



# Deconstructing Whiteness: Visualizing Racial Bias in a Face Recognition Algorithm

Avital Meshi

Performance Studies Graduate Group University of California, Davis, USA

## ABSTRACT

*Deconstructing Whiteness* is an art performance which examines the visibility of race through the lens of AI face recognition technology. The artist interacts with an off-the-shelf face recognition algorithm, while specifically examining the confidence level by which it recognizes her as 'White'. During the performance she changes her appearance by modifying her facial expressions and hair style and by doing so is able to 'trick' the algorithm and affect its confidence regarding her racial visibility.

## CCS CONCEPTS

• **Applied computing** → Arts and humanities; Media arts.

## KEYWORDS

Creative AI, face recognition, machine learning, performance

## ACM Reference Format:

Avital Meshi. 2021. Deconstructing Whiteness: Visualizing Racial Bias in a Face Recognition Algorithm. In *10th International Conference on Digital and Interactive Arts (ARTECH 2021)*, October 13–15, 2021, Aveiro, Portugal, Portugal. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3483529.3483738>

## 1 INTRODUCTION

Face recognition algorithms are becoming increasingly prevalent in our environment. They are embedded in products and services we use on a daily basis [7]. Mostly we are unaware of when and where our faces are detected and classified by them, and we hardly ever know how we are seen through their lens [11]. Recent findings show that AI algorithms have the capacity to harshly impact people's lives due to social disparities and biases which are embedded into the models. This is especially true for people who come from minority and underrepresented groups [3], [1]. MIT scholars Joy Buolamwini and Timnit Gebru specifically examined commercially used face recognition products and documented existing biases against female subjects with darker skin color [5]. Such findings come as no surprise as the conceptual logic behind similar methods was identified before by theorists and critical race scholars as fundamentally discriminatory and racist [10]. While a handful of cities across the US have already banned government use of facial recognition technology due to concerns over its accuracy [9], it is still being used by both governments and private businesses all

over the world. Hence, it is essential for us as individuals to become more familiar with how these algorithms see our faces and to what extent we can control this visibility.

In *Deconstructing Whiteness*, the artist interacts with a face recognition algorithm while specifically examining her racial visibility. By changing her external appearance she manages to 'trick' the algorithm and affect the level of confidence by which it classifies her as 'White'. A video documentation of the performance is further analyzed and the images of all the classifications of the artist's face are displayed to reveal how she is being 'seen' by the machine (Figure 1).

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

## 2 IMPLEMENTATION DETAILS

*Deconstructing Whiteness* is a ten minute performance during which a face recognition algorithm detects and classifies the artist's face. The algorithm used in this performance is an off-the-shelf, pre-trained computer vision model, developed and made freely available [2]. In order to specifically examine racial visibility, we decided to focus only on the label 'White' (The model also predicts other identity related labels such as gender, facial expressions, make-up and accessories and more). The label 'White' was chosen over other racial labels because it was consistently showing higher confidence levels when detecting the artist's face.

During the performance, the artist changes her facial expressions and hair style in an attempt to modify the confidence levels by which the algorithm classifies her as 'White'. A collage showing all the images of all the classifications reveals that when the artist uses her hands to make her eyes seem narrower, confidence levels regarding the label 'White' tend to drop. This is also true for when the artist's hair style seems more frizzy. Minimum confidence level which is achieved during the performance stands on 12.09 percent. In contrast, when the artist covers her eyebrows with her fingers along with keeping her hair straighter, the algorithm classifies her as 'White' with higher confidence levels (Figure 2). Maximum confidence level which is achieved during the performance stands on 94.13 percent. Maximum and minimum confidence levels clearly demonstrate the different appearances which modulate the way the algorithm classifies the artist's face as 'White' (Figure 3).

