

Address Bias through Machine Learning

Companion Lesson to X-STEM All Access Episode “[Exploring Our Minds](#)”

Grade/ Grade Band: 9-12	Topic: Machine Learning	
Brief Lesson Description: Brain Engineer, Dr. Kafui "Kaf" Dzirasa is a psychiatrist and engineer who uses machine learning in his research to uncover changes in our brain circuits that lead to psychiatric illnesses. This lesson shows students what machine learning is, illustrates the constraints, and demonstrates how to design solutions for tackling them.		
Performance Expectation(s): NGSS HS-ETS1-2. Design a solution to a complex real-world problem by breaking it down into smaller, more manageable problems that can be solved through engineering. NGSS HS-ETS1-4. Use a computer simulation to model the impact of proposed solutions to a complex real-world problem with numerous criteria and constraints on interactions within and between systems relevant to the problem.		
Specific Learning Outcomes: 1. Define machine learning bias 2. Use computer simulation to design a solution 3. Describe solutions for addressing bias when using AI		
Narrative / Background Information		
Prior Student Knowledge: Students might get more from the lesson if they understand algorithms and can define the term bias. It would be helpful to students if they know how to upload digital images and how to use the computer’s webcam to take photographs.		
Science & Engineering Practices: Using Mathematics and Computational Thinking Mathematical and computational thinking in 9-12 builds on K-8 experiences and progresses to using algebraic thinking and analysis, a range of linear and nonlinear functions including trigonometric functions, exponentials and logarithms, and computational tools for statistical analysis to analyze, represent, and model data. Simple computational simulations are created and used based on mathematical models of basic assumptions. <ul style="list-style-type: none"> Use mathematical models and/or computer simulations to predict the effects of a design solution on systems and/or the interactions between systems. (HS-ETS1-4) 	Disciplinary Core Ideas: Developing Possible Solutions <ul style="list-style-type: none"> Both physical models and computers can be used in various ways to aid in the engineering design process. Computers are useful for a variety of purposes, such as running simulations to test different ways of solving a problem or to see which one is most efficient or economical; and in making a persuasive presentation to a client about how a given design will meet his or her needs. (HS-ETS1-4) Optimizing the Design Solution <ul style="list-style-type: none"> Criteria may need to be broken down into simpler ones that can be approached systematically, and decisions about the priority of certain criteria over others (trade-offs) may be needed. (HS-ETS1-2) 	Crosscutting Concepts: Influence of Science, Engineering, and Technology on Society and the Natural World <ul style="list-style-type: none"> New technologies can have deep impacts on society and the environment, including some that were not anticipated. Analysis of costs and benefits is a critical aspect of decisions about technology. (HS-ETS3-3)
Possible Preconceptions/Misconceptions: Students may believe machine learning can predict the future. This is partly true in that machine learning models can predict the future if future events have similarities with the past events.		
LESSON PLAN – 5-E Model		
ENGAGE: Opening Activity – Access Prior Learning / Stimulate Interest / Generate Questions: Students will watch Exploring Our Minds- Brain Engineer Dr. Kafui Dzirasa video. To keep students engaged with the video, have them focus on what Dr. Kaf does as a psychiatrist, what he does as an engineer, how he merges the two fields and how it benefits the world. Students can record notes in their notebooks, on post-its, or use back channeling tools like: Google Classroom’s “stream” feature that allows students to chat in real-time.		
After the video, discuss Dr. Kaf’s research and the reason why he uses machine learning in his brain study.		

EXPLORE: Lesson Description – Materials Needed / Probing or Clarifying Questions: Prior to students completing this section, the teacher should ensure all laptops can access the [Teachable Machine](#) and either ask students to bring in stuffed animals or create a collection of digital photos of stuffed animals.

Teachers say to students “Let’s explore AI with Teachable Machine. Go to the [Teachable Machine new project web page](#) and use the webcam to train the tool to recognize if you or a stuffed animal is looking at the camera”. If the laptops don’t have access to a webcam, the teacher or students can upload files of stuffed animal pictures. Students should watch the tutorials if they are stuck.

- a. The [Teachable Machine Tutorial 1 video](#) explains how to gather learning data. Go ahead and gather learning data.
- b. If your computer does not have a webcam, take digital pictures, and upload those to the tool instead. Then, watch the [Teachable Machine Tutorial 2 video](#) to learn how to train the model.
- c. The [Teachable Machine Tutorial 3 video](#) shows how to understand the predictions.

