Raspberry Pi Foundation

AI safety – Media literacy in the age of AI

This guide and the associated slides contain all the guides to the activities you will need to run an AI safety session on the topic of media literacy in an AI-powered world.

This document is not designed to be read from start to finish, it is recommended that you read the introduction and outline that follows and then use the following table of contents to jump to the documents for the activities.

Please give us your feedback!

We'd love to hear how you have used the Experience AI resources and what you thought about them. After using the resources, please take a few minutes to:

- Share your feedback in our user survey: rpf.io/exai-2mf
- If you are an educator, ask your learners to complete a short survey: rpf.io/exai-st

Your feedback supports us to make our AI resources accessible to everyone, and we really appreciate you giving your time to share what you think.

Introduction

The objective of this session is to highlight the role of both users and AI tools in perpetuating misinformation, but also to consider the ways AI applications can help combat misleading claims.

Misleading information is not a problem specific to AI, but this technology has the potential to accentuate existing media literacy problems. There are also many ways that AI might help shed light on the truth of situations. In this session, young people reflect on the roles of different stakeholders when misinformation involves AI tools and the strategies they have to confirm information they see online.

Learning objectives

- Describe different types of media that generative AI tools can produce
- Determine how generative AI will affect the need to check information before sharing it
- Build a set of expectations of fairness, accountability, and transparency around AI content on a social platform

Key vocabulary

Generative AI, misinformation, disinformation, fact-checking, prompt, bias, deep fakes

Preparation

You should be confident discussing online misinformation and the reasons someone might want to influence the opinions of others. The subject knowledge delivered in the video, that generative AI can be used to create content and is not always factually accurate, will be enough to lead this session.

Learners should be familiar with the idea of AI — that it stands for artificial intelligence, and that it's a type of computer system that's becoming more common — but do not need a precise definition to access the learning in this session. Optionally, you could complete Lesson 1 of the Foundations of AI unit to provide an introduction to the topic.

Activities for this topic

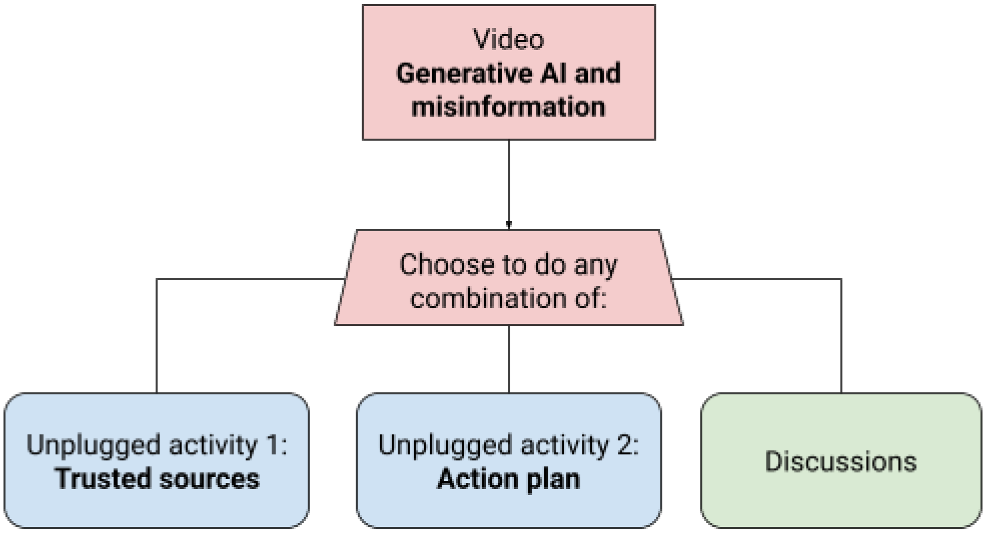
Activity	Description	Suggested timings
Video	Generative AI and misinformation	3 mins
	Short recap activity for concept in the video	5 mins
Unplugged	Trusted sources – Young people review the sources they use for information, and whether they would allow the use of AI	25 mins
	Action plan – Young people decide what to do with information they find online, and whether transparency in AI use would change their minds	25 mins
Discussion topics	A series of discussions on the impact AI tools will have on media literacy and the spread of misinformation	10–30 mins

Combining activities for your session

It is recommended that every session begin with the Generative AI and misinformation video.

After that you can either:

- Choose a topic to discuss with the learners
- Complete one or both of the unplugged activities



Video
Generative AI and misinformation

Choose to do any combination of:

Unplugged activity 1:
Trusted sources

Unplugged activity 2:
Action plan

Discussions

Example activity combinations

For a 30 minutes session you can complete the following activities:

- Generative AI and misinformation (8 mins)
- Action plan (remove two examples) (15 mins)

For a 60 minute session

- Generative AI and misinformation (8 mins)
- Trusted sources (25 mins)
- Discussion (~15 mins)
    - Who is responsible for fact checking information in various types of media?

Raspberry Pi Foundation

Generative AI and misinformation – Video activity guide

Introduction

This short animation will serve as a launch pad for your learners to explore the ways generative AI tools will change the ways they interact with information.

