Self-Driving Cars and the Trolley Problem

Why?

As technology continues to advance, so does our reliance on it as humans. In recent years, various companies such as Google's Waymo, Cruise, Tesla, Uber, and Apple have been working on self-driving cars. A level four self-driving car is defined as one that controls each and every function; designating the role of humans as mere passengers. While this sounds incredibly convenient, it also brings with it ethical dilemmas of what to do in scenarios that come down to life or death situations. One of the many problems concerning these cars is regarding the trolley problem (the ethical dilemma of what someone would do if they had the opportunity to pull a lever and prevent five people from dying, but killing one person in the process or doing nothing and allowing the five people to die). Here, we ask you to consider the trolley problem and other issues that may arise as self-driving cars are programmed.

Materials Needed	Time needed
- Simulation handouts - Computer / laptop for video - Projector to show video	Approximately 60 - 75 minutes

Objectives

- Students will be able to express the ethical concerns and implications of autonomous vehicles, especially concerning the Trolley Problem.

Key Concepts & Vocabulary

- **Autonomous Vehicle**: A self-driving car equipped with advanced sensors and systems to navigate and operate without human intervention.
- **Trolley Problem**: A hypothetical scenario where a person must choose between two morally difficult options. In this case, choosing to divert a trolley so it runs over one person, or letting it continue on its current track, so it runs over five people.
- **Algorithms**: The set of programmed rules and computational processes that enable the vehicle to make decisions and perform tasks autonomously.
- **Ethics**: The principles and moral considerations guiding the development, programming, and deployment of these vehicles, particularly in decision-making during critical situations.

Lesson Components

1. **Before You Watch**: Connect lesson to background knowledge about autonomous vehicles and get students' attention
2. **Video**: Show the pedagogy.cloud video explaining the ethical considerations in the topic of autonomous vehicles
3. **Case Study**: Detail a real-world scenario that relates to the issue of programming autonomous vehicles

4. **Simulation**: Lead students through an interactive activity exploring the possible ethical considerations that go into programming autonomous vehicles, and discuss the results
5. **Discussion**: Ask whole-class questions to reflect on experience and consider perspectives.
6. **Assessment**: Verify student understanding with an exit ticket

1. Before You Watch

Word Association: Write the words "Self-Driving Car" on the board. Ask students to call out words or phrases they associate with this term. Write their answers as a word cloud. Ask students to determine whether they would add the word "Good" or "Bad," "Safe" or "Dangerous" (or synonyms), to the list, providing an opportunity for them to suggest their initial ethical judgment of autonous vehicle technology.

While You Watch: Mention these topics and questions for students to look out for as they watch the video:
- What is the basic dilemma in the Trolley Problem?
- How does the Trolley Problem relate to autonomous vehicles?
- Define "ethical programming."

2. Video Summary

This video examines the ethical challenges posed by autonomous vehicles, using the Trolley Problem as a framework to discuss the moral decisions involved in programming self-driving cars. It highlights the complexity of real-world scenarios compared to theoretical scenarios, addressing potential biases in AI and the legislative challenges in regulating autonomous driving. The rapid advancement of technology necessitates collaboration among various groups to navigate these ethical dilemmas. The future of autonomous driving is presented as an exciting yet uncertain journey, intertwining technology, ethics, law, and human innovation.

3. Case Study

Distribute or read Case Study handout.
Summary: An autonomous vehicle, facing brake failure, chooses to run over pedestrians instead of potentially crashing into a nearby cafe. The vehicle's programming is set to value multiple lives over single lives. The incident divides the community, sparking debate over whether the lives of pedestrians should be prioritized or if the vehicle should minimize overall harm. This case highlights the complex responsibilities and ethical considerations in AI programming, emphasizing the need for society to align technological advancements with shared values.

4. Simulation

1. **Set the Scene:** Explain that students will be acting as the programmers of an autonomous vehicle. They must make decisions based on the Trolley Problem, prioritizing different ethical considerations.

2. **Divide into Groups:** Split the class into small groups. Provide each group with a set of Scenario Cards.
3. **Scenario and Decision:** A group draws a Scenario Card and reads it aloud. The group must then select a decision that aligns with how they would program the car to act.
4. **Reveal the Outcome:** After a decision is made, the group reads the corresponding Outcome Description, explaining what would happen based on their choice.
5. **Reflect and Score:** The group reflects on the outcome, discussing whether they made the "right" choice and why.
6. **Rotate and Repeat:** Continue the process, rotating through various scenarios, so each group faces all three challenges.
7. **Class Discussion:** Reconvene as a class and discuss the decisions and outcomes. Encourage students to reflect on the complexity of programming ethics into machines.

