

# Critical Colorblind Technologies

Essay for the Ethics for Nerds Lecture  
Summer Term 2020

Ursela Barteczko, 2579123

Word Count: 2025

August 18, 2020

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>The Argument</b>	<b>3</b>
2.1	The Basic Idea . . . . .	3
2.2	Tabular Form . . . . .	4
2.3	Soundness Reasoning . . . . .	4
2.4	Attack . . . . .	8
2.5	Defense . . . . .	8
<b>3</b>	<b>Conclusion</b>	<b>8</b>
	<b>Bibliography</b>	<b>10</b>

# 1 Introduction

Due to the #blacklivesmatter demonstrations, the topic "racism" once again moved into public focus[1]. Racism is discrimination based on "race", often closely linked to the skin color of a person, and has a long history starting with slavery in the early 17th century[3]. Ever since then, persons of color (POC) had to face disadvantages in European and American societies[2]. Racism, be it individual, institutional or structural, is a problem in nowadays society, leading to protests and deaths. It is clearly against the Universal Declaration of human rights, which states that "All human beings are born free and equal in dignity and rights." in Article 1[5].

Certain technologies perform worse for POC than for white persons or reinforce racist stereotypes already present in society. An example is the COMPAS software, also seen in the lecture, which predicted a higher probability of recidivism for black people than for whites. A UK visa application reviewing software that was heavily biased by nationality was scrapped just a few days ago[6]. Software of this kind can have a grave impact on the lives of persons, by staying in jail longer than necessary or not being able to work in the UK. Technologies with a grave impact on a person's life are called critical technologies here.

The topic of biased data is not new, and a racist bias might simply be due to true technical limitations or because a white developer team was simply not aware of the issue and forgot to test the performance of their software for all types of skin colors. In these cases, the development team might or might not be aware of the flaw of their software.

However, the responsibility of employing a certain technology normally lies with additional decision-makers from politics, administration, or business. This essay will focus on the normative status of this person's or entities' actions.

Intuitively one might say, that employing a new technology is permissible if everyone benefits from it, even if that benefit is not of equal size for every member in society.

However, I will argue, that we should not employ critical technologies that perform worse for people of color than for white people (which I will abbreviate with ECT) taking into account a broad range of moral theories.

The structure of this essay contains the basic idea behind my argument, the argument in extended standard form, followed by soundness reasoning, an attack, and a counterattack. The main findings will be summarized in the conclusion.

## 2 The Argument

In this essay, I argue in favor of the claim "It is morally forbidden to employ critical technologies that perform worse for people of color (POC) than for white people".

Underlying assumptions are the following:

1. The decision-maker can be held accountable for his or her act, either because they knew about the worse performance for POC, or because they omitted extensive testing to exclude this possibility.
2. The only alternative to ECT is not employing that technology at all.

### 2.1 The Basic Idea

I argue that ECT is a racist act. If this is the case and all major families of normative theories agree on racist acts being morally forbidden, it is strong evidence that racist acts are indeed morally forbidden, and therefore also ECT is morally forbidden since it is an instance of a racist act.

The major families of normative theories illuminated here will be a subset of consequentialist theories, deontological theories, and virtue theories. The following section presents my argument in tabular form.

## 2.2 Tabular Form

P1:	Employing critical technologies that perform worse for POC than for white persons imposes a systematic disadvantage for POC relative to white people.
P2:	If an act imposes a systematic disadvantage for POC relative to white people, it is a racist act.
C1:	<i>Therefore:</i> Employing critical technologies that perform worse for POC than for white persons is a racist act. (P1 – P2)
P3:	Deontological theories forbid racist acts.
P4:	Virtue theories forbid racist acts.
P5:	Most forms of consequentialist theories forbid racist acts.
C2:	<i>Therefore:</i> A broad range of moral theories forbids racist acts. (P3 – P5)
P6:	If C2, <i>probably</i> racist acts are morally forbidden.
<hr/>	
C:	<i>Therefore:</i> Employing critical technologies that perform worse for POC than for white persons is <i>probably</i> morally forbidden. (C1, C2, P6)

## 2.3 Soundness Reasoning

A possible logical form in first-order logic could be the following:

