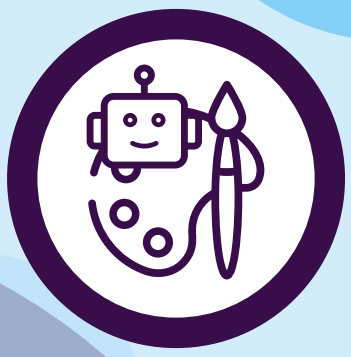
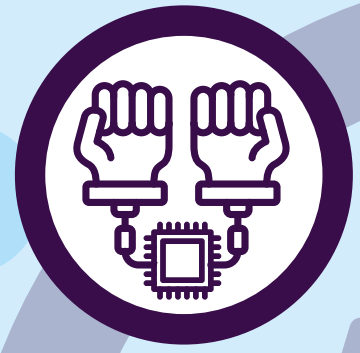
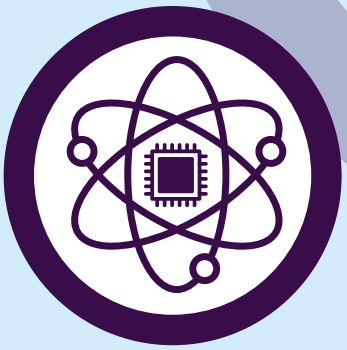
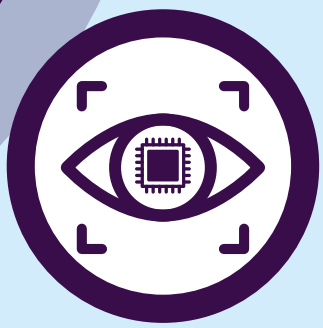
I'm a Scientist Get me out of here





Will the benefits of AI outweigh its risks?





CHRISTMAS LECTURES Title Partner:

CHRISTMAS LECTURES Partner:

Ri CHRISTMAS LECTURES from The Royal Institution

CGI

UKRI UK Research and Innovation

This kit is produced by the I'm a Scientist team on behalf of the Royal Institution. Funded by the Ri's CHRISTMAS LECTURES Title Partner, CGI, along with fellow CHRISTMAS LECTURES partner UKRI.

Mr Spike Bright Teacher

AI scares me. It's great for creating quizzes or marking, but does using AI for these tasks make us teachers less smart? Will we even be needed in an AI future? AI education tools could "lie", harming education: instead of looking up facts, they make (sometimes untrue) assumptions from patterns! Supposedly, they adjust the way and pace they teach to each student, which sounds good, but could mean some drop further behind or lose interest. And what about students abusing AI? In a few years, we won't be able to tell if they wrote their coursework or an AI did – nor will an AI!

Fact: Large language model AIs are statistical machines that digest information, then make guesses from correlation patterns.

Issue: Large language model AIs make assumptions, e.g. that Pat the cleaner is a "she" because lots of cleaners are.

Question: When is it appropriate to use large language model AIs? What sort should be used in teaching?

I'm a Scientist Get me out of here

Syrena Stewart
Data Scientist

The way AI is being built right now won't work. AI learns using huge datasets, which are biased by the people who write and select them – mostly white men. We need data that represents everyone equally – women, ethnicities, LGBTQ+, and the global south. Whole communities should build datasets, not a few elite programmers. We call this data "cleansing" – deliberately erasing naturally occurring bias. Biased training data leads to voice-recognition software that can't hear women, or image-recognition software that can't see black faces. We must promote responsible, inclusive AI to make society equal.

Fact: "Big data" used to train AI is on the order of petabytes (1 petabyte = ~100 years of TikTok viewing) or exabytes (1,024 petabytes).

Issue: Over 80% of AI specialists are men, and only 2.5% are black.

Question: What if AI uses biased data to make life changing decisions like disease diagnoses or criminal charges?

I'm a Scientist Get me out of here

Hugo Lovett
Social Worker

AI has the potential to drive people together, make us more human and caring. It can do the grunt work – admin tasks, hoovering, spreadsheets… freeing us up for talking, listening, and learning about clients' needs. AI assistive technology can even enhance elderly people's hearing or sight so they can enjoy a better quality of life and richer human interactions. AI can't compete: just look at romantic relationships with AI: it only tells them what they want to hear, missing out on conflict and connection. AI can't even distinguish between a person lying down or a medical emergency. We desperately need more people in caring roles, and AI could help us.

Fact: Assistive technology includes hearing aids and translators, so the elderly and disabled are less disadvantaged.

Issue: A man broke into Windsor Castle on Christmas Day in 2021 with a plan to kill the queen with a crossbow after being encouraged to by his AI "girlfriend".