## 3 CONCEPTUAL OVERVIEW

*Deconstructing Whiteness* draws inspiration from the theory of the New Aesthetic. The artist 'waves at the machine' and waits for it to 'wave back' at her [4]. In the Manifesto for a Theory of the



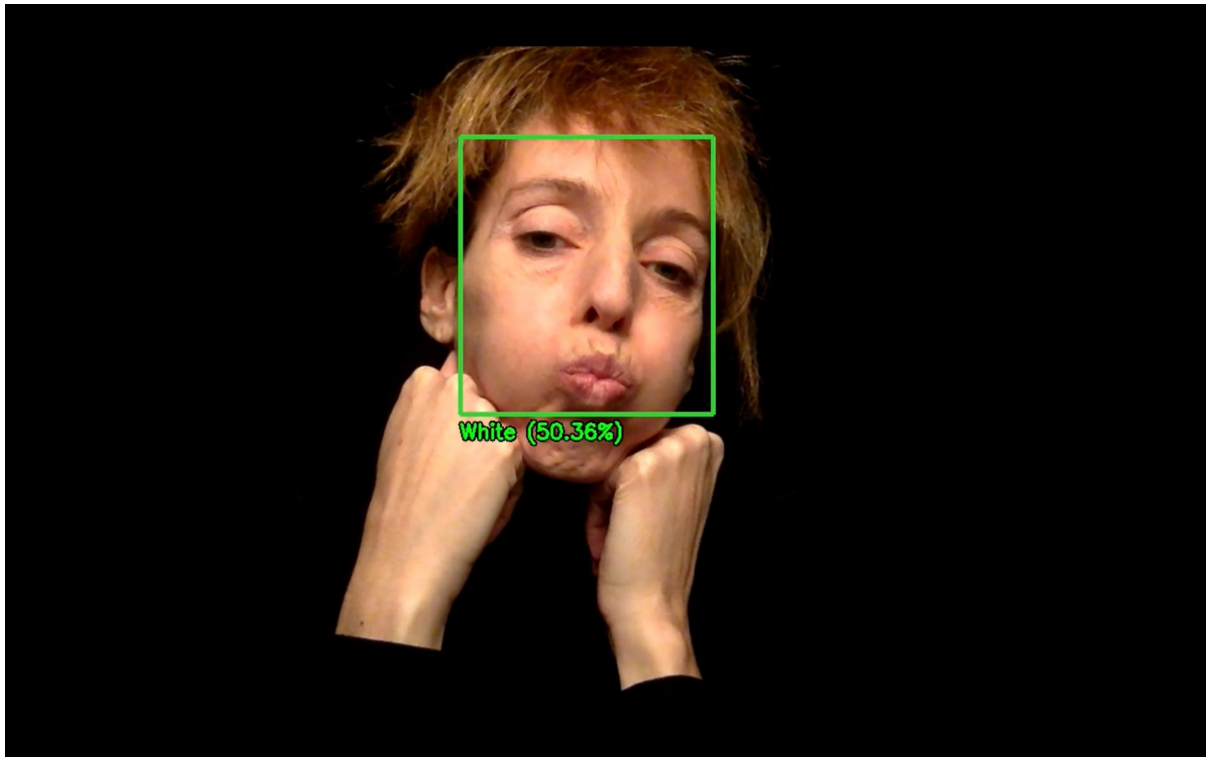
This work is licensed under a Creative Commons Attribution International 4.0 License.

ARTECH 2021, October 13–15, 2021, Aveiro, Portugal, Portugal

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8420-9/21/10.

<https://doi.org/10.1145/3483529.3483738>



**Figure 1:** A screen capture from the performance’s documentation. During the performance the artist changes her facial expressions and her hair style while examining the confidence level by which the algorithm classifies her as ‘White’

‘New Aesthetic’, Curt Cloninger advises us to resist letting such systems dazzle our minds by showing us a real-time documentary of ourselves. Instead, Cloninger claims, “we should spend some time figuring out how these systems flow and function so we can more effectively modulate them (or sabotage them)” [6]. The ‘dialogue’ that is formed between the human and the machine in *Deconstructing Whiteness* utilizes a performative behavior as a tool. Within this entanglement agency emerges and the artist learns how the algorithm ‘sees’ her and how she can control this visibility.

The performance also follows conceptions of cyber-feminism which are demanding liberation from a techno- chauvinist culture. Donna Haraway claims in her *Cyborg Manifesto* that in a cyborg world, people are not afraid of their kinship with machines. This reality of the cyborg exhibits partial identities and contradictory standpoints [8]. *Deconstructing Whiteness* illustrates such standpoint by showing how the algorithm literally lets go of the White/Non-White binary and replaces it with a flow of probabilities. By doing so, it turns racial visibility into a problem in coding. Therefore, we ask, what is the place of race in networks which yield such cyborg identities? What is the meaning of race in a system which measures geometries of difference and contradiction in order to postulate one’s identity?

