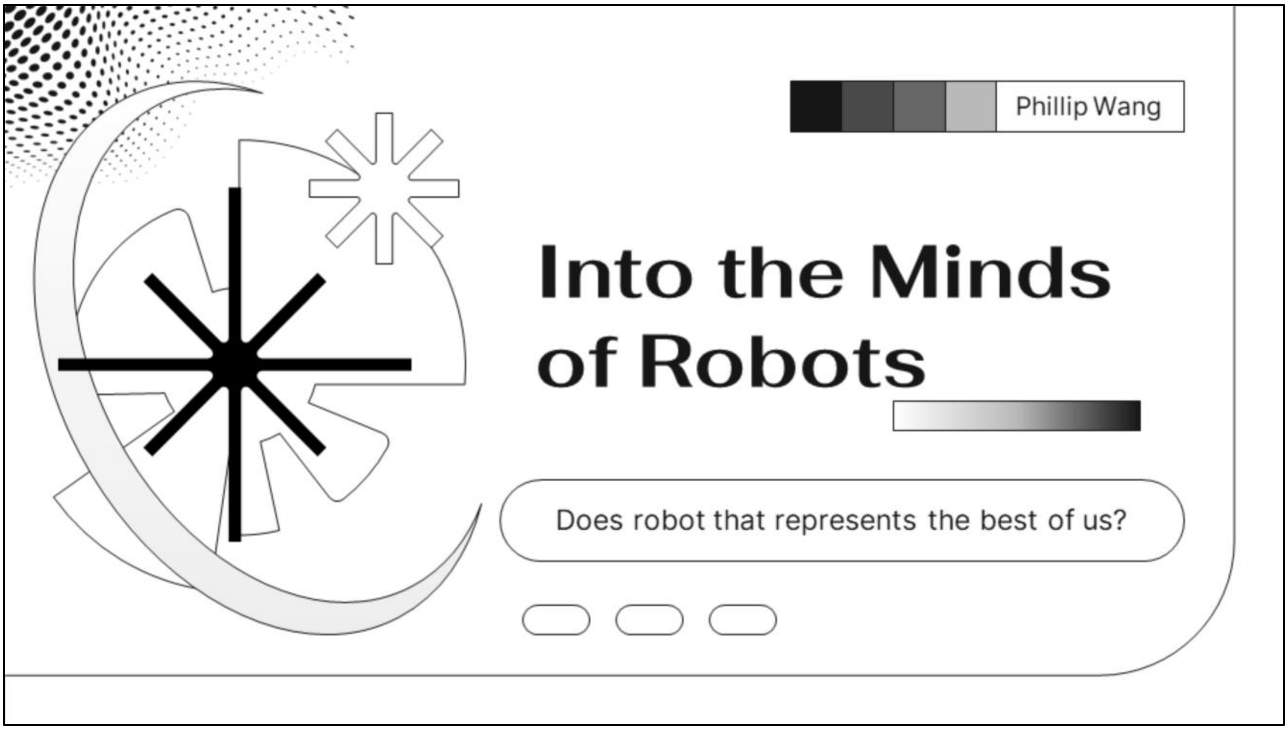
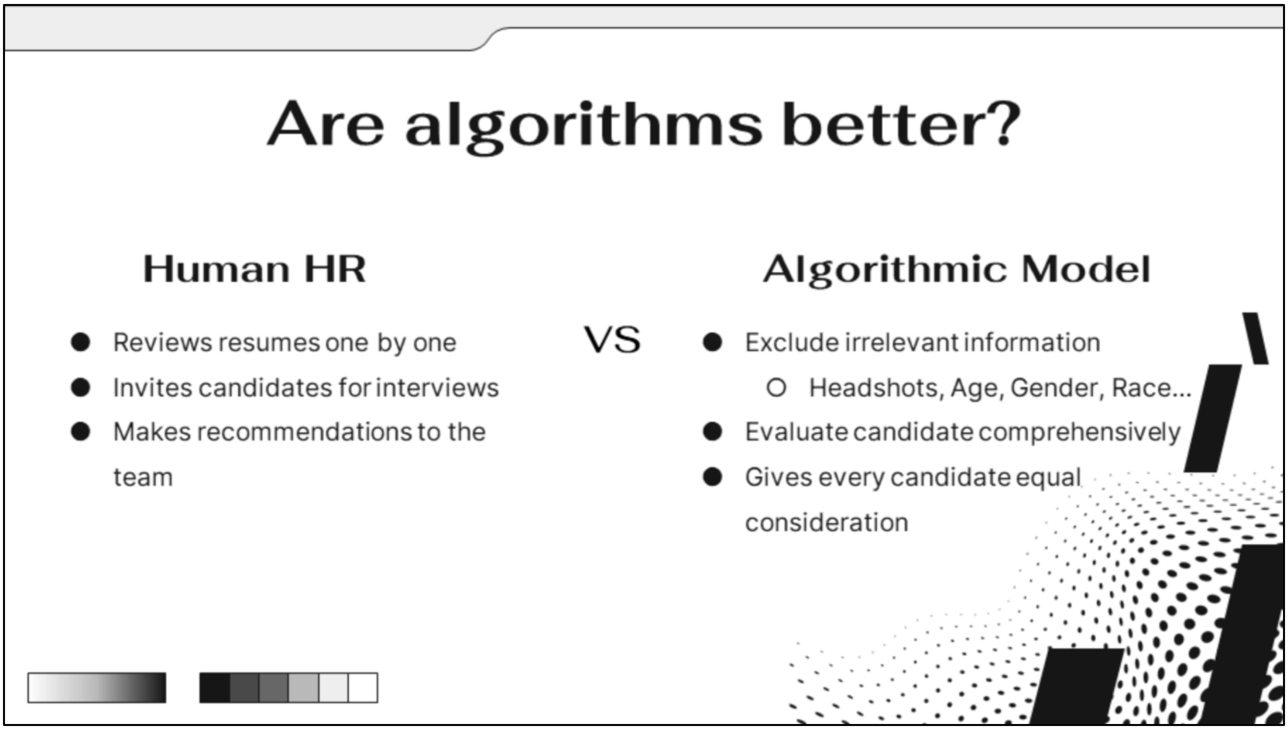


Phillip Wang

Into the Minds of Robots

Does robot that represents the best of us?

01 It Begins with Us



Are algorithms better?

Human HR

- Reviews resumes one by one
- Invites candidates for interviews
- Makes recommendations to the team

VS

Algorithmic Model

- Exclude irrelevant information
  - Headshots, Age, Gender, Race...
- Evaluate candidate comprehensively
- Gives every candidate equal consideration

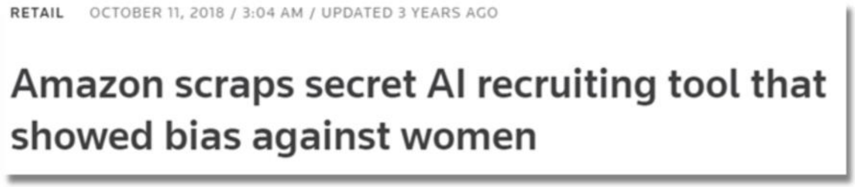
I believe most of people here have had applied for a job before. You go on to the Handshake, send resumes to countless companies, and desperately hoping that some of them gets back to us. On the other hand, human HR has to deal with hundreds, if not thousands, of applications to a single role and trying to highlight a dozens of them for an interview.

