

Lesson 2: How AI uses data – Fairness and Bias

Lesson aims & overview

Aim

For children to understand that **AI systems need lots of good quality data to work well**. To understand that **AI can be biased** if the data does not represent everybody.

Overview

In this session the children will learn about **how AI systems use data to produce outcomes** and what **types of data** an AI system might use. They will explore the **risks** there are of AI systems producing **unfair, inaccurate or biased outcomes** if the data set is not large enough or properly representative, and consider the **importance of including diverse perspectives** – including those of children – when developing AI systems and making decisions about their use.

CfE Experiences & Outcomes

HWB 2-05a; LIT 2-07a;
LIT 2-09a; RME 2-02b;
SOC 2-16b; TCH 2-13a;
TCH 2-14a;
TCH 2-14b.

This lesson plan is aimed at Second Level but may be adaptable to First and Third.

Preparation for the lesson

The definitions cards resource will need to be printed and ideally cut out beforehand. Alternatively, the children can cut them out themselves – the words and definitions have not been arranged next to each other on the page so will not be giving the answers away. You will need 4 to 6 copies depending on the size of your class (enough for one per group of 3-5 children). Keep this resource for future sessions – the activity is included again in Lesson 6 and can also be used to refresh the children's memories at other points if desired.

The 6 images of cats and other animals will need to be printed and cut out (1 copy).

The main activity in this lesson focusses on how a type of AI system can be trained to recognise things in images. This article explains how 'captcha' authentication online also allows for huge amounts of data to be labelled, meaning it can then be used to train AI systems. You may find it useful as background reading: [techradar.com/news/captcha](https://www.techradar.com/news/captcha)

Lesson 2: How AI uses data – Fairness and Bias

Lesson Slides: [To access slides, click here.](#)

Resources needed:

- Slideshow (via link above - slides include embedded video with sound)
- Colouring pens or pencils.
- Blank paper cut into smaller pieces (roughly A5 or A6, you can use scrap paper for this) so you have at least 2 or 3 pieces of paper per child.
- Definitions cards - printed (see additional PDF files)
- Animal images for AI catflap – printed (see additional PDF files)
- Masking tape

Introduction



Estimated completion time: 20 minutes

Step 1 (10 MINUTES)

Remind the children what this series of lessons is about and read them the lesson aims and key questions.

Start the lesson by splitting the class into groups of 3-5 children and giving each group a set of the definitions cards. Explain that they will have 5 minutes to match the words or phrases to their definitions.

Regroup as a class and check everyone agrees on the definitions. Display the 'answers' slide, talk about any differences of opinion and support children to understand the correct definitions.

Step 2 (5 MINUTES)

Introduce today's session: explain to the children that today they will be finding out more about data and why it's so important for AI systems. We'll learn about how AI systems use data and what can go wrong if that data isn't good enough.

Show children the slide on data and read the text out to them: "AI systems are 'trained' with huge amounts of data. This means that people put lots of data into the AI system so that it can spot patterns in that data and use it to make new outcomes. The type of data an AI system needs depends on what it is designed to do."

Helpful notes:

Definitions cards

Explanations:

1. Alexa has to be taught to recognise human voices in order to respond to commands, so the AI system was given lots of recordings of human voices.
2. ChatGPT has been taught how to write like a human, by being given huge amounts of text written by people so it can learn how to form sentences that make sense. It then finds information from the internet and very quickly summarises it.
3. Facial recognition systems are trained with millions of images of human faces so they can understand what the features of a face are, and recognise what makes each face unique. For example, this allows FaceID to only accept **your** face when you go to unlock your device.

Lesson 2: How AI uses data – Fairness and Bias

Click through to the 'types of data' slide and read out the different types of data. Ask the children if they can name the examples of AI systems above. Then ask the children if they can match the types of data that an AI system would need to be trained on.

Answers (displayed on following slide):

Audio recordings – Alexa

Text – ChatGPT

Images – Facial Recognition

Step 3 (5 MINUTES)

Show the children the video explainer '*How does AI use data?*' (embedded in the slides).

Spend a couple of minutes talking about the children's thoughts or questions.

Display the slide on bias. Recap the definition with the children and ask them if they can quickly name a way a person might be biased. Discuss how some forms of bias are unfair but might not have really serious consequences (like bias against a sports team), but others can have real life negative impacts for individuals (like bias against a gender, religion, or skin colour). Ask the children if they've heard of any biases in AI systems – gather any suggestions if the children have any and make a note of these so they can be discussed at the end of the lesson.

Helpful notes:

This conversation will need to be approached sensitively and you may wish to remind your class that all children are likely to have experienced some kind of bias based on their age, and that some children will also have experienced bias on the basis of their skin colour, gender, religion, nationality, ability etc.

