

Safer Internet Day 2026

Resources for use with
14-18 year olds



**Smart
tech,
safe
choices**

**Exploring
the safe
and
responsible
use of AI**

Safer Internet Day

Safer Internet Day is celebrated globally in February each year to promote the safe and positive use of digital technology for children and young people, and to inspire a national conversation about using technology responsibly, respectfully, critically, and creatively. Safer Internet Day 2026 is on Tuesday 10 February.

In the UK, Safer Internet Day is organised by the UK Safer Internet Centre (UKSIC), a partnership of three charities – Childnet International, the Internet Watch Foundation (IWF) and SWGfL.



With kind thanks to the Welsh Government for their support in the development of these resources in Welsh.

For more resources and advice visit:

hwb.gov.wales/keeping-safe-online



Cefnogir gan
Lywodraeth Cymru
Supported by
Welsh Government

With kind thanks to the Safeguarding Board for Northern Ireland for their support in the development of these resources in Irish.

For more resources and advice visit:

onlinesafetyhub.safeguardingni.org



Smart tech, safe choices

The theme for Safer Internet Day 2026 is artificial intelligence or AI. From voice assistants, to chatbots, to the ethics of generative AI, our resources will help you and your learners explore how AI can be used safely and responsibly and ensure everyone feels empowered to adapt alongside changing technology.

We understand that talking about new technology with children and young people can be challenging, especially if you're less tech confident yourself.

This is why we've also compiled a list of further reading and resources that you may wish to use alongside our resources to develop your own or your learners' understanding:

saferinternet.org.uk/further-resources

We are calling on government and industry to ensure safety and children's rights are at the heart of advancements in AI.

For young people who are already using this technology, we want to support them to develop the skills and confidence to make safe and responsible choices.

Exploring the safe and responsible use of AI

3 to 7 year olds

Focus on voice assistants, limited references to AI

7 to 11 year olds

Explores a range of safe and responsible uses of AI, including chatbots

11 to 14 year olds

Explores safe and responsible use of AI, including chatbots and options to discuss AI generated nude images

14 to 18 year olds

Focus on ethics of AI and exploring individual and societal responsibilities



Contents

6 Using these resources

7 **Activity:** ethical AI

13 **Activity:** costs and benefits

20 **Activity:** bias in AI

25 **Activity:** AI manifesto



Using these resources

Adapt our resources to best suit your learners needs

We know that not every young person is the same, and that special educational or additional learning needs can make children more vulnerable to online harm. This is why all our resources have been created with accessibility in mind. We don't reference specific ages, except in the file names and on front covers, so pick and mix activities from across the resources, and feel free to make changes where you need to. All our activities include suggestions to scaffold and support learners, as well as ideas to challenge and extend.

Practitioners are encouraged to incorporate these activities, or selected elements, into their curriculum, while thoughtfully adapting the resources to meet the unique needs of their learners.

Establish a safe space

Consider how to establish a supportive environment for young people to learn and talk about potentially sensitive topics. For ideas visit:

childnet.com/learning-environment

While discussing their online lives, it is possible that children and young people will raise concerns about things they have experienced online. Make sure you are up to date with your school or setting's safeguarding procedures and for further advice visit:

childnet.com/safeguarding

Know where to get more information or help

Remember that there's plenty of help and advice available if you need more information about anything online safety related. The UK Safer Internet Centre's 'Need Help?' page contains further information on reporting specific concerns to organisations outside of your school or setting:

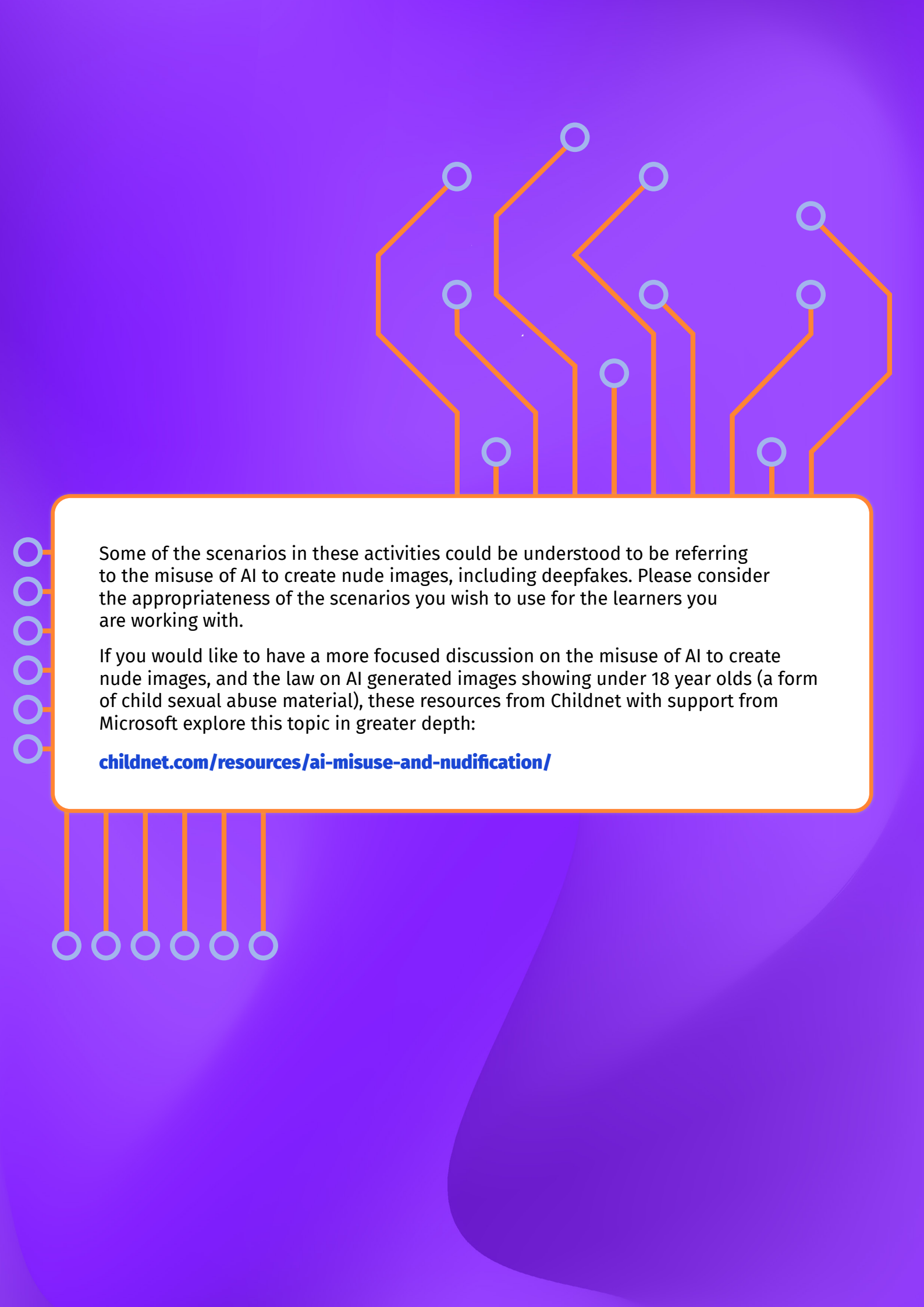
saferinternet.org.uk/advice-centre/need-help

