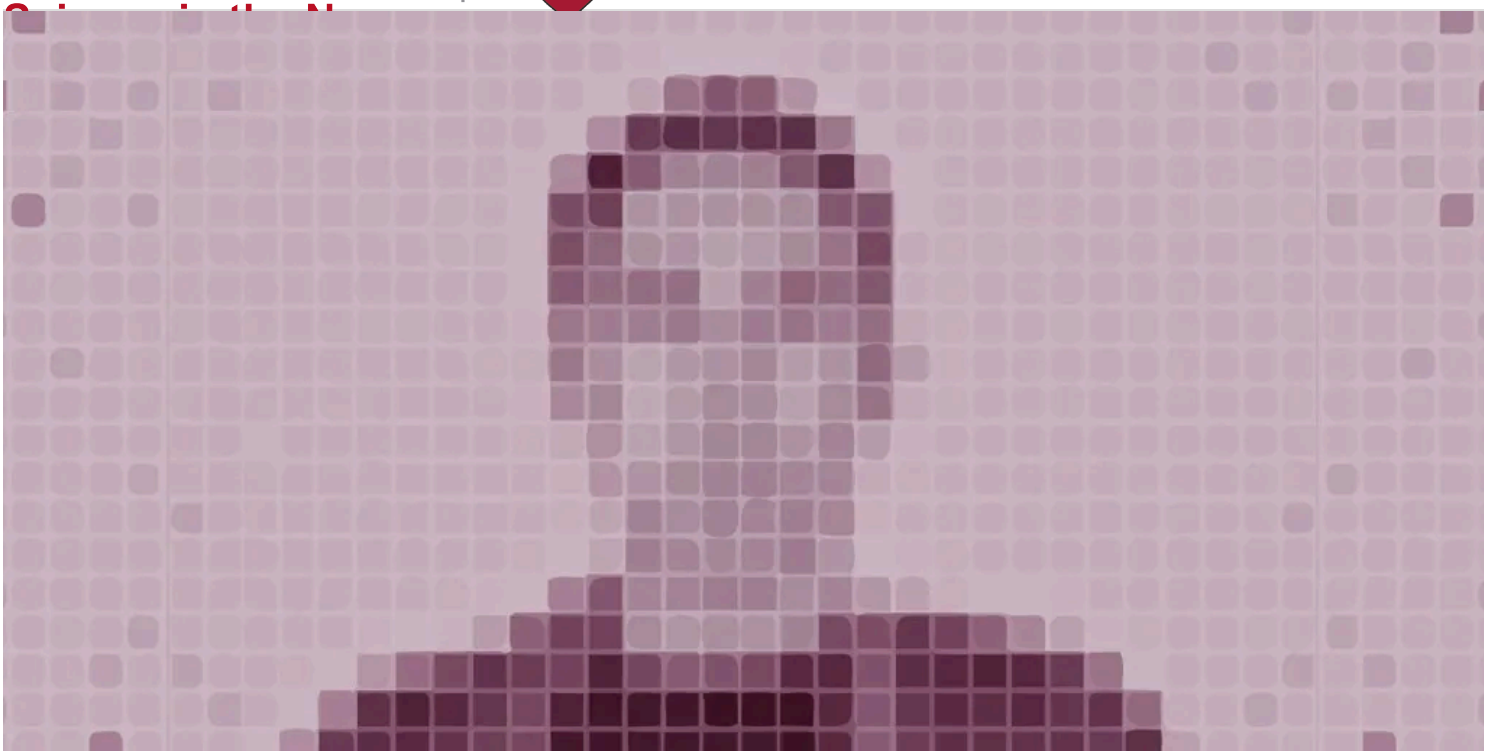

OCTOBER 24, 2020

BLOG, SCIENCE POLICY, SPECIAL EDITION: SCIENCE POLICY AND SOCIAL JUSTICE

Racial Discrimination in Face Recognition Technology

By Alex Najibi

We unlock our iPhones with a glance and wonder how Facebook knew to tag us in that photo. But face recognition, the technology behind these features, is more than just a gimmick. It is employed for **law enforcement surveillance**, **airport passenger screening**, and **employment and housing decisions**. Despite widespread adoption, face recognition was recently **banned** for use by police and local agencies in several cities, including Boston and San Francisco. Why? Of the

dominant **biometrics** in use (fingerprint, iris, palm, voice, and face), face recognition is the least accurate and is rife with privacy concerns.

