AIR UNIVERSITY | AIR FORCE CYBER COLLEGE

CASE #P02-2109R0
TEACHING NOTES

Driving Artificial Intelligence

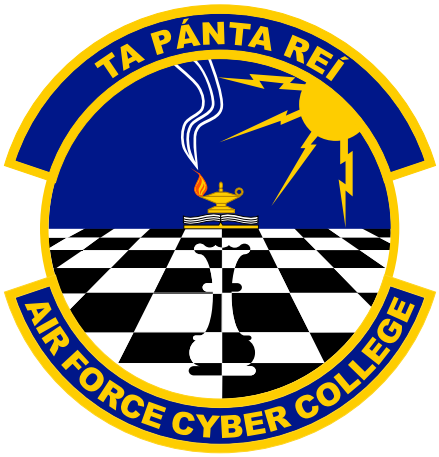
Dr. Jim Chen
National Defense University

CYBER COLLEGE CASE STUDY SERIES

Air Force Cyber College

Col David B. Bosko, USAF, Commandant

Dr. Karen Guttieri, Academic Dean



Please send inquiries or comments to:

Ms. Kay Adams
Air Force Cyber College
60 W. Shumacher Ave., Bldg 803
Maxwell AFB AL 36112

Tel: (334) 953-5591
E-mail: lillian.adams.1@au.af.edu

Driving Artificial Intelligence[1]

ABSTRACT

This is a hypothetical decision-forcing case. It explores a realistically simulated decision dilemma that both artificial intelligence (AI) systems and human beings who design or operate AI systems have to deal with at certain points. It illustrates the complexity in decision-making under extreme circumstances. The case not only has a significant impact upon the AI world as more and more AI systems are rendered with the responsibility of decision-making but also has a policy-level implication for any environment in which a tough choice has to be made with no desirable options available. It challenges students with several aspects, including but not limited to legal, ethical, policy, operational, and technical aspects. The frustration and uneasiness generated in handling this case may motivate a student to search for outside-the-box solutions, thus promoting paradigm-shift thinking.

TARGET AUDIENCE

This case exercise is targeted for the Master of Science students in the Joint Professional Military Education (JPME) II Program. This case can be used in courses on leadership, ethics, decision-making, cyberspace, national security, emerging technologies, etc. It can also be used in other programs whenever it is deemed appropriate.

LEARNING OBJECTIVES

1. Identify ethical, legal, policy, operational, and technical dilemmas in AI technology development
2. Evaluate options/solutions available and their consequences
3. Assess decision-making processes in AI systems
4. Create outside-the-box solutions to solve or dissolve dilemmas
5. Formulate strategies and policy recommendations for dilemmas in which all options available are undesirable

THE PURPOSE OF THE CASE

The purpose of this case as a teaching tool is to place students at the center of a difficult decision concerning saving human lives in vehicles on roads as all the options currently available are not desirable. Students are forced to consider all relevant perspectives, especially the ethical perspective, before making their decisions. Frustration caused by the uncomfortable choice that students select can lead them to paradigm-shift thinking and new ways of solving or dissolving problems. This exercise will also help students to make ethical decisions in other operations or in designing AI systems when human lives are involved.

1. This case study scenario is fictitious but realistically plausible. The views presented are those of the author and do not necessarily represent the views of Department of Defense or its components.

BACKGROUND

AI-enabled self-driving vehicles have been tested on roads now. Soon, roads will be shared by self-driving vehicles and non-self-driving vehicles, i.e., vehicles driven by human beings.[2] Hence, AI-enabled self-driving vehicles should be intelligent enough to make decisions to avoid collisions. Many researchers maintain that ethical principles should be incorporated into AI systems. Articles listed in references from Anderson and Anderson as well as Dignum are representatives of this view; McLaren also proposes the computational models of ethical reasoning. However, no matter how intelligent AI systems are, they will run into dilemmas, just like human beings. This hypothetical case explores responses and their consequences during an extremely tough environment when a collision seems unavoidable. Decisions have to be made with respect to saving lives as well as minimizing damages.

Readings listed in references, especially the articles by Nowak, Wamsley, Anderson and Anderson, may help students to better understand the background.

