Creating Bias in Artificial Intelligence

RESET AN AEOP EFFORT aeop

Authors

Tracy Davis

SUBJECT/GRADE LEVEL

Subject Area: Science

Grade Level: Middle School

Materials

- Slideshow
- Skill sheets by stage (attached at end of the document and linked throughout the lesson)
- Chromebooks + chargers
- Dog/cat cards
- Account in https://teachablemachine.withgoogle.com/
- Peanut Butter, various types and brands (follow allergy procedures at your school)
- Jelly/Jam, various types, and brands
- Bread, various types and brands

STANDARDS

Next Generation Science Standards

MS-ETS1-1. Define the criteria and constraints of a design problem with sufficient precision to ensure a successful solution, taking into account relevant scientific principles and potential impacts on people and the natural environment that may limit possible solutions.

MS-ETS1-2. Evaluate competing design solutions using a systematic process to determine how well they meet the criteria and constraints of the problem

LESSON OBJECTIVE(S)

Students will train a type of AI, machine learning algorithm, using data sets then revise the learning data to reduce bias and increase accuracy. Students will recognize that new AI inventions can help people but can also have unintended effects.

CAREER CONNECTION(S)

Mentor Background: Dr. Benjamin D. Nye is the Director of Learning Sciences at the University of Southern California, Institute for Creative Technologies. Ben's major research interest is to identify best practices in advanced learning technology, particularly for frontiers such as distributed learning technologies (e.g., cloud-based, device-agnostic) and socially situated learning (e.g., face-to-face mobile use). His research interests include modular intelligent tutoring system (ITS) designs, modeling social learning and memes, cognitive agents, and educational tools for the developing world and low-resource/low-income contexts. He received his Ph.D. in Systems Engineering from the University of Pennsylvania in 2011. Ben currently leads projects focused on areas such as mobile adaptive learning ecosystems (Personal Assistant for Life-Long Learning), open-source and user-authorable tutoring chat dialogs (OpenTutor), a studio for career mentors to self-record their own virtual mentors (CareerFair.ai), and augmented reality experiences with the La Brea Tar Pits to understand complex concepts (Re-Living Paleontology) In his prior work as a Research Professor at the University of Memphis, Ben led work on the Shareable Knowledge Objects (SKO) framework integrating ITS services such as AutoTutor for the ONR ITS Grand Challenge, helped data mine effort a corpus of 250k human-to-human online tutoring dialogs (part of the ADL PAL initiative), collaborated on ONR's PAL3 tutoring architecture for supporting life-long learning, and is an advisor and book editor for the ARL Generalized Intelligent Framework for Tutoring (GIFT) advisory panel. Ben's research tries to remove barriers to the development and adoption of ITS so that they can reach larger numbers of learners, which has traditionally been a major roadblock for these highly-effective interventions.

STEM Careers: Artificial Intelligence Engineer, Computer Vision Engineer, Machine Learning Engineer, Robotics Engineer, Software Engineer, Analytics Consultant, Business Analyst, Data Scientist

ENGAGEMENT

Brief Description: By creating the perfect PBJ, students understand that AI algorithms can have various goals and motives.

Procedures:

Provide varying types of bread, varying brands and types of peanut butter, and varying brands and flavors of jelly/jam, paper plates, and plastic silverware.

Have students (without peer persuasion) create their own perfect PBJ sandwich. On the provided engage worksheet (below), have them record the ingredients and the steps they followed to create the perfect PBJ sandwich. The teacher will relate this to input data and algorithms where students learn that algorithms, like recipes, are a set of instructions that modify an input to produce an output. Students are then asked to write an algorithm to make the "best" peanut butter and jelly sandwich. Students then explore what it means to be "best" and see how their opinions are reflected in their algorithms.

Teacher Script (Engage:)

- Remember that an algorithm needs some input data and needs to follow specific steps or instructions to give us the desired output. Computers use algorithms, but so do humans. Algorithms are a lot like a recipe.

- Okay, now I want you to write your own algorithms. I want you to take the next 5-10 minutes to write an "algorithm" (or recipe) for the BEST peanut butter and jelly sandwich. Be sure to specify what your inputs are.
- Okay, now I want you to turn to your partner and share your algorithms. I want you to talk about what your algorithms have in common and how they are different. ALLOW TIME FOR CONVERSATION
- SHARE OUT:
    - Who can share with the class what their algorithms had in common?
    - What was different?
- Possible questions:
    - Did any of you include instructions to put away your ingredients after you used them?
    - Then you were optimizing for tidiness in your algorithm!
    - Did any of you cut your sandwich into fun shapes? Cut off the crust?
    - Then you were optimizing for playfulness or aesthetics!
- As you close out the engagement section, make sure you emphasize that personal opinion and preference define the BEST PBJ and that our own personal opinions about things can be inadvertently included in our writing of computer algorithms if care is not taken and awareness is not taught.