Police use face recognition to compare suspects' photos to mugshots and driver's license images; it is estimated that almost **half** of American adults – over 117 million people, as of 2016 – have photos within a facial recognition network used by law enforcement. This participation occurs without consent, or even awareness, and is bolstered by a lack of legislative oversight. More disturbingly, however, the current implementation of these technologies involves significant racial bias, particularly against Black Americans. Even if accurate, face recognition empowers a law enforcement system with a long history of **racist and anti-activist surveillance** and can widen **pre-existing inequalities**.

Inequity in face recognition algorithms

Face recognition algorithms boast high classification accuracy **(over 90%)**, but these outcomes are not universal. A growing body of **research** exposes **divergent error rates** across demographic groups, with the poorest accuracy **consistently found** in subjects who are **female, Black, and 18-30 years old**. In the landmark 2018 **"Gender Shades" project**, an intersectional approach was applied to appraise three gender classification algorithms, including those developed by IBM and Microsoft. Subjects were grouped into four categories: darker-skinned females, darker-skinned males, lighter-skinned females, and lighter-skinned males. All three algorithms performed the worst on darker-skinned females, with error rates up to 34% higher than for lighter-skinned males **(Figure 1)**. **Independent assessment** by the National Institute of Standards and Technology (NIST) has confirmed these studies, finding that face recognition technologies across 189 algorithms are least accurate on women of color.

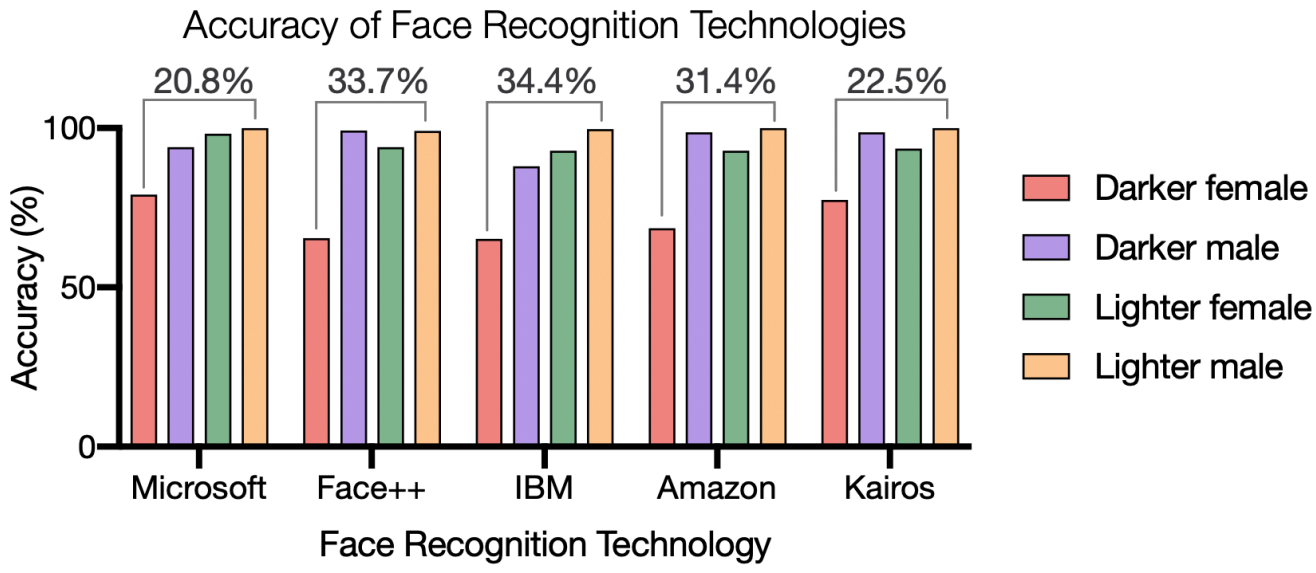

Figure 1: Auditing five face recognition technologies. *The **Gender Shades project** revealed **discrepancies** in the classification accuracy of face recognition technologies for different skin tones and sexes. These algorithms consistently demonstrated the poorest accuracy for darker-skinned females and the highest for lighter-skinned males.*

These compelling results have prompted immediate responses, shaping an ongoing discourse around equity in face recognition. **IBM** and **Microsoft** announced steps to reduce bias by modifying testing cohorts and improving data collection on specific demographics. A **Gender Shades re-audit** confirmed a decrease in error rates on Black females and investigated more algorithms including Amazon's Rekognition, which also showed racial bias against darker-skinned women (31% error in gender classification). This result corroborated an earlier assessment of Rekognition's face-matching capability by the American Civil Liberties Union (ACLU), in which 28 members of Congress, disproportionately people of color, were **incorrectly matched** with mugshot images. However, Amazon's **responses** were **defensive**, alleging issues with auditors' methodology rather than addressing racial bias. As Amazon has marketed its technology to law enforcement, these discrepancies are concerning. Companies that provide these services have a responsibility to ensure that they are equitable – both in their technologies and in their applications.