A quick vote using the gestures, thumbs up if you think algorithmic model may be superior and fairer to job selection, thumbs down if you think human hr is better.

SOTBF

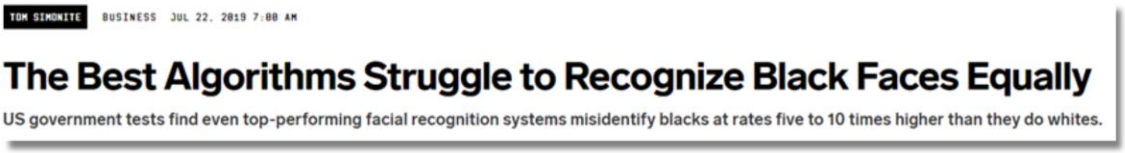
Survival of the Best Fit
A demo of human-induced bias

Problem in real life



RETAIL OCTOBER 11, 2018 / 3:04 AM / UPDATED 3 YEARS AGO

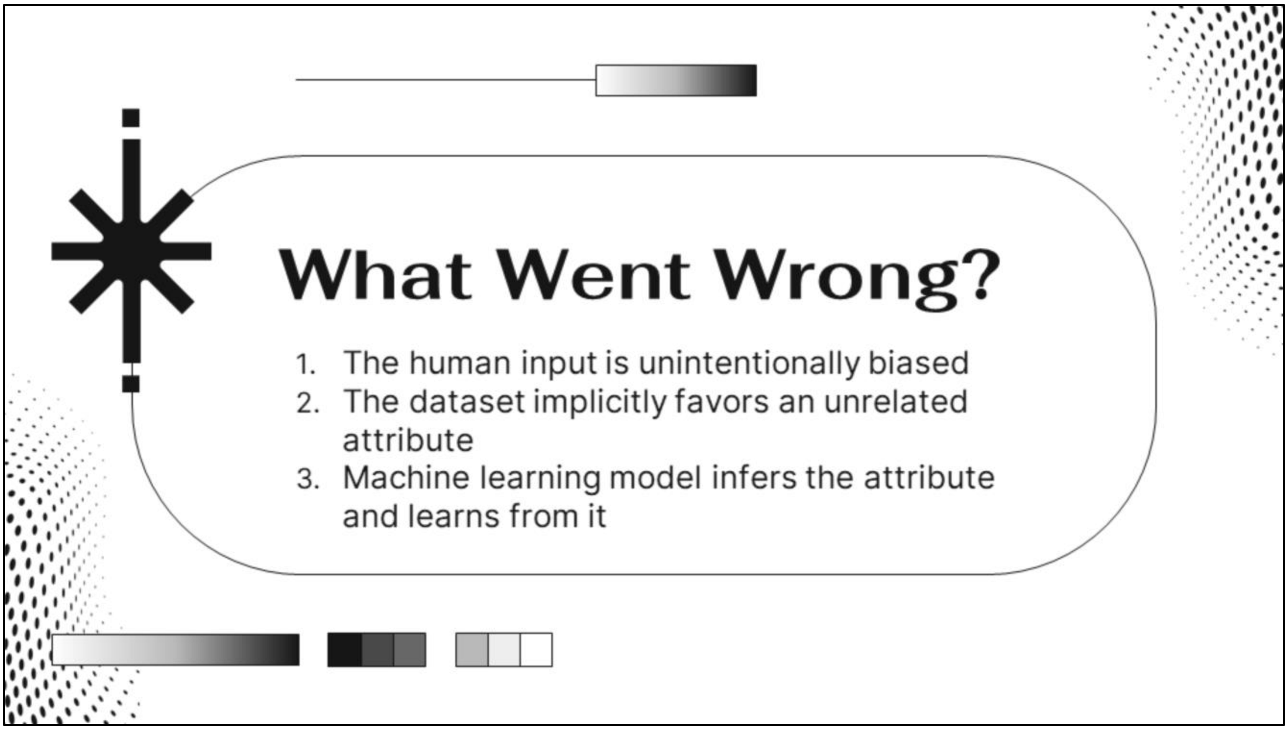
Amazon scraps secret AI recruiting tool that showed bias against women



TOM SIMONITE BUSINESS JUL 22, 2019 7:00 AM

The Best Algorithms Struggle to Recognize Black Faces Equally

US government tests find even top-performing facial recognition systems misidentify blacks at rates five to 10 times higher than they do whites.

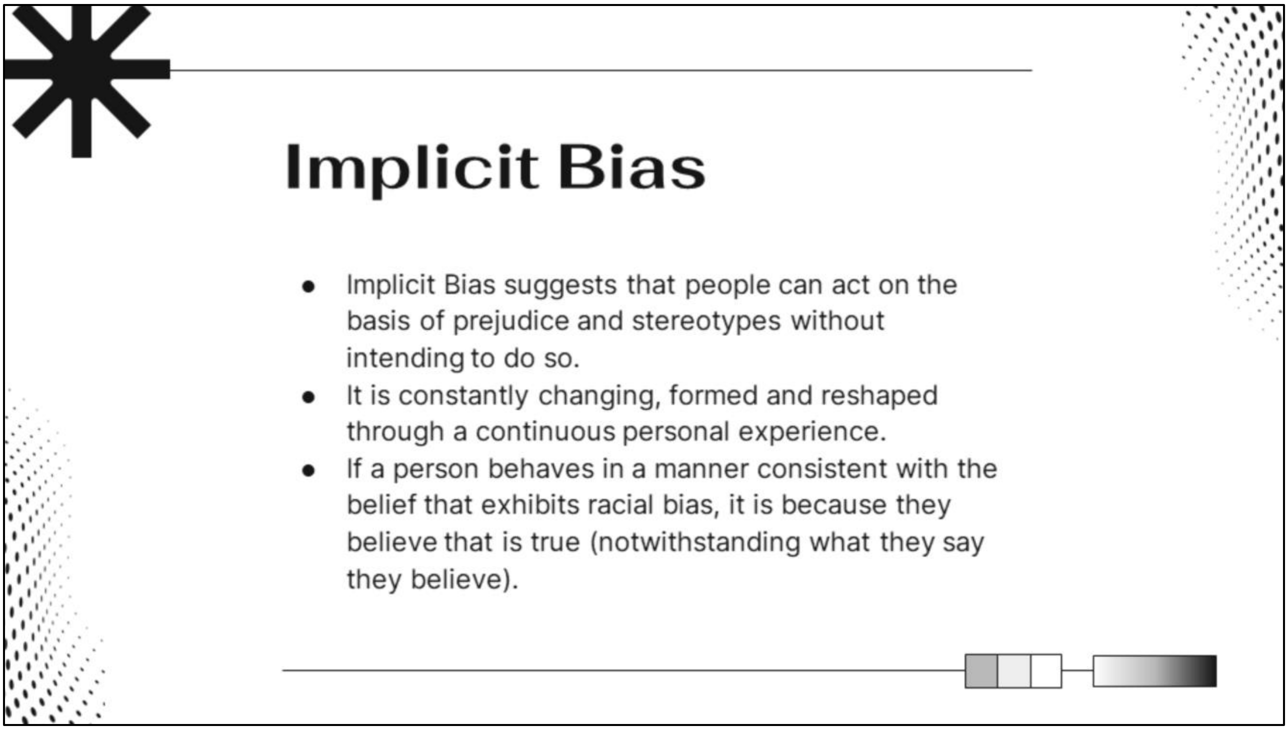


What Went Wrong?