The video explains the concept of generative AI, and then poses questions about the ways these tools might be used to create misinformation — either accidentally or intentionally to sway the learners opinions.

Key vocabulary

Generative AI, misinformation

Preparation

This activity requires learners to watch a video, either all together or on their own devices. The video is hosted on YouTube — if this is blocked, you can use the downloaded version.

You will need:

- Slides (4–9)

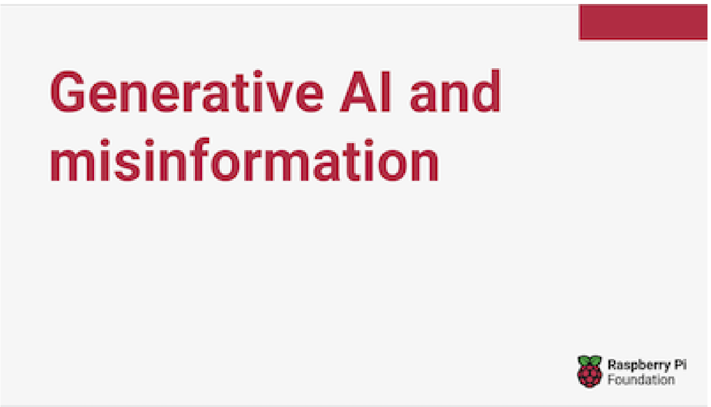
Outline plan

**Timings are rough guides. Adjust to suit your environment.*

Introduction (Slides 4–5) – 2 minutes

Display: Slide 4

Explain: That the learners are going to watch a video about the ways AI tools might change the way they think about online information.

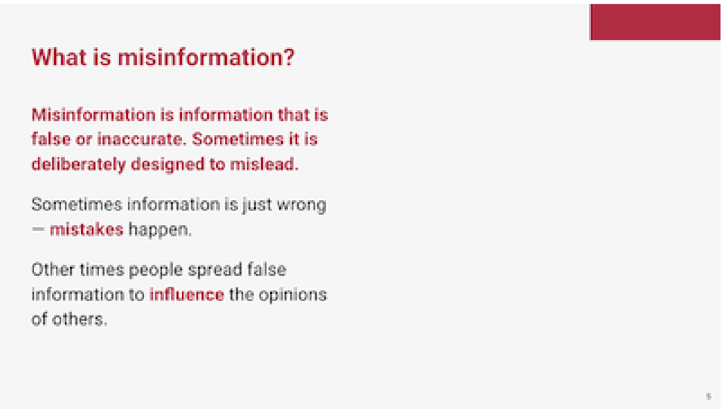


Generative AI and misinformation

Raspberry Pi Foundation

Display: Slide 5

Explain: What misinformation is and how it spreads, both unintentionally and purposefully to influence the opinions of others.



What is misinformation?

Misinformation is information that is false or inaccurate. Sometimes it is deliberately designed to mislead.

Sometimes information is just wrong – mistakes happen.

Other times people spread false information to influence the opinions of others.

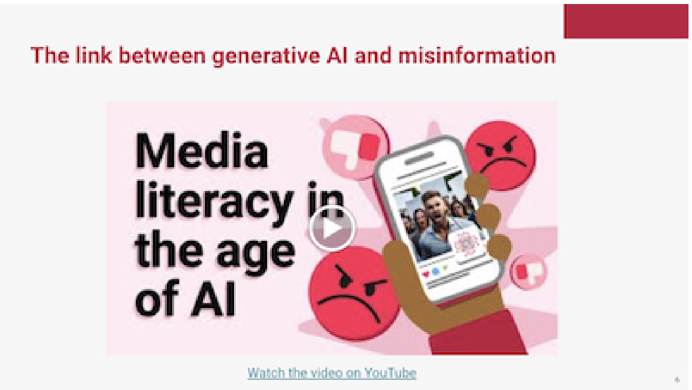
Guidance

Having a shared definition of misinformation will be useful for the learners throughout the rest of the activities. It also serves as a bit of familiar knowledge amongst new concepts in the video.

Video (Slide 6) – 3 minutes

Display: Slide 6

Do: Play the video for the learners, or have them watch the video on their own devices.



The link between generative AI and misinformation

Watch the video on YouTube

Guidance
Make sure the video is in your local language, as there are versions with translated voiceover to engage your learners as much as possible.

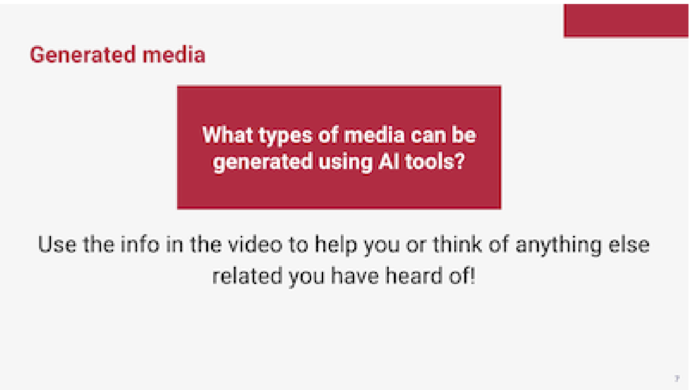
While watching the video have learners consider the following questions:

- What is misinformation, and how can AI be used to create it?
- What does generative AI prioritise when producing content?
- Why might generative AI not always produce factual content?

