

4511 Writing 1

Aunya Mukherjee

February 07 2021

1 Do you agree with the paper on the importance of ethics in AI? Why?

I absolutely agree on the importance of ethics in AI. As we progress as a society and further integrate AI into our daily lives, ethics is one of the most important dilemmas we must face to make sure that the technology around us is aiding us more than it is putting us in danger. The engineers who create the algorithms that AI then will implement, are responsible for accounting for all sorts of scenarios; in which the answers may not be straight forward. I have read some of *“Algorithms of Oppression”* by Safiya Noble, which details a lot of the racial and gender bias in google’s use of AI and Machine Learning algorithms. While reading and researching about this I have become more aware of the countless areas of AI in which ethics is the most important deciding factor. Ethics absolutely must be put at the forefront of conversations that surround AI, because AI that is unethical will hurt society more than it will help.

2 In the example you have chosen, do you agree with the Social Norm presented in the paper? Why or why not? Would you add any?

I chose the example titled *“Trust and Ethics for Autonomous Vehicles.”* There were several Norms presented in the paper that relate to this example, all relating to the ways in which robots can avoid hurting human beings. The first and most simple is that

(SN-1) A robot will never deliberately harm a human being.

I think that this is clearly an important and basic Social Norm that we must include. Moving forward, the next Social Norm that was presented was

(SN-2) In a given situation, a robot will be no more likely than a skilled and alert human to accidentally harm a human being.

I think this norm is interesting because in the formulation of the norm, we also have to take into account the human propensity to err. The paper outlines several ways that we can measure human ability. We can compare different drivers against each other and record the time that it takes them to make decisions, as well as their ability to try to avoid hurting others. I agree with this Social Norm fundamentally, however I am interested in how they choose to set their measure of a skilled and alert human. I think the way in which we define a “skilled and alert human” could change the expectations for the robot, and I am curious to see if our expectations for robots will surpass our expectations of skilled and alert humans, in terms of autonomous vehicles.

Finally the last Social Norm that is presented deals with Deadly Dilemmas. The popular philosophy trolley problem is one of these dilemmas in which all options cause harm. This forces us to choose the “lesser of evils,” and make judgement calls based on what we deem important. The Social Norm that was constructed in relation to Deadly Dilemmas is

(SN-3) A robot must learn to anticipate and avoid Deadly Dilemmas.

I agree with the contents of this Social Norm, and think that the best way to deal with Deadly Dilemmas as a robot is to avoid them. However, I think that there should also be a set of Social Norms that give some sort of reasoning when faced with a Deadly Dilemma for a robot to follow. When humans are faced with Deadly Dilemmas, they take into account what their values are before making a decision (if they have enough time to). I wonder whether it is even possible to try to instill values and moral hierarchies into the algorithms that guide robots. I also understand that when we try to teach robots values and morals, we risk teaching it the prejudice that humans have. As always, these solutions are as complex as the questions they are answering. I’m grateful for the thought we have put into ethics in autonomous vehicles so far, and am hopeful for the conversations that we will continue to have regarding it in the future.