media/rep/

Repositorium für die Medienwissenschaft

Alexander Monea
Race and Computer Vision
2019

https://doi.org/10.25969/mediarep/13540

Veröffentlichungsversion / published version
Sammelbandbeitrag / collection article

DFG Deutsche Forschungsgemeinschaft

Philipps Universität Marburg

Race and Computer Vision

Alexander Monea

Introduction

Any analysis of the intersection of democracy with AI must first and foremost engage the intersection of AI with pre-existing practices of marginalization. In the United States, perhaps no intersection is more salient than that of AI and race. As AI is increasingly positioned as the future of the economy, the military, state bureaucracy, communication and transportation across time and space, in short, as the bedrock of humanity's future, questions of how AI intersects with pre-existing practices of racial marginalization become central. These questions are particularly difficult to answer given the black boxed nature of most contemporary AI systems. While it is certainly a worthwhile endeavor to push for increasing transparency into the datasets and algorithms powering AI systems, that transparency lies in an anticipated future and cannot help us now to analyze the operations of current AI systems. This picture is only complicated by the fact that AI systems, particularly those operating at web scale, are difficult for even their engineers to understand at later stages in their operation. For instance, a programmer may be able to easily describe the seed data and the machine learning algorithm that she started with, but may be completely unable to explain the rationale behind the subsequent classifications that the system learns to make. Again, it is certainly worthwhile to call for AI explicability—namely, requiring AI programmers and engineers to develop systems that can explain their decision-making processes or, in the most extreme case, *only make decisions that can be explained clearly to a human*— this again is an anticipated future that is of little use to answering the immediate question of race and AI which already has dire consequences at this very moment.

So how can an outsider go about critically analyzing the intersection of race and AI in the contemporary moment? This paper will utilize an interdisciplinary methodology that I am calling 'speculative code studies', which combines archival research into press releases, company blog posts, science and technology journalism, and reported instances of technological irregularities with critical code studies research into the available datasets that machine learning algorithms are trained on, analyses of open-sourced variants of black boxed code, and empirical

studies of the outputs of black boxed systems. The goal of such a study is to produce a rigorous, but speculative, analysis of black boxed code. This analysis must always remain speculative, as the actual systems are obfuscated from direct analysis, but the methodology ensures that the analysis is as rigorous as possible given the peripheral materials, segments of code, and inputs/outputs that are available. In this chapter I use this method to probe the myriad ways in which racial biases that are present both in the boardrooms and research and development wings of technology companies *and* in the broader socio-cultural milieu get hardcoded into the datasets and, subsequently, the machine learning algorithms built atop them. It will be my argument that many of these algorithms constitute a material manifestation of racial bias.

This paper will primarily be concerned with visual or optical media, and computer vision algorithms in particular. I will argue that within this context blackness falls into the notorious dialectic of hypervisibility and invisibility—blackness is too often rendered in stereotypes, at times even visually cartoonish, or it is rendered as systemically invisible. However, it is important to note that while the context of my analysis is systemic racism against predominantly African Americans, these systems have global impacts for people of color—or, to speak more precisely, those darker-skinned individuals who fall within types V or VI of the Fitzpatrick scale (cf. Fitzpatrick 1975, 1988). When I use terms like 'black,' 'person of color,' or 'dark skin,' it is meant to indicate that the problems I am identifying are of global concern and have high stakes impact on people at the darker end of the Fitzpatrick scale across the planet, even though my analysis is contextualized within the history and culture of the United States. It is outside the purview of this chapter to extend this analysis to other conjunctures, but I sincerely hope others will help me to do so by extending, revising, and challenging this work. In the first section of this chapter I will draw on critical race theory to demonstrate how this dialectic is problematic from the perspective of egalitarian democracy. In the next section I will offer a brief overview of the history of racial bias in visual media within the context of the United States that perfectly illustrates this hypervisible/invisible dialectic of blackness. In the following two sections I will look at the hypervisibility and invisibility of blackness in contemporary AI systems and will try to demonstrate the enormity of the stakes of this conjuncture. In conclusion I offer some preliminary thoughts about where we can go from here.

The Hypervisibility and Invisibility Dialectic of Racial Difference

> "Black is … 'an black aint."
> *Ellison 1989, p. 9*

In their theory of racial formations, Michael Omi and Howard Winant (2015) have argued that humans essentially use stereotypes to make sense of the world, even though these stereotypes are constantly changing. People make use of fundamental categories of difference, like race, gender, class, age, nationality, and culture to navigate society, all of which imply a certain politics of "othering" that produces structural marginalization, inequality, exploitation, and oppression. We might productively understand machine learning as engaging in a very similar behavior with similar political stakes. As I have argued elsewhere in the context of machine learned semantic labels, these algorithms engage in an iterative process of learning stereotypical differentiations to categorize the various data that they encounter (Monea 2016, 2019). However, race functions differently than these other stereotypical categorizations because, as Omi and Winant explain, race is crucially 'corporeal' and 'ocular'. What they mean by this is that racial distinctions take hold of a set of phenotypic differences—most noticeably morphological differences like skin tone, lip size, and hair texture in the case of blackness in the United States—and essentializes them, as if they were physical markers of an essential difference of kind (2015: p. 13). It is thus othering, as it establishes the border between an 'us' and a 'them,' and reifies that border by making it appear as a fundamental law of nature, a scientific fact, a marker of a different kind of being. There are two unique aspects of this process, which Omi and Winant refer to as 'racialization'. First, these phenotypic differences are arbitrarily selected, are not understood as having the same denotations and connotations across space and time, and often were previously unconnected to any racial classification. Second, they are written on to the body through morphological distinctions in such a way that racial difference is legible on sight alone.