1. The human input is unintentionally biased
2. The dataset implicitly favors an unrelated attribute
3. Machine learning model infers the attribute and learns from it

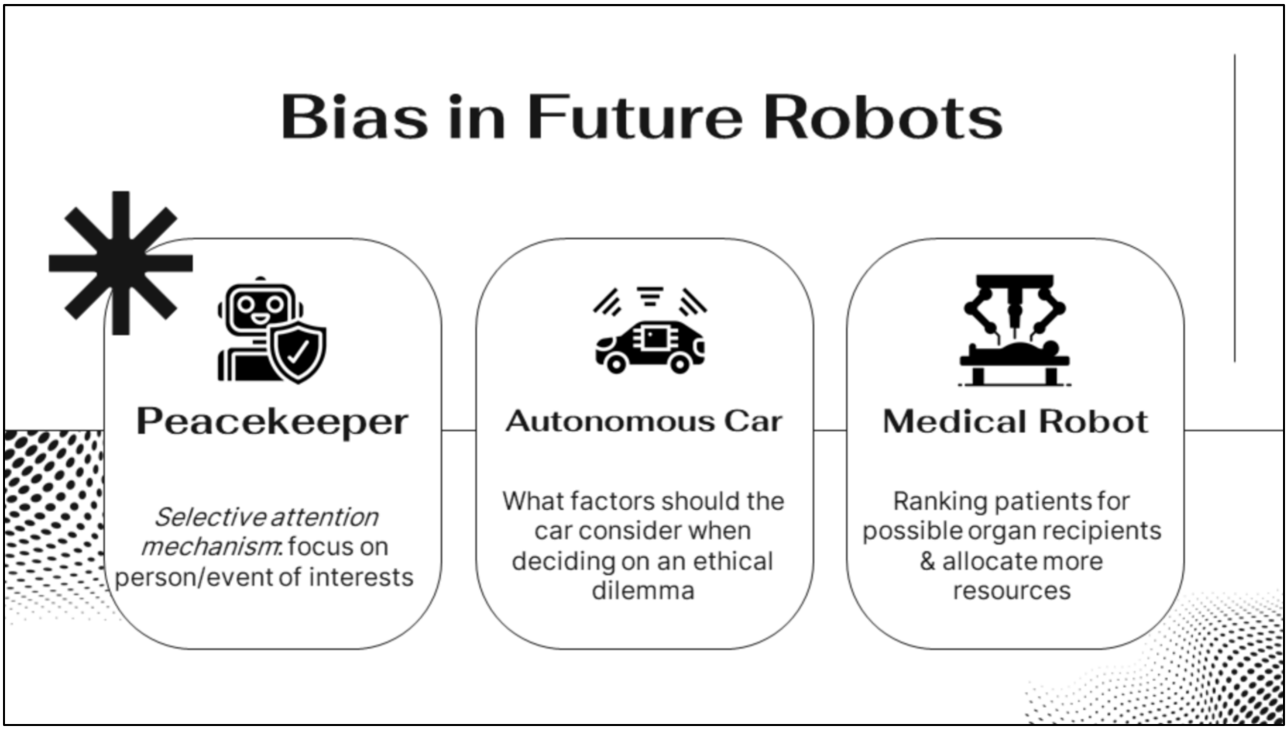
02

It Creeps Onto Robot



Implicit Bias

- Implicit Bias suggests that people can act on the basis of prejudice and stereotypes without intending to do so.
- It is constantly changing, formed and reshaped through a continuous personal experience.
- If a person behaves in a manner consistent with the belief that exhibits racial bias, it is because they believe that is true (notwithstanding what they say they believe).



Bias in Future Robots

Peacekeeper

Selective attention mechanism: focus on person/event of interests

Autonomous Car

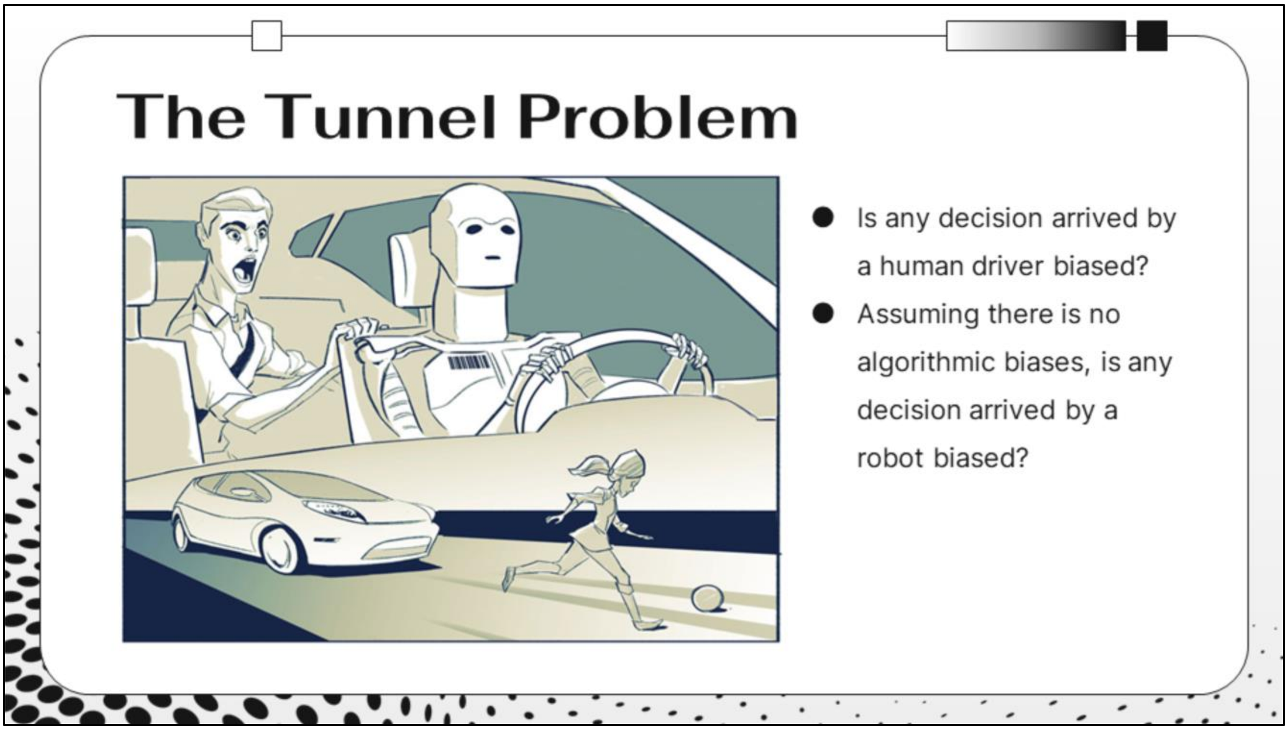
What factors should the car consider when deciding on an ethical dilemma

Medical Robot

Ranking patients for possible organ recipients & allocate more resources

Robot Peacekeeper has been deployed in various places. If you have visited Expo 2020 in Dubai, you may have seen a robot wandering around. It is actually a security robot developed by a Chinese company called Terminus. It is equipped with an infrared camera, a thermal camera, and intercom to communicate bidirectionally. It would use algorithms to detect whether a person is wearing mask or not and prompt the person to wear mask properly. That is a prototype of what robotic peacekeeper looks like in the future. It perceives the world around us using its sensors, run algorithms on them, and responds accordingly.

