beliefs about different groups. The etymology of this term, which is so prominent in everyday conceptions of racism, urges a more sustained investigation of the interconnections between technical and social systems.

To be sure, the explicit codification of racial stereotypes in computer systems is only one form of discriminatory design. Employers resort to credit scores to decide whether to hire someone, companies use algorithms to tailor online advertisements to prospective customers, judges employ automated risk assessment tools to make sentencing and parole decisions, and public health officials apply digital surveillance techniques to decide which city blocks to focus medical resources. Such programs are able to sift and sort a much larger set of data than their human counterparts, but they may also reproduce long-standing forms of structural inequality and colorblind racism.

And 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.[48] Paradoxically, automation is often presented as a solution to human bias – a way to avoid the pitfalls of prejudicial thinking by making decisions on the basis of objective calculations and scores. So, to understand racist robots, we must focus less on their intended uses and more on their actions. Sociologist of technology Zeynep Tufekci describes algorithms as "computational agents who are not alive, but who act in the world."[49] In a different vein, philosopher Donna Haraway's (1991) classic *Simians, Cyborgs and Women* narrates the blurred boundary between organisms and machines, describing how "myth and tool mutually constitute each other."[50] She describes technologies as "frozen moments" that allow us to observe otherwise "fluid social interactions" at work. These "formalizations" are also instruments that enforce meaning – including, I would add, racialized meanings – and thus help construct the social world.[51] Biased bots and all their coded cousins could also help subvert the status quo by exposing and authenticating the existence of systemic inequality and thus by holding up a "black mirror" to society,[52] challenging us humans to come to grips with our deeply held cultural and institutionalized biases.[53]

Consider the simple corrections of our computer systems, where words that signal undue privilege are not legible. The red line tells us that only one of these phenomena, underserved and overserved, is legitimate while the other is a mistake, a myth (Figure 1.3).

But power is, if anything, relational. If someone is experiencing the underside of an unjust system, others, then, are experiencing its upside. If employers are passing up your job application because they associate negative qualities with your name, then there are more

jobs available for more appealing candidates. If, however, we do not have a word to describe these excess jobs, power dynamics are harder to discuss, much less intervene in. If you try this exercise today, your spellcheck is likely to recognize both words, which reminds us that it is possible to change technical systems so that they do not obscure or distort our understanding and experience of social systems. And, while this is a relatively simple update, we must make the same demand of more complex forms of coded inequity and tune into the socially proscribed forms of (in)visibility that structure their design.