This latter aspect of racialization has been a core component of critical race theory for decades, and was perhaps most notably articulated by Frantz Fanon (1967) in his concept of 'epidermalization'. For Fanon, epidermalization is a process by which black people realize their identification as the Other for white people as they encounter the white gaze that dissects and analyzes their body, without permission, to classify them. Fanon writes:

> I am overdetermined from the outside. I am a slave not to the "idea" others have of me, but to my appearance. […] The white gaze, the only valid one, is already dissecting me. I am *fixed*. Once their microtomes are sharpened, the Whites objectively cut sections of my reality. (1967: p. 95)

Stuart Hall succinctly defines Fanon's idea of epidermalization as "literally the inscription of race on the skin" (1996: p. 16). Hortense Spillers similarly writes about the 'hieroglyphics of the flesh', wherein black subjects are transformed into flesh through "the calculated work of iron, whips, chains, knives, the canine patrol, the bullet" (2003: p. 207). For Spillers, Western humanism is built atop these hieroglyphics, as the liberated Man requires definitionally that an other be designated as not fully human. This legacy is passed down through the generations even after black subjects were granted possession of their own bodies and continues to structure our social lives. As Alexander Weheliye describes it, racial categories "carve from the swamps of slavery and colonialism the very flesh and bones of modern Man" (2014: p. 30). Sylvia Wynter (2001) has similarly shown how this happens in her arguments about sociogeny, where a focus on phenotypical differences is just a ruse to essentialize racial difference and divide the Homo sapiens species into humans and nonhuman beings.

We can understand this corporeality, ocularity, epidermalization, and fleshiness of race as a fundamentally *visual* component, and one that makes race hypervisible by stressing phenotypes—especially morphological features like skin tone, lip size, and hair texture—and connoting racial stereotypes that help bolster racial marginalization. Lisa Nakamura (2007) has shown how this hypervisibility of race is perpetuated today by computation and digital visual culture, leading to the production of 'digital racial formations'. John Cheney-Lippold (2017) has similarly shown how algorithms have digitized race into 'measurable types', or statistical probabilities based on user data. Here I'd like to introduce the term 'users of color' to replace the term 'people of color' for this digital context. Users of color are digitally racialized based on algorithmic analysis of big data, which is reifying some old phenotypic stereotypes of racial difference at the same time that it is producing new ones. We might also follow Simone Browne (2015) and think of this process as a 'digital epidermalization'.

Now, we can easily see that even for technology companies that may have little ethical commitment to egalitarian democracy, the public relations nightmare alone of being seen as reifying racial stereotypes in digital culture is a huge deterrent. As we'll see throughout this piece, the most frequent response to criticism of algorithmic racialization is to make race invisible. Rendering blackness invisible is always the flipside of the coin, in dialectical tension with racial hypervisibility. Both options are unsatisfactory, as they both provide safe haven for racism, albeit in different ways. The most recent and visible example of rendering race invisible is the 'color blind' policies that have been pursued in the United States since the 1960s. These policies have been near universally condemned by scholars in critical race studies and related disciplines (e.g. Bonilla-Silva 2017; Brown et al. 2003; Omi/Winant 2015). Color blindness delegitimates affirmative action and similar programs striving for racial equality, allows racism to operate unchecked

provided it uses dog whistles and other careful language to obscure racial malice, and obscures important racial data trails that might otherwise have been used to uncover statistical trends of racism (e.g. in policing, court sentencing, allocation of welfare benefits, etc.). In addition, it has made speaking about race so taboo that the term 'white fragility' was coined to describe contemporary white people's inability to openly talk about race and racism (Dyson 2018). Thus, rather than deal with racialization's roots in colonialism and slavery and doing the hard work of moving towards actual egalitarian democracy, blackness is alternately rendered as hypervisible or invisible, both of which leave much to be desired. In short, as Ralph Ellison put it, "Black is … 'an black ain't" (1989, p. 9).

Accounting for the Visibility of Race in Visual Media

> "Photography is a weapon"
> —*Oliver Chanarin (PhotoQ 2015)*

The United States has a long history of embedding racial stereotypes in its visual media and communications technologies (cf. Dyer 1997, hooks 1992, Nakamura 2007). This legacy spans from analog to digital photography and, as we'll see, continues to impact a number of computer vision applications. For example, Lorna Roth (2009) has shown in detail that Kodak optimized its entire suite of products for white skin. Kodak produced a long series of "Shirley cards", named after Shirley Page, the first studio model for the photos Kodak sent out with its new products. These Shirley cards were marked "normal" and used as test cards for color balancing film stock and printers. The optimization for white skin was to the detriment of people of color, whose features increasingly disappeared in direct correlation with how dark their skin was. One legendary result was French film auteur Jean-Luc Godard's refusal to shoot on Kodak film for an assignment in Mozambique because of the racial bias hardcoded into the film—it literally would not work in Africa. It was not until complaints were made by companies trying to photograph dark objects for advertisements that Kodak developed film that could capture the details of black flesh, a project that was kept quiet at Kodak and referred to via the coded phrase "To Photograph the Details of a Dark Horse in Low Light" (Broomberg & Chanarin n.d.).

Even when film was made with the explicit intent of rendering black skin visible—rather than for rendering chocolate bars and wooden furniture visible in advertisements—deeply rooted racial problems cropped up. Take, for instance, Eric Morgan's (2006) story of the Polaroid ID-2 camera. The ID-2 was designed to take two photos per self-developing print, one portrait and one profile image, of a subject 1.2m from the lens. This was, in essence, a streamlining of Alphonse Bertillon's anthropometric identification system for state policing, which has always intersected in complex

ways with existing practices of racial and gendered marginalization (Browne 2015; Fair 2017; Lyon 2008; Wevers 2018). In the case of the ID-2, this was manifested in a special "boost" button for the flash that would make it around 40 percent brighter, the same amount of light that darker skin absorbs. In and of itself, this feature is rather innocuous. It is actually a step forward in the sense that it allowed the camera to capture the features of dark-skinned people more clearly, although still made it problematic to capture both black and white skin in the same picture. However, in 1970, a Polaroid chemist named Caroline Hunter uncovered evidence that Polaroid was making a lot of money selling ID-2 cameras to the South African government, which used them to make the passbook's that black citizens were forced to carry with them at all times within white areas (see Savage 1986).