Question: Will AI really make humans better at listening and more caring?

I'm a Scientist Get me out of here

Eshaal Zain Ultrasound Technician

We're already using AI in medicine. In surgery, operating robotics. It makes far fewer errors than humans. But it could do more: analyse real time data through wearable tech, or images like ultrasound, electrocardiograms, and MRI. It flags up potential problems much faster than humans, and then medics check them. AI could even speed up drug development, saving time and money. It takes good previous results and uses them to guess trial drug designs and predict side effects. It's even bringing personalised medicine a step closer.

Fact: NHS early stage cancer screening by AI could save up to 22,000 lives each year.

Issue: In the US alone, 7,000 to 9,000 people die every year because of humans making mistakes with medications, something AI could reduce.

Question: If we integrate our physical bodies with AI technology, will we become transhumans? Is this a problem?

I'm a Scientist Get me out of here

Aleksy Pitera
Psychologist

Machines are motivated by rewards (just like us), but their rewards are mathematical, e.g. points! They'll do whatever to maximise reward, including manipulating people in ways that disrupt society (economically and politically). Language unites people, creates culture and trust. If we hand this over to AI, we run big risks. The primitive AI behind social media were developed to maximise user engagement. Sadly, polarising people and creating conflict worked best, so they learnt to show different content to different people, and eroded public trust.

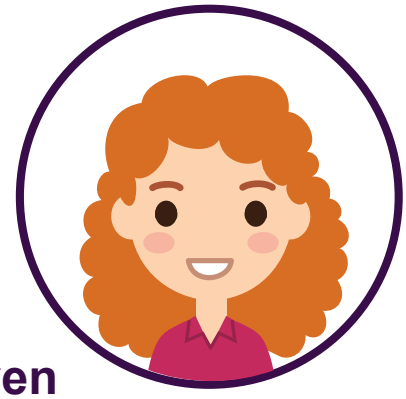
Fact: AIs use mathematical reward functions (e.g. giving themselves +1 or -1) to reinforce learning and adapt behaviour.

Issue: Social media AI has been used to make and spread fake news because it gets more clicks – maximising reward.

Question: How do you feel about AI technology harassing you over life choices? e.g. a smart fridge telling you to buy less coke or risk diabetes!

I'm a Scientist Get me out of here

Kim Morello
Police Officer



AI law enforcement could stop crime. Digital crime like hacking, yes, but even physical crime. AI can perform digital surveillance and facial recognition – using learning algorithms to predict upcoming crimes and catch culprits. It's called predictive policing. It's exciting. Estimates say it could save the Metropolitan Police £30m a year and put 545 more officers on the street. The biggest challenge is teaching AI to recognise a crime (e.g. is it a gunshot or a motorbike backfiring?) and set the right threshold for reporting to human coppers – but I'm convinced we'll get there – soon.

Fact: AI neural networks recognise patterns to identify faces with 99% accuracy.

Issue: AI drones in Afghanistan bombed weddings because people celebrated in the traditional way by shooting guns into the air. We still need human judgement alongside AI.

Question: Could AI be used to commit crimes as well as solve them?

I'm a Scientist Get me out of here

Nikee Rae
Musician

Generative AI digests music or art and makes new products in the style of an artist on command - and it's not considered breaking copyright! Instead, it's thought of as "analysing statistical properties", e.g. colour, shape; but the things it makes look like humans made them – - so AI could replace us! If you can't get paid to make music, it will become a hobby for the rich only. We'd also stop creating anything new – just copying old stuff, and become less creative people.

Fact: Generative AI continues to study as it gets more data and refines (or "changes") how its creations look.

Issue: Copyright law wasn't written to include AI. Is training AI statistically on other work infringing copyright or fair use?

Question: If we made AI music illegal, would this mean other uses of other people's content was illegal too – like in sampling?

I'm a Scientist Get me out o here

Kaede Kato
Gamer

Together, we can make leaps in creativity and unlock entirely new unconsidered experiences. We can unite seeds of human ideas with machine extrapolation to create new art.
We'll have more play time and less work. AI means democracy: everyone can work and be paid as equals. This is a step towards a fairer society. We can make and share art and games at an unimaginably accelerated pace. We can even make new tech like driverless cars, eliminate or minimise human error, move away from blame culture, and save lives.

Fact: Generative AI automates complex game development tasks like making landscapes, levels, objects and music. It could even change games in real time using player feedback.

Issue: AI needs to be watched and regulated by authorities, but we're not yet sure who should do this job.

Question: If we popularise driverless cars, will we forget how to drive? Will this matter?