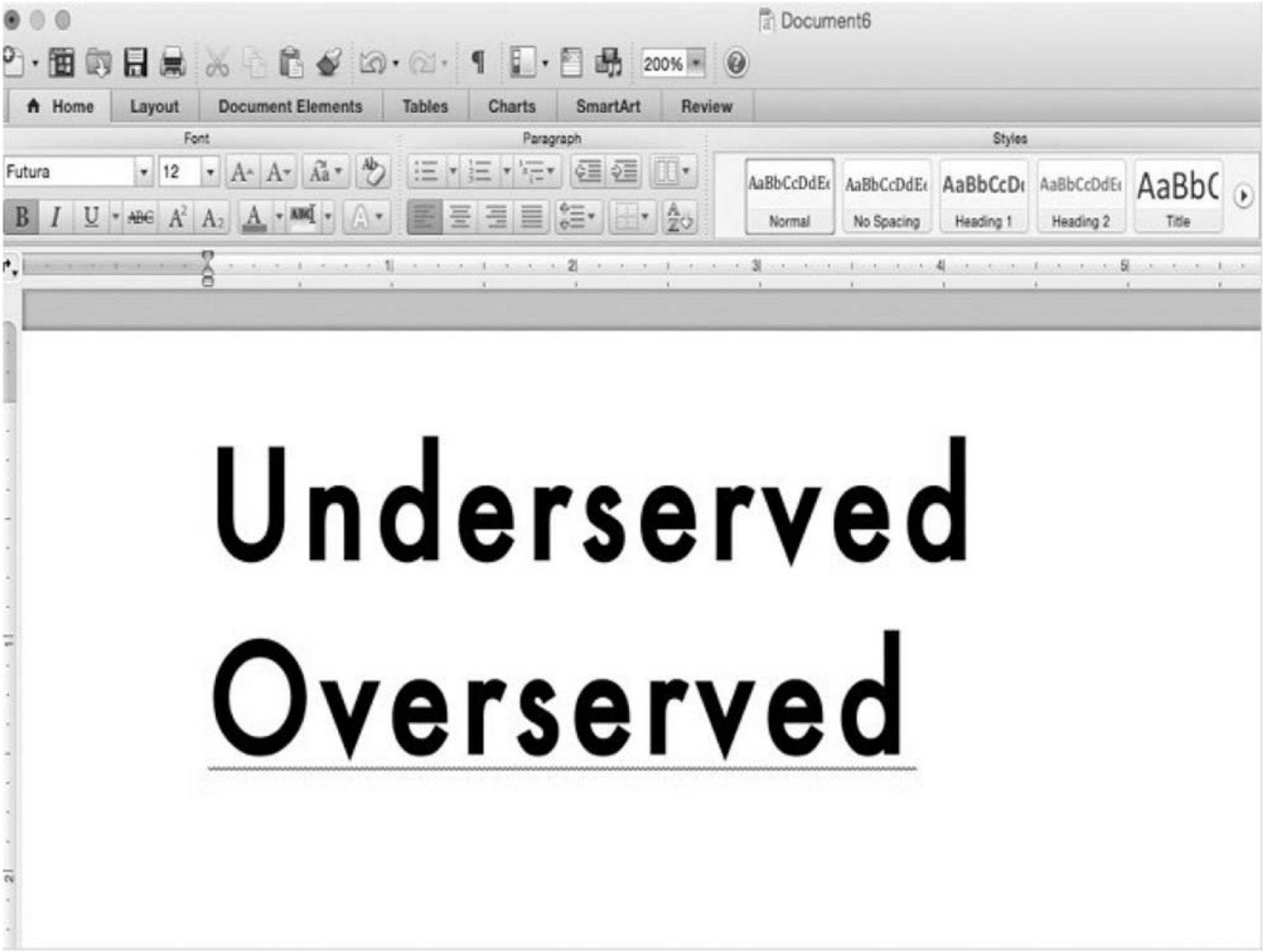

Figure 1.3 Overserved

If we look strictly at the technical features of, say, automated soap dispensers and predictive crime algorithms, we may be tempted to home in on their differences. When we consider the stakes, too, we might dismiss the former as relatively harmless, and even a distraction from the dangers posed by the latter. But rather than starting with these distinctions, perhaps there is something to be gained by putting them in the same frame to tease out possible relationships. For instance, the very idea of hygiene – cleaning one's hands and "cleaning up" a neighborhood – echoes a racialized vocabulary. Like the Beauty AI competition, many advertisements for soap conflate darker skin tones with unattractiveness and more specifically with dirtiness, as did an ad from the 1940s where a White child turns to a Black

child and asks, “Why doesn’t your mama wash you with fairy soap?” Or another one, from 2017, where a Black woman changes into a White woman after using Dove soap. The idea of hygiene, in other words, has been consistently racialized, all the way from marketing to public policy. In fact the most common euphemism for eugenics was “racial hygiene”: ridding the body politic of unwanted populations would be akin to ridding the body of unwanted germs. Nowadays we often associate racial hygienists with the Nazi holocaust, but many early proponents were the American progressives who understood eugenics to work as a social uplift and a form of Americanization. The ancient Greek etymon, *eugeneia* (εὐγένεια), meant “good birth,” and this etymological association should remind us how promises of goodness often hide harmful practices. As Margaret Atwood writes, “Better never means better for everyone … It always means worse, for some.”