Recap activity (Slides 7–9) – 3 minutes

Display: Slide 7

Ask: What media types can be generated using AI? Which ones were mentioned in the video? Are there any others?



Generated media

What types of media can be generated using AI tools?

Use the info in the video to help you or think of anything else related you have heard of!

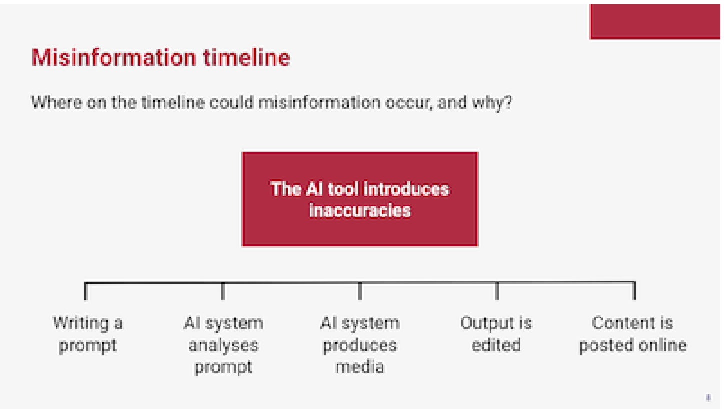
Guidance
Your learners might have more or less experience with generative AI — for less experienced learners, focus on **formats** (text, images, and video) to really make sure they heard it.

For more experienced learners, you might expand to **types** of media (blogs, social media videos, books, etc.).

Display: Slide 8

Explain: The timeline at the bottom of the slide shows the process of generating content with an AI tool. Walk learners through the timeline.

Ask: Where on the timeline could the AI tool introduce inaccuracies?



Misinformation timeline

Where on the timeline could misinformation occur, and why?

The AI tool introduces inaccuracies

Writing a prompt

AI system analyses prompt

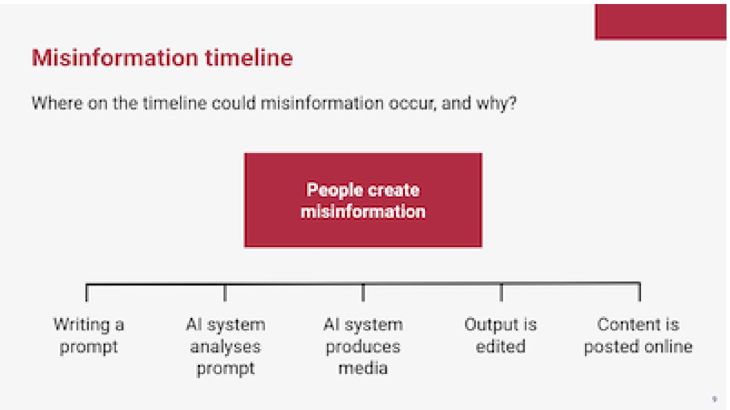
AI system produces media

Output is edited

Content is posted online

Display: Slide 9

Ask: Where on the timeline could a person use generative AI to create misinformation?



Misinformation timeline

Where on the timeline could misinformation occur, and why?

People create misinformation

Writing a prompt

AI system analyses prompt

AI system produces media

Output is edited

Content is posted online

Guidance

The idea here is for young people to reflect on the process of using generative AI tools to produce media, and also the points of the process they need to be careful when using these tools or when they come across generated content.

The AI tool might introduce inaccuracies while analysing the prompt or producing the media — so learners should be careful to double check any outputs.

People might create misinformation when writing the prompt — any information in the prompt will definitely be included in the output. People might also edit the outputs to contain misinformation after generation, or post them online in such a way that incorrect facts are presented as true.

Raspberry Pi Foundation

Unplugged activity 1: Trusted sources – Activity guide

Introduction

In this unplugged activity, learners will reflect on the ways they look for information online and where they find it. At the end, they will decide which sources are most likely to allow the use of generative AI and how that relates to how trustworthy they are.

Key vocabulary

Information sources, trust, generative AI, misinformation

Preparation

This activity requires you to conduct a discussion about the trustworthiness of different sources of information, considering the oversight of the content on each source learners come up with.

You will need:

- Slides (10–20)
- Small cards (flash cards, cardboard, anything about credit card size)

Adaptation

Breaking the activity up: There is the option to split the activity when the learners have finished the fourth type of information (Slide 17) and continue the discussion in the next lesson.

Shortening the activity: You could shorten this activity by only using 2 of the examples of information, picking the 2 that would provoke the most discussion with your learners.

