Ethics in Artificial Intelligence

Ethical dilemmas are a topic that everyone will have some experience with moving forward. This could be ethics involving schoolwork and plagiarism or ethics within a workplace. Ethics is a topic that covers many different places within a person’s lifetime. The question is, can we teach ethics to nonhumans? Ethics is something that we learn through our life, if we think of artificial intelligence and the choices of their actions, can we teach them how to solve ethical dilemmas?

Humans have always worked on methods to make their life and work easier to accomplish. That is why when it comes to solving a problem the saying goes that giving the problem to a lazy person, they will find a way to solve the problem most efficiently. Artificial intelligence was created as a means of helping advance technology and make our job easier. The jobs that are implementing artificial intelligence are also ones that have potential ethical dilemmas.

The trolley problem is an ethical dilemma that we use in school to help explain how there is no black and white answer with ethics. The problem shows two outcomes, and we must determine which would be the lesser evil. The implementation of artificial intelligence within cars, self-driving cars, has brought up the need for a new modern trolley problem. The creation of the moral machine website gives people a way to judge ethical dilemmas of the trolley problem with self-driving cars. The information gathered from the user input will then be averaged and implemented into self-driving cars. The site has spread through the use of media and helps people understand the situations that the artificial intelligence could find themselves and what outcomes would be best.

When a person is making an ethical decision, they are including their relations to the people impacted and how they will be impacted based on their decisions. We have a sense of self preservation which is why when given a choice we will choose an outcome that is less likely to cause us harm or those we care about. The artificial intelligence has yet to implement this within their systems, so the decisions are less about the people themselves and the types of people. We could have two different people where one will be killed no matter the choice. One person is overweight while the other is fit, the collective data shows that the fit person is more likely to be saved based on user input. We only know the outer characteristics and factors such as age and height. What we could find out after is the fit person is wanted for murder within six states while the overweight person is a successful surgeon who has saved countless lives. We may regret the choice after, but we cannot change our decision when it comes to life and death, there is no do over and the ai needs to be able to make decisions even without all the facts.

In the paper by John-Stewart Gordon, he discusses the trouble there is with programming ethics into machines when the programmers may not share the same views or understand ethics themselves. While there is artificial intelligence that learn and grow, we must be able to understand ethics and the dilemmas they cover to be able to teach the ai about them. Humans can learn ethical solutions through their lifetime and comprehend the ethical reasoning behind it. For example, think of how to explain to a child that killing is morally wrong. The child may not comprehend the why as it is a societal rule that we do not kill unless it is in self-defense. Kids may be able to comprehend it with being told that killing someone is hurting them and they can comprehend pain as they have fallen and gotten bruises before. They can have a level of understanding pain and take that idea into killing. Unless it is programmed, most artificial intelligence will not be able to comprehend pain so they would have trouble relaying that to the societal rule of no killing. Adults understand ethics and morals at the base level of “just because”, meaning there is no in-depth reason we just do what we do. To be able to teach ethics to the ai, we need a clearer understanding of how ethics works, or we would be unable to teach them how to handle ethical dilemmas.

Gordon mentions two main types of problems when programming ethics, the “rookie mistake”, where there is a misunderstanding of how to handle ethical problems. The second problem is being unable to agree on an ethical solution to the perceived problem. A note to remember is that there is no single solution to an ethical problem, each person could have a different idea of what is the “correct” choice to make. Ethics takes place in a gray area so the solution to the problem depends on the person and their background.

The paper written by Ron Iphofen and Mihalis Kritikos discusses how artificial intelligence could not follow ethics because they are just using algorithms and their actions are based on their code. The point is true to an extent, but my focus is how the artificial intelligence can learn by being shown different ethical dilemmas and how others account for them. An ai may not be able to determine ethical solutions on their own which is why it falls on their creators to help guide them and teach them morals. Earlier I mentioned how a child may not understand the morals behind taking another’s life. Children are not born with an understanding of “right or wrong”, as they grow, they learn the morals and ethics of society and the area they are. Some places believe that it is respectful to burn the dead while others find that disrespectful and desire the dead buried. By looking at artificial intelligence as just zeros and ones, we take away the need to teach them ethics as we use the excuse that a machine can not comprehend ethical dilemmas. In car commercials we see the person backing up and the car forcing them to stop as a person runs behind them. Technology has advanced to the point that we can start seeing ethical dilemmas that technology will come across and that they will be able to determine the steps once at this stage. The paper about essential factors to create ai for the social good helps explain how we can help artificial intelligence learn ethical reasoning. The writers for this paper give seven essential factors that are used when incorporating ethics into ai.

The first factor is falsifiable which they use to help give the ai a level of trustworthiness. To be able to make a decision when there is an ethical dilemma, the one who is making the decision needs to have a level of trust put into them for their choice to be taken into effect. The second factor is the need for safeguards to avoid being manipulated by others. For this, a person in their place could be swayed to choose one decision instead of another through others like bribery or threats. The third is receiver-contextualized intervention which means that the ai needs to be able to intervene when needed and in a way that respects a person’s autonomy. The fourth factor is being able to explain the third and be transparent with the purpose of their actions. The fifth covers privacy and that they must respect the privacy of others and get consent for data. The sixth is situational fairness which makes sure that the choices are not biased based on the people being affected. The final factor is that the choices they make need to be able to align with the choices humans would make.

The factors are a well thought out way of giving artificial intelligence a limited but easier to implement way of ethical reasoning. When it comes to ethical reasoning humans will make decisions that will help them. The idea with giving artificial intelligence ethical reasoning is that they will use their reasoning to help humanity. Our decisions are affected by our self-preservation which we are not giving to the artificial intelligence. If given the choice between the artificial intelligence being destroyed or a person dying, the person will choose themselves to live each time as they see the artificial intelligence as a machine that can just be rebuilt even if the initial code is lost. Giving the AI the same ethical dilemma, we expect them to choose to save the person instead of themselves because they were created by humans to help humans. To give them the ability to make ethical choices by taking away the need for self-preservation, we are not letting them make a full ethical decision when the choices affect themselves. The trolley problem has a person choosing whether to switch the tracks. While they are the one with the switch, they are not being put into harm with the choices. Looking at the trolley problem of the automated car, if the AI only exists in that car, then they are putting themselves into the problem with whether or not they survive the outcome.

The ethical situations that we are put in help us grow as people and learn about the choices we would make. As we advance in technology, the amount of artificial intelligence will increase. We can already see how ethical dilemmas will appear in relation to the positions that the AI are put into. To be able to handle these situations we need to be able to teach AI about ethics and ethical reasoning so they can make decisions instead of the default decision of inaction. The action of inaction is a choice that a person can make, without knowing about the other options, an artificial intelligence will choose this path each time.

References

Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to Design AI for Social Good: Seven Essential Factors. *Science & Engineering Ethics*, *26*(3), 1771–1796. https://doi-org.links.franklin.edu/10.1007/s11948-020-00213-5

Gordon, J.-S. (2020). Building Moral Robots: Ethical Pitfalls and Challenges. *Science & Engineering Ethics*, *26*(1), 141–157. https://doi-org.links.franklin.edu/10.1007/s11948-019-00084-5

Iphofen, R., & Kritikos, M. (2021). Regulating artificial intelligence and robotics: ethics by design in a digital society. *Contemporary Social Science*, *16*(2), 170–184. https://doi-org.links.franklin.edu/10.1080/21582041.2018.1563803

*Moral machine*. Moral Machine. (n.d.). Retrieved October 24, 2021, from https://www.moralmachine.net/.