common sense education

GRADE	TIME
6–12	20 mins.

TOPICS

- News & Media Literacy
- Digital Citizenship
- Artificial Intelligence

Understanding AI Bias

How does AI bias happen?

Artificial intelligence is trained on real-world data that people have given it, and if that data contains biases (or is incomplete), the AI can end up being biased, too. In this lesson, students will think critically about the training data that informs what AI tools can do, and consider possible ways to reduce AI bias.

For more AI-related lessons, check out the **AI Literacy lesson collection**.

Students will be able to:

- Define AI bias.
- Understand how AI bias happens.
- Reflect on ways to reduce AI bias.

What You'll Need

Some resources below are available in Spanish

IN CLASS

- Lesson Slides

BEFORE THE LESSON

We encourage teaching the following lessons to help set a foundational understanding of how AI works:

- **What Is Artificial Intelligence?**
- **How Is AI Trained?**

Lesson Plan

Practice: **AI Bias** 20 mins.

1. **Say**: *When computer scientists create AI, they use two different types of data:* **training data** *and* **testing data** (**Slide 4**).

   - **Training data** *is the information given to an AI to help it learn how to do specific tasks* (**Slide 5**).
   - **Testing data** *is the information used to check whether the AI that was created is reliable and accurate* (**Slide 6**).

2. **Say**: *Imagine we are computer scientists and w* *ol we're building is to identify different types of fr* **7**).

3. **Ask**: *Based on these examples of training data*

4. **Show Slide 9** and explain that the images he AI is working properly. The labels under each i

   Ask: *Do you notice any mistakes? Why do you*

5. **Explain** that the mistakes the AI made are a *decision that is wrong or problematic because and things accurately* (**Slide 11**).

   Show Slide 12 and **say**: I*n the training data, c shows that the AI learned to identify anything toward thinking that every red fruit is an apple*

6. **Say**: *What are some ways we could reduce the AI bias of this fruit detector?* (**Slide 13**)

   Invite students to share out, and then review the suggestions on **Slide 14**.

7. **Say**: While it's almost impossible to completely eliminate AI bias from a tool, we can do our best to reduce it by coming up with as diverse and complete a set of training data as possible (**Slide 15**).

8. **If time permits**, read **Slide 16** and have students work independently to come up with a list of image descriptors. Then, have them pair up to compare their lists and continue to add any additional image descriptors.

   Review the descriptors on **Slide 17** and continue to add to the list based on any other ideas the students have.

9. **Say**: *Remember that behind every AI tool are humans making decisions on what training data the tool will use. Understanding how AI bias occurs can help us think critically about its potential impacts* (**Slide 18**).

Lesson last updated: October 2023