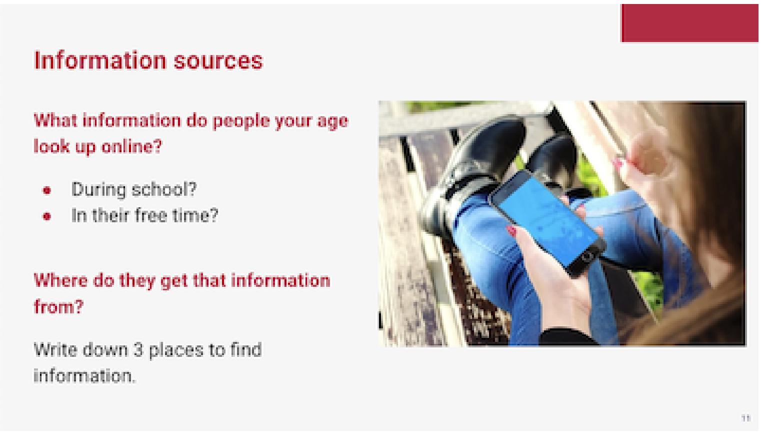
Extension: You can extend this activity by adding your own examples of information and having your learners decide whether they would trust their sources for that information.

Outline plan

**Timings are rough guides. Adjust to suit your environment.*

Introduction (Slides 10–12) – 10 minutes

Display: Slide 11



Information sources

What information do people your age look up online?

- During school?
- In their free time?

Where do they get that information from?

Write down 3 places to find information.

Ask: What kind of information do people your age search for online?

Do: Have learners write down information types for:

- School
- In their free time

Ask: Where do they look for that information?

Do: Get learners to write their top 3 sources for information.

Guidance

Help the learners by prompting them with different situations that might require them to look up information online — this is a primer for the rest of the activity, so you want them to be able to readily recall times they looked for information.

The sources that they come up with can be general or specific — “social media” is just as valid as a particular platform. This activity should be about them thinking critically about the sources they already use and how they label those should be led by them. They might say news websites, blogs, social media platforms or specific creators, Wikipedia, etc.

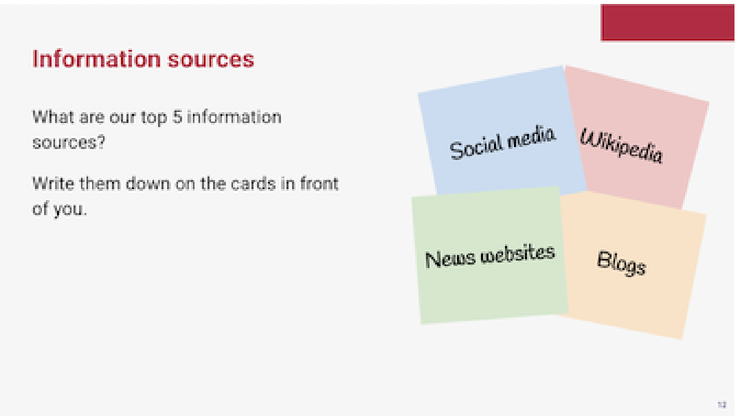
If your learners suggest very specific sources (a specific site, or social media creator, for example), you might ask them if their choice of whether to trust them or not applies to other creators or other platforms — encourage them to use their specific examples to explore their opinions about general information sources.

Display: Slide 12

Ask: For examples of information sources from learners. After each one, ask for a show of hands of others who had that same source.

Pick: The top 5 answers most common across the classroom.

Do: Have learners write each of them on the cards in front of them.



Information sources

What are our top 5 information sources?

Write them down on the cards in front of you.

Social media

Wikipedia

News websites

Blogs

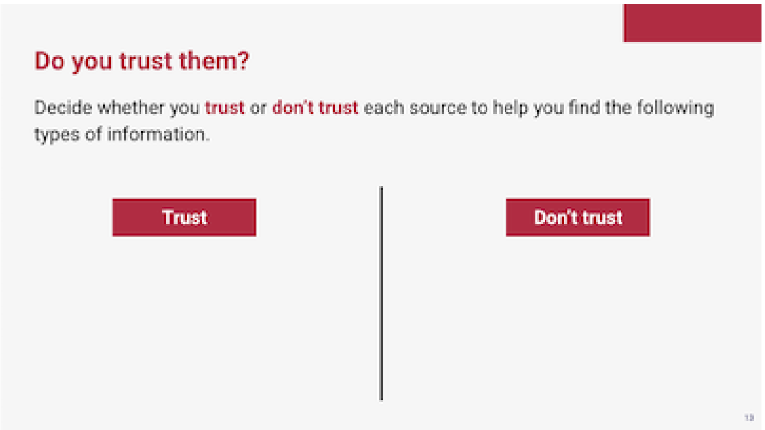
12

Guidance
The objective here is to get 5 sources that the majority of the learners can relate to, so don't worry too much about counting votes and mathematically working it out. If most hands go up for the first 5 mentioned, then just use those.

Trust or don't trust (Slides 13–17) – 10 minutes

Display: Slide 13

Explain: For each of the following types of information, the learners are going to sort their sources into two columns — **trust** or **don't trust** — indicating whether they think it is a good place to get accurate info. They can put them in the middle if they are unsure.



Do you trust them?

Decide whether you trust or don't trust each source to help you find the following types of information.