The Professionals Online Safety Helpline is a free helpline service offering advice, support and guidance to all members of the children's workforce with questions about online safety issues:

saferinternet.org.uk/professionals-online-safety-helpline

Kickstart conversations that last all year round

The online world is changing all the time, but talking with young people about their online lives can help you understand their experiences. Kickstart the conversation today but make it a regular habit all year round.



Some of the scenarios in these activities could be understood to be referring to the misuse of AI to create nude images, including deepfakes. Please consider the appropriateness of the scenarios you wish to use for the learners you are working with.

If you would like to have a more focused discussion on the misuse of AI to create nude images, and the law on AI generated images showing under 18 year olds (a form of child sexual abuse material), these resources from Childnet with support from Microsoft explore this topic in greater depth:

childnet.com/resources/ai-misuse-and-nudification/

Activity: ethical AI

Time:

adaptable, depending on time available.

Learning objective:

- I can explain some of the ethical considerations in responsible use of AI technology.

Materials:

- Accompanying Ethical AI slides (slides 2 to 11)
- Ethical AI educator guidance.

Suggested activity guidelines:

This activity introduces a range of ethical considerations in responsible use of AI technology and can be adapted for use in different ways, depending on the time available.

1. Using slide 3, explain to learners that they will be looking at different uses of AI and thinking about their impact. The focus will be on ethics, which is a way for people to think about what they believe is right and wrong behaviour. Ask learners if they have already heard of any ethical concerns relating to AI.
2. Use slides 4 to 11 to look at each ethical consideration in more detail. Each slide gives an overview of what the issue is and has a short list of discussion questions to help guide reflection. You could:
 3. a. Print the slides and distribute each one to pairs or groups of learners. Ask them to discuss the information they have been given and prepare a short presentation or overview to share with the wider group.
 - b. Use the slides over several days to facilitate a whole group discussion about each issue and the impacts it could have at an individual or societal level.

Note: the text on each slide is also editable, so you could add to or remove the example given and invite learners to think of their own, or debate whether learners think each specific example is an ethical use of AI or not.

Activity: ethical AI

Ways to support and scaffold

Before the session, select the issues that will be most relevant to the group and adapt the language and information on the slides as necessary. You may wish to use these discussion questions as alternatives to those currently on the slides:

1. How might this make someone feel?
2. How might this affect someone like you?
3. How might this affect other people?

You could also search for a news story or video report that covers the issue before they start their discussion.

Ways to challenge and extend

Use the optional discussion statements on slides 12 and 13 to facilitate a debate or group discussion. Alternatively, invite learners to research and write their own arguments for and against.

Ethical issues with AI – **educator guidance**

Issue: dependency

Some would argue that AI was designed to improve our efficiency when completing tasks, but not to just do all the work for us.

If people become dependent on AI this may mean they're not getting the opportunity to practise existing skills or develop new ones. Over time, for example, this could weaken someone's critical thinking and problem-solving skills. They may also miss the mistakes that AI sometimes makes in what it produces, leading to inaccuracies.

Workplaces may become less collaborative as people instead turn to AI, affecting a group of people's interpersonal skills. Some people may also disagree with the use of AI for such tasks and may feel it gives some people an unfair advantage over others.

Some may argue that if large numbers of people become reliant on AI, this could mean a loss of key skills for society, reducing original thought and creativity. However, others may argue that using AI to complete repetitive or mundane tasks could free up time for people to focus on innovating and thinking more

Issue: bias

AI wasn't necessarily designed to be biased but it can only be as accurate as the data on which it was trained. If the data it uses is biased, then AI will also produce biased outcomes.