Take a seemingly mundane tool for enforcing segregation – separate water fountains – which is now an iconic symbol for the larger system of Jim Crow. In isolation from the broader context of racial classification and political oppression, a “colored” water fountain could be considered trivial, though in many cases the path from segregated public facilities to routine public lynching was not very long. Similarly, it is tempting to view a “Whites only” soap dispenser as a trivial inconvenience. In a viral video of two individuals, White and Black, who show that their hotel soap dispenser does not work for the latter, they are giggling as they expose the problem. But when we situate in a broader racial context what appears to be an innocent oversight, the path from restroom to courtroom might be shorter than we expect.

That said, there is a straightforward explanation when it comes to the soap dispenser: near infrared technology requires light to bounce back from the user and activate the sensor, so skin with more melanin, absorbing as it does more light, does not trigger the sensor. But this strictly technical account says nothing about why this particular sensor mechanism was used, whether there are other options, which recognize a broader spectrum of skin tones, and how this problem was overlooked during development and testing, well before the dispenser was installed. Like segregated water fountains of a previous era, the discriminatory soap dispenser offers a window onto a wider social terrain. As the soap dispenser is, technically, a robot, this discussion helps us consider the racism of robots and the social world in which they are designed.

For instance, we might reflect upon the fact that the infrared technology of an automated soap dispenser treats certain skin tones as normative and upon the reason why this technology renders Black people invisible when they hope to be seen, while other technologies, for example facial recognition for police surveillance, make them hypervisible when they seek privacy. When we draw different technologies into the same frame, the distinction between “trivial” and “consequential” breaks down and we can begin to understand how Blackness can be both marginal and focal to tech development. For this reason I suggest that we hold off on drawing too many bright lines – good versus bad, intended versus unwitting, trivial versus consequential. Sara Wachter-Boettcher, the author of *Technically Wrong*, puts it thus: “If tech companies can’t get the basics right … why should we trust them to provide solutions to massive societal problems?”[54] The issue is not simply that innovation and inequity can go hand in hand but that a view of technology as value-free

means that we are less likely to question the New Jim Code in the same way we would the unjust laws of a previous era, assuming in the process that our hands are clean.

Engineered Inequity

In one of my favorite episodes of the TV show *Black Mirror*, we enter a world structured by an elaborate social credit system that shapes every encounter, from buying a coffee to getting a home loan. Every interaction ends with people awarding points to one another through an app on their phones; but not all the points are created equal. Titled "Nosedive," the episode follows the emotional and social spiral of the main protagonist, Lacie, as she pursues the higher rank she needs in order to qualify for an apartment in a fancy new housing development. When Lacie goes to meet with a points coach to find out her options, he tells her that the only way to increase her rank in such a short time is to get "up votes from quality people. Impress those upscale folks, you'll gain velocity on your arc and there's your boost." Lacie's routine of exchanging five stars with service workers and other "mid- to low-range folks" won't cut it if she wants to improve her score quickly. As the title of the series suggests, *Black Mirror* offers a vivid reflection on the social dimensions of technology – where we *are* and where we might be going with just a few more clicks in the same direction. And, although the racialized dimensions are not often made very explicit, there is a scene toward the beginning of the episode when Lacie notices all her co-workers conspiring to purposely lower the ranking of a Black colleague and forcing him into a subservient position as he tries to win back their esteem … an explicit illustration of the New Jim Code.

When it comes to engineered inequity, there are many different types of "social credit" programs in various phases of prototype and implementation that are used for scoring and ranking populations in ways that reproduce and even amplify existing social hierarchies. Many of these come wrapped in the packaging of progress. And, while the idiom of the New Jim Code draws on the history of racial domination in the United States as a touchstone for technologically mediated injustice, our focus must necessarily reach beyond national borders and trouble the notion that racial discrimination is isolated and limited to one country, when a whole host of cross-cutting social ideologies make that impossible.