ACTORS

The actors in this case include but are not limited to advisors, designers, manufacturers, owners, riders of self-driving vehicles, riders of non-self-driving vehicles, judges, members of consumer advocacy groups, journalists from the press/media, AI proponents, and AI opponents. Each of those roles is described below:

- **Advisor**—provides advice in dealing with the dilemma presented in this case so that the emergency response system of self-driving cars can be further improved
- **Designer**—seeks help to further improve the current design, which obviously has limitations
- **Manufacturer**—tries to elicit the improved design for the emergency response system of self-driving cars
- **Rider of the self-driving vehicle**—wants to survive the accident in this dilemma
- **Rider of the non-self-driving vehicle**—wants to survive the accident in this dilemma
- **Judge**—wants to be fair to each one involved in this accident
- **Member of the consumer advocacy group**—tries to hold designers and manufacturers accountable for the inappropriate design
- **Journalist**—intends to accurately report the dilemma
- **AI proponent**—wants to minimize the impact of this dilemma in the application of AI technology in society

2. Alex Kopestinsky, "25 Astonishing Self-Driving Car Statistics for 2021," ***Policy Advice*** (29 April 2021), https://policyadvice.net/insurance/insights/self-driving-car-statistics.

- **AI opponent**—wants to maximize the impact of this dilemma in the application of AI technology in society

ROLE-PLAYING GAME PLAN

This case can be used to support class discussions via the integration of the following two methods: the role-playing method and the multi-round method. (1) In each round, each student is asked to take one role listed above, select an option, and provide reasons for his or her selection. This way, different perspectives from different actors can be captured. (2) In addition, multiple rounds are needed. Each student is asked to play a different role in each round. For instance, in the first round, Student X plays the role of an advisor. In the second round, Student X plays the role of a rider of the self-driving vehicle. In the third round, Student X plays the role of a rider of the non-self-driving vehicle. In the fourth round, Student X plays the role of a manufacturer of the self-driving vehicle. In the fifth round, Student X plays the role of a judge. The purpose of using multiple rounds is to find out different viewpoints from the same student who plays different roles. When all these different viewpoints are collected from the same player and then are compared and contrasted with the viewpoints of other players, relatively objective viewpoints are expected to be revealed. After enough data are collected from many groups of students, statistically significant trends and deviations can thus be recognized. As a result, ethical and responsible decision models as well as relevant quick-and-decisive courses of action can be figured out to reduce damage and to intervene in a potentially deadly situation.

DISCUSSION

The dilemmas in this case consist of the original case dilemma and its corresponding dilemmas in relevant aspects. The relevant aspects include but are not limited to legal, ethical, policy, and technical aspects. Each dilemma is discussed below. Suggested questions for the discussion of these dilemmas are also provided. To answer these questions requires the use of logical reasoning and knowledge of risk management, because what is sought is relatively the most gains and the least losses. The exploration of these dilemmas not only helps students improve their analytical and critical thinking skills but also helps them develop problem-solving skills in a holistic, multi-domain, multi-dimensional, and paradigm-shifting way.

Original Case Dilemma

At the surface level, or at the operational level, the dilemma in this case lies in the selection of one of the three options available: (1) turning slightly right to save the lives of the two people in the self-driving vehicle but take the risk of having the five people in the non-self-driving vehicle injured; (2) driving straightforward to save the lives of the five people in the non-self-driving vehicle but take the risk of having the two people in the self-driving vehicle injured; (3) turning slightly

left to take the risk of having the seven people in both vehicles injured. Beyond doubt, none of the options are desirable. However, Option 3 is the worst one, and if dismissed, the remaining dilemma lies in the choice of saving the lives of the two people in the self-driving vehicle or saving the lives of the five people in another vehicle. Before the tough choice is made, several factors and aspects should be considered. In terms of initial design, the following questions need to be addressed:

- Should the life of a rider in a self-driving vehicle be given the highest preference?
- If not, how can a rider trust a decision made by a self-driving vehicle in an emergency?

In terms of real-time response, the following questions need to be answered:

- Should a self-driving vehicle be empowered to conduct a quick risk assessment and calculate the damage cost for each option before making a decision?
- Should a rider in the self-driving vehicle be allowed to override a decision made by a self-driving vehicle?

To initiate discussions about this dilemma, the following questions may be asked:

- What are the benefits of self-driving vehicles?
- What are the limitations of self-driving vehicles?
- What dilemmas can you identify in this case?
- Which option are you going to select given such a case?
- What are your reasons for your selection?
- Is your selection influenced by the role that you play or is it purely based on the ethical principles that you believe?

After each student answers these questions, another set of questions may be asked. This set of questions can lead to discussion about the analysis of various options.

- Who selected Option 1?
- What is in common with respect to the students who have selected Option 1?
- Which ethical principle, if any, is applied in selecting Option 1?
- Who selected Option 2?
- What is in common with respect to the students who selected Option 2?
- Which ethical principle, if any, is applied in selecting Option 2?
- Who selected Option 3?
- Why is Option 3 seldom selected?
- What can we learn from the selection and application of ethical principles in this case?

In addition to the obvious dilemma in the operational aspect, this case presents dilemmas in other relevant aspects. This is discussed in detail below.