In their exhibit titled after Kodak's coded phrase "To Photograph the Details of a Dark Horse in Low Light," South African artists Adam Broomberg and Alex Chanarin argue that the ID-2 was designed for the purpose of supporting apartheid (PhotoQ 2015; Smith 2013).[1] It is unlikely that this is the case, but it is certainly true that a certain number of Polaroid executives had a damning amount of knowledge of the trafficking of ID-2s in South Africa and worked to intervene too late and with too little energy to actually prevent the use of their new technologies for the support of Apartheid. Neither of these stories is meant to minimize the technical difficulties of capturing darker features and objects on film. These difficulties are inherent to optical media and would likely exist no matter the sociocultural context within which photographic technologies arose. What they do expose, however, is how absent and unimportant the dialectic of black invisibility and hypervisibility was to these companies, a fact that was made materially manifest in their research and development paradigms, the products they took to market, the discourse they used to position these products, and their responses to criticism from the public. It is not that these devices themselves are racist, but instead that racial biases from the context in which they are developed inflect how the companies approach research and development, imagine certain products and not others, prioritize some more highly than others, assess some potential bad applications of their technology as negative press coverage to be accounted for in the design process, etc. All of this is materially manifest in the technology itself, the way it is positioned discursively, and the myriad uses it gets put to in society.

It wasn't until the '90s that American companies really began to take seriously the need to better capture black features on film. Kodak released 'black', 'Asian', and 'Latina' Shirley cards and began marketing Kodak Gold Max film, which had an 'improved dynamic range'—it could finally 'photograph the details of a dark horse in low light'. These achievements were made by new innovations in the chemical composition of film negatives to make them more reactive with darker pigments and modi-

---

1   For more on Broomberg and Chanarin's exhibit, see O'toole 2014.

fications to color balancing techniques to optimize for different skin tones. They all required an increasing focus on the visibility of people of color that spanned from the consumer base to the research and development teams at companies like Kodak. This change in priorities also quickly manifested in digital camera and camcorder technologies. For example, in 1994 the U.S. Philips Corporation filed for a patent on a new 'tint detection circuit' that could automatically adjust the tint in a digital image to white balance for both light and dark skin tones at once. In this instance (see Figure 1 below), there are literally two components soldered onto the board, one for 'SKINR' and one for 'SKINB' that lead to an 'AND CIRCUIT' that combines their outputs to optimize for both. While this solution is not ideal, as the color balancing will be off for both light-skinned and dark-skinned people if they are in the same shot, it is a step in the right direction. It will work equally well for any skin type in isolation and when in the same frame will average towards the middle, rendering both poles of the Fitzpatrick scale equally less visible. We might also take this historical arc as emblematic of the level of intervention necessary if we are to make any given technology democratic as it pertains to racial marginalization. Just as we need to literally hardcode anti-racism measures into the circuit board to get an egalitarian camera, we will need to literally hardcode anti-racism measures into our machine learning algorithms to get egalitarian AI.
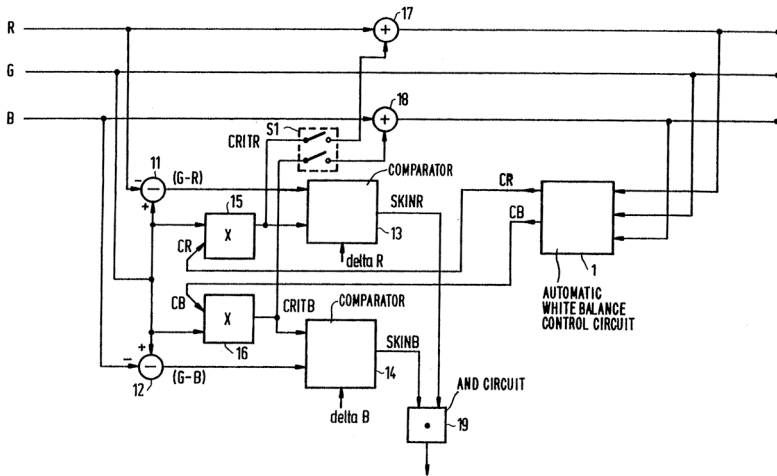

Fig. 1: U.S. Patent No. 5,428,402 (1995)

The Hypervisibility of Race in AI Systems

In a 'humorous' story read for *This American Life*, David Sedaris describes the strange cultural differences surrounding Christmas between the United States and the Netherlands, noting to refined laughter that in the Netherlands, Saint

Nicholas is accompanied by six to eight black men who prior to the mid-1950s were characterized as his personal slaves (Sedaris 2001; cf. Sedaris 2004). In Dutch tradition, these black helpers eventually stabilized into the image of Zwarte Piet, or 'Black Pete', with Dutch men and women dressing up in black face—black face paint, red lipstick, curly black-haired wigs, a golden hoop ear ring, and colorful Spanish/Moorish outfits—to lead Sinterklaas, or Saint Nicholas, in parades, distributing candy and kicks to good and bad children respectively. The majority of Dutch citizens have very positive attitudes about Zwarte Piet, and often downplay the connections between him and black face by noting alternately either that his face is colored such because he is a chimney sweep that now crawls through chimneys to deliver candies to good children who leave their shoes out or because he is a Moor that was adopted by Sinterklaas, who lives in Spain in the offseason rather than at the North Pole, as in American traditions. As Allison Blakely notes, both of these explanations are rather unconvincing (2001: pp. 47-48).

This desperate attempt to preserve a deeply problematic tradition are only complicated by the fact that Zwarte Piet is the name for the Devil in Dutch folklore, who is caught and chained for the celebration every year, and by the Netherland's complicated history with colonialism, the slave trade, and slavery—for example, the term 'apartheid' comes from the Dutch and arose particularly in the context of their colonial occupation of Surinam (Blakely 1980: p. 27). While the Dutch position themselves as a nation apart from colonial and racial marginalization, this has never been the case, especially since a number of Surinamers relocated to the Netherlands after the 1950s and have faced systemic marginalization based on race (Blakely 1980). It is no wonder that these same Surinamers are increasingly unenthused with the Zwarte Piet tradition and argue that it is insulting, and especially damaging to children of color who subsequently face bullying at school (Blakely 2001: p. 48).