Already being implemented, China's social credit system is an exemplar of explicit ranking with far-reaching consequences. What's more, *Black Mirror* is referenced in many of the news reports of China's experiment, which started in 2014, with the State Council announcing its plans to develop a way to score the trustworthiness of citizens. The government system, which will require mandatory enrollment starting from 2020, builds on rating schemes currently used by private companies.

Using proprietary algorithms, these apps track not only financial history, for instance whether someone pays his bills on time or repays her loans, but also many other variables, such as one's educational, work, and criminal history. As they track all one's purchases, donations, and leisure activities, something like too much time spent playing video games marks the person as "idle" (for which points may be docked), whereas an activity like buying diapers

suggests that one is "responsible." As one observer put it, "the system not only investigates behaviour – it shapes it. It 'nudges' citizens away from purchases and behaviours the government does not like."[55] Most alarmingly (as this relates directly to the New Jim Code), residents of China's Xinjiang, a predominantly Muslim province, are already being forced to download an app that aims to track "terrorist and illegal content."

Lest we be tempted to think that engineered inequity is a problem "over there," just recall Donald Trump's idea to register all Muslims in the United States on an electronic database – not to mention companies like Facebook, Google, and Instagram, which already collect the type of data employed in China's social credit system. Facebook has even patented a scoring system, though it hedges when asked whether it will ever develop it further. Even as distinct histories, politics, and social hierarchies shape the specific convergence of innovation and inequity in different contexts, it is common to observe, across this variation, a similar deployment of buzzwords, platitudes, and promises.

What sets China apart (for now) is that all those tracked behaviors are already being rated and folded into a "citizen score" that opens or shuts doors, depending on one's ranking.[56] People are given low marks for political misdeeds such as "spreading rumors" about government officials, for financial misdeeds such as failing to pay a court fine, or social misdeeds such as spending too much time playing video games. A low score brings on a number of penalties and restrictions, barring people from opportunities such as a job or a mortgage and prohibiting certain purchases, for example plane tickets or train passes.[57] The chief executive of one of the companies that pioneered the scoring system says that it "will ensure that the bad people in society don't have a place to go, while good people can move freely and without obstruction."[58]

Indeed, it is not only the desire to move freely, but all the additional privileges that come with a higher score that make it so alluring: faster service, VIP access, no deposits on rentals and hotels – not to mention the admiration of friends and colleagues. Like so many other technological lures, systems that seem to objectively rank people on the basis of merit and things we like, such as trustworthiness, invoke "efficiency" and "progress" as the lingua franca of innovation. China's policy states: "It will forge a public opinion environment where keeping trust is glorious. It will strengthen sincerity in government affairs, commercial sincerity, social sincerity and the construction of judicial credibility."[59] In fact, higher scores have become a new status symbol, even as low scorers are a digital underclass who may, we are told, have an opportunity to climb their way out of the algorithmic gutter.

Even the quality of people in one's network can affect your score – a bizarre scenario that has found its way onto TV shows like *Black Mirror* and *Community*, where even the most fleeting interpersonal interactions produce individual star ratings, thumbs up and down, giving rise to digital elites and subordinates. As Zeynep Tufekci explains, the ubiquitous incitement to "like" content on Facebook is designed to accommodate the desires of marketers and works against the interests of protesters, who want to express dissent by "disliking" particular content.[60] And, no matter how arbitrary or silly the credit (see "meow

meow beenz" in the TV series *Community*), precisely because people and the state invest it with import, the system carries serious consequences for one's quality of life, until finally the pursuit of status spins out of control.

