

Using AI tools responsibly

AI safety

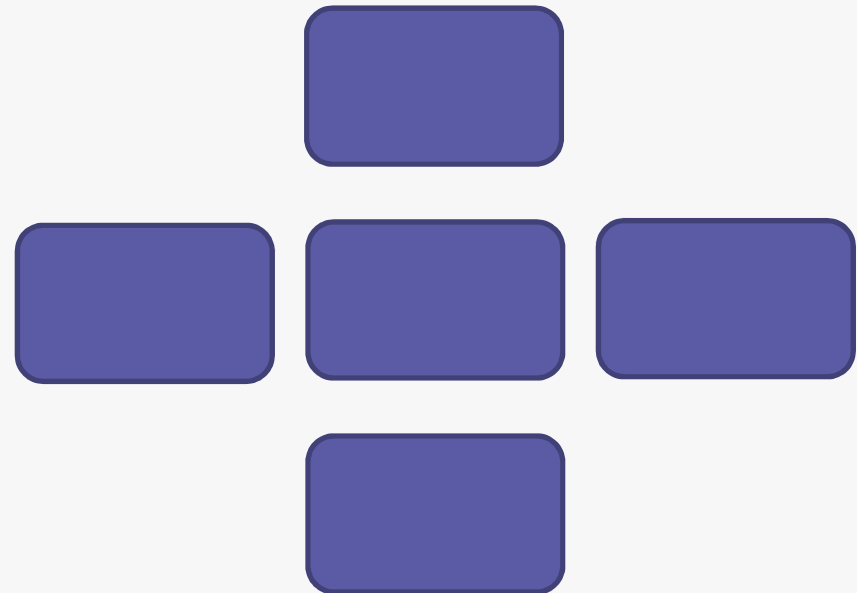
What is AI?

- AI stands for artificial intelligence.
- There are different types of AI:
 - **Narrow** – AI that can perform specific tasks e.g. *Siri, Alexa, Netflix/ Spotify recommendations*
 - **Generative** – AI that can generate new content e.g. *ChatGPT, Co-pilot, Snapchat's 'My AI'*
 - **General** – AI that can perform any intellectual task a human can do e.g. *Baymax from Big Hero 6*. **This type of AI doesn't exist yet!**
- AI needs data in order to work e.g. numerical, visual, audio.
- Some data is personal – this means it can be used to identify an individual (for example, name, date of birth or photograph).

Benefits and Challenges of AI

- Sort the x10 statements into **benefits** and **challenges**.
- Can you prioritise the benefits and challenges into two Diamond 5s?
- The benefit/ challenge you think is most important should be at the top, and the one you think is least important should be at the bottom.

Most important benefit/ challenge



Least important benefit/ challenge

Benefits and Challenges of AI: Answers

Benefits

1. AI can support learning and help people to do their jobs. For example, it can generate ideas and translate languages.
2. AI tools can help people with visual or hearing difficulties to access information.
3. AI can make some online experiences more fun, for example when it is used in games and video recommendations.
4. AI tools can help find ways to deal with climate change, for example by predicting extreme weather events or monitoring endangered species.
5. AI can summarise information quickly and make helpful recommendations.

Challenges

1. AI can sometimes make mistakes, especially if its answers aren't checked by humans
2. AI might replace some jobs currently done by humans, for example creating art or gathering data.
3. AI might be biased or make unfair suggestions depending on what data it has been trained on.
4. It can be difficult to judge what has been generated by AI and what has been created by people.
5. The technology used to power AI can cause harm to the environment by producing carbon emissions and using up lots of water.

Unplugged activity



CheckAI challenge

Using AI tools responsibly

CheckAI Youth Task Force

CheckAI are looking for young people to guide the world towards a responsible AI future.

You have all been invited to try out for the Youth Task Force (YTF).

To join the YTF you will need to complete **three challenges**.

CheckAI Youth Task Force

You will complete these challenges in groups of 3.

The YTF meets at CheckAI's main base, which is kept secure with a secret code.

Each challenge will get you a part of the code needed to enter the base.

Every team who enters the base will join the YTF.

CheckAI Youth Task Force

The challenges are also a competition.

You will receive points for every challenge.

The highest scoring teams will get extra rewards at the end.

Team up

Select who you will work with to complete the challenges.

Get into teams of 3!

Team up

You will have a team leader for every challenge. Everyone will get the chance to lead the team for 1 challenge.

Pick a team leader!

Challenge 1 – Fake accounts

CheckAI is always on the lookout for people who use generative AI to trick people online.

You need to identify the fake profile pictures.

There are 8 pictures — 3 of them are fake.

Rules:

1. The team leader must submit your guesses
2. Each guess must consist of 3 numbers (the pictures you think are fake)
3. You will find out **how many** are correct, but not which ones
4. The team to get all 3 fake pictures with the least guesses gets the most points

Challenge 1 – Fake accounts

The code for this challenge is the 3 correct numbers added together.

Write this down on your worksheet!

Change the leader

Each challenge needs a new team leader — so it's time to switch it up.

Pick a new team leader!

Challenge 2 – AI errors

CheckAI gets lots of requests from teachers worried about students using AI to do their homework for them.

AI applications **do not** always produce factual information.

There are 4 paragraphs each that contain **1 error** — you need to find and correct it.

Rules:

1. The team leader must submit your work
2. You must work as a team — everyone contributes!
3. For each paragraph, you will get:
 - a. 1 point for identifying the error
 - b. 1 point for correcting it

Challenge 2 – AI errors

The code for this challenge is...

9

Write this down on your worksheet!

Change the leader

Each challenge need a new team leader — so it's time to switch it up.

Pick a new team leader!

Challenge 3 – Spot the bias

CheckAI's main job is to help spot and correct bias in AI systems.

Bias is when the output of a machine learning model favours some things and deprioritises or excludes others.

You have 3 scenarios, and as a team you need to match the type, reason and solution for the bias in the AI system.

Rules:

1. The team leader must submit your work
2. You must work as a team — everyone contributes!
3. For each scenario, you will get:
 - a. 1 point for identifying the bias
 - b. 1 point for identifying the explanation
 - c. 1 point for identifying the solution

Challenge 3 – Spot the bias

The code for this challenge is a number hidden in all the scenarios. See if you can find it.

Write it down on your worksheet!

Using your code

It's time to submit your secret code to access the YTF base and join the team!

Submit your code now!

The highest ranking teams!

The winners are...

Transparency, accountability, and fairness

The activities you completed highlight the following important points about responsible use of AI tools.

1. **Transparency** – Using an AI tool is fine, as long as you are not hiding it and you explain how you have used the system. Creating fake images to fool people is never okay.
2. **Accountability** – If you use an AI tool, you are responsible for the output. If your homework contains errors produced by AI, it is your job to fix them.
3. **Fairness** – AI systems can be biased and it is important that everyone gets to experience the benefits and are not treated unfairly.

JCQ – AI Use in Assessments

- *"AI refers to the use of tools to obtain information/ content which might be used in work produced for assessments which lead towards qualifications."*
- *"The misuse of AI tools in relation to qualification assessments constitutes malpractice."*
- Examples of AI misuse:
- Copying or paraphrasing sections of AI-generated content so that the work submitted for assessment is no longer the student's own.
- Copying or paraphrasing whole responses of AI-generated content.
- Failing to acknowledge use of AI tools when they have been used as a source of information.
- Incomplete or poor acknowledgment of AI tools.
- Over reliance on AI tools resulting in student not independent criteria.