If a person is regularly using and consuming biased outputs from AI, without thinking critically about them, it may negatively shape how they see people and can reinforce harmful stereotypes. People who feel underrepresented or misrepresented by AI may feel upset, distressed or frustrated.

Continued exposure to biased AI outputs may reinforce harmful stereotypes in relationships, workplaces, etc., directly affecting people's lives. For example, women or some ethnic groups may receive fewer opportunities at work if biased AI systems are used in recruitment processes, or Black people may be targeted and experience discrimination if biased AI systems are used in policing or judicial processes.

In addition to these individual experiences, bias in AI may contribute to and exacerbate the discrimination against particular groups in society, more widely.

Note to Educators

For a more in-depth exploration of bias in AI, please see activity 3 in these educational resources.

Ethical issues with AI – **educator guidance**

Issue: AI chatbots

If someone is over reliant on AI chatbots, they may overlook potential inaccuracies in the information they are provided. If they are relying on chatbots for support with more personal issues (e.g. mental health), there are risks. AI chatbots do not always signpost to external or expert help, which may mean a user is not prompted to access the support they need. The advice they are given by the AI chatbot may not help them in the long run and may end up being harmful to them. AI chatbots were not designed for this purpose and mental health support should be delivered by a trained professional. Chatbots are often designed to encourage ongoing interaction and respond in a way that pleases or satisfies the user, so they may inadvertently validate unhealthy or harmful ideas expressed by the user.

A dependency on AI chatbots may also affect other people in someone's life – if that person feels like they are getting the interactions they need from the AI chatbot, they may turn to their friends and family less, neglecting their existing relationships.

It could be argued that if large numbers of people seek mental health advice and support only from AI, then health services will not have an accurate understanding of the demand and need for help amongst the general population. This in turn could lead to less funding for support services, creating a cycle where people are unable to access expert advice and are forced to get help from alternative places, like AI.

Issue: creativity and inspiration

AI has been designed to produce outputs that would traditionally be described as 'creative', such as images, stories and music. However, many argue that it cannot fully replace human creative thought.

Some argue that AI can help to boost or enrich someone's creativity, providing inspiration and a new medium for creating digital work. However, there may also be concerns that a reliance on these tools could limit someone's ability or desire to create original work. Additionally, as AI can only generate work that mimics or imitates existing styles, it could be argued that it cannot create any original work at all.

It is difficult to know how AI will impact the arts and creative industries long term. The accessibility and convenience of using AI may mean that some creative tasks will no longer require an expert or professional to complete, which could lead to people in creative industries losing their jobs. However, as AI generated art becomes more common, it may also mean an increase in demand for people with specific creative skills or for human-led original design.

Ethical issues with AI – **educator guidance**

Issue: presenting AI work as your own

AI can generate a range of outputs, including emails, essays, articles, images, videos etc. but producing high quality work can require specific and carefully crafted prompts. Generally, AI is recognised as a tool that can support with the process of writing, but that should not be relied on to produce an output in its entirety.

If someone continuously relies on AI to complete their work and presents this as their own, they will lose the opportunity to practise and develop their own skills. Additionally, if they find themselves in a situation where they are unable to use AI (such as an exam), they may find it difficult to produce the results they have previously. In schools and universities, there is an expectation that AI should not be used in this way, and safeguards are in place to help identify and detect AI generated content. If a student is misrepresenting the work of AI as their own, they could face consequences such as failing a module, expulsion, etc. as it would likely be considered academic misconduct. In a professional environment, overuse of AI or presenting work from AI as their own may impact someone's professional reputation and future opportunities.

Presenting AI work as their own may also impact on someone's friends, family or colleagues, who may assume they are coping well with their workload or be impacted by the consequences if they are discovered to be using AI inappropriately.

More widely, as AI doesn't always correctly cite where it has retrieved its data from, there are concerns about people not receiving credit and recognition for their work. Longer term, if lots of people present AI as their own work, this could be reflected in the skills and knowledge of a workforce or population.

Issue: AI work revolution

There are many concerns about how AI is going to impact the future of jobs, and this is something young people may also be worried about.

For some jobs, AI may be able to effectively replace the role of a human, for example by automating processes that were previously completed manually, or by improving efficiency to a point where fewer people are needed to oversee or complete a task than previously. This could lead to an increase in unemployment, particularly for entry-level roles.

However, there will also likely be new careers that are created, for example to develop and oversee AI technology. Additionally, as AI continues to be developed and society becomes more familiar with it, some industries or roles led by humans may be seen as more valuable.

The wider societal impact may depend on a number of factors, such as how people can be upskilled to adapt to a society alongside AI technology use, how AI use is regulated by the government, and the extent to which different groups of people can access support (for example at school, or through funded training) to ensure that new opportunities are available and accessible to everyone equally.

Ethical issues with AI – **educator guidance**

Issue: environmental impact

AI requires a significant amount of energy and water usage to function. Training and running AI models requires vast amounts of power, often from fossil fuel-based sources.

Whilst individual users may not connect their use of AI with ongoing environmental changes, it could be argued that the normalisation of AI usage will contribute to increased greenhouse gas emissions, making it harder for countries to meet their targets to reduce global emissions and contributing to the ongoing process of global warming and climate change.

However, there are also arguments that the responsibility for mitigating the environmental impact of AI technology should be on the developers of such technologies, as opposed to individuals, and that the rise in popularity of AI may see increased pressure from climate activists and environmental scientists, which ultimately encourages the development of more efficient technology. There are also many examples of AI being used to protect the environment or to support climate science, with technology having the potential to help reduce the global impact of human behaviour and find new solutions to environmental problems.