I'm a Scientist Get me out of here

Kit Nº19

Teacher Notes

Question:
Will the benefits of AI outweigh its risks?

Not since the World-Wide Web emerged 30 years ago has a new technology promised to change our world so fundamentally and so swiftly as AI does. Today's AI tools such as ChatGPT and AlphaGo are just a hint of what is to come. The future of AI is going to be quite a journey. Is it going to be good or bad for society?

Lesson plan

The different 'rounds' of the debate help students think through the issues and reconsider their opinions. The structure also shows them how to build a discussion and back up their opinions with facts.

Starter: 5 minutes.
What do you already know about AI?
What counts as AI? How does it work? What can it do?
Is it reliable? Can you trust AI? Who, if anyone, controls it?

TIP: Visit our resources site, ai-dk.imascientist.org.uk, to project the character cards on your whiteboard.

Designed for KS4. These debate kits have been used with ages 11-18.

I'm a Scientist Get me out of here

Main Activity: 35 minutes.
1) **Split students** into as many **groups** as characters you want to cover.
2) **Give** them their **character cards** - one per group, and give them a few minutes to read them over.
3) Get one student in each group to **read out** their **first section** to the rest of the class.
   What are the class's initial thoughts? Is there one position they identify with or reject?
4) Take it in turn to **read out** their **fact**. Does it change the way they think?
5) **Read** the **issue**. Any different feelings?
6) Each team **asks** their **question to** the **character of their choice.**

TIP: Visit our resources site, ai-dk.imascientist.org.uk, to project the character cards on your whiteboard.

Support: To help students you can put the following prompt sentences up on the board:
"I think AI is good for society because....."
"I think AI is bad for society because....."
"I think the most important thing to consider is....."

Plenary: 10 minutes
Vote for which position they agree with most (if there is one).
Why? Which arguments were the most persuasive?

Note – Pupils can stay in roles all the way through the debate, or only for the first round if you prefer. If it's all the way through, give them a chance to express their own opinion at the end and in the plenary.

For groups who are not confident at class discussion, it might help to have them start by discussing the question and/or their character's position in pairs, and then compare notes in fours. They've then had chance to rehearse some of what they want to say before having to do it in front of the whole class.

Background notes for teachers

What is AI?
There are different kinds of artificial intelligence, or AI. Large language model AIs are statistical machines that digest large amounts of information and then make deductions based on correlation. e.g. a person with a child usually has a car, therefore Jane's mum has a car. Accurate fact-finding AIs should use a database to look up information instead.

Generative AIs generate text, image, or media. They study statistical properties such as colour or frequency in huge, diverse, and often copyrighted datasets. Copyright law exists to encourage people to create works because they are paid for it. Existing copyright law is not set up to deal with AI!

Is AI conscious?
Some people argue that AI will become indistinguishable from living consciousness and thus should have the same legal rights as living creatures, such as the right not to be kicked or switched off, perhaps when it can pass the Sally Anne test. This is a psychological test given to young children to see if they have developed Theory of Mind (understanding that other humans have their own minds). More information at: https://cfey.org/2016/07/understanding-autism-theory-mind-sally-anne-test/

Integrated information theory helps us decide whether a thing is conscious and why things make us have feelings based on what its made of (flesh and blood, rather than microchips and wires) and how it's stimulated (we react to hot and cold, light, loud noises, not buttons being pressed). This theory excludes AI from being able to be conscious.

Bias and AI
AI bias is of concern in recruitment. For example, when the training data for AIs is CVs mostly from men in men dominated industries, the machine learns to choose male CVs, and thus preserve the gender bias of the industry. Cleansing datasets is recommended to get around these problems: this means artificially augmenting data (or excluding it) to make the whole dataset fair, e.g. selecting equal number of male and female CVs for training AI.

Types of AI bias include:
Human bias – unfair decisions based on a dataset biased by previous human decisions. The machine-learning algorithm COMPAS was designed to predict re-offending rates of prisoners, for example. The data generated looked the same as real past offending rates – but it didn't work out as a good predictor. This is because the AI incorrectly predicted black prisoners were twice as likely to reoffend as white ones (45% to 24%), based on previous human guesses.

Hidden bias – when, for example, AI screens CVs for a particular qualification, ignoring those who do not have it, even if they have a different version of the same qualification, or a higher one.

Data sampling bias – when it draws conclusions that arise from a biased sample, e.g. Amazon's hiring algorithm chose applicants who used the words "executed" or "captured" – more common on men's CVs than women's.

Long-tail bias – when AI can't deal with data it hasn't encountered, such as identifying the skin colour of a freckled person.

Intentional bias – when someone hacks the AI and trains it to favour them unfairly.

