

AUSTRALIA

Lesson 4: Ethics and the Responsible Use of AI

2024 JUNIOR PROGRAM

LESSON PLAN

DAYOFIAUSTRALIA.COM

Lesson 4: Ethics and the Responsible use of AI | 60 mins

Lesson Summary

In this lesson, students reflect on the ethical considerations surrounding AI usage. They explore societal impacts, analyse AI-generated content in mainstream media, considering biases, plagiarism and responsible AI use. Through discussions and case studies, they examine the perspectives of various stakeholders in the context of AI-generated art, explore the limits of AI and build an understanding of the ethical implications of AI, fostering critical thinking skills and ethical awareness.

Objectives

- Students will explore **Societal Impact**, the fifth **Big Idea of AI**
- Students will build an awareness of AI-generated content in mainstream media
- Students will reflect on the responsible use of AI by applying ethical concepts like 'harm', 'benefit', 'misrepresentation' and the perspective of multiple stakeholders
- Students will learn about the limits of AI

Curriculum Alignment

This lesson is linked to the following [Australian Curriculum \(Version 9\)](#) content descriptors:

- **Years 5 and 6**
 - [AC9M5P01](#): list the possible outcomes of chance experiments involving equally likely outcomes and compare to those which are not equally likely.
 - [AC9S6H01](#): examine why advances in science are often the result of collaboration or build on the work of others.
 - [AC9TDI6P10](#): explain the creation and permanence of their digital footprint and consider privacy when collecting user data.
- **Years 7 and 8**
 - [AC9M8P01](#): recognise that complementary events have a combined probability of one; use this relationship to calculate probabilities in applied contexts.
 - [AC9M8ST01](#): investigate techniques for data collection including census, sampling, experiment and observation, and explain the practicalities and implications of obtaining data through these techniques.
 - [AC9M8ST02](#): analyse and report on the distribution of data from primary and secondary sources using random and non-random sampling techniques to select and study samples.
 - [AC9TDI8P10](#): evaluate existing and student solutions against the design criteria, user stories and possible future impact.
 - [AC9TDI8P14](#): investigate and manage the digital footprint existing systems and student solutions collect and assess if the data is essential to their purpose.


Vocabulary

- **Ethics**, n. the study of morality and values, specifically what is 'right' and 'wrong'
- **AI Ethics**, n. the application of ethics to AI research and development
- **Beneficial**, adj. something that results in a positive effect for a person, or a group of people
- **Harmful**, adj. something that results in a negative effect for a person, or a group of people
- **Plagiarism**, n. taking someone else's work or ideas and passing them off as your own
- **Misinformation**, n. false or inaccurate information, especially that which is deliberately intended to deceive
- **Bias**, n. a preference or a prejudice towards someone or something, especially in a way considered to be unfair or unrepresentative

Resources

- [Presentation Slides for lesson](#) (available on website once logged in)
- [Puppies video](#) (embedded in the lesson slides)
- [Art stakeholders handout](#) (optional, for students to reference)

Activity Steps

1. **1 min.** Introduce **Societal Impact** as the fifth **Big Idea of AI**.
2. **1 min.** Set up a simple model of ethics using harm and benefit. Harm is something that negatively affects people, benefit is something that positively affects people.
3. **2 mins.** Art gallery scenario. Walk through the example of using AI to generate art rather than commissioning local artists. Point out that the harms and benefits in this scenario apply to different people of interest (the museum, local artists, the visitors to the museum).
4. **4 mins.** Ask students in pairs to discuss whether the way you use technology influences whether or not it is harmful or of benefit.
 Feel free to give an example, such as using the internet to research things you want to learn more about versus using the internet to illegally download movies.
5. **1 min.** Introduce the concept of 'perspective taking', and the idea of a **stakeholder**. Return to the plagiarism example. Point out again how the harms and benefits apply differently to different **stakeholders**.
6. **2 mins.** Step through the slides showing the two art pieces. One of the pieces was created by a human artist and the other was created by AI. Ask students which one they think is which.
7. **1 min.** Reveal the answer: the image on the left is "[Crystalline Maples](#)", a 2021 oil painting by [Erin Hanson](#). The image on the right was generated by an AI called Stable Diffusion, created by a company called [Stability AI](#).
8. **4 mins.** Define plagiarism and ask students the following questions:
 - a. Do you think AI-generated art is plagiarism? Why, or why not?
 - b. Does your answer change if the artist knows their work is being used to train AI?
9. **4 mins.** Introduce the concept of 'bias', and data/algorithmic bias in AI, giving an example of data bias in generating images of people with glasses.
10. **8 mins.** Introduce three imaginary stakeholders in the art industry. In pairs or small groups, students should select one person of interest out of the three stakeholders on the worksheet. For the stakeholder they have chosen, they should discuss:



What sort of things would the stakeholder **care** about?