Many people would argue that overuse of AI will have a long-term detrimental impact on the global environment and will exacerbate the existing climate crisis. You may wish to discuss small changes learners could make to reduce their consumption through AI, such as using a traditional search engine where possible, or making requests of AI short and efficient, including avoiding saying “please” or “thank you”, which will not affect the AI’s responses or functionality, but increases processing requirements, and therefore energy consumption.

Issue: deepfake videos

AI’s capabilities mean that it can create very realistic replicas of real people, saying and doing things that they haven’t done. However, many would consider this a misuse of AI, as it is often done without consent or with malicious intent.

If someone is targeted using deepfake videos, particularly if it makes them appear naked, or like they’re doing something inappropriate or illegal, this can make them feel very attacked, exposed, uncomfortable and vulnerable.

In some cases, this can affect someone’s relationships with others, including if the deepfake has been created or shared by someone they know, or if the deepfake is believed to be real and impacts someone’s opinion of the person shown.

With this sort of AI technology becoming increasingly sophisticated, it may get to a point where people find it difficult to deduce what is AI generated content and what isn’t. It also raises serious concerns about people’s rights to images of themselves and may lead people to become increasingly paranoid about what they share of themselves online. Learners may also observe that deepfakes, particularly sexualised deepfakes, may be seen to disproportionately affect women and that women may experience more serious outcomes when they are targeted in this way. This could perpetuate patterns of sexism and misogyny, where women are objectified or seen as ‘lesser’ than men.

Activity: costs and benefits

Time:

30 minutes.

Learning objective:

- I can evaluate the costs and benefits of different uses of AI, including the environmental costs, data use and ownership/copyright considerations.

Materials:

- 'Weighing scales' activity sheet for each learner or group.
- 'Uses of AI' cards cut into part one and part two of each scenario.

Suggested activity guidelines:

This activity considers how learners can decide if it is responsible to use AI in different circumstances by weighing up the costs and benefits.

1. Explain to learners that to use AI safely and responsibly, we must consider the potential impacts and decide if the benefits outweigh the cost, or if the costs are too high. You may want to remind learners that when we are talking about cost in the session, it means the harms or disadvantages that may occur.
2. Split learners into small groups and give each group a copy of one of the 'Uses of AI' cards. Make sure to only provide them with part one at this point of the activity and retain part two to share with the relevant group later. Give each group a copy of the 'Weighing scales' activity sheet.
3. Ask learners to read through their scenario and decide what benefits there are to using AI in this situation, and what costs it may have. They can record these ideas on their 'Weighing scales' activity sheet.
4. Ask the groups to evaluate whether they think this is a safe and responsible use of AI overall. Do the benefits outweigh the costs? They may need to consider that some benefits or costs may be worth more than others.
5. After learners have discussed and written down their initial ideas, provide them with the second part of their scenario from the 'Uses of AI' cards. Does this change the weighting of their scales?
6. Bring learners together as a whole group to present what they have found. They should share their scenario, the costs and benefits they identified, and the conclusion that they came to overall. The other learners can then vote whether they agree or disagree with the group's conclusion.

Activity: costs and benefits

Ways to support and scaffold

Model the activity by choosing one of the scenarios and working through it as a class. You could also provide learners with the 'Idea Bank' activity sheet which gives suggestions of potential costs or benefits for them to decide which ones apply to their scenario.

Ways to challenge and extend

Ask learners to research the issue raised in their scenario further. Does this change the weighting of their scales? Alternatively, ask learners to create their own AI ethical dilemmas with specific scenarios for other people to consider.



Scenario 1, Part 1

Roisin posts photos and videos about her local area. She has had a lot more views since she started to use AI to add things into her videos. Roisin has found the tools have become much easier to use and look better than when she first started. It generally looks realistic, but she often adds things that aren't believable, like when she put a huge melting ice cream on top of a local ice cream shop. Some platforms have started to mark her posts to say they show AI generated content.

Scenario 1, Part 2

Roisin's posts are getting a much larger audience with lots of people liking and sharing them. She has seen some comments where people are confused about whether the video is true or not. One post, where she had edited the buildings to grow into skyscrapers, even got reshared on another platform where lots of people were commenting quite negatively on it. As she has started to become more popular, she is starting to get brand deals, with local businesses asking her to create an AI video for them. One business wants her to make it seem like a celebrity has visited their shop, whilst another wants it to look like their products are spilling out the door.

Scenario 2, Part 1

Arjan has started his own small business selling a craft he makes. As he started getting more sales, he decided to start his own website. He doesn't know how to code or design graphics so he used an AI tool to build his website for him which was cheaper than paying someone else to make it. It looked great, although he didn't read the whole terms and conditions as they were really long, so he's not quite sure if he owns the rights to the code, or whether the tool can use his images now he has uploaded them.

Scenario 2, Part 2

Arjan has been getting lots of sales through his website but lots of queries as well. It was taking up a lot of his time to reply to each one which meant he had less time to make his crafts, so he decides to put an AI chatbot into his website. The chatbot can answer the basic questions he gets asked a lot, about things like tracking and shipping, or it can pass the question to him if it can't help the customer. He now has more time for making the product so can make more sales.