Warm-up activity: How do suggested videos work?



Estimated completion time: 10 minutes

Display the slide with a graphic of a YouTube/streaming site page – ask children to come up to the board and point out all the different places they could click. Ask the children to try to categorise these types of clicks (e.g. 'play' 'like' 'subscribe' 'search' 'select channel' etc.).

Helpful notes:

Clarifying question: Would you get data from other people if you were logged into your own account on your own device? – A: No, you'd just get suggestions for you individually.

Lesson 2: How AI uses data – Fairness and Bias

Explain to the children that an algorithm is a set of instructions written by a person, a bit like a recipe. It tells the computer programme what to do with the data it has been given. When we use video apps or sites like YouTube, each click creates a piece of data for the website's algorithm to use to then predict what else you might like. That's where suggested videos come from.

Now ask the children by show of hands to vote on which videos they would watch/like and which channels they might subscribe to if they were on this page. Make a note of which options receive the most votes. On the next slide, show the children three suggested videos and ask them which they think would have been suggested to their class as a whole, based on the data (how many of them would have clicked in certain places) from their whole class.

Main Activity – The AI Catflap



Estimated completion time:
45 minutes

Step 1 (5 MINUTES)

Give a brief explanation of how AI systems can recognise images:

One type of AI system can be trained to recognise what things look like.

Ask: **What type of data do you think it might need to be trained on?**

(answer: pictures)

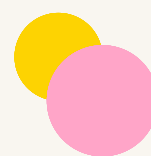
This kind of AI system could be given lots and lots of pictures which have been tagged with the word 'cat'. The more pictures of cats that are fed into it, the better it gets at automatically recognising what cats look like. It does this by recognising similarities in shapes and colours between all of the cat pictures.

Explain that we want to train an AI system for an automatic catflap which can recognise and let in cats but no other animals.

Ask: **What data do you think we need to put into this AI system to train it?** (answer: pictures of cats).

Helpful notes:

The code.org activity 'AI for Oceans' (linked to in the suggested resources below) is a great way for the children to learn more about how AI systems can be trained with data.



Lesson 2: How AI uses data – Fairness and Bias

Step 2 (10 MINUTES)

Explain that in this game, we're going to make our own data set by drawing cats. Give the children 10 minutes to draw cats on the A5/A6 pieces of paper – they can draw one each, or more, one per sheet. These can just be sketches – the aim is to get lots of data so detailed drawings are not required.



Helpful notes:

Prompt question: Do we want all the cats in our data set to look the same? Why not?

Answer: If an AI system's data set only contains one type of something, it won't recognise other types as belonging to that category. So in this case, if it only had pictures of grey cats it would not recognise other cats as it associates the label 'cat' with the colour grey.

The children may well at this point pick up on how this might impact people if an AI system was used for a similar purpose with humans. Be prepared for issues such as racism, ableism etc. to be raised and be sure to make time to explore these discussion points, either when they arise or returning to them at the end of the lesson. AI systems producing discriminatory outcomes based on biased data is a real and legitimate concern.

Step 3 (15 MINUTES)

Set up a 'catflap'. A rectangle marked out on the floor with masking tape will work well for this (roughly A4 size), or anything else which creates a rectangular space to place the pictures on.

Ask the children to lay out all their drawings on the floor next to it. Take a moment as a class to look at the pictures and talk about any similarities between the pictures that an AI system might notice as a pattern in the data.

Prompts:

Are there any common shapes or combinations of shapes?

Are there any common colours?

With the class, create a quick list of common features that can be seen in the pictures that the children have drawn.

i.e. If a majority of the class have included 2 eyes/4

legs/whiskers/a tail in their drawings, add that feature to the list.

If a majority haven't included a feature, it cannot be included on the list.

Helpful notes:

The aim of this activity is to consider how AI cannot take into account anything that is not in the data it has been given. It is important to point out to the children when they are using their own 'human intelligence' to make judgments on behalf of the catflap. For example, they might identify features like 'furry' or 'four legs' as reasons why the AI catflap would let through the cats in the pictures, when these features have not actually been represented in a majority of the children's drawings.

Lesson 2: How AI uses data – Fairness and Bias

Hand a child the first printed animal picture and ask them to place it in the 'catflap'.

Ask: *Do you think our AI system would think this animal is a cat based on the training data we have given it?*

Ask the children to look again at the pictures they drew and compare them to the printed picture.

Prompts:

Remember not to use your human intelligence – the AI system only knows what a cat is from your drawings. Use what you can see in the pictures, not what you know about animals.

