How do we stop machines enslaving us?

This resource was developed by teachers within the Royal Society Schools Network

KS4

Lesson time: 1 hour

Introduction

Science fiction often paints a dismal picture of artificial intelligence (AI) – one in which the world is dominated and tyrannised by computers, making decisions of life or death. But is this likely to happen?

This lesson encourages students to consider the process of making decisions that may have an impact on the lives of others. It centres around the 'moral machine' – an emulator based on driverless cars. Starting from an individual basis, the students will build a picture of how their decisions compare to others with an eventual group decision.

This lesson has been planned for KS4 computer science, with references to the AQA and OCR specifications. However, it can be easily adapted for other key stages, or a PSHE lesson.

Learning objectives:

- Define machine learning and artificial intelligence.
- Describe examples of where machine learning is used.
- Justify the decisions you have made when on the moral machine, and consider different points of view.

Curriculum key words

Algorithms
Artificial Intelligence
Automation
Computer Systems
Machine Learning

Curriculum links

GCSE computer science:

- OCR 1.8
- AQA 3.7

National Computing Programmes of Study:

- Key Stage 4 Teaching Bullet point 3

Equipment needed

– projector;
– computers linked to Internet.

Resources

– robots overlords excel spreadsheet (download with this lesson plan)

How do we stop machines enslaving us?

Starter activity: how do we stop machines enslaving us?

(Approximately 5 minutes)

Watch two short clips from the recorded live stream of the event At the Manchester Science Festival (start 27:50 mins, end 27:55 mins and start 30:30 mins, end 31:40mins) to introduce the topic. The videos describe two experts views on how we stop machines enslaving us.

Discuss with the class "If I were a driverless car, what decisions might I make?". In this hypothetical reality, ask the students to think about the types of decisions they might need to make, and what options there might be.

Key points that should be elicited from the students should include:

- the answers would be yes/no or negative/positive;
- decisions would need to be made on 'smart' answers, rather than emotion based values; and
- decisions would likely be on a 'threat' basis, ie – there is a risk, take a course of action to avoid it.

Activity A: be a robot overlord

(Approximately 15 minutes)

On computers, students can log onto http://moralmachine.mit.edu/ to become a driverless car.

Show the students how to use the 'Moral Machine' – select 'Start Judging' and click on 'Show Description'. The panels give an overview of the composition of the passengers and the pedestrians.

When complete, students will be able to see how they, as robot overlords, make decisions regarding the values of the 9 variables:

1. Saving more lives
2. Protecting passengers
3. Upholding the law
4. Avoiding intervention
5. Gender preference
6. Species preference

7. Age preference

8. Fitness preference

9. Social value preference

Get the students to save their scores. These will be used later.

Activity B: examining our moral decisions

(Approximately 20 minutes)

As a class, or in groups, discuss the following questions:

- What are the results of your moral test?
- Now you have seen them would you make different decisions?
- Is there any bias in your decision making?
- Are your decisions based on judgement of common/most good? For example:
    - A baby has longer to live than an old person?
    - Crossing on a red light is breaking the rules so it is their own fault?

Watch a third short clip from the same video "At the Manchester Science Festival" (start 43:18 mins, end 44:00 mins). This clip describes the problem with human bias creaping into designed AI systems, specifically in the case of driverless cars.

Get the students to individually consider the following question: If systems are making decisions about right and wrong, is there a *correct* decision?

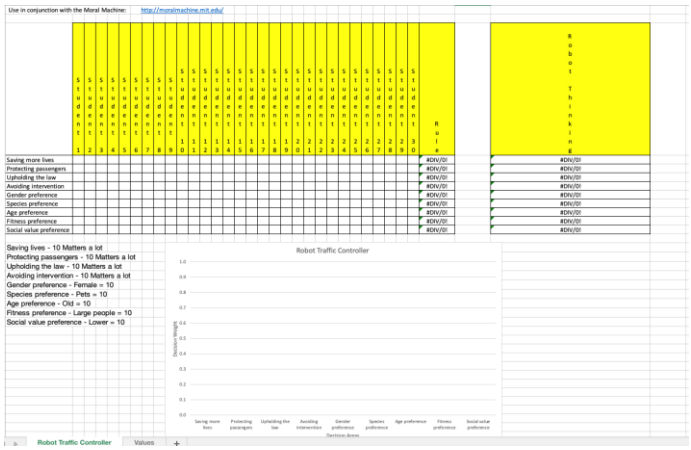
Once the students have had time to consider their own opinion, bring them together in pairs to compare results. Do they have similarities and differences?

As a class discuss if, as driverless cars, they as a group have shown any bias? For example, do they favour the young over the old? If you are homeless are you more likely to be run over by a driverless car?

Using the “Robot Overlords” spreadsheet, enter the data that the students saved from Activity A (a shared document e.g. Google Docs may be easier to manage).

What is the picture that emerges?

What does the graph in the spreadsheet show us?



Activity C: discussion

(Approximately 20 minutes)

Watch a short clip from the recorded live stream of the lecture The Practical Applications of AI (start 14:15 mins, end 15:13 mins). This clip describes issues around liability when decisions are made by computers.

So far, the idea of driverless cars is still under development but is gaining a foothold in society as the hardware and software advances. As a group discuss “Have robots taken over the decision-making process for us?”

Watch a fourth short clip from the recorded live stream of the event At the Manchester Science Festival (start 42:30 mins, end 44:55 mins). This clip describes more fully what implicit or unconscious bias is in a programmed system. From the video clip, ask the students to list the areas in which computer systems make decisions for them.

Answers could include

- The images that we see when we run a search for specific images
- How we recognize people in driverless cars
- Criminal justice decisions
- Insurance claim decisions
- Education decisions

As a class discuss “Do those areas mean that the enslavement has begun and that robots are directing what we do?”

Watch a short clip from the recorded live stream of the lecture The Challenges to Making machines Play Fair (start 5:50 mins, end 8:30 mins) which looks at the bias in computer systems in more detail.

In pairs, as students to create a 10-point policy document that designers of robot decision making systems should follow (consider modelling this to students). Ask the students to share with another pair, and agree from the list of 20 policy points what 10 points should stay.

Plenary

(Approximately 10 minutes)

As a class, ask the students a final question: how safe do we feel about the future and machines making decisions?

Extension question: can you see any other scary or exciting applications of machine learning e.g. your teacher being replaced with a computer teacher?

Opinion lines could be a great way of showing the spread of ideas (see below). You could get students to write their answers on post-its and stick to the board.

Safe Not safe