Uses of AI Cards

Scenario 3, Part 1

When Jay was really stuck on how to start an essay, they used AI to help them. They wrote the majority of the essay and it was just a couple of ideas to get them started. If they hadn't used AI, Jay wouldn't have had time to get such a good grade. Over time, Jay used AI more and more to help them when doing academic writing. Jay feels they wouldn't be doing so well if they weren't using AI and now finds it quite hard to think of their own ideas. Some of Jay's friends are using it too, but some others aren't and seem to spend more time on their writing whereas Jay gets more time to relax.

Scenario 3, Part 2

When Jay was in college, the teachers had never said anything about AI, so Jay wasn't concerned. When they started university, Jay went to a talk about academic misconduct and the university said they would be using AI detection tools, but Jay has heard these don't work that well. Jay's friend Archie has been investigated for academic fraud which Archie was really annoyed by, as he has never used AI and is strongly against it. This caused some tension between Jay and Archie, as he knows Jay uses AI for their work but has never been caught, whereas Archie had been told he could have failed the course or had to leave university. At university, they must reference their writing, so Jay has started to use AI to find references to quote in their writing. Sometimes the tool makes these up, but sometimes they are correct.

Scenario 4, Part 1

Elijah uses AI to answer questions for him. He used to use a search engine, but AI gives him an answer straight away without having to click through to a website. It also tends to put it in a way that is easy to understand, and gives useful suggestions when he has an issue, like when he wanted to work out a skincare routine. Elijah's friend told him that AI takes up way more energy than a search engine, and that this might not be from renewable sources. He also hears it uses lots of water to cool down the data centres.

Scenario 4, Part 2

As the chatbot was so useful, Elijah starts to use it for more personal questions. When he had been feeling down for a while, he talked to AI about it. It gave him suggestions of how to feel better and said he could always talk to it. It made Elijah feel better at the time so now he keeps going back to it when he is feeling these emotions. He does find it can be very agreeable and friendly. Sometimes it can be so agreeable that it goes along with whatever Elijah says, even when he's joking and tells the AI things he knows aren't right. Elijah finds it funny when it agrees with these ridiculous things, but he did think that someone younger or more reliant on AI might not recognise that it's responses could be harmful.