#### 4 RELATED ARTWORKS

The examination and modification of one’s visibility through the lens of facial recognition algorithms is seen in Trevor Paglen’s

project *Machine Readable Hito* (2017), a series of portraits of the artist Hito Steyerl. Steyerl intentional performance in which she changes her facial expressions, modifies the outcomes of her classification with regard to age, emotional expression and gender. In the project *Landmarks* (2021) which examines the visibility of transgender individuals through facial recognition algorithms, the artist Chelsea Thompso asks: “How can I convince this machine of my gender?”. Other artists deliberately engage with face recognition algorithms with the intention to avoid visibility all together. In *CV Dazzle* (2010-2017), Artist Adam Harvey uses unusual make-up designs which prevents the algorithm from detecting a human face. Similarly, artist Zach Blas, in *Facial Weaponization suite* (2011-2014) creates abstract face masks which completely cover the face in order to render it undetectable. In contrast, in *URME Surveillance* (2013) artist Leo Selvaggio offers others a 3D printed mask of his own face as a mode of obfuscation for when they are confronted with face recognition algorithms.

#### 5 CONCLUSION

The potential of face recognition algorithms to perpetuate racism opens up an important discussion. The simple effort taken in *Deconstructing Whiteness* to become familiar with the way a face recognition algorithm classifies a specific face, is enough to create some very disturbing images. These images reveal how the algorithm detects race not just in relation to skin color, but also with



**Figure 2:** All the classifications which were processed during the performance were sorted from lowest to highest confidence level regarding the label 'White'. Confidence levels rise from left to right for each row. The image on the upper left corner side of the collage shows the lowest confidence level. The image on the lower right corner shows the highest confidence level

regard to other facial features such as the shape of one's eyes, the color of one's eyebrows and the straightness of one's hair.

The physiognomic analysis by which one's facial features or expressions are used to indicate a person's character or ethnic origin is a debunked practice which was used in the past to legitimize slavery and to perpetuate Nazi race 'Science'. Face recognition algorithms brush the dust off this pseudo-scientific methods and with that pose essential questions regarding ethical and fair use of this technology. In *Deconstructing Whiteness* the artist finds ways by which she can 'trick' the algorithm but this does not mean to suggest that everyone is able to do the same. It is important to ask who is able to 'trick' the machine and who is unable to do that? Who can control their own visibility and who cannot? and to what extent? Moreover, what does it even mean to control one's visibility? Why do we even want to do that? Are we approaching a 'Minority Report' future in which major changes to our external appearance will be our only escape from systems of surveillance?

*Deconstructing Whiteness* demonstrates how by interacting with a face recognition algorithm we can learn how we are seen through it and whether or not we can control this visibility. Unfortunately, our current relationship with face recognition algorithms is an asymmetrical one; We do not know where, when and why such algorithms are being used and how exactly we are seen through them. This relationship puts us in a passive position in which we are subjected to these systems without really knowing how accurate and fair they are. As long as these systems are out there we must

find a way to develop a more active relationship with them. We have to demand to be more familiar with how they see us and others, and decide whether or not this is how we wish to be seen. It is crucial for us to regain our agency when facing face recognition technology.