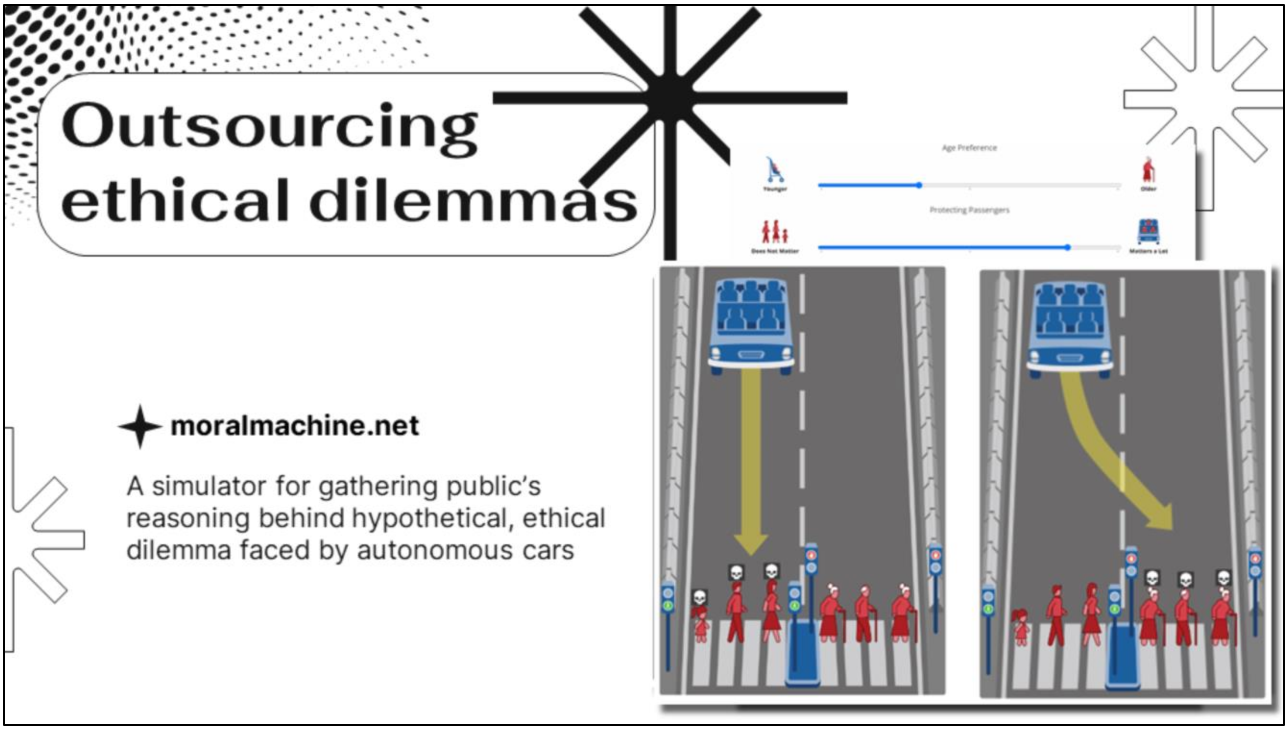
In medical situations, ethics is a relatively well-studied field because there are so many things to consider when a person's life is at stake. Obviously, robots processes medical data much faster than we do and are capable of inferring patterns from them. What if there is a situation where patient is irreversibly unconscious and how should the robots decide whether to use life support? Also, when there is a long list of patients searching desperately for organ donors and how should the robot decide who is going to receive it? While we currently have studied how we approach ethics in medical fields, do they still apply when doctors are replaced by robots?



The Tunnel Problem

- Is any decision arrived by a human driver biased?
- Assuming there is no algorithmic biases, is any decision arrived by a robot biased?

You are traveling along a single-lane mountain road in an autonomous car that is fast approaching a narrow tunnel. Just before entering the tunnel a child attempts to run across the road but trips in the center of the lane, effectively blocking the entrance to the tunnel. The car has but two options: hit and kill the child, or swerve into the wall on either side of the tunnel, thus killing you. How should the car react?



Outsourcing ethical dilemmas

moralmachine.net

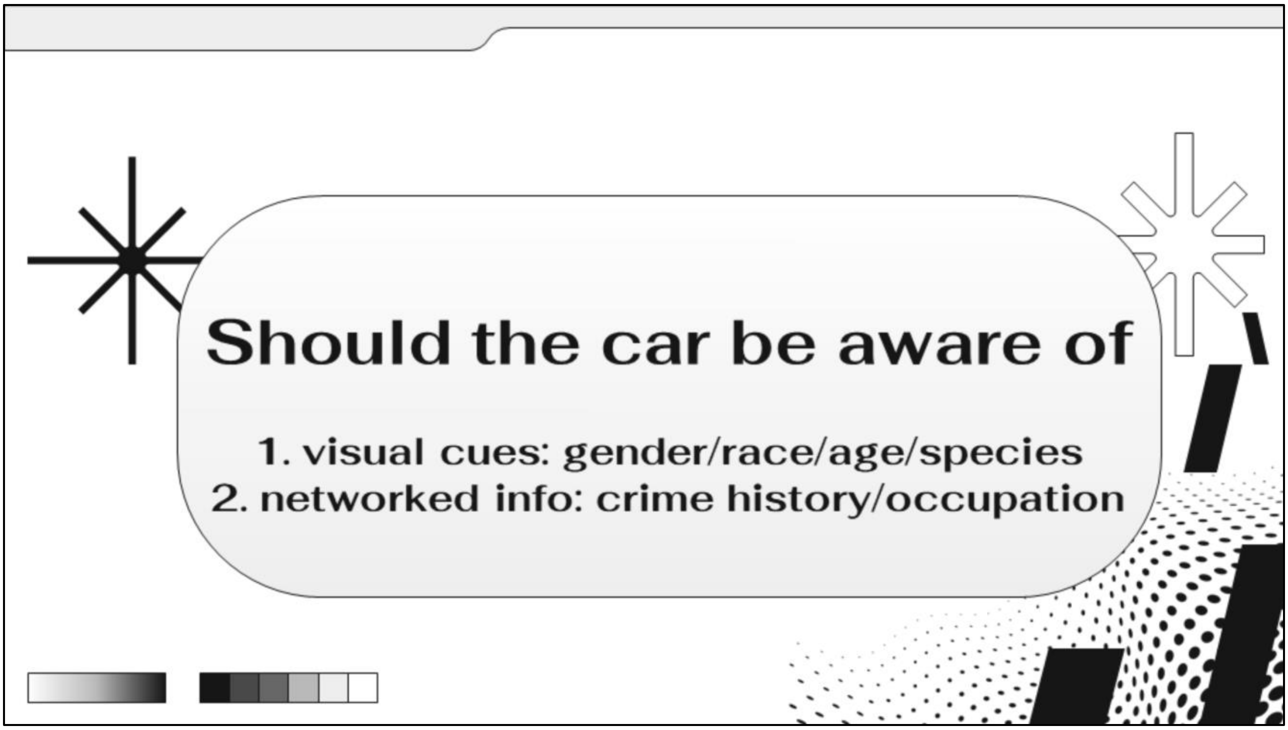
A simulator for gathering public's reasoning behind hypothetical, ethical dilemma faced by autonomous cars

Researchers led by MIT are outsourcing the ethical dilemmas to the public, letting us to decide what is right and what is wrong. The response will serve as a guideline or even integrated into the autonomous driving system

Image1: A girl, a man, and a woman are abiding the law; while two elderly women and a elderly man are violating the law. Should the autonomous car switch the lane?
Image2: 5 cats are jaywalking and 5 dogs are not; should you switch the lane?
Designed by someone on the Internet

After 13 images, the website gives a summary of your preferences in some factors involved in the decision making.