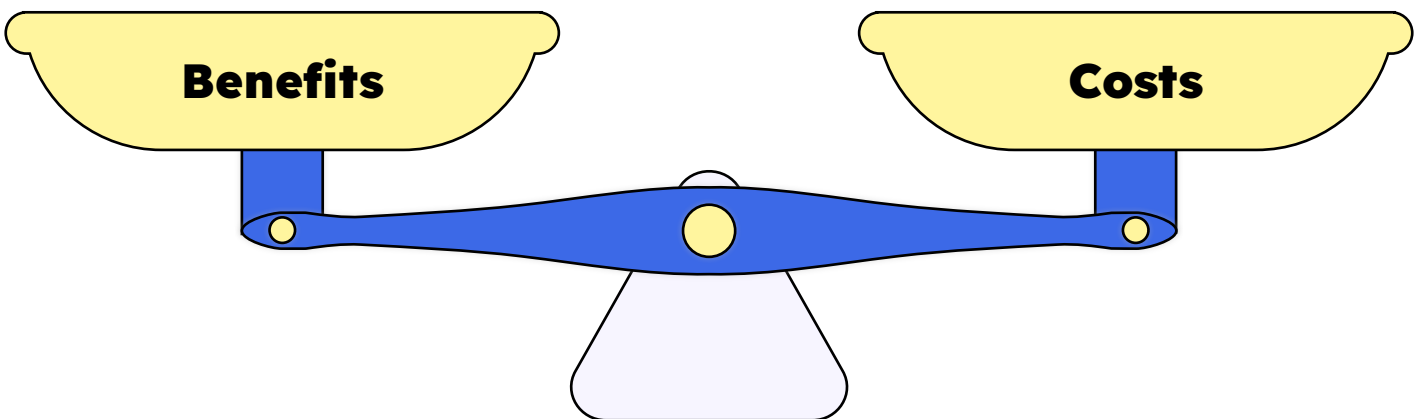
Uses of AI Cards

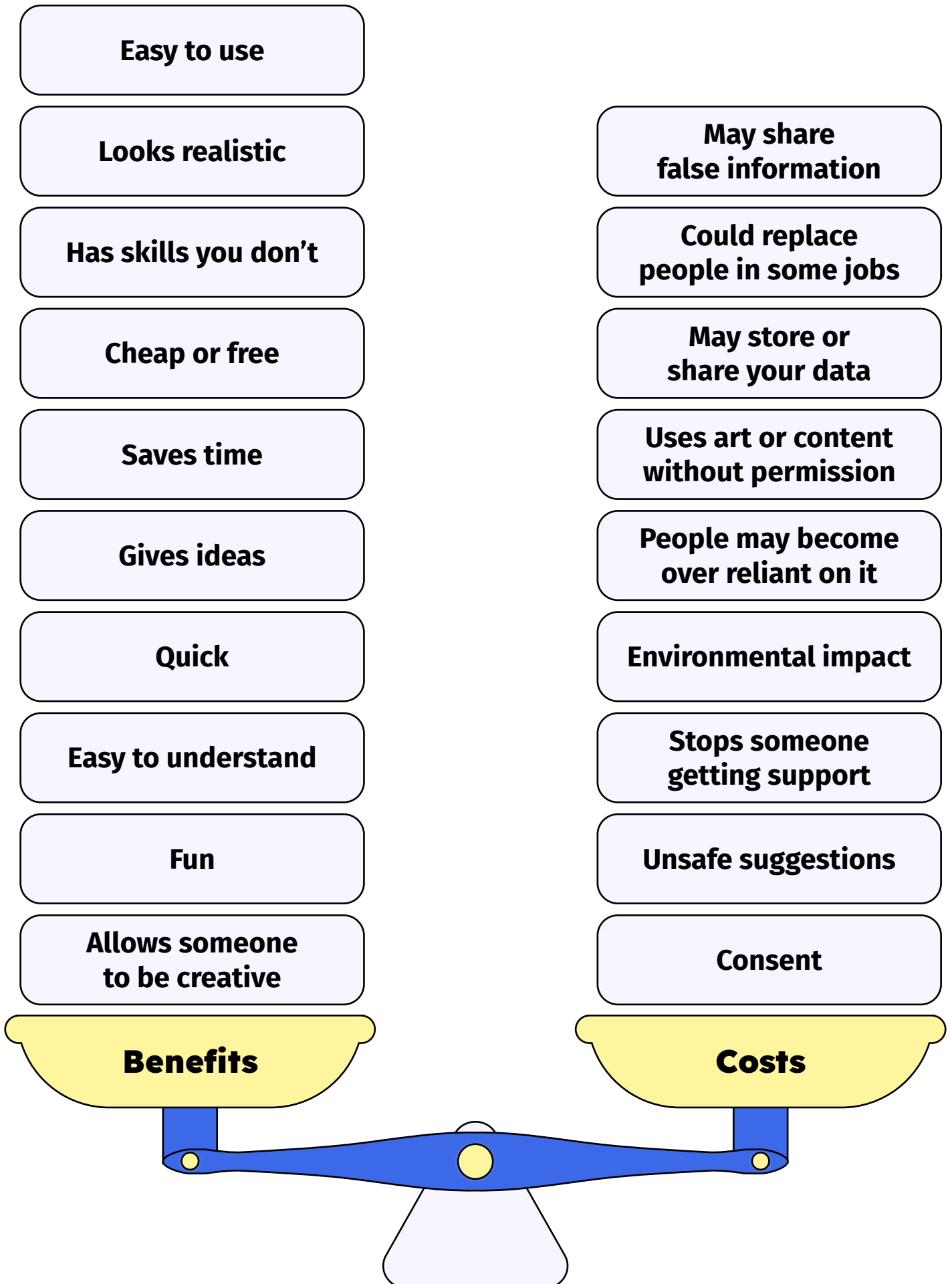
Scenario 5, Part 1

Holly is part of a fan group for her favourite show. People often use it to post fan art of characters from the show. Holly has created some art of her favourite characters using AI, which she said in her caption. Lots of people really enjoyed the style she'd chosen and wanted her to make more with their favourite characters. A few people have commented saying that the art style is stolen from another artist without their permission. Holly learns from the comments that the AI she is using is trained using huge amounts of existing images, and that these are often taken from artists without their knowing. The artist has even signed a petition against AI using their images in this way. Holly isn't making any money from the images though, and it is just for fun.

Scenario 5, Part 2

Some people are posting videos of the TV show that Holly hasn't seen before. Holly thinks from what people are commenting that they use AI to make these, but it isn't tagged as AI so she isn't sure. She remembers there was an actors' strike as they wanted guarantees that their voice and likeness wouldn't be used to train AI models without their consent, as they were worried it would replace them. As some people in the group are using open-source AI, they may have trained it on real photos and videos of the actors to create these without their knowledge. Some of the videos she sees are quite sexual, which has never been part of the TV show. They look just like the actors in the show so it would be easy to think it was real.





Activity: bias in AI

Time:

30 minutes

Learning objective:

- I can describe how AI can perpetuate bias inherited from the datasets on which it was trained and can use this knowledge to inform my own use of AI technology.

Materials:

- Accompanying slides for 'Bias in AI' (slides 14 to 27)
- Bias in AI – Educator Guidance (optional)

Content warning:

This activity includes examples of historic and current discrimination, including racism. These examples have been included to encourage a greater understanding of how bias in new technology may disproportionately affect certain groups. You may wish to consider additional safeguards to support learners who have lived experience of these issues.

Please note, a version of this content warning is also provided on slide 14 and we suggest educators share this with learners before starting the activity.

Suggested activity guidelines:

This activity introduces learners to different examples of bias in AI, why this bias occurs, and encourages them to think about the impact this could have.

1. Display slide 15. Begin by asking if they know what 'bias' is, then click to reveal the definition.
2. Display slide 16 and read out the example of bias to further explain what bias is and to contextualise it in terms of technology design and the harm this can cause.
3. Display slide 17 and ask learners if they know of any examples of bias present in AI. Click to reveal how bias can be present in AI.
4. Divide learners into pairs or small groups, and give each group a printed copy of one of the examples of AI bias on slides 19 to 25. Display slide 18. Ask learners to discuss the questions on this slide in relation to the AI bias example they've been given. Then work your way through slides 19 to 25, asking learners to share what they discussed for each example. You can use the Bias in AI – Educator Guidance to help you with these conversations.

Activity: bias in AI

Suggested activity guidelines:

5. Ask learners to consider the different examples of bias in AI. Ask them to discuss and summarise how bias in AI might impact society more widely. If they are struggling with this conversation, ask them to consider the following points:
 - Inequality in society
 - Reinforcement of stereotypes
 - Ethical and legal concerns
 - Economic impacts
 - Business impacts
 - Health and safety impacts
 - Personal wellbeing
6. Use slide 26 to explain how AI bias might impact society more widely.
7. Explain to learners that it is important for developers and programmers of AI to try and reduce bias in AI. For example, they can train AI models to reflect less bias in their outcomes. However, everyday users of AI can also do their bit to help challenge AI bias. Ask learners what they think everyday users can do to play their part in this. After a discussion and taking some answers, you can use slide 27 to go through the answers.

