CSE487: Cybersecurity, Law and Ethics
[Summer 2022]
Section: 03

Ethical Decision Making in the field of IT/CSE and criticism and justification of actions/stance as per ethical frameworks

Project Report

Submitted to:

Rashedul Amin Tuhin

Senior Lecturer

Department of Computer Science & Engineering

East West University

Submitted by:

Student ID	Student Name
2019-1-60-068	Md. Shahadat Anik Sheikh
2019-1-60-036	Hasib Ar Rafiul Fahim
2018-3-60-088	Rashik Buksh Rafsan

Our Scenario

Suppose there is a fire outbreak in a house. Now, to rescue from the scene, a CFS2H robot (AI-controlled robot used in the fire rescue team) is sent to get victims out of danger. The robot enters a stage where at any moment can be a blast; the robot sees two possible victims that need to be recused: a child and an old man. Suddenly a cat enters the scene. The robot can carry out at a time only two victims.

Ethical Dilemma

Now, if the robot saves the children and the old man, the cat will die. If it chooses to save the old man and the cat, the child will die; similarly, if it chooses to save the child and the cat, the old man will die. What would be the decision of the Rescue robot?

Brainstorming Phase

Stakeholders: In this scenario, we have four stakeholders directly affected by our ethical dilemma. They are the Rescue Robot, the child, the old man and the cat. A few stakeholders have indirectly affected: the robot's programmer, the fire force, and the victim's families.

Risks and consequences: We see three risks have arisen from our scenario. Firstly, the old man would die if the robot chose to save the child and the cat. Secondly, the child will die if it chooses to save the cat and the old man. Lastly, the innocent cat will die if it chooses to save the child and the old man. These are the three risks and consequences.

Who gets each benefit: In this scenario, we considered three different possible events, and if one event occurs, then the other two events will not occur, so those events will not be benefited. For instance, if the robot ignores to save the cat, the child and the old man will be saved, and the child and the old man will be benefited. If the robot chooses not to save the child, the old man and the cat will be saved, and the old man and the cat will be benefited. If the robot chooses not to save the old man, the cat and the child will be saved, and the cat and the child will be benefited.

Analysis Phase

Responsibilities of the decision maker: The autonomous rescue robot in our scenario is the decision-maker. In reality, the person who programmed this system to operate autonomously is the one who ultimately makes decisions. The programmer must make judgments regarding the system, such as whom to save first, what to do in the event of a fire, how to evacuate people safely, and many more. He must consider some obligations while adhering to general and professional ethics while making judgments of this nature. As per general ethics, the programmer shall act consistently in the public interest. He should consider the minority of stakeholders rather than being prejudiced toward a subset of them. He ought to plan his designs so that they do not hurt society or the welfare of people. He must think of all system participants as stakeholders. He must be trustworthy, fair, and unable to discriminate against anyone. According to professional ethics, he must welcome feedback and criticism from stakeholders

and respond with a thorough analysis of the systems and impacts that considers all potential dangers. He should also create and put into place a system that is both reliable and secure.

The rights of stakeholders: In the given scenario, three stakeholders are directly affected in order of identifying rights: A child, an old man and an innocent cat. There are three consequences: (1) If the robot saves the old man and the child, the cat will die, (2) If it saves the cat and the child, the old man will die, and (3) If it saves the cat and the old man, the child will die. In this situation, every stakeholder has the right to be alive. It is their negative right. Sometimes we can think that the robot can ignore the cat and save the two humans, but we cannot do that because it is an innocent animal with little intelligence. We cannot take them as a minority. Also, we cannot kill other humans by following the rules of saving whom the robot saw first. So, if anyone is killed, it is a violation of their negative rights.

Impact of the action options on the stakeholders: From our scenario, the first action will be, if the robot saves the child and the old man, then the negative impact of this action will be that our stakeholder, the cat, will die but the positive impact will be, the other stakeholders, the child and the old man will be safe. In the second act, if the robot saves the cat and the child, the negative impact of this action will be that the old man will die, but the positive impact will be that the child and the cat will be safe. Furthermore, in the last action, if the robot saves the old man and the cat, then the negative impact of this action will be that the child will die, but the positive impact will be that the cat and the old man will be safe in this case.