Trust

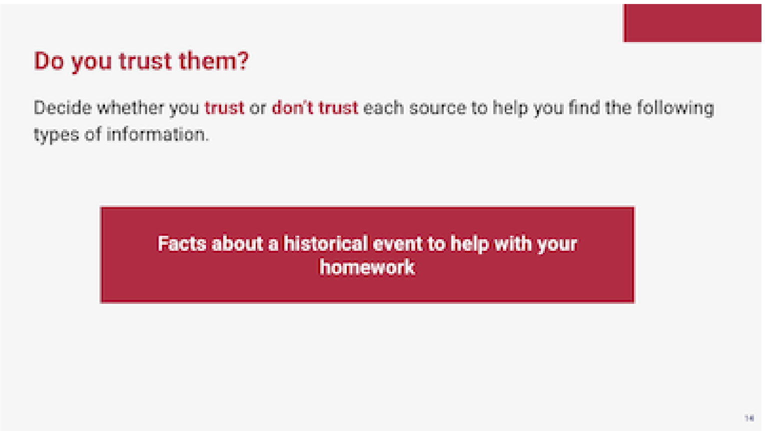
Don't trust

13

Guidance: Have learners write **Trust** and **Don't trust** on two cards to signify their columns.

For example, they might say they trust a news organisation, don't trust a celebrity's social media page, but a source like Wikipedia might be placed in the middle (trusted somewhat, but not entirely).

Display: Slide 14



Do you trust them?

Decide whether you trust or don't trust each source to help you find the following types of information.

Facts about a historical event to help with your homework

Do: Ask the learners to sort their sources into the **trust** and **don't trust** columns for finding facts about a historical event for their homework.

Ask: For each source, choose one person and ask which column they put it in. Give others the chance to disagree and explain their thinking. Repeat this process for all 5 sources.

Repeat: Do this for slides 15, 16, and 17.

Guidance

Putting sources in between the two columns is valid as sometimes it is not a clear-cut choice whether to trust something. You might probe and ask what else they could do to check the information if they did decide to trust the source for that kind of information.

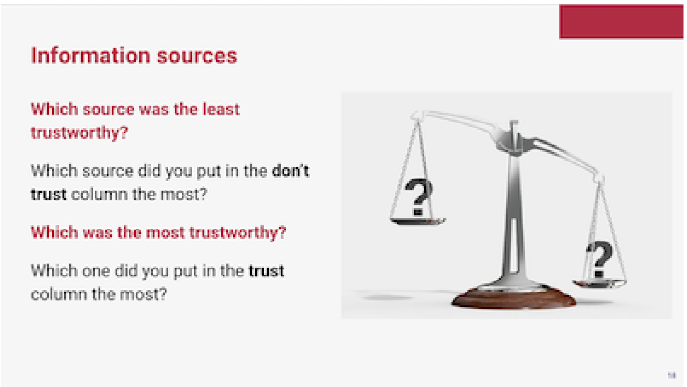
Encourage learners to make their first choices quickly. This replicates real life when we often make snap decisions about whether to trust something. Encourage them to change their minds after thinking about it or hearing arguments — this is an important skill when navigating online information.

Trust and generative AI (Slides 18-20) – 5 minutes

Display: Slide 18

Ask: Which source appears in the **don't trust** column the most or which is their least trustworthy source.

Ask: Which source they had in the **trust** column the most. You could also ask why.



Information sources

Which source was the least trustworthy?

Which source did you put in the **don't trust** column the most?

Which was the most trustworthy?

Which one did you put in the **trust** column the most?

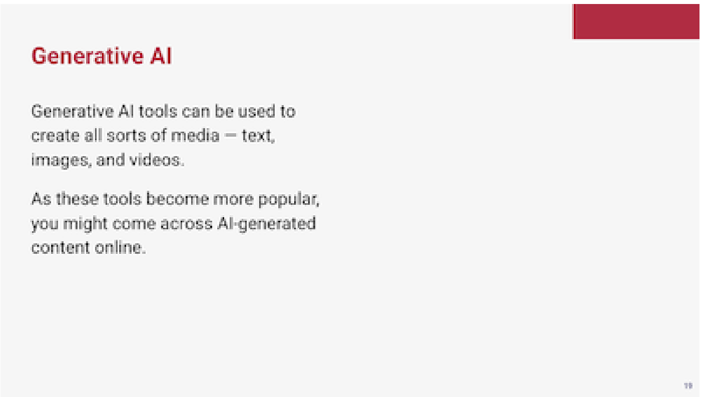
Guidance

This is a great time to get the learners to reflect on the sources they mentioned at the beginning. Do they actually trust them as much as they initially thought?

This might be a product of the examples they have looked at — no source should be considered above reproach and universally trustworthy.

Display: Slide 19

Explain: That generative AI tools can be used to create all kinds of media and that they might see some AI-generated content online.



Generative AI

Generative AI tools can be used to create all sorts of media – text, images, and videos.

