Christopher Underhill

Instructor: David Eve

Course: CSCI-101-01

Date: November 1, 2025

The Color of Recognition: Racial Bias in Facial Recognition Technology

Introduction

Facial recognition technology (FRT) is showing up everywhere: airports, schools, stores, even police departments. It's marketed as something that makes life easier and safer. But beneath all that sleek design and technical confidence is a serious issue: it doesn't recognize everyone equally. Over the past few years, study after study has found that these systems make far more mistakes when identifying people of color, especially Black women. That's not just a technical glitch; it's a social and ethical problem that questions the fairness of the technology we're trusting in so many high-stakes places.

This paper looks at a simple question: why does this bias exist, and how far does it actually reach? The evidence suggests three main reasons: unbalanced training data, technical and environmental issues, and the careless use of these systems by institutions. All three combine to create a cycle that reflects the same inequalities we already see in society. The problem isn't just in the code; it's in the systems and choices behind it. Fixing that means improving the technology, yes, but it also means holding people and organizations accountable for how it's used.

Dataset Imbalance and Algorithmic Bias

At the heart of FRT bias is a simple truth: algorithms learn from the data they're fed. If most of that data comes from light-skinned faces, the algorithm gets better at recognizing light-skinned people and worse at everyone else. In their influential study *Gender Shades* (2018), Joy Buolamwini and Timnit Gebru tested several major gender classification systems and discovered something alarming. The systems misidentified darker-skinned women up to 34.7% of the time, while their error rate for lighter-skinned men was less than 1%. That kind of gap doesn't just happen by chance. It happens because darker-skinned female faces are underrepresented in the data these systems are trained on.

This isn't just a small technical mistake; it shows how unevenly different groups are valued when it comes to collecting data. A lot of commercial image databases are built from pictures found online, and most of those tend to feature Western, male faces. Because of that, the AI ends up learning a pretty narrow version of what a "normal" face looks like. Buolamwini and Gebru's research highlights how race and gender biases layer on top of each other, which makes things especially unfair for women of color. When those kinds of systems are actually used in real life, like by police or in hiring decisions, the errors stop being just numbers on a chart. They turn into real harm: wrongful arrests, lost opportunities, and a growing lack of trust.

Backing up these findings, the U.S. National Institute of Standards and Technology (NIST) ran the largest facial recognition audit to date. Grother, Ngan, and Hanaoka (2019) tested 189 algorithms using millions of images and found that false positives were 10 to 100 times higher for Asian and African American faces compared to white ones. Even more striking, these problems showed up across both commercial and academic systems. That means it's not just one bad company; it's a widespread problem built into how these systems are designed and trained.

Put simply, an algorithm can only be as fair as its data. When entire groups of people are underrepresented or misrepresented in training datasets, those people become invisible to the system. Bias in, bias out.

Technical and Environmental Factors

People usually talk about biased data when it comes to facial recognition, but lighting and image quality matter too. A study in 2022 by Wu and others found that bad lighting makes it harder for these systems to identify people with darker skin. When the lighting is poor, the camera doesn't pick up as much detail, and the software starts to mess up more often. In essence, the darker the photo, the higher the chance the system gets it wrong.

This finding reveals something important: bias isn't just baked into the code. It's also in the environment where data is collected. Most training photos come from ideal conditions; well-lit rooms, professional cameras, and clear angles. But real life isn't like that. Security footage is grainy, lighting varies, and faces aren't always centered or clear. Combine that with the fact that darker skin reflects less light, and the system struggles even more.

Many recognition systems also rely on contrast and texture to identify features. Under uneven lighting, those cues become harder to detect on darker skin tones. Wu and colleagues point out that this isn't a flaw in human appearance; it's a flaw in design. Developers often don't test their systems in varied conditions, so they end up creating tools that work best for people who look like the ones who built them.

Researchers have tried to solve this by training systems with better lighting conditions and more variety in their data. This approach, sometimes called "illumination-aware" training, helps the software handle different environments more accurately. Still, these fixes only go so far. It raises a bigger issue: why do problems like this keep showing up in the first place? Most of the time, it comes down to whose perspectives are centered when technology is built, and whose are left out.