The phenomenon of measuring individuals not only by their behavior but by their networks takes the concept of social capital to a whole new level. In her work on marketplace lenders, sociologist Tamara K. Nopper considers how these companies help produce and rely on what she calls *digital character* – a "profile assessed to make inferences regarding character in terms of credibility, reliability, industriousness, responsibility, morality, and relationship choices."[61] Automated social credit systems make a broader principle of merit-based systems clear: scores assess a person's ability to conform to established definitions of good behavior and valued sociality rather than measuring any intrinsic quality. More importantly, the ideological commitments of dominant groups typically determine what gets awarded credit in the first place, automating social reproduction. This implicates not only race and ethnicity; depending on the fault lines of a given society, merit systems also codify class, caste, sex, gender, religion, and disability oppression (among other factors). The point is that multiple axes of domination typically converge in a single code.

Take the credit associated with the aforementioned categories of playing video games and buying diapers. There are many ways to parse the values embedded in the distinction between the "idle" and the "responsible" citizen so that it lowers the scores of gamers and increases the scores of diaper changers. There is the ableist logic, which labels people who spend a lot of time at home as "unproductive," whether they play video games or deal with a chronic illness; the conflation of economic productivity and upright citizenship is ubiquitous across many societies.

Consider, too, how gender norms are encoded in the value accorded to buying diapers, together with the presumption that parenthood varnishes (and, by extension, childlessness tarnishes) one's character. But one may wonder about the consequences of purchasing too many diapers. Does reproductive excess lower one's credit? Do assumptions about sex and morality, often fashioned by racist and classist views, shape the interpretation of having children and of purchasing diapers? In the United States, for instance, one could imagine the eugenic sensibility that stigmatizes Black women's fertility and celebrates White women's fecundity getting codified through a system that awards points for diapers purchased in suburban zip codes and deducts points for the same item when purchased in not yet gentrified parts of the city – the geography of social worth serving as a proxy for gendered racism and the New Jim Code. In these various scenarios, top-down reproductive policies could give way to a social credit system in which the consequences of low scores are so far-reaching that they could serve as a veritable digital birth control.

In a particularly poignant exchange toward the end of the "Nosedive" episode, Lacie is hitchhiking her way to win the approval of an elite group of acquaintances; and motorists repeatedly pass her by on account of her low status. Even though she knows the reason for being disregarded, when a truck driver of even lower rank kindly offers to give her a ride, Lacie looks down her nose at the woman ("nosedive" indeed). She soon learns that the driver

has purposefully opted out of the coercive point system and, as they make small talk, the trucker says that people assume that, with such a low rank, she must be an "antisocial maniac." Lacie reassures the woman by saying you "seem normal." Finally, the trucker wonders about Lacie's fate: "I mean you're a 2.8 but you don't *look* 2.8." This moment is illuminating as to how abstract quantification gets embodied – that the difference between a 2.8 and a 4.0 kind of person should be self-evident and readable on the (sur)face. This is a key feature of racialization: we take arbitrary qualities (say, social score, or skin color), imbue them with cultural importance, and then act as if they reflected natural qualities in people (and differences between them) that should be obvious just by looking at someone.[62]

In this way speculative fiction offers us a canvas for thinking about the racial vision that we take for granted in our day-to-day lives. The White protagonist, in this case, is barred from housing, transportation, and relationships – a fictional experience that mirrors the forms of ethno-racial exclusions that many groups have actually experienced; and Lacie's low status, just like that of her real-life counterparts, is attributed to some intrinsic quality of her person rather than to the coded inequity that structures her social universe. The app, in this story, builds upon an already existing racial arithmetic, expanding the terms of exclusion to those whose Whiteness once sheltered them from harm. This is the subtext of so much science fiction: the anxiety that, if "we" keep going down this ruinous road, then *we might be next*.

Ultimately the danger of the New Jim Code positioning is that existing social biases are reinforced – yes. But new methods of social control are produced as well. Does this mean that every form of technological prediction or personalization has racist effects? Not necessarily. It means that, whenever we hear the promises of tech being extolled, our antennae should pop up to question what all that hype of "better, faster, fairer" might be hiding and making us ignore. And, when bias and inequity come to light, "lack of intention" to harm is not a viable alibi. One cannot reap the reward when things go right but downplay responsibility when they go wrong.