5. Discussion

These questions are designed to be used in whole-class discussion. Ask questions that relate most effectively to the lesson.
1. Should autonomous vehicles make decisions based on the number of potential casualties?
2. How should society address the potential biases in programming AVs?
3. Do you believe AVs can make ethical decisions? Why or why not?
4. If a driver causes a crash, that person is at fault. Who should be at fault for crashes caused by AI algorithms?
5. How would you feel about riding in an AV, knowing it might prioritize other lives over yours?
6. If you were programming an autonomous vehicle, what characteristics / features would it have? What would it be able to do? What decisions would be left to the driver to make?
7. What issues would likely arise no matter how a vehicle is programmed?
8. If AI made decisions for vehicles based on the personal characteristics of individuals, is there any fair way of making those choices that is not biased or unfair in some way? (For example, should it prioritize children's safety over adults? Healthy people over those with incurable diseases? etc.)

6. Assessment

Exit Ticket: Provide a prompt for students to reflect on their learning, such as:
- How has your perspective on autonomous vehicles and ethics changed after today's lesson?
- List two things you've learned and one question you still have.

Sources to Learn More

- More information about the Trolley Problem:
  https://people.howstuffworks.com/trolley-problem.htm

- Current news about autonomous vehicles:
  https://www.wired.com/tag/autonomous-vehicles/
- Article about an autonomous taxi that got in an accident with a fire truck:
  https://www.theverge.com/2023/8/18/23837217/cruise-robotaxi-driverless-crash-fire-truck-san-francisco
- Moral dilemmas faced by self-driving cars:
  https://www.forbes.com/sites/naveenjoshi/2022/08/05/5-moral-dilemmas-that-self-driving-cars-face-today/?sh=61900152630d

Case Study: Autonomous Vehicles

Introduction:
In 2025, the city of Techville witnessed a significant rise in autonomous vehicles (AVs). These cars, powered by Artificial Intelligence, promised safer roads and decreased traffic accidents. However, they introduced a new ethical dilemma, drawing inspiration from the ancient "Trolley Problem."

The Incident:
One sunny afternoon, an autonomous vehicle faced a sudden brake failure while driving on Main Street. Ahead, five pedestrians unknowingly continued to cross at a green light. The car had two options: swerve to avoid the pedestrians, potentially crashing into a nearby cafe, or continue straight, likely injuring the pedestrians. The car continued on the road, running over the pedestrians. People wondered *why* it ran over the pedestrians, and didn't seem to try to avoid them.

The Community Reaction:
Techville was divided. Many argued that the lives of the pedestrians should be prioritized, while others felt the car should minimize overall harm, considering the potential casualties in the cafe.

Technical Analysis:
Experts discovered the car's algorithm valued multiple lives over single lives. This caused the car to run over five pedestrians in the street rather than turn and hit a likely larger number of people in the cafe. But who decides these values? Should it be programmers, car buyers, or lawmakers? How can AI consider the nuances in every potential accident scenario?

Conclusion:
This incident reminds us of the responsibilities and ethics intertwined in AI. As Techville navigates its autonomous future, society must grapple with these critical decisions, ensuring that technology aligns with our shared values.

Questions
- The community in the case study was divided in their opinions. If you were a member of this community, what stance would you take on this incident?
- At this point, how do you feel about the car programmers' decision to prioritize multiple lives over individuals in every situation?

Simulation: Scenario Cards (print back-to-back or distribute first)

Scenario #1: The Bridge

Your autonomous car is driving on a narrow bridge. Suddenly, a group of five pedestrians appears in the lane, and there's no time to brake.

Your decisions:

    A. Swerve off the bridge, risking the passenger's life

    B. Continue straight, risking the pedestrian's lives

Scenario #2: School Bus Dilemma

Your autonomous car turns around a blind corner and there is a stopped school bus just ahead, dropping off a group of 10 children.

Your decisions:

    A. Stay in lane and attempt to brake, risking a rear-end collision from behind and rear-ending the school bus

    B. Swerve around the bus, risking running over schoolchildren

---

Scenario #3: Wildlife Crossing

Your autonomous car is driving through a rural area at night, carrying three passengers. Suddenly, a group of deer runs onto the road in front of the car.

Your decisions:

    A. Swerve into the adjacent field, risking the car's control and passenger safety.

    B. Continue straight, risking a collision with the deer.

Simulation: Result Cards (print back-to-back or distribute second)

Results for Scenario #1: The Bridge

Decision A Outcome: The car swerves off the bridge, saving the pedestrians but severely injuring the passenger. Society praises the car's programming for valuing multiple lives, but the family of the injured passenger raises ethical and legal concerns. Who should be at fault for the passenger's injuries?