EXPLAIN: Concepts Explained and Vocabulary Defined:

What is machine learning? Unlike traditional computer programs where the decisions and rules are built in, machine learning programs construct their algorithm from data and feedback. This allows machine learning programs to find trends and patterns in enormous quantities of data, including patterns that are hard for humans to catch. They can also improve themselves without human intervention and can make predictions and handle complex, changing environments. But machine learning has its limitations. It requires a neutral and complete set of data to learn from and it uses a lot of computer power. The results need to be taken with some precaution as it is susceptible to systematic errors. Google’s Teachable Machine uses machine learning to classify sounds, pictures, or video and with your help with biases.

What are biases? Prejudice? Bias means a **strong inclination of the mind or a preconceived opinion about something or someone**. A bias may be favorable or unfavorable: bias in favor of or against an idea.

Vocabulary:

Artificial intelligence (AI)- a branch of computer science that tries to build tools that demonstrate intelligence

Machine learning- a sub-division of AI, its goal is to create tools that can learn and improve over time using data

Bias- cause to feel or show inclination or prejudice for or against someone or something

Confusion matrix- a table that is often used to describe the performance of a classification model on a set of test data for which the true values are known; it compares predicted data results to actual data results for what is true

ELABORATE: Applications and Extensions: This lesson is adapted from Science *Buddies Happy or Sad? Use Artificial Intelligence to Classify Faces* Project

Prior to starting the elaboration section of this lesson, you will need to gather the following materials (per student):

- 6 copies of [Face template](#)
- Construction paper
- Colored pencils, crayons, or markers
- Access to a computer with webcam (webcam is not necessary)

Begin this selection by asking students “what details do you think are essential when we draw examples to help the machine recognize a happy face or a sad face? Should you draw identical examples, or do you think it is better to have a variety of different happy faces in our example set? Should you include many details like hair style, earrings, freckles, etc., or should we focus on the features of the face that change when someone is happy or sad?”

Draw examples of happy faces in twenty circles (two sheets of the [template](#)) and sad faces in another set of twenty circles (two sheets of the template). Always use the same pencil or pen to draw your faces.

1. Cut along the lines and make a stack of happy and a stack of sad faces.
2. Use a digital camera like on your phone or scanner to capture the faces. You will be able to easily transfer the files to the computer:
 - a. For each drawing, place the face on a plain background like a sheet of colored construction paper, as shown in Figure 1, so the tool does not get distracted by the environment. Take a picture or scan. Try your best to take every picture the same way, laying the sad/happy face in the same spot, at the same angle, same distance, etc.
 - b. Transfer the files to the computer.

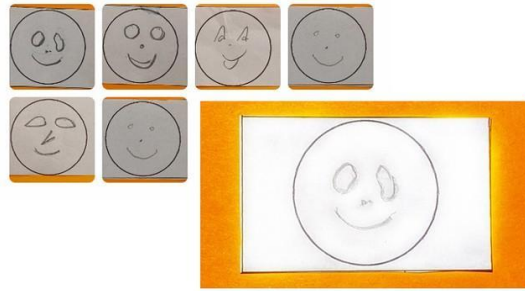


Figure 1. A drawing that is ready to be photographed (lower right corner).

(You can use the webcam to take pictures in Teachable Machine. This is explained in step 8.)

Creating Test Data

3. Repeat steps 1-2 but this time, ask friends and family members to draw faces for you. Collect at least five happy and five sad faces.
4. Repeat steps 1-2 but use coloring pencils, crayons, and markers to draw colorful faces. Some examples are shown in Figure 2. Create at least five happy and five sad faces.



Figure 2. Colorful drawings of happy and sad faces.

Preparing the Tool

5. Go to the [Teachable Machine new project web page](#) for recognizing images. If you decide to classify sounds or poses, go to the [new project page of Teachable Machine](#) instead and choose the project class that fits your goal best.
6. Label the classes. The tool will classify drawings of faces into two groups or classes: happy faces and sad faces. Rename class 1 as "Happy" and class 2 as "Sad."