Face recognition in racial discrimination by law enforcement

Another key source of racial discrimination in face recognition lies in its utilization. In 18th century New York, **"lantern laws"** required enslaved people to carry lanterns after dark to be publicly visible. Advocates fear that even if face recognition algorithms are made equitable, the technologies could be applied with the same spirit, disproportionately harming the Black community in line with existing racist patterns of law enforcement. Additionally, face recognition can potentially target other marginalized populations, such as **undocumented immigrants** by ICE, or **Muslim citizens** by the NYPD.

Discriminatory law enforcement practices were highlighted following the murder of George Floyd by the Minneapolis PD. Black Americans are **more likely to be arrested** and incarcerated for **minor crimes** than White Americans. Consequently, Black people are overrepresented in mugshot data, which face recognition uses to make predictions. The Black presence in such systems creates a feed-forward loop whereby racist policing strategies lead to disproportionate arrests of Black people, who are then subject to future surveillance. For example, the NYPD maintains a database of 42,000 "gang affiliates" – 99% Black and Latinx – with **no requirements** to prove suspected gang affiliation. In fact, certain police departments use gang member

identification as a productivity measure, **incentivizing false reports**. For participants, inclusion in these monitoring databases can lead to **harsher sentencing and higher bails**– or denial of bail altogether.

But how specifically do unjust applications of face recognition and surveillance harm Black Americans? As stated by the **Algorithmic Justice League**, "face surveillance threatens rights including privacy, freedom of expression, freedom of association and due process." Surveillance is linked to **behavioral changes** including self-censorship and avoiding activism for fear of retribution; for example, face recognition was employed to **monitor** and identify Black Lives Matter protestors. The FBI has a **long history** of **surveilling prominent Black activists** and leaders to track and suppress their efforts. Additionally, continual surveillance induces fear and **psychological harm**, rendering subjects vulnerable to **targeted abuses**, as well as physical harm, by expanding systems of government oversight used to deny access to **healthcare and welfare**. In a criminal justice setting, face recognition technologies that are inherently biased in their accuracy can misidentify suspects, **incarcerating innocent** Black Americans.

In a striking example, the model surveillance program **Project Green Light (PGL)** was enacted in 2016, installing high-definition cameras throughout the city of Detroit. The data, which streams directly to Detroit PD, can be tested for face recognition against criminal databases, driver's licenses, and state ID photos; **almost every Michigan resident** is in this system. But PGL stations are not distributed equally: surveillance correlates with majority-Black areas, avoiding White and Asian enclaves **(Figure 2)**. In interviewing residents, a **critical analysis** of PGL reported in 2019 that "surveillance and data collection was deeply connected to diversion of public benefits, insecure housing, loss of employment opportunities, and the policing and subsequent criminalization of the community members that come into contact with these surveillance systems." PGL illustrates how systems of face monitoring can perpetuate racial inequality if their application is not regulated.