Consequences, risks, benefits, harms, and costs for each action: Three risks arise in our scenario. The first risk will be if the robot chooses to save whom it saw first, here the cat and the old man, the consequence of the action would be that the child will die. So, the harm in this action will be one child's life, and the cost of this action will be more than any other action as the child is expected to add more utility to society. The second risk will be if the robot chooses to save the cat and the children; the consequence of this action would be the old man dying. So, the harm in this action will be that an old man will die. Hence, one old man's death will be less costly than a child's death if we consider adding the utility to society, so the cost of this action will be less than the previous action. The last risk will be that if the robot chooses to prioritize human life, the consequence of this action will be that the cat will die. So, the harm in this action will be that an innocent animal will die with some emotional value. We must agree that animal lives are essential, but one animal's death will be less costly than one human's death because human lives are more valuable than animal lives, so the cost of this action will be less than previous actions.

Kant's, Mill's, and Rawls' Approaches

Immanuel Kant's "Categorical Imperative" theory: In this theory, Kant says that one should always respect the humanity of others and that one should only act following rules that could hold for everyone. Kant believes in Absolute moral rule. Absolute moral rules mean the rules that everyone must follow in order to act morally. They are the rules that hold under any circumstances universally. Such as we should never lie or never kill innocent people. In our

scenario, if the robot chooses to save the child and the old man, the cat will die, and if the robot chooses to save the cat and the child, the old man will die, and if the robot chooses to save the cat and old man, then the child will die. In every case in our scenario, an accident will occur and kill someone, but Kantianism does not allow killing someone to save others. The reason is that the decision to kill another rational being is always immoral in the eyes of Kantian ethicists. According to Kant's theory, no decision can be made; in our scenario, at least one life will be lost.

Mill's "Utilitarian Theory" theory: In this theory, Mill focuses on the results or consequences of our actions which produces the greatest good for the most significant number. In our scenario's first case, the old man will die, but the child and the cat will be safe. In the second case, the child will die, but the old man and the cat will be safe. Furthermore, the cat will die in the last case, but the child and the old man will be safe. Hence, we know utilitarianism would favor whatever option in which the more significant numbers of lives are saved. As we see, the last case saves more significant numbers of lives. So, the Utilitarian kills the cat and saves two human lives.

John Rawls's "Theory of Justice": In this theory, Rawls says everyone in the society holds equal fundamental rights whether someone is from a minority group, but it does not matter. They also have equal rights, and no one should deprive them. Based on Utilitarian theory, we agreed to kill the cat and save the child's and the old man's lives because it produces greater good for more significant numbers but based on the theory of justice, we cannot kill the cat. Because the cat is a minority here and its life is less valuable than two human lives, Rawls's "Theory of Justice" does not allow killing the cat. According to Rawl's theory, the robot will save the cat and the old man.

Categorize each potential action: "Ethics" refers to moral ideals. They have an impact on how a person makes choices and lives their lives. "Ethically obligatory" means that it is ethically essential to do the right thing and not to do the wrong thing. We have three options in our scenery, and none satisfy the obligatory ethical class. What is "Ethically accepted" refers to what the majority approves. Moreover, "Ethically prohibited" refers to expressly forbidden activity under an Ethics Code. In our dilemma, we have to kill the cat; if not, the robot must choose between a child or an old man to save from the fire scene. Only these three possibilities are available to us, and all of them are ethically prohibited.

Decision Phase

In the given dilemma, we have only three cases to choose. In the first case, the child will die. In the second case, the old man will die. Furthermore, in the last case, the cat will die. We can see that none of the options is acceptable from the ethics perspective. However, we must choose one out of them as the robot must need to save life. In this case, we will prefer the third option, where the cat will die. Because if the robot saves the cat and the old man, the child will die, although the child has the highest potential to gain more utility for society.

On the other hand, if the robot saves the cat and the child, the old man will die and, in the last case, the cat will die. As per Kant's theory, the robot must save everyone, but this is impossible in our scenario because our scenario does not have other options. Also, the Theory of Justice by Rawls is not practical for our case. Because anyone innocent or minor cannot be deprived of justice and cannot be a victim of a situation. So, we have a chance to choose an option by following "Utilitarian theory". Because here we choose an option which is less harmful than others. If the robot saves the cat, the old man or the child will die, whose life is more valuable than the cat. On the other hand, if the robot decides to save two humans' lives instead of a cat's, it will damage less than any other case. So as per our consideration of all the possible incidents, deciding to save two human life instead of an animal's life is a better option for the robot than any other option.