**Case Study 4**

Moral Machine

Teaching Resources

**Summary**

This unit is focused on getting students thinking more deeply about ethical speculation, unintended consequences, and value tradeoffs with algorithmic design choices. The unit is based on The Trolley Problem as it applies to self-driving cars, also known as The Moral Machine. Students are introduced to the original trolley problem and the dilemma for designing the action of self-driving cars in the event of an unavoidable collision. After the students become familiar with the problem through reading, [making their own self-driving car judgements](https://www.moralmachine.net/), and a class lecture / discussion - they are guided through a 2 part coding activity that allows them to play the role of the coder who designs and deploys their own moral machine algorithm for a self driving car.

**CS Topics That Are Covered in This Unit**

1. File I/O
   1. Reading from csv files:
      1. csv.reader
      2. write()
      3. read()
      4. readlines()
      5. splitlines()
      6. split()
      7. next()
   2. Writing to csv files
2. For loops
3. Calling / testing functions
4. Designing criteria for algorithms
5. Appending to lists

**Ethics Topics That Can Be Covered in This Unit**

1. Designing “fair” algorithms when there is no perfect solution
2. Value judgements and value tradeoffs
3. Ethical speculation
4. Responsibility and accountability of the coder / tech companies
5. The trolley problem

**Resources (Reading / Watching To Assign the Students)**

1. [Introduction to the Trolley Problem](https://www.youtube.com/watch?v=yg16u_bzjPE)
2. [The ethical dilemma of self-driving cars](https://www.youtube.com/watch?v=ixIoDYVfKA0) - Patrick Lin
3. [Moral Machine website](https://www.moralmachine.net/) (navigate to the “judge” tab)

**Discussion Questions**

1. If I was the conductor of the original trolley problem, what would I do?
   1. Save 1 person, let the trolley kill 5 people (inaction)
   2. Save 5 people, force the trolley to kill 1 person (action)
2. How do you feel about the future of self-driving cars?
3. Which of the following statements do you agree with?
   1. I think that self driving cars should not be on the road, there are too many risks
   2. I think that self-driving cars should be on the road, they are much safer than human drivers
   3. I think that self-driving cars shouldn’t be on the road, but it is inevitable that they will be soon
   4. I think that self-driving cars should be on the road but not for a long time
4. Which of the following statements do you agree with?
   1. The trolley problem should be considered when designing self-driving car technology
   2. The trolley problem has no relevance to self-driving car technology, they are entirely separate
5. Would you ever get into a self-driving car that might at some point prioritize the safety of a pedestrian or someone external from the vehicle?
6. Do you agree with Tesla’s trolley problem model? Or would you prefer Waymo’s model?
   1. Tesla’s model (crowd-souring other driver data, non-accountability)
   2. Waymo’s model (hit the smallest object always)
   3. Neither of the above