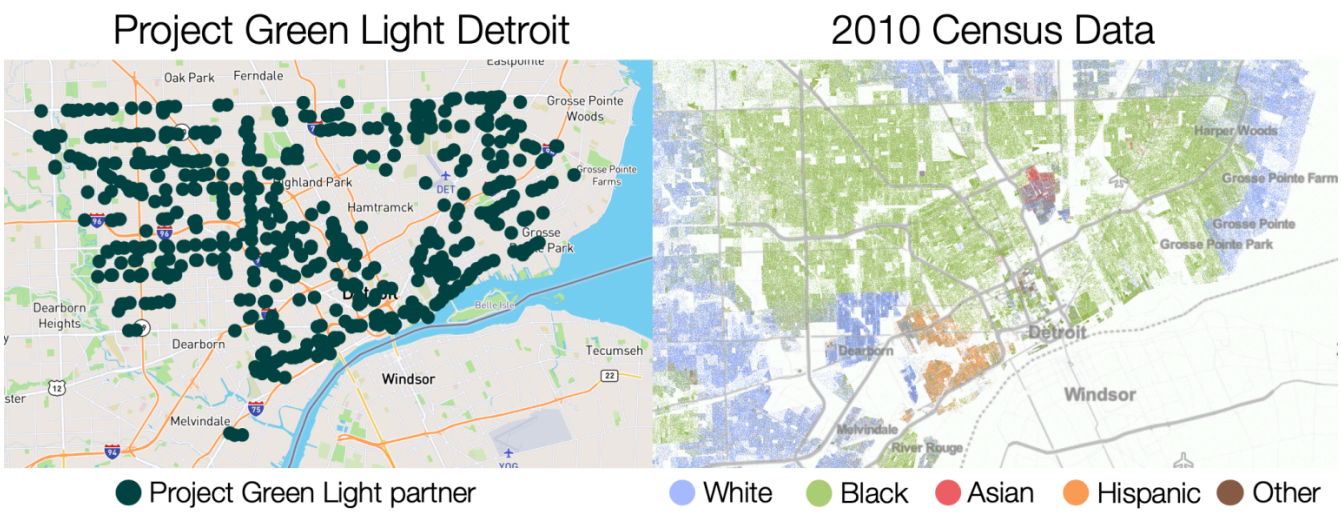

Figure 2: Racial bias in the application of face recognition technology. *Locations of* ***Project Green Light Detroit partners*** *(left) overlap with primarily Black communities in data from the* ***U.S. census*** *(right). In this city-wide program, the brunt of the surveillance falls on Detroit's Black residents.*

Building a more equitable face recognition landscape

Several avenues are being pursued to address these inequities. Some target technical algorithmic performance. First, algorithms can train on diverse and representative datasets, as standard training databases are **predominantly White and male**. Inclusion within these datasets should require consent by each individual. Second, the data sources (photos) can be made more equitable. Default camera settings are often **not optimized to capture darker skin tones**, resulting in **lower-quality** database images of Black Americans. Establishing standards of image quality to run face recognition, and settings for photographing Black subjects, can reduce this effect. Third, to assess performance, regular and **ethical auditing**, especially considering intersecting identities (i.e. young, darker-skinned, and female, for example), by NIST or other independent sources can hold face recognition companies accountable for remaining methodological biases.

Other approaches target the application setting. Legislation can monitor the use of face recognition technology, as even if face recognition algorithms are made perfectly accurate, their contributions to mass surveillance and selective deployment against racial minorities must be curtailed. **Multiple advocacy groups** have engaged with lawmakers, educating on **racial literacy** in face recognition and demanding accountability and transparency from producers. For example, the **Safe Face Pledge** calls on organizations to address bias in their technologies and evaluate their application. Such efforts have already achieved some progress. The 2019 **Algorithmic Accountability Act** empowered the Federal Trade Commission to regulate companies, enacting obligations to assess algorithmic training, accuracy, and data privacy. Furthermore, several **Congressional hearings** have specifically considered anti-Black discrimination in face recognition. The powerful protests following the murder of George Floyd also drove significant change. Congressional Democrats introduced a **police reform bill** containing stipulations to restrain the use of face recognition technologies. More astonishing was the **tech response**: IBM discontinued its system, Amazon announced a one-year freeze on police use of Rekognition, and Microsoft halted sales of its face recognition technology to the police until federal regulations are instituted. These advances have supported calls for more progressive legislation, such as the movements to **reform or abolish policing**. For now, the

movement for equitable face recognition is intertwined with the movement for an equitable criminal justice system.

Face recognition remains a powerful technology with significant implications in both criminal justice and everyday life. Less contentious applications of face recognition exist, for example, **assistive technology** supporting people with visual impairments. While we focus specifically on face recognition in this article, the discussed problems and solutions are part of **broader efforts** to identify and eliminate inequalities in the fields of artificial intelligence and machine learning. So the next time we unlock our phone, let's remember that addressing racial bias within face recognition and its applications is necessary to make these algorithms equitable and even more impactful.

Alex Najibi is a 5th-year Ph.D. candidate studying bioengineering at Harvard University's School of Engineering and Applied Sciences.

For More Information:

- Learn more about facial recognition in the criminal justice system **here**
- Check out **this article** discussing how we can hold AI accountable
- Read **Georgetown Law's report** on the unregulated, discriminatory use of face recognition by law enforcement in the U.S.
- Take a look at the **Gender Shades project's** research on disparities in face recognition accuracy

This article is part of our **special edition on science policy and social justice**.

---

50 thoughts on "Racial Discrimination in Face Recognition Technology"

boby

APRIL 16, 2024 AT 2:11 PM

where !?!?!?!?!?!!?!?