David Leonard has argued that a more appropriate metric for determining culpability in instances of blackface is not whether the person dressing in blackface meant to offend people, but whether that person is causing harm, either to individuals or to society (Desmond-Harris 2014). This is a much smarter way of analyzing the situation, as it not only rids us of complex interrogations of intentionality, but also opens up the analysis of how photographic or videographic media might extend the range of impact of that harm. For instance, barring for the moment the issue of blackface in the Netherlands, it is certainly the case that it is extremely damaging in the context of the United States. As C. Vann Woodward (2001) has shown the campaign of pseudo-scientifically legitimated racism, dehumanization, segregation, disenfranchisement, and terror waged against African Americans that we know as "Jim Crow" was named after a blackface minstrel routine. Blackface thus encapsulates rather neatly the logic of American racism, in that it literally denotes and makes hypervisible phenotypes—and in still images

these are primarily morphological traits—while at the same time always preserving their connotation of inferiority in connection with white supremacy.

If we look at the case of ImageNet, we can clearly see how these Zwarte Piet images escape their context and cause social harm. ImageNet is a large dataset of labeled images first launched out of Princeton University in 2009. The dataset originally drew 80,000 labels from the semantic database WordNet—each label is referred to as a 'synset' which is a set of synonymous terms—with the goal of populating each of these labels with 500-1,000 clean and full resolution images (Deng et al. 2009). In 2010, ImageNet launched its annual ImageNet Large Scale Visual Recognition Challenge (ILSVRC), which has since served as an industry benchmark for success in computer vision applications (Russakovsky et al. 2015). This centrality was cemented by Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton's (2012) groundbreaking success in using neural networks to win the competition to produce an algorithm that could learn to classify images in the ILSVRC (cf. Sudmann 2016, 2018). What is important to know is that ImageNet not only serves as the performance benchmark for nearly all computer vision systems, but that because those systems are trained on and optimized for the ImageNet dataset, any biases in the ImageNet dataset have wide-ranging repercussions since they subsequently become hardcoded through machine learning into a large portion of computer vision systems.

Fig. 2: Zwarte Piets and Sinterklaas (Splinter 2010[2])

---

2   N.b. I could not establish the licensing for any of the Zwarte Piet images contained in ImageNet, but any search on Flickr for the term returns images like this one which are illustrative.

ImageNet gathers images for a synset that contains the terms "Black", "Black person", "blackamoor", "Negro", and "Negroid", which it defines as "a person with dark skin who comes from Africa (or whose ancestors came from Africa)". This synset is interesting for a number of reasons. First, just from browsing it, one can tell that it contains many fewer images as a percentage of the total images than other synsets that would contain useful visual details for building out a classifier. What I mean by this is that an inordinate number of photos are low resolution, don't show facial details, have black people's bodies positioned further away from the camera, and inordinately feature celebrities (about 1 per cent of the entire dataset is pictures of Barack Obama) and memes. While this certainly isn't a smoking gun for racial bias, let alone intentional racial bias, it does reveal that the capacity to accurately identify black facial features is not prioritized by default for any computer vision algorithms trained on ImageNet's data. What is perhaps closer to a smoking gun is that of the 1,286 images for this synset that are still available online, a full 79 of them are of people in blackface.[3] All but one of these images are of people dressed as Zwarte Piet. Thus, this odd Dutch phenomenon has the exact consequences that are feared in critiques of blackface: regardless of the intentions of those wearing blackface, when it enters public discourse, this signifier of blackness quickly escapes any contextualization and instead reifies racism through its connotations of white supremacy and its denotations of blackness being reducible to phenotypical difference. We can see this quite literally in the case of ImageNet, where blackface images have escaped their context to compose just over 6 per cent of the entire dataset, a dataset developed on a continent where the vast majority have never even heard of Zwarte Piet. Even if we allow the extremely dubious argument that these images are harmless in the Netherlands, a tradition beloved by fewer than 20 million has helped to bias a fundamental piece of infrastructure for computer vision.

Take, for example, Google's use of the ImageNet dataset. In 2014, Google researchers won the ILSVRC challenge with their 'Inception' algorithm (also known as 'GoogLeNet'), a 22-layered convolutional neural network (Syzegedy et al. 2015). This CNN was trained on the ImageNet dataset, and thus internalized any biases present in that data.[4] At Google I/O 2015, Google announced the launch of its new

---

3   At this time, I can only work from the publicly accessible list of links to images for given synsets. This means that I cannot access images that have since been taken down from the web, which makes it impossible to access the full set of 1,404 images that make up the dataset for that synset. I have repeatedly submitted requests to register for an account with ImageNet so that I might access the full dataset and have sent multiple emails to ImageNet's contact address, but have yet to get either access or a response.

4   While there is no comparable data for racial biases, it has been demonstrated that in the case of gender biases, neural networks not only internalize the biases in their datasets, but *amplify* them (Zhao et al. 2017).

Google Photos software. Google argued that humans were now taking over a tril-
lion images a year and at this rate would need a second lifetime to label, organize,
and revisit their photos. Google Photos was the solution we had all been waiting
for:

> Google Photos automatically organizes your memories by the people, places, and
> things that matter. You don't have to tag or label any of them, and you don't need
> to laboriously create albums. When you want to find a particular shot, with a simp-
> le search you can instantly find any photo—whether it's your dog, your daughter's
> birthday party, or your favorite beach in Santa Barbara. And all of this auto-grou-
> ping is private, for your eyes only. (Google 2015)

The new Google Photos software was primarily powered by the Inception/Goo-
gLeNet algorithm that was trained on ImageNet data, though it supplemented
the image patterns it learned from ImageNet with a huge database of photos and
nearby text from websites it had crawled and a few other indicators, like looking
at the place and time stamps of both the user during the search and the images
via their metadata (Brewster 2015). The centrality of ImageNet is no wonder, as it
not only serves as the benchmark for computer vision algorithms and is standard
across a large portion of the industry, but Google Photos was developed under
Bradley Horowitz, Google's Vice President of Streams, Photos, and Sharing, who
previously saw the value in Flickr's Creative Commons licensed images when he
helped purchase the company as an executive at Yahoo (Levy 2015).