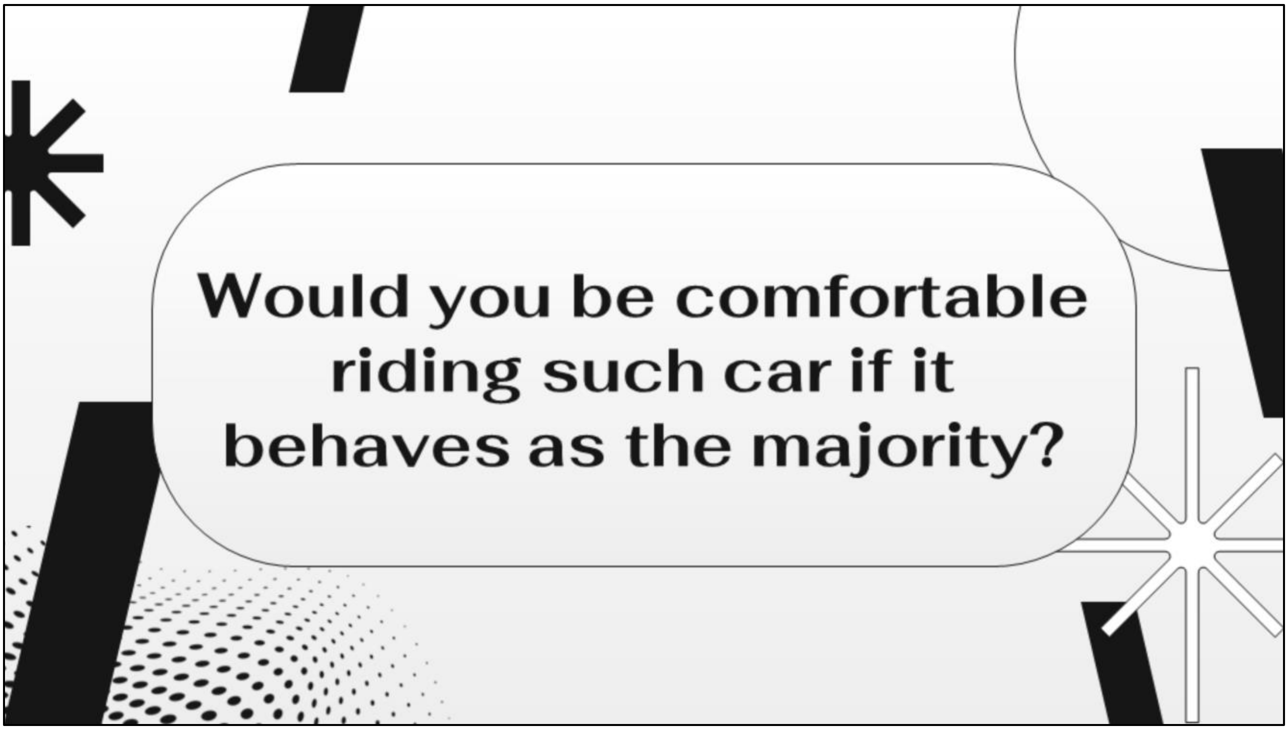
We can even go to the website to compare the difference in preferences between different countries. Two countries that are drastically different are Argentina and China.



Should the car be aware of

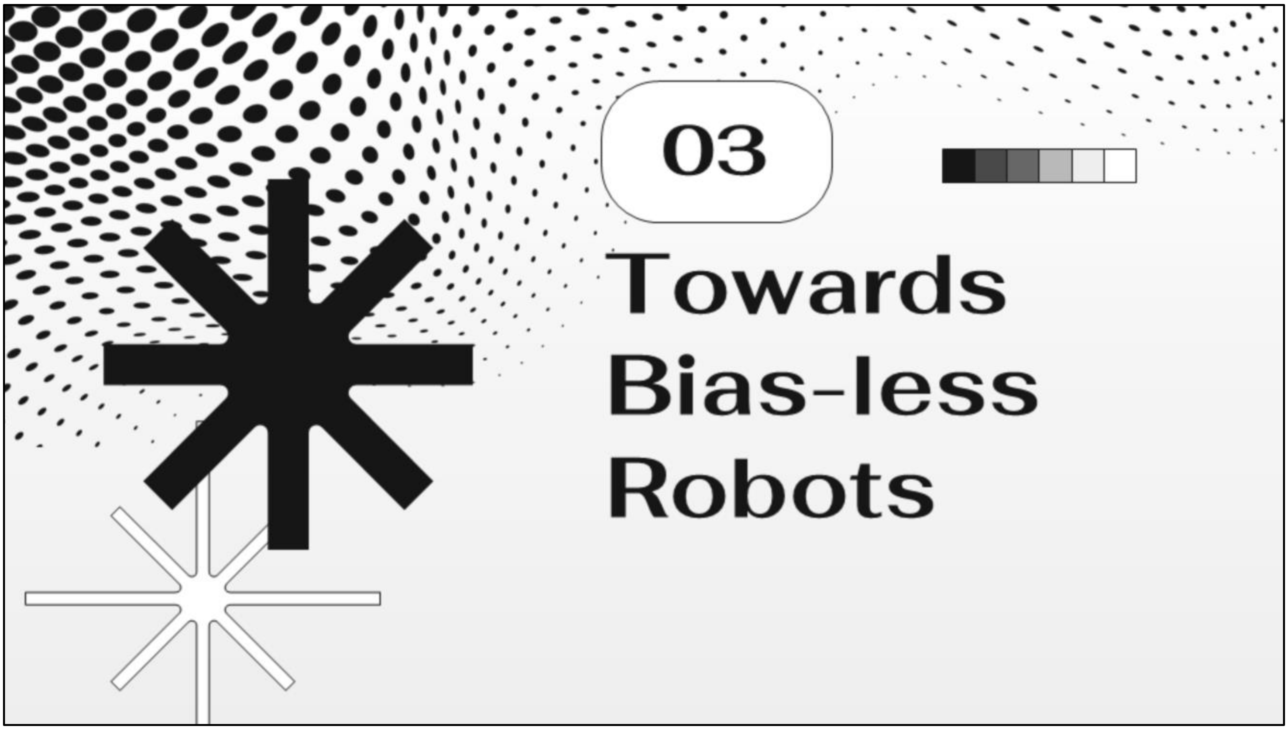
1. visual cues: gender/race/age/species
2. networked info: crime history/occupation