Legal Dilemma

In the legal aspect, the dilemma is caused by ambiguity and uncertainty. It is not clear who should be held liable for any potential human injury, loss of life, or property damage in such a case. Ironically enough, one thing is obvious. It is not possible to hold a vehicle liable. The dilemma in this aspect lies in the selection of the individual(s) or group(s) who should be held liable: (1) the designer(s); (2) the manufacturer(s); (3) the owner(s); (4) the rider(s) of the self-driving vehicle; (5) some of the above; or (6) all of the above. In discussions, the following questions related to the legal dilemma may be asked:

- If property damage or human injury or casualty occur in such a case, who should be held liable?
- Which laws, regulations, rules, and guidance should designers, engineers, manufacturers, owners, and riders of self-driving vehicles follow?
- What new law, regulations, rules, and guidance do you recommend in order to handle a dilemma like this?

Ethical Dilemma

In the ethical aspect, the dilemma lies in the preference assignment in applying ethical principles. This is reflected in the order in which ethical principles are applied. The preferred ethical principles always have a heavy weight in decision-making. They are always applied first. Therefore, of the same list of ethical principles, the outcomes may differ with different orders in application. Specifically, the dilemma exists in the selection of the following options: (1) saving the lives of the people in the self-driving vehicle first; and (2) saving the lives of the people in the non-self-driving vehicle first.

There is another level of uncertainty in this aspect. It is not clear whether it is ethical to assign preference in such a case. It is also not clear who has the authority to assign preference even if preference assignment is allowed. What is more, it is not clear whether one is able to decide preference ethically. In discussions, the following questions related to the ethical dilemma may be asked:

- Is it ethical to have preference consideration in deciding who should be given a chance of survival? Why or why not?
- If preference consideration is allowed, what should be the deciding factors?
- Should the number of people be a deciding factor? Why or why not?
- If preference consideration is allowed, who should have the authority to assign preference?
- Can one decide preference ethically? If so, how?

Policy Dilemmas

In the policy aspect, the dilemmas lie in the adoption of AI systems in human society as well as the assignment of leadership roles in AI systems.

With respect to the former, the dilemma lies in the permission of self-driving vehicles on roads. It is known that there is an issue with the emergency response system. Under such a circumstance, the dilemma exists in the selection of the following options: (1) not allowing self-driving vehicles on roads; (2) conditionally allowing self-driving vehicles on roads; and (3) allowing self-driving vehicles on roads.

With respect to the latter, the dilemma lies in the assignment of leadership roles in AI systems. The dilemma exists in the selection of the following options: (1) having humans in charge; (2) having AI systems in charge; and (3) having both humans and AI systems in charge. In addition, it is not clear who has the authority to decide who is in charge.

In discussions, the following questions related to the policy dilemmas may be asked:

- If there is an issue with the emergency response system, should a self-driving vehicle be allowed to be on roads? Why or why not?
- Who do you think should be in charge of AI systems in normal situations? Humans? AI systems? Both humans and AI systems? What are your reasons?
- Who should be in charge at critical moments—humans, AI systems, or both humans and AI systems? Why?
- How can gains be maximized and losses be minimized for all parties involved? How can a win-win outcome be achieved?
- What policy would you recommend for an environment in which all the options available are undesirable?
- What should be the normal process of making a decision related to life-and-death of other people?
- What should be the normal process of making a decision related to saving the lives of a group of people at the cost of sacrificing the lives of another group of people?
- Who should have the authority of making such a decision?
- What kind of qualification should such a decision-maker possess?
- How should such a human or machine decision-maker be held accountable for a decision that he/she/it makes?
- Can a decision made be revoked? If yes, who should have the authority to revoke it? If no, how can checks and balances be supported?
- Should the riders of the self-driving vehicle be given the authority to take control in a devastating and catastrophic environment?

Technical Dilemma

In the technical aspect, the dilemma lies in the capability of the technology in a self-driving vehicle. Specifically, the current technology only supports a certain number of maneuvers of the vehicle. The dilemma exists in the selection of the following options with respect to vehicle maneuvers: (1) moving straightforward; (2) moving forward with slight right turn; (3) moving forward with slight left turn; (4) coming to a stop; and (5) moving backward. Indisputably, none of the options are desirable. Any of the first three maneuvers may lead to an accident. The last two maneuvers are just not available due to the technology currently utilized. Even if they are available, any of these two maneuvers may lead to an accident if there is a vehicle that is immediately behind the self-driving vehicle. In this sense, they are not good options either. In discussions, the following questions related to the technical dilemma may be asked:

- What is the most appropriate course of action if all the options available are either unavailable or undesirable?
- What are the limitations of the current design of self-driving vehicles?
- Are there other technical options that have not been explored yet? If yes, what are they?
- What should be considered in a new design to address this type of dilemma?