The point to be taken from all this is that Google Photos was designed by peo-
ple who viewed ImageNet as an unquestioned industry standard and who placed
strong faith in the utility of Flickr images. They thus were ill positioned to fore-
see the racial biases inherent in the visual data that their machine learning al-
gorithms had used to develop their classifiers. This problem came to the fore just
a month after Google Photos was released when in June 2015 a black software
engineer named Jacky Alcine posted a set of images run through Google's photo
tagging software to Twitter in which images of him and his girlfriend were labeled
as photos of 'gorillas' (Alcine 2015).[5] The case clearly hit a nerve, as dozens of arti-
cles were published calling Google's algorithms racist within days and it has since
been one of the most frequently cited examples of algorithmic bias in technology
journalism. This comparison is only possible based on an arbitrary definition of
certain phenotypical differences as the sole markers of an essentialized differ-

---

5   It is worth noting that while Google is the highest profile instance of this happening, it is by no
    means the sole incident. Just a month prior to the Alcine incident Flickr made news for mislabel-
    ing a black man (and a white woman) as "animal" and "ape", and also labeled photos of Dachau
    concentration camp as "jungle gym", "sport", and "trellis" (Hern 2015).

ence. In short, when we think of skin tone, lips, and hair as the cornerstone of racial difference, this slippage between classifiers is opened up. And further, when we essentialize these differences and connect them to racist stereotypes, the connotation of this slippage becomes unbearable. It calls into question not only people's intelligence, but their very humanity. This is a problem of the highest stakes for all users of color.

As can be seen from Google's response, the company similarly understood this instance to be a serious problem and one that might threaten the future of their computer vision platforms. Within hours, a Google engineer named Yonatan Zunger was responding to Alcine's tweet asking for permission to examine his Google Photos account to figure out what had gone wrong. The next day, Zunger tweeted that Google had not recognized a face in the images of Alcine and his girlfriend at all and noted, "We're also working on longer-term fixes around both linguistics (words to be careful about in photos of people [lang-dependent]" and in "image recognition itself. (e.g., better recognition of dark-skinned faces) (Zunger 2015). Zunger promised that Google would continue to work on these issues, which included developing systems that could better process the different contrasts for different skin tones and lighting. A few days later a Google spokesperson told the BBC that, "We're appalled and genuinely sorry that this happened. We are taking immediate action to prevent this type of result from appearing" (BBC 2015). Yet, as we'll see in the next section, Google has yet to discover a solution for this problem. The datasets its algorithms are trained on make it so that race must be rendered either as a hypervisible emphasis on phenotypes, which, without a heavily curated new dataset will continue to produce slippage between users of color and apes, or as invisible.

The Invisibility of Race in AI Systems

While the speed and sincerity of Google's initial response seemed promising, after more than two years *WIRED* reported that all Google had managed to do was remove potentially offensive auto-tags for terms from its Photos software (Simonite 2018). In that same report, *WIRED* noted the results of a series of experiments they had done with Google Photos. First, they ran a collection of 40,000 images well-stocked with animals through the system and found that it did not locate any results for the terms "gorilla", "chimp", "chimpanzee", and "monkey". In a second experiment they tried uploading pictures solely of chimpanzees and gorillas and found that it still would not recognize the offending set of terms. In a third and more damning test, *WIRED* uploaded a collection of over 10,000 images used for facial-recognition research. Searching these photos for auto-tags of "black man", "black woman", or "black person" only delivered photos that were in black-and-white, which did correspond to the gender specified, but did *not* sort

people by race. In short, in response to this public relations disaster, blackness has become invisible in Google Photos.[6] This color blindness also extends to Google's Open Images dataset, which contains "30.1M image-level labels for 19.8k concepts, 15.4M bounding boxes for 600 object classes, and 375k visual relationship annotations involving 57 classes" (Kuznetsova et a. 2018). None of these millions of labels, thousands of concepts, or hundreds of classes, from what I can gather after examining the database, explicitly label race. Users of color are only identified by their absence.

The issue of black invisibility has a long history in systems that process visual data for applications like facial recognition and motion-sensing. Take, for example, the viral 2009 YouTube video of an HP laptop designed to use facial recognition to track users' faces and follow them as they move with the webcam that failed to register the movements of 'Black Desi' at all, despite easily following the motions of his white coworker (wzamen01 2012). *Consumer Reports* (2009) tried to debunk the argument that this was a racial bias by arguing that it instead is a factor of lighting conditions, and while they present their results as if the system would work the same for lighter and darker skin tones in the same lighting, in their video it is clear that this is not the case, as they have to better light their user of color's face before the system starts to track him. For another example, take Xbox Kinect, which in 2010 was reported to have trouble recognizing the faces of users of color (Ionescu 2010). This primarily effected their ability to automatically log in to their avatars, as Kinect gameplay largely functions on skeletal movement. In other words, Kinect is capable of seeing black bodies but not black faces, and can facilitate their gameplay so long as it doesn't need to recognize their face, which some games do. *Consumer Reports* (2010) similarly argued that this was merely a lighting issue and claimed to have 'debunked' the idea that Kinect is 'racist'.

It is certainly the case that these machines themselves are not intentional agents engaged in prejudicial thinking, but to merely wave away the claims of racism as 'debunked' after demonstrating that failure to function appropriately when utilized by people of color requires an uncomfortable amount of hubris and near total lack of empathy. There is something clearly going on here and it has a felt impact on racialized bodies so clear in the videos of people attempting and failing to utilize facial recognition and motion-sensing technologies. At the very least, these instances are material embodiments of some combination of the lack of forethought in the research and development phase and a lack of concern over

---

6 Interestingly, this invisibility plays out differently in Google's less publicly visible computer vision projects. Google's Cloud Vision API launched in 2016 as a new tool to make its computer vision algorithms available particularly to developers (Google 2016). Cloud Vision API still uses labels like "chimpanzee" and "gorilla", though I've found no studies to date of whether the same racial hypervisibility of Google Photos persists on Google's Cloud Vision API.

going to market with a product that would fail to operate for a large and protected minority class of citizens. And further, the common arguments that all technologies fail, that these technical constraints are unavoidable, and that it is common sense to design products based on your majority market (and in most instances this is code for 'white' people) are all inadequate at best, and deeply offensive at worst.