P1:	$SystematicDisadvantage(ECT)$
P2:	$\forall x[SystematicDisadvantage(x) \rightarrow RA(x)]$
C1:	$Therefore: RA(ECT) \text{ (P1 - P2)}$
P3:	$MT(Deontological) \rightarrow \forall x\neg RA(x)$
P4:	$MT(Virtue Theories) \rightarrow \forall x\neg RA(x)$
P5:	$MT(Consequentialist) \rightarrow \forall x\neg RA(x)$
C2:	$Therefore: MT(x) \rightarrow \forall y\neg RA(y)$ $  x \in [Deontological, Virtue Theories, Consequentialist]$ (P3 - P5)
P6:	$(\exists x MT(x) \rightarrow \forall y\neg RA(y)) \rightarrow \forall z\neg RA(z)$
<hr/>	
C:	$Therefore: \neg ECT \text{ (C1, C2, P6)}$

Abbreviations:

1. RA: Racist Act
2.  $\neg RA$ : Racist Act is forbidden
3. ECT: Employing critical technologies that perform worse for POC than for white persons
4. MT: Moral theory

In C2 and P6 we *prima facie* see defeasible reasoning when generalizing from the observations on deontological moral theories, consequentialist theories as well as virtue theories to a general deontic status of racist acts. However, this is permissible, since there are no hot candidates for moral theories outside these families.

The following paragraphs will provide additional reasons to believe in the premises.

**P1:** The premise follows directly from the description of the essay claim provided by the lecturers.

**P2:** From the lecture, we know that "discrimination" consists of "acts, practices, or policies that impose a relative disadvantage on persons based on their membership in a salient social group". If the salient social group is based on

the skin color, we have discrimination based on skin color, which is politically correct called shadeism or colorism, but can be seen as a subphenomenon of racism[3].

**P3:** In the lecture, we saw deontological frameworks of Immanuel Kant and Scanlon. I will briefly outline a possible application of their principles to our case. Kant is the more interesting candidate in this case. While he worked for decades on a hierarchy and classification of different human races[7], his moral theory with the categorical imperative can be used to show that racist acts are morally forbidden. This is done by formulating the Maxime and resulting Universal Law, as well as answering the questions if the Maxime is consistently conceivable in a world where everyone always follows the Universal Law and also if anyone can willfully want to live in a world where everyone follows the Universal Law:

*Maxime:* I treat some people worse than others because they are physically less like me.

*Universal law:* Everyone treats humans that are physically different from themselves worse than people that look alike to them.

*Is the Maxime consistently conceivable in a world where everyone always follows the Universal Law?* Yes, it is. A completely racist world is conceivable, but would probably lead to less mixing of different cultures and reluctance against international cooperation as well as an increase in other forms of discrimination, such as sexism, ableism, adultism, and others.

*Can I willfully want to live in that world?* No, international cooperation and mixing of different cultures have brought many advantages to society and humankind. Additionally, one would likely be exposed to huge amounts of physical and verbal violence in a very hostile world.

While Kantianism is hard to apply, and there are also different Maximes and Universal Laws that can be derived for one act, Scanlon paints a clearer picture. Any person that is a victim of racism has to carry an additional burden, that is high enough to reject all sets of principles that could allow a racist act.

**P4:** In the lecture, we saw the eudaemonist virtue ethics of Aristotle, which state that an agent should be virtuous. So, a virtue is a trait, that contributes to *Eudaimonia*, roughly translating to "happiness", "flourishing", "well-being". Since racist acts are often linked to negative emotions such as hate or pride,

we can argue that acting in a racist way, goes against *Eudamonia*. Also, acting racist is against the virtues of "Charity" and "Kindness", as identified by Thomas Aquinas. From Aristotle's list of vices, "Boastfulness", "Spitefulness" and "Insensibility" are vices expressed by acting in a racist way.

However, if the agent is not aware of acting racist (violating one of the assumptions stated as underlying), either by not being aware of the flaw<sup>1</sup> in the software or about the dangers of ECT, it is difficult to state that the agent is expressing some vice. In this case, he or she is still lacking the *Phronesis* or practical wisdom, an agent needs in order to become a virtuous agent.