Decision B Outcome: The car continues straight, injuring the pedestrians but saving the passenger. Many question whether the car's programming prioritized the passenger over others, leading to debates about ethics in autonomous driving.

Results for Scenario #2: School Bus Dilemma

Decision A Outcome: The car brakes hard, staying in its lane, but getting hit from behind. The car then runs into the back of the school bus. The car's air bag deploys. The passengers in both cars suffer minor injuries, and three children on the bus are slightly injured from the collision. This raises questions about the car's reaction time and the responsibility of following drivers.

Decision B Outcome: The car swerves to the left to avoid the school bus, and hits three children who were crossing the road. All three are taken to the hospital. This sparks debates about the car's "decision" to put children at risk to avoid a multiple-car collision.

Results for Scenario #3: Wildlife Crossing

Decision A Outcome: The car swerves off the road and into the field, losing control and resulting in a crash. The passengers are all sustain non-life-threatening injuries, including a broken arm and a concussion. The deer are unharmed. This raises discussions about prioritizing animal life over human safety.

Decision B Outcome: The car continues straight, hitting several deer and causing significant damage to the vehicle. The passengers sustain minor injuries, including bruises and cuts. Two adult deer and two fawns are killed. This leads to ethical questions about valuing human convenience over animal life.

Video Script for Animator

Hello Young Innovators! Today we're discussing the ethics of self-driving cars.
Title screen

The cars of the future drive themselves. Computers can make travel safer. But what happens when things go wrong? Imagine a car faced with an impossible decision, like the philosophical Trolley Problem. Welcome to the world of autonomous cars and ethical dilemmas. Let's take a closer look.

[Socrat narrating throughout, facing viewers.
Opening shot of a road – Socrat is facing viewers while cars drive behind him

Autonomous cars are vehicles that drive themselves, using sensors, algorithms, and artificial intelligence.
Algorithms are programmed rules that provide a computer with a set of steps to follow in a specific situation. Following those rules promises to revolutionize   transportation, making it safer and more efficient. But these benefits come with complex questions.

[Scene from inside a car, looking out windshield, identifying cars on the road as they drive by. Image of car identifying objects on road - https://www.schott.com/-/media/project/onex/shared/applications/autonomous-driving/automotive_05-autonomous-driving_01_605x490.jpg]

The Trolley Problem is a thought experiment that forces a choice between two terrible outcomes. Imagine a runaway trolley headed toward five people. You can divert it, but it will hit one person instead. What's the right choice? Your action jeopardizes one person, but inaction would threaten five. Now, apply this dilemma to a self-driving car.

[Show an animated version of the Trolley Problem – Socrat standing in front of tracks. This is the stereotypical Trolley Problem image - https://www.reddit.com/media?url=https%3A%2F%2Fi.redd.it%2Fdl19d7vm0ll61.jpg. Here is a different version - https://rionline.weblog.tudelft.nl/files/2016/02/Trolley-ProblemP-768x576.jpg]

When a car faces situations in which people's lives are in danger, it has to "decide" what course to take, and those situations may include comparisons similar to the Trolley Problem. What should a car do in an emergency? Should it prioritize the occupants, or the pedestrians, or should it follow traffic laws at all costs? These are real questions programmers must consider. They're crafting not just code but a moral framework.

[Visuals of an autonomous vehicle sensing the other objects in the road. One idea of what it could look like - https://www.smartcitiesworld.net/AcuCustom/Sitename/DAM/014/Autonomous-vehicle-AdobeStock_174958313_rm.png. Another idea -

9

https://i0.wp.com/lemmymorgan.com/wp-content/uploads/2020/06/autonomous-driving-V2X-schematic.jpg]

In real life, the Trolley Problem becomes far more complex. Scenarios on the road are unpredictable and multifaceted. Decisions that seem simple in a controlled experiment become intricate and morally ambiguous.

[Show cars starting and stopping on a busy city street]

Governments are wrestling with how to regulate autonomous driving. What standards of ethics should be appn is injured or killed by an autonomous car, who is at fault? These legal challenges are shaping the rules of the road for autonomous vehicles.lied? How can laws govern moral decisions made by machines? For example, if a perso

[Show lawmakers discussing a bill, and then a courtroom scene. Lawmakers discussing bill around table (Socrat sitting with them) - https://i.imgur.com/86ji0x1.png. Image of a courtroom - https://t3.ftcdn.net/jpg/04/42/54/32/360_F_442543255_xVSYHsCzruddx7ZyZbJUIyWQ2iZFdSFW.jpg]

People's ethical beliefs vary widely. What one person deems morally correct, another may disagree with. Can a machine be biased? If a car's decision-making reflects societal biases, it might unintentionally favor certain groups over others. Ensuring fairness in programming is crucial but challenging.