Ways to support and scaffold

Rather than distributing the scenarios out to learners, choose a few of the scenarios most relevant or accessible to your learners and explore them together.

Alternatively, ask learners, in pairs or in small groups, to draw a scale on a piece of A3 paper, from 'not harmful' to 'very harmful'. Give learners cut out copies of the Bias in AI cards. Ask learners to read each example and decide where they'd place it on this scale. Ask different groups to explain where they placed certain examples and why.

Ways to challenge and extend

Challenge learners to draft a list of principles for AI developers to adhere to that will reduce bias in AI systems.

Bias in AI Teacher Guidance

| | |
|---|--|
| <p>In healthcare, data of women and minority groups is sometimes underrepresented, such as in computer-aided diagnosis (CAD) systems, meaning there sometimes isn't enough data to confidently give a full and accurate diagnosis for them.</p> | <ul style="list-style-type: none"> • Systems may return lower accuracy results for certain patient groups • This may lead to these groups being misdiagnosed • Misdiagnosis might result in the wrong treatment being given, further harm to their health, and avoidable deaths |
| <p>When using AI to streamline CV scanning during a search to find ideal candidates for a job, the information requested and answers screened out can result in disproportionate outcomes across different groups of people. For example, a certain word in the job description might attract or favour more men than women.</p> | <ul style="list-style-type: none"> • This can give men an advantage over women for certain job roles • This reinforces gender gaps in the workplace and disadvantages women – women may be less able to gain opportunities for progression and pay increases • This may result in certain workforces becoming predominantly male • Inclusive workplace policies might be overlooked if they are male dominated • It may create a hostile and unsupportive environment in the workplace for women |
| <p>When using AI image and art generators, they will often perpetuate archaic and harmful stereotypes. For example, job roles like lawyers, doctors and judges will often be generated as white men. It may also generate images of specific groups of people (e.g. LGBTQ+ people) in a stereotypical way (e.g. gay men as physically fit or wearing 'feminine' clothing, gay women as having nose rings and shaved hair, and bisexual people and nonbinary people with dyed hair).</p> | <ul style="list-style-type: none"> • This can reinforce gendered workplace stereotypes, perpetuating the idea that these jobs are only suitable for white men • People who do not feel represented in these types of images may feel disregarded and unseen, and like they don't fit societal norms, and it may negatively affect their sense of self-image and identity • These AI generated images may be published widely, reaching lots of different people very quickly, meaning these harmful stereotypes might be embedded and reinforced more widely in society |
| <p>AI-powered predictive policing tools may be used to predict areas where crime is more likely to happen. These often rely on arrest data from the past. This past data may have included incidents where people of colour were unfairly targeted and prosecuted as a result of prejudice.</p> | <ul style="list-style-type: none"> • Because the data includes patterns of bias, this would mean that bias is perpetuated, with people of colour disproportionately affected and targeted • It may lead to continued assumptions that certain people are more likely to commit crime, or that particular areas will have a higher crime rate because of the identities of the people that live there • People may be accused or prosecuted for crimes they did not commit |

Bias in AI Teacher Guidance

| | |
|--|--|
| <p>Studies show that AI facial recognition technology is least reliable for people of colour, women, and nonbinary individuals.</p> | <ul style="list-style-type: none"> • These impacts might vary depending on how the technology is being used. • If this technology is being used on a personal device, it may be less effective for these groups, causing inconvenience, frustration or a sense of exclusion • If the technology is used in law enforcement or security, this can lead to the people in these groups being unfairly targeted, profiled or facing legal consequences, even if they haven't done anything wrong |
| <p>Voice recognition technology in conversational AI systems can show bias against certain accents or dialects. For example, AI assistants might struggle with non-native speakers or regional accents. There have also been cases of it struggling with more high-pitched or breathier voices, meaning it can more frequently misunderstand female users.</p> | <ul style="list-style-type: none"> • This can make these technologies more difficult to use for certain people, and lead to a frustrating user experience • It can perpetuate the idea that those with more of a neutral accent is the societal norm, and that people with regional accents or dialects are considered as less important, and not the priority when it comes to technology design. • It also perpetuates the idea of a male-centric and male-dominated world, where men are the priority when it comes to user experience |
| <p>A study found that, even after being programmed with anti-racism training, AI was still making prejudiced assumptions about people who speak African American English. The AI model characterised African American English speakers as “suspicious”, “aggressive”, “loud” and “rude.”</p> | <ul style="list-style-type: none"> • This perpetuates racist perspectives and can cause hurt and distress to African American people seeing AI targeting them in this way, as well as potentially normalising these perspectives • This form of racism could be described as “covert” racism that might be harder to identify as racist, or that is more disguised or subtle. This may mean that people fail to recognise these examples as racist, erasing or minimising the lived experience and oppression of marginalised groups. |



Bias in AI cards

In healthcare, medical studies have mainly focused on white men, which means that there is less information about women or minority groups and how they may respond to medication or how they may experience symptoms.

AI systems in healthcare have been found to return lower accuracy results for African-American patients than white patients, for example. This can lead to further harm to their health and the wrong treatments.