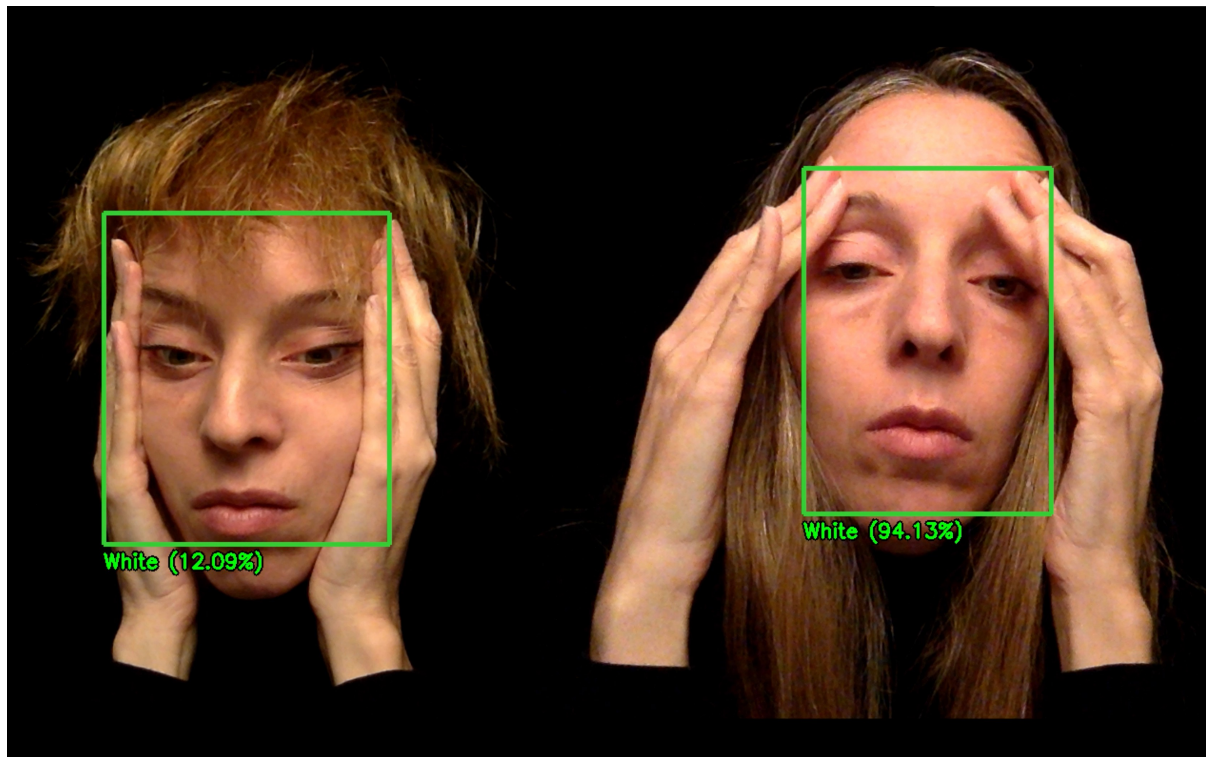
Lastly, assuming a future of constant surveillance and ubiquitous interpretation by machine learning systems, *Deconstructing Whiteness* offers performance as one of the tools with which we may be able to regain such agency. By spending time with the algorithm and becoming familiar with how it 'sees' us we also become engaged in a performative behavior. Such performance may lead to changes in the outcomes of our classifications. In this sense, one might consider face recognition technology as a platform for radical identity transformation with which we can shape and accomplish a performance that will allow us to be seen as we wish to be seen.

## ACKNOWLEDGMENTS

The author wishes to thank Ofer Meshi for his invaluable help in integrating the AI system used in this artwork.

## REFERENCES

- [1] 2018. *Aclu: Amazon's face-recognition software matched members of congress with mugshots*. Retrieved Sep 4, 2021 from <https://www.orlandosentinel.com/politics/os-amazon-rekognition-face-matching-software-congress-20180726-story.html>
- [2] 2018. *face-classification*. Retrieved Sep 4, 2021 from <https://github.com/wondonghyeon/face-classification>



**Figure 3: Maximum and minimum predicted confidence levels of the label 'White'. On the left the artist uses her hands to narrow her eyes while having a frizzy hair. This appearance made the confidence level drop. On the right, the artist covers her eyebrows with her fingers while keeping her hair style straight. This appearance made the confidence level rise**

- [3] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine bias. *ProPublica*, May 23, 2016 (2016), 139–159.
- [4] James Bridle. 2011. Waving at the Machines. *Web Directions* 5 (2011).
- [5] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*. PMLR, 77–91.
- [6] Curt Cloninger. 2012. Manifesto for a Theory of the 'New Aesthetic'. *Mute* 3, 4 (2012).
- [7] Sumit Das, Aritra Dey, Akash Pal, and Nabamita Roy. 2015. Applications of artificial intelligence in machine learning: review and prospect. *International Journal of Computer Applications* 115, 9 (2015).
- [8] Donna Haraway. 2006. A cyborg manifesto: Science, technology, and socialist-feminism in the late 20th century. In *The international handbook of virtual learning environments*. Springer, 117–158.
- [9] R Metz. 2020. Portland passes broadest facial recognition ban in the US. *CNN*.
- [10] Luke Stark. 2019. Facial recognition is the plutonium of AI. *XRDS: Crossroads, The ACM Magazine for Students* 25, 3 (2019), 50–55.
- [11] Meredith Whittaker, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Myers West, Rashida Richardson, Jason Schultz, and Oscar Schwartz. 2018. *AI now report 2018*. AI Now Institute at New York University New York.