Societal Implications and Policy Concerns

The problems with facial recognition aren't just theories anymore; they're happening to real people. When biased systems get used by police or government agencies, things can go really wrong. People have already been misidentified, and some have even been arrested for crimes they didn't commit. In her 2019 report *Garbage In, Garbage Out,* Clare Garvie talks about how many police departments in the U.S. use facial recognition tools that rely on mugshot databases. Since those databases include a higher number of people of color, the technology ends up repeating the same racial bias that already exists in the system.

Garvie mentions several cases where Black men were wrongly arrested because of facial recognition errors. These aren't random mistakes; they show how bias gets built into the process itself. If biased data goes in, bad results come out. And when those results are used in law enforcement, the consequences can be serious and unfair.

Najibi (2020) argues that fixing this isn’t just about improving the tech; it also means putting real rules and limits in place. She points out that cities like San Francisco and Boston have already banned or restricted police use of facial recognition because of concerns about accuracy and ethics. Najibi also notes that the issue isn’t only about data. It’s about the choice to keep using something that people already know is flawed.

All of this shows that facial recognition bias isn’t just a technology problem; it’s a moral one. Companies like to say their systems are neutral, but if they make more mistakes for certain groups, that neutrality doesn’t mean much. And when government agencies use those systems without enough oversight, it becomes hard to tell the difference between progress and injustice.

Counterarguments and Rebuttals

Some argue that facial recognition bias is being exaggerated. They claim algorithms are improving, datasets are becoming more diverse, and that soon FRT will reach demographic parity. Others point out that humans are biased too; eyewitnesses make mistakes, police lineups are flawed, so even imperfect AI might still be an improvement.

It’s true that progress is being made. But that doesn’t erase the underlying issue. As studies by Buolamwini and Gebru (2018) and Grother et al. (2019) show, even the most advanced systems still have noticeable gaps in accuracy across races and genders. The problem goes deeper than just needing “more data.” It’s about the power structures shaping how technology is built and used.

Even if algorithms someday achieve equal accuracy for all faces, that doesn’t automatically make their use ethical. As Garvie (2019) warns, deploying these tools in policing without transparency or regulation risks institutionalizing discrimination. So, while improving the technology is important, it’s only one piece of a much bigger puzzle.

Conclusion

Facial recognition technology really shows both the promise and the pitfalls of AI. It can be powerful and helpful, but it can also be flawed in ways that reflect society’s existing biases. Research from Buolamwini and Gebru (2018), Grother et al. (2019), Wu et al. (2022), Najibi (2020), and Garvie (2019) all point to the same problem: FRT systems tend to make more errors when identifying people of color, mainly because of biased datasets, tricky environmental conditions, and careless or inconsistent use by institutions.

These issues aren’t just technical glitches; they mirror bigger inequalities in society. Fixing them takes more than better algorithms. It means having diverse teams working with diverse data, testing systems in real-world conditions, and putting strong rules in place to prevent misuse. While tools like police databases can play an important role in public safety, other uses, like mass surveillance, carry serious risks and need careful limits and oversight.

At the heart of it, this isn’t just about technology recognizing faces. It’s about recognizing people; fairly, fully, and with the respect everyone deserves.

References

Buolamwini, J., & Gebru, T. (2018). *Gender shades: Intersectional accuracy disparities in commercial gender classification.* Proceedings of Machine Learning Research, 81, 1–15. http://proceedings.mlr.press/v81/buolamwini18a.html

Garvie, C. (2019). *Garbage in, garbage out: Face recognition on flawed data.* Georgetown Law Center on Privacy & Technology. https://www.flawedfacedata.com

Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face recognition vendor test (FRVT) Part 3: Demographic effects.* National Institute of Standards and Technology (NIST) Interagency/Internal Report 8280. https://doi.org/10.6028/NIST.IR.8280

Najibi, A. (2020, October 24). *Racial discrimination in face recognition technology.* Harvard University: Science in the News. https://sciencepolicy.hsites.harvard.edu/blog/racial-discrimination-face-recognition-technology

Wu, H., Albiero, V., Krishnapriya, K. S., King, M. C., & Bowyer, K. W. (2022). *Face recognition accuracy across demographics: Shining a light into the problem. arXiv preprint arXiv:2206.01881.* https://arxiv.org/abs/2206.01881