EXPLORATION

Brief Description: Students are introduced to the concept of classification. By exploring Google's Teachable Machine tool, students learn about supervised machine learning. (During the Explore phase, students will be asked to explore and familiarize themselves with the Teachable Machine website.) Then students are asked to build a cat-dog classifier but are unknowingly given a biased dataset. When the classifier works better on cats than dogs, students have the opportunity to retrain their classifiers with their own new datasets.

Procedures:

Teacher Script (Exploration):

- Today we are going to talk a little bit more about training datasets and dig into the "learning" algorithm aspect of AI a little more. Particularly, we're going to talk about one very common form of artificial intelligence called supervised machine learning.
- In a supervised machine learning system, a computer learns by example. How many of you have ever seen a parent teach a baby or toddler their colors? They'll often sit with their child in a board book with colored shapes or animals and point to the image and say "brown bear" or "red bird." And the idea is that a child will learn from these labeled examples.
- But today we're going to focus on classification, who can tell me some useful examples of classification that you use in your everyday life or that you've learned about in schools?
    - [Wait for students to respond. If need prompting - ask them if they have learned about any classification systems in science class such as animal classification. Ask them how they

find books in their school library (fiction/non-fiction, and then by genre). Song/movie genres are another good answer to hear. Wait for 2-3 answers from students.]

- Great, now I want to talk about a few examples of classification technology:
    - Face detection - is it a face or not
    - Spam detection - is it spam or not
    - Handwriting detection - to deposit checks at the bank - is it one of 26 characters or 10 numbers
- Let's build our own classifiers now. Let me demo Teachable Machines to you now. Take students to this website and demonstrate this user-friendly website or show the first video on the site.

    https://teachablemachine.withgoogle.com/

    *Be sure to note the difference between training data and test data in your demo by saying, "And now I want to test how well my algorithm does with data that is similar but slightly different to what it has seen before"
- Now you try! Follow the tutorial (Explore worksheet below) with your partner. Write what you find on the activity sheets. I'll walk around to see how everyone is doing. [give students time to each try the demo, ~5-8 minutes. Can go around and prompt students to tell you about their training and test datasets. What happens when they change their pose? What happens when both partners are in the frame?]
- Discuss:
    - What happens if you only train one class?
    - What happens as you increase your dataset?
    - What happens when your test dataset is different from your training dataset?

EXPLANATION

Brief Description: Students are introduced to the concept of classification. By exploring Google's Teachable Machine tool, students learn about supervised machine learning. Then students are asked to build a cat-dog classifier but are unknowingly given a biased dataset. (During the Explain phase, students will be building the classifier with teacher scaffolding .) When the classifier works better on cats than dogs, students have the opportunity to retrain their classifiers with their own new datasets.

Procedures:

Teacher Script (Explanation):

- Now we are going to build a cat-dog classifier using the new version of teachable machines.
- Be available for struggling students. With your partner, I want you to build a machine that classifies cats and dogs. (Data set below. Be sure to give them the correct set of data. Sets are clearly labeled at the top of first page). [give students time to build their classifiers and record info.]
- How are your classifiers working? Prompt:
    - Is this classifier useful if it only works well on just cats?

    - Why do you think it works better on cats vs dogs?
    - How could we make it better with our training data? (If students have trouble, ask them to notice similarities in the dataset, e.g., dogs were really fluffy and cat-like/not as diverse)
    - When algorithms, specifically artificial intelligence systems, have outcomes that are unfair in a systematic way, we call that algorithmic bias. We would say that our cat-dog classifier shows algorithmic bias and that it is biased towards cats since it works really well for them and biased against dogs since it doesn't work as well for them.
    - [give students time to re-curate their datasets]
- How are everyone's classifiers performing now? Prompt:
    - What did you do to make it work better?
    - If students say they used less training data, prompt them to think about if it is better to have more vs less data?
    - So we've seen firsthand how algorithmic bias can occur in our supervised machine learning systems. Now I want us to take some time to watch a video about how it can happen in the real world.
    - [play Gender Shades facial detection video]
- Discuss:
    - What problem did Joy identify in the video?
        - [wait for students to say facial recognition system did not recognize darker, female faces as well as lighter, male faces]
        - If students need prompting, ask if the technology Joy talked about worked the same for all people.
    - Why is this a problem?
        - [wait for students to mention unequal user experiences between groups - "technology doesn't work unless it works for everyone"]
        - If students need prompting, ask them if they would all like to be able to use Snapchat filters. Is it wrong if they did not all have access to that technology?
    - How does Joy suggest we can fix this problem?
        - [wait for students to mention better dataset curation]
        - If students need prompting, ask them how they improved their classification algorithms during the Teachable Machines activity.
    - Optional: if time permits and the classroom environment is safe
        - How might you find images to better curate your dataset?
        - [Prompt students to think about where photos exist - social media, mug shots, ID cards. Which of these would be okay to use as sources of information? Which might lead to more bias? Which might breach privacy?]