Take, for another example, the frequent instances of motion-activated devices like soap dispensers in public restrooms failing to recognize users of color (e.g. Fussell; Plenke 2015). Bharat Vasan, the COO of Basis Science, explained to *CNET* that there are systems that can avoid this problem by detecting the darkness of objects beneath their sensors and adjusting a spotlight to match, but these systems are too expensive for many of these lower-end motion-activated devices (Profis 2014). It is here that chronic lack of consideration for users of color becomes most apparent. These systems are designed for *public* use, and thus by default require consideration of users of color. Further, many aspects of the lighting conditions can be predicted in advance (e.g. fluorescent overhead lighting, often with the shadow of the motion-sensing object itself falling over the object to be detected). This picture only gets more complicated when we consider that the principle site of research and development for automated restroom innovation is in prisons, which are disproportionately comprised of people of color in the United States (Edwards 2015).

This is a much more significant problem than simply having automated public utilities that fail to operate for users of color. These same problems extend to medical technologies and limit the effectiveness of new wearable technologies touted as breakthrough technologies for everyday medical monitoring. For instance, Pulse Oximetry, which optically measures arterial hemoglobin oxygen saturation is demonstrably less effective in people with darker skin tones (Bickler/Feiner/Severinghaus 2005). While the FDA requires these devices to meet certain accuracy thresholds before they can go to market, they do not specify where on the Fitzpatrick scale the test subjects must fall. There is thus a financial incentive to use lighter-skinned test subjects, as utilizing and designing for darker skin types slows their path to market. New non-invasive neuroimaging techniques like functional near-infrared spectroscopy (fNIRS) are being used to study and potentially treat medical issues like Alzheimer's disease, Parkinson's disease, epilepsy, traumatic brain injury, schizophrenia, mood disorders, and anxiety disorders (Irani et al. 2007). fNIRS measures brain activity through blood oxygenation and volume in the pre-frontal cortex and is similarly less effective with darker pigmentation and darker, thicker hair (Saikia/Besio/Mankodiya 2019). These same problems of black invisibility extend to the optical heart trackers installed in many contemporary wearable technologies, like FitBit and the Apple Watch (Kim 2017; Profis 2014).

Perhaps the ultimate example of the algorithmic invisibility of users of color though can be found in lidar-based state-of-the-art object detection systems like those used in autonomous vehicles or 'self-driving cars'. Autonomous vehicles have to engage in grisly cost-benefit calculations in crash scenarios to determine how to kill or injure the fewest people (e.g. Roff 2018). New research has come out demonstrating that such systems are statistically less likely to identify darker-skinned pedestrians as humans to be factored into these calculations (Wilson/ Hoffman/Morgenstern 2019). The researchers found that this remained true even when you factor in time of day (i.e., lighting, the go-to excuse for any technological failure to recognize black faces and bodies) and visual occlusion. This could have been predicted, as the training data used for the system they analyzed contained nearly 3.5 times as many images of light-skinned people as dark-skinned people (ibid.: p. 1). Black invisibility is thus not merely a matter of identity politics, but instead can literally have life or death stakes for users of color in our increasingly AI-driven future.

Conclusion

The hypervisibility/invisibility dialectic has historically sheltered the worst forms of racism in the United States and abroad. It has been endemic to visual media since their inception and is currently being cemented into the AI paradigms that we increasingly believe our going to usher in the next stage of human civilization. The approach that AI design takes to this dialectic thus ought to be a central battleground for anyone working towards the democratization of AI. We cannot have egalitarian or democratic technology if we hardcode pre-existing regimes of marginalization into our AI systems. While it is beyond the scope of this chapter to fully articulate an alternative to the hypervisibility/invisibility dialectic in AI systems, I think we look to some of the repeated refrains that have still yet to be initiated in Silicon Valley as a way to at least open the space where we might imagine a better alternative. In her *WIRED* article "How to Keep Your AI from Turning into a Racist Monster," Megan Carcia (2017) offers some common-sense approaches that might be advocated for and instituted broadly to great effect. These systems ought to better empower users to analyze, debug, and flag problematic components of AI systems. This process not only crowdsources the labor, making it much more appealing to tech companies beholden to their shareholders, but also in the process educates users about the importance of this labor as well as the technological functioning of the systems they are helping to analyze. Second, Silicon Valley needs to hire more diverse computer programmers immediately and without making any more excuses. More diverse development teams are much more likely to recognize hardcoded marginalization prior to release and are sta-

tistically likely to generate greater profits. Third, empower more third-parties to engage in high-tech auditing of AI systems. Even if trade secrecy and fear of spam prevent the open sourcing of all code, certain trusted third parties more dedicated to advancing democracy and eradicating marginalization ought to be allowed access to these systems to help monitor hardcoded biases. Lastly, we ought to further develop a public discourse that demands egalitarian AI and institute an inter-company set of tools and standards to help better motivate companies and hold them accountable. While these are Carcia's ideas, they are also ours, writ large, as they somehow echo across internet discourse ad nauseum without ever really being instituted in Silicon Valley. We might need to supplement this with more traditional grass-roots organizing and activism tactics to turn this polite request into a demand. Democracy often requires revolutionaries.

References

Alcine, Jacky (2015); Retrieved from https://twitter.com/jackyalcine/status/6153 29515909156865 (accessed July 9, 2019).

BBC (2015): "Google Apologises for Photo App's Racist Blunder". In: BBC News. Retrieved from https://www.bbc.com/news/technology-33347866 (accessed July 9, 2019).

Bickler, Philip E./Feiner, John R./Severinghaus, John W. (2005): "Effects of Skin Pigmentation on Pulse Oximeter Accuracy at Low Saturation". In: Anesthesiology 4/102, pp. 715-719.

Blakely, Allison (2001): Blacks in the Dutch World: The Evolution of Racial Imagery in a Modern Society, Bloomington: Indiana University Press.

Blakely, Allison (1980): "Santa's Black Aide: A Glimpse of Race Relations in Holland". In: New Directions 7/2, pp. 27–29.

Bonilla-Silva, Eduardo (2017): Racism Without Racists: Color-Blind Racism and the Persistence of Racial Inequality in America (Fifth Edition), Lanham: Rowman & Littlefield.