What **concerns** might the stakeholder have about generative AI?



How does the stakeholder stand to **benefit (positive)** from generative AI?



What might be a potential **harm (negative)** from generative AI?



How could the stakeholder be affected by potential **bias** in generative AI?



NOTE

This handout transcribes each of the stakeholders, which you can print and share with students in advance if you would like them to have a physical copy to refer to.

11. **6 mins.** Have one student from each group act out their stakeholder and talk through their perspective on AI technology. They can speak freely or follow a script that could look like:
 - a. *I think this technology could be beneficial/harmful for me because...*
 - b. *From my point of view, this technology should/should not be used because...*
12. **1 min.** [Play this video of puppies](#) generated by [OpenAI's Sora model](#). Try and avoid showing the video's title, and don't share that it's made by AI just yet.
13. **2 mins.** At the conclusion of the video, ask students to discuss in pairs the following:



Do they think the video is real? If so, what features appear real?



Do they think the video is AI? If so, what features seem artificially generated?

14. **4 mins.** Reveal that the video was produced using a [new AI model called Sora](#), created by OpenAI (the same company that created ChatGPT). The AI model can create realistic and imaginative scenes from a text-prompt.
 - a. The text-prompt for the puppies video was *"a litter of golden retriever puppies playing in the snow. Their heads pop out of the snow, covered in."*
 - b. A more complex and specific text prompt generated the 2nd example, where *"A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a black leather jacket, a long red dress, and black boots, and carries a black purse. She wears sunglasses and red lipstick. She walks confidently and casually. The street is damp and reflective, creating a mirror effect of the colourful lights. Many pedestrians walk about."*
15. **8 mins.** In the same pairs or small groups they were in before, ask students to revisit the stakeholder they chose earlier and think about how Sora might affect them using these discussion points:
 - a. What **concerns** might they have about Sora?
 - b. How do they stand to **benefit (positive)** from Sora?

- c. What might be a potential **harm (negative)** from Sora?
 - d. How could they be affected by potential **bias** in Sora?
16. **5 mins. CASE STUDY:** Researchers at the [University of Queensland \(UQ\)](#) have developed an AI tool which has been trained on input from human experts and trusted datasets to identify misinformation. When the system looks at a new piece of information, it classifies it as potential misinformation or not, which is then confirmed and refined by humans to ensure quality, agency and accountability.
- a. **Here's how it works:**
 - i. Humans provide expert inputs, including expert datasets like the [RMIT ABC Fact Check dataset](#), to train the AI. The RMIT ABC Fact Check dataset is information on statements made by Australian politicians and whether those statements are true, false or in-between.
 - ii. The AI can then review large amounts of online information and generate labels and explanations as to why a claim may be misleading. These labels and explanations are provided to users to help them make better choices about the information they trust and use.
 - iii. Humans are then also involved to check and refine the labels and explanations, as both experts and crowd workers (people who have volunteered to work on these sorts of projects) to help maintain the quality, agency, and accountability of the process.
 - iv. Using AI in this way means huge amounts of information can be processed - much more than could be done by humans alone.
 - b. Ask students to discuss in their groups if they think AI-generated media created with tools like Sora could be considered **misinformation**?
 - c. You may need to define **misinformation** in simpler terms for your students.
17. **6 mins.** Discuss the limits of AI and Large Language Models. Encourage students to consider other limits of AI and how they should think about using them. These are a few prompts to guide the discussion:
- a. **AI models are just making predictions.** All forms of AI are ultimately just guessing at what you want. Sometimes they're right, sometimes they're not. They don't really understand the things they're predicting.
 - b. **AI models do not always give the same answer,** even for the same prompt. The AI generates a new prediction every single time, and generating two identical responses is extremely unlikely.
 - c. **AI models can be biased.** They can be trained on data that can lead to inaccurate or skewed predictions, just like humans.

END OF LESSON PLAN