

Case Study 1: The Trolley Problem and Self-Driving Cars

The use of ethical AI frameworks surrounding the Trolley Problem, specifically in regards to AI autonomous driving vehicles and the decisions that must be pre-built into AI vehicles including the impact that certain decision trees will have on adoption of this technology as well as the liability that AI designers will have for design flaws when there is an accident.

In this case study, students are put in the position of being executives in an automobile company that seeks to manufacture self-driving cars and are asked whether they should allow their programmers to handle the so-called “trolley problem” and what the ramifications of doing so are from an ethical standpoint. In addition, they will explore whether self-driving cars should have the ability for those in the car to take over in order to respect the Principle of Responsibility from the Montreal Declaration.

Goals

- Introduce students to the “Trolley Problem,” which has become a key conversational topic for AI designers of autonomous vehicles
- Identify the issues from an ethical perspective with attempting to solve the Trolley Problem as well as whether they should be solving this problem
- Identify whether full autonomous driving with absolutely no override would violate the Montreal Declaration’s Principle of Responsibility
- Have students actually test their own ethical beliefs using the Moral Machine from MIT Labs

Tags

- Principle of Responsibility from the Montreal Declaration
- Consequentialist Ethics (both short- and long-term)*
- Deontological Ethics*

*Consequentialist, Deontological, and Virtue Ethics are major frameworks throughout this course and this module really needs to be placed either within the framework of a business ethics or philosophy course or these concepts need to be introduced prior to using this case study.

Assigned Readings

Renda, Andrea, Ethics, Algorithms and Self-Driving Cars – A CSI of the ‘Trolley Problem’ (January 17, 2018). CEPS Policy Insight - No 2018/02, January 2018 , Available at SSRN: <https://ssrn.com/abstract=3131522>

[Major legal changes needed for driverless car era \(bbc.com\)](#)

[Why self-driving cars still require a lot of human supervision : NPR](#)

[AI and machine learning has its own trolley problem debate | World Economic Forum \(weforum.org\)](#)

<https://www.turing.ac.uk/blog/ais-trolley-problem-problem>

[The folly of trolleys: Ethical challenges and autonomous vehicles | Brookings](#)

[The problem with the trolley problem \(qz.com\)](#)

[Autonomous Cars Don't Have a 'Trolley Problem' Problem | The Drive](#)

In addition to looking at the case study itself that will be developed, students will interact with the modules at:

<https://www.moralmachine.net/>

Students will engage in a debate on this topic and additionally will write a paper on this topic.

Implementation and Assessment

This topic will be covered in two days, within one week. Prior to the first class for the week, students are to complete the reading assignments. They will also come into class with a written statement answering the question, “Is it ethical to use AI to solve the trolley problem?”

After handing in their written statements, students are to break up into small groups on each of the three key philosophical questions:

1. Can self-driving cars be programmed to deal with the trolley problem and, to the degree that they can be, should they be? (drawing on the readings from the Brookings Institution and Quartz)
2. Who, if anyone, should be held accountable for deaths due to self-driving cars that encounter a Trolley Problem? (drawing on the readings from the BBC and NPR)
3. To what extent should we prioritize getting people into self-driving cars as opposed to how to deal with the Trolley Problem? (drawing on the reading from the Drive)

In between the first class and the second class, students are asked to use the Moral Machine to see how they would “solve” various Trolley Problems.

During the second class, students are to discuss what they learned about themselves with respect to their own moral frameworks. This class session is devoted to debating the question: Is there a fundamental issue with the Montreal Declaration’s Principle of Responsibility with respect to self-driving cars and the Trolley Problem?

Students will then do a short (2 to 4 page) paper answering the four questions that were detailed for themselves with reference to both the debate and class readings that will be due the following week. Note that students are effectively answering the question “Is it ethical to use AI to solve the trolley problem?” as part of their answer to the first question.