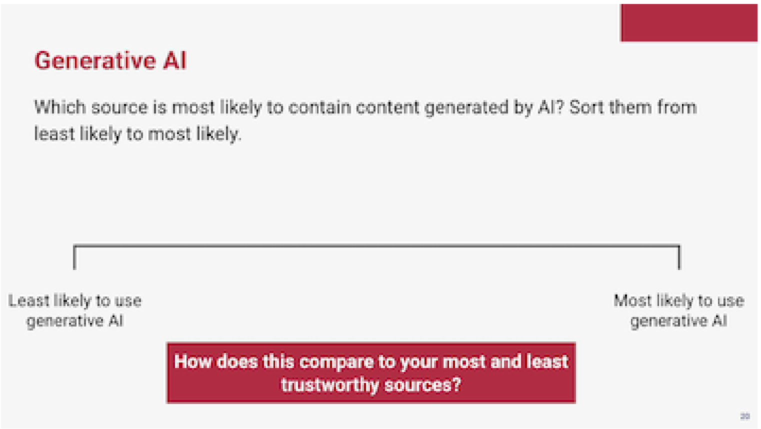
As these tools become more popular, you might come across AI-generated content online.

Display: Slide 20

Do: Get the learners to order the sources from least likely to most likely to use generative AI.

Ask: Learners to help create a class ranking. Ask why they put certain sources where they did.

Ask: Whether the ranking is related to how trustworthy a source is?



Generative AI

Which source is most likely to contain content generated by AI? Sort them from least likely to most likely.

Least likely to use generative AI
Most likely to use generative AI

How does this compare to your most and least trustworthy sources?

Guidance
It might be worth explaining that oversight is an important factor in determining the trustworthiness of content published online. News organisations have standards they must stick to whereas social media accounts are more free to post anything they like. The risks in publishing something inaccurate might play into whether a source is likely to allow the use of generative AI.

Raspberry Pi Foundation

Unplugged activity 2: Action plan – Activity guide

Introduction

In this unplugged activity, learners will reflect on how AI might change the way they interact with information they come across online. They will decide whether to share, double-check, or ignore a piece of information they find online. In an AI-powered twist, a fictional AI-checking company will tell them how generative AI was used in creating that content.

Suggested activity length

25 minutes

Key vocabulary

Information sources, trust, generative AI, misinformation

Preparation

Confidence in conducting a discussion about the reasons why a young person might share a piece of information. Knowledge of how to double-check information and the appropriate time to do so.

You will need:

- Slides (21–35)

Adaptation

Breaking the activity up: There is the option to split the activity in the middle of the examples of online information (Slide 30), and then continue the activity in the next lesson.

Shortening the activity: You could shorten this activity by only using 2 of the examples, picking the 2 that would provoke the most discussion with your learners.

Outline plan

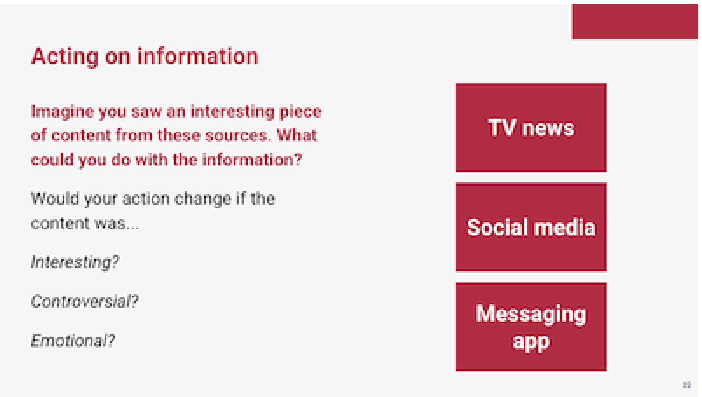
Part 1 (Slides 22–23) – 10 minutes

Display: Slide 22

Explain: That learners come across information in their lives all the time, from sources like those on the right.

Ask: What actions they take with the information they find in the three sources on the slide: TV news, social media, and messaging apps.

Ask: If their choice of action would change if the content was: interesting, controversial, or emotional.



Acting on information

Imagine you saw an interesting piece of content from these sources. What could you do with the information?

Would your action change if the content was...

Interesting?

Controversial?

Emotional?

TV news

Social media

Messaging app

22

Guidance

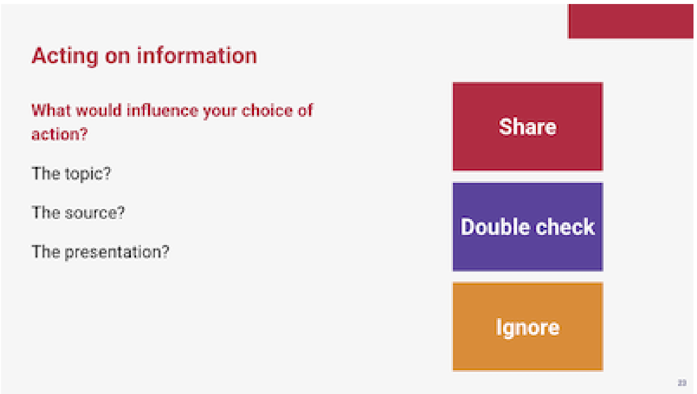
Let the learners explain how they deal with information, and inquire as to what makes them share things with their friends. For every action they suggest, discuss what makes them do that — what types of information receive that treatment? Some example actions might be: send to a friend, ignore, like, comment, read, copy for schoolwork.

Display: Slide 23

Explain: That they are going to use three simple actions for the rest of the activity.

Ask: What would influence their choice of what to do with the information?

Do: Get learners to write a rule for each of the three actions.