Uploading Learning Data

7. If you already saved the learning data files to your computer in step 5:
 - a. Use the "Upload" button to upload the learning data files for happy faces to the "Happy" class, and the learning data files for the sad face to the "Sad" class.
8. If you will use the webcam in Teachable Machine to take pictures:
 - a. Start the webcam for the "Happy" class by clicking the "Webcam" button.
 - b. Place a happy face drawing on a plain background like a sheet of colored construction paper, so the machine does not get distracted by the environment.
 - c. Hold the drawing close to the lens, so the drawing fills most of the space.
 - d. Take all the pictures with the camera at the same angle so the lighting stays consistent.
 - e. Try to keep your fingers out of the picture. Teachable Machine has a function that crops the pictures as you are taking them. This can help you crop your fingers out of the pictures.

- f. Repeat step b–e for all happy drawings in your learning data. Try your best to take every picture in the same way.
- g. Once all happy faces of the learning data are uploaded, move on to the "Sad" class, repeat steps b–e for uploading the sad faces learning data. Try your best to take these pictures the same way as you took the happy face pictures.

You now have twenty pictures in the learning data for each class. If this is not the case, remove doubles or add pictures where needed. It is fine if the learning data shows some variation in lighting, distance, etc. It is important though that these variations occur in both classes. For example, it is not OK to have all the pictures in the "Happy" class to be close-ups and all the pictures of the "Sad" class to be taken from further away. But having about the same portion of close-ups and pictures taken from further away in both classes is fine.

Training the Model

9. Start the tool's learning process by clicking on the "Train Model" button. The machine will take less than a minute to complete this step. In that time, the tool searches patterns in the learning data it can use to distinguish between happy and sad drawings.

Saving the Tool

10. All the uploaded data on the machine can be saved to Google Drive and later uploaded from the Drive to continue. Look for the commands "Save project to Drive" and "Open project from Drive" under the "Teachable Machine" menu. It will ask you to log in to a Google account to access your Google Drive.

The project can also be saved as a file on your computer and uploaded later. Look for the commands "Download project as file" and "open project from file" under the "Teachable Machine" menu.

Testing the Tool

11. To see how the tool performs, you can use the webcam or files. Figure 3 shows how to toggle between the two.

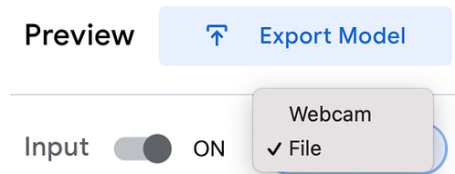


Figure 3. Teachable Machine allows you to use files you upload or the webcam to test the tool.

One by one, choose a few happy and a few sad drawings from the learning data, and test how the tool classifies these drawings. The bars under the test picture in the "output" box inform you how the tool classifies the drawing. Your tool will probably classify its learning data with high confidence, as shown in the left image of Figure 4 where we see the tool is certain the image belongs to the "Happy" class.

Sometimes, the tool is not clear about how to classify a drawing. The right image in Figure 4 is an example. The orange bar next to "Happy" shows that the tool classifies this drawing with a confidence of 59% in the "Happy" class. The red bar next to "Sad" shows it is 41% confident the drawing belongs to the "Sad" class.

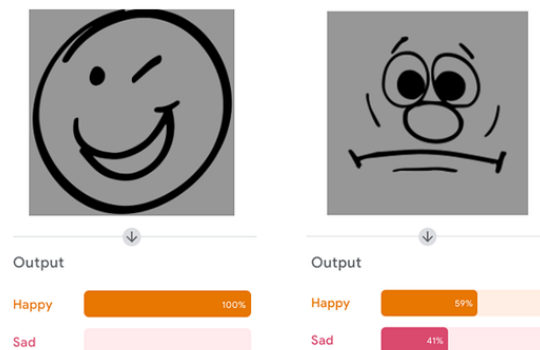


Figure 4. Two examples of an AI tool that learned to classify happy and sad face drawings.

In this experiment, you classify a drawing in the class for which the tool shows a confidence of over 50%. With this rule, the picture on the right in Figure 4 is classified under "Happy" because 59 is higher than 50. Because this drawing is in fact a sad face, the tool receives a mark in "Sad drawing misclassified as "Happy."

If you use the webcam, try your best to show the drawings the same way as you took pictures in step 8.