Notes

1. Visit Beauty.AI First Beauty Contest Judged by Robots, at http://beauty.ai.

2. Pearson 2016b.

3. Pearson 2016b.

4. Levin 2016.

5. Both Harcourt quotations are from Levin 2016.

6. See http://beauty.ai.

7. See https://machinelearningmastery.com/what-is-deep-learning.

8. Metz 2013.

9. Field note, Jack Clark's Keynote Address at the Princeton University AI and Ethics Conference, March 10, 2018.

10. The flip side of personalization is what Eubanks (2018) refers to as an "empathy override." See also Edes 2018.

11. Fox 2012, n.p.

12. "Homelessness is not a systems engineering problem, it's a carpentry problem" (Eubanks 2018, p. 125).

13. The term "uncanny valley" was coined by Masahiro Mori in 1970 and translated into English by Reichardt (1978).

14. But it is worth keeping in mind that many things dubbed "AI" today are, basically, just statistical predictions rebranded in the age of big data – an artificial makeover that engenders more trust as a result. This point was made by Arvind Narayanan in response to a Microsoft case study at a workshop sponsored by the Princeton University Center for Human Values and Center for Informational Technology Policy, October 6, 2017.

15. Truitt 2016.

16. Richardson 2015, p. 5.

17. Richardson 2015, p. 2.

18. As Imani Perry (2018, p. 49) explains, "Mary Shelley's *Frankenstein* provided a literary example of the domestic anxiety regarding slavery and colonialism that resulted from this structure of relations … Frankenstein's monster represented the fear of the monstrous products that threatened to flow from the peculiar institutions. The novel lends itself to being read as a response to slave revolts across the Atlantic world. But it can also be read as simply part of anxiety attendant to a brutal and intimate domination, one in which the impenetrability of the enslaved was already threatening."

19. Richardson 2015, p. 2.

20. Everett 2009, p. 1.

21. Binder 1957.

22. These passages come from a PoliceOne report that cautions us: "as wonderful an asset as they are, they cannot provide a complete picture. The camera eye can only see so much, and there are many critical elements of information that may go undiscovered or unrecognized … Throwable robots provide such an advance in situational awareness that it can be easy to forget that our understanding of the situation is still incomplete" (visit https://www.policeone.com/police-products/police-technology/robots/articles/320406006–5-tactical-considerations-for-throwable-robot-deployment).

23. Rorty 1962.

24. Daniels 2015, p. 1379. See also Crain et al. 2016; Gajjala 2004; Hossfeld 1990; Pitti 2004; Shih 2006.

25. Nakamura 2002, p. 24.

26. Daniels 2013, p. 679.

27. Noble and Tynes 2016.

28. Field note from the Princeton University Center for Human Values and Center for Informational Technology Policy Workshop, October 6, 2017.

29. The notion of "racist robots" is typically employed in popular discourse around AI. I use it as a rhetorical device to open up a discussion about a range of contemporary technologies, most of which are not human-like automata of the kind depicted in films and novels. They include forms of automation integrated in everyday life, like soap dispensers and search engines, bureaucratic interventions that seek to make work more efficient, as in policing and healthcare, and fantastical innovations first imagined in science fiction, such as self-driving cars and crime prediction techniques.

30. McWhorter 2016.

31. Field note from the Princeton University Center for Human Values and Center for Informational Technology Policy Workshop, October 6, 2017.

32. The famed android Lieutenant Commander Data of the hit series *Star Trek* understood well the distinction between inputs and outputs, intent and action. When a roughish captain of a small cargo ship inquired whether Data had ever experienced love, Data responded, "The act or the emotion?" And when the captain replied that they're both the same, Data rejoined, "I believe that statement to be inaccurate, sir." Just as loving behavior does not require gushing Valentine's Day sentiment, so too can discriminatory action be fueled by indifference and disregard, and even by good intention, more than by flaming hatred.