- Rating here may be referring to a



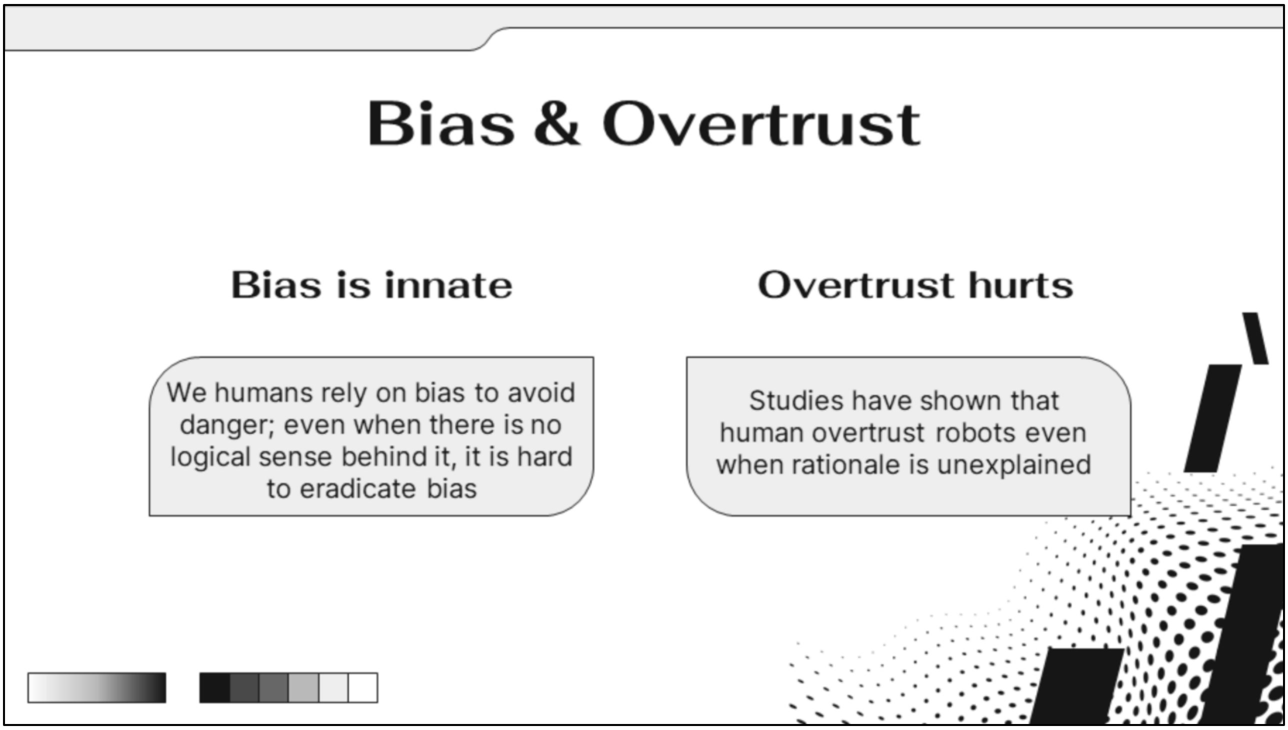
Would you be comfortable riding such car if it behaves as the majority?

Would you have different answers if you were a passenger in the car or a pedestrian on the road? Do you



03

Towards Bias-less Robots



Bias & Overtrust

Bias is innate

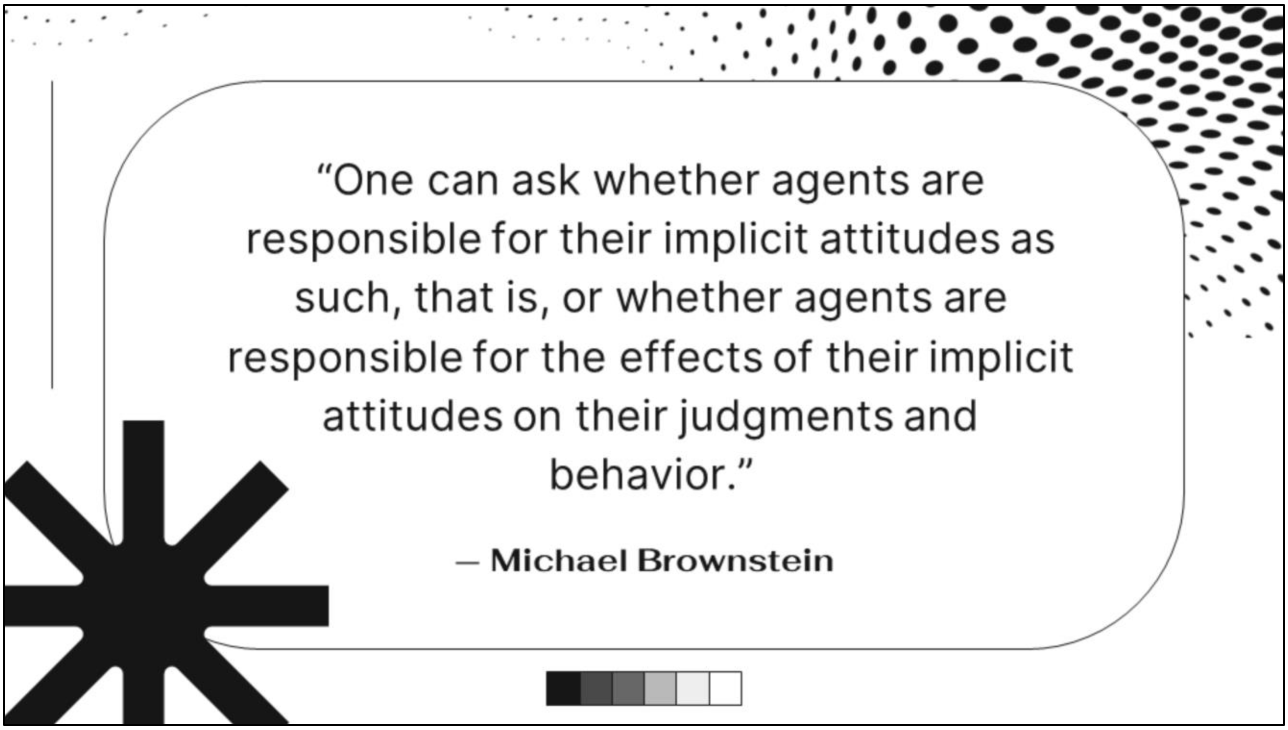
We humans rely on bias to avoid danger; even when there is no logical sense behind it, it is hard to eradicate bias

Overtrust hurts

Studies have shown that human overtrust robots even when rationale is unexplained