Brewster, Signe (2015): "How Google's New Photos App Can Tell Cats From Dogs". In: Backchannel. Retrieved from https://medium.com/backchannel/how-google-s-new-photos-app-can-tell-cats-from-dogs-ffd651dfcd80 (accessed July 9, 2019).

Broomberg, Adam/Chanarin, Oliver (n.d.): "To Photograph a Dark Horse in Low Light". Retrieved from http://www.broombergchanarin.com/to-photograph-a-dark-horse-in-low-light-1-1/ (accessed July 9, 2019).

Brown, Michael K./Carnoy, Martin/Currie, Elliott/Duster, Troy/Oppenheimer, David B./Schultz, Marjorie M./Wellman, David (2003): Whitewashing Race: The Myth of a Color-Blind Society, Berkeley and Los Angeles: University of California Press.

Browne, Simone (2015): Dark Matters: On the Surveillance of Blackness, Durham: Duke University Press.

Consumer Reports (2010): "Consumer Reports Debunks the 'Racist' Kinect". In: Consumer Reports. Retrieved from https://www.consumerreports.org/cro/news/2010/11/consumer-reports-debunks-the-racist-kinect/index.htm (July 9, 2019).

Consumer Reports (2009): "Are HP Webcams Really Racist? Consumer Reports Weighs In". Retrieved from https://youtu.be/NphmOV0lrBg (July 9, 2019).

Deng, Jia/Dong, Wei/Socher, Richard/Li, Li-Jia/Li, Kai/Li, Fei-Fei (2009): "ImageNet: A Large-Scale Hierarchical Image Database". In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition 1, pp. 248-255.

Desmond-Harris, Jenée (2014): "Don't Get What's Wrong With Blackface? Here's Why It's So Offensive". In: Vox. Retrieved from https://www.vox.com/2014/10/29/7089591/why-is-blackface-offensive-halloween-costume (July 9, 2019).

Dyer, Richard (1997): White: Essays on Race and Culture, New York: Routledge.

Dyson, Michael Eric (2018): White Fragility: Why It's So Hard for White People to Talk About Racism, Boston: Beacon Press.

Edwards, Philip (2015): "The Truth About Bathroom Sensors (And Why They Fail So Often)". In: Vox. Retrieved from https://www.vox.com/2015/9/3/9255805/bathroom-motion-sensors (accessed July 9, 2019).

Ellison, Ralph (1989): Invisible Man, New York: Vintage.

Fanon, Frantz (1967): Black Skin, White Masks (Charles Lam Markmann, trans.), New York: Grove Press.

Fair, Freda L. (2017): "Surveilling Social Difference: Black Women's 'Alley Work' in Industrializing Minneapolis". In: Surveillance & Society 15/5, 655-675.

Fitzpatrick, Thomas B. (1975): "Soleil et Peau". In Journal de Médecine Esthétique 2, 33-34.

Fitzpatrick, Thomas B. (1988): "The Validity and Practicality of Sun-Reactive Skin Types I through VI". In Archives of Dermatology 124/6, 869-871.

Fussell, Sidney (2017): "Why Can't This Soap Dispenser Identify Skin?". In: Gizmodo. Retrieved from https://gizmodo.com/why-cant-this-soap-dispenser-identify-dark-skin-1797931773 (accessed July 9, 2019).

Google (2016): "Google Cloud and Autodesk Enable 10x Improvement in Media Rendering Efficiency". In: Google Cloud Platform Blog. Retrieved from https://cloudplatform.googleblog.com/2016/04/Google-Cloud-and-Autodesk-enable-10x-improvement-in-media-rendering-efficiency.html (accessed July 9, 2019).

Google (2015): "Picture This: A Fresh Approach to Photos". In: Google Blog. Retrieved from https://googleblog.blogspot.com/2015/05/picture-this-fresh-approach.html (accessed July 9, 2019).

Hall, Stuart (1996): "The After-Life of Frantz Fanon: Why Fanon? Why Now? Why Black Skin, White Masks?". In: Read, Alan (ed.), The Fact of Blackness: Frantz

Fanon and Visual Representation, London: Institute of Contemporary Arts, pp. 12-37.

Hern, Alex (2015): "Flickr Faces Complaints Over 'Offensive' Auto-Tagging for Photos". In: The Guardian. Retrieved from https://www.theguardian.com/technology/2015/may/20/flickr-complaints-offensive-auto-tagging-photos (accessed July 9, 2019).

hooks, bell (1992): Black Looks: Race and Representation, Boston: South End Press.

Ionescu, Daniel (2010): "Is Microsoft's Kinect Racist?". In: PCWorld. Retrieved from https://www.pcworld.com/article/209708/Is_Microsoft_Kinect_Racist.html (accessed July 9, 2019).

Irani, Farzin/Platek, Steven M./Bunce, Scott/Ruocco, Anthony C./Chute, Douglas (2007): "Functional Near Infrared Spectroscopy (fNIRS): An Emerging Neuroimaging Technology with Important Applications for the Study of Brain Disorders". In: The Clinical Neuropsychologist 21/1, pp. 9-37.

Kim, Meeri (2017): "Wearables Still Haven't Solved the Problems of Skin Science, But New Ideas are Coming". In: Wearable News. Retrieved from https://www.wareable.com/health-and-wellbeing/skin-science-complex-wearables-4441 (accessed July 9, 2019).

Krizhevsky, Alex/Sutskever, Ilya/Hinton, Geoffrey E. (2012): "ImageNet Classification with Deep Convolutional Neural Networks". In: Advances in Neural Information Processing Systems 25, pp. 1097-1105.

Kuznetsova, Alina/Rom, Hassan/Alldrin, Neil/Uijlings, Jasper/Krasin, Ivan/Pont-Tuset, Jordi/Kamali, Shahab/Popov, Stefan/Malloci, Matteo/Duerig, Tom/Ferrari, Vittorio (2018): "The Open Images Dataset V4: Unified Image Classificaiton, Object Detection, and Visual Relationship Detection at Scale". In: arXiv. Retrieved from https://arxiv.org/abs/1811.00982 (accessed July 9, 2019).