12. In your notebook, draw a confusion matrix like the one shown in Figure 5. Do not forget to fill in the amount of learning data you used in the title.

Confusion matrix
Tool trained on drawings per class.

	Classified as "Happy"	Classified as "Sad"
Happy drawing	Happy drawings that are correctly classified as "Happy"	Happy drawings that are misclassified as "Sad"
Sad drawing	Sad drawings that are misclassified as "Happy"	Sad drawings that are correctly classified as "Sad"

Figure 5. A confusion matrix helps organize the test results.

13. Test new data

The real test is seeing how well the tool can classify drawings it has never seen. Test on all the cases listed below. Always tally the results in your confusion matrix.

- Use five happy and five sad drawings of the pile you collected in step 4 (drawings made by other people).
- Test five happy and five sad drawings of the pile you drew in step 5 (more detailed and colorful drawings).
- Use the webcam or a camera to take more pictures to test what happens if you change the background. Use four drawings of the learning data for each class to perform this test.
- Use the webcam or a camera to take more pictures to test what happens if drawings are shown close-up or further away from the lens compared to the learning data. Use six drawings of the learning data for each class to perform this test.
- Test any other variations you want to test.
- Write down observations in your notebook. Does the tool perform well on a subset of the tests, or very poorly on a specific subset? Do you notice any trends?

EVALUATE:

Formative Monitoring (Questioning / Discussion): Confusion Matrix where students record results of the test data to see if they provided enough variety in their data points to overcome their bias

Summative Assessment (Quiz / Project / Report): Students write a report explaining which types of faces were initially recognized and what types of faces the program learned to recognize based on the data they used. The report should include data from the confusion matrix.

Elaborate Further / Reflect: Enrichment: Students can design a new project, teaching the Teachable Machine to classify poses, shapes, or sounds.

SOCIAL EMOTIONAL LEARNING ACTIVITY

SELF AWARENESS and SOCIAL AWARENESS

Understanding biases will help students develop both self and social awareness as well as responsible decision-making. Everyone has an invisible bias and this lesson shows students how to recognize theirs.

[Bias Mad-Lib](#) by Jacquelyn Whiting

Student Instructions: Working independently fill-in the blanks with words/short phrases that make sense in the context and make the paragraph cogent. There is no right answer as long as the paragraph makes sense to the reader. After completing the task, pair up with a classmate, each reads the paragraph they've written and discusses the task, agreeing on which of the words/short phrases should appear in this second draft.

The discrepancies reveal students have different points of view even if they live in similar neighborhoods, take multiple classes together, or have been classmates since 2nd grade. As the teacher you want to note that their choice of words indicates subtle and even unconscious beliefs and things that they thought were facts might actually be opinions.

The next step is to focus on this one sentence from the paragraph:

“They _____ to keep up, while others are _____ up on elevators.”

Ask students to share how they completed this sentence. (Possible ans.: they *struggle* to keep up, while others are *racing* up...; they *try* to keep up, while others are *rising* up...; they *attempt* to keep up, while others are *moving* up...) Ask students what is the difference implied by “*struggle*”, “*try*”, and “*attempt*”? What do these words say about the people doing the action? Difficulty, obstacles, or lack of will or commitment? Noting that these words imply a lack of agency. Do the same for the second words: “*racing*”, “*rising*”, and “*moving*”. Noting progress, accomplishment, and action, that the “others” have power and agency.

Finally share Ms. Carter, the original author’s, words:

“They work to keep up, while others are zooming up on elevators.”

Noting how the author’s word choices convey “they” have agency and skill.

Discuss with students how where they live, how they are brought up, their background, and exposure to media contributes to their world view and this is what makes up the biases people have.

INTERDISCIPLINARY CONNECTIONS/IDEAS

As students create a written report for this lesson, they are also developing their writing skills on the following standard(s): Write informative/explanatory texts to examine and convey complex ideas, concepts, and information clearly and accurately through the effective selection, organization, and analysis of content. CCSS.ELA-LITERACY.W.9-10.2

Materials Required for This Lesson/Activity	
Quantity	Description
Per student	Laptops
6 copies/student	Face template
	Construction paper
	Colored pencils/crayons/markers



Lesson Created by Stacy Douglas
For questions, please contact info@usasciencefestival.org