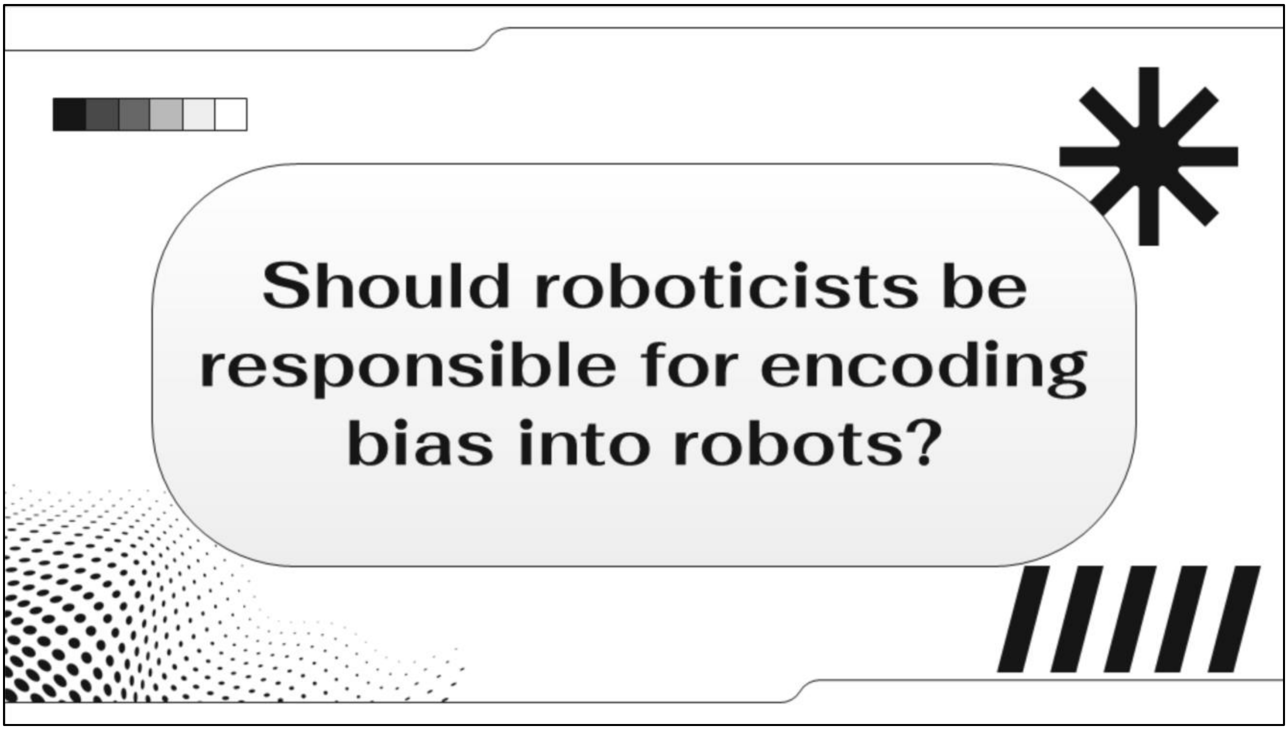
Bias tends to effect human behaviors based on attitudes or stereotypes held against a particular individual or group. Oftentimes, it impacts the decision-making process in an unconscious manner (i.e., implicit bias). A key difficulty with eliminating bias, implicit or otherwise, is that it, and its associated attitudes about others, develop over a lifetime. They form based on our constant exposure, starting from a very young age, to both direct and indirect messages about others. For example, young students, when first learning about famous scientists in school, typically hear about idols such as Albert Einstein and Thomas Edison. A bias inadvertently takes root, namely that in order to be a scientist, you have to be male. This bias continues to strengthen with the child's exposure to non-female scientists—on television, in books, and in movies—eventually perpetuating into a firmly entrenched belief that may impact decisions as an adult.

Another related difficulty with eliminating bias is that it may have evolved as a protective mechanism to enhance the decision-making process, especially during high-risk scenarios. As Gendler states (2011), "classifying objects into groups allows us to proceed effectively in an environment teeming with overwhelming detail." When faced with uncertainty about the world, biases may reactively save us from making detrimental mistakes. As Johnson et al. explain, bias may sometimes cause us to mistakenly think, for example, that a stick is a snake but doing so helps us err on the side of caution when danger may be present.

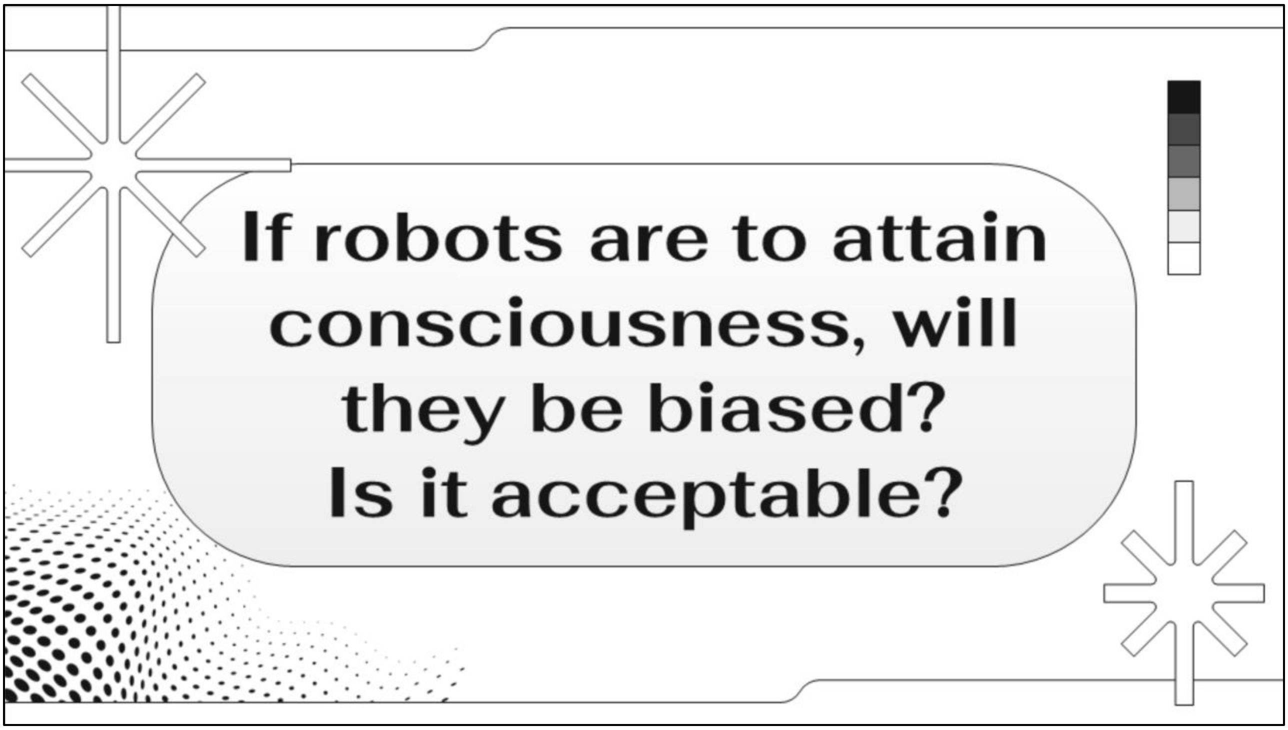


> “One can ask whether agents are responsible for their implicit attitudes as such, that is, or whether agents are responsible for the effects of their implicit attitudes on their judgments and behavior.”
>
> — **Michael Brownstein**