Levy, Steven (2015): "Bradley Horowitz Says That Google Photos is Gmail for Your Images. And That Google Plus Is Not Dead". In: Backchannel. Retrieved from https://medium.com/backchannel/bradley-horowitz-says-that-google-photos-is-gmail-for-your-images-and-that-google-plus-is-not-dead-54be1d641526 (accessed July 9, 2019).

Lyon, David (2008): "Biometrics, identification and surveillance". In: Bioethics 22/9, pp. 499-508.

Monea, Alexander (2019): "From Aristotle to Computational Topoi". In: Sundvall, Scott (ed.), Rhetorical Speculations: The Future of Rhetoric, Writing, and Technology, Logan: Utah State University Press, pp. 203-225.

Monea, Alexander (2016): The Graphing of Difference: Numerical Mediation and the Case of Google's Knowledge Graph". In: Cultural Studies ↔ Critical Methodologies 16/5, pp. 452-461.

Morgan, Eric (2006): "The World is Watching: Polaroid and South Africa". In: Enterprise & Society 7/3, pp. 520-549.

Nakamura, Lisa (2007): Digitizing Race: Visual Cultures of the Internet, Minneapolis: University of Minnesota Press.

Omi, Michael/Winant, Howard (2015): Racial Formation in the United States (Third Edition), New York: Routledge.

O'toole, Sean (2014): "Making, Refusing, Remaking: Adam Broomberg and Oliver Chanarin's Recent Photography." In: Safundi 15/ 2-3, pp. 1-14.

PhotoQ (2015): "Broomberg & Chanarin: Low Light". Retrieved from https://vimeo.com/123396189 (accessed July 9, 2019).

Plenke, Max (2015): "The Reason This 'Racist Soap Dispenser' Doesn't Work on Black Skin". In: Mic. Retrieved from https://mic.com/articles/124899/the-reason-this-racist-soap-dispenser-doesn-t-work-on-black-skin#.1hOLrb9JR (accessed July 9, 2019).

Profis, Sharon (2014): "Do Wristband Heart Trackers Actually Work? A Checkup". In: CNET. Retrieved from https://www.cnet.com/news/how-accurate-are-wristband-heart-rate-monitors/ (accessed July 9, 2019).

Roth, Lorna (2009): "Looking at Shirley, the Ultimate Norm: Colour Balance, Image Technologies, and Cognitive Equity." In: Canadian Journal of Communication 34/1, pp. 111-136.

Russakovsky, Olga/Deng, Jia/Su, Hao/Karuse, Jonathan/Satheesh, Sanjeev/Ma, Sean/ Huang, Zhigeng/Karpathy, Andrej/Khosla, Aditya/Bernstein, Michael/ Berg, Alexander C./Fei-Fei, Li (2015): "ImageNet Large Scale Visual Recognition Challenge". In: International Journal of Computer Vision 115/3, pp. 211-252.

Savage, Michael (1986): "The Imposition of Pass Laws on the African Population in South Africa, 1916-1984". In: African Affairs 85/339, pp. 181-205.

Sedaris, David (2001): "Don't They Know It's Christmas After All". In: This American Life 201. Retrieved from https://www.thisamericanlife.org/201/them (accessed July 9, 2019).

Sedaris, David (2004): "Six to Eight Black Men". In: Dress Your Family in Corduroy and Denim, New York: Little Brown and Co., pp. 157-165.

Simonite, Tom (2018): "When it Comes to Gorillas, Google Photos Remains Blind". In: WIRED. Retrieved from https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/ (accessed July 9, 2019).

Smith, David (2013): "'Racism' of Early Colour Photography Explored in Art Exhibition." In: The Guardian. Retrieved from https://www.theguardian.com/artanddesign/2013/jan/25/racism-colour-photography-exhibition (accessed July 9, 2019).

Spillers, Hortense (2003): "'Mama's Baby, Papa's Maybe': An American Grammar Book". In: Black, White, and in Color: Essays on American Literature and Culture, Chicago: University of Chicago Press, pp. 203-229.

Splinter, Hans (2010): "Sinterklaas 2010". In: Flickr. Retrieved from https://www.flickr.com/photos/archeon/5214550043/ (accessed July 9, 2019).

Szegedy, Christian/Liu, Wei/Jia, Yangqing/Sermanet, Pierre/Reed, Scott/Anguelov, Dragomir/Erhan, Dumitru/Vanhoucke, Vincent/Rabinovich, Andrew (2015): "Going Deeper with Convolutions." In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-9.

U.S. Patent No. 5,428,402 (1995) Retrieved from https://patents.google.com/patent/US5428402A/en (accessed July 9, 2019).

Weheliye, Alexander (2014): Habeas Viscus: Racializing Assemblages, Biopolotics, and Black Feminist Theories of the Human, Durham: Duke University Press.

Wevers, Rosa (2018): "Unmasking Biometrics' Biases: Facing Gender, Race, Class and Ability in Biometric Data". In: Journal for Media History 21/2, pp. 89-105.

Wilson, Benjamin/Hoffman, Judy/Morgenstern, Jamie (2019): "Predictive Inequity in Object Detection". In: arXiv. Retrieved from https://arxiv.org/abs/1902.11097 (accessed July 9, 2019).

Woodard, C. Vann (2001): The Strange Career of Jim Crow, New York: Oxford University Press.

Wynter, Sylvia (2001): "Towards the Sociogenic Principle: Fanon, Identity, the Puzzle of Conscious Experience, and What It Is Like to Be 'Black'". In: Durán-Cogan, Mercedes F./Gómez-Moriana, Antonio (eds.), National Identities and Sociopolitical Changes in Latin America, New York: Routledge, pp. 30-66.

Wzamen (2012): "HP computers are racist". Retrieved from https://youtube/t4DT3tQqgRM (accessed July 9, 2019).

Zhao, Jieyu/Wang, Tianlu/Yatskar, Mark/Ordóñez, Vincente/Chang, Kai-Wei (2017): "Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-Level Constraints". In: arXiv. Retrieved from https://arxiv.org/abs/1707.09457 (accessed July 9, 2019).

Zunger, Yonatan (2015); Retrieved from https://twitter.com/yonatanzunger/status/615585375487045632 (accessed July 9, 2019).