Acting on information

What would influence your choice of action?

The topic?

The source?

The presentation?

Share

Double check

Ignore

23

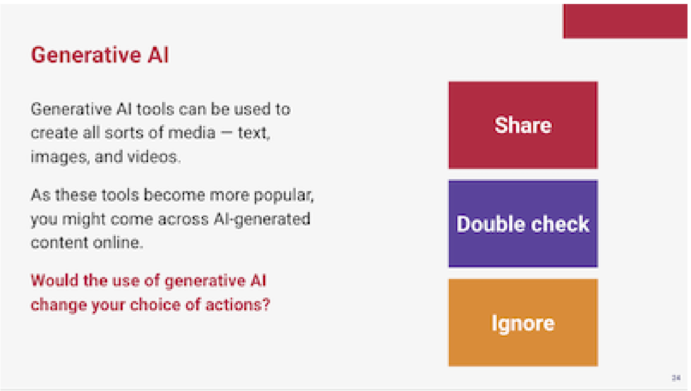
Guidance
For simplicity, the options have been reduced to three broad categories: share, double-check, or ignore. Most things the learners suggest can be simplified to these. For their rules, have them create if…then… style rules about the types of information they would do for each of those actions.

Example rules:
If the information is important, **then** I would share it with a friend.
If the source seems untrustworthy, **then** I would double-check it.

Display: Slide 24

Explain: Generative AI tools can be used to make media, and some of the content they see online might use generative AI.

Ask: If the use of generative AI tools would change the way they act on information and why?



Generative AI

Generative AI tools can be used to create all sorts of media — text, images, and videos.

As these tools become more popular, you might come across AI-generated content online.

Would the use of generative AI change your choice of actions?

Share

Double check

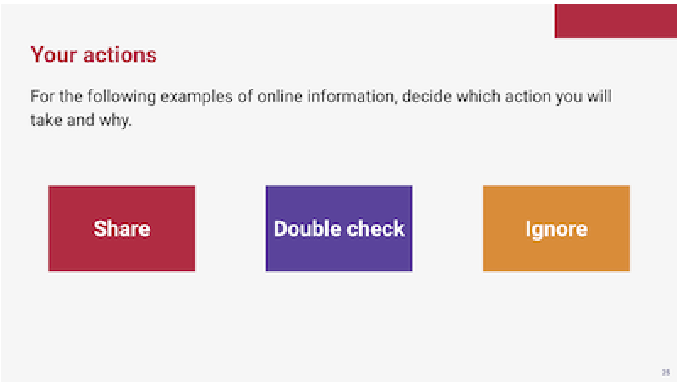
Ignore

24

Actions and AI (Slides 25–34) – 10 minutes

Display: Slide 25

Explain: For each of the examples of online information, the learners will decide which action to take and come up with a reason why.



Your actions

For the following examples of online information, decide which action you will take and why.

Share

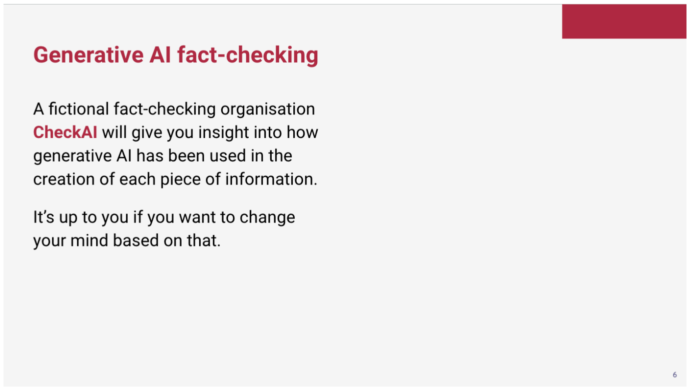
Double check

Ignore

25

Display: Slide 26

Do: Explain that after each piece of information, the learners will get insights from a fictional "AI tool check" company — CheckAI — detailing how generative AI was used in the creation of the information. They will then have a chance to change their decision.



Generative AI fact-checking

A fictional fact-checking organisation **CheckAI** will give you insight into how generative AI has been used in the creation of each piece of information.

It's up to you if you want to change your mind based on that.

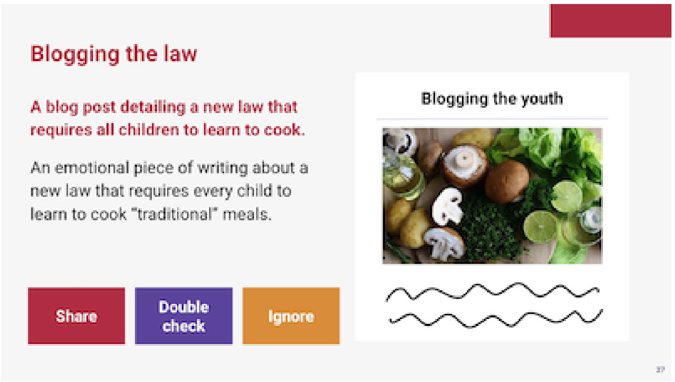
6

Display: Slide 27

Ask: What would you do with a blog post about a new law requiring every child to learn to cook?