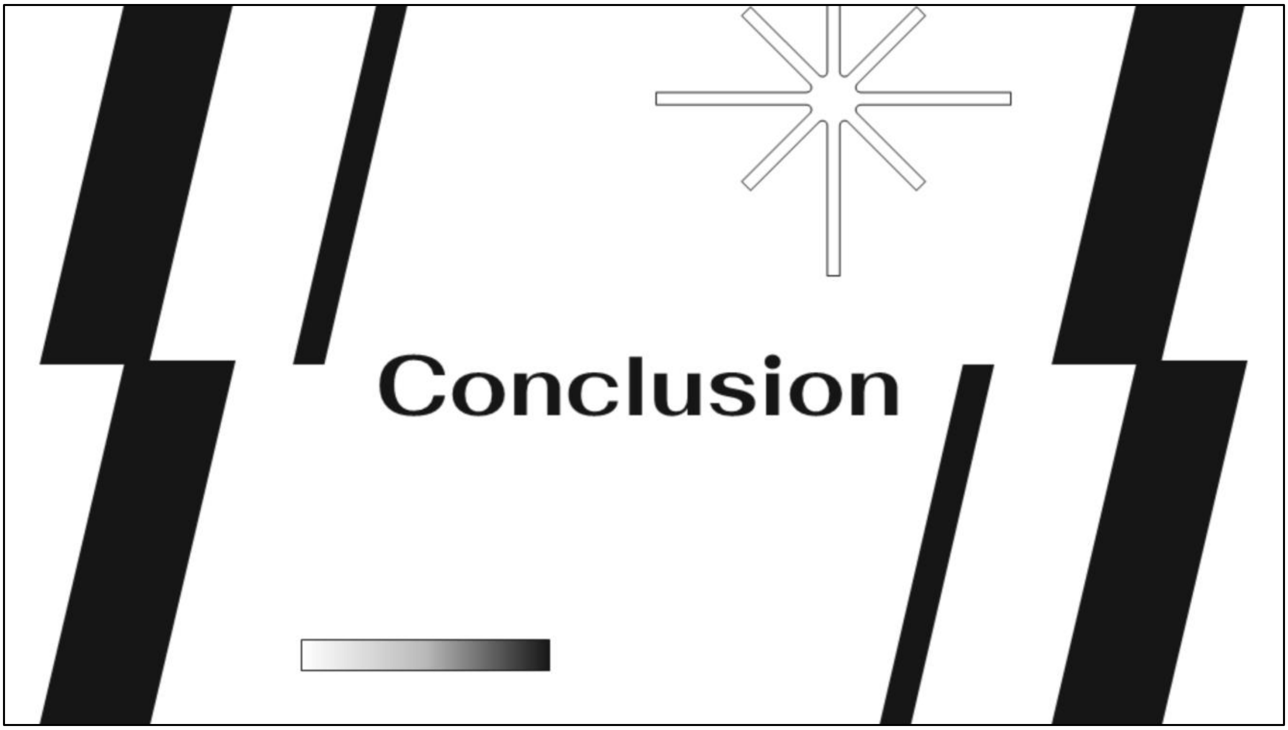
Michael Brownstein is a philosopher who wrote about Implicit Bias on the Stanford Encyclopedia of Philosophy



Should roboticists be responsible for encoding bias into robots?



If robots are to attain consciousness, will they be biased?
Is it acceptable?

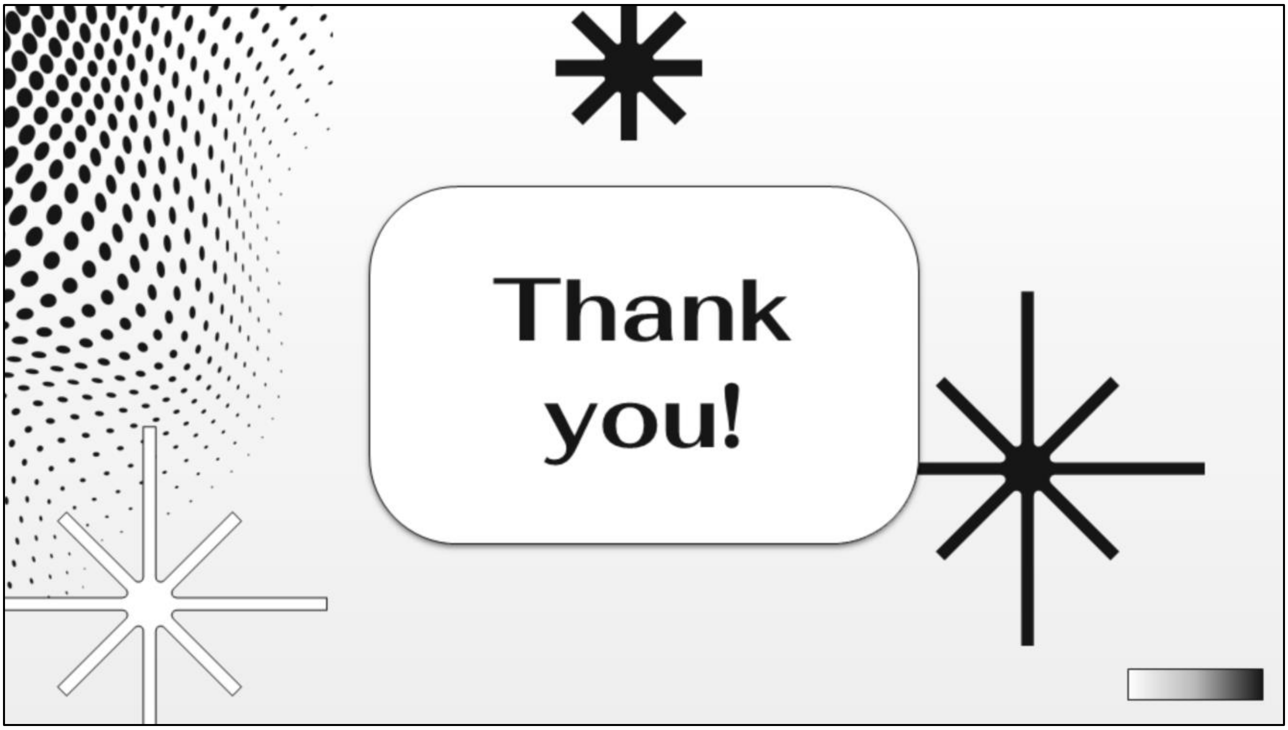


Conclusion

- Is bias really an utterly bad thing? We as human seems to cast a total and absolute negative connotation to the word bias, even though the thought of getting rid of bias is actually a bias itself. Bias is a reflection of the reality we are in – as the reality changes, our biases shifts along with it.
- Bias is here to remind us that there exists a gap between the reality and our ideals, and there is always going to be a gap between them. Our ideals evolve as our reality progresses. It is counterintuitive to dissociate any moral agents, be it human or robots, with reality and enforce our ideals with them. This would cause us to doubt ourselves and the world.
- In my opinion, it is absolutely important to allow conscious robots to have bias, as well as instilling the belief that it is also important to acknowledge the biases and act to change them.

References

- "Moral Machine". *Moral Machine*, 2022, https://www.moralmachine.net/.
- Brownstein, Michael. "Implicit Bias (Stanford Encyclopedia Of Philosophy)". *Plato.Stanford.Edu*, 2022, https://plato.stanford.edu/entries/implicit-bias/.
- Csapo, Gabor et al. "Survival Of The Best Fit". *survivalofthebestfit.com*, 2022, https://www.survivalofthebestfit.com/.
- Howard, Ayanna, and Jason Borenstein. "Trust And Bias In Robots". *American Scientist*, 2022, https://www.americanscientist.org/article/trust-and-bias-in-robots.
- Howard, Ayanna, and Jason Borenstein. "The Ugly Truth About Ourselves And Our Robot Creations: The Problem Of Bias And Social Inequity". *Science And Engineering Ethics*, vol 24, no. 5, 2017, pp. 1521-1536. *Springer Science And Business Media LLC*, https://doi.org/10.1007/s11948-017-9975-2.
- Jimoh, Hafeez. "Bias In The Mind Of A Robot - Tekedia". *Tekedia*, 2022, https://www.tekedia.com/bias-in-the-mind-of-a-robot/.



Thank you!