Employers will sometimes use AI to quickly scan for CVs and to search for and advertise to ideal candidates. However, certain words in these job descriptions may attract or favour men more than women.

This can give men an unfair advantage for this job, and may also result in a workforce being mainly male.

AI image generators can create an image of something quickly after being given a prompt (e.g. a doctor). However, they often reproduce harmful stereotypes. For example, job roles like lawyers, doctors and judges will often be shown as white men. It may also generate images of specific groups of people (e.g. LGBTQ+ people) in a stereotypical way (e.g. gay men as physically fit or wearing 'feminine' clothing, gay women as having nose rings and shaved hair, and bisexual people and nonbinary people with dyed hair).

This can reinforce harmful stereotypes about people, and can make those groups being misrepresented by AI feel insecure and unseen.

AI-powered predictive policing tools may be used to predict areas where crime is more likely to happen. These often rely on arrest data from the past. This past data may have included incidents where people of colour were unfairly targeted and prosecuted as a result of prejudice.

This can reinforce existing patterns of racial profiling and disproportionate targeting of minority communities, even if they haven't done anything wrong.

Studies show that AI facial recognition technology is least reliable for people of colour, women, and nonbinary individuals.

This can lead to these people in these groups being unfairly targeted by law enforcement and facing legal consequences, even if they haven't done anything wrong, which can also lead to distress and to them being treated badly.

Voice recognition technology in conversational AI systems can show bias against certain accents or dialects. For example, AI assistants might struggle with non-native speakers or regional accents.

This can make these technologies more difficult to use for certain people, and lead to a frustrating user experience.

A study found that, even after being programmed with anti-racism training, AI was still making prejudiced assumptions about people who speak African American English. The AI model characterised African American English speakers as "suspicious", "aggressive", "loud" and "rude."

This perpetuates racist thoughts and can cause hurt and distress to African American people seeing AI targeting them in this way, as well as potentially normalising these perspectives.

Activity: AI manifesto

Time:

45 minutes, with time afterwards to collect information and complete their manifesto.

Learning objective:

- I can support others to use AI safely and responsibly.

Materials:

- Accompanying slides for AI manifesto. (slides 28 to 34)
- AI manifesto template. (optional)

Suggested activity guidelines:

This activity allows learners to gather youth voice from other learners in their setting and amplify what they have found throughout the community.

Note to educators:

This activity requires learners to talk to other young people in your setting so make sure to follow any of your setting's safeguarding procedures that may relate to this. It is important to consider how best to facilitate these peer-to-peer discussions and what steps you will take to protect the young people in your setting, e.g. ensuring a member of staff is always present. You can consult with the designated safeguarding lead/person to help you with this.

1. Display slide 29 and ask learners who they think should have input on how new technology is developed and the rules around this. They may suggest ideas like the developers and governments. Draw out that the people who use the technology should be able to give feedback on it, and hence if a technology is being used by young people, they should be able to have a say.
2. Display slide 30 and explain that they have a right to be listened to on issues (including technology) that affect them. The slide shows the parts of the UN Convention on the Rights of a Child that relate to this. Explain to learners these rights were written by the UN and came into effect in the UK in 1992. Read learners the rights they have and explain that it also says that the best interests of a child must be a top priority in all decisions and actions that affect children. When this was written, they didn't know how the online world would play such a big part in children's lives, so in 2021, the UN also decided that all children's rights apply in the digital world too.

Activity: AI manifesto

Suggested activity guidelines:

3. Display slide 31 and explain to learners that they are going to gather young people's thoughts and experiences of AI and any support they want. They will use this to create a manifesto of what the future of AI should look like, which they will have opportunity to share more widely.

NB: You may wish to explain further that a manifesto is a collection of thoughts or actions about a topic and show them an example of a **children's manifesto** for the future of AI produced at a summit hosted by the **Children and AI Team at The Alan Turing Institute**.

4. Display slide 32. Explain that the first thing learners will need to do is find out how young people in your setting feel about AI and any support they would want. You may wish to organise which young people they will be speaking to for them or get their ideas of how they would talk to other young people in your setting. They then need to decide what questions they want to ask them. The slide gives ideas of topics if they need some support with this.
5. Once learners have decided what questions they will ask, they should then decide how they will share the information they find. Display slide 33 which has some ideas how they may want to share this information. Remind them that they should think about which audiences they want to see it, and how best to reach them.
6. Display slide 34 and remind learners that when they are talking to other young people, if somebody says something that makes them worried, or suggests they may be at risk of harm, then they shouldn't keep it a secret, they should report it to you or to your safeguarding lead/person and talk to a trusted adult if they need more support with what they heard. You can edit slide 34 to include the name of your safeguarding lead.
7. Learners will then require further time to gather their youth voice from other learners in your setting and create their manifesto.

Ways to support and scaffold

Provide learners with the manifesto template to collect their ideas. This has questions to guide them with what they could ask.

Ways to challenge and extend

Ask learners to decide two audiences they would like to see their manifesto e.g. parents and carers, and the government. They should then create two versions of their manifesto that are pitched at each of their audiences. They should consider what information would be most useful for each audience, and if they need to change their tone.



AI manifesto template

..... Manifesto on the Future of AI

What do young people enjoy about AI?

.....

.....

.....

.....

.....

What are young people's concerns about AI?

.....

.....

.....

.....

.....

What changes would young people like to see in AI use and development?

.....

.....

.....

.....

.....

Who could help young people with AI? What could they do?

.....

.....

.....

.....

.....