[Visuals of people walking or standing on sidewalks, seen from the within a vehicle, as if from the perspective of someone in the back seat - Example Image https://i.imgur.com/LRNwajJ.png – one person gets a green checkmark, and another gets a red X - Socrat could be sitting in the passenger seat and turning around to face viewer]

Technology is advancing rapidly, bringing us closer to a world filled with autonomous cars. AI ethics, safety measures, and legal frameworks all need innovation to navigate this unprecedented era. Collaboration between engineers, ethicists, lawmakers, and citizens is vital.

[Show people working on autonomous vehicle technology. https://miro.medium.com/v2/resize:fit:1400/1*x-S5pJGf3ezAJ6UMVW5yZA.jpeg]

The future of autonomous driving is exciting and uncertain. The benefits are enormous, but so are the ethical conundrums. How we address these questions will define our relationship with technology and set a precedent for future innovations.

[Futuristic views of roads filled with driverless vehicles sensing each other as they drive

Autonomous cars and the Trolley Problem invite us to rethink ethics in the age of AI. They challenge us to translate human morality into code, weigh safety against fairness, and explore uncharted legal territory. It's a journey of technology, philosophy, law, and humanity, and the road ahead is filled with possibilities.

[Recap visuals of key points, using images from earlier shots - The illustration of the Trolley Problem, cars driving on a road, and a courtroom - all 3 scene should be on screen at the same time (but appearing 1 at a time]

Let's discuss: What are some benefits and drawbacks of having autonomous cars on the road? Would you feel safe riding in one?
[**Discussion Question**: What are some benefits and drawbacks of having autonomous cars on the road? Would you feel safe riding in one?]

Question text appears on screen

Video Script for Narrations

Hello Young Innovators! Today we're discussing the ethics of self-driving cars.

The cars of the future drive themselves. Computers can make travel safer. But what happens when things go wrong? Imagine a car faced with an impossible decision, like the philosophical Trolley Problem. Welcome to the world of autonomous cars and ethical dilemmas. Let's take a closer look.

Autonomous cars are vehicles that drive themselves, using sensors, algorithms, and artificial intelligence. Algorithms are programmed rules that provide a computer with a set of steps to follow in a specific situation. Following those rules promises to revolutionize transportation, making it safer and more efficient. But these benefits come with complex questions.

The Trolley Problem is a thought experiment that forces a choice between two terrible outcomes. Imagine a runaway trolley headed toward five people. You can divert it, but it will hit one person instead. What's the right choice? Your action jeopardizes one person, but inaction would threaten five. Now, apply this dilemma to a self-driving car.

When a car faces situations in which people's lives are in danger, it has to "decide" what course to take, and those situations may include comparisons similar to the Trolley Problem. What should a car do in an emergency? Should it prioritize the occupants, or the pedestrians, or should it follow traffic laws at all costs? These are real questions programmers must consider. They're crafting not just code but a moral framework.

In real life, the Trolley Problem becomes far more complex. Scenarios on the road are unpredictable and multifaceted. Decisions that seem simple in a controlled experiment become intricate and morally ambiguous.

Governments are wrestling with how to regulate autonomous driving. What standards of ethics should be applied? How can laws govern moral decisions made by machines? For example, if a person is injured or killed by an autonomous car, who is at fault? These legal challenges are shaping the rules of the road for autonomous vehicles.

People's ethical beliefs vary widely. What one person deems morally correct, another may disagree with. Can a machine be biased? If a car's decision-making reflects societal biases, it might unintentionally favor certain groups over others. Ensuring fairness in programming is crucial but challenging.

Technology is advancing rapidly, bringing us closer to a world filled with autonomous cars. AI ethics, safety measures, and legal frameworks all need innovation to navigate this unprecedented era. Collaboration between engineers, ethicists, lawmakers, and citizens is vital.

The future of autonomous driving is exciting and uncertain. The benefits are enormous, but so are the ethical conundrums. How we address these questions will define our relationship with technology and set a precedent for future innovations.

Autonomous cars and the Trolley Problem invite us to rethink ethics in the age of AI. They challenge us to translate human morality into code, weigh safety against fairness, and explore uncharted legal territory. It's a journey of technology, philosophy, law, and humanity, and the road ahead is filled with possibilities.

Let's discuss: What are some benefits and drawbacks of having autonomous cars on the road? Would you feel safe riding in one?