CONCLUSION

This decision-forcing case, which implements best practices for creating teaching cases provided by Franke, listed in references, enables an active learning exercise that can be conducted in multiple ways.[3] It can place students at the center of a difficult decision, force them to wrestle with frustration caused by the uncomfortable choice that they have to select, and lead them to innovative solutions. As a result, new knowledge can be constructed, and dilemmas can be tamed.

As shown above, this hypothetical case has its unique value in discussions because the dilemmas in it exist in multiple aspects, which include but are not limited to the legal, ethical, policy, operational, and technical aspects. Hence, this case can be used for discussions in different fields. For instance, JPME students may look into the dilemmas in all the aspects discussed above so that they may learn how to deal with challenges and make good decisions in a dilemma. AI system designers, engineers, and manufacturers may scrutinize the dilemmas in all the aspects so that they can design and build next-generation trustworthy AI systems, including ethical and responsible self-driving vehicles. Policy-makers may examine the dilemmas in the legal, ethical, and policy aspects so that they can come up with policies that promote trustworthy AI systems but discourage

3. Volker Franke, "Security by Contractor: Outsourcing in Peace and Stability Operations," in Case Study No.1, Complex Operations Case Studies Series, ed. Karen Guttieri (Washington DC: National Defense University Center for Complex Operations, 2010) 1-32.

unethical and irresponsible AI systems. In this sense, this case can help students, leaders, military commanders, AI system designers and engineers make tough but ethical and responsible decisions in a dilemma or during times of complexity, ambiguity, and uncertainty.

REFERENCES

Kopestinsky, Alex. "25 Astonishing Self-Driving Car Statistics for 2021." *Policy Advice* (29 April 2021). https://policyadvice.net/insurance/insights/self-driving-car-statistics/.

This article lists 25 astonishing self-driving statistics for 2021. One of them is that "[t]here are over 1,400 self-driving cars in the US". Another is that "55% of small businesses believe that they are going to have a fully autonomous fleet in the next two decades."

Anderson, Michael and Susan Anderson. "Machine Ethics: Creating an Ethical Intelligent Agent." *AI Magazine* 28, no. 4 (2007): 15-26.

This article argues for the necessity of creating "a machine that itself follows an ideal ethical principle or set of principles". It maintains that by so doing a machine can make good decisions about possible courses of action it can take.

Dignum, Virginia. "Responsible Artificial Intelligence: Designing AI for Human Values." *ITU Journal: ITU Discoveries*, special issue, no.1 (25 September 2017): 1-8.

This article holds that ethical principles should be incorporated in the design of AI systems in order to uphold human values. It proposes the accountability, responsibility, and transparency (ART) design principles for the development of AI systems.

McLaren, Bruce. "Computational Models of Ethical Reasoning: Challenges, Initial Steps, and Future Directions." *IEEE Intelligent Systems* 21, no. 4 (2006): 297-315.

This article argues that ethical reasoning is different from reasoning in more formalized domains, because in ethical reasoning, conditions or premises are not precise, are subject to interpretation, and may even have different meanings in different contexts. Hence, computational models of ethical reasoning can only be used as teaching aids.

Nowak, Peter. "The Ethical Dilemmas of Self-Driving Cars." *The Globe and Mail*, 2 February 2018. https://www.theglobeandmail.com/globe-drive/culture/technology/the-ethical-dilemmas-of-self-drivingcars/article37803470/.

This article shows the significance of understanding the challenges that AI behind the machine is facing. It calls for a debate over how self-driving cars should behave in situations they cannot anticipate.

Wamsley, Laurel. "Should Self-Driving Cars Have Ethics?" *NPR Technology*. 26 October 2018. https://www.npr.org/2018/10/26/660775910/should-self-driving-cars-have-ethics.

This article discusses the importance of having ethics in self-driving cars as these cars "will need to make split-second decisions to avoid endangering human lives--both inside and outside of the vehicles."

Franke, Volker. "Security by Contractor: Outsourcing in Peace and Stability Operations." In *Case Study No.1*, *Complex Operations Case Studies Series*, edited by Karen Guttieri, 1-32. Washington DC: National Defense University Center for Complex Operations, 2010.

This article serves as an example for case writing and case teaching. It demonstrates dilemmas as well as ways of using dilemmas in active learning. One method that it suggests is to place students at the center of a difficult decision and force them to wrestle with frustration caused by the uncomfortable choice that they have to select.

Please send inquiries or comments to:

Ms. Kay Adams
Air Force Cyber College
60 W. Shumacher Ave., Bldg 803
Maxwell AFB AL 36112

Tel: (334) 953-5591
E-mail: lillian.adams.1@au.af.edu