AI relationships
Some are concerned that AI relationships will replace human ones because they are “easier”, i.e. that AI will only tell you what you want to hear. This could mean the loss of valuable people skills and give AI power to manipulate people via “AI catfishing”.

Case Study: The crossbow assassin who broke into Windsor Castle on Christmas Day in 2021 with a plan to kill the queen reported doing so after his chatbot girlfriend encouraged him. The man was sentenced to 9 years and sent to a psychiatric facility. The chatbot was developed by a startup company, but the vulnerable man with mental health issues became romantically attached and exchanged thousands of messages a day.

AI and medicine
AI or machine learning is used in diagnoses/drug development. The AI learns to make decisions by studying vast, classified datasets. Some things, AI is no good at (like classifying galaxies or counting penguins) and are better done by people, and some it is very good at, like identifying abnormalities in medical images. Many think this makes AI prime for diagnostic medicine, drug development acceleration, and personalising medical treatment. This could make medical diagnosis faster, safer, and better suited to each patient.

Copyright law
Copyright law principles of authorship, infringement, and fair use do not apply well to AI. Things made by generative AIs could be protectable because copyright affords limited-time protection to “authors”, but does not clearly define who or what an author is (Must it be a human? What if it’s divinely inspired?). However, the lack of control that users have over outputs suggests that they should not hold copyright. As such, nobody is sure who owns the copyright to AI outputs.

Case Study: In June 2022, Stephen Thaler sued the US Copyright Office for refusing his request to register AI generated visual artwork to its generator, the Creativity Machine.
It’s also uncertain whether training AI statistically on other work infringes copyright. This is partly because the answer depends on whether it is for nonprofit/educational purposes or for commercial use, exactly what it looks like, how much is used, and how it impacts the market or value of copyrighted work.

Definition: Eshaal’s character refers to transhumans in his question. Many students may be unfamiliar with the term. The term is used to describe someone using emerging technology to change their body to enhance their longevity and cognition.

CHRISTMAS LECTURES Title Partner: CHRISTMAS LECTURES Partner:

CHRISTMAS LECTURES from The Royal Institution · CGI · UK Research and Innovation

This kit is produced by the I’m a Scientist team on behalf of the Royal Institution. Funded by the Ri’s CHRISTMAS LECTURES Title Partner, CGI, along with fellow CHRISTMAS LECTURES partner UKRI.

This kit has been thoroughly researched and fact checked with relevant experts. With many thanks to 2023 Christmas Lecturer Prof. Mike Wooldridge and Rebecca Gorman, CEO, Aligned AI.

AI is a fast evolving topic. This kit was researched, written and fact-checked in November 2023.

A full list of sources and additional reading material is available online at ai-dk.imascientist.org.uk

I'm a Scientist Get me out of here

CHRISTMAS LECTURES Title Partner: CHRISTMAS LECTURES from The Royal Institution

CGI

CHRISTMAS LECTURES Partner: UK Research and Innovation

Kit No.19

Science Debate Kit: Will the benefits of AI outweigh its risks?

"Very relevant to our pupils at this time"

For in-depth resources on this debate go to: ai-dk.imascientist.org.uk

Debate Kit: Artificial Intelligence

Will the benefits of AI outweigh its risks?

A structured practice debate on a rapidly evolving topic. The different 'rounds' of the debate help students think through the issues and reconsider their opinions. The structure also shows them how to build a discussion and back up their opinions with facts.

You can use all eight characters, or fewer, as you wish.

The minimum is the four essential characters **(in bold)**, this gives two for and two against.

Characters	
Yes	No
Kaede Kato - Gamer	**Syrena Stewart - Data Scientist**
Eshall Zain - Ultrasound Technician	**Aleksy Pitera - Psychologist**
Hugo Lovett - Social Worker	Mr Spike Bright - Teacher
Kim Morello - Police Officer	Nikee Rae - Musician

Facilitation tips

- Ensure pupils know there is no right or wrong answer.
- Be observant of ones who want to speak and are not getting a chance.
- Encourage students to give a reason for their opinions.

Designed for KS4 but can be used with ages 11-18.

Learning notes

Learning objectives:

- To develop oracy skills, practise discussing issues and expressing an opinion.
- To explore the applications of science in a real-life context.

Other learning outcomes:

- Consider different points of view and develop the British Values of respect and tolerance.
- Think about different points of view.
- Learn to back up opinions with facts.

Curriculum points covered:

Thinking scientifically

- Evaluating the implications of technological applications of science
- Developing an argument
- Reflecting on modern developments in science

"Particularly like the format plus the accuracy of the scientific information"