Decide collectively what decision you think the AI catflap would make. Repeat this step with the other printed pictures. **It is likely (and preferable, from a learning perspective) that the AI system will have made some wrong decisions.**

Step 4 (15 MINUTES)

Once the children have decided what decisions they think the AI catflap would have made based on their training data (drawings), reflect on how well it has worked:

Ask: **How successful was the AI system based on the data you gave it? What do you think you could do to improve its accuracy?**

Give the children another 5 minutes to improve their training data – they may wish to edit existing drawings or add more drawings to address issues they identified from errors the AI system made.

Have another quick go at 'testing' the AI system with the printed images, inviting children to notice and comment on how specific improvements to the training data might have improved the likelihood of the AI system working correctly for specific images.

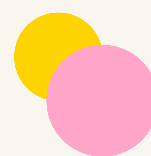
Helpful notes:

Similarly, they might use their own knowledge of features of the other animals, like size, to explain why they wouldn't be allowed in, but without including other things in their drawings to compare their cats to, the AI system would not be able to judge the size of a cat.

There are lots of ways these conversations might go. There are no strict rules about how the activity should be run – it is more important that these points of confusion or disagreement are used to discuss the limitations of AI if it is not trained on enough good quality data.

Helpful notes:

For example, if your class has drawn lots of pictures of cats' faces without the bodies, or all in the same stance, you might point out that it won't be very good at recognising cats as having bodies/in a different position or from a different perspective.



Lesson 2: How AI uses data – Fairness and Bias

Final Discussion



Estimated completion time: 15 minutes

Display the 'Final Discussion' slide with the cat images and prompt Qs. Use the slide to help prompt a discussion with the class about data and bias.

Prompt Questions:

- **If your catflap was only trained on pictures of black cats and your cat was ginger, would it categorise your cat as a cat?**
- **Is this fair?**
- **How do you feel about it?**

Ask children the 'key question' at the bottom of the slide:
Can you explain how AI systems might create biased outcomes?

Sit in a circle & discuss the following questions, as well as any other observations the children have:

- In our version, as this was just an example/game, we only gave our imagined AI system drawings of cats – what problems might this create for a real AI system?
- What other problems might a real AI system have with this data set? (Can prompt to consider things like colour; accuracy of representations; whether cats portrayed lying/sitting; whether backgrounds included in children's drawings – e.g. If the AI system only has pictures of cats with white backgrounds in its data set, would it recognise a cat in a setting like in these photos or would it think that for it to be a cat it has to be on a white background?)
- Can you think of any issues that might happen if you were using a similar AI system to recognise different people?
- What have we learned about how AI systems work?
- What have we learned about what can go wrong with an AI system and why? (If the children suggested biases in AI that they had heard of earlier in the lesson, include these examples in your discussion here.)

Helpful notes:

There is evidence that some facial recognition AI systems (which work like our theoretical AI catflap) are better at recognising white adult men than other groups as this group was over-represented in the training data (as a result of human bias and the over-representation of white men in images online). You may wish to include this example in your discussion. Concerns have been raised about the use of AI systems such as these in law enforcement and the risk of them increasing the risk of marginalised communities being unfairly targeted by police.

Any AI system which does not include diverse data risks creating outcomes which are unfair, exclusionary, or discriminatory.

Key takeaways

When recapping at the end, be sure to remind children of these points:

- AI systems 'learn' by spotting patterns in data. The more data they have, the better they get at carrying out their task.
- For an AI system to learn to recognise something in a picture, for example, it will need to have had lots and lots of examples of this thing given to it and labelled as that thing (like the pictures of cats tagged as cats).
- If they do not have enough data, or the data does not represent all of the groups (of people/animals/anything relevant to the AI system's function) it should, it might not work for some groups or might create unfair outcomes for them.

Lesson 2: How AI uses data – Fairness and Bias

Additional activities and resources

You may find the following useful in order to explore this session's themes in greater depth.

- [AI for Oceans - Code.org](#) – This online activity can form the basis for an additional lesson. It explains how AI systems are trained with data and allows children to play a game where they train a robot to recognise rubbish in the ocean. It is based on a real-world use of AI and the children who worked on the project found it an enjoyable and memorable way of learning about AI and data. Levels 1-4 cover the basics described here, levels 5-8 go into more detail about wider implications and some more complex issues. With access to a computer or other device (we have not tested how well it works on tablets), children can go through the levels at their own pace, though it is advisable to watch the first video together as a class.
- [Home - Tech We Trust](#) . 'Tech We Trust' is an educational project that looks at digital skills, data skills, and the ethics of using technology to make decisions. It features an online live lesson activity which asks children 'Can you make a fair and unbiased algorithm?'. The activity supports children to think about fairness and bias in data. The interactive lesson is for learners age 10+ interested in tech, coding, and being ethical digital citizens.