**P5:** Consequentialist theories examine relevant qualities of the consequences of an action and check if those fulfill a specific condition. The focus lies on Hedonism or Preference Theory, assuming that we want to maximize pleasure or preference satisfaction, rather than on the niche Objective List Accounts. Any consequentialist theories that want the consequences to have an equalizing effect would clearly forbid any racist acts. The ones that aim for an individual threshold of utility that needs to be surpassed would be against racism if the threshold were between the utility for POC and white persons. For collective satisfaction, the permissibility of the action would probably depend on the composition of the beings affected. If the portion of POC is small enough, their relative disadvantage might be outweighed by the advantage of white people. Prima facie this also seems to be true for theories that aim at the maximization of pleasure and pain.

Generally speaking, racism is reinforced by any racist acts. And racism inflicts pain on the individual level and leads to preference frustration of discriminated groups on various levels of discrimination since their access to societal resources is limited. So if we abstract from the consequences of one specific action to the overall picture of society, consequentialist theories are likely to forbid racist acts.

**P6:** I show here, that arguing from the standpoints of diverse normative theories with different basic assumptions racist acts should be forbidden. In my considerations I covered all three mayor families of normative theories, so this is strong evidence, that probably racist acts are morally forbidden.

---

<sup>1</sup>Flawed" in this context means the worse performance for POC compared to white persons.

## 2.4 Attack

A promising point of attack here could be that the selection of moral theories was not broad enough and missing the influential Classical Utilitarianism, thereby attacking P5 and P6. This moral theory *prima facie* leads to the very intuitive counterargument mentioned in the introduction:

If the additional pleasure by the introduction of the technology, even though it is flawed, is bigger than the pleasure in the status quo without the technology, it should be introduced. It is beneficial for white people and also to a less degree beneficial for POC so we should employ the technology to maximize overall pleasure.

## 2.5 Defense

Classical Utilitarianism is very clearly not applicable in this case, since it is based on the objective consequences of an action. We cannot know the objective consequences of any action, therefore classical Utilitarianism is not suitable as a guide for a decision procedure. However, applying the Principle of Charity, we might extend the argument to any form of Utilitarianism maximizing the expected pleasure.

Racism is constituted by racist acts. I already pointed out, that racism causes pain and preference frustration. This happens on a global scale. It is highly unlikely, that the pleasure that comes from ECT can outweigh the pain that comes from living in a racist society. The fight against racism is a daily one, the technology might just be used once in a while. Also, the pleasure invoked through the technology will come with a huge grain of salt for the POC if they are aware of being discriminated against by exactly that technology. Additionally, people that discriminate one group of humans, are prone to start discriminating against other groups of humans too[4]. Therefore, the pain caused by racism could multiply, as other forms of discrimination reach the middle of society.

Overall, in the vast majority of the cases, Pleasure Maximizing Utilitarianism can also be used to argue against ECT.

## 3 Conclusion

I argued for the claim that we should not employ critical technologies that perform better for white persons than for POC by showing that that would be racist



and arguing that normative theories forbid racist acts. This is especially clear from the standpoint of deontological theories, but also virtue theories and consequentialist theories give strong evidence against employing critical technologies that perform better for white persons than for POC.

#blacklivesmatter

## References

- [1] BLACK LIVES MATTER MOVEMENT. Black lives matter. <https://blacklivesmatter.com/about/> (Accessed: 2020-18-08).
- [2] BUNDESZENTRALE FÜR POLITISCHE BILDUNG. Rassismus. <https://www.bpb.de/politik/extremismus/rechtsextremismus/213670/rassismus> (Accessed: 2020-08-17), 2015.
- [3] HURFORD, K. Shadeism and the politics of skin tone. <http://socialistreview.org.uk/425/shadeism-and-politics-skin-tone> (Accessed: 2020-08-17), 2017.
- [4] SHOOK, N. J., FITZGERALD, H. N., AND ET AL. Sexism, racism, and nationalism: Factors associated with the 2016 u.s. presidential election results? <https://doi.org/10.1371/journal.pone.0229432> (Accessed: 2020-08-18), 2020.
- [5] UNITED NATIONS. Universal declaration of human rights. <https://www.un.org/en/universal-declaration-human-rights/index.html>) (Accessed: 2020-18-08), 1948.
- [6] WALDRON, D., AND ALI, S. Racist uk visa software algorithm was scrapped. <https://workpermit.com/news/racist-uk-visa-software-algorithm-was-scrapped-7-august-20200809>.
- [7] ZACK, N. Ideas of race in the history of modern philosophy. In *The Oxford Handbook of Philosophy and Race* (2017), Oxford University Press.