ELABORATION

Brief Description: Provide a concise overview of the proposed activity.

Procedures:

Teacher Script (Elaboration):

- During this phase of the lesson, the students will retrain their teachable machine with a data set (below) that does not promote bias.

EVALUATION

Brief Description: Students will summarize learning of bias and machine learning.

Procedures:

The student will create a visual story map at https://miro.com/templates/visual-story-map/ or similar site. The story map should summarize bias and who AI can also contain bias if the author of the AI does not include ways to exclude bias from the data sets.

REFERENCES

- https://teachablemachine.withgoogle.com/
- https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/

ENGAGE Activity

PB&J Sandwich Activity Sheet

Write an "algorithm" to make the BEST Peanut Butter and Jelly sandwich:

a) What *input data* (or ingredients!) do you need? (Be specific with brands and descriptions of ingredients)

b) Write out the steps in your algorithm:

EXPLORE Activity

Introduction to Supervised Machine Learning Activity

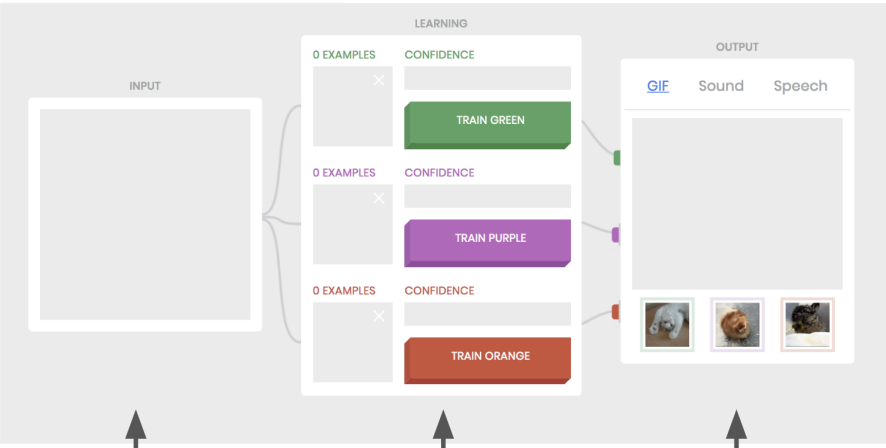
Description

In this exercise, you will learn about the three components of an artificial intelligence (AI) system. You will then learn about the role of training data in an AI system.

Instructions

1. Go to: https://teachablemachine.withgoogle.com/
2. Click "Let's go!"
3. Follow the tutorial.

Identify the three parts of an AI system in the teachable machine as discussed in class:



Hit refresh. This time click "skip the tutorial." Train the same classifier with your face and hands. What happens when:

(1) You only train one class?

(2) What happens when you increase the number of images in your dataset? Make sure both classes have at least ten images.

(3) If you've mainly been training with one hand up, try using the other hand. What happens when your test dataset is different from your training dataset?

DATA SETS FOR ACTIVITY

Initial Training Dataset

Dogs

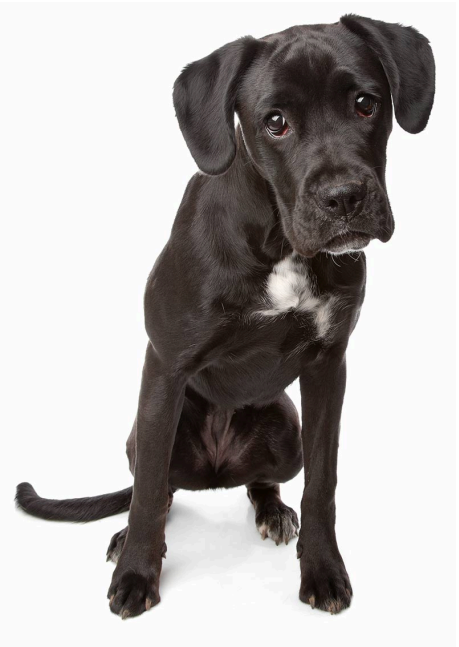














Cats





























Test Dataset










©Warren Photographic





Recurating Dataset

Dogs



Cats