Do: Get the learners to make their decision using their rules.

Ask: Learners to share their answers.



Blogging the law

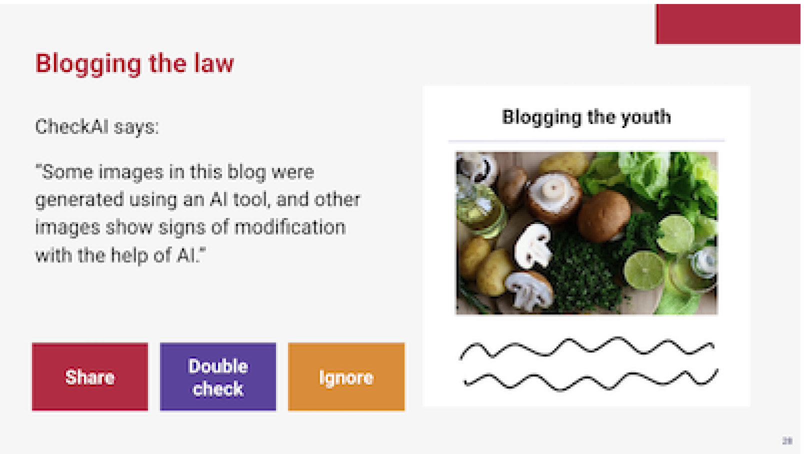
A blog post detailing a new law that requires all children to learn to cook.

An emotional piece of writing about a new law that requires every child to learn to cook "traditional" meals.

Blogging the youth

27

Display: Slide 28



Blogging the law

CheckAI says:

"Some images in this blog were generated using an AI tool, and other images show signs of modification with the help of AI."

Explain: That the blog post used AI-generated images and images edited using AI.

Do: Learners can decide to change their vote.

Ask: Learners if their choice has changed and why.

Repeat: This process with slides 29 and 30, 31 and 32, and 33 and 34.

Guidance
You can get answers however works for you:

- Whiteboards
- A show of hands
- Have learners write them down
- Learners stand in the area that reflects their views

The important part is to get a consensus.

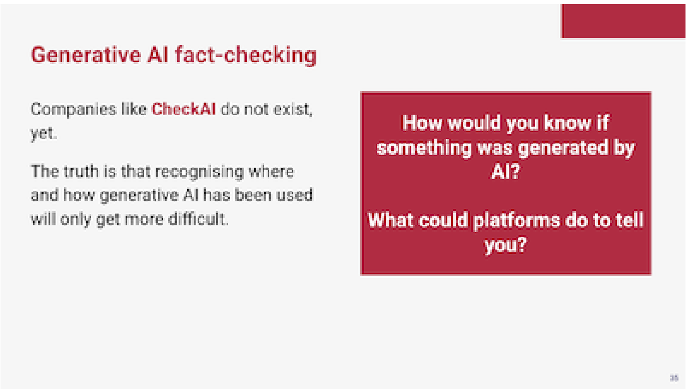
You should choose at least one person for each answer and get their reasoning. When the learners get the chance to change their minds, make sure to pick someone who changed their opinion to find out why.

As each use of generative AI is revealed, discuss the impact on the **credibility** of the information in the content. Do generated images used to add decoration to a blog post make the information untrustworthy? How about using AI to create graphics that contain data in research?

Young people need to decide their own line for trusting generative AI — this is a personal choice and cannot be dictated for them.

Transparency (Slide 35) – 5 minutes

Display: Slide 35



Generative AI fact-checking

Companies like **CheckAI** do not exist, yet.

The truth is that recognising where and how generative AI has been used will only get more difficult.

How would you know if something was generated by AI?

What could platforms do to tell you?

35

Explain: There is no company like CheckAI, and recognising where generative AI has been used will only get more difficult as AI tools improve.

Ask: How would the learners know if generative AI was used?

Ask: How could platforms tell you where and how AI has been used?

Guidance

Ask them how easy they think it would be to spot AI-generated content online? How could they know? They might say it is easy to spot, in which case remind them the tools are always improving. They might not have any idea how to spot AI-generated content, in which case they are like most people and transparency is even more important.

This is their chance to come up with ideas for how they would like to see transparency on platforms — how would they want the information to be presented? What do they want to know about the use of generative AI? Remind them of misinformation tags that already exist, and ask them if they would be effective.

They might not mind or want transparency, so you might have to give them some emotive examples (such as an article about their favourite food being discontinued, or a photo in a school newspaper spreading gossip) to showcase why sometimes you definitely want to know if AI has been used.

Discussion guidance

Potential discussion topics:

- Who creates misinformation, and why might someone want to sway your opinion?
- Who is responsible for fact checking information in various types of media?
- What advice would you give someone who wants to use an AI tool to help them draft an outline for an essay?
- What advice would you give someone who wants to use an AI tool to create a poster?

Draw out

Within this session, draw out the following with the learners:

1. The need to separate fact from fiction.
2. The need for transparency by companies and individuals in declaring what content has been AI generated.
3. The need to be a critical consumer of digital artefacts.

Key message

The key message for this session is:

1. Be a critical consumer in order to separate fact from fiction.
2. Be an ethical consumer.