33. Baldwin 1998, p. 129.

34. See https://www.nclc.org/images/pdf/credit_discrimination/InsuranceScoringWhitePaper.pdf.

35. Policeone.com, at https://www.policeone.com/police-products/police-technology/robots.

36. This is brought to life in the 2016 HBO series *Silicon Valley*, which follows a young Steve Jobs type of character, in a parody of the tech industry. In a segment at TechCrunch, a conference where start-up companies present their proof of concept to attract venture capital investment, one presenter after another exclaims, "we're making the world a better place" with each new product that also claims to "revolutionize" some corner of the

industry. See https://longreads.com/2016/06/13/silicon-valley-masterfully skewers-tech-culture.

37. McWhorter 2016.

38. Sociologist Eduardo Bonilla-Silva (2006) argues that, "if racism is systemic, this view of 'good' and 'bad' whites distorts reality" (p. 132). He quotes Albert Memmi saying: "There is a strange enigma associated with the problem of racism. No one, or almost no one, wishes to see themselves as racist; still, racism persists, real and tenacious" (Bonilla-Silva 2006, p. 1).

39. Dobush 2016.

40. Perry explains how racial surveillance does not require a "bogeyman behind the curtain; it is a practice that emerges from our history, conflicts, the interests of capital, and political expediency in the nation and the world … Nowhere is the diffuse and individuated nature of this practice more apparent than in the fact that over-policing is not limited to White officers but is instead systemic" (Perry 2011, p. 105).

41. Calling for a post-intentional analysis of racism, Perry argues that intent is not a good measure of discrimination because it "creates a line of distinction between 'racist' and 'acceptable' that is deceptively clear in the midst of a landscape that is, generally speaking, quite unclear about what racism and racial bias are, who [or what] is engaging in racist behaviors, and how they are doing so" (Perry 2011, p. 21).

42. Schonbrun 2017.

43. Field note from the Princeton University Center for Human Values and Center for Informational Technology Policy Workshop, October 6, 2017.

44. Field note from the Princeton University Center for Human Values and Center for Informational Technology Policy Workshop, October 6, 2017.

45. Richardson 2015, p. 12.

46. Richardson 2015, p. 12; see also Helmreich 1998.

47. See s.v. "stereotype" at https://www.etymonline.com/ word/stereotype (Online Etymology Dictionary).

48. "It is to say, though, that all those inhabiting subject positions of racial power and domination – notably those who are racially White in its various formulations in different racially articulated societies – project and extend racist socialities by default. But the default is not the only position to occupy or in which to invest. One remains with the default because it is given, the easier to inhabit, the sociality of thoughtlessness" (Goldberg 2015, pp. 159–60).

49. Tufekci 2015, p. 207.

50. Haraway 1991, p. 164.

51. Haraway 1991, p. 164.

52. This potential explains the name of the provocative TV series *Black Mirror*.

53. According to Feagin and Elias (2013, p. 936), systemic racism refers to "the foundational, large-scale and inescapable hierarchical system of US racial oppression devised and maintained by whites and directed at people of colour … [It] is foundational to and engineered into its major institutions and organizations."

54. Wachter-Boettcher 2017, p. 200. On the same page, the author also argues that "[w]e'll only be successful in ridding tech of excesses and oversights if we first embrace a new way of seeing the digital tools we rely on – not as a wonder, or even as a villain, but rather as a series of choices that designers and technologists have made. Many of them small: what a button says, where a data set comes from. But each of these choices reinforces beliefs about the world, and the people in it."

55. Botsman 2017.

56. Nguyen 2016.

57. Morris 2018.

58. State Council 2014.

59. State Council 2014.

60. Tufekci 2017, p. 128.

61. Nopper 2019, p. 170.

62. Hacking 2007.